IN SILICO SINGLE POLYMORPHISM ANALYSIS OF *RPMS1* AND *A73* GENE IN NASOPHARYNGEAL CARCINOMA

QUANG TRONG MINH[1], NGUYEN HOANG DANH[1], THIEU HONG HUE[2], LAO DUC THUAN[1,2]*

[1]Ho Chi Minh City Open University, Vietnam
[2]Vietnam National University Ho Chi Minh City, Vietnam
*Corresponding author, email: thuan.ld@ou.edu.vn

(Received: April 21, 2019; Revised: May 10, 2019; Accepted: May 21, 2019)

ABSTRACT

Nasopharyngeal carcinoma (NPC) is one of the most frequent cancer types in Vietnam, with high mortality rate. Therefore, the early diagnosis and detection of NPC is urgently needed to improve patient survival. Recent studies have confirmed that the infection of Epstein-Barr virus (EBV) and polymorphism in *RPMS1* and *A73* are considered as the etiological factor associated with NPC. However, in Vietnam, there are no studies relevant to the identification of polymorphism of *RMPS1* and *A73*. With the aims, in future, to develop a technique based on detecting the frequencies of *RPMS1* and *A73* variants as biomarkers of prognosis and early diagnosis of NPC, we conducted the initial *in silico* analysis (1) Data collection, statistical analysis the frequencies detection of *RPMS1* and *A73* variants from various previous published studies; (2) Determine experimental methods to predict and diagnose early NPC and examine necessary steps *in silico.* As the results, we established the systematic databases of *RPMS1* and *A73* polymorphism, and evaluation the primer set for the amplification of *RPMS1* and *A73*, which could be applied in further studies related to the identification of *RPMS1* and *A73* gene polymorphism to find out the potential biomarkers for screening, diagnosis as well as NPC treatment.

Keywords: *A73*; NPC**;** Polymorphism; *RPMS1*.

1. Introduction

Nasopharyngeal carcinoma (NPC), which has a striking geographic and ethic distribution, has been considered as the most common and highly incident cancer of head and neck cancer in Asian countries, especially in Vietnam (Regaud et al., 1921, Schmincke et al., 1921). According to Globocan (2018), the high prevalence of NPC cases was observed in reached to 164,671 cases (ASR = 5.7/100,000) and deaths were 114,871 cases (ASR = 3.9/100,000) in Vietnamese population. Notably, the symptoms of NPC were unclear such as hearing loss, nosebleeds, headache, feeling of fullness in the ear, etc., thus, NPC often presents at an advance stage when first diagnosis (stage 3 or 4). Therefore, there is a challenge for finding an early diagnosis and biomarker to achieve favorable treatment and increasing of patient's survival. Up to date, many studies have been demonstrated that multiple etiological factors, including Epstein-Barr virus (EBV) infection, genetics/or genetic susceptibility and epigenetics factors, have been suggested to be strongly linked with NPC (Hildesheim et al., 1993). Moreover, there are growing evidences demonstrating that many single nucleotide polymorphisms (SNPs)

in EBV, including *RPMS1* and *A73*, may have associated to NPC risk through the contribution to many cell signal pathways, such as cell-cycle regulation, apoptosis, cell differentiation,... (Brooks et al., 1993; Li et al., 2005; Li et al., 2006).

RPMS1 gene, belongs to the *BamHI-A rightward transcripts* (*BARTs*) family, starts at position 138,325 and ends at position 160,531 on the wild EBV gene (Smith et al., 2000; Sadler et al., 1995; de Jesus et al., 2003). Up to now, a number of studies have been demonstrated that SNPs of *RPMS1* G155391A, was specific to NPC, based on the amino acid substitution at position 155391 (Genbank, NC_007605). Its substitution resulted the decrease of degradation of the oncogenic RPMS1 protein. (Feng et al., 2015; Cui et al., 2017; Wu et al., 2018). According to Wu et al. (2018), they observed four significant variants: G155325T, G155326A, C155389T and G155391A, on the RPMS1 gene, and classified as the four subtypes of RPMS1: RPMS1-A, RPMS1-B, RPMS1-C, and RPMS1-D. Among them, RPMS1-C, RPMS1-D were reported to be strongly associated with NPC tumor formation. (Wu et al., 2018). *A73* (also called *RB2*), one of the primary members in EBV *BARTs* family, is reported to be related to the NPC tumorigenesis (Zhang et al., 2007). The study indicated that A157154C in *A73* gene, was detected almost exclusively in Chinese populations and preferentially exists in biopsies of nasopharyngeal carcinoma (Li et al., 2006). Not only that, three polymorphisms A157154C, G159188C, and G159209C were also identified in *A73* gene by Zhang et al. (2007). Among them, A157154C showed high frequency and might be correlated to the occurrence and development of NPC (Zhang et al., 2007; Shen et al., 2015).

In Vietnam, there is still no research about the variants of *RPMS1* and *A73* in NPC, thus, in current report, we summarized the current evidences about the polymorphism of *RPMS1* and *A73* in NPC based on the various previous publication via a systematic literature revision. The identification of SNPs of these genes may serve as targets for our further experiment and might become a potential biomarker for NPC in Vietnamese population.

2. Materials and Methods

2.1. Literature analysis, eligibility criteria and data extraction

A systematic literature analysis was conducted based on the comprehensive search of observational studies. Previous studies were obtained from the following databases using validated search strategies: PubMed, Web of Science databases, Embase Database, etc. The following keywords were used: Nasopharyngeal carcinoma, *A73*, *RPMS1*, single nucleotide polymorphism, etc.; conducted up to April, 2019. For these articles fulfilling the following criteria were used in the systematic review. Inclusion criteria: [1] Case-control study or cohort study that published the SNPs of *RPMS1* and *A73* gene profile. [2] Studies that investigated SNPs of *RPMS1* and *A73* in NPC. [3] Study that mention detection, analysis method of SNPs of *RPMS1* and *A73*.

Using these articles, we extracted the frequency of detection *RPMS1* and *A73* variants, country, method, etc. When data were available, we reported the ranged and computed average weight frequencies of candidate gene in studies.

2.2. Evaluation primer for PCR-sequencing

The sequence and information of *RPMS1* and *A73* gene were downloaded from Genbank (https://www.ncbi.nlm.nih.gov/genbank/) with accession number NC_007605, V01555 respectively. For evaluation of PCR primers, primer's physical characteristics were computed by IDT OligoAnalyzer 3.1 (http://sg.idtdna.com/calc/analyzer), Annhyb and primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) were tools for finding specific primers, as well

as PCR product length.

2.3. Building phylogenetic tree

For the *RPMS1* and *A73* databases, the variants' sequences were collected from Genbank (NCBI) based on the previous publications. The relationship between *RPMS1* subtypes, as well as *A73* were evaluated by Phylogenetic Tree was built by MEGA6 software based on using Joining Neural algorithm, the optimal evolutionary model was Jukes-Cantor and bootstrap values repeated 1000 times.

3. Results

3.1. Database of variants in RPMS1 and A73 gene

Up to April 2018, there was only three published studies, investigated about SNPs of *RPMS1*. According these studies, *RPMS1* was classified into four subtypes: RPMS1-A, RPMS1-B, RPMS1-C, RPMS1-D (Wu et al., 2018). In particular, RPMS1-A includes random mutants in positions from nt 155277 – 155546 of *RPMS1* gene, for example: G155284T, G155307T, C155358T, C155481A, C155525T, A155545G (Wu et al., 2018). The RPMS1-B consisted of a mutant in the nucleotides G155377A (RPMS1-B1) and C155389T (RPMS1-B2) (Wu et al., 2018). The RPMS1-C included mutant in position G155391A (Wu et al., 2018. RPMS1-D variant encompassed mutant position G155325T and G155326A (Wu et al., 2018) (Table 1).

By using many keywords described above, a systematic literature analysis was conducted based on the comprehensive search of *A73* variants studies. In results, only one the polymorphism of *A73* included a characteristic variant was A157145C, related to NPC (Li et al., 2006; Zhang et al., 2007; Shen et al., 2015; Han et al., 2013). All polymorphism studies in the *A73* gene included: Li et al. (2006); Han et al. (2013); Zhang et al. (2007); Shen et al., (2015), certainly definite only polymorphism point A157145C was significantly associated in NPC. This variant was located on exon V in the *A73* gene also in the *A73* coding region, but did not cause any amino acid alteration in protein A73 (CCA: Pro → CCC: Pro) (Table 1).

Table 1

Location of variants in the gene *RPMS1* and *A73* in NPC

	Variants	**Positions**	**Changed nucleotide**	**Changed amino acid**
RPMS1	RPMS1-A	*	-	*
	RPMS1-B1	155377	G → A	S → N
	RPMS1-B2	155389	C → T	P → L
	RPMS1-C	155391	G → A	D → N
	RPMS1-D	155325 155326	G → T G → A	G → Stop codon
A73	A157154C	157154	A → C	Not change

Note: "*" The changed nucleotides were randomized in the *RPMS1* gene's region (from nucleotide 155277 to 155546), except for nucleotides positions of the RPMS1-B, C, D.

3.2. Characteristics of included studies

Overall, total of three previous published studies were identified, accessed for eligibility from inclusion criteria, enrolled into systematic revision by using keywords described above. The characteristics of those studies were shown in Table 2.

Table 2

Studies included in the systematic review about *RPMS1* variants in NPC

Studies	SCS	SC	RPMS1-A		RPMS1-B		RPMS1-C		RPMS1-D	
			Case	Control	Case	Control	Case	Control	Case	Control
Wu et al., 2018	biopsy	WB	74/151	104/136	17/151	22/136	48/151	10/136	12/151	0/136
Feng et al., 2015	biopsy	WB	-	-	-	-	42/50	25/54	-	-
			-	-	-	-	914/1109	878/2052	-	-
Cui et al., 2017	biopsy	WB	-	-	-	-	17/100	33/54	-	-
Average weight frequency			49%	76%	11%	16%	74%	37%	8%	0%

Note: SCS: Source of cancer samples; SC: source of control; WB: wash brush; "-": not data

Overall, average weight frequency of RPMS1-A were 49% and 76% in NPC samples and non-cancerous samples; the average weight frequency of RPMS1-B were 11% and 16% in NPC samples and non-cancerous samples. Meanwhile, the frequencies of RPMS1-C variant were quite high in NPC samples (74%) and low in non NPC samples (34%). In contrast to RPMS1-C, the variant of RPMS1-D appears with low frequency in both case and control samples (8% and 0%, respectively).

According to *A73*, after exclusion studies that did not meet the inclusion criteria, up to now, only four studies carried out for analysis polymorphism of *A73* gene in NPC. The characteristics of those studies were shown in Table 3.

Table 3

Studies included in the systematic review about *A73* variant (A157154C) in NPC

Studies	SCS	SC	AA		AC		CC	
			Case	Control	Case	Control	Case	Control
Shen et al., 2015	PB	PB	54/510	98/520	214/510	208/520	242/510	214/520
Zhang et al., 2007	biopsy	PB	15/162	24/99	0/162	9/99	147/1162	66/99
Li et al., 2006	biopsy	WB	0/7	-	0/7	-	7/7	-
Han et al., 2013	biopsy	WB	2/51	20/52	0/51	10/52	49/51	22/52
Average weight frequency			9%	21%	5%	28%	78%	45%

Note: SCS: Source of cancer samples; SC: source of control; PB: peripheral blood; WB: wash brush; "-": not data

In general, at nt 157154 in *A73* gene, the frequencies of CC genotype in NPC samples and non-cancerous samples were 78% and 45%, respectively, higher than AA genotype (9% in case, 21% in control) and AC genotype (5% in case, 28% in control). The CC genotype

frequency in the NPC group was significantly higher than that in the control group, the difference was significant (78% vs. 45%; $p < 0.001$). Therefore, A157154C polymorphism of the *A73* gene in EBV was associated with NPC susceptibility.

3.3. Polymerase chain reaction (PCR) - sequencing method and selection primer

The sequences of *RPMS1* and *A73* genes was obtained from Genbank with the accession number NC_007605, V01555, respectively. The primers that identified variants in the *RPMS1* genes was collected from the research by Feng et al. (2015). At the same time, primers identified variants in the *A73* gene were obtained from the study of Shen et al. (2015). Parameters of primer pairs, were shown in Table 4 such as melting temperature, length, GC-base pair ratios, Gibbs Free Energy (ΔG) for secondary structures (hairpin, self-dimer, and heterodimer).

Table 4

Primer sequences and parameters

Primer	Sequence (5'-3')	L	%GC	Tm	ΔG (Kcal/mol^{-1})			(4)
					(1)	(2)	(3)	
RPMS1-F	GCTGGGTTGATGCTGTAGATG	21	52,4	55,9	0,32	-3,14	-3,17	733
RPMS1-R	AGGGTCTGGACGTGGAGTTTG	21	57,1	59,2	0,27	-6,3		
A73-F	AACATTGGGAATGAGAC	17	41,2	43,6	-1,78	-5,37	-3,43	711
A73-R	TTTGACCTGGAGGGCATC	18	55,6	54,7	-1,14	-4,67		

Note: L: Length of primer (bps); Tm: melting temperature (°C); %GC: GC-base pair ratios; Gibbs free energy (Kcal/mole^{-1}) for hairpin loop (1); homodimer (2) and heterodimer (3) structure formations; (4) Product sizes (bps).

The primer was obtained from the study of Shen et al. (2015), used to amplify the sequence of *A73* gene by PCR method. The amplification product of the *A73* gene was shown in Figure 1.

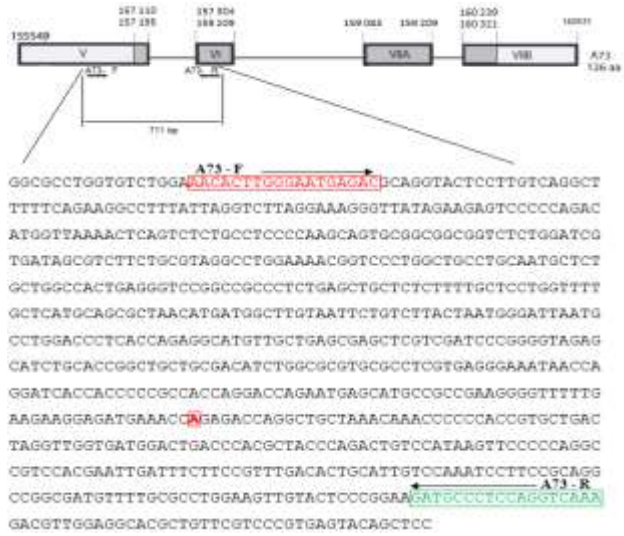

Figure 1. Position of primers used to amplified *A73* gene, and the location of SNPs

Note: The red nucleotide is located at polymorphism point 157154 on the strain B95-8 sequence (Accession number: V01555). The Figure 1 showed the position on the genome, structure of the *A73* gene and location of primers. *A73* gene is described by exon (box) and intron (straight line). The open reading frame (ORF) of *A73* genes was also shown as a dark gray box. The number of amino acids in the open reading frame was shown in the figure. Specific primer pairs were displayed by arrows and size of PCR products were also shown.

In 2015, Feng et al used SNP-RPMS1-F, R primer to conduct a position analysis of G155391A variant on NPC biopsy samples and benign brush sample in the Chinese population. Most recently in 2018, Wu et al had also used this primer in his study to detect G155391A and other SNP variants of *RPMS1* from NPC tissue samples and negative NPC brush samples. The amplification product of the primers, containing the sequence of variants in *RPMS1* gene, were shown in Figure 2.

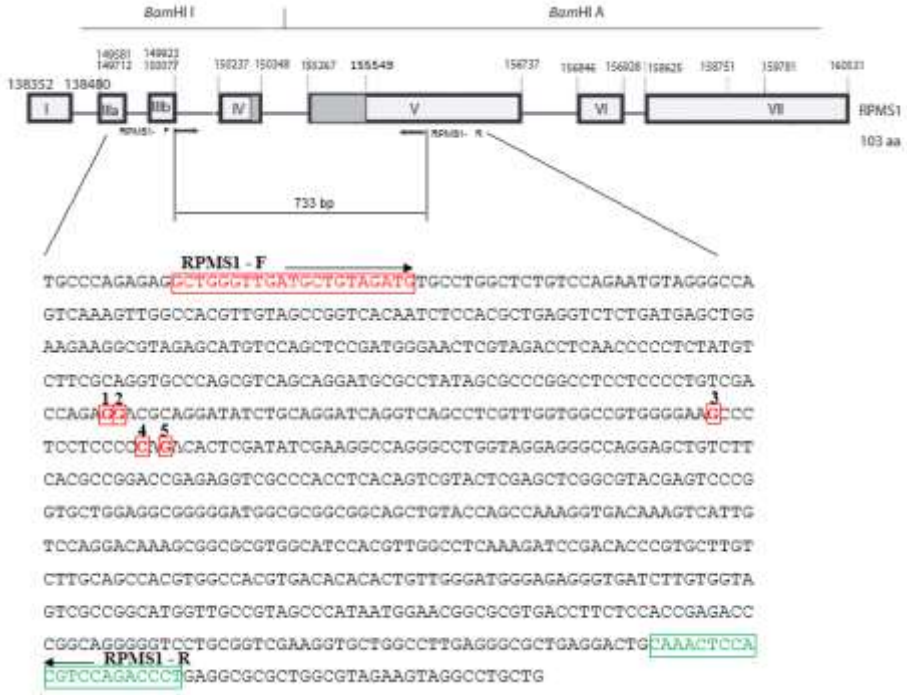

Figure 2. Position of primers used to amplify *RPMS1* gene, and the location of SNPs

SNP at positions 155325, 155326, 155377, 155389 and 155391 on the Wild-type EBV sequence (Accession number: NC_007605). The Figure 2 shows the position on the genome, the structure of *RPMS1* gene and primers. *RPMS1* is described by exons (box) and intron (straight line). The open reading frame *RPMS1* was also shown as a dark gray box. The number of amino acids of *A73* protein and sizes of PCR products were also shown. Specific primer pairs were displayed by arrows.

The sequences of *RPMS1* and *A73* genes were obtained from different accession number based on the distinguished origin of strains or types of infections. These sequences were conducted to align for compare sequences, the results were shown in Figure 3 & 4.

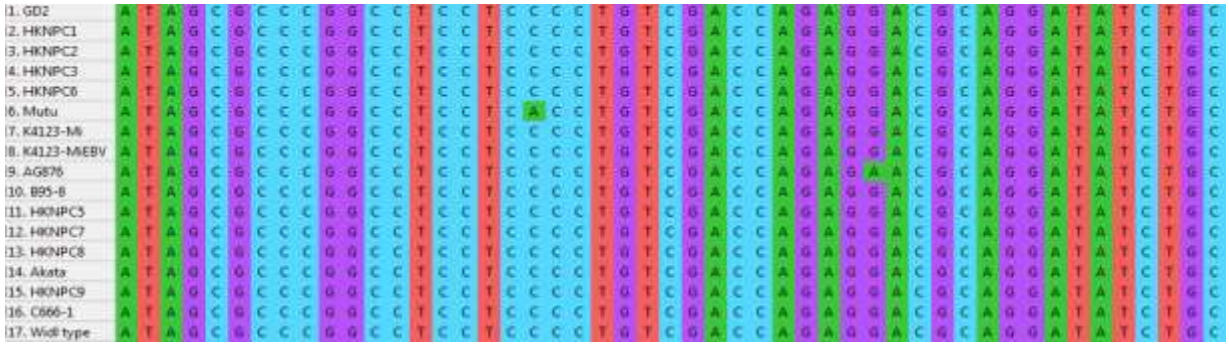

Figure 3. The sequence alignment of *RPMS1* sequences from different strains

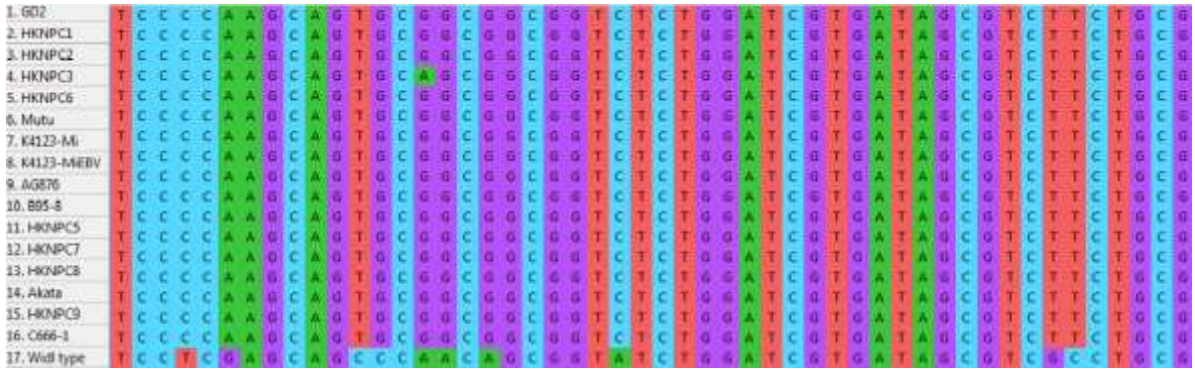

Figure 4. The sequence alignment of *A73* sequences from different strains

The results of sequence alignment of *RPMS1* and *A73* showed that the conservational ratio of this gene is quite high. In addition, the conservation sequences of this gene are very specific to EBV via BLAST (data not shown). Therefore, it can be concluded that conservational gene region of *RPMS1* and *A73* has sufficient basis for the target sequences to analysis polymorphism in *RPMS1* and *A73* genes.

3.4. Phylogenetic tree of RPMS1 and A73 gene

With the aim of surveying of the distribution of variants of *RPMS1* and *A73* gene in EBV strains in the world, thus, the database was collected from *RPMS1* and *A73* gene sequences of EBV belonging to different strains in the world, such as: Asia (China, Hong Kong, Japan), Africa (Kenya, Ghana), America (USA) was collected to build phylogenetic trees. The database of *RPMS1* and *A73* gene included: GD1, GD2, HKNPC1, HKNPC2, HKNPC3, HKNPC4, HKNPC5, HKNPC6, HKNPC7, HKNPC8, HKNPC9, Akata, C666-1, Mutu, K4123-Mi, K4123-MiEBV, AG876, B95-8, Wild type, Macacine.

Table 5

Database of *RPMS1* and *A73* gene

STT	Accession number	Strain	Subtype (RPMS1)	Subtype (A73)	Country/ region	
1	AY961628	GD1	RPMS1-B	-	China	Asia
2	HQ020558	GD2	RPMS1-C	A157154C	China	
3	JQ009376	HKNPC1	RPMS1-C	A157154C	Hong Kong	
4	KF992564	HKNPC2	RPMS1-C	A157154C	Hong Kong	
5	KF992565	HKNPC3	RPMS1-C	A157154C	Hong Kong	
6	KF992567	HKNPC4	RPMS1-C	A157154C	Hong Kong	
7	KF992567	HKNPC5	RPMS1-C	A157154C	Hong Kong	
8	KF992568	HKNPC6	RPMS1-C	A157154C	Hong Kong	
9	KF992569	HKNPC7	RPMS1-C	A157154C	Hong Kong	
10	KF992570	HKNPC8	RPMS1-C	A157154C	Hong Kong	

STT	Accession number	Strain	Subtype (RPMS1)	Subtype (A73)	Country/ region	
11	KF992571	HKNPC9	RPMS1-C	A157154C	Hong Kong	
12	KC207813	Akata	RPMS1-B	-	Japan	
13	KC617875	C666-1	RPMS1-C	A157154C	Hong Kong	
14	KC207814	Mutu	RPMS1-B	-	Kenya	Outside of Asia
15	KC440851	K4123-Mi	RPMS1-C	-	USA	
16	KC440852	K4123-MiEBV	-	-	USA	
17	DQ279927	AG876	RPMS1-D	-	Ghana	
18	V01555	B95-8	-	-	USA	
19	NC_007605	Wild type	-	-		
20	NC_006146	Macacine	Out-group		USA	

Note: "-": The sequence has not yet detected variant.

Phylogenetic tree of *RPMS1* gene was built, based on local database, aligned and synchronized. The length of the sequences after synchronization, was 733 bps, started from the nucleotide coding for amino acids 10 to amino acid 103 in each sequence.

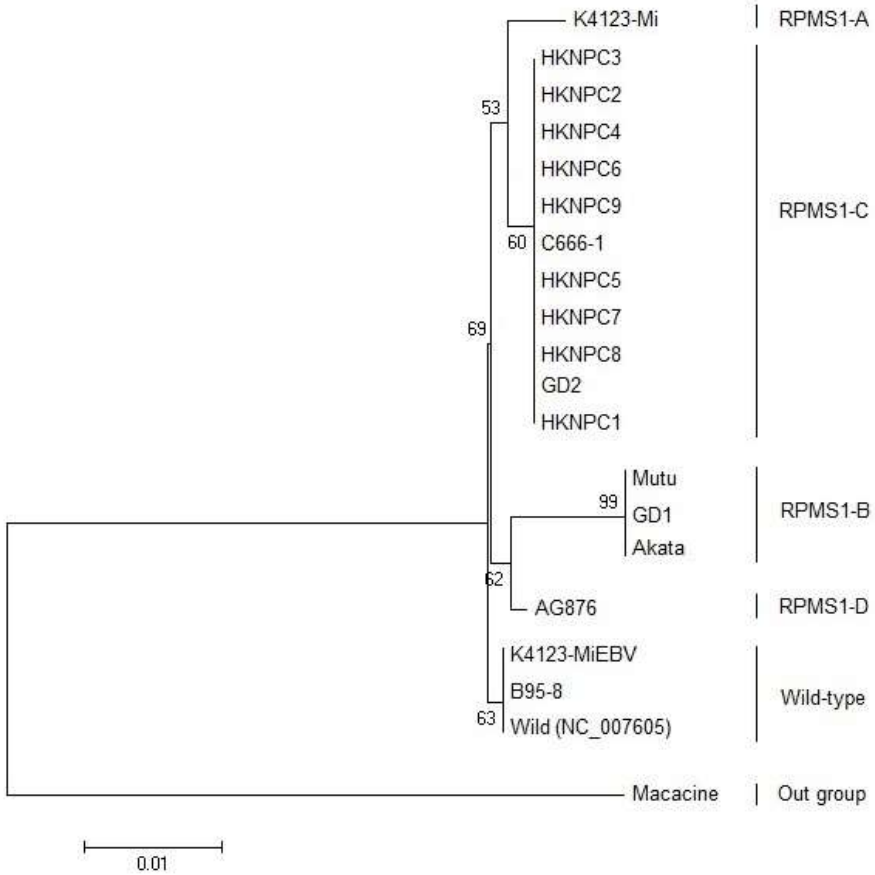

Figure 5. Phylogenetic tree of *RPMS1* gene

In the phylogenetic tree, 19 nucleotide sequence of EBV strains was classified into four genotypes: RPMS1-A, RPMS1-B, RPMS1-C, and RPMS1-D. Line segment lengths proportional to evolutionary distance of EBV strains. The results showed that RPMS1-C was detected in most of EBV strains in the world. These strains are mostly in Asia, such as HKNPC1, 2, 3,... of Hong Kong, GD2 of China. The RPMS-B, RPMS1-A, RPMS1-D variant were detected less and scattered in different regions of the world. For example, RPMS1-D was detected in Ghana (Europe). The remaining variants (RPMS1-B, C, D) are variants that appeared with low frequency or are characteristic geographic variants.

Phylogenetic tree of *A73* gene was built, based on local database, aligned and synchronized. The length of the sequences after synchronization, was 711 bps, started from the nucleotide coding for amino acids 1 in each sequence.

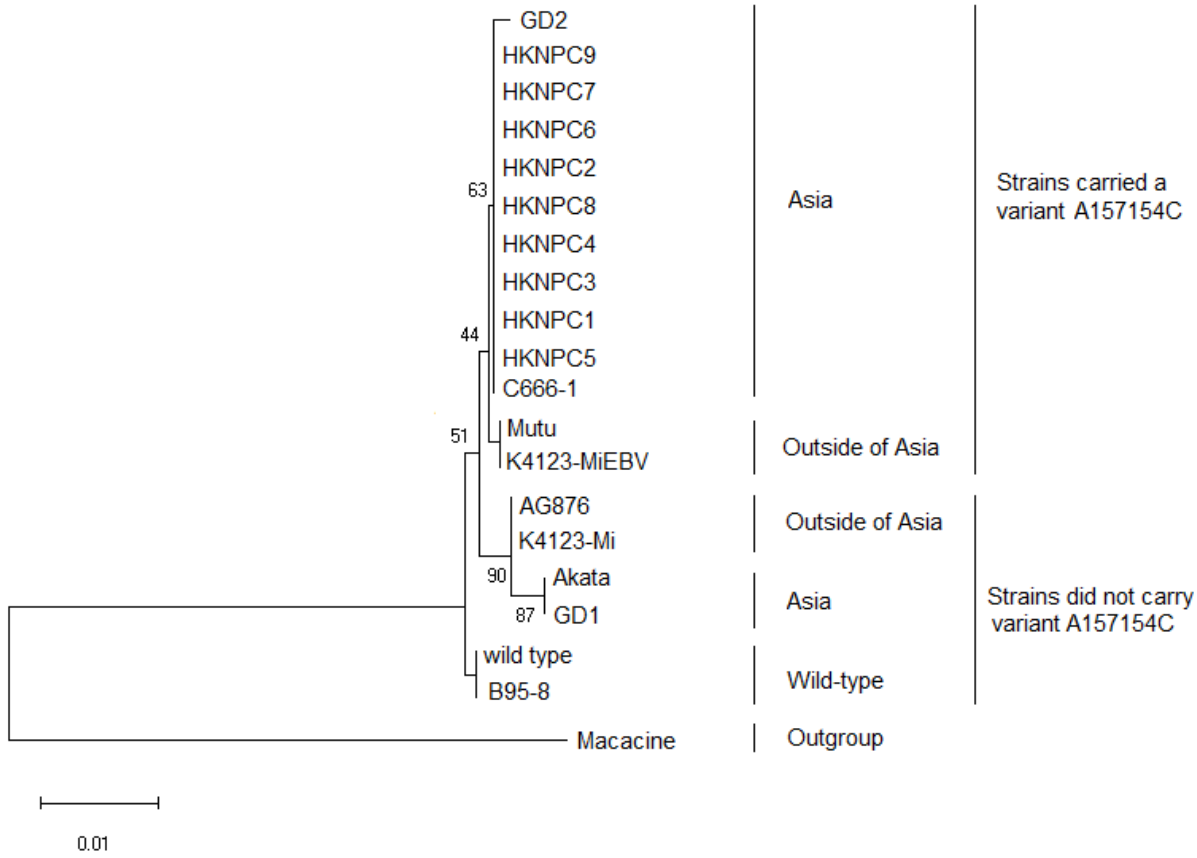

Figure 6. Phylogenetic tree of *A73* gene

The results of phylogenetic tree of *A73* gene showed that the A157154C variant was detected in most strains from Asia (C666-1, GD2, HKNPC1, etc.) and a few of strains in Africa (Mutu), America (K4123-Mi). The EBV strains group did not carry out A157154 variant was relatively less than the group carried A157154C variant, these strains were scattered in different regions of the world such as Japan (Akata), China (GD1), and the United States (K4123-Mi), Ghana (AG876). Consequence, the A157154C variant was mainly detected in EBV strains from Asia (especially Hong Kong, China), might be significantly associated with NPC in Asia - an endemic area for NPC.

4. Discussion

Nasopharyngeal carcinoma is one of the

commonly occurring cancers among Asian region, including Vietnam with the high prevalence. Within the unclear symptoms, nasopharyngeal cancer often presents in the last stage when first diagnosis. Therefore, the detection of NPC at the early stage based on a potential biomarker (Epstein, 1993). For the past few years, in addition to the viral infections, which has been indicated certainly will led to improve treatment and outcome for early diagnosis to be strongly related to human NPC, the polymorphism of *RPMS1* and *A73*, is now regarded as one of the important mechanisms in the cancer development. In Vietnam, up to date, there was no study was conducted on evaluation an association between patterns of *RPMS1* and *A73* led to risk of NPC. Therefore, in our initial study, we have to integrate the previous knowledge and findings to have a SNPs profiles of RPMS1 and *A73* gene by analysis frequencies of *RPMS1* and *A73* variants.

According to a study by Wu et al. 2018, RPMS1-C and RPMS1-D exhibited a strong association with tumor formation. RPMS1-A could be detrimental to tumor formation, and RPMS1-B2 had a strong correlation with tumor formation in northern China. However, the mechanism that affects tumor formation related to these variants have not been clarified.

In the remaining studies of variants in *RPMS1*, these studies focused on analysis the polymorphism point (G155391A) in RPMS1-C variants (Cui et al., 2017, Feng et al., 2015). The results show that this SNP position is a characteristic variant for NPC, especially in Cui et al (2017), the frequencies of G155391A were significantly higher in the 50 matched samples from NPC patients (84% in NPC biopsy samples and 82% in NPC washing brush samples) than that in 54 healthy throat washing samples (39%) (Cui et al., 2017). Consequence, G155391A variant was determined significantly associated in NPC.

In the study of Feng et al (2015), the authors indicated that the variant of G155391A at polymorphism point 155391, changed G → A, led to aa changed (Asp → Asn), could be related to the transcription or expression of *RPMS1*. The function of SNP G155391A in *RPMS1* has been proven in Feng's study. *RPMS1* genes carried a variant 155391A and did not carry variant (EBV wild-type) was grafted into pBABE-Puro vector. Then, these vectors were transferred and cultured on NP69 cell lines. Protein expression in these cells was examined by Western blot method. Results showed that cells treated with cycloheximide (CHX) (which inhibited protein formation in eukaryote organisms), the degradation of RPMS1 protein was observed after 0.5 hours in NP69 cells with wild-type RPMS1 (155391G), while the degradation of protein in NP69 cells with mutated RPMS1 (155391A) was slower (2.5 hours). At the same time, the half-life of the mutant RPMS1 protein 155391A was significantly longer than the RPMS1 protein without mutation (3.2 vs 0.6 h), indicating that SNP G155391A has a stabilizing function in RPMS1 protein or the stability of RPMS1 protein corresponds to the stability of the functional tumorigenesis.

Statistical analysis of *A73* variant showed that mutation at polymorphism point (A157145C) was a common SNP of *A73* on NPC samples. The average weight frequencies of CC genotype on NPC samples were 80%. In the healthy sample, the average weight frequencies of CC genotype were 45%, lower than the frequencies found in cancer samples. Therefore, it can be seen that the CC genotype at position 157145 is characteristic of NPC. It is obvious that all various studies on the polymorphisms of *RPMS1* and *A73* were concentrated in China - the country with the highest prevalence of NPC in the world. Therefore, based on our data, we concluded *RPMS1* and *A73* variants were significant association and contribution to the risk of NPC,

might serve as molecular targets for screening test for NPC, which was affirmed again based on our systematic literature analysis, especially, in further study, applied in Vietnamese population. Not only that, this study will be a basic detailed database, give a hand to contribute the polymorphism data of EBV gene in the world, along with providing necessary data on gene polymorphism EBV, the distribution (epidemiology) of EBV strains in Vietnam. Additionally, based on molecular phylogenetic analysis, in further study, we could proceed to determine the relationship between polymorphism of *RPMS1* and *A73*, identified in Vietnamese population, with the world.

5. Conclusion

We successfully performed a systematic literature revision and computed average detection frequencies of *RPMS1* and *A73* variants. A significant association between RPMS1-C, RPMS1-D in *RPMS1* gene; A157154C polymorphism of the *A73* gene and NPC was shown and confirmed by systematical of previous studies. The location of variants were carried out and definitely confirmed the specificity and referent PCR primers. These databases will be the useful information for the identification of variants in *RPMS1* and *A73* genes that may serve as biomarker for the further experiment and screening test for NPC in Vietnamese population■

References

Brooks, L. A., Lear, A. L., Young, L. S., & Rickinson, A. B. (1993). Transcripts from the Epstein-Barr virus BamHI A fragment are detectable in all three forms of virus latency. *Journal of virology*, *67*(6), 3182-3190.

Cui, Q., Feng, F. T., Xu, M., Liu, W. S., Yao, Y. Y., Xie, S. H.,... & Bei, J. X. (2017). Nasopharyngeal carcinoma risk prediction via salivary detection of host and Epstein-Barr virus genetic variants. *Oncotarget*, *8*(56), 95066.

de Jesus, O., Smith, P. R., Spender, L. C., Karstegl, C. E., Niller, H. H., Huang, D., & Farrell, P. J. (2003). Updated Epstein–Barr virus (EBV) DNA sequence and analysis of a promoter for the BART (CST, BARF0) RNAs of EBV. *Journal of General Virology*, *84*(6), 1443-1450.

Epstein JB, Jones CK (1993). Presenting signs and symptoms of nasopharyngeal carcinoma. *Oral Surg Oral Med Oral Pathol*, *75*, 32-6.

Feng, F. T., Cui, Q., Liu, W. S., Guo, Y. M., Feng, Q. S., Chen, L. Z.,... & Middeldorp, J. M. (2015). A single nucleotide polymorphism in the Epstein-Barr virus genome is strongly associated with a high risk of nasopharyngeal carcinoma. *Chinese journal of cancer*, *34*(3), 61.

GLOBOCAN (2018). Estimated cancer incidence, mortality and prevalence worldwide in 2018, http://globocan.iarc.fr/Pages/fact_sheets_population.aspx.

Han, R. M., Research Progress In The CDKN2A Gene And EBV A73 Gene Polymorphism With nasopharyngeal Carcinoma In Yunnan China. (2013), *Globe Thesis*, link: https://www. globethesis.com/?t=2214330374955373.

Hildesheim, A., & Levine, P. H. (1993). Etiology of nasopharyngeal carcinoma: a review. *Epidemiologic reviews*, *15*(2), 466-485.

Li, A., Zhang, X. S., Jiang, J. H., Wang, H. H., Liu, X. Q., Pan, Z. G., & Zeng, Y. X. (2005). Transcriptional expression of RPMS1 in nasopharyngeal carcinoma and its oncogenic potential. *Cell Cycle*, *4*(2), 303-308.

Li, A., Zhao, S., Hu, W., Zhao, G. F., & Zhang, X. S. (2006). Cloning of A73 gene and its coding sequence analysis. *Journal of Southern Medical University*, *26*(6), 826-7.

Regaud, C. (1921). Lympho-épitheliome de l'hypopharynx traité par la roentgenthérapie. Bull Soc Franc *Otorhinolaryngol*, *34*, 209-214.

Sadler, R. H., & Raab-Traub, N. (1995). Structural analyses of the Epstein-Barr virus BamHI A transcripts. *Journal of virology*, *69*(2), 1132-1141.

Schmincke, A. (1921). Uber lymphoepitheliale geschwulste. *Beitr Path Anat*, *78*, 161-170.

Shen, J. J., Niu, W. N., Zhou, M., Zhou, F., Zhang, H. Y., & Wang, L. (2015). Association of Epstein Barr virus A73 gene polymorphism with nasopharyngeal carcinoma. *Genetic testing and molecular biomarkers*, *19*(4), 187-190.

Smith, P. R., De Jesus, O., Turner, D., Hollyoake, M., Karstegl, C. E., Griffin, B. E.,... & Farrell, P. J. (2000). Structure and coding content of CST (BART) family RNAs of Epstein-Barr virus. *Journal of virology*, *74*(7), 3082-3092.

Wu, S., Liu, W., Li, H., Zhao, Z., Yang, Y., Xiao, H.,... & Luo, B. (2018). Conservation and polymorphism of EBV RPMS1 gene in EBV-associated tumors and healthy individuals from endemic and non-endemic nasopharyngeal carcinoma areas in China. *Virus research*, *250*, 75-80.

Zhang, L. L., Li, D. J., Li, Z. H., Zhang, X. S., Zhang, R. H., Yu, X. J.,... & Jia, W. H. (2007). Correlation of Epstein-Barr virus A73 gene polymorphisms to susceptibility to nasopharyngeal carcinoma. *Chinese journal of cancer*, *26*(10), 1047-1051.