



Expression Analysis of *uS19* and its Association with
Nasopharyngeal Carcinoma

Cassandra Chee Sheau Mei

Faculty of Resource Science and Technology
Universiti Malaysia Sarawak
2020

Expression Analysis of *uS19* and its Association with Nasopharyngeal Carcinoma

Cassandra Chee Sheau Mei

A thesis submitted

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

(Molecular Biology)

Faculty of Resource Science and Technology

UNIVERSITI MALAYSIA SARAWAK

2020

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: Cassandra Chee Sheau Mei

Matric No.: 15020839

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Date :

ACKNOWLEDGEMENT

First of all, I would like to thank God for giving me the strength to complete this research to a huge success. When in doubt, He has supported me and gave me courage. Through Him anything is possible. I would also like to express gratitude and thanks to my supervisor, Prof. Dr. Edmund Sim Ui Hang who has been supervising and guiding me as well as sharing his knowledge and giving me advice which leads to a successful end. Without his help, I would have been lost therefore I would like to express my highest appreciation to him.

Furthermore, I would like to pay my highest gratitude to my fellow colleagues, Mrs. Ng Kher Lee, Ms. Felicia Kavita Thomas, Ms. Yew Keh Li and Ms. Stella Chan Li Li, whom have been so patient and consistently guiding me. It is always nice to know that I am not alone in this and with these many people giving me the emotional support I need whenever I am in doubt was the thing I needed to get through the day.

Last but not least, I would like to thank my mother, Ms. Choong Peck Yoon for her endless support and encouragement not only throughout the duration of this research but also throughout my entire life. Her moral support and trust towards me have always boosted me up when I am at my lowest. I would like to give a big shout out to thank my grandparents for supporting me and believing in me that I can do this. Their constant support and the way their face lit every time they ask about me, knowing that I am studying to pursue my dreams is the best feeling in the world. Thank you so much for following me through this journey.

ABSTRACT

Nasopharyngeal carcinoma (NPC) is currently the fourth most common cancer in Malaysia. Ribosomal protein genes are crucial in producing functional ribosomes that assist in cell functions and have shown to be associated in pathways that link to the formation of cancer. Previous studies have shown that *uS19* gene is involved in the Mdm2-p53-MdmX pathway, where it inhibits the p53, a tumor suppressor. Furthermore, expression of ribosomal proteins have been studied to observe its expression levels, in which potential biomarkers could be discovered that assist in the early diagnosis of cancers. Notably, uS19 is currently understudied, though studies have been done on the expression of *uS19* gene in NPC, where it was found to be upregulated in poorly differentiated NPC tissue samples and HK1 cell line. However, further validation is required. Studies on the post-translational effect of uS19 in NPC carcinogenesis are yet to be done. In this study, mRNA and protein expression of uS19 data obtained revealed a differential expression at both gene and protein levels in NPC cell lines compared to normal nasopharyngeal epithelial (NPE) cells. The gene expression of *uS19* was found to be down-regulated following a trend of cell type and differentiation, while protein expression of uS19 was found to be upregulated in all NPC cell lines, with C666-1 cell line being the highest. The data attained showed that uS19 might be involved in the NPC tumorigenesis based on the difference in transcription and translational level. Protein-protein interaction via *in silico* approach revealed uS19 could possibly bind with 10 proteins, which includes LMP2 and EBNA1, a known Epstein-Barr virus (EBV) protein. Our findings strengthen the potential role of uS19 in NPC tumorigenesis based on the expression data obtained and calculated interactions through bioinformatics analysis that

revealed a high possible chance of uS19 protein interacting with EBV proteins involved in the carcinogenesis of NPC.

Keywords: Nasopharyngeal carcinoma, uS19, expression, protein-protein interaction

Ekspresi Analisa uS19 dan Hubungannya dengan Karsinoma Nasofarink

ABSTRAK

Karsinoma nasofarinks (NPC) merupakan kanser keempat di Malaysia. Gen protein ribosom adalah penting dalam menghasilkan ribosom berfungsi untuk membantu dengan fungsi sel dan laluan yang membentuk tumor. Kajian terdahulu telah menunjukkan bahawa gen uS19 terlibat dalam laluan Mdm2-p53-MdmX, di mana ia menghalangkan p53, penindas tumor. Tambahan pula, ekspresi protein ribosom telah dikaji untuk memperolehi potensi biomarker yang dapat membantu dalam prognosis awal kanser. Walaubagaimanapun, uS19 di NPC pada masa kini belum lagi difahami, tetapi kajian telah dilakukan pada ekspresi gen uS19 dalam sampel tisu dan sel NPC, di mana ia menunjukkan ekspresi yang tinggi, bagaimanapun pengesahan selanjutnya diperlukan. Kajian mengenai kesan pasca translasi dari uS19 dalam perbentukan kanser NPC masih belum dilakukan. Dalam kajian ini, ekspresi gen dan protein telah dilakukan di mana perbezaan dalam segi ekspresi di antara gen dan protein dalam sel NPC berbanding dengan sel normal berbeza antara satu sama lain. Ekspresi gen uS19 didapati mengikuti trend menurun dalam perbezaan ciri-ciri sel NPC, sementara ekspresi protein uS19 menunjukkan ekspresi yang tinggi dalam semua sel NPC dibandingkan dengan sel normal dengan ekspresi uS19, di mana sel C666-1 mempunyai ekspresi paling tinggi. Data yang diperoleh menunjukkan kemungkinan penglibatan uS19 dalam perbentukan kanser NPC berdasarkan perbezaan transkripsi dan tahap translasi. Interaksi protein-protein melalui kaedah in-silico mendedahkan bahawa uS19 didapati untuk mungkin berinteraksi dengan 10 protein, termasuk LMP2 and EBNA1, protein Epstein-Barr virus (EBV). Penemuan kami

menguatkan potensi peranan uS19 dalam pembentukan kanser NPC berdasarkan data eksperimen yang diperoleh dan analisa bioinformatic, yang menunjukkan peluang yang tinggi untuk uS19 berinteraksi dengan protein EBV yang terlibat dalam karsinogenesis NPC.

Kata kunci: *karsinoma nasofarinks, uS19, interaksi protein-protein*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	v
TABLE OF CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xiii
LIST OF ABBREVIATION	xv
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	4
2.1 Nasopharyngeal Carcinoma	4
2.1.1 Histopathology of Nasopharyngeal Carcinoma	5
2.1.2 Nasopharyngeal Carcinoma Cell Lines	7
2.1.3 Epidemiology of Nasopharyngeal Carcinoma	10
2.1.4 Etiology of Nasopharyngeal Carcinoma	11
2.1.5 Clinical Prevention and Treatment of Nasopharyngeal Cancer	14
2.2 Ribosomal Protein	15
2.2.1 Ribosomal Protein in Cancer	16
2.2.2 uS19	18
2.3 Protein Interaction Prediction through Bioinformatics Analysis	20
2.3.1 SWISS-MODEL	22

2.3.2 Vector Alignment Search Tool (VAST)	23
2.3.3 IntAct	24
2.3.4 ClusPro	25
2.3.5 PatchDock and FireDock	26
CHAPTER 3: MATERIALS AND METHODS	28
3.1 Cell Culture	28
3.2 Gene Expression	29
3.2.1 Total RNA Extraction	29
3.2.2 cDNA Synthesis	30
3.2.3 Real-time PCR	31
3.3 Protein Expression	33
3.3.1 Total Protein Extraction	33
3.3.2 SDS-PAGE	34
3.3.3 Western Blot	35
3.4 Bioinformatics Analysis	36
3.4.1 3D Protein Model Construction	36
3.4.2 Identification of Protein Interactors	37
3.4.3 Docking of Protein Partners	37
CHAPTER 4: EXPRESSION ANALYSIS OF <i>uS19</i>	38
4.1 Introduction	38
4.1 Results	38
4.2 Discussion	49
CHAPTER 5: BIOINFORMATICS ANALYSIS	55
5.1 Introduction	55

5.2 Results	56
5.3 Discussion	96
CHAPTER 6: GENERAL DISCUSSION	103
CHAPTER 7: CONCLUSIONS AND FUTURE WORK	107
REFERENCES	108
APPENDIX	126

LIST OF TABLES

	Page
Table 2.1 Long-term tissue cultures derived from NPC tumours	8
Table 3.1 Passage number of cell line used	29
Table 3.2 Primer sequences for <i>GAPDH</i> and <i>uS19</i>	31
Table 4.1 ΔC_T calculation for validation experiment	39
Table 4.2 PCR efficiency, R^2 , and m-value of target and reference gene	39
Table 4.3 Normalized expression of <i>uS19</i> using the $\Delta\Delta C_T$ method	40
Table 4.4 Average fold difference of <i>uS19</i> in NPC relative to NP69	42
Table 4.5 Cumulative fold difference of <i>uS19</i> in NPC versus NP69	43
Table 4.6 Average band intensity of beta actin and uS19 protein	46
Table 4.7 Normalized band intensities of uS19 to beta actin protein	47
Table 4.8 Fold change of uS19 protein in NPC cell lines	47
Table 4.9 Average fold change of uS19 in NPC cell lines	47
Table 4.10 Cumulative fold change of uS19 in NPC compared to NP69	48
Table 5.1 Quality estimations of uS19 model	57
Table 5.2 Predicted interaction partners with uS19 protein.	58
Table 5.3 Quality estimates for selected protein models interacting with uS19 protein	60
Table 5.4 ClusPro and PatchDock/FireDock energy profile and RMSD values	62
Table 5.5 Probable interface residues of uS19 and TMEM237 protein	64
Table 5.6 Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-TMEM237 docked complex	64

Table 5.7	Possible interface residues of uS19 and SLC22A23 protein	66
Table 5.8	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- SLC22A23 docked complex	70
Table 5.9	Possible interface residues of uS19 and SLC13A4 protein	72
Table 5.10	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- SLC13A4 docked complex	74
Table 5.11	Possible interface residues of uS19 and HSPA8 protein	78
Table 5.12	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- HSPA8 docked complex	79
Table 5.13	Possible interface residues of uS19 and CUBN protein	81
Table 5.14	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- CUBN docked complex	82
Table 5.15	Possible interface residues of uS19 and Cobalamin protein	84
Table 5.16	Predicted interfacial residues involved in hydrophobic interactions within the uS19- Cobalamin docked complex	84
Table 5.17	Possible interface residues of uS19 and SLC7A1 protein	86
Table 5.18	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-SLC7A1 docked complex	86
Table 5.19	Possible interface residues of uS19 and RRP1B protein	88
Table 5.20	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-RRP1B docked complex	89
Table 5.21	Possible interface residues of uS19 and LMP1 protein	91
Table 5.22	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-LMP1 docked complex	92

Table 5.23	Possible interface residues of uS19 and EBNA1 protein	93
Table 5.24	Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-EBNA1 docked complex	96

LIST OF FIGURES

	Page
Figure 2.1 Anatomy of the pharynx	4
Figure 2.2 Histopathology of Type I NPC	5
Figure 2.3 Histopathology of Type II NPC	6
Figure 2.4 Location of <i>uS19</i> in S19 chromosome	19
Figure 4.1 Validation plot of ΔC_t (crossing threshold) versus log input amount of RNA	39
Figure 4.2 Relative fold difference of <i>uS19</i> in NPC cell lines versus NPE, NP69	43
Figure 4.3 Cumulative fold difference of <i>uS19</i> in overall NPC cell lines versus NPE, NP69	44
Figure 4.4 Western blot analysis of uS19 protein expression in NPC cell lines	45
Figure 4.5 Relative fold change of uS19 in NPC cell lines versus NPE	48
Figure 4.6 Cumulative fold difference of uS19 protein in overall NPC cell lines compared to NPE, NP69	49
Figure 5.1 Ribbon representation of uS19 protein	57
Figure 5.2 Ribbon representation of the five predicted proteins	59
Figure 5.3 Ribbon representation of uS19-TMEM237 docked complexes	63
Figure 5.4 Ribbon representation of uS19-SLC22A23 docked complexes	66
Figure 5.5 Ribbon representation of uS19-SLC13A4 docked complexes	71
Figure 5.6 Ribbon representation of uS19-HSPA8 docked complexes	77
Figure 5.7 Ribbon representation of uS19-CUBN docked complexes	80

Figure 5.8	Ribbon representation of uS19-Cobalamin docked complexes	83
Figure 5.9	Ribbon representation of uS19-SLC7A1 docked complexes	85
Figure 5.10	Ribbon representation of uS19-RRP1B docked complexes	87
Figure 5.11	Ribbon representation of uS19-LMP1 docked complexes	90
Figure 5.12	Ribbon representation of uS19-EBNA1 docked complexes	93
Figure 5.13	Schematic diagram of uS19 interaction with predicted protein	103

LIST OF ABBREVIATION

AJCC	American Joint Committee on Cancer
APS	Ammonium persulfate
ASR(W)	Age-world standardized incidence rate
BLAST	Basic local alignment search tool
BCL-2	B-cell lymphoma 2
CLL	Chronic lymphocytic leukemia
CRT	Chemoradiation therapy
DARS	Decoys as the Reference State
D-KSFM	Defined-keratinocyte serum-free medium
DNA	Deoxyribonucleic acid
EBNA1	Epstein-Barr nuclear antigen 1
EBNA-LP	Epstein-Barr nuclear antigen leader protein
EBV	Epstein-Barr Virus
ECL	Enhanced chemiluminescent
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
FBS	Fetal bovine serum
FFT	Fast fourier transform
FMRP1	Fragile X Mental Retardation Protein 1
HRP	Horse-radish peroxide
IMRT	Intensive-modulated radiation therapy

LMP1	Latent membrane protein 1
LMP2	Latent membrane protein 2
LRRK2	Leucine-rich repeat kinase 2
M-MLV-RT	Moloney Murine Leukemia Virus Reverse Transcriptase
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
NaCl	Sodium Chloride
NCBI	National Centre for Biotechnology Information
NPC	Nasopharyngeal carcinoma
NPE	Nasopharyngeal epithelial
PML	Promyelocytic leukemia
PMSF	Phenylmethylsulfonyl fluoride
PPI	Protein-protein interaction
PSI-BLAST	Position-Specific Iterated BLAST
PVDF	Polyvinylidene difluoride
QMEAN	Qualitative model energy analysis
RIPA	Radioimmunoprecipitation assay
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RP	Ribosomal proteins
RPS	Ribosomal protein subunit
RT-PCR	Reverse-transcription Polymerase Chain Reaction
rRNA	Ribosomal ribonucleic acid
SCC	Squamous cell carcinoma

SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TBS	Tris-Buffered Saline
TBST	Tris-Buffered Saline Tween 20
TEMED	Tetramethylethylenediamine
UV	Ultraviolet
V	Voltage
VAST	Vector alignment search tool
VEGF	Vascular endothelial growth factor
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

Cancers are progressive diseases that worsen and spread due to the accumulation of defected genes (Devereux et al., 1998). Early diagnosis on cancer diseases is crucial thus research regarding the possible causes of oncogenesis is looked into to provide an insight into the development of cancers, which could reduce the fatality of this disease. Recent studies have associated ribosomal protein (RP) genes with oncogenesis and tumour progression in humans through the activation or silencing of tumour suppressors, though the exact molecular mechanism remains a challenge. Genes in the ribosomal protein subunit (RPS) family gives instructions for producing ribosomal proteins used in protein synthesis and possesses extra-ribosomal functions such as cell proliferation and apoptosis. Studies revealed that changes in the RP expression levels are involved and regulates an extensive variety of activities correlating with cell growth and death and has been known to be implicated in carcinogenesis such as colon cancer, breast cancer, and nasopharyngeal carcinoma. In Malaysia, nasopharyngeal carcinoma (NPC) is currently the fourth most common cancer and is known to have an unbalanced geographical distribution, which makes it hard to diagnose especially due to its subtle symptoms. To this date, early diagnosis of NPC cases is hard to be detected causing the mortality rate for NPC patients to be high. Nasopharyngeal cancer is a type of cancer that starts at the nasopharynx, which is the upper part of the throat behind the nose, near the base of the skull. It is a malignant tumour that begins in the epithelial nasopharyngeal cells. Poorly differentiated squamous cell carcinoma (SCC) cell line is common in NPC patients in South China and has a high level of NPC incidence and mortality reported, according to Wei et al. (2014). The factors known to cause

NPC incidence are mainly Epstein-Barr virus infection (EBV), gender, age, race, geographical, environment and diet.

Though EBV is one of the main suspects in NPC carcinogenesis, current studies have suggested that RP genes are candidates to cancer-causing genes as it could be involved in certain pathways that coincides with EBV leading to NPC. Multiple studies have looked into the expression patterns of RP genes to obtain a further understanding regarding the molecular mechanism of how RP genes can be linked to NPC formation. Findings by Fang et al. (2008) through microarray analysis revealed an up-regulation of *uS19* gene, which was previously known as *RPS15*, in the poorly differentiated SCC of the NPC primary tumour when compared to normal nasopharyngeal tissue. Expression of *uS19* gene was also found to be upregulated in HK1 cell line in one of our previous studies (Sim et al., 2018). Furthermore, the *uS19* gene is known to be involved in regulating the Mdm2-p53-Mdmx network that inhibits the p53, which is a tumour suppressor in cells, from allowing the cell to undergo cell arrest or apoptosis, therefore, allowing the proliferation of the cells (Daftuar et al., 2013). This gene also exhibits expression patterns across the liver cancer samples (Wong et al., 2014) illustrating that *uS19* gene among many other genes has extra-ribosomal functions and is vital in the biological processes in humans but also can be highly active in cancerous and tumour cells. Furthermore, p53 has been known to be involved in the apoptotic pathway of NPC (Tulalamba & Janvilisri, 2012).

Although the *uS19* gene possesses a substantial case to be further looked into past transcription, however, no studies have been done on the expression pattern of uS19 protein in NPC cell lines compared to normal nasopharyngeal epithelial (NPE) cell lines. Hence,

whether the expression level of protein and mRNA are similar to one another or different is unknown. This could play a role in understanding whether the expression level of *uS19* gene during the transcription could affect the translation level of uS19 protein and lead to tumorigenesis. Moreover, identifying protein partners that associate with uS19 protein could lead to an understanding of the molecular pathway involved in the formation of NPC. Hence, we hypothesised that (1) *uS19* will be differentially expressed in NPC compared to NPE in both transcript and translation levels, (2) uS19 could play a role in NPC progression, and (3) uS19 might be interacting with EBV proteins and possibly pathways in cancer survival. The present study aimed to compare the transcription and translation level of *uS19* gene from both normal nasopharyngeal cell line, which is NP69, and six different NPC cell lines, which are HK1, HONE-1, SUNE-1, TW 01, TW 04, and C666-1, and to identify the protein-protein interaction between uS19 and its candidate protein partners through structural neighbouring and docking simulation.

The objectives of this study were:

- 1) To detect and quantify the transcription level of *uS19* gene in NPC cell lines compared to normal nasopharyngeal cell line
- 2) To detect and quantify the translation level of uS19 protein in NPC cell lines compared to normal nasopharyngeal cell line
- 3) To identify candidate protein partners associated with uS19 in NPC formation via structural neighbouring approach and docking simulation.

CHAPTER 2

LITERATURE REVIEW

2.1 Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma (NPC) is a rare malignant tumour that begins in the epithelial nasopharyngeal cells. It occurs in the nasopharynx located in the upper part of the throat behind the nose and near the base of the skull (Figure 2.1). Early diagnosis of NPC is rarely obtained due to its hidden location and subtle signs and symptoms. Common diagnosis to test NPC is physical examination using nasal endoscopy, blood testing for Epstein-Barr virus (EBV) and biopsy to observe the histopathology of the cells.

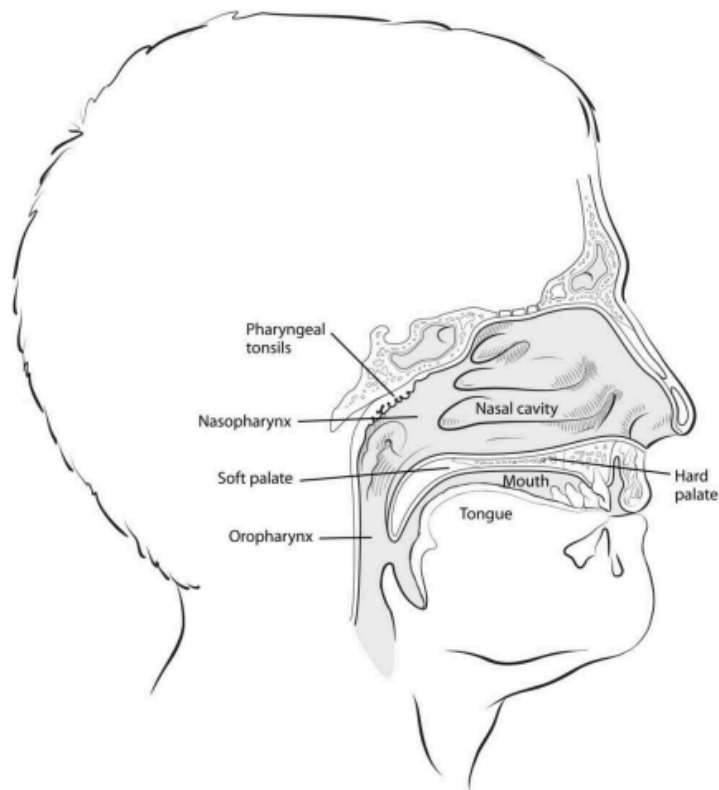


Figure 2.1: Anatomy of the pharynx. [Adapted from Nasopharyngeal cancer. (2013). *American Cancer Society*. Retrieved from <http://www.cancer.org/acs/groups/cid/documents/webcontent/003124-pdf.pdf>]

2.1.1 Histopathology of Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma (NPC) is a non-lymphomatous, squamous cell carcinoma (SCC) of an undifferentiated type. Histopathology studies on NPC are crucial in distinguishing prognosis and treatment. According to the World Health Organization (WHO), NPC is subcategorized into keratinizing carcinoma (Type I) and non-keratinizing carcinoma (Type II). Type 1, which is a keratinizing, well-differentiated SCC, is seen in 25% of NPC cases (Gullo et al., 2008). Type 1 NPC shows keratinization of the cell which is differentiated and possesses intracellular bridges. In this study, TW 01 cell line is used as a representative of Type 1 NPC (Gullo et al., 2008).

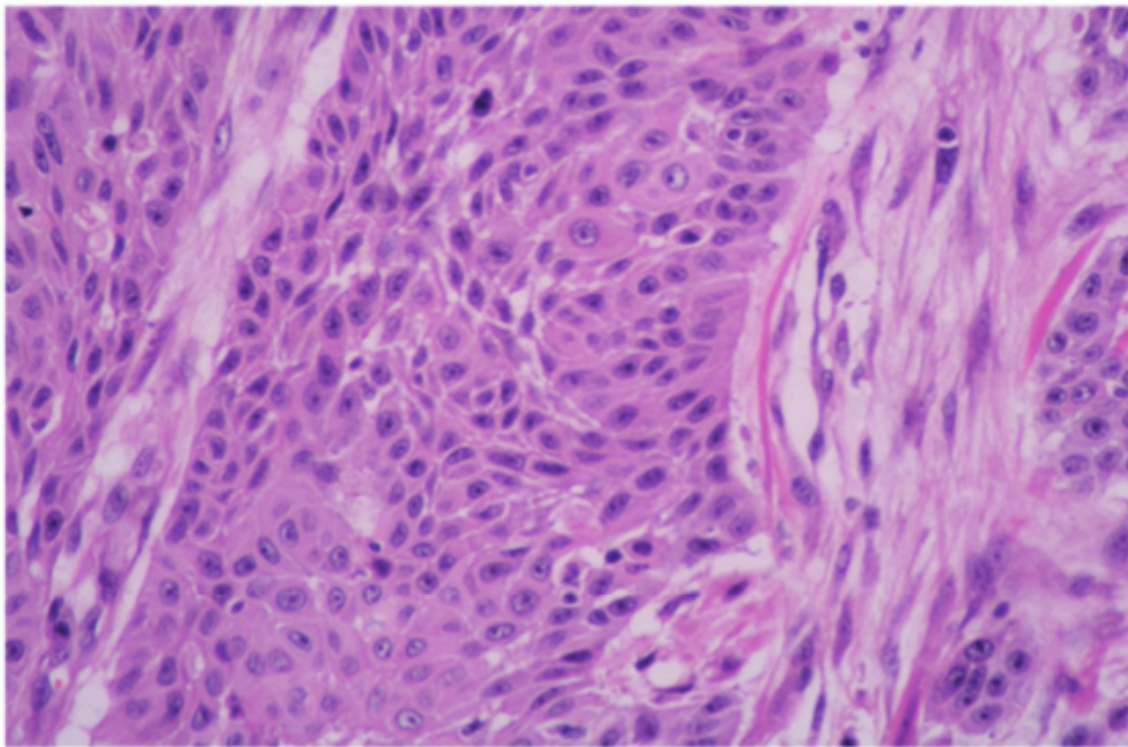


Figure 2.2: Histopathology of Type I NPC. Cells display individual cell keratinization and intracellular bridges Adopted from <https://ars.els-cdn.com/content/image/1-s2.0-S2090074015000055-gr2.jpg> (Lahuri et al., 2015).

Type II NPC, which is a non-keratinizing SCC, is subtyped into differentiated and undifferentiated carcinoma. Type IIa is a non-keratinizing, differentiated carcinoma and constitutes approximately 20% of NPC cases (Gullo et al., 2008). The cell exhibits obvious cell differentiation along with well-defined cell borders. In this study, HK1, HONE-1, and SUNE-1 cell lines are used as a representative for Type IIa NPC. HK1 cell line is more differentiated compared to HONE-1 and SUNE-1 cell lines (Gullo et al., 2008). Type IIb is a non-keratinizing, undifferentiated carcinoma, which represents 55% of NPC cases and is the most common NPC cases (Gullo et al., 2008). Type IIb NPC is commonly found to be associated with Epstein-Barr virus (EBV). The histopathology for this type is similar to Type IIa NPC but in a more scattered manner. In this study, TW 04 and C666-1 cell lines are used as a representative for Type IIb NPC (Gullo et al., 2008). C666-1 cell line is an EBV infected cell line.

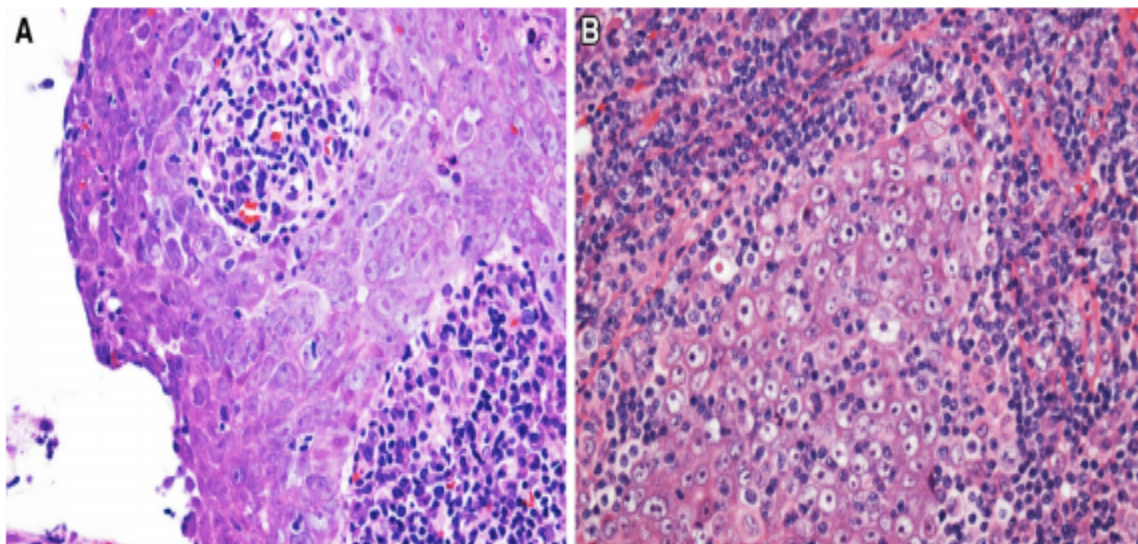


Figure 2.3: Histopathology of Type II NPC. (A) depicts the Type IIa and (B) Type IIb based on the clear differentiation of the epithelial. Adopted from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5340728/figure/Fig1/> (Stelow and Wenig, 2017).

However, due to NPC's hidden location and various cell characteristics, the prognosis for NPC patients can be challenging. Nevertheless, histopathology of the biopsy samples allows distinguishing the cell's classifications that could assist further research on its etiology and aid in the response to certain clinical treatments.

2.1.2 Nasopharyngeal Carcinoma Cell Lines

Nasopharyngeal carcinoma (NPC) cell lines that are currently used are mostly derived from primary tumours tissue samples obtained from the nasopharynx. However, there are a few cell lines which are derived from metastatic NPC tissue from other sites, commonly the colon and lymph nodes (Lahuri et al., 2015). For example, the NPC-BM1 cell line, which was established from a bone marrow biopsy sample. Also, C17 cell line from a skin biopsy sample, and C18 cell line from a cervical lymph node biopsy. Epstein-Barr virus (EBV) have been commonly known to be a trigger to NPC formation. However, most of the cell lines which were EBV-positive when first established, loses their EBV-positivity after long-term culture, while some cell lines do not contain the EBV genome. This could be an indication that EBV may initiate the pathogenesis of NPC but is not necessary for the survival and proliferation of the cell *in vitro*. However, some cell lines (for example, CAO, X666 and C666-1 cell lines) have been shown to maintain its EBV-positivity after long-term culture (Gullo et al., 2008). A list compiled by Gullo et al. (2008) illustrates 33 NPC-derived cell lines that have been established and classified according to its origin, NPC WHO classification, and EBV status (Table 2.1).

Table 2.1: Long-term tissue cultures derived from NPC tumours (Gullo et al., 2008).

Cell Line	Origin of cell line	NPC WHO classification	EBV status
C-15	Epithelial	Type II	Negative
C-17	Epithelial (from skin)	Type II	Negative
C-18	Epithelial (from lymph node)	Type II	Negative
X666	Epithelial	Type IIb	Positive
C666-1	Epithelial	Type IIb	Positive
CG1	Epithelial	Type II	Negative
CNE-2	Epithelial	Type IIb	Negative
CNE-2Z	Epithelial	Type IIb	Negative
CNE-3	Epithelial	Type IIb	Negative
HNE-1	Epithelial	Type IIb	Negative
HONE-1	Epithelial	Type IIa	Negative
NPC-TW 03	Epithelial	Type IIb	Negative
NPC-TW 06	Epithelial	Type IIb	Negative
NPC-TW 07	Epithelial	Type IIb	Negative
SUNE	Epithelial	Type IIb	Negative
SUNE-1	Epithelial	Type IIa	Unknown
NPC-TW 04	Epithelial	Type IIb	Negative
NPC-TW 09	Epithelial	Type IIb	Negative
CNE	Epithelial	Type I	Negative
NPC-HK1	Epithelial	Type IIa	Negative
NPC-TW 02	Epithelial	Type I	Negative
NPC-TW 08	Epithelial	Type IIb	Negative
NPC-TW 01	Epithelial	Type I	Negative
NPC-TW 05	Epithelial	Type IIb	Negative
NPC69SV40T	Epithelial	Type I	Positive
NP39E6/E7	Epithelial	Type I	Positive
NPC-BM1	Epithelial (from bone marrow)	Type IIb	Unknown
NPC-F/L	Epithelial	Type IIb	Positive

Type I NPC used in this study is represented by TW 01 and HK1 cell line. TW 01 is established from a biopsy sample of the nasopharynx of a 64-year old Chinese male from Taiwan and is a keratinizing squamous cell carcinoma in nature. In addition, it depicts a well-differentiated phenotype as well (Lin et al., 1990). HK1 cell lines is established from a biopsy sample from a recurrent squamous carcinoma of the nasopharynx of a 58-year old Chinese male living in Southern China, who relapsed after 17.5 years since his last radiotherapy (Huang et al., 1980). Type IIa NPC is represented by 2 NPC cell lines. The first is HONE-1, established from a biopsy sample of a poorly differentiated SCC of the nasopharynx of a patient living in Southern China (Glaser et al., 1989). The second cell line is SUNE-1, established from a biopsy sample of the nasopharynx of a 38-year old Cantonese female with Stage III squamous cell carcinoma (Dong et al., 2012). Type IIb NPC is represented by 2 NPC cell lines. The first is TW 04 cell line, which was established from biopsy specimens of the nasopharynx of a 36-year old Chinese female living in Hong Kong (Lin et al., 1993). The second cell line is C666-1, established from Xeno-666 at passage 18, which derived from an NPC xenograft of an undifferentiated NPC tumour biopsy from a Southern Chinese patient (Cheung et al., 1999).

The normal nasopharyngeal epithelial (NPE) cell (NP69) is an immortal NPE cell line that serves as a standard premalignant nasopharyngeal epithelial (NPE). In this study, the NP69 cell line acts as a negative control. It was established from a biopsy sample of patients with symptoms such as nasal obstruction and excessive bleeding. The primary culture of NPE cells was then immortalized by virally-transfecting them with SV40T antigen (Tsao et al., 2002). Studies by Tsao et al. (2002) suggests that NP69 cell line provides *in vitro* models

for studying tumorigenesis of NPC as they maintained characteristics of primary nasopharyngeal epithelial cells such as keratin profile and are non-tumorigenic.

2.1.3 Epidemiology of Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma (NPC) is the fourth most common cancer among adults in Malaysia after breast, colorectal, and lung cancer. NPC has shown to have an unbalanced geographical distribution where it is mostly occurring in Asia, especially in the Southeast Asian countries, which accounts for 71% of the global burden of cancer (Mahdavifar et al., 2016). In Asia, Malaysia has the highest age-world standardized incidence rate (ASR(W)) of nasopharynx cancer with a standardized rate of 7.2 per 100,000 people followed by Singapore, Indonesia, Vietnam and Brunei (Mahdavifar et al., 2016). In Peninsular Malaysia, the Chinese population has been shown to be at a higher risk of NPC (Tiong & Selva, 2005). In Sarawak, the Bidayuh race has the highest incidence of NPC as reported by Devi et al. (2004). However, recent findings by Aziz et al. (2017) has shown that young NPC incidence cases between the age of 16 and 20 years is predominantly high in the Malay population. Due to its hidden and subtle nature of its symptoms, presentation for medical treatments is usually late and at an advanced stage of the disease.

Nasopharyngeal carcinoma occurs at any age. However, according to the Mahdavifar et al. (2016), the peak age of incidence is between 50 and 60 years with a high risk of relapse and low survival rate in elderly people. Aziz et al. (2017) showed that NPC risk by age is divided into two peaks. Where the first peak is in late adolescence or early adulthood, roughly around 15 to 24 years of age followed by a second peak later in life, roughly around 65 to 79 years of age. Furthermore, gender distribution seems to play a role in NPC development as the

occurrence of NPC in male is two to three times more likely than in females. According to the findings by Mahdavi et al. (2016), the ASR(W) in Malaysian men is 10.6 as compared to women which are 3.9.

Genetically, NPC has been known to be associated with Epstein-Barr virus (EBV), which has been known to be involved with many different ribosomal proteins in pathways that result in the formation of NPC. However, EBV infections are nearly ubiquitous and NPC development is rare. It is commonly known that EBV infections are not sufficient to induce cancer, implying that other cofactors play a significant role in NPC pathogenesis (Hildesheim & Wang, 2012). Many other factors play a role in NPC pathogenesis apart from EBV infection, including exposure to carcinogens, dietary, genetic factors and many others. Understanding its etiology allows a much better understanding of the progression of cancer which could assist in the improvement of the prognosis of NPC.

2.1.4 Etiology of Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma (NPC) is presented as a complex disease caused by an interaction with the Epstein-Barr virus (EBV) infection and environmental factors and genetic susceptibility that involves a multistep carcinogenic process (Tabuchi et al., 2011). The latest findings have shown that EBV infection alone is not able to sustain NPC proliferation which indicates that other non-viral factors could be involved in maintaining cell proliferation. Generally, when suspected with NPC, EBV detection through blood samples is a standard procedure. EBV has been well known to have caused NPC especially in undifferentiated carcinoma (Type IIb), which is the most common NPC. Multiple studies have shown a strong association between NPC and EBV infection, where the EBV genes are found in the tumour.

EBV latent genes, example EBNA1, and LMP2 proteins have been shown to have an essential role in the replication and segregation of the EBV episomes and maintaining its survival rate (Young & Dawson, 2014). EBNA1 protein was found to protect the infected cells from apoptosis and directly contribute to the tumorigenic phenotype by destabilizing p53 (a tumour suppressor), disruption of promyelocytic leukaemia (PML) nuclear bodies, and modulation of various signalling pathways (Young & Dawson, 2014). LMP2 gene encodes 2 proteins which are LMP2A and LMP2B and they both play a role in resistance to the antiviral effects of interferons in the epithelial cells (Young & Dawson, 2014). EBNA1 and LMP2 proteins expression are reported to be consistent in NPC, mainly in the non-keratinizing carcinoma, especially Type IIb. Other factors have been discovered to induce or trigger EBV infection as in some NPC types, the EBV alone is not able to maintain cell survival.

Naturally, people with a family history of cancer, in particular, a family history of NPC, they are 4-10 times higher at risk to NPC (Salehiniya et al., 2018). Multiple studies have been conducted that showed the role of genetic combination with environmental factors contributes to NPC development. Based on the studies done by Fung et al. (2000) on gene expression between malignant NPE cells and non-malignant NPE cells, the development of NPC is likely to involve multiple genetic events including activation of oncogenes and inactivation of tumour suppressor genes. Genetic polymorphism in NPC have been studied in depth due to its strong association in NPC progression. Studies by Wee et al. (2010) showed that NPC is caused by genetic inheritance. This was observed when they looked into the number of cancer incidence among Southern Chinese and found similarities with Bai Yue people, that was proposed to be the origin of NPC. Furthermore, they identified the

association of the X-linked recessive mutation with EBV infections causing the HLA inability to detect the defected cell, resulting in an uncontrollable proliferation of EBV-infected cells (Wee et al., 2010). Hsu et al. (2011) further looked into the familial aspect of NPC and found that common genetics and/or factors (EBV, environment) plays a role in NPC susceptibility. However it is not necessary that NPC cases are mediated by factors alone. Identification of genes involved in NPC has been looked into extensively in order to understand its mechanism for the purpose of identifying biomarkers for NPC diagnosis and prognosis . Regardless, ethnicity appear to be an alarming factor due to its high incident of NPC in specific ethnic groups in specific geographical distribution indicating that genetic susceptibility plays a major role in NPC. For example, in Malaysia most NPC cases are reported to be from the Chinese, Malays and Bidayuh groups which could be due to their traditional dietary consumption, which is mostly preserved food.

Dietary consumption of preserved fish and other preserved food, is a common diet in several NPC-endemic areas like in Southern China and Malaysia. Multiple studies have been done to establish that consumption of salt-preserved fish since childhood can lead to the development of NPC (Huang et al., 1978; Yu et al., 1989). Moreover, the way the salt-preserved fish is prepared for consumption compared to consuming the fish by frying, grilling or boiling the salted fish (Zheng et al., 1994), increases the risk of NPC. Preserved food contains volatile nitrosamine that is a known carcinogen for NPC development. The active carcinogens metabolite of these nitrosamines might induce DNA damage and chronic inflammation in nasopharyngeal mucosa of genetically susceptible individuals (Tsao et al., 2014). Furthermore, the use of traditional herbal medicine has been reported to be associated with the elevated NPC risk in Asian population. This is due to the usage of Chinese herbal

plants, which induces the EBV lytic antigen expression (Chang & Adami, 2019). Studies by Hildesheim et al. (1992) confirmed the association of anti-EBV antibodies and herbal medicines with NPC in the Philippines. However further validation was not done further to be substantial enough to support the theory.

Non-dietary risk factors associated with increased NPC risk has also been reported, including occupational exposure and unhealthy habits. Exposure to formaldehyde, wood dust, fumes and chemicals have been recognized as NPC risk factors by causing chronic inflammation in the nasopharynx. Cigarette smoking has been reported as an NPC risk factor but mostly associated with well-differentiated NPC (Type I). Studies by Hildesheim et al. (2001) revealed an increased risk of NPC due to prolong exposure to formaldehyde. Furthermore, wood dust has been shown to be a trigger for EBV infection as long as the patients possessed at least one or more of the five anti-EBV antibodies that are associated with NPC (Hildesheim et al., 2001). Understanding its etiology allows earlier prognosis and treatments to reduce the mortality rate of NPC patients.

2.1.5 Clinical Prevention and Treatment of Nasopharyngeal Cancer

Diagnosis of nasopharyngeal carcinoma (NPC) is usually made at a later stage due to its location and subtle symptoms hence higher mortality rate. Furthermore, the rate of metastasis of NPC is the highest compared to other head and neck cancers. However, preventions and treatments are available for NPC patients. Early detection of NPC in the high-incidence area can be done by testing for antibodies against EBV or other suspected human proteome (Liu, 2016, p.9-10). Moreover, detection of the tumour using the NPC stages by the American Joint Committee on Cancer (AJCC) 2002 staging system allows

early-stage detection in which allows early treatment to be strategized. Symptoms and signs of NPC include neck masses, epistaxis, nasal obstruction and discharge, headache and other nonspecific indicators (Liu, 2016, p.4). Patients who possess such symptoms should opt to be clinically assessed especially if they are in high-incidence area or have a family member previously diagnosed with NPC. An indirect nasopharyngoscope will be performed to examine the lump in the nasopharynx, followed by a direct nasopharyngoscope using fiberoptic endoscope. A biopsy of the tumour will be performed if a suspicious growth in the nasopharynx is detected. A cross-sectional imaging by computed tomography (CT) scan or magnetic resonance imaging (MRI) will be done if the suspected tumour is unable to be visualized with the endoscopic examination. Current treatments available for NPC patients are the intensive-modulated radiation therapy (IMRT) for early stages of NPC and chemoradiation therapy (CRT) for advanced stages. However, biological target treatments are available to treat squamous cell cancer, by using cetuximab, a monoclonal antibody that inhibits epidermal growth factor receptor (EGFR), and bevacizumab, a monoclonal antibody that inhibits vascular endothelial growth factor (VEGF), though these treatments are still under research (Spratt & Lee, 2012).

2.2 Ribosomal Protein

Genes in the ribosomal protein subunit (RPS) family encode ribosomal proteins that makeup part of the cellular structure called ribosomes. Ribosomes catalyse protein synthesis. Ribosomes consist of 2 subunits, which are the small (40s) and large (60s) subunits. Ribosomal subunit accumulation needs a deposition of many ribosomal proteins on ribosomal RNA (rRNA) to form two ribonucleoprotein particles, a small and a large subunit that assists in protein translation (Wild et al., 2010). Ribosomal proteins genes have a unified

control mechanism for transcription and translation and are able to manage its own synthesis at the translational level and are involved in the translational regulation (Stelzl et al., 2001; Ishii et al., 2006). Ribosomal proteins are individual proteins and are separated by their gene expression level and tissue-specificity, as they are controlled by different regulators, optimal amino acids and codon (Ishii et al., 2006). Furthermore, ribosome biogenesis and protein translation coordinates with cell growth, proliferation, differentiation, which is vital for animal development. Any damage to any of these two processes can implement retardation to cell growth and disrupt animal development (Zhou et al., 2015). Many studies have shown mutation or differential expression of genes leads to multiple birth defects, diseases or cancer. Several tumour suppressors were shown to interact with ribosomal proteins, which interferes with cell proliferation or cell death (for example, their involvement with p53 in the MDM2 pathway). Their further involvement in other than ribosome biogenesis further concludes that they are able to perform extra-ribosomal functions as independent polypeptides.

2.2.1 Ribosomal Proteins in Cancer

Many studies have been conducted on the relationship between ribosomal genes and cancer progression. Ribosomal protein genes have been shown to be involved directly or indirectly in the proliferation of cancer cells. When compared between non-malignant and malignant nasopharynx tissue samples, differential expression of genes was observed, which indicates possible involvement in cancer progression (Fung et al., 2000). The relationship between expression of genes was further supported by Fang et al. (2008), where they suggested that NPC and specific pattern of gene expression plays a role in its oncogenesis and tumour progression. Dysregulation of ribosomal proteins leads to abnormal phenotype. Ribosome

assembly and function is reliant on the stoichiometric stability of ribosomal proteins, whereby a fall in the levels of one or more ribosomal proteins may possibly decrease the efficiency of, or even reduce in operational, the translational apparatus, and thus signal the cell to self-destruct (Naora & Naora, 1999).

The involvement of ribosomal protein eS1 in the cell transformation and apoptosis reveals that individual ribosomal proteins and their expression level regulates an extensive variety of activities correlated with cell growth and death (Naora & Naora, 1999). However, the expression of an individual ribosomal protein gene can vary depending on the type and state of the cell and also the presence of other signals (Naora & Naora, 1999). A down-regulation or up-regulation of an individual ribosomal protein gene does not necessarily lead to cell death. According to Naora and Naora (1999), the expression level of a ribosomal component that is close to the required threshold in a particular cell type may reduce. Subsequently the cells that are more sensitive to the reduction in the expression will produce a phenotype that is cell-type specific. The study by Naora and Naora (1999) shows that the increase in eS1 expression is able to induce transformation and is involved in the formation of tumours in nude mice. However not all of the ribosomal protein genes whose expression is up-regulated plays a direct role in oncogenesis. Nevertheless there is significant evidence that some translation initiation factors can be involved in the oncogenic transformation (Naora & Naora, 1999).

Normally, overexpression of protein is correlated with the presence of disease and cancer. Studies by Arthurs et al. (2017) showed an overexpression of eS19, eS21 and eS24 in prostate cancer. Knockdown of eS24 inhibits colorectal cancer cells consistent with its direct

relation with the formation of colorectal carcinoma (CRC) (Wang et al., 2015). Furthermore, the differential expression of *eS19* (Kondoh et al. (1992)) reveals links between the level of expression between normal and tumour colorectal cell lines that further indicates its association with cancer progression. Mutation in multiple genes (*eS19*, *eS24*, *eS17*, *eL33*, *uL18*, *uL5*, and *eL15* to name a few) causes Diamond-Blackfan anemia (DBA), which is an heritable red blood cell aplasia (Landowski et al., 2013).

As previously stated, identification of differentially expressed genes in normal and malignant nasopharyngeal epithelial cells will help in understanding the pathways involved in the development and progression of NPC (Fung et al., 2000). Apart from that, further understanding of ribosomal proteins and their association in cancer pathways allows biomarkers in early cancer detection and prognosis. For example, the detection of EBV allows early detection of NPC. Studies done by Lo et al. (2003) reveals that the LMP1 protein enhances cell survival and promotes tumour invasion in nasopharyngeal epithelial (NPE) cells, which was supported by Pathmanathan et al. (1995), where they observed a high expression of LMP1 in premalignant lesion of NPC (which facilitates their conversion into invasive NPC cells). Other genes reported to be associated with pathways involving NPC tumorigenesis includes *eL27*, *eL41* and *eL43* in the MAPK/ERK pathway (Sim et al., 2016).

2.2.2 uS19

In this study, the *uS19* gene, which was previously known as *RPS15*, will be further looked into based on its transcriptional and translational expression levels in NPC compared to

NPE. The *uS19* gene belongs to the S19 protein family located and expressed at the chromosome 19p13.3 (Figure 2.2A).

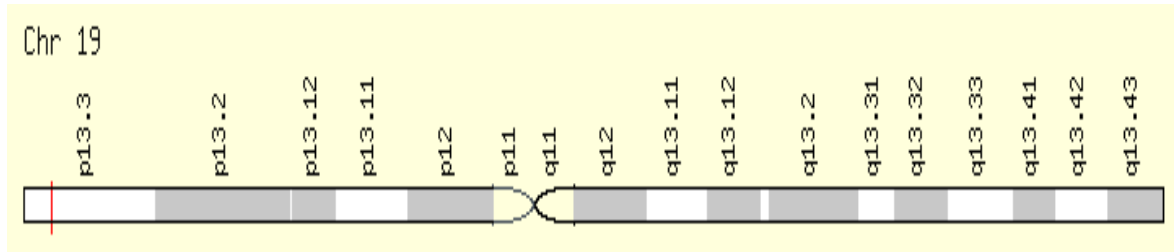


Figure 2.4: Location of *uS19* in S19 chromosome (Retrieved from <http://www.genecards.org/cgi-bin/carddisp.pl?gene=RPS15>)

This gene is similar to the S19 protein in bacteria and archeons, situated in the head of the 40S subunit and is suggested to be involved in the initiation and elongation steps of translation (Kitagawa et al., 1991; Rouquette et al., 2005). According to Ishii et al. (2006), *uS19* were expressed in skin, fetal brain and spleen. The *uS19* gene has been observed in many tumour cells and is highly expressed such as insulinomas, oesophageal and colon cancers and also during the differentiation phase of liver regeneration (Kitagawa et al., 1991). Findings by Marion and Marion (1988), reveals that the deterioration of ribosomal proteins S15 can cause the segmentary orientation of the 40S subunit. The *uS19* gene has proven to be vital in the final cytoplasmic maturation and assembly of the 40S subunit as the reduction of *uS19* in human cell lines causes the decrease in 40S subunits and 80S mature ribosomes (Robledo et al., 2008). The *uS19* gene is required for nuclear transport of the pre-40S particles and may directly or indirectly communicate with the nuclear transport machinery (Leger-Silvestre et al., 2004; Rouquette et al., 2005). The halt on the expression of *uS19* causes the decrease and eventually vanishing of the free 40S subunit pool and an increase in the free 60S subunit (Rouquette et al., 2005). The disappearance of *uS19*

intervenes with 40S biogenesis in human cells and leads to nuclear 40S biogenesis defect (Wild et al., 2010). Studies by Wong et al. (2014) have stumbled upon *uS19* gene exhibiting a heterogeneous expression pattern across the hepatocellular carcinoma samples, which indicates that they may have a substantial role in tissue-specific gene expression control. The *uS19* gene was found to be involved in the LRRK2 pathway increasing its pathogenicity of Parkinson's disease (Martin et al., 2014). Furthermore, this gene has also been reported to be involved in the Mdm2-p53-MdmX pathway as reported by Daftuar et al. (2013). They discovered that *uS19* gene binding of the pathway inhibits p53, a tumour suppressor that allows cell proliferation to take place. Besides that, uS19 protein has been shown to interact with the Fragile X Mental Retardation Protein 1 (FMRP1) through *in silico* analysis (Sim et al., 2016). Analysis of possible interactions through bioinformatics analysis allows further studies of the possible extra-ribosomal functions of the gene in which will allow a better understanding of cancer formation.

2.3 Protein Interaction Prediction through Bioinformatics Analysis

Protein-protein interactions studies are increasing due to the importance of understanding their roles within the cell. However, proteins do not function in isolation but with other macromolecules. Nasopharyngeal carcinoma (NPC) has a low rate of detection at an early stage. Through bioinformatics analysis, understanding of its molecular mechanism can identify potential biomarkers that can facilitate early diagnosis, better prognosis, and allow a more curative treatment. Bioinformatics acts as a platform where a large amount of data are tabulated and stored for the convenience of other researchers to look at for further refinement of studies. According to Benton (1996), bioinformatics is outlined as a field of studies that includes biological data acquisition, processing, storage, distribution, analysis,

and interpretation. It mixes the tools and techniques of arithmetic, engineering science, and biology with the aim of understanding the biological significance of a spread of data. Chen et al. (2017) stated the importance of performing bioinformatics analysis with all the collected data from microarray analysis for the purpose of predicting the progression of NPC, understanding the mechanism of the occurrence and development of tumour and interaction network. The aims of bioinformatics are divided into three-folds, as depicted by Luscombe et al. (2001). First, the manner of how the bioinformatics systematises data allows researchers to access existing information and to submit new entries as they are produced, e.g. The Protein Data Bank for 3D macromolecular structures. However, data stored are essentially useless unless they are further analysed. Thus the purpose of bioinformatics is to further establish the data stored. The second aim is to develop tools and resources that aid in the analysis of data, e.g. comparing sequences of a particular protein of interest and previous characteristic sequences using programs such as FASTA and PSI-BLAST. The third aim is to use these tools to evaluate the data and interpret the results in a biologically significant manner to highlight novel features, which could benefit the development of research. Through bioinformatics analysis, preliminary studies are able to be done on a variety of novel proteins of a molecular pathway to further investigate their interactions. The uS19 protein is known to be involved in ribosome biogenesis. However, recent studies have shown that it has extra-ribosomal functions that are involved in some diseases and cancer pathways. As an individual protein model structure is unavailable, a 3D model structure is done to dock with other interacting proteins to observe their possible functions in NPC.

2.3.1 SWISS-MODEL

To fully understand the structural function of a protein on a molecular level whether it functions individually or within a complex require the necessary tools to construct a 3D protein model for further assessment. As we know, protein is involved in ribosome biogenesis and also pathways which could lead to diseases and cancer. Hence studies are now revolved around protein-protein interaction to further validate and determine targeted novel proteins which could assist as potential biomarkers for future studies. SWISS-MODEL is a server which allows not only the construction of 3D protein model but also the tools to analyse and compute the best protein model structures for further analysis. SWISS-MODEL allows only input of sequences, which will be used to build a 3D model. SWISS-MODEL server uses a PROMOD3 modelling engine to generate actual models based on the highest ranked template, and additional models are progressively generated if new segments of the target sequence can be covered, or if alternative models with significantly different conformations can be generated (Bienert et al., 2017). From the site itself, the model built can be further assessed to obtain its QMEAN score, while Ramachandran plot will be plotted to observe the quality of the structure whether it is viable for further analysis. Qualitative model energy analysis (QMEAN) is a composite scoring function describing the major geometrical aspects of protein analysis that is able to provide both a global quality estimate as well as a per-residue local quality estimate (Benkert et al., 2008; Bienert et al., 2017). It uses a statistical potential of mean force to produce a quality valuation on a global and local scale (Waterhouse et al., 2018). QMEAN score point to structural descriptors contributes most to the final score allowing the users to distinguish good from bad models, through the three-residue torsion angle potential (Benkert et al., 2008; Benkert et al., 2011). Ramachandran plot analysis uses PROCHECK to evaluate the stereochemical quality and

accuracy of the protein structure models built using the SWISS-MODEL workspace. The analysis consists of ϕ - Ψ angles, which is divided into three categories: ϕ dependent, Ψ dependent, and ϕ - Ψ co-dependent distances that acts as a parameterization of the protein backbone in determining protein structures and defining its secondary structure (Ho et al., 2003). According to Carugo and Djinovic-Carugo (2013), there are three clusters of protein structures that can be identified on the ϕ - Ψ plot (PR plot), which correspond to the structure rich in helices, in strands, and in both helices and strands, and are connected by a sigmoid curve that delineates the fraction of the plot that can be populated by folded proteins. Each protein structure on the PR plot are according to the average values of its ϕ and Ψ backbone torsion angles and tend to cluster in three regions that correspond to the structures rich in α -helices and β -strands, and in both types of secondary structure (Carugo & Djinovic-Carugo, 2013). The Ramachandran plot shows the predicted favoured and forbidden regions on the plot based on its α -helices and β - strands regions, where the disfavoured regions are due to the crowding of the H atoms due to its prevention of the formation of H-bond with the solvent (Lovell et al., 2003). Models with more favourable regions are taken for further analysis to determine their interacting proteins. As no known individual structural model of uS19 protein is available online, this tool allows us to build a predictive 3D structural model of uS19 protein, which is used to obtain predictive structure models that possibly interacts with it.

2.3.2 Vector Alignment Search Tool (VAST)

Vector Alignment Search Tool (VAST) is a server that is hosted by NCBI. VAST allows users to input their protein structure and search onto the NCBI database for similar structures. This is useful especially for novel proteins structures as it can shed some light on their

evolutionary relationship. Proteins interacting with one another tend to interact at the same domain when putting through VAST, users are able to identify similar protein structures, which could indicate possible interaction with their complexes. The 3D structure is highly conserved in protein evolution and the interpretation of the similarities allows insights to many extended families or their possible functions related to that specific pathway (Gibrat et al., 1996). This allows us to predict their possible functions as most of the conserved regions tend to play the same role (Panchenko & Madej, 2004). The input of the PDB file of targeted 3D protein structures will be run through a series of algorithm to produce an alignment of secondary structure elements that are similar to the targeted structure. The output will then be run through to observed their interacting partners to be further docked together with the targeted structure.

2.3.3 IntAct

The IntAct molecular interaction database is an open-source database that allows users to obtain data on its interacting partners through a series of protein-protein interaction (PPI) archival curation of literature and deposited publication data (Kerrien et al., 2012). The database is created on a relational database management system Hibernate object-relational mapping, Java-based middleware, Lucene/SOLR indexing and a web-based front end for public search, visualisation, and download of the data (Orchard et al., 2014). The site will detect and run through a series of published data and those with similar domain sites are said to be interacting with the input protein. Links for further identification of the interacting protein are available for users to view through external resources like UniProt KB. The information such as the PDB ID obtained from the list of interactors or their protein sequences in FASTA format can be obtained and taken for further analysis like docking.

Through this database, possible interacting partners with uS19 protein was obtained to be docked to attain the objective in this study.

2.3.4 ClusPro

Protein interactions are crucial in understanding their functions in molecular complexes and subsequently their importance in pathways. Many experimental studies have been done through molecular techniques such as pull-down and Mammalian di-Hybrid. Through bioinformatics, we are able to uncover a prediction of possible functions in complexes that can be further confirmed through molecular experiments. ClusPro is an open-source web-based server that docks two interacting proteins. According to Kozakov et al. (2017), the server performs three computational steps. Firstly, through rigid body docking using PIPER, which is a docking program based on the Fast Fourier Transform (FFT), by sampling billions of conformations. Secondly, the root-mean-square deviation (RMSD) based clustering of the 1000 lowest energy structures are generated to find the largest clusters that will represent the most likely models of the complex. Lastly, refinement of selected structures using energy minimization. The FFT based docking program includes a pairwise potential as part of its scoring function, where the approach is based on DARS (Decoys as the Reference State). This is a novel approach that allows large decoy set of docked conformations to be generated using only shape complementarity as the scoring function. This uses the frequencies of interacting atom pairs in these decoy structures as the reference state that increases the accuracy of docking, resulting in more near-native structures (Kozakov et al., 2017; Vadjia et al., 2018). The receptor and ligand protein will be put through the program and evaluated based on its RMSD and energy minimization to obtain clusters with the most members that will be further evaluated and refined.

2.3.5 PatchDock and FireDock

Clusters obtained through docking have to be further evaluated and refined as all data obtained are done through predictions. Unlike standard molecular experiments, uncertainties of whether the interactions between two proteins will definitely arise and must be evaluated through a series of algorithms. Hence, PatchDock and FireDock approach is done to further evaluate and refine the docking complexes obtained from ClusPro. All models in this study were further validated through this approach. PatchDock is a geometry-based molecular docking algorithm that is designed to locate docking transformations that generate decent molecular structure complementarity. Once applied, it generates a wide interface areas and slight amount of steric clashes that includes numerous matching native qualities of docked molecules that have complementary characteristics (Schneidman-Duhovny et al., 2005). According to Schneidman-Duhovny et al. (2005), the algorithm for PatchDock divides the Connolly dot surface representation of the molecules into concave, convex, and flat patches. Those that are complementary are then matched in order to generate candidate transformations where each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy and the RMSD clustering is applied to the candidate solutions to discard redundant solutions (Schneidman-Duhovny et al., 2005). The top 10 from PatchDock are then put through FireDock for further refinement. FireDock is a web server used for further refinement of protein-protein docking solutions that comprises a side-chain optimization that allows a high-throughput refinement of up to 1000 solution candidates. This method simultaneously focuses the problem of flexibility and scoring of solutions produced by fast rigid-body docking algorithms. These algorithms provides a list of refined complexes, ranked by a binding energy function and a

3D visualisation for observing and comparing between the refined complexes (Mashiach et al., 2008).

CHAPTER 3

MATERIALS AND METHODS

3.1 Cell Culture

Cell lines were chosen to be used in this study to obtain an insight of the expression levels of uS19 in NPC. Six NPC cell lines (HK1, HONE-1, SUNE-1, TW 01, TW 04, and C666-1) and one NPE cell line (NP69) were used in this study (Table 3.1). The cell lines were provided by the Faculty of Medicine and Health Sciences of Universiti Malaysia Sarawak and Universiti Malaya, with the consent from Prof. George S.W. Tsao, University of Hong Kong. All NPC cell lines were cultured using RPMI 1640 (Nacalai Tesque, Japan) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Nacalai Tesque, Japan) and 100 U/ml penicillin-streptomycin (GIBCO, USA). The NPE cell line was cultured in defined-keratinocyte serum-free medium (D-KSFM) (GIBCO, USA), supplemented with 5% dialyzed FBS, 100 U/ml penicillin-streptomycin and 0.2 ng/ml recombinant epidermal growth factor, as provided by the manufacturer GIBCO, USA. All seven cell lines were incubated at 37 °C, in humidified 5% carbon dioxide (CO₂) environment. Cells are taken for extraction when it reaches confluency of 70-80%, which assures the sufficiency of cells extracted for further application. The passage of cell lines used in the study is summarized in Table 3.1.

Table 3.1: Passage number of cell lines used

Cell Lines	mRNA analysis			Protein analysis		
	BR1	BR2	BR3	BR1	BR2	BR3
NP69	P4	P2	P6	P5	P3	P2
HK1	P2	P2	P3	P3	P2	P2
HONE-1	P4	P3	P2	P3	P2	P4
SUNE-1	-	-	-	P3	P5	P2
TW 01	P2	P4	P2	P2	P2	P2
TW 04	P4	P4	P2	P3	P4	P2
C666-1	P5	P2	P4	P3	P4	P5

3.2 Gene Expression

3.2.1 Total RNA Extraction

The total RNA was extracted from all five cell lines using TRIzol (Chomczynski and Sacchi, 1987). The flask containing the cultured cell lines were taken and placed inside the RNA hood. The media inside the flask was removed and the cell monolayer was rinsed with 1 ml of PBS. 1 ml of TRIzol reagent (Invitrogen™, USA) per 3.5 cm diameter dish was added into the flask. Using a cell scraper, the cells were scraped off the flask thoroughly for 1 minute and then mechanically lysed by pipetting up and down several times and then incubated for 5 minutes. The TRIzol mixture was transferred into a clean sterile tube and 0.2 ml of chloroform was added and vortexed for 15 seconds and then incubated for 5 minutes. The sample was centrifuged at 10,000 rpm for 15 minutes at 4 °C and the upper aqueous layer was transferred into a new tube. 0.5 ml of isopropanol was added and incubated for 5 minutes. The tube was then centrifuged at 10,000 rpm for 10 minutes at 4 °C and then the supernatant was removed. The pellet was washed with 1 ml of 75% ethanol and left to dry for 10 minutes. The RNA pellet was immersed with 30 µl of RNase-free water and pipetted up and down to dissolve the RNA and then stored at -80 °C. The concentration and purity of

the RNA extracted were measured using a UV spectrophotometer (Ultrospec 1100 pro, UK) at the absorbance of 260 nm and 260/280 ratio respectively. The total RNA extracted from the cell lines had sufficient concentration and high purity (260/280 ratio is 1.9-2.1) for a single tube reverse transcription. The RNA samples were quantitated so that the amount of RNA is approximately the same and can be used to compare between different cell lines.

3.2.2 cDNA Synthesis

Prior to cDNA synthesis, the total RNA extracted from all cell lines were treated with DNase treatment. A total of 2 ng of RNA is diluted in distilled water (dH₂O) up to 7 µl. Then, 1 µl of RQ1 buffer (Promega, USA) and 2 µl of RQ1 DNase (Promega, USA) were added into the mixture and incubated for 37 °C for 30 minutes. After that, 1 µl of RQ1 DNase Stop Solution (Promega, USA) was added in and then incubated for 65 °C for 10 minutes. The tube of the treated RNA sample is then used for cDNA synthesis. Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV-RT) was used to synthesize cDNA from the extracted RNA samples. The treated RNA sample was added with 1 µl of random primers and 3 µl of RNase-free water (Promega, USA) in an RNase-free microcentrifuge. The mixtures were then incubated at 70 °C for 5 minutes to melt the secondary structure within the template and immediately cooled on ice for 1 minute to prevent the secondary structure from forming. A volume of 5 µl of M-MLV 5x reaction buffer (Promega, USA), 1.25 µl of dNTP (Promega, USA), 0.6 µl of Recombinant RNasin® Ribonuclease Inhibitor (Promega, USA) and 1 µl of M-MLV RT (PROMEGA, USA) were added to the previous mixture and topped up with nuclease-free water to a total volume of 25 µl. The samples were mixed gently by flicking the tubes and then incubated for 60 minutes at 37 °C and later on 70 °C for 15 minutes. The cDNA samples were then labelled and stored in -20 °C.

3.2.3 Real-time PCR

Real-time PCR (qPCR) was performed using Quantinova SYBR Green PCR Kit (QIAGEN, Germany). Primers were predesigned and tested its compatibility with the kit. In this study, the gene of interest was *uS19* while *GAPDH* was used as a reference gene (Table 3.2).

Table 3.2: Primer sequences for *GAPDH* and *uS19*

Description	Sequences (5'-3')	Amplicon size (bp)
<i>GAPDH</i> (Forward)	CTGGGCTACACTGAGCACC	101
<i>GAPDH</i> (Reverse)	AAGTGGTCGTTGAGGGCAATG	
<i>uS19</i> (Forward)	GAAGACGCACCTGCGGGACA	115
<i>uS19</i> (Reverse)	GTGGCCGATCATCTCGGGCTTG	

The reference gene was used for normalization in the qPCR analysis. The real-time PCR assay was carried out and reported according to MIQE (Minimum Information for Publication of Quantitative Real-time PCR Experiments). The assay was carried out as suggested according to the manufacturer's protocol. The reaction mix was prepared with 6.25 µl of 2X QuantiNova SYBR Green PCR Master Mix, 0.7 µl of both forward and reverse primers each, 3.85 µl dH₂O and 1 µl of cDNA template. The reaction mix was mixed well and dispensed into the Rotor-Gene Q machine. The cycling condition is a three-melt curve, where the initial temperature hold for 2 minutes at 95 °C, denaturation for 95 °C for 5 seconds, annealing for 10 seconds at 60 °C then 15 seconds for 72 °C for 35 cycles. The fluorescence was detected at the green channel for both *uS19* and *GAPDH* gene. Reactions without template were used as a negative control. Standard curve analysis using the standard curve method (Livak and Schmittgen, 2011) was done to determine PCR efficiency and R²

value for both the gene. The validation of primer efficiency was determined using the slope of calibration curve as follow:

$$\text{Eq 1: Amplification efficiency} = [10^{-1/\text{slope}} - 1]$$

The PCR assay efficiencies were evaluated by generating a five-fold serial dilution over five log magnitude. A standard curve was plotted using the average CT values of duplicates against log input amount (80, 16, 3.2, 0.64, 0.128 ng/μl). From the standard curve, the amplification efficiency of the reactions was calculated where the ideal amplification range is 0.8-1.1 (80-110%) and m value between -3.6 to -3.1. Comparative CT method ($\Delta\Delta\text{CT}$ method) was applied to relatively quantitate *uSI9* in NPC compared with NPE cell line. This method requires the relatively equivalent amplification efficiencies between target and reference genes. In the validation experiment, the absolute value of the slope of Δ CT (reference-target genes) vs log input was as required ($m < 0.1$).

The qPCR was performed in triplicate with three biological replicates for each cell line along with duplicate technical replicates. The biological replicates are used to test the variability between samples, while the technical replicates are to test the variability in the protocol or pipetting error. Student t-test was used to determine the significance of overall differential expression. The target gene (*uSI9*) expression was normalized to the reference gene (*GAPDH*) and the normalized target gene expression was calculated using the method of $\Delta\Delta\text{CT}$ method as follow.

$$\text{Eq 2: } \Delta\text{C}_\text{T} = \text{C}_\text{T} \text{ target} - \text{C}_\text{T} \text{ reference}$$

$$\text{Eq 3: } s = \sqrt{s_1^2 + s_2^2}, s \text{ is standard deviation}$$

The standard deviation for $\Delta\Delta C_T$ was carried from standard deviation for ΔC_T of target genes. The relative gene expression to the normal sample (NP69) was determined using the formula,

$$\text{Eq 4: } \Delta\Delta C_T = \Delta C_T \text{ test sample} - \Delta C_T \text{ normal sample}$$

The $\Delta\Delta C_T$ is then transformed to log form, $2^{-\Delta\Delta C_T}$ to present the target gene fold difference in relative to the normal sample. The $2^{-\Delta\Delta C_T}$ values > 1.0 are known as overexpressed while $2^{-\Delta\Delta C_T}$ values < 1.0 is known as under expressed. Significant readings for each run and individual gene expression (p-value <0.05) enabled the fold differences for the two runs to be averaged. The error bar was plotted using the standard error of the fold difference mean.

3.3 Protein Expression

3.3.1 Total Protein Extraction

The protein from all six NPC cell lines and one NPE cell line was extracted using Radioimmunoprecipitation assay (RIPA buffer) (1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 150 mM NaCl, 2 mM EDTA, 1 mM PMSF). The 25 cm² flask containing cells with confluency about 70-80% were put on ice before the extraction begins. The media was discarded and the flask was washed twice with 1ml of precooled 1X TBS (Tris-Buffered Saline). 1 ml of RIPA buffer was added in and cells scraped with a cell scraper. The flask containing the lysis buffer was put on ice and rocked on a rocker for 30 minutes to allow the cells to be lysed further. After that, the mixture was transferred to a 1.5 ml centrifuge tube,

vortexed for a few seconds and then centrifuged at 16,000 xg for 20 mins at 4 °C. The supernatant was carefully removed to a new 1.5ml centrifuge tube and stored at -80 °C. Due to low concentration of protein for NP69, protein precipitation was done by first precooling acetone in -20 °C overnight. A 1:4 ratio of sample to acetone was mixed together by vortex and then incubated for 1 hour at -20 °C. The mixture was then centrifuged at 15,000 xg for 15 minutes at 4 °C. The supernatant was carefully discarded and the tube containing the protein pellet was dried for about 15-30 minutes. Next, 30 µl of RIPA buffer was added and vortexed till the pellet dissolve. The precipitated protein was then stored at -80 °C.

3.3.2 SDS-PAGE

In preparation of denatured protein samples, 20 µg of each sample were mixed with 4X Laemlli sample buffer and denatured for 10 minutes at 95 °C. Equal amounts of 20 µg for each sample was loaded into wells of a mini format Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE) gel along with an ExcelBand High Range Protein Marker (SMOBIO, Taiwan). The running gels were made prior with 12% resolving gel (4 ml 30% acrylamide/bis, 2.5 ml 1.5M Tris (pH 8.8), 100 µl 10% SDS, 3.35 ml dH₂O, 60 µl 10% APS and 7 µl TEMED) and 4% stacking gel (0.65 ml 30% acrylamide/bis, 1.25 ml 0.5 M Tris (pH 6.8), 50 µl 10% SDS, 3.05 ml dH₂O, 66 µl 10% APS and 13 µl TEMED). The ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) were added last as these two are the solidifying reagent. The gel electrophoresis was run for 30 minutes at 60V and then the voltage was increased to 100V for an hour.

3.3.3 Western Blot

Total protein from the SDS-PAGE gel was transferred onto a polyvinylidene difluoride (PVDF) membrane (0.2 mm pore size). The membrane was presoaked in methanol to activate it. The foam pads and mini filter papers were presoaked in 1X transfer buffer. The transfer system used was a wet system, therefore all equipment used were soaked properly beforehand. The stacking gel was removed before assembling the transfer sandwich. Air bubbles were gently rolled out using a roller as the sandwich is being assembled to ensure no air bubbles were trapped in between the gel and the membrane. The cassette was put into the transfer tank with ice blocks and filled with precooled 1X transfer buffer. The ice block was used to maintain a cold environment (4 °C). The transfer was done at 50V for 45 minutes. The membrane was stained with Ponceau S to check transfer quality and destained with 1X TBST (Tris-Buffered Saline Tween 20) buffer. The membrane was blocked using blocking buffer (3% Bovine Serum Albumin (BSA) in 1X TBST) for an hour at room temperature. The blocking step is to prevent non-specific background binding of the primary and/or secondary antibody to the membrane. The blocking buffer was then discarded and the membrane was incubated with blocking buffer containing primary antibody (Thermo Fisher Scientific, USA) against uS19 overnight at 4 °C. Beta-Actin antibody (Santa Cruz Biotechnology, USA) was used as a loading control. After primary incubation, the membrane was washed with 1X TBST 3-5 times for 5 minutes each on a rocker. The secondary antibodies anti-rabbit (uS19) and anti-mouse (Beta-Actin) are horse-radish peroxidase (HRP)-conjugated. They were poured over the membrane specific to the targeted protein and incubated in room temperature for 2 hours. The membrane was then washed again with 1X TBST 3-5 times for 5 minutes each on a rocker before incubating with a 1:1 ratio of the enhanced chemiluminescent (ECL) solution (Thermo Fisher Scientific, USA) for

about 2 minutes. After incubation, the membrane was removed and placed into a plastic wrap and the chemiluminescent signals were captured using ImageQuant TL instrument (GE Healthcare, United Kingdom). This western blot was carried out in three biological replicates. The fold difference for this analysis was calculated according to the formula as shown (Eq5). The target protein and loading control represents the uS19 and beta actin protein expression for the cell line respectively. Loading control for normal represents the uS19 protein expression in NP69 cell line. The average fold change for each cell line was calculated where NP69 represents as the normal (values >1 represents over expression and <1 represents under expression).

$$\text{Eq 5: } \frac{\text{Target protein (TP)} \times \text{Loading control (LP)}}{\text{Loading control for normal (LC*)}}$$

3.4 Bioinformatics Analysis

3.4.1 3D Protein Model Construction

Due to unavailability of an individual uS19 protein structure on UniProt KB, a predictive 3D protein model structure was created. The amino acid sequence of the target sequence, the human uS19 (accession no. NP_001295155.1) was retrieved from National Centre for Biotechnology Information (NCBI). The FASTA format of the uS19 (40S ribosomal protein S15 isoform 1) was chosen as it is the longest amino acid sequence and isoform 2 is included in the isoform 1. Multiple analysis was done to construct the best 3D protein model structure and all lead to constructing a model with just using the amino acid sequence obtained from NCBI. The sequence obtained was then input into SWISS-MODEL Workspace and a model was build. The model build then undergoes quality check using QMEAN score and

Ramachandran plot that is available at the SWISS-MODEL server. RMSD score for the structure build was obtained from Swiss PdbViewer v4.1.1 (SPDBV 4.1.1) (Guex and Peitsch, 1997) when score < 5 Å shows a good structure.

3.4.2 Identification of Protein Interactors

The model constructed was then submitted to VAST to search through the NCBI database for similar structures. The results were filtered down to human only and then input into IntAct to search for its interacting proteins. The PDB ID for each protein were obtained and those without a PDB ID or singular structure were constructed using the FASTA format of the amino acid sequence that was available under their accession number in the UniProt KB server. The construction of the 3D protein model was done using SWISS-MODEL Workspace and checked for its quality using the same method as mentioned.

3.4.3 Docking of Protein Partners

Dual docking analysis were done for each uS19 and its probable interacting protein. Each interacting proteins were docked with uS19 model that was previously constructed using ClusPro, where uS19 act as a receptor while the interacting protein as ligand. The pdb file of the most clusters was downloaded. The docking analysis was then further evaluated and refined using PatchDock and FireDock. The RMSD scores for each docked complexes from both docking analyses were obtained using SPDBV 4.1.1. The interface residue contacts (<5 Å) were obtained using SPDBV 4.1.1 and the highest number of interface residues were chosen to portray the docked complexes viewed using Pymol molecular graphics version 2.3. The hydrophobic interaction and ionic interactions were then evaluated using Protein Interaction Calculator (PIC) to assess the stability of the predictive docking structures.

CHAPTER 4

EXPRESSION ANALYSIS OF *uS19*

4.1 Introduction

The transcription level of *uS19* was quantitated using qPCR according to MIQE guideline. The assay was analyzed using quantitative method ($2^{-\Delta\Delta C_t}$) to obtain the fold difference of the target gene relative to their normal cell line. Protein level of uS19 was obtained through western blot and the data obtained was tabulated and calculated to obtain its fold difference compared to normal cell line.

4.2 Results

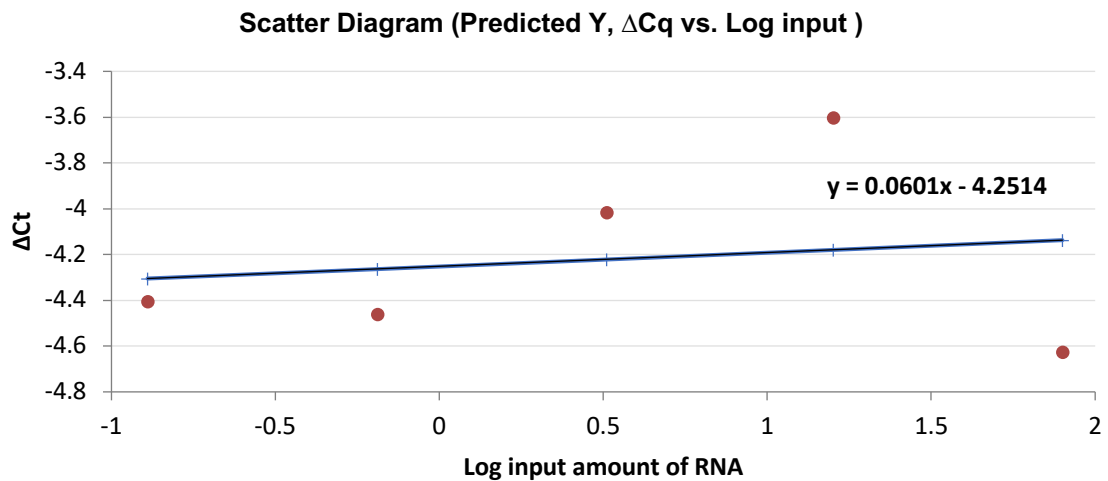
The results for the validation experiment are summarized in Table 4.1 and Table 4.2 where the 5-fold dilution range was used. Both genes are within ideal amplification range with PCR efficiency of 0.95-1.1 (95%-110%), where the R^2 for *uS19* and *GAPDH* genes are 0.95 (95%) and 1.0 (100%) respectively (Table 4.2) revealing a high level of confidence in the C_T values correlating with the log output of cDNA. The absolute values from the slope of standard curve were evaluated prior to relative quantification qPCR analysis. The absolute value for the slope for the validation experiment have to be where $m < 0.1$. Based on Figure 4.1, *uS19* primers has a slope of 0.0601, which is less than 0.1. Therefore the validation experiment can be concluded to be successful.

Table 4.1: ΔC_T calculation for validation experiment

Input amount (ng)	Log output of cDNA amount	ΔC_T <i>uS19-GAPDH</i>
80	1.9	-4.625
16	1.2	-3.6
3.2	0.51	-4.015
0.64	-0.19	-4.46
0.128	-0.89	-4.405

Table 4.2: PCR efficiency, R^2 , and m-value of target and reference gene

Gene	PCR Efficiency	R^2	m-value
<i>uS19</i>	0.95	0.99	-3.439
<i>GAPDH</i>	1.00	0.98	-3.318

**Figure 4.1:** Validation plot of ΔC_T (crossing threshold) versus log input amount of RNA.

Biological triplicate with technical duplicate were used to obtain a reliable data of transcription expression level to be quantitated using student *t*-test analysis. The average C_T values of *uS19* was normalized to the reference gene, *GAPDH*, where ΔC_T *uS19-GAPDH* and standard deviation (SD) were calculated (Table 4.3). The relative gene expression of the target gene, $\Delta\Delta C_T$ values were calculated along with the standard deviation (SD), where the $\Delta\Delta C_T$ for NP69 cell line is 0 as it is calculated in comparison to NP69 cell line (Table 4.3). The fold difference using the arithmetic formula, $2^{-\Delta\Delta C_T}$ is then calculated for each biological replicates to obtain the fold difference in relative to the normal cell line (NP69), where values >1.0 is overexpressed while values <1.0 is underexpressed. The fold difference is expressed as a range where the values in the bracket represent the lower and upper limit of the fold difference considering the standard deviation (Table 4.3).

Table 4.3: Normalized expression of *uS19* using the $\Delta\Delta C_T$ method

Cell line	Biological replicate	<i>uS19</i> Average $C_T \pm SD$	<i>GAPDH</i> Average $C_T \pm SD$	ΔC_T (<i>uS19</i> - <i>GAPDH</i>) $\pm SD$	$\Delta\Delta C_T$ (ΔC_T NPC- ΔC_T NP) $\pm SD$	Fold difference of <i>uS19</i> in NPC
NP69	1	22.26 $\pm$ 0.099	22.8 $\pm$ 0.184	-0.54 $\pm$ 0.209	0.000 $\pm$ 0.209	1.01 (0.865- 1.156)
	2	18.065 $\pm$ 0.007	18.65 $\pm$ 0.127	-0.585 $\pm$ 0.127	0.000 $\pm$ 0.127	1.004 (0.915- 1.092)
	3	18.605 $\pm$ 0.035	19.41 $\pm$ 0.000	-0.805 $\pm$ 0.035	0.000 $\pm$ 0.035	1.00 (0.976- 1.025)
HK1	1	18.515 $\pm$ 0.092	16.18 $\pm$ 0.057	2.335 $\pm$ 0.108	2.875 $\pm$ 0.108	0.137 (0.126- 0.147)
	2	16.555 $\pm$ 0.053	16.625 $\pm$ 0.007	-0.07 $\pm$ 0.036	0.515 $\pm$ 0.036	0.7 (0.683- 0.718)

Table 4.3 continued

HONE-1	3	18.81 ± 0.028	17.34 ± 0.042	1.47 ± 0.051	2.275 ± 0.051	0.207 (0.199- 0.214)
	1	21.155 ± 0.064	20.865 ± 0.035	0.29 ± 0.073	0.83 ± 0.073	0.563 (0.535- 0.592)
	2	19.765 ± 0.021	17.555 ± 0.035	2.21 ± 0.041	2.795 ± 0.041	0.144 (0.14- 0.148)
TWO 1	3	18.425 ± 0.007	18.335 ± 0.106	0.09 ± 0.106	0.895 ± 0.106	0.539 (0.5- 0.579)
	1	19.775 ± 0.007	17.16 ± 0.170	2.615 ± 0.170	3.155 ± 0.170	0.113 (0.1- 0.126)
	2	18.935 ± 0.007	18.13 ± 0.085	0.805 ± 0.085	1.39 ± 0.085	0.382 (0.36- 0.405)
TWO 4	3	18.715 ± 0.028	18.06 ± 0.085	0.655 ± 0.085	1.46 ± 0.085	0.364 (0.343- 0.386)
	1	22.22 ± 0.085	22.11 ± 0.000	0.11 ± 0.085	0.65 ± 0.085	0.638 (0.601- 0.676)
	2	17.935 ± 0.092	18.06 ± 0.071	-0.125 ± 0.116	0.46 ± 0.116	0.729 (0.671- 0.788)
C666-1	3	18.04 ± 0.028	18.435 ± 0.049	-0.395 ± 0.057	0.41 ± 0.057	0.753 (0.723- 0.783)
	1	21.73 ± 0.255	22.13 ± 0.948	-0.4 ± 0.981	0.14 ± 0.981	1.126 (0.46- 1.791)
	2	17.2 ± 0.778	16.595 ± 0.778	0.605 ± 0.782	1.19 ± 0.782	0.504 (0.255- 0.754)
	3	28.54 ± 0.212	28.46 ± 0.085	0.08 ± 0.228	0.885 ± 0.228	0.548 (0.462- 0.634)

The average fold difference along with the standard deviation of the fold difference between biological replicates is tabulated in Table 4.4. This was calculated by normalizing to average $2^{\Delta\Delta C_t}$ value of *uS19-GAPDH* gene in NP69 cell line and the expression trend showed a downregulation in NPC cell lines when compared to NP69 as shown in Figure 4.2. Student *t*-test with a *p* value <0.05 were calculated and shown in Table 4.4. It revealed a significant underexpression of *uS19* in NPC cell line when compared to the NPE cell line, NP69. The cumulative fold difference of *uS19* in NPC cell lines were calculated and shown in Table 4.5 and displayed in Figure 4.3. The overall cumulative fold difference of *uS19* in 6 NPC cell lines was significant in comparison to the normal cell line, NP69. A student *t*-test was also done for all NPC cell lines compared to the NPE cell line (NP69) and the result showed that the down-regulation of *uS19* gene in NPC cell lines is statistically significant with a *p*-value of 0.041 (*p*<0.05). This statistical analysis concludes that *uS19* gene is down regulated in NPC cell line compared to NPE cell line.

Table 4.4: Average fold difference of *uS19* expression in NPC relative to NP69

Cell Line	Mean fold difference	SD	p-value
NP69	1.005	0.005	
HK1	0.348	0.307	0.028
HONE-1	0.416	0.235	0.046
TWO 1	0.286	0.151	0.017
TWO 4	0.707	0.061	0.019
C666-1	0.726	0.347	0.035

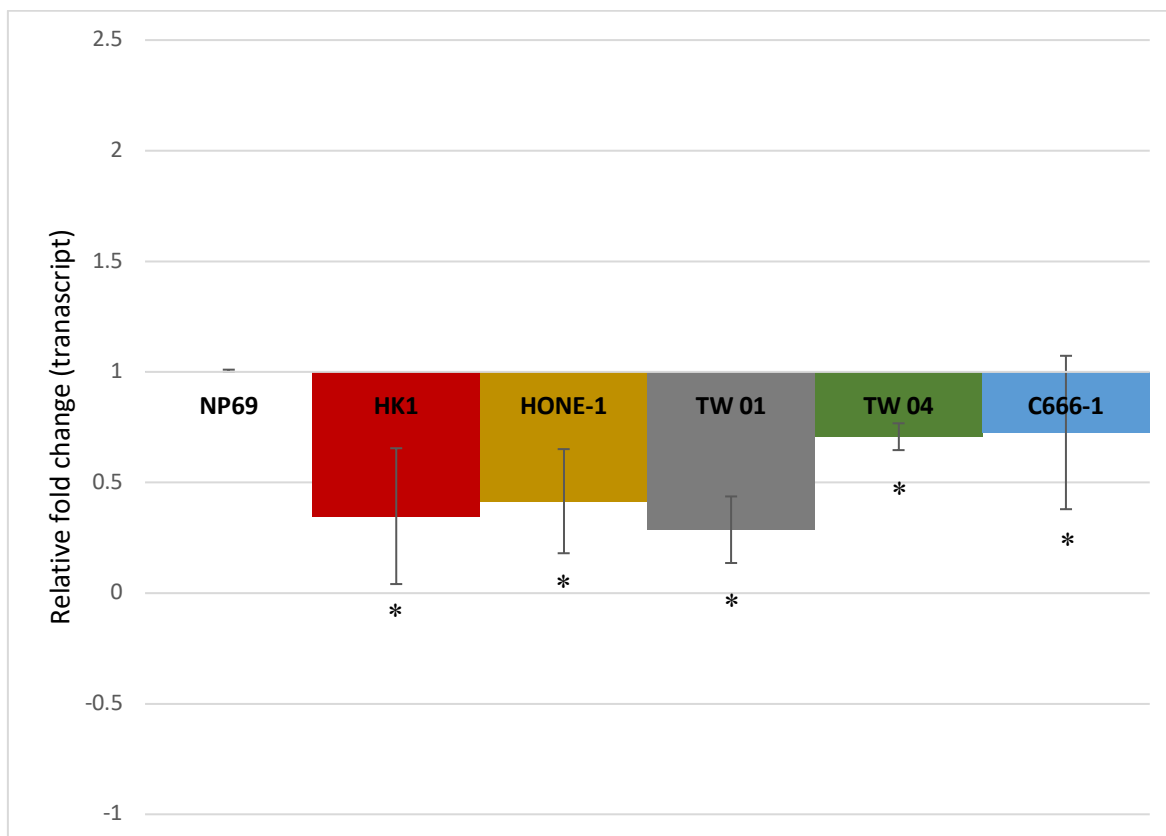


Figure 4.2: Relative fold difference of *uS19* expression in NPC cell lines versus NPE, NP69. The y- and x- axis represents the relative fold change and cell lines respectively. Data from NPC cell lines were compared to the NPE cell line (NP69) respectively. The asterisk symbol (*) represents p-value <0.05. The error bars were included and represents the standard deviation of the expression level of *uS19* in each cell line.

Table 4.5: Cumulative fold difference of *uS19* expression in NPC versus NP69

Cell line	Mean fold difference	SD	p-value
NP69	1.005	0.005	
NPC	0.497	0.116	0.041

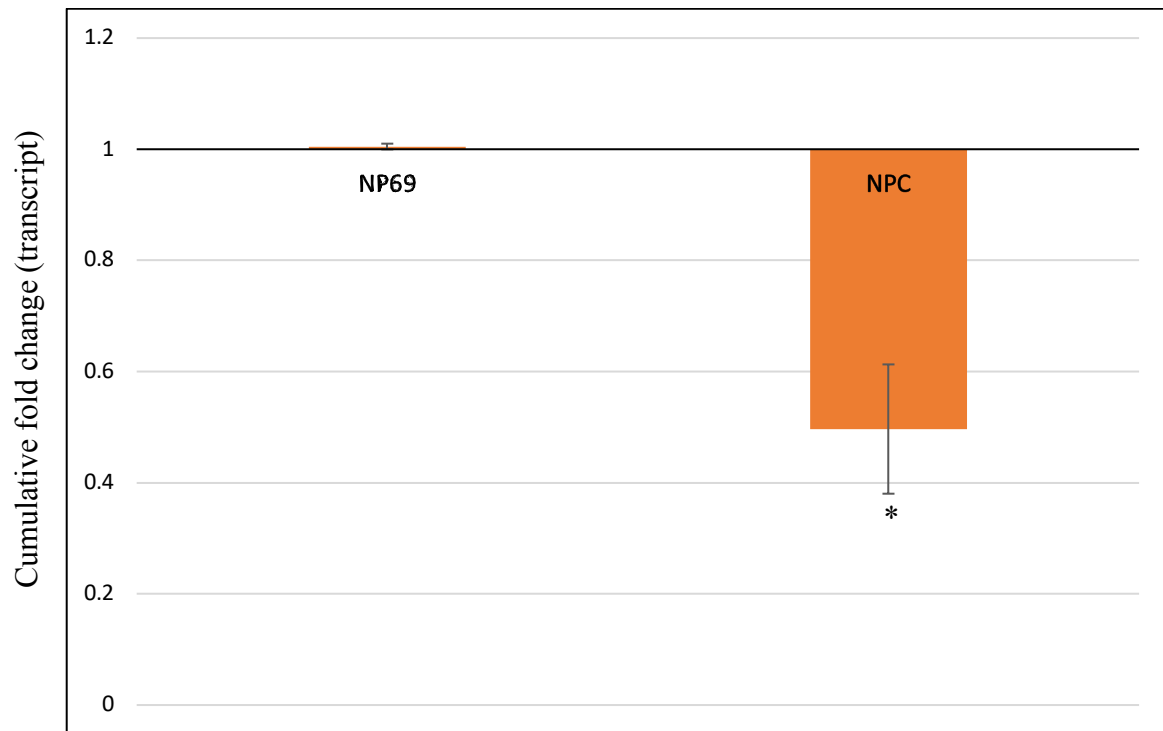


Figure 4.3: Cumulative fold difference of *uS19* expression in overall NPC cell lines versus NPE, NP69. All 6 NPC cell lines were compared to the NPE cell line (NP69) respectively. The y- and x- axis represents the cumulative fold change and cell lines respectively. The asterisk symbol (*) represents p-value < 0.05. The error bars were included and represents the standard deviation of the expression level of *uS19* in NP69 cell line and average SD for 5 NPC cell lines.

Based on the transcription analysis, significant underexpression of *uS19* between NPC and NPE cell lines were observed. A noticeable trend of down regulation of *uS19* was observed across NPC cell types. Following the transcriptional analysis, western blot analysis of *uS19* was carried out in three biological replicates. The protein expression of *uS19* showed a differential expression trend in NPC compared to NP69 cell line (Figure 4.4). Beta actin was used as a reference protein and loading control to obtain the normalization factors. The western blot band intensity values for *uS19* and beta actin in each cell lines was measured using ImageQuant TL instrument (GE Healthcare, United Kingdom). The band intensity obtained by beta actin was consistent throughout the cell lines, with range of 53000-56000

units, varying around 4000 units. The consistency of the band intensity from beta actin indicates that differences of more than 4000 units indicates difference in the protein expression among NP69 to NPC cell lines. The average band intensity for beta actin and uS19 protein were calculated and tabulated in Table 4.6.

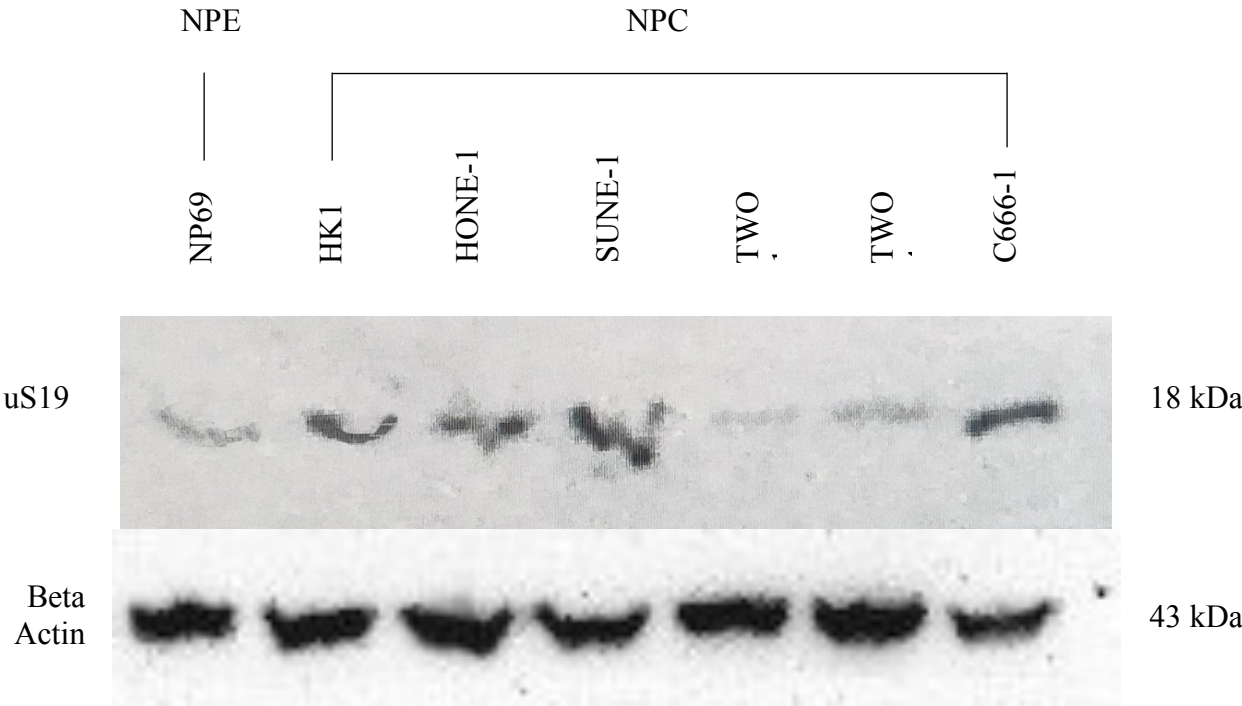


Figure 4.4: Western blot analysis of uS19 protein expression in NPC cell lines with NPE cell line as normal and beta actin as reference and loading control used in normalization.

Based on the average band intensities in Table 4.6, uS19 protein is overexpressed in all NPC cell lines compared to NP69. The normalized values were calculated using the formula mentioned in Materials and Methods and shown in Table 4.7. The fold changes of uS19 protein in Table 4.8 were calculated by normalizing to NP69 to obtain the fold changes in NPC cell lines compared to normal, where values >1.0 indicates overexpression while values

<1.0 represents underexpression. The average fold change along the standard deviation between each biological replicates are tabulated in Table 4.9 and displayed in Figure 4.5 revealed an overexpression of uS19 protein in NPC cell lines at least 2-7X higher than in NP69 cell line. Student *t*-test with a p-value <0.05 for each cell line against NP69 were calculated and shown in Table 4.9. The p-value calculated shows a significant overexpression of uS19 in NPC cell line when compared to the NPE cell line, NP69. The cumulative fold change of uS19 in NPC cell lines were calculated and shown in Table 4.10 and presented in Figure 4.6. A student *t*-test was done for all NPC cell lines compared to NPE cell line, NP69. The statistical analysis showed the overexpression of uS19 protein in NPC cell lines is significant as shown in Table 4.10.

Table 4.6: Average band intensity of beta actin and uS19 protein

Cell line	Average $\pm$ SD	
	Beta Actin	uS19
NP69	54250.013 $\pm$ 6349.895	1045.4 $\pm$ 37.906
HK1	55099.627 $\pm$ 4875.445	2722.123 $\pm$ 257.473
HONE-1	53264.123 $\pm$ 7883.546	4985.653 $\pm$ 116.004
SUNE-1	54099.683 $\pm$ 6391.876	4484.617 $\pm$ 446.422
TWO 1	54258.633 $\pm$ 7818.592	1167.047 $\pm$ 12.035
TWO 4	55837.013 $\pm$ 5173.044	2192.38 $\pm$ 41.492
C666-1	53180.283 $\pm$ 9778.573	8131.69 $\pm$ 782.976

Table 4.7: Normalized band intensities of uS19 to beta actin protein

Cell line	uS19		
	BR1	BR2	BR3
NP69	1088.86	1019.17	1028.17
HK1	3183.749	2553.828	2593.048
HONE-1	4813.998	4866.599	4959.057
SUNE-1	4980.071	4209.049	4222.242
TWO 1	1120.833	1189.936	1183.06
TWO 4	2369.339	2263.986	2153.008
C666-1	8064.869	7990.903	7616.576

Table 4.8: Fold change of uS19 protein expression in NPC cell lines

Cell line	uS19		
	BR1	BR2	BR3
NP69	1.000	1.000	1.000
HK1	2.924	2.506	2.522
HONE-1	4.421	4.775	4.823
SUNE-1	4.574	4.130	4.107
TWO 1	1.029	1.168	1.151
TWO 4	2.176	2.221	2.094
C666-1	7.407	7.841	7.408

Table 4.9: Average fold change of uS19 expression in NPC cell lines

Cell line	Average fold change	SD	p-value
NP69	1.000	0.000	
HK1	2.651	0.237	0.003
HONE-1	4.673	0.220	0.001
SUNE-1	4.270	0.263	0.001
TWO 1	1.116	0.0754	0.028
TWO 4	2.164	0.065	0.001
C666-1	7.552	0.250	0.0002

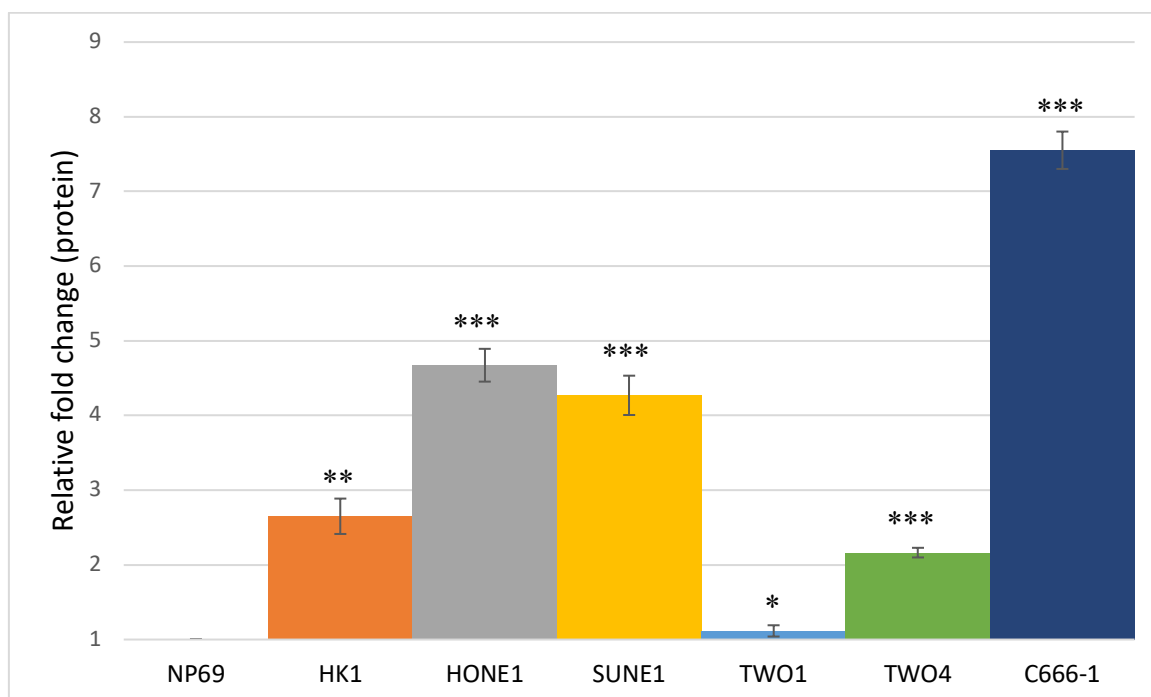


Figure 4.5: Relative fold change of uS19 expression in NPC cell lines versus NPE cell line. The y- and x- axis represents the relative fold change and cell lines respectively. Data from NPC cell lines were compared to the NPE cell line (NP69) respectively. The asterisk symbol (*) represents p-value <0.05 (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). The error bars were included and represents the standard deviation of the expression level of uS19 protein in 6 NPC cell line.

Table 4.10: Cumulative fold change of uS19 expression in NPC compared to NP69 cell line

Cell line	Average fold difference	SD	p-value
NP69	1.000	0.000	
NPC	3.738	2.291	0.022

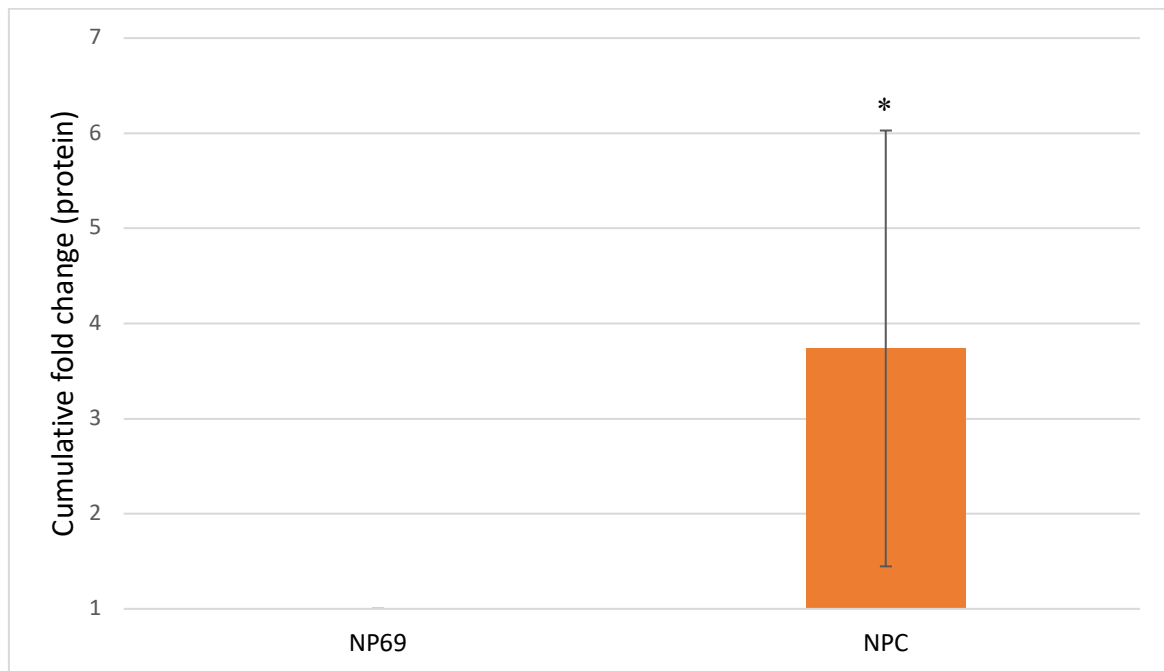


Figure 4.6: Cumulative fold difference of uS19 protein expression in overall NPC cell lines compared to NPE, NP69. All 6 NPC cell lines were compared to the NPE cell line (NP69) respectively. The y- and x- axis represents the cumulative fold change and cell lines respectively. The asterisk symbol (*) represents p-value <0.05. The error bars were included and represents the average standard deviation of the expression level of uS19 protein in 6 NPC cell lines.

4.3 Discussion

From the transcriptional expression analysis, a significant underexpression was observed in *uS19* gene in all 5 NPC cell lines when compared to NP69 cell line. This, however, contradicts previous studies done by Fang et al (2008), where they observed an overexpression of *uS19* gene in poorly differentiated squamous cell carcinoma that was obtained from primary NPC tissue sample. In their study, microarray analysis was done to obtain an abundant number of genes. Through statistical and bioinformatics analysis, they discovered 435 genes to be upregulated and *uS19* gene was among the top 70 upregulated genes with a fold difference of 6.91 (Fang et al., 2008). Though they discovered that *uS19* gene to be overexpressed, further analysis using conventional PCR or qPCR was not done to further validate its expression in the tissue sample. Furthermore, preliminary assessment

from our previous study (Sim et al., 2018) shows the expression of *uS19* gene to also be upregulated in NPC-HK1 cell line when using Reverse Transcription PCR (RT-PCR) but it was not statically significant (refer to Appendix).

Kothapalli et al. (2002) revealed an inconsistent sequence accuracy of spotted cDNA microarrays, where it showed an inaccuracy of differential expression data between differentially expressed genes spotted on the microarray that were excised and used as probes in Northern blots. Besides that, the low specificity of cDNA microarray probes are common in microarray analysis due to the probes having to closely identical DNA sequences (Kothapalli et al., 2002). Their studies suggests further validation is required using techniques such as semi-quantitative RT-PCR or real-time PCR (qPCR) due to the fact that microarray is hybridization-based and therefore the orders of magnitude are limited (Kothapalli et al., 2002). The upregulation of *uS19* gene shown by Fang et al. (2008) requires further validation whereas the upregulation of *uS19* gene showed by Sim et al. (2018) was using RT-PCR. Semi-quantitative RT-PCR is a conventional method that uses agarose gel for the detection of PCR amplification at the final phase or end-point of the PCR reaction. Data obtained using qPCR is much more reliable due to its early detection of PCR amplification during the early phases of the reaction. Furthermore, our previous study was only done in technical duplicates. Biological replicates are crucial in investigating the significance of an expression data that allows a consistent observation of expression throughout the same type of sample, whereas technical replicates are to minimize human error throughout the investigation. Although one of the data obtained in this study was looking at the same cell line as Sim et al. (2018), the lack of biological replicates along with the use of conventional PCR could be the cause in the difference in data obtained. While the

results obtained by Fang and his colleagues showed an upregulation of *uS19* mRNA expression, they were looking at poorly differentiated squamous cell carcinoma obtained from a NPC tissue sample. The cell lines used in this study are established cell lines that possess specific characteristic to represents different NPC cell types. This could be one of the reasons for the different expression of *uS19* gene between our study and theirs. Tissue samples are heterogeneous, which means that there is variation in growth rate and pattern between samples. In contrast, cell lines are of clonal lineage and have a higher homogeneity, resulting in a uniform growth rate and pattern. Cell lines used in this study are established and used in numerous studies resulting in a substantial amount of data that help in understanding the roles of ribosomal proteins. Cell lines with different storage period and low passage number were used in this study to ensure that the cells maintain their lineage. These cell lines are sufficient enough for preliminary studies in obtaining the expression analysis to identify possible roles of ribosomal proteins, (in this case *uS19*) in NPC progression.

As shown in Figure 4.2, *uS19* gene appears to be showing a downregulated pattern across the different cell types. The expression pattern is more in a downregulated pattern as the differentiation level of the cell increases. The underexpression is highest in TWO 1 followed by HK1, which are a Type I NPC and keratinized and well-differentiated cell line. This is followed by HONE-1, a Type IIa, then TWO 4 and C666-1, both a Type IIb NPC. Hence the data shows the more differentiated the cell line is, the greater the underexpression level of *uS19* gene. The difference between Type I and Type II is based on the keratinizing of the cell due to the infiltration of human papillomavirus (HPV) and Epstein-Barr virus (EBV) infection that leads to the non-keratinizing of the cell (Broccolo et al., 2018). The

differentiation of the cells between the Type IIa and Type IIb shows how far it differentiates away from its origin. This study shows the expression of *uS19* gene is downregulated in the better differentiated NPC cell types.

Gene expression plays a crucial role in the morphological changes of the cells. The same gene could play different roles and be expressed at different levels in cells of different differentiation stages. Studies conducted by Trapnell et al. (2010) revealed changes in isoform dominance upon cell differentiation, where many genes switched between major and minor transcripts. In a different study conducted by Owen et al. (1991), they observed that some genes not expressed during the beginning of the proliferation period were then expressed significantly later on in the proliferation period. Hence the differentiation of cells could play a role in the expression of the genes, which could cause changes in the expression pattern of genes. Furthermore, *uS19* gene was shown to be expressed during different differentiation phase of liver regeneration and also presented a non-consistent expression pattern across liver cancer tissue samples. (Kitagawa et al., 1991; Wong et al., 2014). Changes in the protein expression pattern of *uS19* gene could reveal the involvement of this gene in the tumorigenesis of NPC. However the mechanism and pathway involving other genes and proteins remain unclear. Hence, protein expression analysis was done to further understand the role of *uS19*.

The *uS19* protein in this study was significantly overexpressed in NPC cell lines compared to NPE cell line (Figure 4.5). However, the fold difference obtained for TW 01 was less than 2-fold which is considered not so significant. Nevertheless, the protein analysis obtained showed an overexpression. This is in contrast to underexpression seen at the transcriptional

level. However, it is not unusual for mRNA and protein expression to differ as modes of regulation during the translation process might affect the levels of proteins. Vogel and Marcotte (2013) conducted a study on the comparison between mRNA and protein expression level and suggested that the regulation at the level of mRNA acts as a switch, where the regulation downstream functions as a rheostat for further tuning of protein abundance. This theory was supported by Eswaran et al. (2013) where they showed intron retention affecting expression-regulating mechanisms by regulating the expression of functional wild type proteins in cancer cells compared to normal cells. Prasanth et al. (2005) showed that some RNA molecules that translate to proteins are usually in low quantities and are not needed unless required during physiological stress. Their studies demonstrated that when the cells were treated with α -amanitin, a RNA polymerase II inhibitor, the expression of the CTN-RNA was reduced drastically compared to untreated (control). However upon cellular stress, the protein expression increased revealing that the translation-competent mRNA was cleaved during post-transcriptional (Prasanth et al., 2005). Further investigation in intron retention in cancer reveals its association with cancer and diseases due to the different splicing of genes. Intron retention is often observed as potential cancer biomarkers due to its ability to suppress gene expression leading to cancer (Ren et al., 1995; Dvinge & Bradley, 2015). This theory was validated by Braunschweig et al. (2014), where they reported the relationship between intron retention with lower gene expression. Intron retention leads to alternative splicing in cancer cells that favors mechanisms that assist in malignancy (Eswaran et al., 2013). Studies done by Ren et al. (2012) showed that intron retention in prostate cancer showed deregulation in cancer compared to normal cells, thus linking it to the causal role of tumorigenesis. Typically, intron retention is due to mutation due to cellular stress exhibited in most tumor formation leading to alternate splicing of

specific genes. However, Dvinge & Bradley (2015) found intron retention in cancer tissue in comparison to normal tissue without mutation towards the splicing mechanism. This could explain the differential expression observed between the mRNA and protein expression levels seen in this study.

Based on the results from the protein analysis, uS19 protein expression level was observed to be the highest in C666-1 cell line. Interestingly, C666-1 is the only cell line that possesses the EBV antigens that are reported to be one of the main causes of NPC to this date. Many ribosomal protein genes have been reported to interact with the EBV pathways that resulted in the formation of NPC. As stated by Hildesheim and Wang (2012), EBV infections are not sufficient to induce cancer on its own, implying that other cofactors are required for NPC pathogenesis. Multiple studies have shown ribosomal protein genes to be associated with EBV and is crucial for its survival. For example, an increase in RPL4 expression allows the survival of EBNA1 in EBV-related malignancies. Also, uL22 interaction with EBER-1 protecting the cells from EBV transformation (Elia et al., 2004; Shen et al., 2016). In silico approach has been done by Sim and Talwar (2019) showing multiple ribosomal proteins interacting with EBV proteins. They found 27 human ribosomal proteins interacting with EBV proteins (EBNA1, LMP1, LMP2A and LMP2B). Although no further validation using molecular methods such as pull-down, this reveals possible interactions of ribosomal proteins with the EBV protein that could lead to NPC formation. Based on the protein analysis results, uS19 protein could possibly be interacting with one or more of the EBV proteins which sustains the pathogenicity in NPC through EBV pathway. Hence, further analysis through bioinformatics analysis could give a promising insight into the possible protein interaction partners of the uS19 protein.

CHAPTER 5

BIOINFORMATICS ANALYSIS

5.1 Introduction

Protein-protein interaction (PPI) analysis is crucial in understanding cellular processes and functions in pathways that assist in the development of diseases and cancers. While ribosomal proteins function in regulating essential cellular functions, most ribosomal proteins have shown to possess extra ribosomal functions that assist in maintaining and causing carcinogenicity of the cell. The uS19 has been known to be associated in a few pathways, namely the LRRK2 pathway, Mdm2-p53-MdmX pathway and Fragile X Mental Retardation Protein 1 (Daftuar et al., 2013; Martin et al., 2014; Sim et al., 2016). It has been known to interact with p53 (a tumor suppressor) although no studies were done verifying its interactions in nasopharyngeal carcinoma (NPC) but making it a relevant suspect in NPC development as p53 is known to be associated in NPC apoptosis pathway (Tulalamba & Janvilisri, 2012). The expression analysis conducted in this study has shown uS19 to be differentially expressed in tumor cell lines compared to a normal cell line. The uS19 protein is highly expressed in an Epstein-Barr virus (EBV) known cell line, C666-1. This raises the question of whether uS19 interacts with any known EBV antigens that could be linked to NPC formation. There is no known literature report of uS19's direct association with NPC. Nevertheless this study suggests a possible involvement of uS19 based on its expression data when compared to normal cell line. Previous studies have shown ribosomal protein genes interacting with EBV antigens by initiating, maintaining or inhibiting EBV infections in NPC. This study aims to identify potential candidate of any cancer-causing proteins associating with uS19 that can link to NPC tumorigenesis. *In silico* approach via

bioinformatics has assisted in many research pertaining to molecular pathways and functions associated with tumor progression and diseases. Although identification of PPI through bioinformatics only predicts interactions and requires further molecular techniques to validate the interactions, nevertheless, this approach helps predict which proteins might be linked to tumorigenesis and provide information regarding their extra ribosomal functions that may help in early diagnosis. Thus uS19 was used as a receptor to identify potential interacting proteins that could be linked to any known NPC-causing proteins via an *in silico* approach.

5.2 Results

A 3D protein model was built using the FASTA sequence of human uS19 (accession no. NP_001295155.1) retrieved from National Centre for Biotechnology Information (NCBI). It was generated using SWISS-MODEL Workspace server manually and then further refined using ModRefiner. The model was then viewed using Pymol 2.3 as presented in Figure 5.1, where the N-terminus and C-terminus were colour coded in blue and red respectively. The 3D model constructed underwent quality check using PROCHECK, ERRAT, VERIFY3D, QMEAN6 and RMSD value (Table 5.1). The RMSD values represents the quality and stability of the protein complexes in which the lower the RMSD value (less than 1 Å) the more stable the protein structure. The RMSD value of uS19 protein model reveals that the model is stable and is at the best possible quality as it has a value of 0.577 Å, which is less than 1 Å.



Figure 5.1: Ribbon representation of uS19 protein viewed using Pymol molecular graphics system (version 2.3) Each is coloured from N-terminus (blue) to C-terminus (red) as a spectrum.

Table 5.1: Quality estimations of uS19 model

Model Quality Evaluations	uS19
RAMPAGE	
Residues in favoured region (%)	89.43
Residues in allowed region (%)	13.1
Residues in outlier region (%)	4.07
ERRAT	76.58
Verify 3D (%)	68.0
QMean6	-1.63
RMSD (Å)	0.577

The 3D model structure was then further analysis using VAST revealing one human structural neighbour protein, 3KQ4, a transport protein bound to Cubilin. The results obtained was then input to IntAct to identify its possible interactions , which revealed an

interaction with 10 proteins listed in Table 5.2. These 10 proteins are possible interaction partners with uS19 protein.

Table 5.2: Predicted interaction partners with uS19 protein.

Accession	Identifier	Description	Pdb ID
EBI-10982110	Q96Q45-2	Transmembrane protein 237 isoform A (TMEM237)	3lo8
EBI-12081840	A1A5C7-2	Solute carrier family 22 member 23 (SLC22A23)	4dmu
EBI-12808018	Q9UKG4	Solute carrier family 13 member 4 (SLC13A4)	4f35
EBI-351896	P11142	Heat shock cognate 71 kDa protein (HSPA8)	-
EBI-3953632	O60494	Cubilin (CUBN)	-
EBI-3953638	P27352	Cobalamin binding intrinsic factor (Cobalamin)	-
EBI-4289564	P30825	Cationic amino acid transporter (SLC7A1)	3k2k
EBI-5280110	Q14684-1	rRNA processing protein 1 homolog B (RRP1B)	4wj2
EBI-7181113	P13285	Latent membrane protein 2 (LMP2)	1uxs
EBI-996522	P03211	EBNA1	5wmf

Out of the 10 proteins identified, three proteins did not have an individual protein structure model nor a pdbID, hence the structures were generated using the same format in building the uS19 protein model using SWISS-Model Workspace and then refined using ModRefiner and its quality assessment were checked. In addition, LMP2 and EBNA1 models were built due to the unavailability of individual structure as the structure provided were in a cluster. The FASTA information used to build these models were obtained from the website linked to its description. The information was then used to build each individual structure. Table 5.3 shows the quality estimates for the protein models built. The root mean square deviation

(RMSD) values for each protein built is listed in Table 5.3. The RMSD values obtained that are less than 1 Å show that the protein structures built are stable. Figure 5.2 shows the ribbon protein structure of the five proteins structure predicted to interact with uS19 protein model that was built viewed using Pymol molecular graphics system (version 2.3). The N-terminus and C-terminus are colour coded in blue and red respectively.

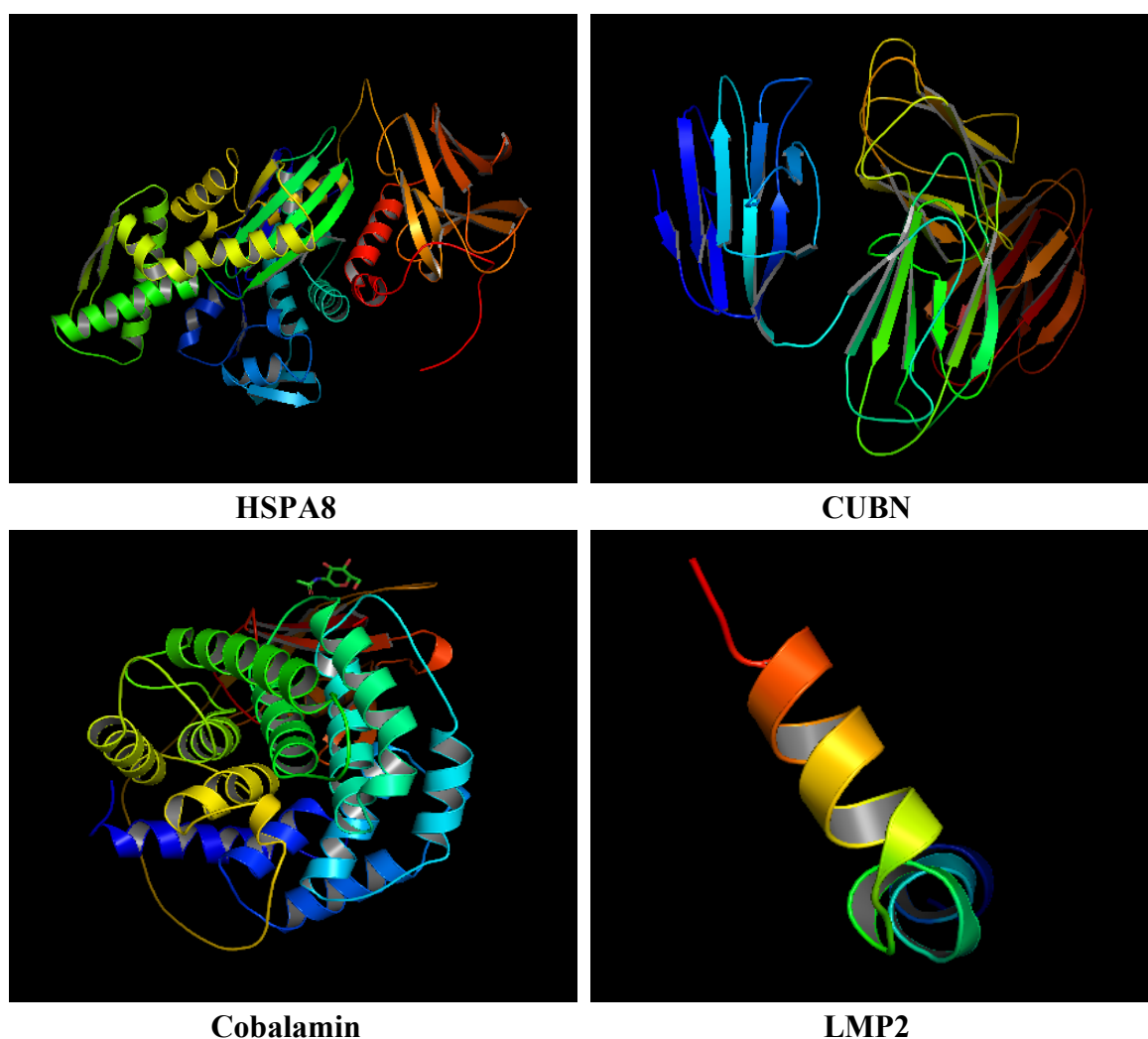
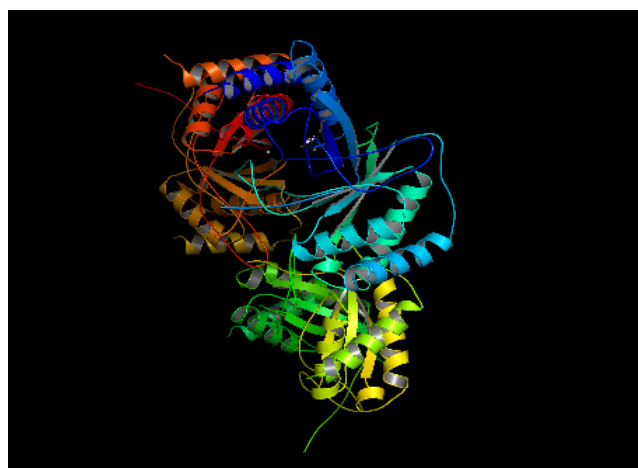


Figure 5.2: Ribbon representation of the five predicted proteins viewed using Pymol 2.3. Each is coloured from N-terminus (blue) to C-terminus (red) as a spectrum.



EBNA1

Figure 5.2 continued

Table 5.3: Quality estimates for selected protein models interacting with uS19 protein

Model Quality Evaluations	Proteins Models (Identifiers)				
	HSPA8	CUBN	Cobalamin	LMP2	EBNA1
RAMPAGE					
Residues in favoured region (%)	92.57	89.01	92.82	92.0	98.52
Residues in allowed region (%)	9.9	12.2	8.7	8.3	4.1
Residues in outlier region (%)	1.81	12.56	12.87	0.0	0.62
ERRAT	96.8	75.0	90.33	78.95	91.82
Verify 3D (%)	91.16	97.37	80.36	0.00	89.68
QMean6	0.07	-1.80	-2.13	-1.60	0.82
RMSD (Å)	1.185	0.373	0.652	0.097	0.527

Docking analysis of uS19 using the two docking system was performed against the ten predicted interacting proteins. ClusPro was first performed. The ClusPro results were obtained based on four categories (balanced, electrostatic-favoured, and Van der Waals combined with electrostatics) based on the weighting of the interactions calculated. The docked clusters with the most number of cluster size were chosen to be further analysed

using the second docking system, PatchDock (Refer to Appendix, Table A.4). For the analysis using PatchDock (with a default set-up), the pdb files obtained from ClusPro was entered to obtain the scores. The scoring function considers both geometric fit and atomic desolvation energy to evaluate each candidate transformation. The top 10 structures were further subjected through high-throughput refinement and scoring by FireDock (Refer to Appendix, Table A.5 & Table A.6). The RMSD values for each docked complex using both the docking analyses were obtained using SPDBV 4.1.1 after superimposition. The RMSD values represents the quality and stability of the dock complexes in which the lower the RMSD value, the more stable the complexes are and has a high likeliness to interact. An RMSD value of less than 1.5 Å is considered to have the highest possibility of complex formation. The information obtained for all ten docked complexes with uS19 protein are as tabulated below according to its uS19-docked complexes. The interface residues of each docked complex were obtained using SPDBV 4.1.1. The complexes with the most number of interfacial residues were selected to portray the uS19-protein complex. Based on the maximum number of contact residues of the interfacial region of all 10 docked complexes, each docking complex with uS19 was then submitted to Protein Interaction Calculator (PIC). This was to assess its interface residues involving in stabilizing the complex via hydrophobic interactions and ionic interactions within 5 Å.

Table 5.4: ClusPro and PatchDock/FireDock energy profile and RMSD values

Dock model	ClusPro			PatchDock/FireDock		
	Cluster size	Lowest free energy (kcal/mol)	RMSD (Å)	Desolvation energy (kcal/mol)	Global free energy (kcal/mol)	RMSD (Å)
uS19-3lo8	78	-697.3	0.90	0.65	11.57	0.00
uS19-4dmu	55	-1276.0	0.86	-0.93	-0.34	0.00
uS19-4f35	66	-1387.8	0.91	-2.65	5.45	0.00
uS19-P11142	63	-794.3	0.98	5.89	2.74	0.00
uS19-O60494	232	-317.1	0.94	14.71	-12.03	0.00
uS19-P27352	109	-902.5	0.90	1.20	-20.41	0.00
uS19-3k2k	68	-1010.0	0.93	5.63	10.36	0.00
uS19-4wj2	132	-779.1	0.93	-1.06	-4.73	0.00
uS19-LMP2	184	-1111.4	0.98	13.29	-1.50	0.00
uS19-EBNA1	59	-938.1	0.89	0.39	13.15	0.00

Table 5.4 shows the summarized energy profile and RMSD obtained from both docking analyses. The energy profiles obtained from both analyses revealed a fair to good energy profile, while the RMSD values for both docked models reveals uS19 did not change its conformation after docking. Both sets of data indicates stability of each docked complex.

uS19-TMEM237 docked complex

The uS19-TMEM237 complex generated by ClusPro was selected from the balanced category on the basis of higher cluster size and lowest free energy, which were 78 and -697.3 kcal/mol respectively (Refer to Appendix, Table A.5). The top ten solutions predicted and refined through PatchDock/FireDock revealed desolvation energy of 0.65 kcal/mol and global free energy of 11.57 kcal/mol for the best docked conformation (Refer to Appendix,

Table A.7). The energy profile from both analyses obtained revealed a fair energy profile, however the ClusPro binding mode demonstrated high probability for uS19 and TMEM237. The RMSD values for both docked models were 0.90 Å and 0.00 Å for ClusPro and PatchDock respectively. This indicates that the conformational changes occurred during binding were relatively low, hence showing reliability for subsequent investigation. Interface residues of the complex listed in Table 5.5 were further analysed to identify residues in the interaction which were within 5 Å of each other. Based on the number of interacting residues, the PatchDock model was selected as the most likely conformation to portray the uS19-3lo8 complex. The potential interface contact residues of uS19 to TMEM237 protein obtained from SPDBV 4.1.1 and the ribbon representation of the uS19-TMEM237 protein docked complex obtained from PatchDock is displayed using Pymol molecular graphics system (version 2.3) in Figure 5.3. PIC evaluation results of protein interactions within the complex to assess the interface residues involved in stabilizing the complex via hydrophobic interactions and ionic interactions are as listed in Table 5.6.

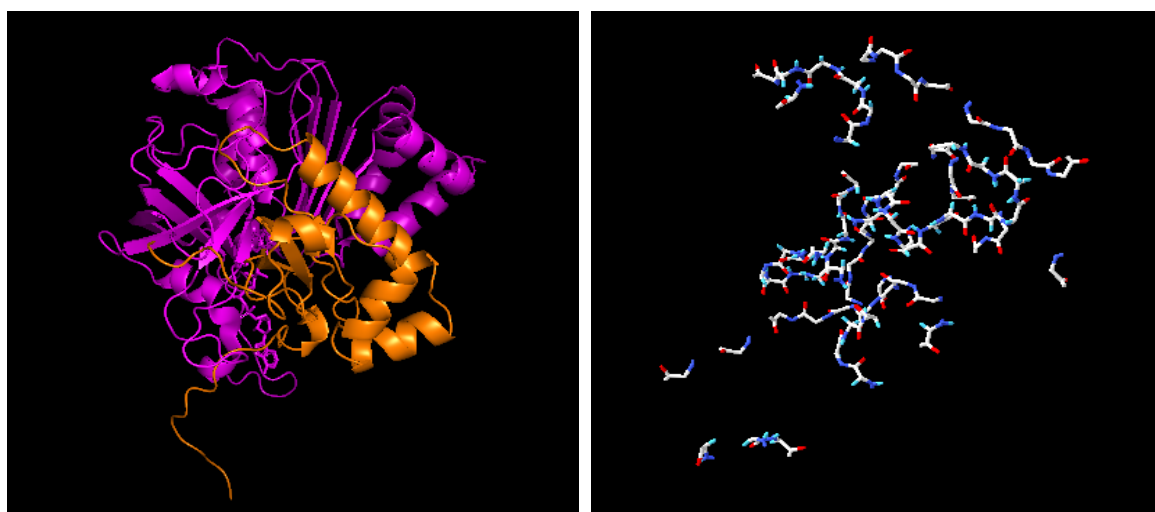


Figure 5.3: Ribbon representation of uS19-TMEM237 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by TMEM237. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.5: Probable interface residues of uS19 and TMEM237 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5 Å)
uS19	ClusPro	ARG88, ASP89, ASN105, LYS107, THR108, PHE109, ASN110, GLN111, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, HIS135, ARG137, PRO138, GLY139, ILE140, GLY141,
	PatchDock	ARG17, ARG66, GLU73, ALA74, PRO75, PRO76, GLU78, LYS79, PRO80, GLU81, VAL82, LYS84, HIS86, GLY98, MET100, TYR104, LYS107, THR108, PHE109, ASN110, GLN111, VAL112, GLU113, ILE114, LYS115, MET118, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, LYS134
TMEM237	ClusPro	GLU42, SER43, GLY46, PRO47, LYS48, ALA49, PRO50, GLY51, GLU52, ARG114, ARG115, ALA116, VAL117, TYR118, TYR119, ASP120, PRO121, GLU122, THR123, GLY124, LYS125, GLU126, LYS130, SER135, ASN136, CYS139, CYS139, ASN140, LYS142, LEU276, LYS277, MET280, PRO281, GLN284, LEU302, LYS306, TRP311, VAL131, TYR316
	PatchDock	GLU22, PRO23, LEU25, ASN26, THR27, TYR28, PRO30, LYS31, GLU32, GLY73, VAL74, GLU79, ASN80, PRO81, LYS82, LYS83, GLN88, ASN89, VAL90, ARG91, LEU92, ARG114, ALA116, PRO128, SER129, LYS130, ASN131, GLY132, VAL133, THR150, GLY151, SER153, GLY154, LYS155, ILE156, LYS277, GLU314, VAL315, TYR316, OXT316

Table 5.6: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-3lo8 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	75	PRO	uS19	25	LEU	TMEM237
	82	VAL	uS19	92	LEU	TMEM237
	109	PHE	uS19	316	TYR	TMEM237
	109	PHE	uS19	92	LEU	TMEM237
	128	ILE	uS19	81	PRO	TMEM237
	130	TYR	uS19	81	PRO	TMEM237
	132	PRO	uS19	133	VAL	TMEM237
	132	PRO	uS19	81	PRO	TMEM237
Ionic	17	ARG	uS19	32	GLU	TMEM237
	73	GLU	uS19	155	LYS	TMEM237
	81	GLU	uS19	155	LYS	TMEM237
	84	LYS	uS19	314	GLU	TMEM237
	113	GLU	uS19	29	LYS	TMEM237
	113	GLU	uS19	31	LYS	TMEM237

uS19-SLC22A23 docked complex

From the docked conformation generated by ClusPro, a potential uS19-SLC22A23 complex was selected based on the balanced category on the basis of higher cluster size and lowest free energy, which were 55 and -1276.0 kcal/mol respectively (Refer to Appendix, Table A.5). The top ten solutions predicted and refined through PatchDock/FireDock reveals desolvation energy of -0.93 kcal/mol and global free energy of -0.34 kcal/mol for the best docked conformation (Refer to Appendix, Table A.7). The energy profiles obtained from both analyses indicates high probability for uS19-SLC22A23 interaction. The RMSD values for both docked models were 0.86 Å and 0.00 Å for ClusPro and PatchDock respectively indicating reliability for subsequent investigation. Interface analysis of the complex listed in Table 5.7 was done to identify interface residues within the interactions. PatchDock model was selected based on the most number of interface residues in the uS19- SLC22A23 complex. The potential interface contact residues of uS19 to SLC22A23 protein obtained from SPDBV 4.1.1 and the ribbon representation of the uS19- SLC22A23 protein docked complex obtained from PatchDock are displayed using Pymol molecular graphics system (version 2.3) in Figure 5.4. PIC evaluation results of protein interactions within the complex to assess the interface residues involved in stabilizing the complex via hydrophobic interactions and ionic interactions are as shown in Table 5.8.

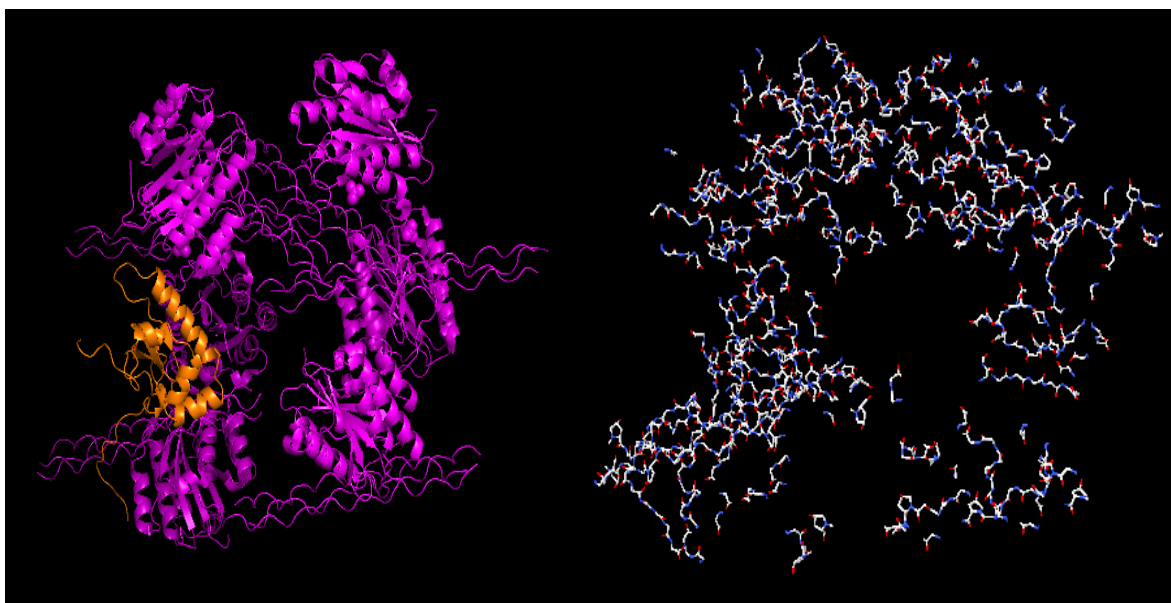


Figure 5.4: Ribbon representation of uS19- SLC22A23 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by SLC22A23. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.7: Possible interface residues of uS19 and SLC22A23 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5 Å)
uS19	ClusPro	ARG17, PHE19, ARG20, LYS21, PHE22, TYR24, LEU43, TYR44, SER45, ALA46, ARG47, GLN48, ARG49, ARG50, ARG54, ARG88, ASP89, VAL103, ASN105, LYS107, THR108, ASN110, GLN111, VAL112, LYS115, PRO116, GLU117, MET118, GLY120, HIS121, TYR122, GLY124, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, LYS134, HIS135, GLY136, ARG137
	PatchDock	ARG51, ARG54, GLY55, LEU56, ARG57, ARG58, LYS59, GLN60, HIS61, SER62, LEU63, LEU64, LYS65, ARG66, LYS69, ALA70, LYS72, GLU73, ALA74, PRO75, PRO76, LYS79, PRO80, GLU81, LYS84, THR85, HIS86, LEU87, ARG88, ASP89, MET90, TYR104, ASN105, GLY106, LYS107, THR108, PHE109, GLY124, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140, GLY141, OXT142

Table 5.6 continued

SLC22A23	ClusPro	GLN14, GLY16, VAL17, MET18, GLY19, PHE20, GLY22, PRO23, GLY110, PRO111, ARG112, GLY210, PRO211, ARG212, GLN214, GLY216, MET218, SER1699, SER1700, SER1701, GLN1734, SER1737, ILE1738, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, VAL1747, HIS1753, LEU1757, VAL1760, MET1761, GLN1762, GLU1764, GLY1766, PRO1767, SER1768, GLN1769, ASP1772, PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786, ASP1802, VAL1803, SER1804, VAL1805, ASP1806, SER1807, ASP1809, ALA1810, ARG1830, GLY14, VAL17, MET18, GLY19, PHE20, GLY22, PRO23, GLY110, ARG112, PHE120, GLY210, PRO211, ARG212, GLY213, GLN214, GLY216, MET218, GLN1734, ILE1738, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, VAL1747, LEU1757, VAL1760, MET1761, GLN1762, ARG1763, GLU1764, PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786, ASP1845, SER1846, VAL1848, VAL1849, LYS1850, LEU1851, GLN1852, ARG1853, GLY1855, ASP1856, THR1859, MET1860, LEU1863, GLY1864, ASN1865, SER1866, PHE1867, GLY4, PRO5, GLY7, PRO8, GLY10, ARG12, GLN14, VAL17, MET18, PHE20, GLY22, PRO23, PRO26, GLY110, PRO111, ARG112, GLY113, GLN114, GLY116, VAL117, MET118, GLY119, PHE120, GLY122, PRO123, GLY125, PRO126, GLY201, PRO202, GLY210, PRO121, ARG122, GLN214, GLY216, VAL217, MET218, PHE220, GLY222, PRO223, GLY225, PRO226, PRO1725, ARG1726, GLY1734, ILE1738, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, VAL1748, GLU1750, HIS1753, LEU1757, VAL1760, MET1761, GLN1762, GLU1764, PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786, PRO1790, SER1817, ASN1818, ARG1819, PRO2, PRO5, GLY7, PRO8, GLY10, ARG12, GLN14, VAL17, MET18, GLY19, PHE20, GLY22, PRO23, GLY101, PRO102, GLY104, PRO105, GLY110, PRO111, ARG112, GLN114, GLY116, VAL117, MET118, GLY119, PHE120, GLY201, PRO202, GLY210, PRO211, ARG212, GLN214, GLY216, VAL217, MET218, GLN1734, ILE1738, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, ASN1746, VAL1747, VAL1760, MET1761, GLN1762, GLU1764,
-----------------	----------------	--

Table 5.6 continued

	PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786,
PatchDock	HYP9, GLY10, PRO11, ARG12, GLY13, GLN14, HYP15, GLY16, VAL17, MET18, GLY19, PHE20, HYP21, PRO26, HYP27, HYP106, HYP109, GLY110, PRO111, ARG112, GLY113, GLN114, HYP115, GLY116, VAL117, MET118, GLY119, PHE120, HYP121, HYP127, PRO205, HYP206, GLY207, PRO208, HYP209, GLY210, PRO211, ARG212, GLY213, GLN214, HYP215, GLY216, VAL217, MET218, GLY219, PHE220, HYP221, GLY222, PRO223, HYP224, GLY225, HYP227, SER1700, SER1701, LYS1715, SER1737, ILE1738, THR1739, THR1740, ILE1741, ASP1742, PRO1744, LYS1751, ALA1752, HIS1753, LEU1755, SER1756, LEU1757, VAL1758, ASP1759, VAL1760, MET1761, GLN1762, GLU1764, GLY1766, PRO1767, GLN1769, ASP1772, PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786, PRO1790, GLY1791, ASP1802, VAL1803, SER1804, VAL1805, ASP1806, SER1807, ASP1809, HYP15, GLY16, VAL17, MET18, GLY19, PHE20, HYP21, HYP109, GLY110, ARG112, GLN114, HYP115, GLY116, VAL117, MET118, GLY119, PHE120, HYP121, HYP124, HYP127, PRO208, HYP209, GLY210, PRO211, ARG212, GLY213, GLN214, HYP215, MET218, PHE220, PRO223, HYP277, CYS1686, SER1687, GLN1688, PRO1689, SER1719, LYN1720, ALA1721, ASN1722, ILE1723, GLY1724, PRO1725, ARG1726, LEU1727, ILE1738, THR1739, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, ASN1746, PRO1749, GLY1750, LYS1751, ALA1752, LEU1755, LEU1757, VAL1760, MET1761, GLN1762, GLU1764, PHE1776, ARG1779, TYR1780, SER1783, GLY1784, MET1785, HIS1786, PRO1790, GLY1791, ASP1845, SER1846, VAL1848, VAL1849, LYS1850, LEU1851, GLN1852, ARG1853, ASP1856, THR1859, MET1860, VAL1861, THR1862, LEU1863, GLY1864, ASN1865, SER1866, PHE1867, LEU1868, HIS1869, SER1873, HYP15, VAL17, MET18, GLY19, PHE20, HYP21, GLY22, HYP109, ARG112, GLN114, HYP115, GLY116, VAL117, MET118, GLY119, PHE120, HYP121, HYP124, HYP127, PRO208, HYP209, GLY210, PRO211, ARG212, GLY213, GLN214, HYP215, MET218, PHE220, PRO223, HYP277, SER1700, SER1701,

Table 5.6 continued

SER1737, ILE1738, THR1739, THR1740, ILE1741, ASP1742, PRO1744, LEU1757, VAL1760, MET1761, GLN1762, GLU1764, GLY1766, PRO1767, GLY1769, ASP1772, PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786, PRO1790, ASP1802, VAL1803, SER1804, VAL1805, ASP1806, SER1807, ASP1809, ARG183, VAL17, MET18, GLY19, PHE20, HYP21, GLY22, PRO23, HYP109, ARG112, PHE120, HYP127, PRO208, HYP209, GLY210, PRO211, ARG212, GLY213, GLN214, HYP215, MET218, HYP227, ILE1738, THR1739, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, LEU1757, VAL1760, MET1761, GLN1762, GLU1764, PHE1776, ARG1779, TYR1780, SER1783, GLU1784, MET1785, HIS1786, PRO1790, GLY1791, ASP1845, SER1846, VAL1848, VAL1849, LYS1850, LEU1851, GLN1852, ARG1853, ASP1856, THR1859, MET1860, LEU1863, GLY1864, ASN1865, SER1866, PHE1867, HIS1869, HYP3, GLY4, PRO5, HYP6, GLY7, PRO8, HYP9, ARG12, HYP15, GLY16, VAL17, MET18, GLY19, PHE20, HYP21, HYP109, GLY110, PRO111, ARG112, GLN114, HYP115, GLY116, VAL117, MET118, GLY119, PHE120, HYP121, GLY201, PRO202, HYP203, HYP206, PRO208, HYP209, GLY210, PRO211, ARG212, GLY213, GLN214, HYP215, VAL217, MET218, GLN1688, ASN1722, ILE1723, GLY1724, PRO1725, ARG1726, LEU1727, ILE1738, THR1739, THR1740, ILE1741, ASP1742, VAL1743, PRO1744, ASN1746, VAL1747, VAL1748, PRO1749, GLU1750, LYS1751, LEU1757, VAL1760, GLN1762, GLU1764, PHE1776, ARG1779, TYR1780, SER1783, MET1785, HIS1786, HYP3, GLY4, PRO5, HYP6, GLY7, PRO8, HYP9, ARG12, HYP15, VAL17, MET18, GLY19, PHE20, HYP21, HYP109, PRO111, ARG112, GLN114, HYP115, GLY116, VAL117, MET118, GLY119, PHE120, HYP121, GLY201, PRO202, HYP203, HYP206, PRO208, HYP209, GLY210, PRO211, ARG212, GLY213, GLN214, HYP215, VAL217, MET218, ILE1738, THR1739, THR1740, ILE1741, ASP1742, PRO1744, LEU1757, VAL1760, MET1761, GLN1762, GLU1764, PHE1776, ARG1779, TYR1780, SER1783, GLU1784, MET1785, HIS1786,

Table 5.8: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-SLC22A23 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	5	PRO	uS19	217	VAL	SLC22A23
	17	VAL	uS19	1741	ILE	SLC22A23
	17	VAL	uS19	1744	PRO	SLC22A23
	17	VAL	uS19	1780	TYR	SLC22A23
	18	MET	uS19	1785	MET	SLC22A23
	18	MET	uS19	117	VAL	SLC22A23
	18	MET	uS19	1785	MET	SLC22A23
	20	PHE	uS19	1776	PHE	SLC22A23
	20	PHE	uS19	1780	TYR	SLC22A23
	20	PHE	uS19	1741	ILE	SLC22A23
	63	LEU	uS19	1727	LEU	SLC22A23
	87	LEU	uS19	220	PHE	SLC22A23
	90	MET	uS19	220	PHE	SLC22A23
	104	TYR	uS19	217	VAL	SLC22A23
	111	PRO	uS19	111	PRO	SLC22A23
	117	VAL	uS19	117	VAL	SLC22A23
	117	VAL	uS19	118	MET	SLC22A23
	117	VAL	uS19	120	PHE	SLC22A23
	117	VAL	uS19	18	MET	SLC22A23
	118	MET	uS19	117	VAL	SLC22A23
	120	PHE	uS19	1785	MET	SLC22A23
	120	PHE	uS19	117	VAL	SLC22A23
	140	ILE	uS19	1755	LEU	SLC22A23
	208	PRO	uS19	1760	VAL	SLC22A23
	217	VAL	uS19	5	PRO	SLC22A23
	220	PHE	uS19	202	PRO	SLC22A23
	223	PRO	uS19	8	PRO	SLC22A23
	1738	ILE	uS19	1863	LEU	SLC22A23
	1767	PRO	uS19	1851	MET	SLC22A23
	1805	VAL	uS19	1849	VAL	SLC22A23
	1805	VAL	uS19	1851	LEU	SLC22A23
	1805	VAL	uS19	1860	MET	SLC22A23
Ionic	65	LYS	uS19	1750	GLU	SLC22A23
	112	ARG	uS19	1764	GLU	SLC22A23
	112	ARG	uS19	1742	ASP	SLC22A23
	1802	ASP	uS19	1853	ARG	SLC22A23

uS19-SLC13A4 docked complex

The best docked conformation generated by ClusPro, a potential uS19-SLC13A4 complex was selected based on the balanced category on the basis of higher cluster size and lowest free energy which were 66 and -1387.8 kcal/mol respectively (Refer to Appendix, Table

A.5). The top ten solutions predicted and refined through PatchDock/FireDock reveals desolvation energy of -2.65 kcal/mol and global free energy of 5.45 kcal/mol for the best docked conformation (Refer to Appendix, Table A.7). Both complexes showed a high probability of interaction between uS19 and SLC13A4. The RMSD values for both docked models were 0.91 Å and 0.00 Å for ClusPro and PatchDock respectively, signifying reliability for subsequent investigation. Interface analysis of the complex to identify interface residues within the interactions are as tabulated in Table 5.9. PatchDock model was selected based on the most number of contact interface residues to portray the uS19-SLC13A4 docked complex. The potential interface contact residues of uS19 to SLC13A4 protein obtained from SPDBV 4.1.1 and the ribbon representation of the uS19- SLC13A4 protein docked complex obtained from PatchDock is displayed using Pymol molecular graphics system (version 2.3) in Figure 5.5. PIC evaluation results of protein interactions within the complex to evaluate the interface residues involved in stabilizing the complex via hydrophobic interactions and ionic interactions are as shown in Table 5.10.

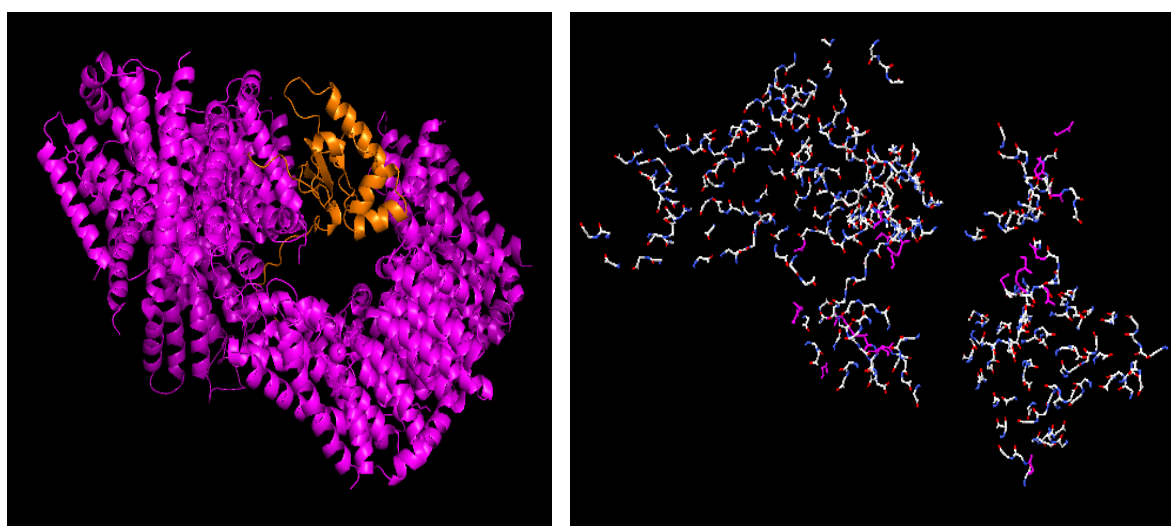


Figure 5.5: Ribbon representation of uS19- SLC13A4 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by SLC13A4. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.9: Possible interface residues of uS19 and SLC13A4 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	PHE19, PHE22, THR23, TYR24, ARG25, GLN42, SER45, ALA46, ARG47, ARG49, ARG50, ARG51, ARG54, ARG57, LYS59, HIS86, LEU87, ARG88, ASP89, ASN105, GLY106, GLN111, VAL112, GLU113, LYS115, GLU117, MET118, HIS121, TYR122, LEU123, GLY124, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133
	PatchDock	ARG17, THR18, PHE19, ARG20, PHE22, THR23, TYR24, ARG25, GLY26, VAL27, ALA46, ARG47, ARG49, ARG50, ARG51, ASN53, ARG54, GLY55, LEU56, ARG57, ARG58, LYS59, HIS61, MET77, GLU78, LYS79, LYS107, THR108, ASN110, GLN111, VAL112, GLU113, LYS115, GLU117, MET118, HIS121, TYR122, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140, GLY141, OXT141
SLC13A4	ClusPro	HIS19, SER22, VAL25, LEU26, LEU64, HIS65, THR67, ILE71, LEU72, VAL75, VAL78, PHE79, GLY81, PHE83, GLU84, THR85, GLN86, LEU89, ASN90, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, LYS117, VAL225, TRP257, LEU285, PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, ILE300, LEU301, SER304, PHE305, ALA306, ARG307, HIS310, TRP311, LYS312, GLN315, ALA318, ASP319, TRP320, GLY321, LEU324, HIS19, ASN21, SER22, LEU23, VAL25, LEU26, ASP28, VAL29, PHE32, LEU33, TYR36, HIS37, PHE38, HIS43, ASN44, LEU47, PHE54, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, PHE80, GLY81, ILE82, PHE83, GLU84, THR85, GLN86, LEU89, ASN90, ASN91, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, LEU285, PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, ILE300, LEU301, SER304, PHE305, ALA306, ARG307, TRP311, GLN315, LYS316, THR317, ALA318, ASP319, TRP320, GLY321, LEU324, SER22, VAL25, LEU26, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, PHE80, GLY81, GLU84, THR85, GLN86, LEU89, ASN90, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, LYS122, GLN128, LEU159, ASP170, PRO254, VAL255, ASN256, TRP257, ASP258, LYS259, VAL263, GLY266, ILE267, GLY269, LEU270, PHE273, LEU274, PHE277, PRO280, ILE281, ASN282, ALA283, ALA284, LEU285, GLY286, PHE288, LYS289,

Table 5.8 continued

	SER290, PHE291, THR293, LEU294, VAL295, LEU297, GLY288, LEU301, MSE302, SER304, PHE305, ALA306, TRP311, GLN315, LYS316, ALA318, ASP319, TRP320, GLY321, LEU324, ALA425, ILE426, PHE428, ALA429, SER430, GLY431, LYS434, LEU18, SER22, VAL25, LEU26, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, PHE80, GLY81, PHE83, GLU84, THR85, GLN86, LEU89, ASN90, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, LEU159, LEU163, LEU285, GLY287, , PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, LEU301, MSE302, SER304, PHE305, TRP311, GLN315, ALA318, ASP319, TRP320, GLY321, LEU324, PHE375PRO418, VAL419, ALA425, PHE428, SER430, GLN435
PatchDock	SER22, VAL25, LEU26, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, GLU84, THR85, GLN86, LEU89, ASN90, ASN91, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, HIS111, HIS112, GLN113, LEU115, LYS117, VAL118, ILE119, ALA120, LYS122, VAL123, LEU124, MSE126, ALA127, GLN128, MSE137, GLY140, VAL141, ALA143, LEU144, MSE147, TRP148, ALA156, LEU159, PRO160, LEU222, ALA225, MSE226, LEU229, ASN256, TRP257, ASP258, LYS259, GLY260, LYS261, VAL262, VAL263, LEU265, GLY266, ILE267, PHE268, GLY269, LEU270, VAL272, PHE273, PHE277, LEU285, PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, ILE300, LEU301, SER304, PHE305, ARG307, VAL308, HIS310, TRP311, LYS312, GLN315, ALA318, ASP319, TRP320, GLY321, LEU324, SER22, VAL25, LEU26, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, PHE80, GLU84, THR85, GLN86, LEU89, ASN90, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, HIS111, HIS112, GLN113, LEU115, LYS117, VAL118, LYS122, LYS259, GLY260, LYS261, VAL262, VAL263, PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, ILE300, LEU301, SER304, PHE305, ALA306, ARG307, VAL308, TRP311, GLN315, LYS316, ALA318, ASP319, TRP320, GLY321, LEU324, HIS19, ARG20, ASN21, SER22, LEU23, VAL25, LEU26, ALA63, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, PHE80, GLU84, THR85, GLN86, LEU89, ASN90, PHE92, ALA93, ASN94, SER95, PHE98, GLN113, GLY114, LEU115, LYS117, VAL118, ILE119,

Table 5.8 continued

ASP121, LYS122, PRO160, LEU161, ASP170, ASP172, ALA253, PRO254, VAL255, ASN256, TRP257, ASP258, LYS159, GLY260, LYS261, VAL262, LEU265, PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, ILE300, LEU301, SER304, PHE305, TRP311, GLN315, LYS316, ALA318, ASP319, TRP320, GLY321, LEU324, PHE428, ALA429, SER430, LEU18, HIS19, ASN21, SER22, LEU23, VAL25, LEU26, LEU64, HIS65, THR67, VAL68, ILE71, LEU72, VAL75, VAL78, PHE79, PHE80, GLU84, THR85, GLN86, LEU89, ASN90, PHE92, ALA93, ASN94, SER95, PHE98, LEU101, PHE288, LYS289, SER290, PHE291, THR293, LEU294, LEU297, ILE300, LEU301, SER304, PHE305, TRP311, GLN315, ALA318, ASP319, TRP320, GLY321, LEU324, PRO418, ALA425, PHE428, ALA429, ILE433, GLN435

Table 5.10: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- SLC13A4 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	19	PHE	uS19	123	VAL	SLC13A4
	19	PHE	uS19	144	LEU	SLC13A4
	25	VAL	uS19	144	PHE	SLC13A4
	26	LEU	uS19	305	PHE	SLC13A4
	64	LEU	uS19	305	PHE	SLC13A4
	68	VAL	uS19	301	LEU	SLC13A4
	68	VAL	uS19	311	TRP	SLC13A4
	68	VAL	uS19	300	ILE	SLC13A4
	71	ILE	uS19	297	LEU	SLC13A4
	71	ILE	uS19	300	ILE	SLC13A4
	71	ILE	uS19	301	LEU	SLC13A4
	71	ILE	uS19	311	TRP	SLC13A4
	72	LEU	uS19	301	LEU	SLC13A4
	75	VAL	uS19	297	LEU	SLC13A4
	75	VAL	uS19	301	LEU	SLC13A4
	75	VAL	uS19	294	LEU	SLC13A4
	77	MET	uS19	277	PHE	SLC13A4
	78	VAL	uS19	288	PHE	SLC13A4
	78	VAL	uS19	294	LEU	SLC13A4
	79	PHE	uS19	288	PHE	SLC13A4
	79	PHE	uS19	291	PHE	SLC13A4
	79	PHE	uS19	294	LEU	SLC13A4
	89	LEU	uS19	297	LEU	SLC13A4
	89	LEU	uS19	93	ALA	SLC13A4

Table 5.9 continued

	89	LEU	uS19	98	PHE	SLC13A4
	92	PHE	uS19	93	ALA	SLC13A4
	92	PHE	uS19	98	PHE	SLC13A4
	93	ALA	uS19	93	ALA	SLC13A4
	93	ALA	uS19	92	PHE	SLC13A4
	98	PHE	uS19	324	LEU	SLC13A4
	98	PHE	uS19	89	LEU	SLC13A4
	98	PHE	uS19	92	PHE	SLC13A4
	101	LEU	uS19	324	LEU	SLC13A4
	112	VAL	uS19	262	VAL	SLC13A4
	118	MET	uS19	265	LEU	SLC13A4
	126	PHE	uS19	257	TRP	SLC13A4
	138	PRO	uS19	118	VAL	SLC13A4
	138	PRO	uS19	254	PRO	SLC13A4
	140	ILE	uS19	115	LEU	SLC13A4
	140	ILE	uS19	118	VAL	SLC13A4
	140	ILE	uS19	257	TRP	SLC13A4
	285	LEU	uS19	79	PHE	SLC13A4
	288	PHE	uS19	78	VAL	SLC13A4
	288	PHE	uS19	79	PHE	SLC13A4
	291	PHE	uS19	79	PHE	SLC13A4
	294	LEU	uS19	75	VAL	SLC13A4
	294	LEU	uS19	78	VAL	SLC13A4
	294	LEU	uS19	75	VAL	SLC13A4
	297	LEU	uS19	71	ILE	SLC13A4
	297	LEU	uS19	75	VAL	SLC13A4
	297	LEU	uS19	89	LEU	SLC13A4
	300	ILE	uS19	71	ILE	SLC13A4
	301	ILE	uS19	68	VAL	SLC13A4
	301	ILE	uS19	71	ILE	SLC13A4
	301	ILE	uS19	72	LEU	SLC13A4
	301	ILE	uS19	75	VAL	SLC13A4
	305	PHE	uS19	25	VAL	SLC13A4
	305	PHE	uS19	26	LEU	SLC13A4
	305	PHE	uS19	64	LEU	SLC13A4
	311	TRP	uS19	324	LEU	SLC13A4
	311	TRP	uS19	68	VAL	SLC13A4
	311	TRP	uS19	71	ILE	SLC13A4
	320	TRP	uS19	320	TRP	SLC13A4
	320	TRP	uS19	324	LEU	SLC13A4
	324	LEU	uS19	101	LEU	SLC13A4
	324	LEU	uS19	311	TRP	SLC13A4
	324	LEU	uS19	320	TRP	SLC13A4
	324	LEU	uS19	98	PHE	SLC13A4
Ionic	59	LYS	uS19	116	ASP	SLC13A4
	289	LYS	uS19	84	GLU	SLC13A4

uS19-HSPA8 docked complex

The prospective uS19-HSPA8 complex acquired from ClusPro was picked based on the balanced category on the basis of higher cluster size and lowest free energy, which were 63 and -794.3 kcal/mol respectively (Refer to Appendix, Table A.5). The top ten solutions predicted and refined through PatchDock/FireDock reveals desolvation energy of -0.61 kcal/mol and global free energy of 2.74 kcal/mol for the best docked conformation (Refer to Appendix, Table A.7). The ClusPro binding mode demonstrated high probability for uS19 and HSPA8 interaction. The RMSD values for both docked models were 0.98 Å and 0.00 Å for ClusPro and PatchDock respectively, thus validating the complex. Interface analysis of the complex to identify interface residues within the interactions are as tabulated in Table 5.11. PatchDock model was selected based on the maximum number of contact interface residues to represent the uS19- HSPA8 docked complex. The potential interface contact residues of uS19 to HSPA8 protein generated from SPDBV 4.1.1 and the ribbon representation of the uS19- HSPA8 protein docked complex obtained from PatchDock is presented using Pymol molecular graphics system (version 2.3) in Figure 5.6. PIC evaluation results of protein interactions within the complex to show the interface residues involved in stabilizing the complex via hydrophobic interactions and ionic interactions are as shown in Table 5.12.

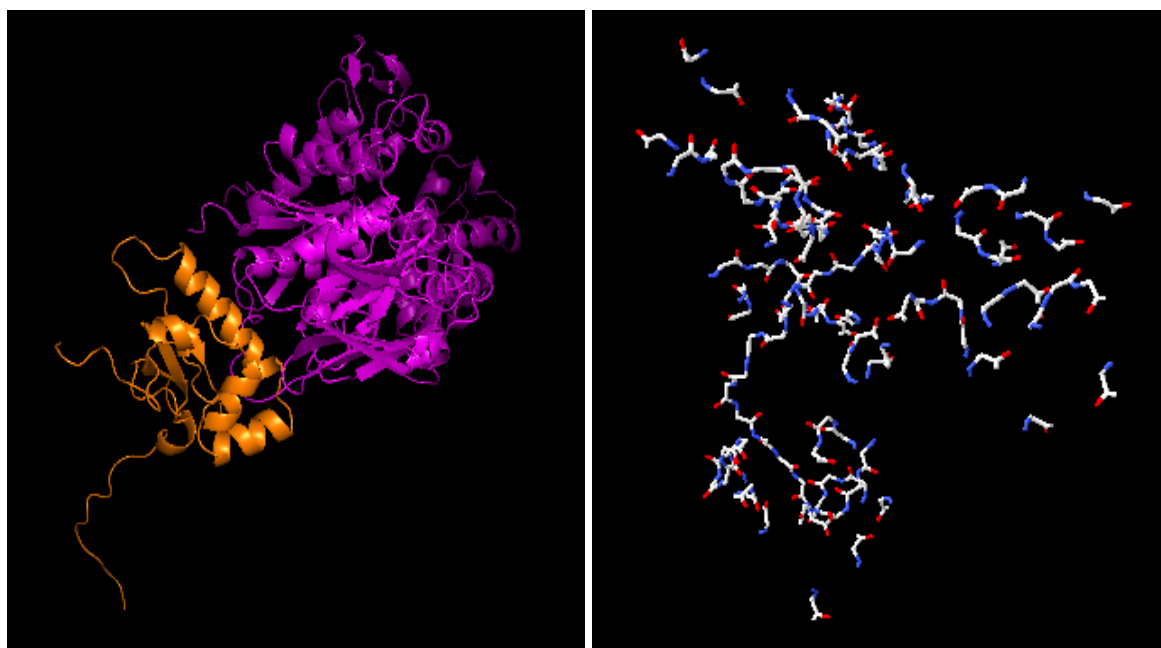


Figure 5.6: Ribbon representation of uS19- HSPA8 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by HSPA8. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.11: Possible interface residues of uS19 and HSPA8 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	TYR37, MET41, LEU52, ASN53, ARG54, GLY55, LEU56, ARG57, ARG58, LYS59, GLN60, HIS61, SER62, LEU63, LEU64, LYS65, ARG66, ARG68, LYS84, THR85, HIS86, LEU87, ASP89, PHE109,
	PatchDock	ARG17, PHE19, PHE22, THR23, ARG25, GLN42, LEU43, SER45, ALA46, ARG47, ARG49, ARG50, ASN53, ASN110, GLN111, VAL112, GLU113, LYS115, GLU117, MET118, GLY120, HIS121, TYR122, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, WAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140, GLY141, OXT141
HSPA8	ClusPro	VAL189, ASP214, GLU386, ASN387, VAL388, GLN389, ASP390, LEU391, LEU392, LEU393, LEU394, ASP395, VAL396, ALA482, ASN483, ILE485, ASP506, LYS507, ARG509, LEU510, SER511, LYS512
	PatchDock	PHE21, GLN22, HIS23, GLY24, LYS25, VAL26, GLU27, ILE28, ILE29, ALA30, ASN31, ASP32, GLN33, GLY34, ASN35, ARG36, ALA133, TYR134, GLY136, ILE181, ALA182, GLY184, LYS187, ASP214, GLY215, ILE216, ARG272, THR273, SER275, SER276, SER277, ARG299, ARG342, GLN347, LYS361, SER362, ILE363, ASN364, GLU367, TYR371, VAL375, ALA378, ILE379, GLY382, ASP383, LYS384, SER385, GLU386, ASN387, ASP395, VAL396, THR397, PRO398, LEU401, ILE420, ILE480, ALA482, ASN483, GLY484, LYS507, GLY508, ARG509, LEU510, SER511, LYS512, ILE515, GLUE516

Table 5.12: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- HSPA8 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	22	PHE	uS19	30	ALA	HSPA8
	126	PHE	uS19	26	VAL	HSPA8
	126	PHE	uS19	28	ILE	HSPA8
	126	PHE	uS19	371	TYR	HSPA8
	128	ILE	uS19	21	PHE	HSPA8
	130	TYR	uS19	375	VAL	HSPA8
	132	PRO	uS19	378	ALA	HSPA8
	132	PRO	uS19	379	ILE	HSPA8
	133	VAL	uS19	379	ILE	HSPA8
	138	PRO	uS19	398	PRO	HSPA8
	138	PRO	uS19	510	LEU	HSPA8
	138	PRO	uS19	515	ILE	HSPA8
	140	ILE	uS19	420	ILE	HSPA8
Ionic	121	HIS	uS19	367	GLU	HSPA8
	135	HIS	uS19	516	GLU	HSPA8

uS19-CUBN docked complex

The best uS19-CUBN complex generated by ClusPro was selected from the balanced category on the basis of higher cluster size and lowest free energy, which were 232 and -317.1 kcal/mol respectively (Refer to Appendix, Table A.5). The top ten solutions predicted and refined through PatchDock/FireDock exhibited a desolvation energy of -7.24 kcal/mol and global free energy of -12.03 kcal/mol for the best docked conformation (Refer to Appendix, Table A.7). Both servers indicates a high probability of interaction between both proteins and shown to have low conformational changes as the RMSD values obtained were 0.94 Å and 0.00 Å for ClusPro and PatchDock respectively. Interface residues of the complex within 5 Å of each other are shown in Table 5.13. PatchDock model was selected to portray the uS19-CUBN docked complex. The potential interface contact residues of uS19 to CUBN protein obtained from SPDBV 4.1.1 and the ribbon representation of the uS19-CUBN protein docked complex generated from PatchDock is displayed using Pymol

molecular graphics system (version 2.3) in Figure 5.7. PIC evaluation involving the interface residues stabilizing the complex via hydrophobic interactions and ionic interactions are as listed in Table 5.14.

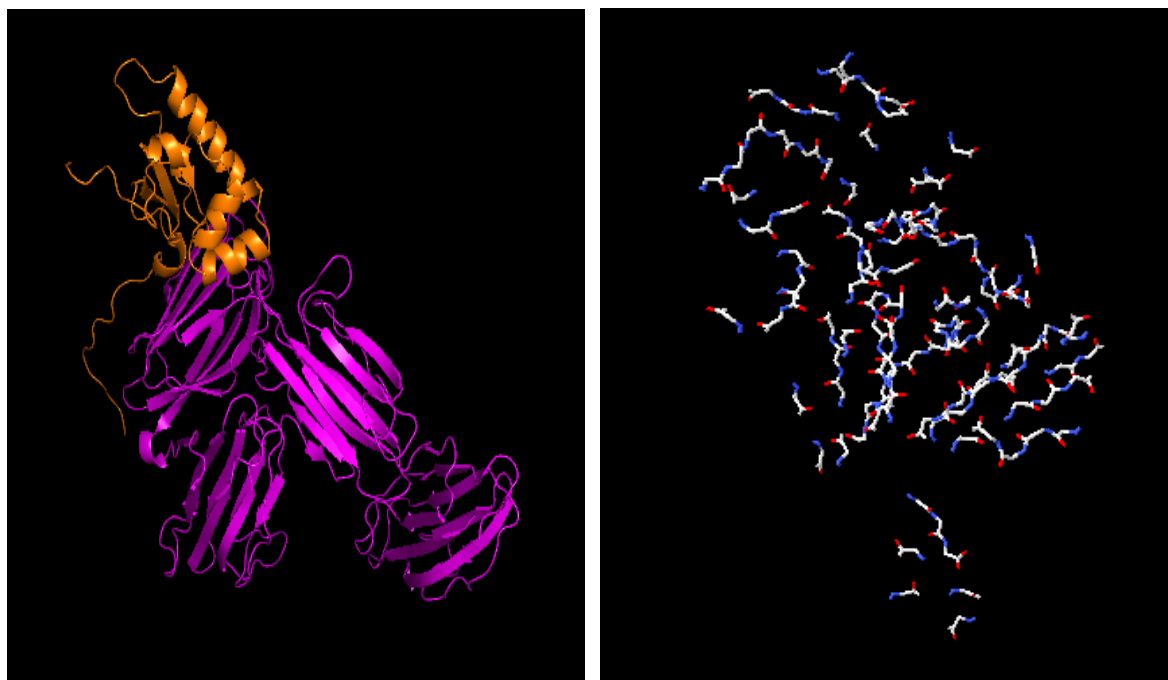


Figure 5.7: Ribbon representation of uS19-CUBN docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by CUBN. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.13: Possible interface residues of uS19 and CUBN protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	ARG47, ARG50, ARG51, ASN53, ARG54, GLY55, ARG57, ARG58, LYS59, GLN60, HIS86, LEU87, ARG88, ASP89, MET90, TYR104, ASN105, GLY106, LYS107, THR108, TYR122, GLY124, ILE128, THR129, TYR130, LYS131, VAL13, HIS135, ARG137, PRO138, GLY139, ILE140, GLY141
	PatchDock	ARG17, PHE19, ARG20, ARG47, ARG51, MET77, GLU78, LYS79, VAL82, HIS86, ARG88, ASP89, MET100, ASN105, GLY106, LYS107, THR108, PHE109, ASN110, GLN111, VAL112, GLU113, LYS115, GLU117, MET118, HIS121, TYR122, GLY124, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, LYS134, HIS135, GLY139, ILE140, GLY141, OXT141
CUBN	ClusPro	CYS932, GLY933, GLU934, GLY955, ILE956, ASN957, CYS958, THR959, HIS961, GLN965, PRO965, GLU980, HIS982, TYR983, ASN984, CYS985, THR986, ASN987, TYR989, GLU991, TYR993, THR995, ASP996, SER997, ARG1003, SER1019, MET1021, VAL1023, VAL1025, THR1026, ASP1027, SER1028, ASP1029, LEU1030, HIS1187, SER1188, GLU1213, LEU1220, ASP1221, ASP1262, GLU1263, GLY1264, GLN1265, GLN1266
	PatchDock	THR942, GLU975, LEU1036, ASN1038, LEU1085, ALA1087, VAL1088, HIS1089, PRO1118, LEU1119, LEU1120, GLY1121, PHE1123, TYR1124, PRO1129, PRO1130, THR1131, ILE1132, ILE1133, SER1134, HIS1135, TRP1158, ASP1159, GLY1160, SER1161, GLY1166, GLY1167, ASN1168, LEU1169, THR1170, THR1171, SER1172, SER1173, GLY1174, THR1175, PHE1176, TRP1194, LYS1196, SER1197, SER1198, HIS1199, GLU1206, LYS1208, PRO1229, SER1230, ASN1232, ASP1253, SER1254, LYS1271, GLU1273, TYR1274, ARG1275, ASN1285, GLN1286, THR1287, TYR1288, GLY1289, ILE1290, GLU1292, GLY1295, TYR1296, PRO1297, ASN1298, PRO1299, THR1321, LEU1323, LYS1380, GLN1383, GLN1385, TRP1386, PHE1387

Table 5.14: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- CUBN docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	77	MET	uS19	1387	PHE	CUBN
	82	VAL	uS19	1297	PRO	CUBN
	109	PHE	uS19	1297	PRO	CUBN
	118	MET	uS19	1194	TRP	CUBN
	122	TYR	uS19	1133	ILE	CUBN
	128	ILE	uS19	1120	LEU	CUBN
	130	TYR	uS19	1118	PRO	CUBN
	130	TYR	uS19	1119	LEU	CUBN
Ionic	17	ARG	uS19	1253	ASP	CUBN
	47	ARG	uS19	1159	ASP	CUBN
	113	GLU	uS19	1196	LYS	CUBN
	113	GLU	uS19	1199	HIS	CUBN
	115	LYS	uS19	1253	ASP	CUBN
	134	LYS	uS19	1206	GLU	CUBN
	134	LYS	uS19	1273	GLU	CUBN

uS19-Cobalamin docked complex

The uS19-Cobalamin complex acquired from ClusPro with the highest cluster size and lowest free energy, which were 109 and -902.5 kcal/mol respectively was selected (Refer to Appendix, Table A.5). The best docked conformation predicted and refined through PatchDock/FireDock reveals desolvation energy of 0.00 kcal/mol and global free energy of -20.41 kcal/mol (Refer to Appendix, Table A.7). Both the energy profile obtained showed a relatively good energy profile, which indicates high chance of interaction between uS19 and Cobalamin. The RMSD values for both docked models were 0.90 Å and 0.00 Å for ClusPro and PatchDock respectively revealing low conformation changes in the complex. Interface residues within the interactions are as shown in Table 5.15. PatchDock model was chosen due to the most number of contact interface residues to represent the uS19- Cobalamin docked complex. The potential interface contact residues of uS19 to Cobalamin protein taken from SPDBV 4.1.1 and the ribbon representation of the uS19- Cobalamin protein

docked complex acquired from PatchDock is displayed using Pymol molecular graphics system (version 2.3) in Figure 5.8. The PIC evaluation only predicts the hydrophobic interaction of the interface residue of uS19 and Cobalamin (Table 5.16).

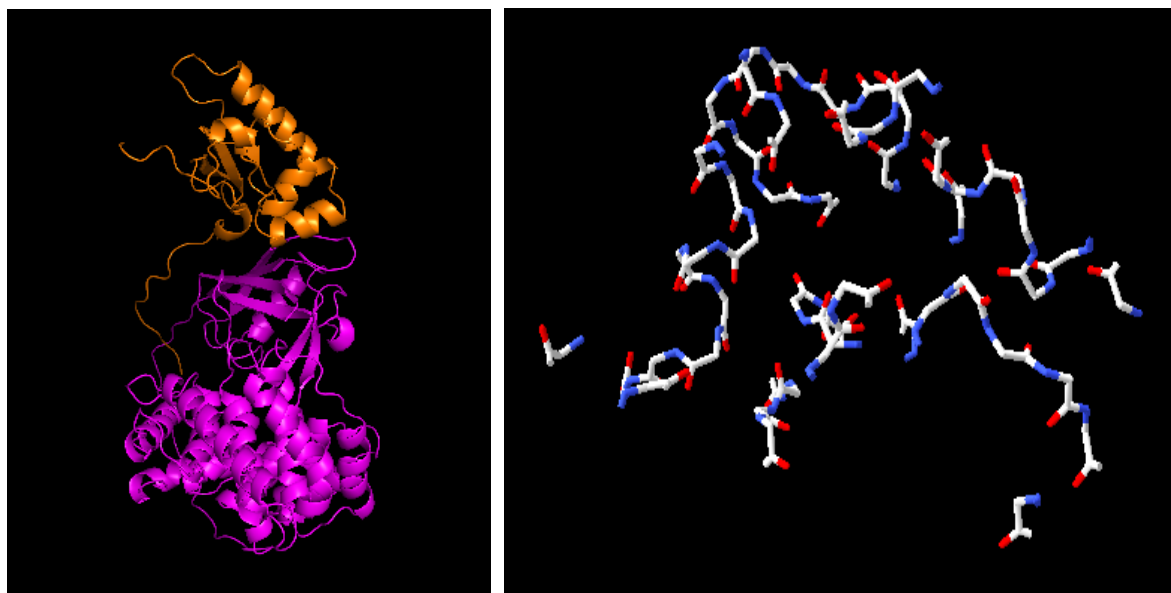


Figure 5.8: Ribbon representation of uS19- Cobalamin docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by Cobalamin. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.15: Possible interface residues of uS19 and Cobalamin protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	ARG47, ARG50, ARG51, ARG54, GLY55, ARG57, ARG88, ASP89, TYR122, GLY124, GLU125, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, ARG137, PRO138, GLY139, ILE140, GLY141,
	PatchDock	ARG17, ARG20, LEU29, ASP30, LEU33, ASP34, MET35, SER36, TYR37, LEU40, ARG68, LYS71, LYS72, GLU73, ALA74, PRO75, PRO76, MET77, GLU78, LYS79, PRO80, GLU95, VAL97, GLY98, SER99, MET100, GLU113, ILE114, LYS115, PRO116
Cobalamin	ClusPro	PRO302, GLY303, PRO304, GLY305, PRO306, THR307, SER308, ALA309, ILE315, THR317, ASN319, GLN321, LEU322, GLU324, VAL325, GLU326, LEU327, LEU328, PHE329, GLU331, ILE333, LYS338, GLY391, PHE405, ASN406, HIS407, GLU408, HIS409, ASN413, PHE414, THR415, GLN416
	PatchDock	LEU219, GLY221, ILE222, TYR223, VAL224, LEU225, VAL226, LEU228, VAL229, LEU230, ALA234, TYR235, ARG237, ARG238, TRP239, ARG240, ARG241, LEU242, THR243, VAL244, CYS245, OXT245

Table 5.16: Predicted interfacial residues involved in hydrophobic interactions within the uS19- Cobalamin docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	29	LEU	uS19	239	TRP	Cobalamin
	33	LEU	uS19	226	VAL	Cobalamin
	33	LEU	uS19	230	LEU	Cobalamin
	37	TYR	uS19	219	LEU	Cobalamin
	37	TYR	uS19	222	ILE	Cobalamin
	40	LEU	uS19	222	ILE	Cobalamin
	74	ALA	uS19	235	TYR	Cobalamin
	76	PRO	uS19	234	ALA	Cobalamin
	80	PRO	uS19	242	LEU	Cobalamin
	97	VAL	uS19	239	TRP	Cobalamin
	116	PRO	uS19	239	TRP	Cobalamin

uS19-SLC7A1 docked complex

The uS19-SLC7A1 docked complex was selected from ClusPro with the largest cluster size of 68 and lowest free energy, -1010.0 kcal/mol among the other complexes (Refer to Appendix, Table A.5). The complex was refined through PatchDock/FireDock obtaining desolvation and global free energy of -0.57 kcal/mol and 10.36 kcal/mol respectively (Refer to Appendix, Table A.7). The ClusPro binding mode demonstrated a higher probability for both protein to interact. Structural superimposition via SPDBV 4.1.1 shows the RMSD values of 0.93 Å and 0.00 Å for ClusPro and PatchDock models respectively, validating the prediction. The interface residues within 5 Å from each other was obtained from SPDBV 4.1.1 (Table 5.17) and PatchDock was selected to represent the docked complex, where the ribbon representation of the complex along with the potential interface residues are displayed in Figure 5.9. Table 5.18 reveals the hydrophobic and ionic interaction in the binding site interface of the uS19 and SLC7A1 via the PIC evaluation.

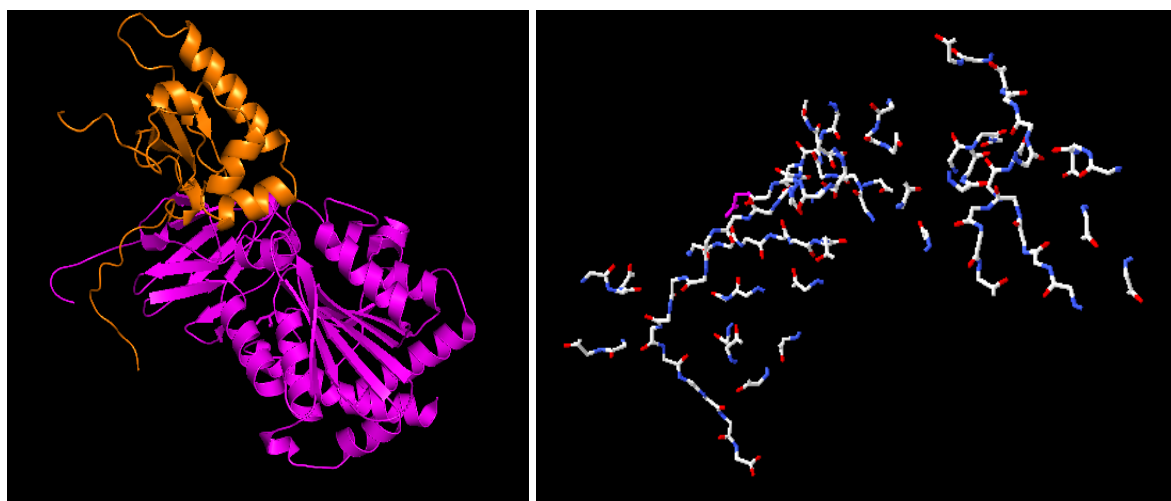


Figure 5.9: Ribbon representation of uS19- SLC7A1 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by SLC7A1. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.16: Possible interface residues of uS19 and SLC7A1 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	PHE19, PHE22, THR23, TYR24, ARG25, GLY26, SER45, ALA46, ARG47, ARG50, ARG51, LYS59, HIS86, ARG88, ASP89, TYR104, ASN105, GLY106, LYS107, HIS121, TYR122, GLY124, GLU125, PHE126, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140
	PatchDock	ARG47, ARG50, ARG51, ARG54, GLY55, LEU56, ARG57, LYS59, GLN60, LEU87, ARG88, ASP89, TYR104, ASN105, GLY106, LYS107, THR108, GLY124, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140, GLY141, OXT141
SLC7A1	ClusPro	HIS7, GLU6, ASN5, TYR3, PHE2, THR6, SER7, ASN8, PHE9, ASP10, ALA11, GLY12, ARG30, ASP32, ASN33, ARG34, SER35, GLU36, ARG44, THR46, TYR78, ASP79, ARG80, VAL81, ASN82, GLU102, PHE103, ASP104, SER105, HIS107, GLU116, GLU117, SER120, GLU121, GLU142, GLY143, ARG144, VAL221, HIS222, GLY223, ASN224, GLU237, ARG245
	PatchDock	HIS7, GLU6, TYR3, PHE2, GLN1, THR2, LEU3, SER4, ILE5, THR6, SER7, ASN8, ILE14, ASP15, VAL16, VAL17, SER18, CYS19, CLU20, ARG26, ARG28, ASN33, ARG34, SER35, ARG44, THR46, GLY47, ARG49, THR100, GLU102, ASP104, GLY138, ARG139, THR140, VAL141, GLU142, GLY143, ARG144, HIS222, ALA229, ALA230, ALA243, GLU244, ARG245, PRO247, VAL251

Table 5.17: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- SLC7A1 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	130	TYR	uS19	5	ILE	SLC7A1
Ionic	51	ARG	uS19	142	GLU	SLC7A1
	57	ARG	uS19	244	GLU	SLC7A1
	89	ASP	uS19	222	HIS	SLC7A1
	135	HIS	uS19	104	ASP	SLC7A1

uS19-RRP1B docked complex

Potential ClusPro-generated complexes of uS19-RRP1B was selected based on the highest cluster size (132) and the lowest free energy (-779.1 kcal/mol) (Refer to Appendix, Table A.5). The complex was then entered through PatchDock/FireDock obtaining a desolvation and global energy of -0.45 kcal/mol and -4.73 kcal/mol (Refer to Appendix, Table A.7). Thus, the energy profiles obtained from both analyses indicates high probability that uS19 and RRP1B to interact. Using SPDBV 4.1.1, the RMSD values for each docked complex reveals structures that have low conformational changes between them. The RMSD values for the uS19-RRP1B docked complex generated by ClusPro and PatchDock are 0.93 Å and 0.00 Å respectively. The possible contact interface residues obtained from SPDBV 4.1.1 are as tabulated in Table 5.19. PatchDock was selected to represent the uS19- RRP1B complex due to high number of contact interface residues and is displayed in Figure 5.10, where both the ribbon representation of the complex and the potential interface residue are exhibited. The hydrophobic and ionic interactions of uS19 and RRP1B via the PIC evaluation are shown in Table 5.20.

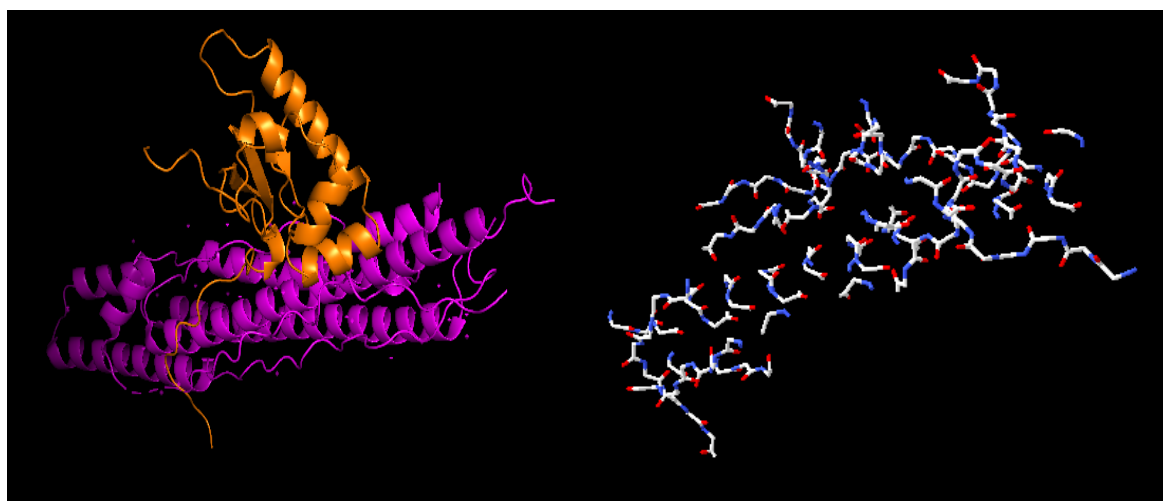


Figure 5.10: Ribbon representation of uS19- RRP1B docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by RRP1B. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.19: Possible interface residues of uS19 and RRP1B protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	SER36, TYR37, LEU40, LEU52, ASN53, ARG54, GLY55, LEU56, ARG57, ARG58, HIS61, LEU64, LYS65, ARG68, LEU93, GLU95
	PatchDock	ALA46, ARG47, ARG49, ARG50, ARG51, LEU52, ASN53, ARG54, GLY55, LEU56, ARG57, ARG58, LYS59, HIS86, LEU87, ARG88, ASO89, MET90, TYR104, ASN105, GLY106, LYS107, THR108, PHE109, TYR122, GLY124, GLU125, ILE128, THR129, TYR130, LYS131, VAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140
RRP1B	ClusPro	GLU241, TRP242, ALA243, SER244, GLU245, TRP246, ASP247, GLN249, ASN250, ALA253, MET254, LEU255, GLU257, GLU260,
	PatchDock	GLU13, ALA16, HIS17, GLU18, CYS19, GLU20, SER21, THR22, ARG23, PRO24, GLU25, GLY26, PRO27, GLY28, ASP29, ALA30, ALA37, THR38, SER41, THR43, SER44, LYS45, MET47, ALA48, ALA50, ALA51, LYS52, LYS54, ALA55, PHE57, ALA58, ALA59, ASP61, ARG62, GLU65, ALA68, ALA69, ARG72, ALA73, VAL107, PRO108, ASP109, MET110, SER111, GLY112, ILE113, PRO114, GLY115, LEU117, PRO120, MET122, GLU123, TYR124, PRO125, SER126, ALA127, GLN159, TYR163, ALA267, GLN271, ILE272, GLN273

Table 5.20: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19- RRP1B docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	122	TYR	uS19	117	LEU	RRP1B
	128	ILE	uS19	120	PRO	RRP1B
	128	ILE	uS19	55	ALA	RRP1B
	130	TYR	uS19	120	PRO	RRP1B
	130	TYR	uS19	122	MET	RRP1B
	130	TYR	uS19	51	ALA	RRP1B
	130	TYR	uS19	55	ALA	RRP1B
	133	VAL	uS19	122	MET	RRP1B
	133	VAL	uS19	125	PRO	RRP1B
	133	VAL	uS19	48	ALA	RRP1B
	133	VAL	uS19	51	ALA	RRP1B
	138	PRO	uS19	127	ALA	RRP1B
Ionic	50	ARG	uS19	109	ASP	RRP1B
	51	ARG	uS19	65	GLU	RRP1B
	54	ARG	uS19	109	ASP	RRP1B
	57	ARG	uS19	13	GLU	RRP1B
	59	LYS	uS19	20	GLU	RRP1B
	88	ARG	uS19	61	ASP	RRP1B
	89	ASP	uS19	62	ARG	RRP1B
	131	LYS	uS19	29	ASP	RRP1B
	135	HIS	uS19	123	GLU	RRP1B

uS19-LMP2 docked complex

The ClusPro docked model for uS19-LMP2 complex were chosen based on their cluster size (184) and lowest free energy (-1111.4 kcal/mol) (Refer to Appendix, Table A.5). The complex model was then submitted to be refined via PatchDock/FireDock with a desolvation and global energy of -5.60 kcal/mol and -1.50 kcal/mol respectively (Refer to Appendix, Table A.7). Both analyses obtained indicates high probability for both proteins to interact. Structural superimposition using SPDBV 4.1.1 indicates RMSD values of 0.98 Å for ClusPro model and 0.00 Å for the PatchDock model, thus validating the complex. Interface analysis was done to identify the contact interface residues (Table 5.21). Based on the interface analysis, PatchDock model was used to represent the uS19-LMP1 docked complex

as the number of contact interface residues is the highest (Figure 5.11). PIC evaluation was then done to obtain the hydrophobic and ionic interaction between uS19 and LMP2 (Table 5.22).

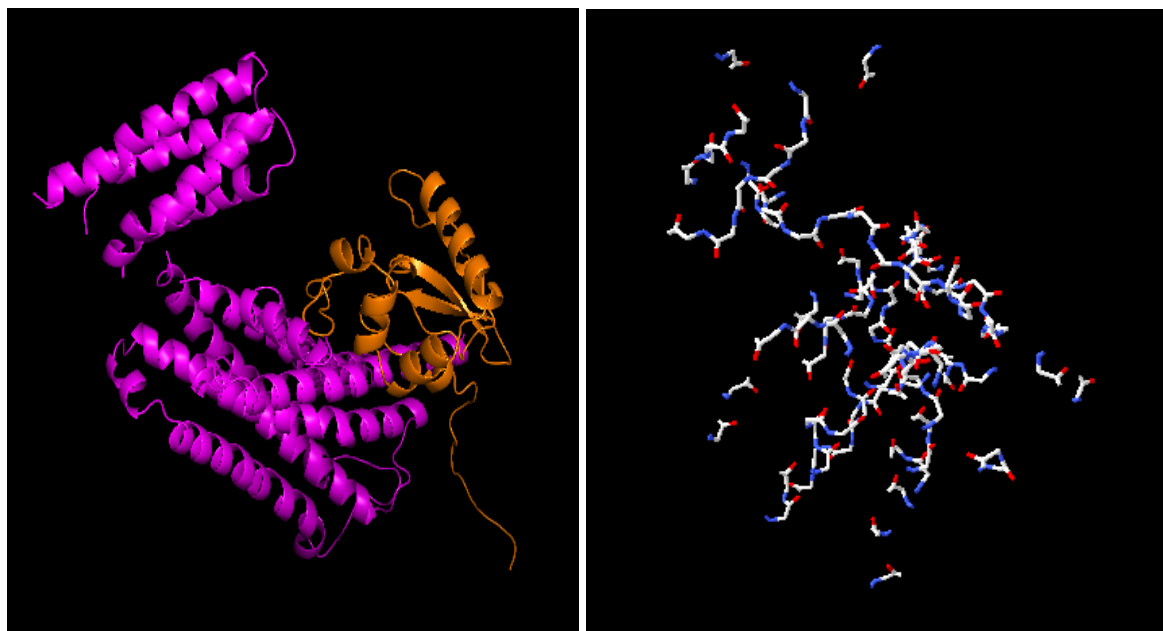


Figure 5.11: Ribbon representation of uS19-LMP2 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by LMP1. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.21: Possible interface residues of uS19 and LMP2 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	ARG17, THR18, PHE19, LYS21, PHE22, THR23, TYR24, ARG25, GLY26, VAL27, ASP28, GLN31, MET35, GLN42, LEU43, ILE92, ASN105, THR108, ASN110, GLN111, VAL112, GLU113, LYS115, GLU117, MET118, HIS121, LEU123, PHE126, SER127, ILE128, THR129, TYR130, PRO132, VAL133, LYS134, HIS135
	PatchDock	ARG17, THR18, PHE19, ARG20, LYS21, PHE22, THR23, TYR24, ARG25, GLY26, SER45, ARG47, MET77, ASN110, GLN111, VAL112, GLU113, ILE114, LYS115, PRO116, GLU117, MET118, ILE119, GLY120, HIS121, TYR122, GLU125, PHE126, ILE128, TYR130, LYS131, PRO132, VAL133, LYS134, HIS135,
LMP2	ClusPro	THR22, PHE23, VAL25, VAL26, PHE27, VAL29, LEU30, VAL31, GLY33, SER34, LEU35, LEU37, VAL38, ILE41, LEU45, VAL87, ILE88, TYR90, ALA344, ILE347, LEU377ILE384, ALA387, CYS388, ALA390, LEU391, LYS397, PHE400, LYS401, LEU404, MET405, CYS406
	PatchDock	GLU118, TRP120, PRO125, ASN126, ALA129, LEU151, PRO152, VAL155, ARG156, PHE157, ALA158, LYS159, THR160, LEU162, ALA163, ASN164, SER165, SER166, PRO167, PHE168, ASN169, VAL170, GLU192, TYR195, LEU198, GLN201, VAL202, LYS204, ASP205, GLU208, LYS209, ILE210, MET212, LYS213, ILE214, LYS215, ASP216, ASN217, GLY221, ASP222, ILE223, LYS241, LYS242, GLN321, TYR385, GLN387, PHE388, LEU389, VAL392, THR393, PRO394, LEU395, ASN396, GLU397, GLN416, TYR417

Table 5.22: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-LMP2 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	19	PHE	uS19	223	ILE	LMP1
	24	TYR	uS19	212	MET	LMP1
	118	MET	uS19	167	PRO	LMP1
	126	PHE	uS19	162	LEU	LMP1
	128	ILE	uS19	162	LEU	LMP1
	128	ILE	uS19	163	ALA	LMP1
	130	TYR	uS19	162	LEU	LMP1
	130	TYR	uS19	198	LEU	LMP1
	133	VAL	uS19	155	VAL	LMP1
Ionic	25	ARG	uS19	216	ASP	LMP1
	117	GLU	uS19	213	LYS	LMP1
	135	HIS	uS19	192	GLU	LMP1

uS19-EBNA1 docked complex

The best ClusPro docked complex of uS19-EBNA1 was selected based on the most number of cluster size (59) and lowest free energy (-938.1 kcal/mol) (Refer to Appendix, Table A.5).

The model was refined using PatchDock/FireDock to obtain the desolvation and global energy of -1.12 kcal/mol and 13.15 kcal/mol respectively (Refer to Appendix, Table A.7).

The ClusPro binding mode demonstrated high probability for uS19 and EBNA1 to interact.

The RMSD values obtained via SPDBV 4.1.1, through superimposition, were 0.89 Å and 0.00 Å for ClusPro and PatchDock models respectively. This validates that the structure has

low conformational changes between them allowing further investigation. The possible

interface residues contact for both ClusPro and PatchDock complex between uS19 and

EBNA1 were obtained using SPDBV 4.1.1 (Table 5.23). Based on Table 5.22, PatchDock

model of uS19-EBNA1 was chosen to represent the complex as it has the most contact

interface residues. Figure 5.12 portrays the ribbon representation of the PatchDock model of

uS19-EBNA1 complex and the potential interface residue of the complex. The model is then

input via PIC to obtain the hydrophobic and ionic interaction between uS19 and EBNA1 as

shown in Table 5.24.

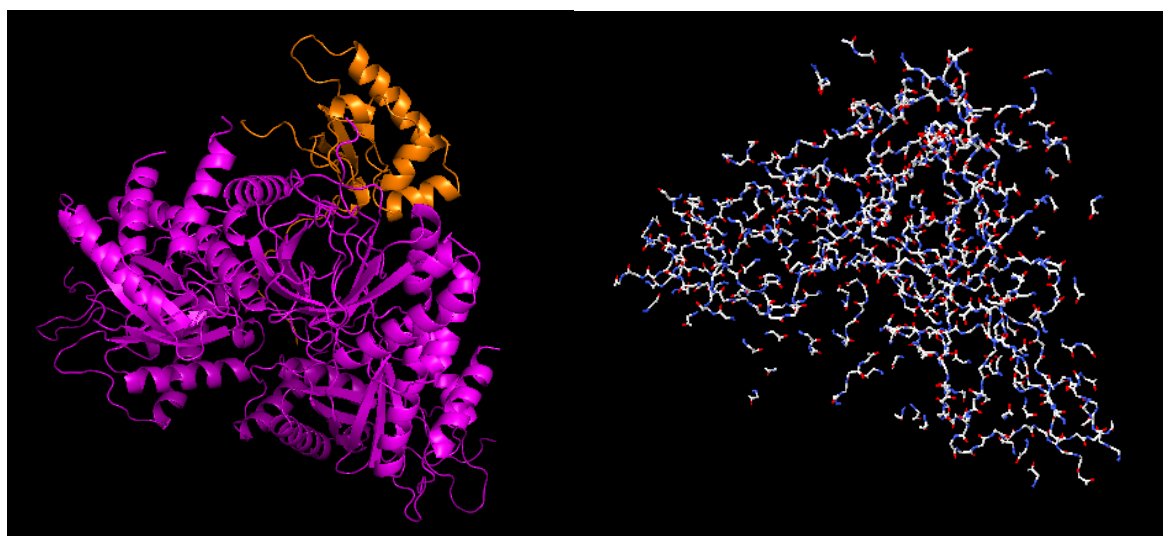


Figure 5.12: Ribbon representation of uS19-EBNA1 docked complexes predicted by PatchDock viewed using Pymol 2.3. Receptor (orange) is represented by uS19 and ligand (magenta) is represented by EBNA1. Potential interface residues only within 5 Å of each other are portrayed.

Table 5.23: Possible interface residues of uS19 and EBNA1 protein explored through the dual docking protocols.

Protein	Docking Servers	Interface Residues (<5Å)
uS19	ClusPro	ARG17, PHE22, ARG47, ARG50, ARG51, ARG54, GLY55, LEU56, ARG57, ARG58, LYS59, GLN60, THR85, HIS86, LEU87, ARG88, ASP89, TYR104, ASN105, GLY106, LYS155, GLU117, MET118, HIS121, TYR122, GLY124, GLU125, PHE126, SER127, ILE128, THR129, TYR130, LYS131, PRO132, VAL133, HIS135
	PatchDock	THR18, PHE19, PHE22, THR23, TYR24, ARG25, GLY26, VAL27, GLN39, MET41, GLN42, LEU43, SER45, ALA46, ARG47, ARG49, ARG50, GLU117, GLY120, HIS121, TYR122, GLU125, PHE126, ILE128, TYR130, VAL133, LYS134, HIS135, GLY136, ARG137, PRO138, GLY139, ILE140, GLY141, OXT141
EBNA1	ClusPro	PHE43, VAL44, TYR45, GLY46, THR69, PRO70, SER72, ARG73, LEU74, PRO75, MET78, PRO84, GLN85, PRO86, GLY87, PRO88, LEU89, ARG90, TYR96, MET98, PHE100, CYS126, ASN127, ARG129, VAL130, THR131, VAL132, CYS133, SER134, PHE135, ASP136, ASP137, GLY138, VAL139, ASP140, LEU141, PRO142, PRO143, TRP144, PHE145
	PatchDock	ARG486, HIS493, VAL494, GLU495, ARG496, GLY501, TRP503, PHE508, TYR510, ARG521, ARG522, ALA525,

Table 5.22 continued

ILE528, PRO529, GLN530, CYS531, ARG532, LEU533, THR534, PRO535, LEU536, SER537, ARG538, LEU539, PRO540, PHE541, GLY542, MET543, ALA544, PRO545, PRO549, GLN550, PRO551, GLY552, PRO553, LEU554, ARG555, GLU556, SER557, VAL559, TYR561, MET563, PHE565, GLN567, GLU573, LYS576, LYS580, ARG594, VAL595, THR596, VAL597, CYS598, SER599, PHE600, ASP601, ASP602, GLY603, VAL604, ASP605, LEU606, PRO607, PRO608, TRP609, PHE610, PRO611, OXT611, LEU488, ARG491, GLY501, TRP503, PHE508, TYR510, GLY511, GLY512, SER513, LYS514, THR515, SER516, ARG521, ARG522, ALA525, ILE528, PRO529, CYS531, ARG532, LEU533, THR534, PRO535, LEU536, SER537, ARG538, LEU539, PRO540, PHE541, GLY542, MET543, ALA544, PRO545, GLY546, PRO547, GLY548, PRO549, GLN550, PRO551, GLY552, PRO553, LEU554, ARG555, GLU556, SER557, ILE558, VAL559, TYR561, MET563, PHE565, GLN567, GLU573, LYS576, LYS580, ASP581, MET584, THR585, LYS586, PRO587, ALA588, PRO589, CYS591, ASN592, ARG594, VAL595, THR596, VAL597, CYS598, SER599, PHE600, ASP601, ASP602, GLY603, VAL604, ASP605, LEU606, PRO607, PRO608, TRP609, PHE610, PRO611, GLY501, TRP503, PHE508, TYR510, GLY511, GLY512, ARG521, ARG522, ALA525, ILE528, PRO529, CYS531, ARG532, LEU533, THR534, PRO535, LEU536, SER537, ARG538, LEU539, PRO540, PHE541, GLY542, MET543, ALA544, PRO545, GLY546, PRO547, GLY548, PRO549, GLN550, PRO551, GLY552, PRO553, LEU554, ARG555, GLU556, SER557, VAL559, TYR561, MET563, PHE565, GLN567, GLU573, LYS576, LYS580, VAL583, MET584, THR585, LYS586, PRO587, ALA588, CYS591, ASN592, ARG594, VAL595, THR596, VAL597, CYS598, SER599, PHE600, ASP601, ASP602, GLY603, VAL604, ASP605, LEU606, PRO607, PRO608, TRP609, PHE610, PRO611, ARG486, HIS493, VAL494, GLU495, ARG496, GLY501, TRP503, PHE508, TYR510, ARG521, ARG522, THR524, LEU526, ILE528, PRO529, GLN530, CYS531, ARG532, LEU533, THR534, PRO535, LEU536, SER537, ARG538, LEU539, PRO540, PHE541, GLY542, MET543, ALA544, PRO545, PRO549, GLN550, PRO551, GLY552, PRO553, LEU554, ARG555, GLU556, SER557, VAL559, TYR561, MET563, PHE565, GLN567, GLU573,

Table 5.22 continued

LYS576, LYS580, ARG594, VAL595, THR596, VAL597, CYS598, SER599, PHE600, ASP601, ASP602, GLY603, VAL604, ASP605, LEU606, PRO607, PRO608, TRP609, PHE610, PRO611, OXT611, PHE478, ARG486, VAL494, GLU496, ARG496, THR497, ASP499, GLY501, TRP503, PHE508, TYR510, LYS514, THR515, TYR518, ARG521, ARG522, ALA525, ILE528, PRO529, GLN530, CYS531, ARG532, LEU533, THR534, PRO535, LEU536, SER537, ARG538, LEU539, PRO540, PHE541, GLY542, MET543, ALA544, PRO545, PRO549, GLN550, PRO551, GLY552, PRO553, LEU554, ARG555, GLU556, SER557, VAL559, TYR561, MET563, PHE565, GLN567, GLU573, LYS576, LYS580, ARG594, VAL595, THR596, VAL597, CYS598, SER599, PHE600, ASP601, ASP602, GLY603, VAL604, ASP605, LEU606, PRO607, PRO608, TRP609, PHE610, PRO611, LYS477, PHE478, ARG491, GLY501, TRP503, PHE508, TYR510, GLY511, GLY512, SER513, LYS514, THR515, SER516, TYR517, ASN519, ARG521, ARG522, ALA525, , ILE528, PRO529, CYS531, ARG532, LEU533, THR534, PRO535, LEU536, SER537, ARG538, LEU539, PRO540, PHE541, GLY542, MET543, ALA544, PRO545, GLY546, PRO547, GLY548, PRO549, GLN550, PRO551, GLY552, PRO553, LEU554, ARG555, GLU556, SER557, ILE558, VAL559, TYR561, MET563, PHE565, GLN567, GLU573, LYS576, LYS580, ASP581, VAL583, MET584, THR585, LYS586, PRO587, ALA588, CYS591, ASN592, ARG594, VAL595, THR596, VAL597, CYS598, SER599, PHE600, ASP601, ASP602, GLY603, VAL604, ASP605, LEU606, PRO607, PRO608, TRP609, PHE610, PRO611

Table 5.24: Predicted interfacial residues involved in hydrophobic and ionic interactions within the uS19-EBNA1 docked complex.

Interaction	Position	Residue	Chain	Position	Residue	Chain
Hydrophobic	19	PHE	uS19	478	PHE	EBNA1
	19	PHE	uS19	518	TYR	EBNA1
	46	ALA	uS19	518	TYR	EBNA1
	126	PHE	uS19	558	ILE	EBNA1
	133	VAL	uS19	611	PRO	EBNA1
	138	PRO	uS19	549	PRO	EBNA1
	138	PRO	uS19	588	ALA	EBNA1
	138	PRO	uS19	589	PRO	EBNA1
	140	ILE	uS19	526	LEU	EBNA1
	503	TRP	uS19	543	MET	EBNA1
	508	PHE	uS19	563	MET	EBNA1
	508	PHE	uS19	565	PHE	EBNA1
	508	PHE	uS19	600	PHE	EBNA1
	508	PHE	uS19	604	VAL	EBNA1
	510	TYR	uS19	604	VAL	EBNA1
	510	TYR	uS19	547	PRO	EBNA1
	525	ALA	uS19	553	PRO	EBNA1
	525	ALA	uS19	554	LEU	EBNA1
	529	PRO	uS19	588	ALA	EBNA1
	533	LEU	uS19	553	PRO	EBNA1
	553	LEU	uS19	554	LEU	EBNA1
	539	LEU	uS19	565	PHE	EBNA1
	539	LEU	uS19	604	VAL	EBNA1
	539	LEU	uS19	606	LEU	EBNA1
	539	LEU	uS19	607	PRO	EBNA1
	539	LEU	uS19	547	PRO	EBNA1
	540	PRO	uS19	565	PHE	EBNA1
	540	PRO	uS19	606	LEU	EBNA1
	541	PHE	uS19	607	PRO	EBNA1
	541	PHE	uS19	609	TRP	EBNA1
	541	PHE	uS19	611	PRO	EBNA1
	543	MET	uS19	503	TRP	EBNA1
	544	ALA	uS19	609	TRP	EBNA1
	545	PRO	uS19	610	PHE	EBNA1
	545	PRO	uS19	608	PRO	EBNA1
	547	PRO	uS19	510	TYR	EBNA1
	547	PRO	uS19	559	VAL	EBNA1
	547	PRO	uS19	607	PRO	EBNA1
	549	PRO	uS19	609	TRP	EBNA1
	549	PRO	uS19	611	PRO	EBNA1
	551	PRO	uS19	609	TRP	EBNA1
	553	PRO	uS19	525	ALA	EBNA1
	553	PRO	uS19	533	LEU	EBNA1
	554	LEU	uS19	525	ALA	EBNA1

Table 5.23 continued

	554	LEU	uS19	533	LEU	EBNA1
	559	VAL	uS19	547	PRO	EBNA1
	559	VAL	uS19	607	PRO	EBNA1
	561	TYR	uS19	561	TYR	EBNA1
	563	MET	uS19	508	PHE	EBNA1
	563	MET	uS19	563	MET	EBNA1
	563	MET	uS19	561	TYR	EBNA1
	563	MET	uS19	600	PHE	EBNA1
	565	PHE	uS19	508	PHE	EBNA1
	565	PHE	uS19	539	LEU	EBNA1
	565	PHE	uS19	540	PRO	EBNA1
	588	ALA	uS19	529	PRO	EBNA1
	600	PHE	uS19	508	PHE	EBNA1
	604	VAL	uS19	508	PHE	EBNA1
	604	VAL	uS19	510	TYR	EBNA1
	604	VAL	uS19	539	LEU	EBNA1
	606	LEU	uS19	539	LEU	EBNA1
	606	LEU	uS19	540	PRO	EBNA1
	607	PRO	uS19	539	LEU	EBNA1
	607	PRO	uS19	541	PHE	EBNA1
	607	PRO	uS19	547	PRO	EBNA1
	607	PRO	uS19	559	VAL	EBNA1
	608	PRO	uS19	545	PRO	EBNA1
	609	TRP	uS19	541	PHE	EBNA1
	609	TRP	uS19	544	ALA	EBNA1
	609	TRP	uS19	551	PRO	EBNA1
	610	PHE	uS19	545	PRO	EBNA1
	610	PHE	uS19	610	PHE	EBNA1
	611	PRO	uS19	549	PRO	EBNA1
Ionic	117	GLU	uS19	514	LYS	EBNA1
	125	GLU	uS19	538	ARG	EBNA1
	493	HIS	uS19	581	ASP	EBNA1
	521	ARG	uS19	556	GLU	EBNA1
	556	GLU	uS19	521	ARG	EBNA1

5.3 Discussion

Overall, this study revealed ten possible protein interactions with uS19 protein (Table 5.2).

After further investigation, the data obtained demonstrated that the proteins predicted have a strong interaction with uS19 protein based on the low desolvation and global energy score

and RMSD value of less than 1 Å. The uS19 protein has already been established as a crucial protein in ribosome biogenesis and with extra ribosomal function. It is associated with disease pathways including cancer though no studies have reported its association with nasopharyngeal carcinoma (NPC). Thus, these ten potential interacting protein are helpful for understanding the potential extra ribosomal functions of uS19.

Transmembrane protein 237 isoform A (TMEM237) is a human protein (pdb ID: 3lo8) that is a component of the transition zone in primary cilia and required for ciliogenesis. TMEM237 protein also assists in the transport of substances across the membrane. Mutation in the TMEM237 protein is known to be linked with Joubert syndrome-related disorder (Huang et al., 2011). Studies done by Huang et al. (2011) reveals that when TMEM237 is knocked down, the normal ciliary function is affected, suggesting that the role of the protein in ciliogenesis. Their studies show the importance of TMEM237 as knockdown or mutation of the gene causes deregulation of Wnt Signaling and developmental defects in zebrafish. Thus, TMEM237 functions with other transition zone proteins in ciliogenesis, which is crucial for human development. uS19 protein based on the results of this study has a probable interaction with TMEM237 protein. The TMEM237 protein in the Wnt signalling pathway was found to elevate the β -catenin to disrupt the pathway causing carcinogenesis. Hence uS19's interaction with TMEM237 protein and subsequently might have a role in the NPC tumorigenesis since the Wnt signaling pathway is known to be linked to NPC carcinogenesis.

The uS19 protein has been known to be involved in ribosomal biogenesis including nuclear transports. From the data obtained through bioinformatics analysis, uS19 has been found to interact with 3 solute carriers, SLC22A23 (Solute carrier family 22 member 23; pdb ID:

4dmu), SLC13A4 (solute carrier family 13 member 4; pdb ID: 4f35) and cationic amino acid transporter (SLC7A1; pdb ID: 3k2k). Both SLC22A23 and SLC13A4 function to transport ions across the membrane, while SLC7A1 functions to transport amino acid across the membrane in the liver. The interaction of uS19 with these solute carriers could be possible due to its similar functions. SLC22A23 and SLC13A4 are involved in cancer progression as they are shown to be overexpressed in laryngeal squamous cell carcinoma (SCC) and differentially expressed in oral SCC respectively (Ekizoglu et al., 2018; Cha et al., 2011). Furthermore, uS19 protein was found to interact with RRP1B (rRNA processing protein 1 homolog B; pdb ID: 4jw2) which mediates amino acids for RNA stability, translation and subcellular localization in ribosome biogenesis. The *uS19* gene is located in the 40S subunit as mentioned previously, and assist in protein translation. RRP1B was shown to be highly expressed in breast cancer and underexpressed in laryngeal SCC (Kolenda et al., 2017; Lee et al., 2014). The interaction of uS19 with these proteins seems plausible due to its role in ribosome biogenesis and cancer.

CUBN (Cubilin) is an intrinsic receptor for vitamin B. It is crucial for kidney function and low protein levels leads to renal cancer. Cubilin alongside with Cobalamin (cobalamin binding intrinsic factor) assists in the transport of vitamin B. Cubilin is a membrane protein expressed in the intestinal epithelium and proximal renal tubule epithelium, and used as a predictive marker for treatment of renal cancer patients (Niinivirta et al., 2017). Studies done by Niinivirta et al. (2017) revealed the potential of cubilin used as a marker for renal cancer that interacts with sorafenib to assist in the progression-free survival of patients suffering from kidney cancer. This allows monitoring of the cancer progression based on the cubilin level (Niinivirta et al., 2017). High levels of cobalamin was reported to increase the risks of

cancers (Arendt et al., 2013). Studies conducted by Liang et al. (2018) suggested a role of uS19 in kidney cancer when its knockdown led to inhibition of kidney cancer progression. Hence uS19 protein's possible interaction with both cubilin and cobalamin binding intrinsic factor which have roles in renal cancer, indicates a potential association with their pathway.

The main focus on this study is to identify uS19 protein's potential role in nasopharyngeal carcinoma (NPC). Bioinformatics analysis was performed to find potential protein partners known to be involved in NPC progression. From the results obtained, uS19 protein was found to have a possible interaction with one heat shock protein (HSPA8) and two known Epstein-Barr virus (EBV) proteins [LMP2 (pdb ID: 1uxs) and EBNA1 (pdb ID: 5wmf)]. Furthermore, the RMSD value between uS19 and these 3 proteins were less than 1 Å, indicating likelihood of interaction.

HSPA8 (also known as HSC70) facilitates protein folding and signalling due to its location on the cell's surface. This protein has a role in both cell viability and tumor growth. HSPA8 has been found to mediate the anti-NPC effects of *Radix Ophiopogonis* through the P13K-AKT, Wnt and cAMP signalling pathway (Feng et al., 2019). Furthermore, EBNA1 induces the overexpression of HSPA8 in EBV-infected B cell surface during EBV infection (Stricher et al., 2013). This protein also plays a role in the mechanism of chemo-resistance in CNE1 and CNE1DDP cell line (Wang et al., 2018). HSPA8 was found to interact with EBNA-LP (known as EBNA-5) via Affinity Capture-MS and Western method (Forsman et al., 2008).

EBNA1 protein is expressed in all NPC tumours and is the only EBV protein that is required for the stability of EBV episomes (Frappier, 2012). This protein maintains NPC immortality

by facilitating the initiation of DNA synthesis, ensuring the even distribution of the viral episomes and activation of transcription of other viral latency genes crucial for cell immortalization (Frappier, 2012). The EBNA1 protein protects the cell from undergoing apoptosis and maintaining its survival by binding to various signalling pathways (Young & Dawson, 2014). It infects the cells, replicates, spreads to B cells resulting in latent infection in them, which then allows the cell to proliferate due to absence of T-cell (natural killer cells) (Kang & Kieff, 2015). According to Kang and Kieff (2015), EBNA1 is a sequence-specific DNA-binding protein to EBV element, a sequence-nonspecific chromosome association protein. It is a transactivator of viral latent genes and host genes responsible for episome replication, segregation and persistence of viral genome and oncogenesis. EBNA1 interacts with various pathways to protect the cell from undergoing apoptosis. It suppresses TGF- β 1 signaling in the p53 apoptosis pathway leading to p53 degradation, hence allowing cell proliferation (Teow et al., 2017). Bioinformatics simulation studies done by Sim and Talwar (2019), reported that EBNA1 protein interacted with several ribosome proteins. Hence the EBNA1-uS19 protein is plausible since EBNA1 has been reported to interact with ribosomal proteins that lead to NPC tumorigenesis (through *in silico* analysis).

The LMP2 proteins (LMP2A and LMP2B) are viral proteins whose mRNA are transcribed from different promoters, and are expressed during EBV latency in EBV-infected B cells (Sample et al., 1988; Rovedo & Longnecker, 2007). LMP2A is a major tyrosine kinase substrate in both transfected B-lymphoma cells and EBV-transformed B-lymphocytes. It inhibits B-cell receptor signalling and induces EBV lytic replication resulting in maintenance of cell survival allowing EBV to gain access to manipulate the memory B cell pool (Longnecker et al., 1991; Caldwell et al., 1996; Scholle et al., 2000; Rovedo &

Longnecker, 2007). Studies conducted by Revedo and Longnecker (2007) indicates that LMP2B associates with and significantly regulates LMP2A activity. The LMP2B protein restores the cellular tyrosine kinases Lyn levels when LMP2B is coexpressed with LMP2A, suggesting that LMP2B may be counteracting the association of Lyn (Revedo & Longnecker, 2007). The protein was also suggested to function in the initial EBV infection of B cells. It may be crucial for the expression of LMP2B in ensuring the cells do not undergo lytic replication by regulating the LMP2A activity (Revedo & Longnecker, 2007). As LMP2A is mainly involved in the B cells, LMP2B may be important in regulating the activation of lytic virus production from latently infected memory B cells (Revedo & Longnecker, 2007). Furthermore, LMP2B expression coordinates with LMP1 expression to allow for the expansion of cells harbouring EBV by preventing LMP2A-induced lytic replication, hence allowing proliferation of the EBV-infected cells (Revedo & Longnecker, 2007). LMP2 was also found to activate certain pathways in promoting NPC malignancies. It activates the Hedgehog signalling pathways by inducing cancer stem-like function, mTOR pathway by inhibiting cellular differentiation and promoting cancer cell survival, and P13K-AT pathway by encouraging cell survival and proliferation by counteracting with the TGF- β 1 effects (Teow et al., 2017).

Figure 5.13 shows an overall schematic diagram of the interactions with uS19 protein. Although the data obtained through this analysis reveals promising possible interactions, it is obtained through computational prediction. Computational predictions only allows a predictive insight based on similarities in structural form. The analysis obtained does not represent the physiological situation of the cell *in vivo* but based on its similarities in functions and highly predicted interaction results, there could be a possible link between

them. Further investigation using molecular techniques on the direct or indirect association of uS19 with these proteins should be done to fully understand uS19's role in NPC oncogenesis.

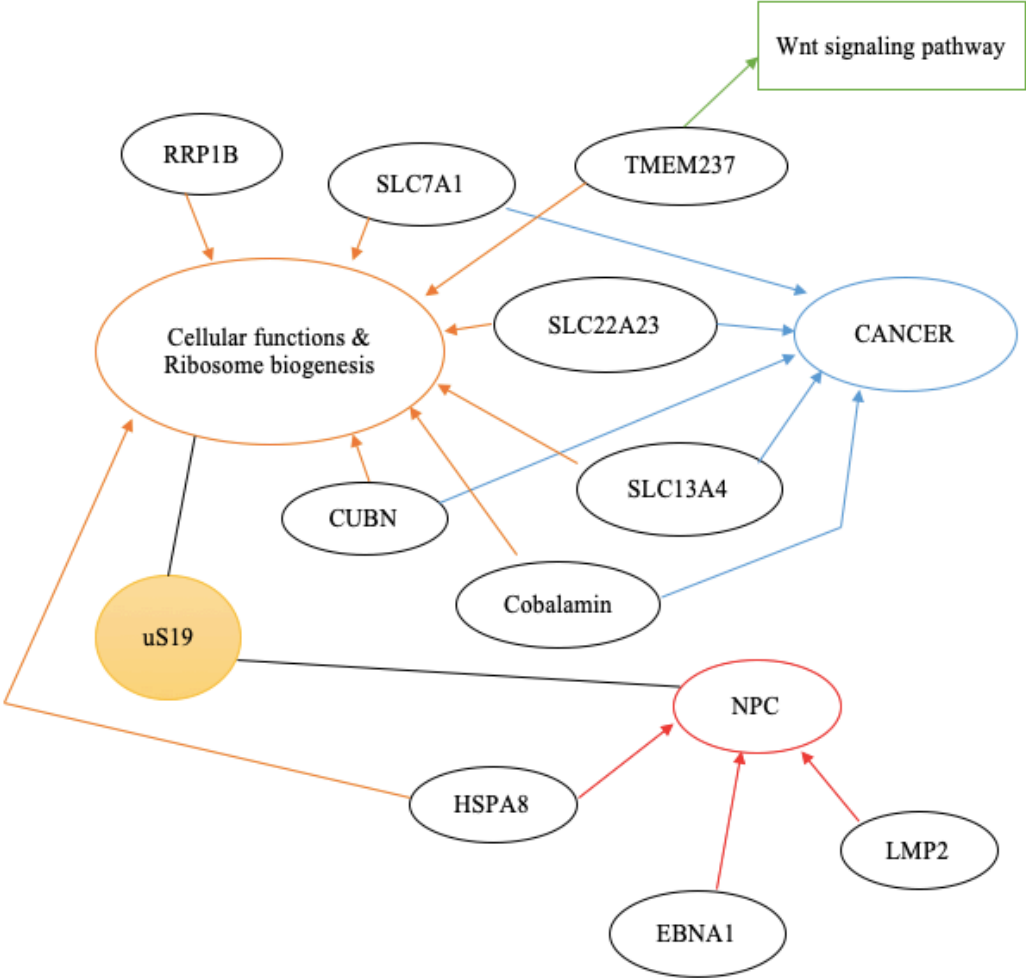


Figure 5.13: Schematic diagram of uS19 interaction with predicted partner proteins and their known association. Diagram illustrates uS19 protein possible involvement with predicted interacting proteins involved in known cellular and cancer pathways.

CHAPTER 6

GENERAL DISCUSSION

The *uS19* gene to this date is still understudied especially in relation to nasopharyngeal carcinoma (NPC). Previous research have found upregulation of this gene in NPC tissue (poorly differentiated squamous cell carcinoma) and cell line (HK1), but further validation is needed (Fang et al., 2008; Sim et al., 2018). Despite these studies of *uS19* gene at the transcriptional level, the protein expression of uS19 in NPC was never investigated. The results obtained in this study revealed a significant differential expression between transcriptional and translational levels, are indicated possible involvement of uS19 in NPC. The *uS19* transcriptional level presented a significant under expression that could have resulted from intron retention due to mutagenesis or change in the cell undergoing carcinogenesis. Studies have shown that intron retention causes low gene expression as shown in tumor tissues compared to normal tissues (Ren et al., 1995; Braunschweig et al., 2014; Dvinge & Bradley, 2015). Furthermore, a trend of downregulation of *uS19* gene with increased cell differentiation suggested cells deviating from its origin might have *uS19* gene involvement of in NPC tumorigenesis. When comparing *uS19* translational level, it was found to be significantly overexpressed, revealing a potential role in the cause of NPC. This gene has been known to interact in pathways leading to cancer and notably known to interact with a tumor suppressor, p53, although not in NPC cases. In NPC, p53 protein has been known to be involved in the apoptosis pathway where an upregulation of this protein was found to be associated with EBV infection and high levels of LMP1 (Tulalamba & Janvilisri, 2012). The uS19 protein in this study was significantly over expressed in an EBV cell line, which further validates its plausible role in NPC. This protein might interact with p53 in

NPC. Since p53 was highly expressed, this could explain uS19's overexpression at the protein level. Based on the results obtained in this study, uS19 might have a potential role in NPC tumorigenesis, possible interacting proteins should be identified before further conclusion about its role in NPC.

In silico approach was used to identify its potential interacting partners to obtain an insight into possible association of uS19 with NPC. Based on the results obtained, uS19 protein appears to be potentially interacting with TMEM237 protein, which is known to be involved in the Wnt signaling pathway. The dysregulation of the Wnt signaling pathway is famously known to be involved in carcinogenesis of many types of cancer, including NPC. This is potential evidence that uS19 might be associated with NPC. uS19 protein was found to interact with HSPA8 that functions in anti-NPC mechanisms but is also known to allow cell proliferation when interacting with EBNA1. Forsman et al. (2008) found EBNA-LP (EBNA5) to be interacting with HSPA8 through molecular methods. They also found uS19 interacting in the same complex, which further validates the interaction between HSPA8 and uS19. Moreover, the analysis also revealed that apart from the 10 predicted proteins, 2 known Epstein-Barr virus (EBV) antigens were found to have a high plausible interaction with uS19. Although the data obtained through this analysis is just a prediction, based on previous studies, uS19 has been shown to be involved in several pathways of diseases and cancer. Hence it is not surprising if uS19 has the potential to be involved in NPC carcinogenesis. The EBV-antigens are usually not sufficient to sustain tumorigenesis on its own, requiring other factors to support in its survivability. However multiple studies have shown that ribosomal proteins does play a role in NPC tumorigenesis.

This study revealed a novel potential role of uS19 in NPC that might be associated with EBV proteins and subsequently in NPC pathways. The fact that uS19 from previous studies have shown to be involved in the Mdm2-p53-MdmX pathway reveals potential role of uS19 with p53 deregulation in NPC involving the apoptosis pathway. The apoptosis pathways in NPC involves B-cell lymphoma 2 (BCL-2) protein that functions alongside LMP1 and p53 in NPC tumorigenesis by advancing cell growth (Tulalamba & Janvilisri, 2012). Studies have also shown mutated uS19 protein to be interacting with BCL-2 protein in chronic lymphocytic leukemia (CLL) (Landau et al., 2015). Furthermore, LMP2 have been known to work closely with LMP1 in tumorigenesis by encouraging cell proliferation of EBV-infected cells. The bioinformatics analysis obtained in this study revealed uS19 to be potentially associating with LMP2 protein. Based on published literature, the uS19 interaction with LMP2 could be possible due to its known association with the proteins involved in other cancers. The uS19 protein have been found to be assisting in cancer cell stabilization in other cancers, and could play the same role in NPC. Nevertheless, further molecular validation should be done as the results obtained is just a prediction. The EBNA1 protein, one of the main cause of NPC via EBV-antigen, functions in maintaining NPC immortality by infecting B cells. It was also found to potentially interact with uS19. As previously mentioned, uS19 protein was found to be potentially interacting with LMP2 protein, which is also known to interact with BCL-2 protein. Considering that both EBNA1 and LMP2 have been known to attack B cells, it is plausible that uS19 is involved with EBNA. No known studies have found any interaction of uS19 with EBNA1, though Forsman et al. (2008) have found uS19 to be interacting with Epstein-Barr nuclear antigen leader protein (EBNA-LP) using Affinity Capture and Western method. This study corroborates with my predictive interaction of uS19 with EBNA1. uS19 might be interacting with EBV

proteins through certain pathways involving BCL-2 protein or through EBNA complexes in inducing NPC malignancies. All docked complexes were refined to the optimum to obtain the best interacting scores. Though it is a computational prediction, the likeliness of its interaction based on the energy scores and RMSD values seem plausible for their interactions at the molecular level.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

The main objective of this study is to obtain an insight into the potential extra ribosomal role of uS19 in nasopharyngeal carcinoma (NPC) tumorigenesis. The *uS19* gene has been known to assist in not only ribosome biogenesis but also disease and cancer formation. The potential of this gene as a biomarker for early diagnosis is a possibility based on its wide variety of roles in cancer. The results obtained revealed differential expression level in mRNA and protein between NPC cell lines and normal nasopharyngeal epithelial (NPE) cell line, indicating a potential role of uS19 in carcinogenesis. To further conclude its potential, involvement *in silico* analysis was done which revealed 2 known Epstein-Barr virus (EBV) antigens along with TMEM237 protein (known to be involved in the Wnt signalling pathway) and HSPA8 protein (involved in both modifying cell viability and inhibition) to be interacting with uS19 protein (though it is just a computational prediction). The uS19 involvement in known pathways involving p53 makes it plausible for its involvement in NPC as p53 protein is also widely known to be associated with the apoptotic pathways involved in NPC formation. Furthermore, the Wnt signalling pathway is frequently studied in NPC as it is known to be one of the cancer causing pathways. Although the results obtained seems promising, further validation must be done to further affirm uS19's exact role in NPC tumorigenesis. Molecular techniques such as pulldown method and Mammalian di-Hybrid method should be done to verify their interactions. However, based on what that has been done in this study, it is safe to conclude that uS19 might have a potential role in NPC.

REFERENCES

- Arendt, J. F. B., Pederson, L., Nexø, E., Sørensen, H. T. (2013). Elevated plasma vitamin B12 level as a marker for cancer: A population-based cohort. *Journal National Cancer Institute*, 105(23), 1799-1805.
- Arthurs, C., Murtaza, B. N., Thomson, C., Dickens, K., Henrique, R., Patel, H. R. H., Beltran, M., Millar, M., Thrassivoulou, C., & Ahmed, A. (2017). Expression of ribosomal proteins in normal and cancerous human prostate tissue. *PLOS One*, 12(10), e0186047. <https://doi.org/10.1371/journal.pone.0186047>.
- Aziz, A., Ramli, R. R., Mohamad, I., & Bhavaraju, V. M. K. (2017). Young nasopharyngeal carcinoma: a review of an 8-year experience in the East Coast Malaysia Hospital. *The Egyptian Journal of Otolaryngology*, 33(2), 490-494.
- Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function and Bioinformatics*, 71(1), 261-277.
- Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Structural Bioinformatics*, 27(3), 343-350.
- Benton, D. (1996). Bioinformatics – principles and potential of a new multidisciplinary tool. *Introductory Overview*, 14(8), 261-272.

- Bienert, S., Waterhouse, A., de Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository- new features and functionality. *Nucleic Acids Research*, 45, D313-D319.
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., & Blencowe, B. J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Research*, 24(11), 1774-1786.
- Broccolo, F., Ciccarese, G., Rossi, A., Anselmi, L., Drago, F., & Toniolo, A. (2018). Human papillomavirus (HPV) and Epstein-Barr virus (EBV) in keratinizing versus non-keratinizing squamous cell carcinoma of oropharynx. *Infectious Agents and Cancer*, 13(32). <https://doi.org/10.1186/s13027-018-0205-6>.
- Caldwell, R. G., Wilson, J. B., Anderson, S. J., & Longnecker, R. (1998). Epstein-Barr Virus LMP2A Drives B Cell Development and Survival in the Absence of Normal B Cell Receptor Signals. *Immunity*, 9(3), 405-411.
- Carugo, O., & Djinojic-Carugo, K. (2013). A proteomic Ramachandran plot (PR plot). *Amino Acids*, 44(2), 781-790.
- Cha, J. D., Kim, H. J., & Cha, I. H. (2011). Genetic alterations in oral squamous cell carcinoma progression detected by combining array-based comparative genomic hybridization and multiplex ligation-dependent probe amplification. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics*, 111(5), 594 – 607.

- Chang, E. T., & Adami, H. O. (2006). The Enigmatic Epidemiology of Nasopharyngeal Carcinoma. *Cancer Epidemiol Biomarkers*, 15(10), 1765-1777.
- Chen, G., Shen, C., Wang, X., Wang, H., Liu, Y., Yu, C., Iv, J., He, J., & Wen, Z. (2017). Identification of genes and pathways in nasopharyngeal carcinoma by bioinformatics analysis. *Oncotarget*, 8(38), 63738-63749.
- Cheung, S. T., Huang, D. P., Hui, A. B., Lo, K. W., Tsang, Y. S., Wong, N., Whitney, B. M., & Lee, J. C. (1999). Nasopharyngeal carcinoma cell line (C666-1) consistently harbouring Epstein-Barr virus. *International Journal of Cancer*, 83(1), 121-126.
- Closs, E. I., Boissel, J. P., Habermeier, A., & Rotmann, A. (2006). Structure and function of cationic amino acid transporters (CATs). *Journal of Membrane Biology*, 213(2), 67-77.
- Daftuar, L., Yan, Z., Jacq, X., & Prives, C. (2013). Ribosomal proteins RPL37, RPS15 and RPS20 regulate the Mdm2-p53-MdmX network. *PLOS One*, 8(7), e68667. <https://doi.org/10.1371/journal.pone.0068667>.
- Debi, B. C. R., Pisani, P., Tang, T. S., & Parking, D. M. (2004). High Incidence of Nasopharyngeal Carcinoma in Native People of Sarawak, Borneo Island. *Cancer Epidemiology, Biomarkers and Prevention*, 13(3), 482-486.
- Devereux, T. R., Risinger, J. I., & Barrett, J. C. (1998). Mutations and altered expression of the human cancer genes: what they tell us about causes. *IARC Scientific Publications*, (146), 19-42.

- Dong, J. Q., Li, M. Z., Liu, Z. G., Zhong, Q., Xiong, D., Xu, L. H., Du, Y., Xia, Y. F., & Zeng, M. S. (2012). Establishment and characterization of a novel nasopharyngeal carcinoma cell line (SUNE2) from a Cantonese patient. *Chinese Journal of Cancer*, 31(1), 36-44.
- Dvinge, H., & Bradley, R. K. (2015). Widespread intron retention diversifies most cancer transcriptomes. *Genome Medicine*, 7(45). doi 10.1186/s13073-015-0168-9.
- Elia, A., Vyas, J., Laing, K. G., & Clemens, M. J. (2004). Ribosomal protein L22 inhibits regulation of cellular activities by the Epstein-Barr virus small RNA EBER-1. *European Journal of Biochemistry*, 271(10), 1895-1905.
- Ekizoglu, S., Seven, D., Ulutin, T., Guliyev, J., & Buyru, N. (2018). Investigation of the *SLC22A23* gene in laryngeal squamous cell carcinoma. *BMC Cancer*, 18(477). <https://doi.org/10.1186/s12885-018-4381-y>.
- Eswaran, J., Horvath, A., Godbole, S., Reddy, S. D., Mudvari, P., Ohshiro, K., Cyanam, D., Nair, S., Fuqua, S. A. W., Polyak, K., Florea, L. D., & Kumar, R. (2013). RNA sequencing of cancer reveals novel splicing alterations. *Scientific Reports*, 3(1689). doi: 10.1038/srep01689.
- Fang, W., Li, X., Jiang, Q., Liu, Z., Yang, H., Wang, S., Xie, S., Liu, Q., Liu, T., Huang, J., Xie, W., Li, Z., Zhao, Y., Wang, E., Marincola, F. M., & Yao, K. (2008). Transcriptional patterns, biomarkers and pathways characterizing nasopharyngeal carcinoma of Southern China. *Journal of Translational Medicine*, 6(32). <https://doi.org/10.1186/1479-5876-6-32>.

- Feng, X., Shi, H., Chao, X., Zhao, F., Song, L., Wei, M., & Zhang, H. (2019). Deciphering the pharmacological mechanism of the herb *Radix Ophiopogonis* in the treatment of nasopharyngeal carcinoma by integrating iTRAQ-coupled 2-D LC-MS/MS analysis and network investigation. *Frontiers in Pharmacology*, 10(253). doi: 10.3389/fphar.2019.00253.
- Forsman, A., Ruetschi, U., Ekholm, J., & Rymo, L. (2008). Identification of Intracellular Proteins Associated with the EBV-Encoded Nuclear Antigen 5 Using an Efficient TAP Procedure and FT-ICR Mass Spectrometry. *Journal of Proteome Research*, 7(6), 2309-2319.
- Frappier, L. (2012). Role of EBNA1 in NPC tumorigenesis. *Seminars in Cancer Biology*, 22(2), 154-161.
- Fung, L. F., Lo, A. K. F., Yuen, P. W., Liu, Y., Wang, X. H., & Tsao, S. W. (2000). Differential gene expression in nasopharyngeal carcinoma cells. *Life Sciences*, 67(8), 923-936.
- Gibrat, J. F., Madej, T., & Bryant, S. H. (1996). Surprising similarities in structure comparison. *Current Opinion in Structural Biology*, 6(3), 377-385.
- Glaser, R., Zhang, H. Y., Yao, K. T., Wang, F. X., Li, G. Y., Wen, D. S., & Li, Y. P. (1989). Two epithelial tumor cell lines (HNE-1 and HONE-1) latently infected with Epstein-Barr virus that were derived from nasopharyngeal carcinomas. *Proceedings of National Academy of Sciences of the United States of America*, 86(23), 9524-9528.

- Gullo, C., 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 Academy of Medicine Singapore*, 37(9), 769-777.
- Hildesheim, A., West, S., DeVeyra, E., De Guzman, M. F., Jurado, A., Jones, C., Imai, J., & Hinuma, Y. (1992). Herbal Medicine Use, Epstein-Barr Virus, and Risk of Nasopharyngeal Carcinoma. *Cancer Research*, 52(11), 3048-3051.
- Hildesheim, A., Dosemeci, M., Chan, C. C., Chen, C. J., Mittl, B. F., Sun, B., Levine, R. H., Chen, J. Y., Brinton, L. A., & Yang, C. Z. (2001). Occupational Exposure to Wood, Formaldehyde, and Solvents and Risk of Nasopharyngeal Carcinoma. *Cancer Epidemiology, Biomarkers and Prevention*, 10(11), 1145-1153.
- Hildesheim, A., & Wang, C. P. (2012). Genetic Predisposition Factors and Nasopharyngeal Carcinoma Risk: A Review of Epidemiological Association Studies, 2000– 2011. *Seminars in Cancer Biology*, 22(2), 107-116.
- Ho, B. K., Thomas, A., & Brasseur, R. (2003). Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the α -helix. *Protein Science*, 12(11), 2508-2522.
- Hsu, W. L., Yu, K. J., Chien, Y.C., Chiang, C. J., Cheng, Y. J., Chen, J. Y., Liu, M. Y., Chou, S. P., You, S. L., Hsu, M. M., Lou, P. J., Wang, C. P., Hong, J. H., Leu, Y. S., Tsai, M. H., Su, M. C., Tsai, S. T., Chao, W. Y., Ger, L. P., Chen, P. R., Yang, C. S., Hildesheim, A., Diehl, S. R., & Chen, C. J. (2010). Familial tendency and risk of nasopharyngeal carcinoma in Taiwan: Effects of covariates on risk. *American Journal of Epidemiology*, 173(3), 292-299.

- Huang, D. P., Ho, J. H., Saw, D., & Teoh, T. B. (1978). Carcinoma of the nasal and paranasal regions in rats fed Cantonese salted marine fish. *IARC Scientific Publications*, (20), 315-328.
- Huang, D. P., Ho, J. H. C., Poon, Y. F., Chew, E. C., Saw, D., Lui, M., Li, C. L., Mak, L. S., Lai, S. H., & Lau, W. H. (1980). Establishment of cell line (NPC/HK1) from a differentiated squamous carcinoma of the nasopharynx. *International Journal of Cancer*, 26(2), 127-132.
- Huang, L., Szymanska, K., Jensen, V. L., Janecke, A. R., Innes A. M., Divis, E. E., Frosk, P., Li, C., Willer, J. R., Chodirker, B. N., Greenberg, C. R., McLeod, D. R., Bernier F. P., Chudley, A. E., Muller, T., Shboul, M., Logan, C. V., Loucks, C. M., Beaulieu, C. L., Bowie, R. V., Bell, S. M., Adkins, J., Zuniga, F. I., Ross, K. D., Wang, J., Ban, M. R., Becker, C., Numberg, P., Douglas, S., Craft, C. M., Akimenko, M. A., Hegele, R. A., Ober, C., Utermann, G., Bolz, H. J., Bulman, D. E., Katsanis, N., Blacque, O. E., Doherty, D., Parboosingh, J. S., Leroux, M. R., Johnson, C. A., & Boycott, K. M. (2011). TMEM237 is mutated in individuals with a Joubert syndrome related disorder and expands the role of the TMEM family at the ciliary transition zone. *American Journal of Human Genetics*, 89(6), 713-730.
- Ishii, K., Washio, T., Uechi, T., Yoshihama, M., Kenmochi, N., & Tomita, M. (2006). Characteristics and clustering of human ribosomal protein genes. *BMC Genomics*, 7(37). Published 2006 Feb 28. doi:10.1186/1471-2164-7-37.
- Kang, M. S., & Kieff, E. (2015). Epstein-Barr virus latent genes. *Experimental & Molecular Medicine*, 47(131), e131; doi:10.1038/emm.2014.84.

- Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R. C., Khadake, J., Mahadevan, U., Patrick, M., Perduzzi, I., Pfeifferberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., & Hermjakob, H. (2012). The IntAct molecular interaction database in 2012. *Nucleic Acids Research*, 40, D841-D846.
- Kitagawa, M., Takasawa, S., Kikuchi, N., Takako, I., Teraoka, H., Yamamoto, H., & Okamoto, H. (1991). *Rig* encodes ribosomal protein S15 The primary structure of mammalian ribosomal protein S15. *FEBS Letters*, 283(2), 210-214.
- Kolenda, T., Guglas, K., Rys, M., Bogaczynska, M., Teresiak, A., Blizniak, R., Lasinska, I., Mackiewicz, J. & Kamperska, K. M. (2017). Biological role of long non-coding RNA in head and neck cancers. *Reports of Practical Oncology and Radiotherapy*, 22(5), 378-388.
- Kondoh, N., Schweinfest, C. W., Henderson, K. W., & Papas, T. S. (1992). Differential expression of S19 ribosomal protein, laminin-binding protein, and human lymphocyte antigen class I messenger RNAs associated with colon carcinoma progression and differentiation. *Cancer Research*, 52(4), 791-796.
- Kothapalli, R., Yoder, S.J., Mane, S., & Loughran, T.P. (2002). Microarray results: how accurate are they? *BMC Bioinformatics*, 3(22).
<http://www.biomedcentral.com/1471-2015/3/22>.

- Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*, 12(2), 255-278.
- Lahuri, Y. S., Mohamad, I., & Hashim, H. (2015). Nasopharyngeal carcinoma with metastases to colon. *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*, 16(2), 201-204.
- Landau, D. A., Tausch, E., Taylor-Weiner, A. N., Stewart, C., Reiter, J. G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Bottcher, S., Carter, S. L., Cibulskis, K., Mertens, D., Sougnez, C. L., Rosenberg, M., Hess, J. M., Edelmann, J., Kless, S., Kneba, M., Ritgen, M., Fink, A., Fischer, K., Gabriel, S., Lander, E. S., Nowak, M. A., Dohner, H., Hallek, M., Neuberg, D., Getz, G., Stilgenbauer, S., & Wu, C. J. (2015). Mutations driving CLL and their evolution in progression and relapse. *Nature*, 526, 525-530.
- Lawdowski, M., O'Donohue, M. F., Buross, C., Ghazvinian, R., Montel-Lehry, N., Vlachos, A., Sieff, C. A., Newburger, P. E., Niewiadomska, E., Matysiak, M., Gader, B., Atsidaftos, E., Lipton, J. M., Beggs, A. H., Gleizes, P. E., & Gazda, H. T. (2013). Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond-Blackfan Anemia. *Human Genetics*, 132(11), 1265-1274.
- Lee, M., Dworkin, A. M., Lichtenberg, J., Patel, S. J., Gildea, D., Bodine, D. M., & Crawford, N. P. S. (2014). Metastasis-associated protein ribosomal RNA processing 1 Homolog B (RRP1B) modulates metastasis through regulation of histone methylation. *Molecular Cancer Research*, 12(12), 1818-1828.

- Léger-Silvestre, I., Milkereit, P., Ferreira-Cerca, S., Saveanu, C., Rousselle, J. C., Choismel, V., Guinefoleau, C., Gas, N., & Gleizes, P. E. (2004). The ribosomal protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *The EMBO Journal*, 23(12), 2336-2347.
- Liang, J., Liu, Z., Zou, Z., Wang, X., Tang, Y., Zhou, C., Wu, K., Zhang, F., & Lu, Y. (2018). Knockdown of ribosomal protein S15A inhibits human kidney cancer cell growth *in vitro* and *in vivo*. *Molecular Medicine Reports*, 19(2), 1117-1127.
- Lin, C. T., Wong, C. I., Chan, W. Y., Tzung, K. W., Ho, J. K., Hsu, M. M., & Chuang S. M. (1990). Establishment and characterization of two nasopharyngeal carcinoma cell lines. *Laboratory Investigation*, 62(6), 713-724.
- Lin, C. T., Chan, W. Y., & Chen, W. (1993). Characterization of seven newly established nasopharyngeal carcinoma cell line. *Laboratory Investigation*, 68(6), 716-727.
- Liu, Z. (2016). *Etiology and early detection of nasopharyngeal carcinoma – an epidemiological approach*. Stockholm: Karolinska Institute.
- Lo, A. K. F., Liu, Y., Wang, X. H., Huang, D. P., Yuen, P. W., Wong, Y. C., & Tsao, G. S. W. (2003). Alterations of biologic properties and gene expression in nasopharyngeal epithelial cells by the Epstein-Barr virus-encoded latent membrane protein 1. *Laboratory Investigation*, 83(5), 697-709.
- Longnecker, R., Drunker, B., Roberts, T. M., & Kieff, E. (1991). An Epstein-Barr Virus Protein Associated with Cell Growth Transformation Interacts with a Tyrosine Kinase. *Journal of Virology*, 65(7), 3681-3692.

- 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(4), 346-358.
- Mahdavifar, N., Ghoncheh, M., Mohammadian-Hafshejani, M., Khosravi, B., & Salehiniya, H. (2016). Epidemiology and Inequality in the Incidence and Mortality of Nasopharynx Cancer in Asia. *Osong Public Health and Research Perspectives*, 7(6), 360-372.
- Marion, M. J., & Marion, C. (1988). Ribosomal proteins S2, S6, S10, S14, S15 and S25 are localized on the surface of mammalian 40 S subunits and stabilize their conformation A study with immobilized trypsin. *FEBS Letters*, 232(2), 281-285.
- Martin, I., Kim, J. W., Lee, B. D., Kang, H. C., Xu, J. C., Jia, H., Stankowski, J., Kim, M. S., Zhong, J., Kumar, M., Andrabi, S. A., Xiong, Y., Dickson, D. W., Wszolek, Z. K., Pandey, A., Dawson, T. M., & Dawson, V. L. (2014). Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. *Cell*, 157(2), 472-485.
- Nasopharyngeal cancer. (2013). *American Cancer Society*.
<http://www.cancer.org/acs/groups/cid/documents/webcontent/003124-pdf.pdf>.
- Naora, H., & Naora, H. (1999). Involvement of ribosomal proteins in regulating cell growth and apoptosis: Translational modulation or recruitment for extraribosomal activity? *Immunology and Cell Biology*, 77(3), 197-205.

- Niinivirta, M., Enblad, G., Edqvist, P. H., Ponten, F., Dragomir, A., & Ullenhag, G. J. (2017). Tumoral cubilin is a predictive marker for treatment of renal cancer patients with sunitinib and sorafenib. *Journal of Cancer Research and Clinical Oncology*, 143(6), 961-970.
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N. H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R. C., Meldal, B., Melidoni, A. N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., Roey, K. V., Cesareni, G., & Hermjakob, H. (2014). The MIntAct project – IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, 42, D358-D363.
- Owen, T.A., Aronow, M.S., Barone, L.M., Bettencourt, B., Stein, G.S., & Lian, J.B. (1991). Pleiotropic Effects of Vitamin D on Osteoblast Gene Expression Are Related to the Proliferative and Differentiated State of the Bone Cell Phenotype: Dependency upon Basal Levels of Gene Expression, Duration of Exposure, and Bone Matrix Competency in Normal Rat Osteoblast Cultures*. *Endocrinology*, 128(3), 1496-1504.
- Panchenko, A. R., & Madej, T. (2004). Analysis of Protein Homology by Assessing the (Dis)similarity in Protein Loop Regions. *Proteins*, 57(3), 539-547.

- Pathmanathan, R., Prasad, U., Sadler, R., Flynn, K., & Raab-Traub, N. (1995). Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesion related to nasopharyngeal carcinoma. *The New England Journal of Medicine*, 333(11), 693-698.
- Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q., & Spector, D. L. (2005). Regulating Gene Expression through RNA Nuclear Retention. *Cell*, 123(2), 249-263.
- Ren, S., Peng, Z., Mao, J. H., Yu, Y., Yin, C., Gao, X., Cui, Z., Zhang, J., Yi, K., Xu, W., Chen, C., Wang, F., Guo, X., Lu, J., Yang, J., Wei, M., Tian, Z., Guan, Y., Tang, L., Xu, C., Wang, L., Gao, X., Tian, W., Wang, J., Yang, H., Wang, J., & Sun, Y. (2012). RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Research*, 22(5), 806-821.
- Robledo, S., Idol, R. A., Crimmins, D. L., Ladenson, J. H., Mason, P. J., & Bessler, M. (2008). The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA*, 14(9), 1918-1929.
- Rouquette, J., Choismel, V., & Gleizes, P. E. (2005). Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *The EMBO Journal*, 24(16), 2862-2872.
- Rovedo, M., & Longnecker, R. (2007). Epstein-Barr Virus Latent Membrane Protein 2B (LMP2B) Modulates LMP2A Activity. *Journal of Virology*, 81(1), 84-94.

- Salehiniya, H., Mohammadian, M., Mohammadian-Hafshejani, A., & Mahdavifar, N. (2018). Nasopharyngeal Cancer in the World: Epidemiology, Incidence, Mortality and Risk Factors. *World Cancer Research Journal*, 5(1), e1046. <https://www.wcrj.net/article/1046>.
- Sample, J., Liebowitz, D., & Kieff, E. (1989). Two Related Epstein-Barr Virus Membrane Proteins are Encoded by Separate Genes. *Journal of Virology*, 63(2), 933-937.
- Scholle, F., Bendt, K. M., & Raab-Traub, N. (2000). Epstein-Barr Virus LMP2A Transforms Epithelial Cells, Inhibits Cell Differentiation, and Activates Akt. *Journal of Virology*, 74(22), 10681-10689.
- Shen, C. L., Lie, C. D., You, R. I., Ching, Y. H., Liang, J., Ke, L., Chen, Y. L., Chen, H. C., Hsu, H. J., Liou, J. W., Kieff, E., & Peng, C. W. (2016). Ribosome Protein L4 is essential for Epstein-Barr Virus Nuclear Antigen 1 function. *Proceedings of the National Academy of Sciences of the United States of America*, 113(8), 2229-2234.
- Sim, E. U. H., Chan, S. L. L., Ng, K. L., Lee, C. W., & Narayanan, K. (2016). Human Ribosomal Proteins RPeL27, RPeL43, and RPeL41 are Upregulated in Nasopharyngeal Carcinoma Cell Lines. *Disease Markers*, 2016(5179594). <http://dx.doi.org/10.1155/2016/5179594>.
- Sim, E. U. H., Ma, X. R., Chan, S. L. L., Lee, C. W., & Narayanan, K. (2016). Predicted interaction of human Ribosomal Protein S15 with Fragile X Mental Retardation Protein. *Journal of Applied Biology and Biotechnology*, 4(2), 38-45.

- Sim, E. U. H., Chee, C. S. M., Vasudevan, L., Ng, K. L., & Chan, S. L. L. (2018). Selective Differential Expression of the Ribosomal Protein Genes *eL14* and *uS19* in a Well-Differentiated Epithelial Cell Line of Nasopharyngeal Carcinoma. *Malaysian Applied Biology*, 47(1), 247-253.
- Sim, E. U. H., & Talwar, S. P. (2019). In silico evidence of de novo interactions between ribosomal and Epstein-Barr virus proteins. *BMC Molecular and Cell Biology*, 20(34). <https://doi.org/10.1186/s12860-019-0219-y>.
- Spratt, D. E., & Lee, N. (2012). Current and emerging treatment options for nasopharyngeal carcinoma. *Onco Targets and Therapy*, 5, 297-308.
- Stelow, E. B., & Wenig, B. M. (2017). Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: Nasopharynx. *Head Neck Pathology*, 11(1), 16-22.
- Stelzl, U., Connell, S., Nierhaus, K. H., & Wittmann-Liebold, B. (2001). Ribosomal proteins: Role in ribosomal functions. *Encyclopedia of Life Science*. <https://doi.org/10.1002/9780470015902.a0000687.pub4>.
- Stricher, F., Macri, C., Ruff, M., & Muller, S. (2013). HSPA8/HSC70 chaperone protein: Structure, function, and chemical targeting. *Autophagy*, 9(12), 1937-1954.
- Tabuchi, K., Nakayama, M., Nishimura, B., Hayashi, K., & Hara, A. (2011). Early Detection of Nasopharyngeal Carcinoma. *Hindawi Publishing Corporation*, 2011(638058). doi:10.1155/2011/638058.

- Teow, S. Y., Yap, H. Y., & Peh, S. C. (2017) Epstein-Barr virus as a promising immunotherapeutic target for nasopharyngeal carcinoma treatment. *Hindawi*, 2017.
- Tiong, T. S., & Selva, K. S. (2005). Clinical presentation of nasopharyngeal carcinoma in Sarawak Malaysia. *Medical Journal Malaysia*, 60(5), 624-628.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., & Pachter, L., (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), 511-515.
- Tsao, S. W., Wang, X., Liu, Y., Cheung, Y. C., Feng, H., Zheng, Z., Wong, N., Yuen, P. W., Lo, A. K., Wong, Y. C., & Huang, D. P. (2002). Establishment of two immortalized nasopharyngeal epithelial cell lines using SV40 large T and HPV16E6/E7 viral oncogenes. *Biochimica et Biophysica Acta*, 1590(1-3), 150-158.
- 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.
- Tulalamba, W., & Janvilisri, T. (2012). Nasopharyngeal Carcinoma Signaling Pathway: An Update on Molecular Biomarkers. *International Journal of Cell Biology*, 2012(594681). doi:10.1155/2012/594681.
- Vogel, C., & Marcotte, E. M. (2013). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4), 227-232.

- Wang, K., Chen, Z., Long, L., Tao, Y., Wu, Q., Xiang, M., Liang, Y., Xie, X., Jiang, Y., Xiao, Z., Yan, Y., Qiu, S., & Yi, B. (2018). iTRAQ-based quantitative proteomic analysis of differential expressed proteins in chemoresistant nasopharyngeal carcinoma. *Cancer Biology & Therapy*, 19(9), 809-824.
- Wee, J. T., Ha, T. C., Loong, S. L., & Qian, C. N. (2010). Is nasopharyngeal cancer really a “Cantonese cancer”? *Chinese Journal of Cancer*, 29(5), 517-526.
- Wei, K., Zheng, R., Zhang, S., Liang, Z., Ou, Z., & Chen, W. (2014). Nasopharyngeal carcinoma incidence and mortality in China in 2010. *Chinese Journal of Cancer*, 33(8), 381-387.
- Wild, T., Horvath, P., Wyler, E., Widmann, B., Badertscher, L., Zemp, I., Kozak, K., Csucs, G., Lund, E., & Kutay, U. (2010). A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60 S subunit export. *PLOS Biology*, 8(10), e1000522. <https://doi.org/10.1371/journal.pbio.1000522>.
- Wong, Q. W., Li, J., Ng, S. R., Lim, S. G., Yang, H., & Vardy, L. A. (2014). RPL39L is an example of a recently evolved ribosomal protein paralog that shows highly specific tissue expression patterns and is upregulated in ESCs and HCC tumors. *RNA Biology*, 11(1), 33-41.
- Young, L. S., & Dawson, C. W. (2014). Epstein-Barr virus and nasopharyngeal carcinoma. *Chinese Journal of Cancer*, 33(12), 581-590.

- Yu., M. C., Nichols, P. W., Zou, X. N., Estes, J., & Henderson, B. E. (1989). Induction of malignant nasal cavity tumours in Wistar rats fed Chinese salted fish. *British Journal of Cancer*, 60(2), 198-201.
- Zheng, Y. M., Tuppin, P., Hubert, A., Jeannel, D., Pan, Y. J., Zeng, Y., & de The, G. (1994). Environmental and dietary risk factors for nasopharyngeal carcinoma: a case- control study in Zangwu County, Guangxi, China. *British Journal of Cancer*, 69(3), 508-514.
- Zhou, X., Liao, W. J., Liao, J. M., Liao, P., & Lu, H. (2015). Ribosomal protein: functions beyond the ribosome. *Journal of Molecular Cell Biology*, 7(2), 92-104.

APPENDIX

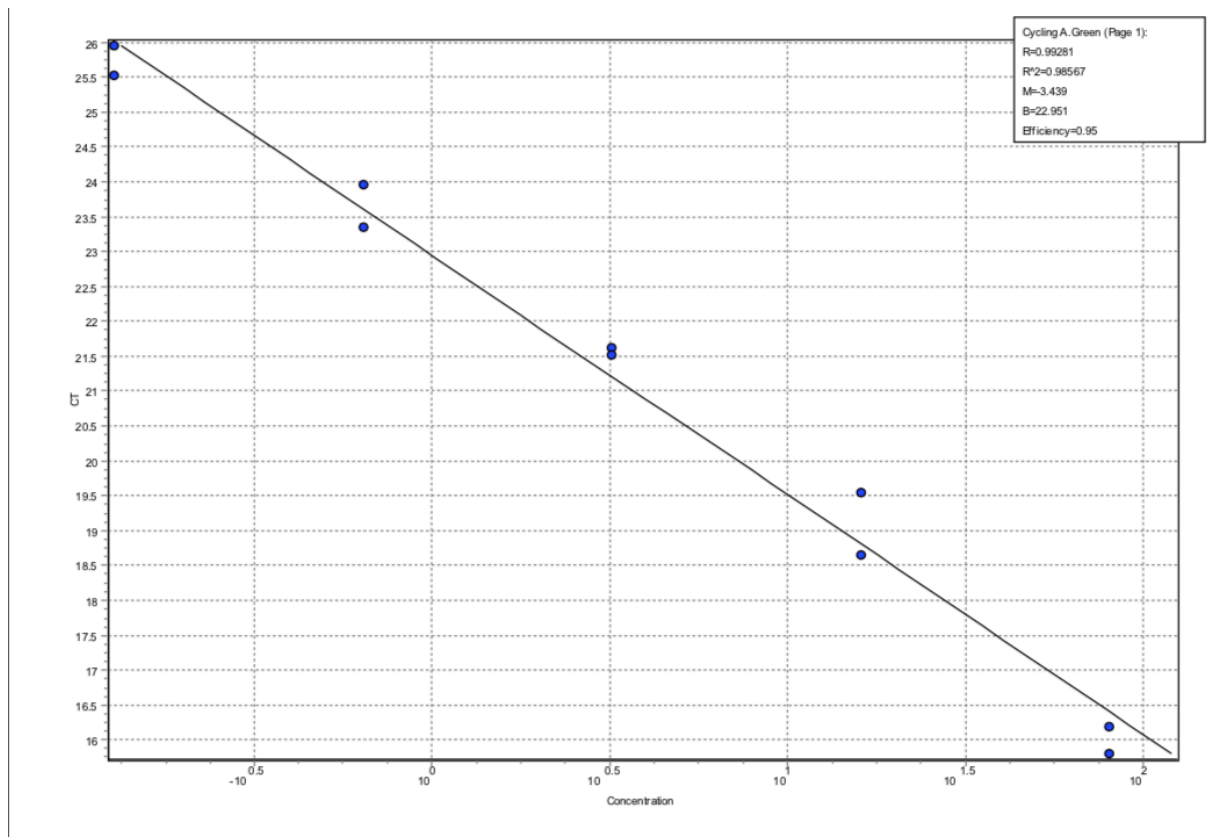


Figure A.1: Standard curve for *uS19* primer validation

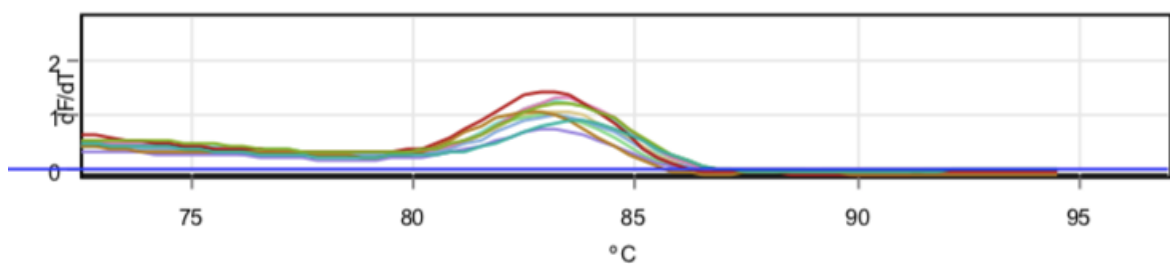


Figure A.2: Melting curve for *uS19*

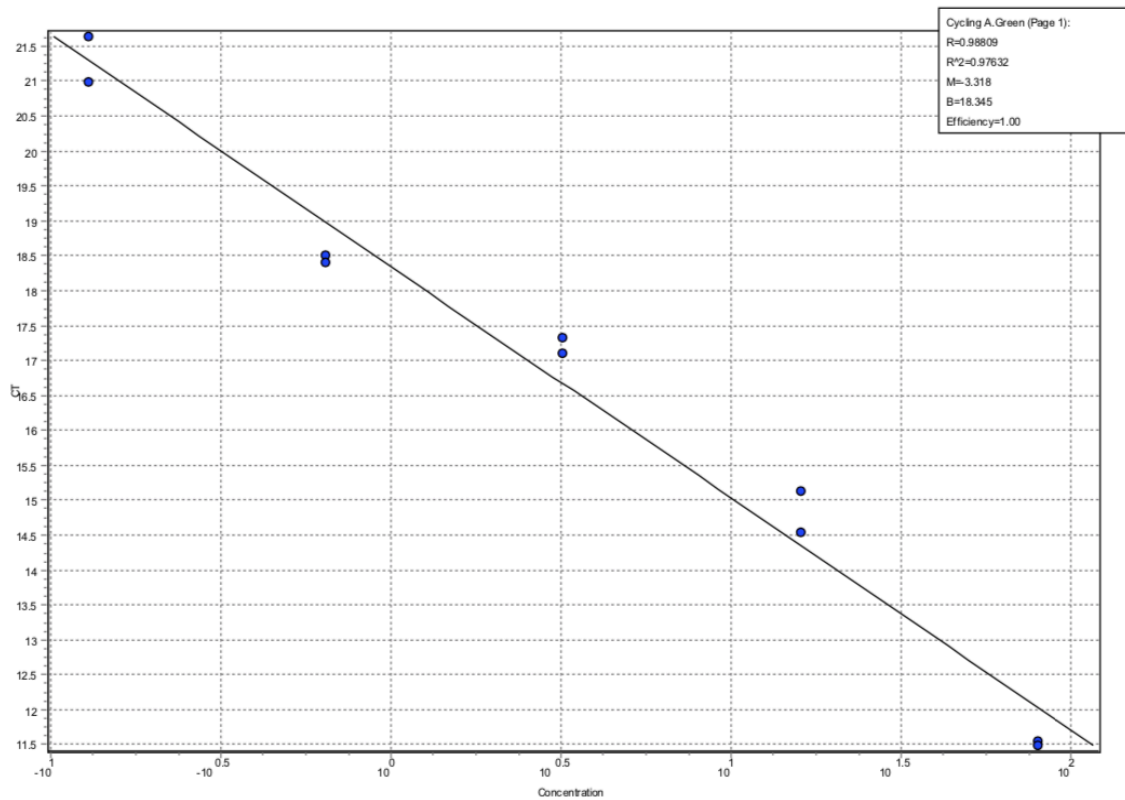


Figure A.3: Standard curve for *GAPDH* primer validation

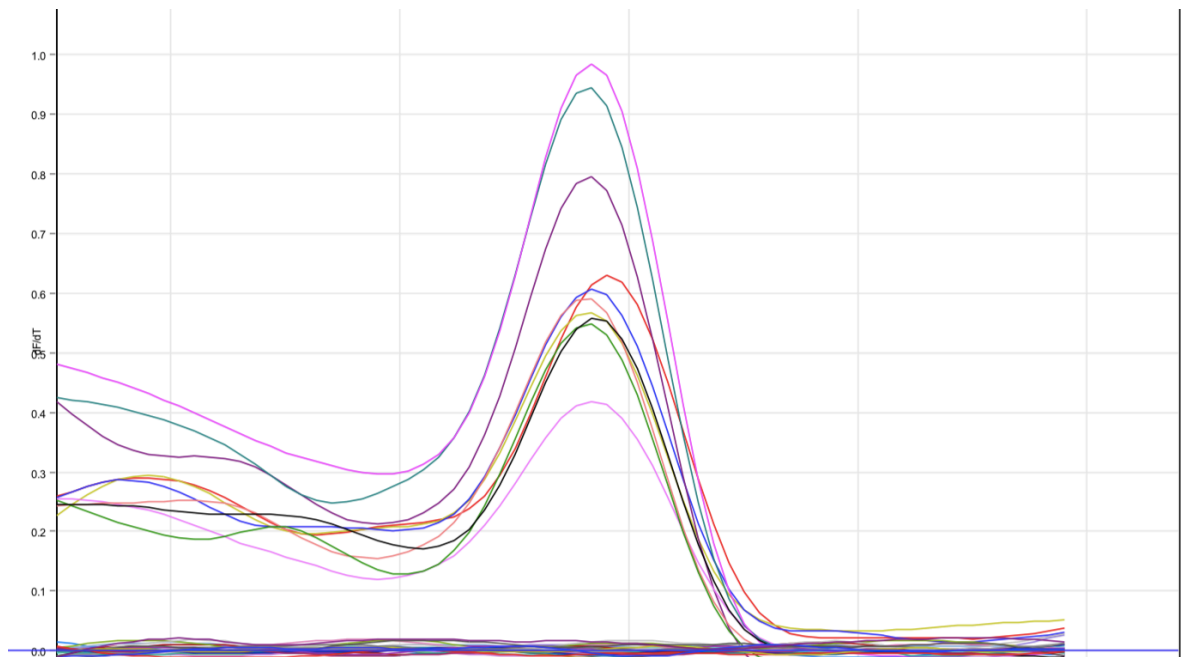


Figure A.4: Meting curve for *GAPDH*

Table A.1: Raw Ct value data for *GAPDH* (TR stands for technical replicate)

	NP69	HK1	HONE1	TW 01	TW 04	C666-1
Biological Replicate 1						
TR1	22.93	16.14	20.89	17.04	22.11	21.46
TR2	22.67	16.22	20.84	17.28	22.11	22.8
Biological Replicate 2						
TR1	18.74	16.62	17.58	18.07	18.11	16.65
TR2	18.56	16.63	17.53	18.19	18.01	16.54
Biological Replicate 3						
TR1	19.41	17.31	18.41	18	18.47	28.52
TR2	19.41	17.37	18.26	18.12	18.4	28.4

Table A.2: Raw Ct value data for *uS19* (TR stands for technical replicate)

	NP69	HK1	HONE1	TW 01	TW 04	C666-1
Biological Replicate 1						
TR1	22.19	18.58	19.78	22.28	21.11	21.55
TR2	22.33	18.45	19.77	22.16	21.2	21.91
Biological Replicate 2						
TR1	18.06	16.53	18.93	18	19.78	17.75
TR2	18.07	16.58	18.94	17.87	19.75	16.65
Biological Replicate 3						
TR1	18.58	18.83	18.71	18.02	18.43	28.39
TR2	18.63	18.79	18.72	18.06	18.42	28.69

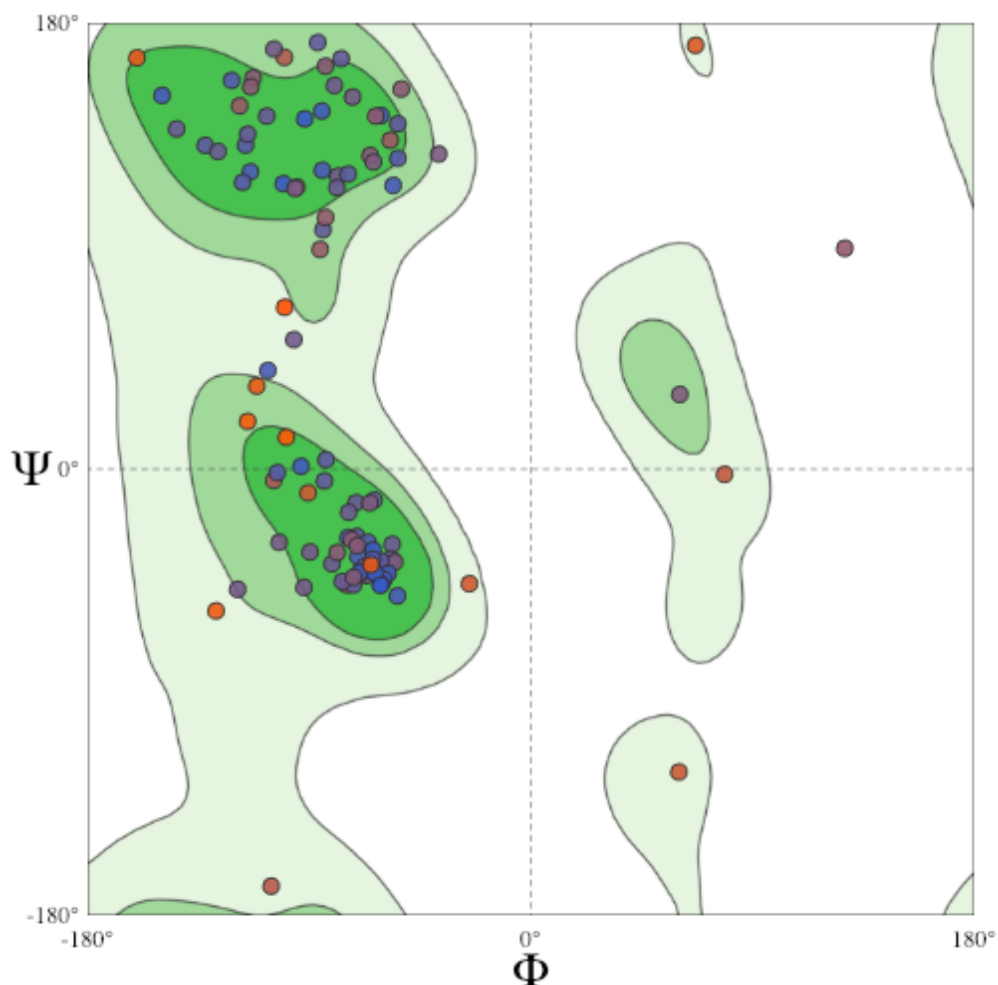
Table A.3: Raw band intensity value data for GAPDH protein (BR stands for biological replicate)

	NP69	HK1	HONE-1	SUNE-1	TWO 1	TWO 4	C666-1
BR1	46923.82	49485.64	44172.95	46736.14	45247.41	49889.36	41904.91
BR2	57655.55	57542.85	57405.7	57345.85	59242.55	59289.85	59337.55
BR3	58170.67	58270.39	58213.72	58217.06	58285.94	58331.83	58298.39

Table A.4: Raw band intensity value data for uS19 protein (BR stands for biological replicate)

	NP69	HK1	HONE-1	SUNE-1	TWO 1	TWO 4	C666-1
BR1	1088.86	3018.93	5113.79	5000.07	1162.36	2228.5	9030.79
BR2	1019.17	2558.83	4887.78	4231.78	1158.06	2201.58	7764.39
BR3	1028.17	2588.61	4955.39	4222	1180.72	2147.06	7599.89

Figure A.5: Ramachandran plot and PROCHECK analysis of uS19 protein

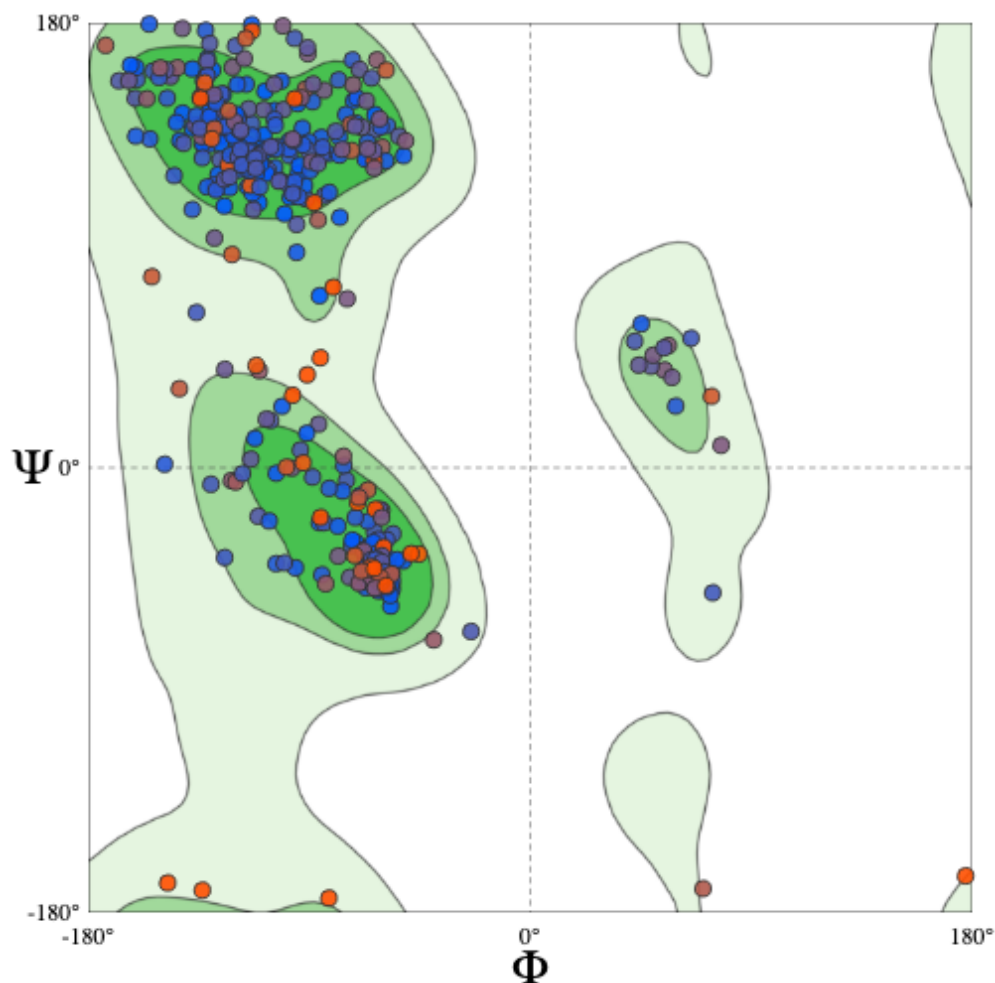


```

+-----<<< P R O C H E C K       S U M M A R Y >>>-----+
| /var/www/PROCHECK/Jobs/8393674/8393674.pdb   1.5           125 residues |
| * Ramachandran plot:   84.1% core   13.1% allow   1.9% gener   0.9% disall |
| * All Ramachandrans:   8 labelled residues (out of 123) |
| * Chi1-chi2 plots:    4 labelled residues (out of 87) |
| + Side-chain params:   4 better    0 inside    1 worse |
| * Residue properties: Max.deviation:    4.0           Bad contacts:    0 |
| *                   Bond len/angle:    8.2       Morris et al class:  1  3  2 |
| G-factors              Dihedrals:  -0.34   Covalent:  -0.25   Overall:  -0.29 |
| * Planar groups:       79.6% within limits  20.4% highlighted    2 off graph |
+-----+-----+
| + May be worth investigating further.  * Worth investigating further. |

```

Figure A.6: Ramachandran plot and PROCHECK analysis of HSPA8 protein



```

+-----<<< P R O C H E C K       S U M M A R Y >>>-----+
| /var/www/PROCHECK/Jobs/2246038/2246038.pdb   1.5           554 residues |
| * Ramachandran plot:   92.7% core    6.7% allow    0.4% gener    0.2% disall |
| + All Ramachandrans:   12 labelled residues (out of 552) |
| + Chi1-chi2 plots:     1 labelled residues (out of 329) |
|   Side-chain params:   5 better     0 inside     0 worse |
| * Residue properties: Max.deviation:   12.6           Bad contacts:   11 |
| *                   Bond len/angle:   13.9       Morris et al class:  1  1  2 |
| + G-factors           Dihedrals:    0.00  Covalent:  -1.04   Overall:  -0.43 |
|   Planar groups:      100.0% within limits  0.0% highlighted |
+-----+
| + May be worth investigating further.  * Worth investigating further. |

```

Figure A.7: Ramachandran plot and PROCHECK analysis of CUBN protein

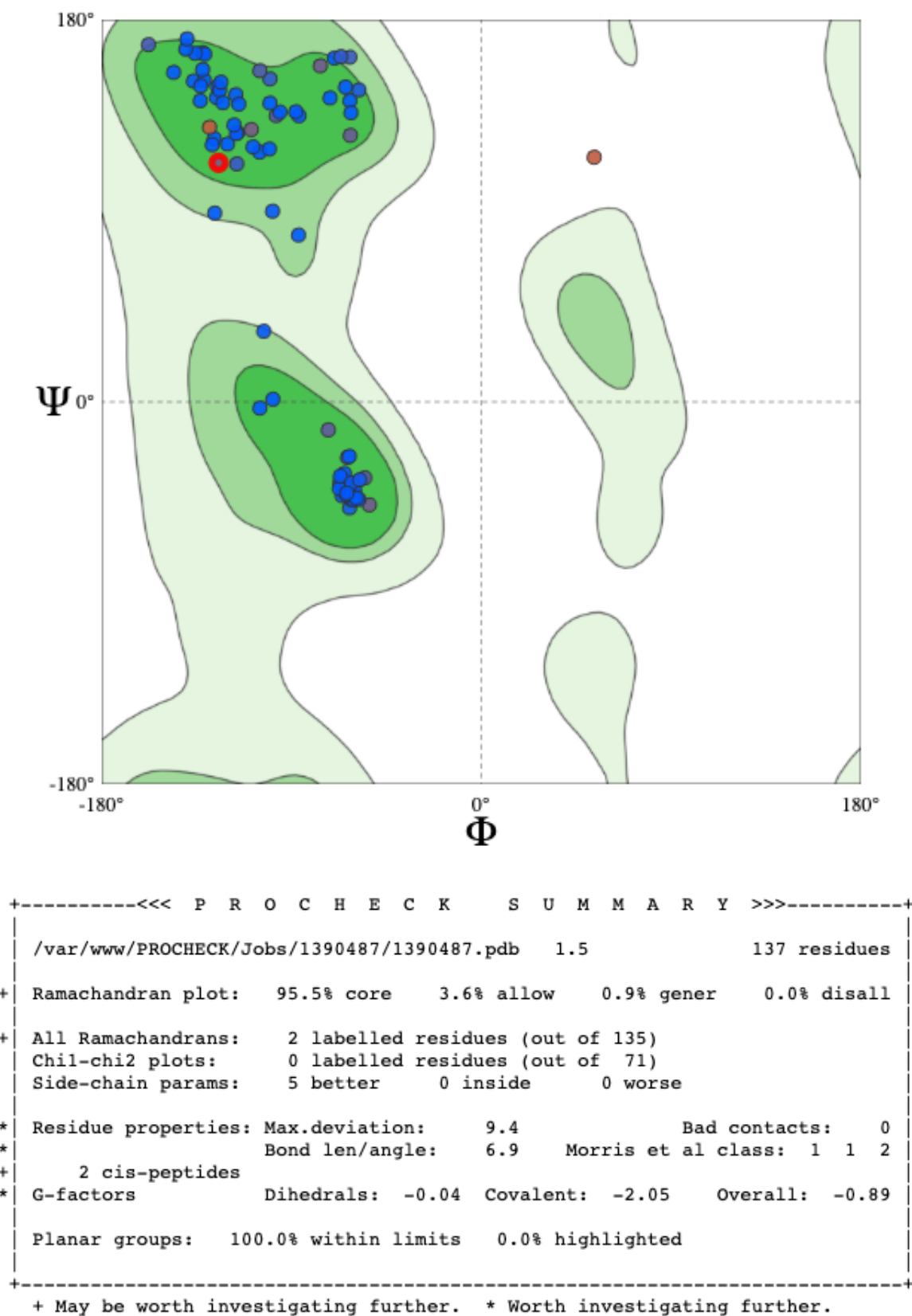
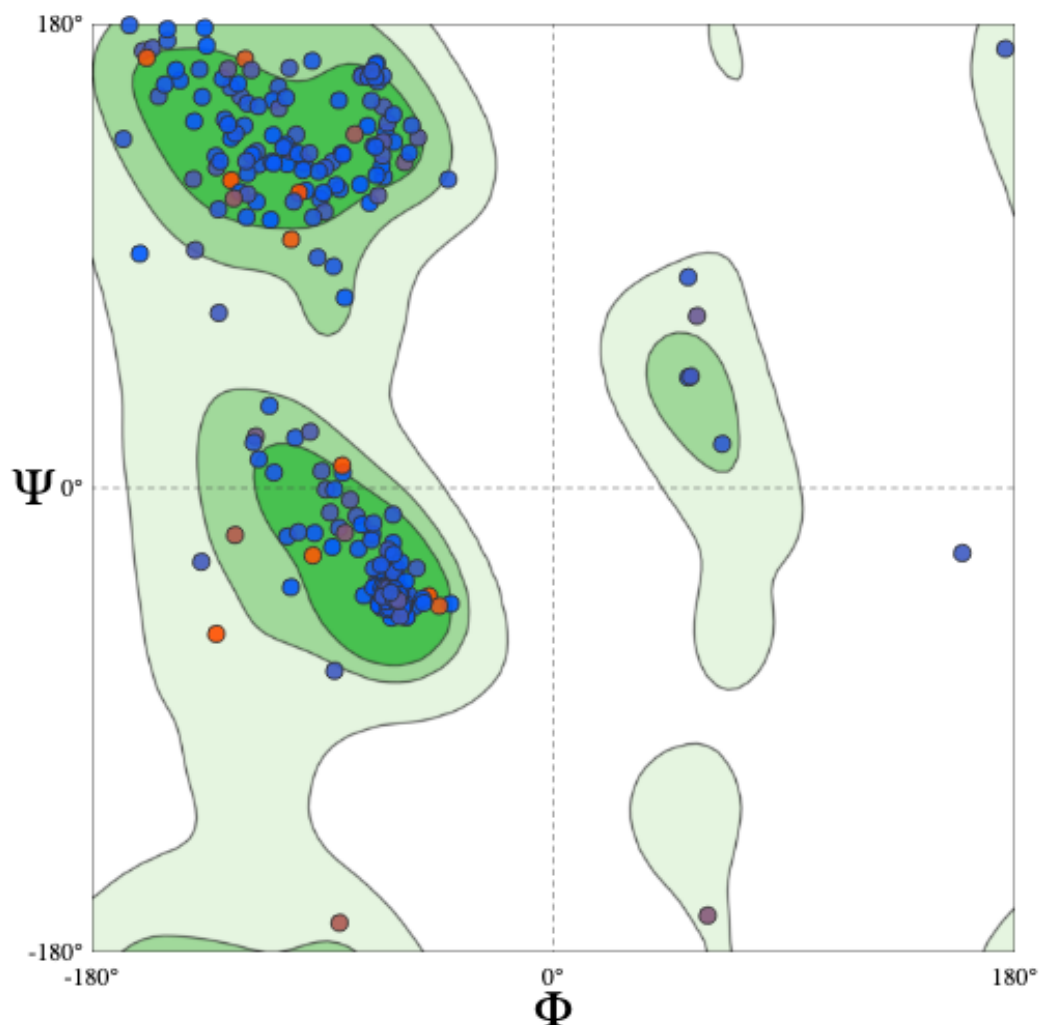


Figure A.8: Ramachandran plot and PROCHECK analysis of Cobalamin protein



```

+-----<<< P R O C H E C K       S U M M A R Y >>>-----+
| /var/www/PROCHECK/Jobs/6357988/6357988.pdb   1.5           392 residues |
+ Ramachandran plot:  93.6% core    6.1% allow    0.3% gener    0.0% disall |
+ All Ramachandrans:  10 labelled residues (out of 390) |
+ Chi1-chi2 plots:    3 labelled residues (out of 222) |
+ Side-chain params:  5 better      0 inside      0 worse |
| P5_Maths_CA1_2018_Nan_Hua_Exam_Pap |
+ Residue properties: Max.deviation:  10.6           Bad contacts:  7 |
+                      Bond len/angle:  6.8       Morris et al class: 1 1 2 |
+ G-factors           Dihedrals:  0.05  Covalent:  -1.31  Overall:  -0.52 |
| Planar groups:      100.0% within limits  0.0% highlighted |
+-----<<< P R O C H E C K       S U M M A R Y >>>-----+
+ May be worth investigating further.  * Worth investigating further.

```

Figure A.9: Ramachandran plot and PROCHECK analysis of LMP2 protein

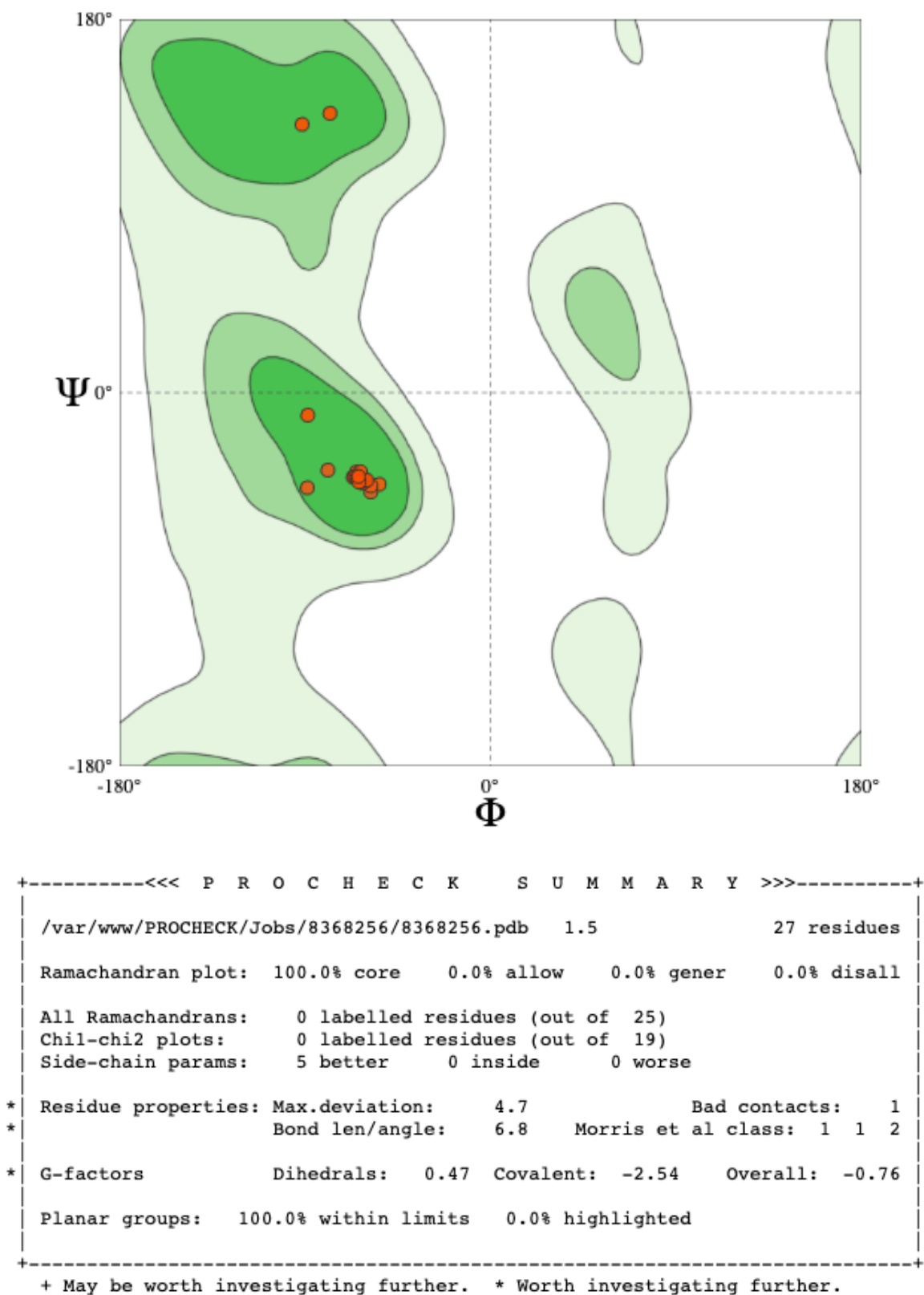
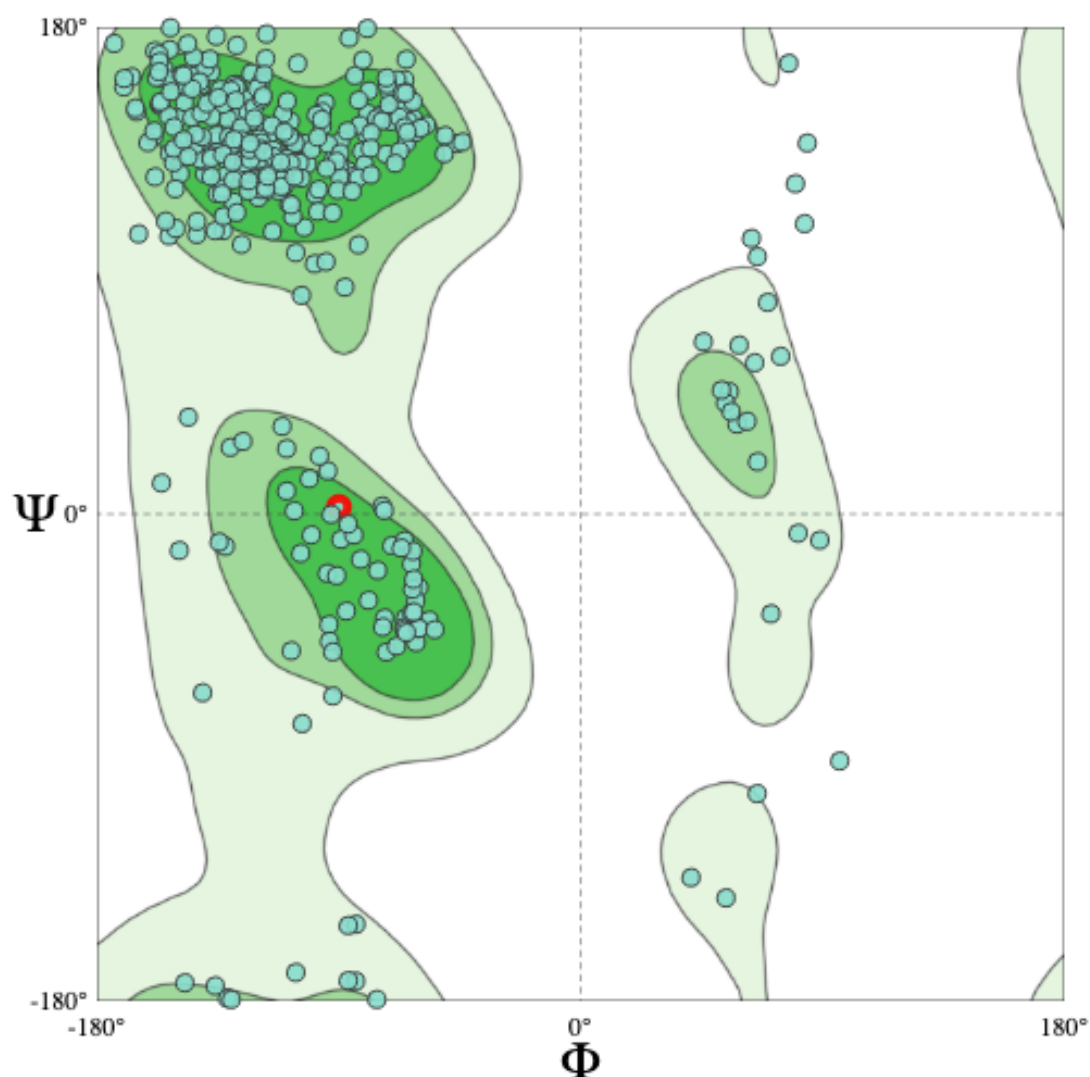


Figure A.10: Ramachandran plot and PROCHECK analysis of EBNA1 protein



```

+-----<<< P R O C H E C K       S U M M A R Y >>>-----+
| /var/www/PROCHECK/Jobs/3523801/3523801.pdb   1.5           457 residues |
| * Ramachandran plot:   85.7% core   12.0% allow   0.8% gener   1.5% disall |
| * All Ramachandrans:   20 labelled residues (out of 455) |
| + Chi1-chi2 plots:     2 labelled residues (out of 277) |
|   Side-chain params:   5 better     0 inside     0 worse |
| * Residue properties: Max.deviation:   11.6           Bad contacts:   22 |
| *                       Bond len/angle:  12.4       Morris et al class: 1 1 2 |
| +       2 cis-peptides |
| + G-factors           Dihedrals:  -0.41  Covalent:  -1.58   Overall:  -0.90 |
| Planar groups:        100.0% within limits   0.0% highlighted |
+-----+
| + May be worth investigating further.  * Worth investigating further. |

```

Table A.5: ClusPro scores of uS19 docked complex models

Dock model	ClusPro scores			RMSD (Å)
	Cluster size	Centre free energy (kcal/mol)	Lowest free energy (kcal/mol)	
uS19- TMEM237	74	-601.8	-697.3	0.90
uS19- SLC22A23	55	-1144.9	-1276.0	0.86
uS19- SLC13A4	66	-1167.9	-1387.8	0.91
uS19- HSPA8	63	-763.3	-794.3	0.98
uS19- CUBN	232	-266.1	-317.1	0.94
uS19- Cobalamin	109	-902.5	-902.5	0.90
uS19-SLC7A1	68	-957.7	-1010.0	0.93
uS19-RRP1B	132	-710.6	-779.1	0.93
uS19-LMP2	184	-1111.4	-1111.4	0.98
uS19-EBNA1	59	-797.0	-938.1	0.89

Table A.6: PatchDock scores

Dock model	Score	Area	ACE	RMSD (Å)
uS19- TMEM237	16170	2257.50	485.86	0.00
uS19- SLC22A23	20788	3567.80	247.34	0.00
uS19-SLC13A4	21484	3715.70	-68.01	0.00
uS19- HSPA8	19624	2557.20	151.05	0.00
uS19- CUBN	20480	2772.20	453.21	0.00
uS19- Cobalamin	12820	1869.40	-59.79	0.00
uS19-SLC7A1	17002	2303.50	358.75	0.00
uS19-RRP1B	18046	2677.10	494.61	0.00
uS19-LMP2	18408	2566.40	164.85	0.00
uS19-EBNA1	18388	2347.60	434.59	0.00

Table A.7: FireDock scores

Dock model	Global energy	Attractive VdW	Repulsive VdW	ACE	Hydrogen Bond
uS19-TMEM237	11.57	-1.27	0.34	0.65	-0.84
uS19-SLC22A23	-0.34	-1.11	0.00	-0.93	0.00
uS19-SLC13A4	5.45	-10.04	19.25	-2.65	-0.76
uS19- HSPA8	2.74	-7.30	1.60	5.89	-0.61
uS19- CUBN	-12.03	-37.59	14.41	14.71	-7.24
uS19-Cobalamin	-20.41	-15.85	4.88	1.20	0.00
uS19-SLC7A1	10.36	-6.05	0.42	5.63	-0.57
uS19-RRP1B	-4.73	-6.33	3.62	-1.06	-0.45
uS19-LMP2	-1.50	-23.09	22.77	13.29	-5.60
uS19-EBNA1	13.15	-14.69	5.34	0.39	-1.12

SELECTIVE DIFFERENTIAL EXPRESSION OF THE RIBOSOMAL PROTEIN GENES *eL14* AND *uS19* IN A WELL-DIFFERENTIATED EPITHELIAL CELL LINE OF NASOPHARYNGEAL CARCINOMA

EDMUND UI-HANG SIM^{1*}, CASSANDRA SHEAU-MEI CHEE¹, LISHA VASUDEVAN¹,
KHER-LEE NG¹ and STELLA LI-LI CHAN²

¹Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia ²National Cancer Centre, 11 Hospital Drive, 169610 Singapore *E-mail: uhsim@unimas.my

Accepted 14 February 2018, Published online 31 March 2018

ABSTRACT

Besides ribosome biogenesis and protein synthesis, ribosomal proteins (RP) are associated with congenital diseases and cancers. A small subset of ribosomal protein genes has shown expression pattern indicative of their association with nasopharyngeal carcinoma (NPC). Nevertheless, the list of RP genes that are NPC-associated factors is largely incomplete. Herein we report the expression patterns of *eL14* and *uS19* in NPC normal nasopharyngeal epithelium cell lines. Expression levels of *eL14* and *uS19* in the NPC-HK1 cell line was comparatively analysed with a normal nasopharyngeal cell line (NP69) using Reverse Transcription – Polymerase Chain Reaction (RT-PCR). We revealed that the transcript level of *eL14* was significantly down-regulated in HK1 when compared to NP69. The expression behaviour of *eL14* is demonstrated for the first time in the NPC context. In contrast, the transcript level of *uS19* was up-regulated in NPC/HK1 compared to NP69, but not to a statistically significant extent. This study provides new evidence of differential expression of the ribosomal protein gene, *eL14* in an NPC cell line derived from well-differentiated squamous cell carcinoma of human nasopharynx. It adds to the list of NPC-associated ribosomal protein genes amenable for development of biomarkers for improved molecular diagnosis of nasopharyngeal cancer.

Key words: Ribosomal protein genes, *eL14*, *uS19*, nasopharyngeal carcinoma, gene expression analysis

INTRODUCTION

The basic function of RP genes constitutes ribosome-mediated protein biosynthesis. Eukaryotic RP genes are divided into two broad types. In human, these comprise genes for the small (40S) ribosomal subunits (*e/uS1 – S15*, *e/uS17*, *e/uS19*, *eS21*, *eS24 – S28*, *eS30*, *eS31*, and *RACK1*), and those of the large (60S) ribosomal subunit (*e/uL1 – L8*, *uL10*, *uL11*, *e/uL13 – L16*, *e/uL18 – L24*, *e/uL27 – L34*, *eL36 – L43*, and *PI/P2*) (Ban *et al.*, 2014). Despite a unified mechanism for control of transcription and translation, each is distinguishable by its expression level and tissue-specificity (Ishii *et al.*, 2006). Their functions often extend beyond ribosome biogenesis and protein synthesis to include extra-ribosomal roles such as interactions with oncogenic factors and/or tumour suppressors in

the tumorigenesis of a variety of cancer types (deLas-Heras-Rubio *et al.*, 2014). The association of some RP genes with nasopharyngeal carcinoma (NPC) cells has been reported (Fang *et al.*, 2008; Sim *et al.*, 2008, 2010, 2016, 2017).

The *uS19* gene is up-regulated in non to poorly differentiated squamous cell carcinoma of the nasopharynx (Fang *et al.*, 2008). Its activity in well-differentiated type of NPC cells is unknown. Located at 19p13.3, it functions in the maturation and assembly of the 40S ribosome subunit (Robledo *et al.*, 2008), and is presumably involved in the initiation and elongation steps of translation (Kitagawa *et al.*, 1991; Rouquette *et al.*, 2005). Its nuclear respiratory factor (NRF) acts as a transcriptional activator, and its activator protein (AP-1) is associated with cellular proliferation, apoptosis and differentiation (Ishii *et al.*, 2006). Expression of *uS19* is high in tumour cells of insulinomas, and oesophageal and colon cancers

* To whom correspondence should be addressed.

(Kitagawa *et al.*, 1991). It can be detected in skin, foetal brain, and spleen (Ishii *et al.*, 2006). *uS19* regulates p53 via direct interaction with the Mdm2-p53-MdmX network (Daftuar *et al.*, 2013).

The expression of *eL14* has never been studied in NPC before this. The *eL14* protein is a component of the 60S ribosomal subunit and belongs to L14E family of ribosomal proteins. The gene contains a trinucleotide repeat that highly polymorphic constituting loss of heterozygosity at 3p21.3 in oral cancer (Shriver *et al.*, 1998). Non-random allelic loss at 3p21 has also been observed in cancers of the lung, breast, kidney, cervical and nasopharynx (Liu *et al.*, 2003; Chow *et al.*, 2004). In fact, *eL14* is under-expressed, but not mutated, in tumours of lung and oesophageal cancers (Huang *et al.*, 2006). However, its activity has never been investigated in NPC cells.

NPC is a malignancy of the nasopharyngeal mucosal epithelium specifically in the fossa of Rosenmüller. WHO Classification categorises NPC as three histopathological types, namely Type I – well differentiated keratinising squamous cell carcinoma; Type II – non-keratinising differentiated squamous cell carcinoma; and Type III – undifferentiated carcinoma (Pathmanathan *et al.*, 1995; Chan *et al.*, 2002). Types II and III NPC are related to Epstein-Barr virus (EBV) infection of affected cells (Pathmanathan *et al.*, 1995) with Type III having the highest level of association (Cote *et al.*, 2002). Type III NPC is also the most common carcinoma in NPC patients (Cote *et al.*, 2002), with high incidence and mortality rate in Southern China (Huang *et al.*, 1980; Wei *et al.*, 2014). In contrast, Type I NPC is rare and has no reported association with EBV infection (Huang *et al.*, 1980).

The involvement of RPs in NPC is still undefined despite a few studies on NPC-associated RP genes (Sim *et al.*, 2010, 2016, 2017). This is largely due to the incomplete list of NPC-associated RP genes. This paper reveals the selective differential expression pattern of *uS19* and *eL14* in a well-differentiated non-keratinising NPC cell line, HK1, and adds to list of RP genes associated with the cancer.

MATERIALS AND METHODS

Cell culture

The HK1 cell line used in this study is a non-keratinised, well-differentiated epithelial cell line derived from Type II-NPC (Huang *et al.*, 1980). Normal control cell line (NP69) is from non-malignant nasopharyngeal epithelium (Tsao *et al.*, 2002). Both cell lines were provided by Dr Ching-Ching Ng (University of Malaya). HK1 was grown in RPMI media with 10% (v/v) foetal bovine

serum (FBS), 2 mM L-glutamine, and 100 U/ml Penicillin-Streptomycin mixture; and NP69 was in K-SFM media with 0.2 ng/ml growth factors, 5% heat-inactivated FBS, and 100 U/ml Penicillin-Streptomycin mixture. All cells were cultured at 37°C in a humidified condition with 5% purified CO₂ gas. Cells were grown to at least 70% confluence prior to RNA extraction.

RNA extraction and quantification

Total RNA was extracted using the TRIzol method (Chomczynski & Sacchi, 1987). Basically, cells rinsed with ice cold Phosphate Buffer Saline solution was added with TRIzol reagent (InvitrogenTM, USA), dislodged using a cell scraper, and homogenised. The homogenate was incubated for 5 min, and aliquots were transferred to 1.5 ml tube where chloroform (0.2 ml per 1 ml TRIzol) was added. This was vortexed, incubated (5 min), and centrifuged (10000 × g for 15 min at 4°C). The upper aqueous layer was transferred into a new 1.5 ml tube, and the RNA precipitated using iso-propanol. The RNA was pelleted via centrifugation, and pellet was washed with 75% ethanol, air-dried, and dissolved in 30 µl of nuclease-free water. The extracted RNA was first assessed agarose gel electrophoresis (AGE) assay. Concentration of RNA was determined by taking absorbance value at 260 nm using a UV spectrophotometer (Shimadzu, USA or Ultrospec 1100 pro, UK), and purity was assessed by the ratio of values taken at 260 nm to 280 nm.

Reverse transcriptase – Polymerase Chain Reaction (RT-PCR)

Reverse transcription of RNA was done according to protocol by Promega. Basically, 2 µg of RNA was reverse transcribed using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) (Promega, UK) in a reaction mixture (25 µl) containing 0.5 µg random primers, 1× reaction buffer, 40 mM dNTPs, 25 U Recombinant RNasin® Ribonuclease Inhibitor, and 200 U M-MLV RT. The reaction was incubated for 60 min at 37°C, and then at 70°C for 15 min. Synthesised complementary DNAs (cDNAs) were checked using AGE assay, and kept at -20°C until use. PCR was done based on protocol from Fermentas (Fermentas, USA). Each reaction mixture (25 µl) consists of 1 µM forward and reverse primers each, 1.25 U *Taq* DNA polymerase, 1 mM MgCl₂, 0.2 mM dNTPs, 1× reaction buffer, and 0.5 µg template cDNA. Thermal cycling parameters are initial denaturation at 95°C for 2 min, and then 30 cycles of denaturation (95°C for 30 sec), annealing (51 to 54°C for 30 sec) and extension (72°C for 1 min), with a final extension of 5 minutes at 72°C. The annealing temperatures for analysis of *uS19*, *EL14*, and *GAPDH* are 51.4, 53.5, and 51.7°C respectively. Forward and reverse

primers for *uS19* (5'-GAAGACGCACCTGCGGACA-3' and 5'-GTGGCCGATCATCTCGGGC TTG-3'), *EL14* (5'-TTCTTCCTTCTCGCCT AACG-3' and 5'-CCTCCTAACTTCAGCCTCAA-3') and glyceraldehyde-3-phosphate dehydrogenase, *GAPDH* (5'-AAGATCATCAGCAATGG CTC-3' and 5'-TACCAGGACAT GAGCTTGAC-3') genes were designed using Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). *GAPDH* gene was used as internal control. Expected amplicons of *uS19*, *eL14* and *GAPDH* are 115, 773, and 511 bps respectively. PCR results were assessed by AGE assay and verified using sequence analysis. Sequence alignment was via blast2q programme (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). GenBank data of *uS19* transcript (accession no. NM_001018.3) and *eL14* transcript 1 (accession no. NM_001034996.2) were used as reference.

Data analysis

Band intensities on AGE gel were documented and analysed using TotalLab Quant software (TotalLab, USA). Normalisation of expression data was carried out using *GAPDH* data as reference. Differential expression pattern was expressed as fold difference (FD) between data of either *uS19* or *EL14* in NPC cell line (HK1) relative to NP69. FD >1 and < 1 is considered over- and under-expression respectively. Statistical analysis was done using a 2-tailed unpaired Student's *t*-Test with significance taken at $p < 0.05$.

RESULTS

The amplicons are within the vicinity of expected size range (Figure 1) and sequence analysis shows at least 99% similarity with GenBank data (Figure 2), hence validating our tests. Band intensity values for expression of *eL14*, *uS19*, and *GAPDH* in the two

cell lines are presented in Table 1, and Figure 3. For *eL14*, expression was undetectable in the HK1 cell line (Figure 1), and supported by the band intensity data from image analysis (Table 1 and Figure 3). This indicated the significant under-expression of *eL14* in HK1 compared to NP69. For *uS19*, up-regulated expression was observed for HK1 cell line in comparison to NP69 (Table 1 and Figure 3). Generally, the expression of *uS19* is higher than *eL14* in both cell lines (Figure 3) and this is indicative of the distinct (or independent) expression behaviour between the two RP genes.

The under-expression of *eL14* in HK1 cells relative to NP69 is obvious based on its undetected expression in the NPC cell line. Hence, the fold difference calculation for this case is not needed. The statistical analysis via Student's *t*-test is also not necessary, and its under-expression is considered as significant. For *uS19*, fold difference of over-expression in HK1 cells when compared to NP69 is 1.15 (312.02/271.78). Student's *t*-test analysis showed that there is no significant difference ($p = 0.736$) in the expression level between HK1 and NP69 cell lines.

DISCUSSION

Our assessment of *eL14* non-expression in HK1 implies its inactivity in this NPC cell line. There has been observation of decreased expression for genes such as *calgranulin A*, *calgranulin B*, *ENA-78*, *interleukin-8*, *integrin beta-1*, *integrin beta-4*, *p21* and *RET* in NPC cell lines (including HK1) relative to normal controls (Fung *et al.*, 2000). Reduced expression of *eL14* accompanied by allelic loss in the trinucleotide repeat region has been found in oesophageal squamous cell carcinomas (Huang *et al.*, 2006). The under-expression of *eL14* may be common in cancers of squamous cell epithelium

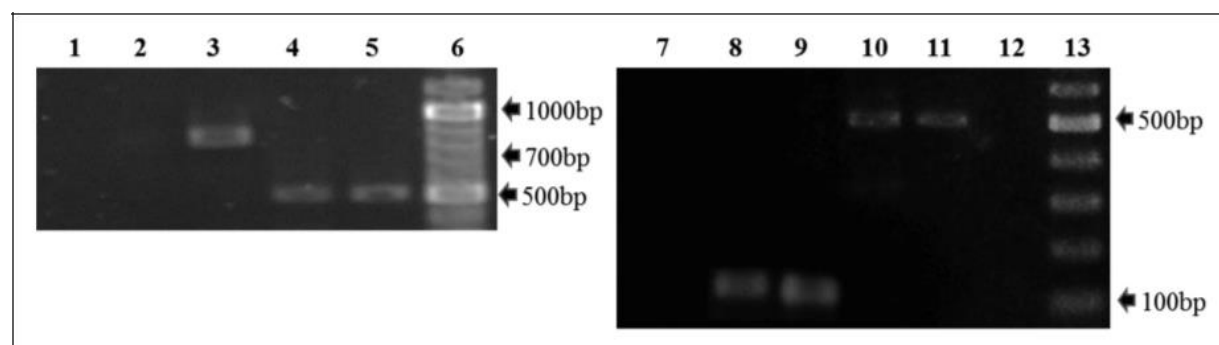
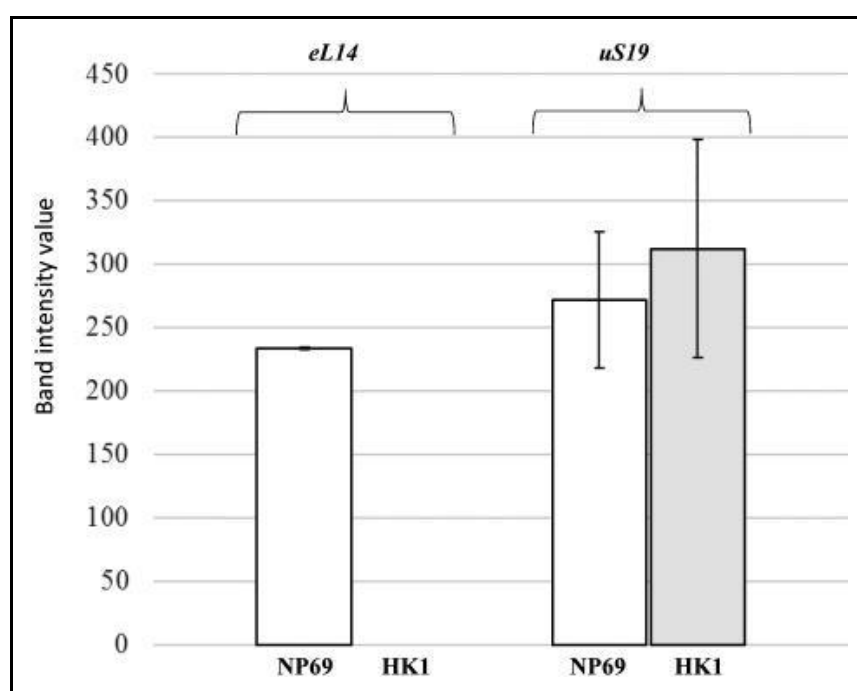


Fig. 1. Representative AGE-assayed results of RT-PCR analysis. RT-PCR results of *eL14* in HK1 and NP69 cell lines are in lanes 2 and 3 respectively. The internal control (*GAPDH* amplicons) for the analysis of *eL14* expression in HK1 and NP69 are in lanes 4 and 5. Amplicons of *uS19* in HK1 and NP69 are in lanes 8 and 9, and the *GAPDH* amplicons of associated analysis are in Lanes 10 and 11. Negative controls (lanes 1, 7 and 12) are included. Lane 6 and 13 are DNA size marker (GeneRuler™ 100 bp Plus DNA Ladder, Fermentas USA).

Table 1. Expression data of RP genes obtained from image analysis (via TotalLab Quant software) of RT-PCR products on AGE-assayed gel

Gene	Cell line	Replicate	Band intensity (pixel value)	<i>GAPDH</i> band intensity	Normalised band intensity	Mean
<i>eL14</i>	NP69	1	234.80	232.92	232.67	233.63
		2	232.45	228.70	234.59	
	HK1	1	0.00	232.75	0.00	0.00
		2	0.00	230.63	0.00	
<i>uS19</i>	NP69	1	303.61	232.18	325.26	271.78
		2	232.83	265.29	218.30	
	HK1	1	369.49	214.49	398.07	312.02
		2	242.21	247.67	225.98	

**Fig. 3.** Bar chart of gene expression data comparison between cell lines. The y-axis represents normalised band intensity value and the x-axis is the types of cell line. Error bars are included.

origin, and in our case, even in cells of well-differentiated squamous cell carcinoma. It is unclear if there is any allelic loss in the NPC situation. Nonetheless, we provide the first evidence of *eL14* under-expression in a cell line derived from Type II-NPC.

uS19's pattern of over-expression in HK1 cell line relative to NP69 is not statistical significant. *uS19* has been found to be up-regulated in primary NPC tumours that comprises poorly differentiated squamous cell carcinoma (Fang *et al.*, 2008). Hence, there seems to be some consistency to its activity in NPC, despite the fact that our study focus on cells

of differentiated origin. Besides NPC, it is over-expressed in hepatocellular carcinoma cell lines (Yoon *et al.*, 2006). Further analysis of *uS19* in HK1 cells is still needed, perhaps with respect to its expression at protein level.

CONCLUSION

Significant under-expression of the RP gene, *eL14* is evident in the non-keratinising well-differentiated cell line (HK1) derived from squamous cell carcinoma of Type II-NPC.

ACKNOWLEDGEMENT

This study is funded by the Ministry of Higher Education, Malaysia under the Fundamental Research Grant Scheme [grant code: FRGS/ST03(02)/1299/2015(16)].

REFERENCES

- Ban, N., Beckmann, R., Cate, J.H., Dinman, J.D., Dragon, F., Ellis, S.R., Lafontaine, D.L., Lindahl, L., Liljas, A., Lipton, J.M., McAlear, M.A., Moore, P.B., Noller, H.F., Ortega, J., Panse, V.G., Ramakrishnan, V., Spahn, C.M., Steitz, T.A., Tchorzewski, M., Tollervey, D., Warren, A.J., Williamson, J.R., Wilson, D., Yonath, A. & Yusupov, M. 2014. A new system for naming ribosomal proteins. *Current Opinions in Structural Biology*, **24**:165-169.
- Chan, A.T.C., Teo, P.M.L. & Johnson, P.J. 2002. Nasopharyngeal carcinoma. *Annals of Oncology*, **13**: 1007-1015.
- Chomczynski, P. & Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, **162**(1): 156-159.
- Chow, L.S., Lo, K.W., Kwong, J., To, K.F., Tsang, K.S., Lam, C.W., Dammann, R. & Huang, D.P. 2004. RASSF1A is a target tumor suppressor from 3p21.3 in nasopharyngeal carcinoma. *International Journal of Cancer*, **109**(6): 839-847.
- Cote, R., Suster, S., Weiss, L. & Weidner, N. (Eds.). 2002. Modern Surgical Pathology (2 Volume Set). W B Saunders, London.
- Daftuar, L., Zhu, Y., Jacq, X. & Prives, C. 2013. Ribosomal proteins RPL37, RPS15 and RPS20 regulate the Mdm2-p53-MdmX network. *PLoS One*, **8**(7): e68667.
- deLas Heras-Rubio, A., Perucho, L., Paciucci, R., Vilardell, J. & Leonart, M. 2014. Ribosomal proteins as novel players in tumorigenesis. *Cancer and Metastasis Reviews*, **33**(1): 115-141.
- Fang, W., Li, X., Jiang, Q., Liu, Z., Yang, H., Wang, S., Xie, S., Liu, Q., Liu, T., Huang, J., Xie, W., Li, Z., Zhao, Y., Wang, E., Marincola, F.M. & Yao, K. 2008. Transcriptional patterns, biomarkers and pathways characterizing nasopharyngeal carcinoma of Southern China. *Journal of Translational Medicine*, **6**(1): 32.
- Fung, L.F., Lo, A.K., Yuen, P.W., Liu, Y., Wang, X.H. & Tsao, S.W. 2000. Differential gene expression in nasopharyngeal carcinoma cells. *Life Sciences*, **67**(8): 923-936.
- Huang, D.P., Ho, J.H., Poon, Y.F., Chew, E.C., Saw, D., Lui, M., Li, C.L., Mak, L.S., Lai, S.H., Lau, W.H. 1980. Establishment of a cell line (NPC/HK1) from a differentiated squamous carcinoma of the nasopharynx. *International Journal of Cancer*, **26**(2): 127-132.
- Huang, X.P., Zhao, C.X., Li, Q.J., Cai, Y., Liu, F.X., Hu, H., Xu, X., Han, Y.L., Wu, M., Zhan, Q.M. & Wang, M.R. 2006. Alteration of RPL14 in squamous cell carcinomas and preneoplastic lesions of the esophagus. *Gene*, **366**(1): 161-168.
- Ishii, K., Washio, T., Uechi, T., Yoshihama, M., Kenmochi, N. & Tomita, M. 2006. Characteristics and clustering of human ribosomal protein genes. *BMC Genomics*, **7**(37).
- Kitagawa, M., Takasawa, S., Kikuchi, N., Itoh, T., Teraoka, H., Yamamoto, H. & Okamoto, H. 1991. Rlg encodes ribosomal protein S15. The primary structure of mammalian ribosomal protein S15. *FEBS Letters*, **283**(2): 210-214.
- Liu, X.Q., Chen, H.K., Zhang, X.S., Pan, Z.G., Li, A., Feng, Q.S., Long, Q.X., Wang, X.Z. & Zeng, Y.X. 2003. Alterations of BLU, a candidate tumour suppressor gene on chromosome 3p21.3, in human nasopharyngeal carcinoma. *International Journal of Cancer*, **106**(1): 60-65.
- Pathmanathan, R., Prasad, U., Chandrika, G., Sadler, R., Flynn, K. & Raab-Traub, N. 1995. Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx. Variants of Epstein-Barr virus-infected neo-plasia. *The American Journal of Pathology*, **146**(6): 1355-1367.
- Robledo, S., Idol, R.A., Crimmins, D.L., Ladenson, J.H., Mason, P.J. & Bessler, M. 2008. The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA*, **14**(9): 1918-1929.
- Rouquette, J., Choismel, V. & Gleizes, P.E. 2005. Nuclear export and cytoplasmic processing of precursors to the 40S ribosomal subunits in mammalian cells. *The EMBO Journal*, **24**(16): 2862-2872.
- Shriver, S.P., Shriver, M.D., Tirpak, D.L., Bloch, L.M., Hunt, J.D., Ferrell, R.E. & Siegfried, J.M. 1998. Trinucleotide repeat length variation in the human ribosomal protein L14 gene (RPL14): localization to 3p21.3 and loss of heterozygosity in lung and oral cancers. *Mutation Research/Mutation Research Genomics*, **406**(1): 9-23.
- Sim, E.U.H., Toh, A.K.L. & Tiong, T.S. 2008. Preliminary findings of down-regulated genes in nasopharyngeal carcinoma. *Asia-Pacific Journal of Molecular Biology and Biotechnology*, **16**: 79-84.

- Sim, E.U.H., Ang, C.H., Ng, C.C., Lee, C.W. & Narayanan, K. 2010. Differential expression of a subset of ribosomal protein genes in cell lines derived from human nasopharyngeal epithelium. *Journal of Human Genetics*, **55**: 118-120.
- Sim, E.U.H., Chan, S.L.L., Ng, K.L., Lee, C.W. & Narayanan, K. 2016. Human ribosomal proteins RPeL27, RPeL43, and RPeL41 are upregulated in nasopharyngeal carcinoma cell lines. *Disease Markers*, **2016**: Article ID 5179594.
- Sim, E.U.H., Ng, K.L., Lee, C.W. & Narayanan, K. (2017). The uS8, uS4, eS31, and uL14 ribosomal protein genes are dysregulated in nasopharyngeal carcinoma cell lines. *BioMed Research International*, **2017**: Article ID 4876954.
- Tsao, S.W., Wang, X., Liu, Y., Cheung, Y.C., Feng, H., Zheng, Z., Wong, N., Yuen, P.W., Lo, A.K., Wong, Y.C. & Huang, D.P. 2002. Establishment of two immortalized nasopharyngeal epithelial cell lines using SV40 large T and HPV16E6/E7 viral oncogenes. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research*, **1590**(1): 150-158.
- Wei, K.R., Zheng, R.S., Zhang, S.W, Liang, Z.H., Ou, Z.X. & Chen, W.Q. 2014. Nasopharyngeal carcinoma incidence and mortality in China in 2010. *Chinese Journal of Cancer*, **33**(8): 381-387.
- Yoon, S.Y., Kim, J.M., Oh, J.H., Jeon, Y.J., Lee, D.S., Kim, J.H., Choi, J.Y., Ahn, B.M., Kim, S., Yoo, H.S., Kim, Y.S. & Kim, N.S. 2006. Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags. *International Journal of Oncology*, **29**(2): 315-327.