Supplementary Methods

Quantification of EBV DNA

TaqMan Fast Advanced Master mix were used (Applied Biosystems, USA). In each well, 5 µl of DNA was used in a total reaction volume of 20 µl. qPCR was performed with the following conditions: 50°C for 2 min, 95°C for 20 s, and 40 cycle of 95°C for 3 s and 56°C for 30 s.

Evaluation of miRNA expression

MiRNA expression profiling data were obtained from National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) DataSets (http://www.ncbi.nlm.nih.gov/gds) by using the search terms "(microRNA) AND Nasopharyngeal Carcinoma) AND RNA[Sample Type]". Only datasets which studied tissue samples with a sample size of at least 20 and published prior to August 2015 were considered. Five GEO datasets (GSE70970, GSE43039, GSE32960, GSE36682 and GSE32906) which fulfilled these criteria were analysed using the web tool GEO2R at the NCBI website (http://www.ncbi.nlm.nih.gov/geo/geo2r/). Differentially expressed miRNAs with significance (adjusted p value with false discovery rate correction, $p < 0.05$) were obtained by using limma R package as the default analysis. A total of 27 miRNAs were shortlisted based on this analysis for further validation in RT-qPCR (Supplementary Table 1). RNA samples of NW were subjected to reverse transcription (RT) using MicroRNA Reverse Transcription Kit (Applied Biosystems, USA) according to previous study[1]. Briefly, RT primer pool was prepared by mixing equal volume of 32 stem-loop specific 5X RT primers. Primer pool consists of 27 miRNAs of interest, three small nucleolar RNAs (U6, U44 and U48)

which are potential reference miRNAs and two *C. elegans* miRNA (cel-miR-39 and cel-miR-54) which are spike in controls for data normalisation. Each RT reaction consists of 4.65 µl of RNA sample, 0.2 µl 100 mM dNTPs, 0.15 µl RNase inhibitor (20 U/µl), 1 µl 10X Reverse Transcription Buffer, 1µl MultiScribe Reverse Transcriptase (50 U/µl) and 3 µl of RT primers pool. RT was performed with thermal condition as follows: 16°C for 30 min, 42°C for 30 min, 85°C for 5 min and 4°C until the reaction is removed from incubation. Preamplification primer pool was prepared by mixing 32 TaqMan primers to a final concentration of 0.02X each. Preamplification reaction consisted of 2.5 µl diluted RT products (1:4 dilution), 5 µl TaqMan PreAmp Master Mix (2X) and 2.5 µl TaqMan preamplification primer pool. Preamplification was carried out with the following conditions: denaturation at 95 ºC for 10 min, followed by 16 cycles of preamplification at 95 ºC for 15 s and 60 ºC for 4 min. Preamplified product was further diluted 1:5 before qPCR was carried out using a 96.96 dynamic array in BioMark (Fluidigm, USA). No-template-control and a series of standard points from pooled RNA of human cell/xenograft were run in the same dynamic array as control samples. qPCR for all samples were carried out in triplicate wells and average C_q were obtained from at least duplicate wells.

Pre-analysis quality control (QC) was performed to exclude any miRNA assays that do not exhibit linear amplification. Average C_q for NW were normalised according to normalisation factor calculated from the C_q of spiked-in synthetic oligonucleotides to remove technical bias[2,3].

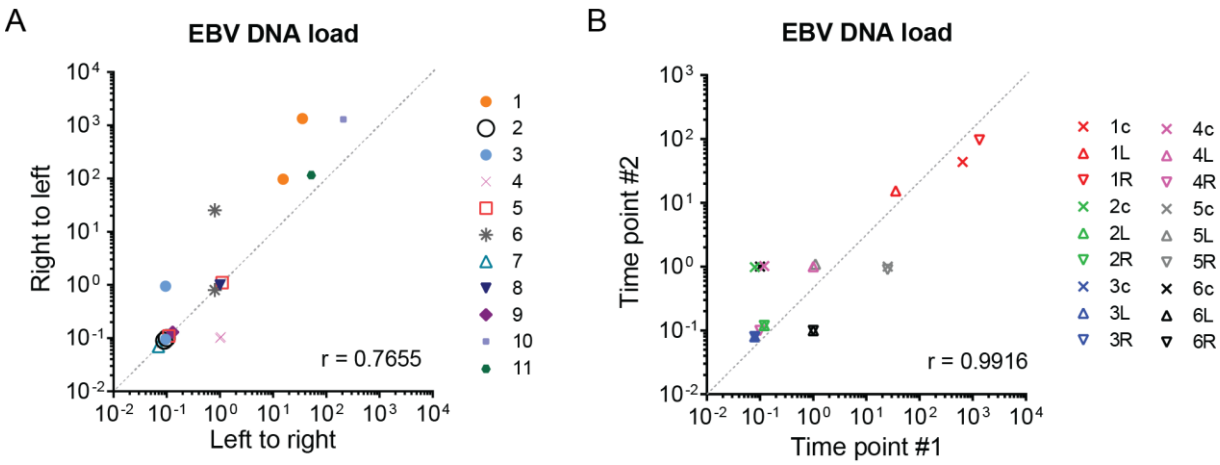
References

1. Heegaard NHH, Schetter AJ, Welsh JA, Yoneda M, Bowman ED, Harris CC. Circulating micro-RNA expression profiles in early stage nonsmall cell lung

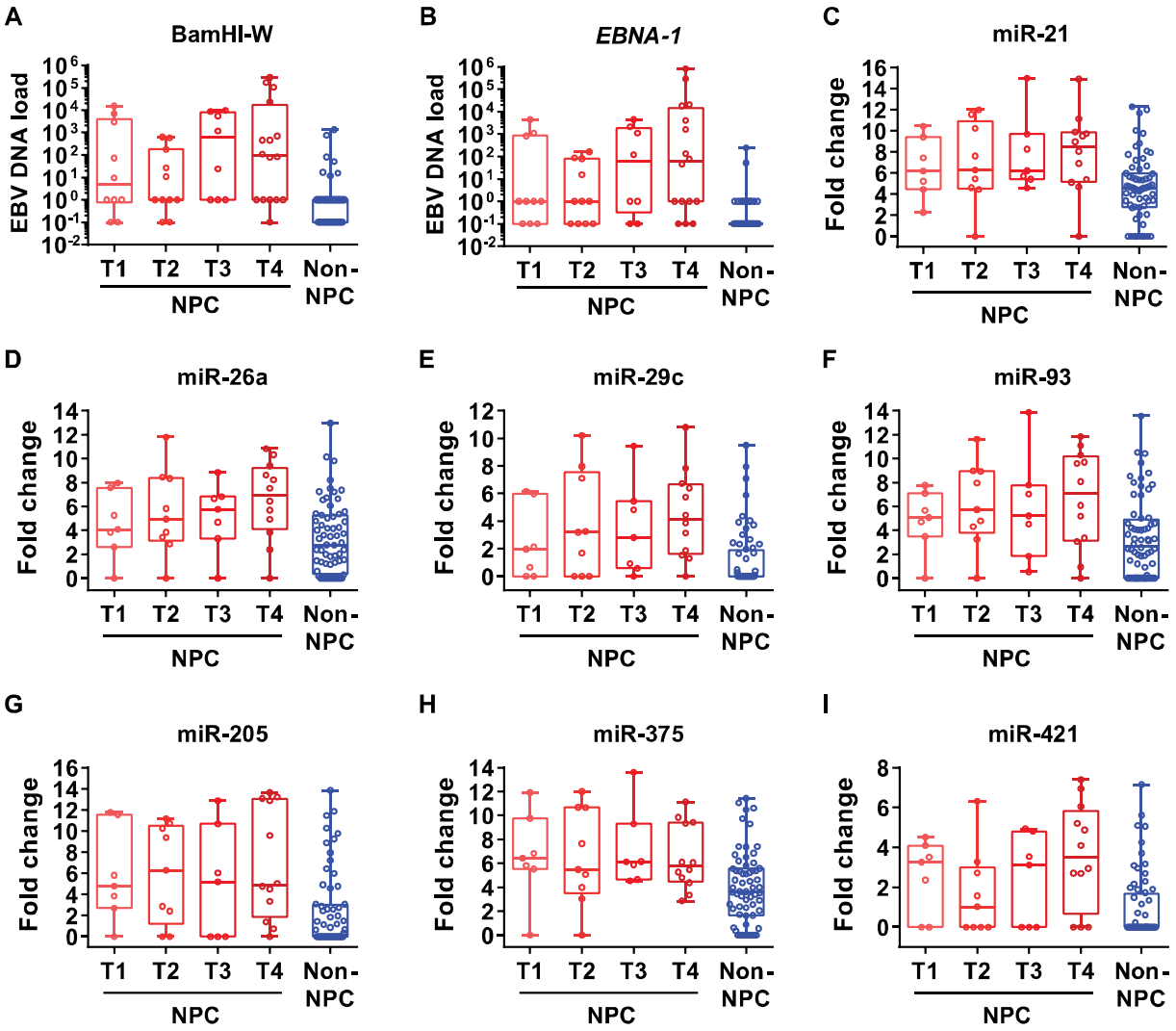
cancer. *Int J Cancer* 2012;130:1378–86.

2. Tan GW, Khoo ASB, Tan LP. Evaluation of extraction kits and RT-qPCR systems adapted to high-throughput platform for circulating miRNAs. *Sci Rep* 2015;5:9430.

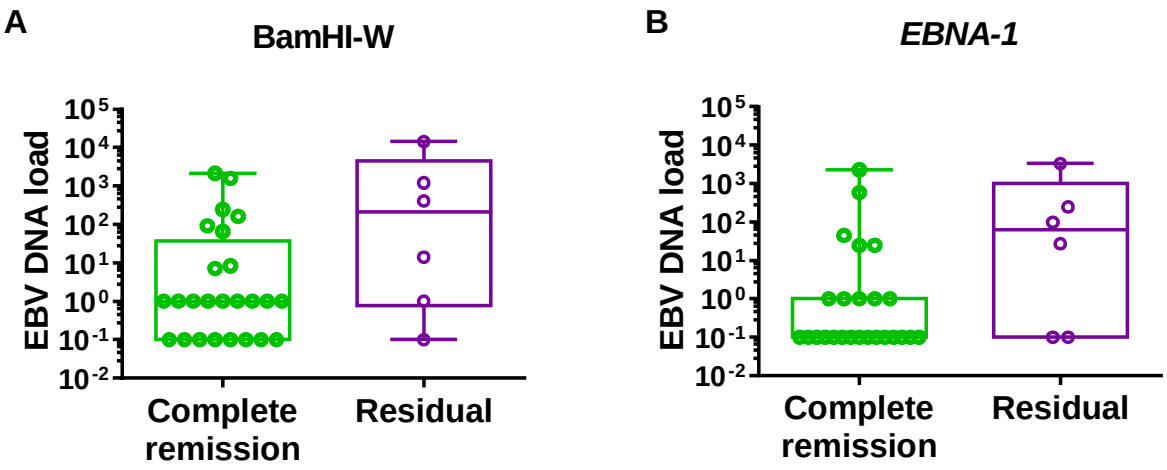
3. Kroh EM, Parkin RK, Mitchell PS, Tewari M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods* 2010;50:298–301.



Supplementary Figure S1. Comparison of EBV DNA load in nasal washings collected differently. (A) NW collected from the right nostril by introducing saline into the left nostril and vice versa. Patients with EBV DNA showed high EBV DNA load regardless of the side of NW samples being collected while patients that were negative for EBV DNA consistently had no EBV DNA on both sides. Each number indicates one patient with some patients having two samples collected at different time points. (B) NW collected from patients at two different time points during the course of treatment. All data points that were negative were given arbitrary values of 0.07 to 0.13 EBV DNA copy while samples with low EBV DNA load were given arbitrary values of 0.7 to 1.3. *c, combined lysate of left to right and right to left samples; L, NW samples from left to right; R, NW samples from right to left.*



Supplementary Figure S2. EBV DNA load and miRNA expression level in nasal washings of NPC patients with different T stages in comparison to non-NPC patients. Patients with T3 and T4 tumours had higher median EBV DNA loads compared to T1 and T2 tumours, as measured by (A) BamHI-W and (B) *EBNA-1* assay. NPC cases with higher T stages showed higher expression of (C) hsa-miR-21, (D) hsa-miR-26b, (E) hsa-miR-29c, and (F) hsa-miR-93, while no obvious trend is observed for (G) hsa-miR-205, (H) hsa-miR-375 and (I) hsa-miR-421.



Supplementary Figure S3. EBV DNA load in nasal washings. Higher median NW EBV DNA load is seen in residual samples compared to remission samples as measured by (A) BamHI-W assay and (B) *EBNA-1* assay.

Supplementary Table S1. Differential expression of miRNAs in different studies.

miRNA	GSE70970	GSE43039	GSE32960	GSE36682	GSE32906	this study
ebv-miR-BART4	UP*	UP	UP*	UP*	NT	NE
ebv-miR-BART6-3p	UP*	UP	UP*	UP*	NT	NE
ebv-miR-BART6-5p	UP*	UP	UP*	UP*	NT	NE
ebv-miR-BART7	UP*	UP*	UP*	UP*	NT	NA
ebv-miR-BART8	UP*	UP*	ND	UP*	NT	NE
ebv-miR-BART9	UP*	UP*	UP*	UP*	NT	ND
ebv-miR-BART10	UP*	UP*	UP*	UP*	NT	NE
ebv-miR-BART17-5p	UP*	UP	ND	UP*	NT	NE
ebv-miR-BART19-3p	UP*	ND	UP*	UP*	NT	NE
hsa-miR-142-3p	UP	ND	DOWN*	ND	ND	UP
hsa-miR-143	ND	DOWN	DOWN*	DOWN*	ND	ND
hsa-miR-145	DOWN*	DOWN	DOWN*	DOWN*	ND	UP
hsa-miR-155	UP*	ND	ND	ND	ND	UP
hsa-miR-196b	UP*	ND	ND	ND	ND	NE
hsa-miR-205	UP	UP*	UP*	UP*	ND	UP*
hsa-miR-21	UP*	ND	DOWN*	UP*	ND	UP*
hsa-miR-26a	ND	ND	DOWN*	DOWN*	ND	UP*
hsa-miR-26b	ND	DOWN*	DOWN*	DOWN*	ND	UP
hsa-miR-29b	UP*	DOWN*	DOWN*	DOWN	ND	ND
hsa-miR-29c	ND	DOWN*	DOWN*	DOWN*	ND	UP*
hsa-miR-34c-5p	DOWN*	DOWN*	DOWN*	DOWN*	DOWN	ND
hsa-miR-375	DOWN*	ND	ND	DOWN*	ND	UP*
hsa-miR-421	UP*	ND	DOWN*	DOWN*	ND	UP*
hsa-miR-451	ND	ND	DOWN*	DOWN*	ND	ND
hsa-mir-9	UP	ND	ND	ND	UP	ND
hsa-miR-93	UP*	ND	UP*	ND	ND	UP*
hsa-miR-99b	DOWN	ND	ND	DOWN*	UP	UP

*statistically significant

UP, >2-fold upregulation; DOWN, >2-fold downregulation; NT, not tested; NE, no expression in >80% of samples; NA, not analysed due to assay failing pre-analysis quality control; ND, no difference in expression between NPC

Supplementary Table S2. Clinical and demographics features of study subjects.

	All	NPC[a]	non-NPC	p value
	n = 119	n = 46	n = 73	
Sex				
Male	88 (73.9%)	37 (80.4%)	51 (69.9%)	> 0.05[b]
Female	31 (26.1%)	9 (19.6%)	22 (30.1%)	
Age				
mean years (range)	48.9 (15 - 85)	51.5 (15 - 85)	44.8 (19 - 78)	> 0.05[c]
Ethnicity				
Chinese	68 (57.1%)	30 (65.2%)	38 (52.1%)	> 0.05[b]
Malay	44 (37.0%)	15 (32.6%)	29 (39.7%)	
Others	7 (5.9%)	1 (2.2%)	6 (8.2%)	
AJCC Staging				
I		1 (2.2%)		
II		2 (4.3%)		
III		10 (21.7%)		
IVA		7 (15.2%)		
IVB		7 (15.2%)		
IVC		6 (13%)		
Incomplete staging		12 (26.1%)		
Unknown		1 (2.2%)		
Nasopharynx (T)				
T1		10 (21.7%)		
T2		11 (23.9%)		
T3		8 (17.4%)		
T4		16 (34.8%)		
Unknown		1 (2.2%)		
Regional node (N)				
N0		6 (13.0%)		
N1		5 (10.9%)		
N2		20 (43.5%)		
N3		4 (8.7%)		
N3a		4 (8.7%)		
N3b		5 (10.9%)		
Unknown		2 (4.3%)		
Metastasis (M)				
M0		27 (58.7%)		
M1		6 (13.0%)		
Mx		12 (26.1%)		
Unknown		1 (2.2%)		

[a]Only NPC patients that contributed to pre-treatment samples (n = 46) are included here for statistical analyses of differences in sex, age and ethnicity.

[b]Statistical significance as determined by Chi-square test.

[c]Statistical significance as determined by Mann-Whitney U test

Supplementary Table S3. Performance of biomarkers in the classification of NPC.

	Area under ROC curve				Sensitivity	Specificity	Positive predictive value	Negative predictive value
	Area	95% C.I.		p value				
		Lower	Upper					
BamHI-W[a]	0.774	0.684	0.864	<0.001	54.3%	90.4%	78.1%	75.9%
BamHI-W[b]	0.786	0.689	0.883	<0.001	48.6%	95.3%	85.0%	77.2%
EBNA-1 [a]	0.805	0.718	0.893	<0.001	71.7%	83.6%	73.3%	82.4%
EBNA-1 [b]	0.810	0.712	0.908	<0.001	74.3%	81.3%	68.4%	85.2%
hsa-miR-21[b]	0.736	0.633	0.839	<0.001	77.1%	62.5%	52.9%	83.3%
hsa-miR-26a[b]	0.708	0.600	0.817	0.001	74.3%	60.9%	51.0%	81.3%
hsa-miR-29c[b]	0.762	0.661	0.862	<0.001	80.0%	70.3%	59.6%	86.5%
hsa-miR-93[b]	0.716	0.609	0.823	<0.001	71.4%	70.3%	56.8%	81.8%
hsa-miR-205[b]	0.736	0.630	0.842	<0.001	65.7%	78.1%	62.2%	80.6%
hsa-miR-375[b]	0.733	0.630	0.835	<0.001	82.9%	57.8%	51.8%	86.0%
hsa-miR-421[b]	0.705	0.592	0.819	0.001	60.0%	84.4%	67.7%	79.4%
Regression model[b,c]	0.860	0.783	0.936	<0.001	80.0%	78.1%	66.7%	87.7%

C.I., confidence interval.

[a]Calculation was done for all samples evaluated (n = 119).

[b]Calculation was done for a subset of samples (n = 99) that were available for miRNA analyses.

[c]This model includes EBV DNA measured by the EBNA-1 assay and hsa-miR-21.

Supplementary Table S4. Markers in NPC analyzed by simple and multiple logistic regression.

	Simple logistic regression					Multiple logistic regresion[b]				
				95% C.I.					95% C.I.	
	B[a]	p value	Crude odd ratio	Lower	Upper	B[a]	p value	Adjusted odd ratio	Lower	Upper
BamHI-W[c]	0.287	<0.001	1.333	1.170	1.518					
EBNA-1[c]	0.363	<0.001	1.438	1.205	1.716	0.370	0.000	1.447	1.192	1.757
hsa-miR-21	0.262	<0.001	1.300	1.124	1.504	0.220	0.009	1.246	1.056	1.471
hsa-miR-26a	0.236	0.001	1.266	1.096	1.463					
hsa-miR-29c	0.345	<0.001	1.412	1.182	1.686					
hsa-miR-93	0.208	0.001	1.231	1.085	1.397					
hsa-miR-205	0.197	<0.001	1.218	1.098	1.350					
hsa-miR-375	0.262	<0.001	1.300	1.123	1.503					
has-miR-421	0.406	<0.001	1.500	1.201	1.874					

C.I., confidence interval.

[a]Regression coefficient

[b]Multiple logistic regression with forward and backward methods. All methods yielded the same results.

[c]Calculation for EBV DNA was done based on subset (n=99) for comparison with miRNA.

Supplementary Table S5. Comparison of EBV DNA tests using different minimally/non-invasive sample types.

Study	Population	Sample type	Requirement of otolaryngologist for sampling	Storage conditions		Quantified region/ gene	Sensitivity	Specificity	Positive predictive value	Negative predictive value
				Buffer	Temperature					
Tong *et al* , 2002	Hong Kong	NPB	Yes	NA	NA	BamHI-W	96.4%	96.2%	96.4%	96.2%
Stevens *et al* , 2006	Indonesia	NPB	Yes	NucliSens Lysis Buffer	-80°C	EBNA-1	91.0%	98.0%	97.0%	91.0%
Adham *et al* , 2013	Indonesia	NPB	Yes	NucliSens Lysis Buffer	-80°C	EBNA-1	94.3%	90.0%	84.4%	80.0%
Ng *et al* , 2014	Hong Kong and Canada	TOB	Yes	Proprietary preservation agent	Room temperature	EBNA-1	98.9%	99.3%	96.9%	99.7%
Chen *et al* , 2015	China	NPS	Yes	Saline	-80°C	BamHI-W	87.5%	98.9%	41.2%	99.9%
Zheng *et al* , 2015	China	NPB	Yes	RNAlater	-80°C	BamHI-W	96.0%	97.0%	97.0%	96.0%
Chan *et al* , 2008[a]	Hong Kong	urine	No	NA	NA	BamHI-W	56.8%	95.7%	93.3%	67.7%
Tan *et al.* (this study)	Malaysia	NW	No	Saline	-20°C	EBNA-1	71.7%	83.6%	73.3%	82.4%
						BamHI-W	54.3%	90.4%	78.1%	75.9%
Yip *et al.* 2018	various	blood	No	plasma processed from EDTA blood tube	-80°C	various	69-99%	87-100%	NM	NM

NPB, nasopharyngeal brushings; TOB, trans-oral brushings; NPS, nasopharyngeal swab; NW, nasal washings; NA, not applicable (sample tested directly after sampling); NM, not mentioned.

[a]Sensitivity, specificity, positive predictive value and negative predictive value were calculated based on reported results from the studies.

Tong JHM, Tsang RKY, Lo KW, Woo JKS, Kwong J, Chan MWY, Chang AR, Van Hasselt CA, Huang DP, To KF. Quantitative Epstein-Barr virus DNA analysis and detection of gene promoter hypermethylation in nasopharyngeal (NP) brushing samples from patients with NP carcinoma. Clin Cancer Res 2002;8:2612–9.

Stevens SJC, Verkuijlen SAWM, Hariwiyanto B, Harijadi, Paramita DK, Fachiroh J, Adham M, Tan IB, Haryana SM, Middeldorp JM. Noninvasive diagnosis of nasopharyngeal carcinoma: Nasopharyngeal brushings reveal high Epstein-Barr virus DNA load and carcinoma-specific viral BARF1 mRNA. Int J Cancer 2006;119:608–14.

Adham M, Greijer AE, Verkuijlen SAWM, Juwana H, Fleig S, Rachmadi L, Malik O, Kurniawan AN, Roezin A, Gondhowiardjo S, Atmakusumah D, Stevens SJC, et al. Epstein-barr virus DNA load in nasopharyngeal brushings and whole blood in nasopharyngeal carcinoma patients before and after treatment. Clin Cancer Res 2013;19:2175–86.

Ng RHW, Ngan R, Wei WI, Gullane PJ, Phillips J. Trans-oral brush biopsies and quantitative PCR for EBV DNA detection and screening of nasopharyngeal carcinoma. Otolaryngol Head Neck Surg 2014;150:602–9.

Chen Y, Zhao W, Lin L, Xiao X, Zhou X, Ming H, Huang T, Liao J, Li Y, Zeng X, Huang G, Ye W, et al. Nasopharyngeal Epstein-Barr Virus Load: An Efficient Supplementary Method for Population-Based Nasopharyngeal Carcinoma Screening. PLoS One 2015;10:e0132669.

Zheng X-H, Lu L-X, Li X-Z, Jia W-H. Quantification of Epstein-Barr virus DNA load in nasopharyngeal brushing samples in the diagnosis of nasopharyngeal carcinoma in southern China. Cancer Sci 2015;106:1196–201.

Chan KCA, Leung SF, Yeung SW, Chan ATC, Lo YMD. Quantitative analysis of the transrenal excretion of circulating EBV DNA in nasopharyngeal carcinoma patients. Clin Cancer Res 2008;14:4809–13.

Yip TTC, Ngan RKC, Fong AHW, Law SCK. Application of circulating plasma/serum EBV DNA in the clinical management of nasopharyngeal carcinoma. Oral Oncol 2014;50:527–38.