



## Evaluation of circulating EBV microRNA BART2-5p in facilitating early detection and screening of nasopharyngeal carcinoma

Chen Jiang<sup>1,4</sup> \*, Jinna Chen<sup>1\*</sup>, Shanghang Xie<sup>2\*</sup>, Lifang Zhang<sup>2</sup>, Yanqun Xiang<sup>3</sup>, Maria Lung<sup>1</sup>, Ngar-Woon Kam<sup>1</sup>, Dora Lai-wan Kwong<sup>1</sup>, Sumei Cao<sup>2</sup>, Xin-Yuan Guan<sup>1,4</sup>

<sup>1</sup>Department of Clinical Oncology, The University of Hong Kong, Hong Kong, China; <sup>2</sup>Department of Epidemiology, Cancer Prevention Center, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in Southern China, Guangzhou, China; <sup>3</sup>Department of Nasopharyngeal Carcinoma, Sun Yat-Sen University Cancer Center, Guangzhou, China; <sup>4</sup>State Key Laboratory of Oncology in Southern China, Sun Yat-sen University Cancer Center, Guangzhou 510060, China.

**Correspondence:** XY Guan, Department of Clinical Oncology, Li KaShing Faculty of Medicine, The University of Hong Kong, Room L10-56, Laboratory Block, 21 Sassoon Road, Pokfulam, Hong Kong. Tel: 852-3917-9782; E-mail: [xyguan@hku.hk](mailto:xyguan@hku.hk); or Sumei Cao, MD, PhD, Sun Yat-sen University Cancer Center, 651 Dongfeng Road, East, Guangzhou 510060, People's Republic of China, Fax: t86-20-87343535, E-mail: [caosumei@mail.sysu.edu.cn](mailto:caosumei@mail.sysu.edu.cn)

\*These authors contributed equally to this work.

**Keywords:** EBV, BART microRNA, early detection, nasopharyngeal carcinoma.

**Article category:** Tumor markers and signatures

**Novelty and Impact:** Nasopharyngeal carcinoma is an EBV-related malignancy which is highly prevalent in South China. As some EBV BART microRNAs are abundant in patient's circulation, the authors systematically evaluated their diagnostic value in early diagnosis of NPC. Here, they reported BART 2-5p could successfully distinguish NPC patients from controls. More importantly, BART 2-5p began to rise even before clinical diagnosis of NPC as shown in a nested case-control study, suggesting its value in NPC early diagnosis.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/ijc.31642

## Abstract

Nasopharyngeal carcinoma is an Epstein–Barr Virus (EBV) associated malignancy which is highly prevalent in Southeast Asia. EBV-related antibodies have been widely used as screening markers for early nasopharyngeal carcinoma (NPC) detection. However, due to low positive predictive rate, it is essential to develop new biomarkers to facilitate NPC early diagnosis or triage EBV serological high-risk individuals to improve the chance of NPC early detection. BART microRNAs, which are encoded by BamHI region of EBV, were reported to be abundant in NPC and have potential value in early diagnosis of NPC. Here, we quantified the circulating level of 17 BART microRNAs in discovery stage based on previous microarray and sequencing data and, in particular, BART 2-5p, the sole candidate whose AUC was higher than 0.8, has been chosen for further study. In validation stage, the sensitivity, specificity and AUC of BART 2-5p was 93.9%, 89.8%, 0.972(95%CI: 0.954-0.989) respectively in cohort 1 constituted by NPC patients and controls from Hong Kong. For validation cohort 2 consisting of patients and controls from Guangzhou, the sensitivity, specificity and AUC was 94.2%, 83.5%, 0.959(95%CI: 0.939-0.980) respectively. To evaluate its ability to distinguish preclinical NPC patients, we established a nested case-control study with serum samples prospectively collected from 22 NPC patients prior to their clinical diagnosis and 88 matched healthy high-risk controls in a screening trial. The sensitivity and specificity were 90.9% and 54.5%. Collectively, EBV microRNA BART2-5p may be a valuable biomarker for early detection of NPC.

## Introduction

Nasopharyngeal carcinoma is highly prevalent in Southeast Asia and Southern China.<sup>1</sup> In endemic areas, especially Guangdong Province and Hong Kong, the incidence rate is about 20-50/100,000 person-years, causing serious health problems.<sup>2</sup> As nasopharyngeal cancer cells are quite sensitive to radiation, radiotherapy has been served as First-line therapy for NPC patients. If the disease is diagnosed at early stages, the five-year survival rate is above 90%.<sup>3</sup> However, for patients with advanced stages, less than half of them could survive longer than 5 years. Furthermore, radiotherapy is more likely to induce serious complications in patients with advanced disease as they have to receive larger dosage than early stage patients.<sup>4</sup> Unfortunately, due to lack of obvious clinical symptoms in early stages, most NPC patients have their disease diagnosed when it has already reached an advanced stage.<sup>4</sup> Therefore, it is essential to conduct screening and early detection to reduce mortality in endemic areas of NPC.

NPC is closely associated with Epstein-Barr virus infection,<sup>5</sup> and since EBV is specifically detected in cancer cells, it is possible to use EBV products as biomarkers for disease diagnosis.<sup>6</sup> Up to present, quantification of EBV-related serologic antibodies is the main means for early detection and screening of NPC.

For EBV serologic test, quantification of the titer of serum EBV VCA IgA and EBNA1-IgA by ELISA has been widely used for NPC screening with high diagnostic performance.<sup>7</sup> However, due to relatively low incidence for NPC in the screening population, the positive predictive value for EBV serologic high risk individuals is still less than 5%, leading unnecessary financial and psychological

burden to those individuals.<sup>8</sup> So, it requires new biomarkers to facilitate NPC early diagnosis or triage EBV serologic high risk individuals to increase NPC early detection rate and screening effect.

Another method reported for NPC diagnosis in the clinic is quantifying the level of plasma/serum EBV DNA from the patients.<sup>9</sup> Although the clinical application of plasma/serum EBV DNA has been extensively investigated and reported to have good performance for NPC monitoring of treatment and prognostication,<sup>10-12</sup> it has less satisfactory results in the early detection of NPC, especially for those high risk individuals who developed NPC one year after the test.<sup>13</sup> Another problem for plasma EBV DNA quantification is the fluctuation of EBV DNA level. Only 29% participants showed consistent results in follow-up samples, thus requiring repeated sampling to ensure sensitivity if not combined with other tests.<sup>14</sup> Therefore, it is necessary to develop new biomarkers for early detection of NPC.

MicroRNA is a short, single-stranded RNA which is about 22 nt, serving as gene regulators by degrading mRNA or repress translation.<sup>15</sup> The dysregulation of microRNAs was found to be associated with various human cancers.<sup>16</sup> However, most researchers were focused on the role of human microRNAs in cancer before the identification of EBV microRNA in 2004.<sup>17</sup> In NPC, BART region (BamHI A Rightward Transcripts), which encodes more than 40 mature microRNAs, was found to be actively transcribed.<sup>18</sup> Several groups have compared the levels of these microRNAs in NPC patients and healthy controls and found some microRNAs were highly abundant in NPC patients, suggesting it might be helpful in early diagnosis of NPC or act as a supplement for EBV-related antibody test in screening.<sup>19-22</sup> Up to present, there is no systematic evaluation of the potential value of these microRNAs as biomarkers in identifying early stage patients. In this study, we evaluated the potential of BART microRNAs as biomarkers for NPC diagnosis; in particular, we also elucidated their capability in distinguishing early stage patients from healthy high-risk individuals.

## Materials and Methods

### Sample collection

In total, we collected 811 serum samples from 3 groups of participants: NPC patients and non-cancer controls from Hong Kong, NPC patients and healthy controls from Guangzhou, NPC patients and healthy controls identified from the high-risk population in a screening scheme in Sihui, Guangdong. All clinicopathological characteristics of these participants were recorded and listed in Table 1, including age, gender, clinical stage, sample origin and risk assessment by EBV serologic test. For all NPC patients, they were diagnosed with pathological evidence and the clinical stage of which were determined with MRI scanning according to the 7<sup>th</sup> UICC/AJCC staging system. The non-cancer controls from Hong Kong were patients without cancerous disease in Queen Mary Hospital. The healthy controls from Guangzhou were recruited from the physical examination center of Cancer Center of Sun Yat-sen University.

All samples were distributed to discovery stage or validation stage.

Discovery stage

In this stage, serum sample of 24 NPC patients and 24 non-cancer controls were obtained from Queen Mary Hospital in Hong Kong between Jan 1, 2010 and Dec 31, 2014.

#### Validation stage

Three validation cohorts were involved in this stage. In validation cohort 1, there were another 148 NPC patients and 118 non-cancer controls recruited from Queen Mary Hospital between Jan 1, 2010 and Dec 31, 2014. Among the non-cancer controls, there were 106,8,4 individuals were defined as low, medium and high risk to develop NPC respectively based on quantifying the titer of serum EBV VCA IgA and EBNA1-IgA by ELISA. Validation cohort 2 was composed of 103 NPC patients and 284 healthy controls collected from Cancer Center of Sun Yat-sen University between Jan 8, 2010 and Oct 11, 2013. The 284 healthy controls included 183 low risk individuals, 93 medium risk individuals and 8 high risk individuals. The validation cohort 3 was designed as a nested case-control cohort. The 22 NPC patients and 88 healthy high-risk controls in the nested case-control study were from a screening scheme.<sup>7</sup> Eligible residents in Sihui were enrolled in this scheme and received the ELISA test as mentioned above between June 2007 to June 2017. The high-risk individuals were followed up continuously and received annual check by EBV antibody test, endoscopy and/or MRI to exclude NPC development until Dec 31,2017. Their entry date and clinical diagnosis date were recorded. (Supplemental Table 1) The median follow-up time for these high-risk controls was 32 months. The serum samples of NPC cases were collected from the patients identified in this population, 2-79 months (median:12.5 months) prior to their clinical diagnosis. The controls were age and gender matched high risk participants without NPC development. Each NPC case was matched to 4 high risk controls by age varying within 4 years old, gender, the year of enrollment and follow-up years. Also, all the high-risk controls were examined by endoscopy to exclude NPC development when they were taken blood.

This study was approved by Institutional Review Board of the University of Hong Kong and written informed consent was obtained from all participants. All serum samples were obtained by leaving whole blood at room temperature at 15minutes without disturbing, followed by centrifuging at 1,000-2,000 x g for 10 minutes in a refrigerated centrifuge.

#### Study Design

Figure 1 shows the design of this study.

In discovery stage, we selected 17 BART microRNAs based on our previous microarray data and sequencing data from another group.<sup>23, 24</sup> Among them, 12 of 17 were selected from our microarray data based on abundance in NPC tissue and significant fold change between tumor tissue and non-tumor tissue and had been detected in patients' sera. In order to make the profiling more comprehensive, we also included 5 more microRNAs that were most frequently detected in NPC tissue by deep sequencing. Then we quantified their levels in serum samples from 24 NPC patients and 24 non-cancer controls in Hong Kong. The BART microRNAs with AUC (Area Under Curve) greater than 0.8 in discovery stage would be tested in two cohorts consisting of participants from Hong Kong and Guangzhou respectively for further validation. To test the ability of the microRNA to identify preclinical NPC patients from high risk population, we designed the nested case-control study mentioned above.

### **RNA extraction**

Total RNA was extracted from serum samples with miRNeasy RNA isolation kit (Qiagen, Inc). Specifically, add 100ul serum to 1ml Qiazol and vortex for complete denature. Then add 5ul mixture of three *C.elegans* microRNAs (5fmol/  $\mu$  L for cel-39,cel54,cel-238) to each sample and mix well immediately. To achieve aqueous phase, add 220  $\mu$  L chloroform to each sample and vortex for 30 seconds, followed by centrifugation at 12000g at 4°C for 15minutes. Then transfer the upper layer to a new tube quickly and perform the following step as described in the product manual. Total RNA will be eluted from the column with 30ul Rnase-Free water and stored at -80°C.

### **Reverse transcription**

Reverse transcription of microRNA was performed with the Taqman miRNA Transcription Kit and microRNA-specific stem-loop primers (Part No. 4366597, ABI). The specific information of microRNA-specific stem-loop primers, including primers for reverse transcription and QPCR, were attached in Supplemental Table 2. Each reaction consisted of 1.387  $\mu$  L water, 0.5ul 10X Reverse Transcription Buffer, 0.063ul Rnase Inhibitor (20U/  $\mu$  L), 0.05ul 100mM dNTP with dTTP, 0.33ul Multiscribe Transcriptase, 1ul RT primer and 1.67  $\mu$  L total RNA. The program for RT is set as the following conditions: 16°C for 30min,42°C 30 min, 85°C 5min, keep at 4°C. Add 28.9ul to each tube to dilute the RT product.

### **Real-time PCR**

Real-time PCR was done in triplicate. Each reaction mix contains 2.5ul Taqman 2X Universal Master Mix with No AmpErase UNG, 0.25ul microRNA-specific Taqman Probe and 2.25ul diluted RT product. QPCR was performed on ABI 7900 Real-time PCR machine with the following conditions: 95°C for 5min, 40 cycles of 15 sec at 95°C and 1 min at 60°C.

### **Absolute quantification of microRNA with synthetic miRNA mimics**

To estimate the copy number of particular microRNA in serum samples, a standard curve is required by performing QPCR with serial diluted synthetic miRNA mimics. For the first dilution, adding 4ul of 1nM synthetic miRNA mimics to 118.1ul Rnase-free water to create 8.192fM solution. Perform 10 serial 4-fold dilutions by adding 20ul of previous solution containing mimics to 60  $\mu$  L Rnase-free water to get the 11<sup>th</sup> tube which contains about 10 copies of miRNA mimics. The 12<sup>th</sup> tube contains water only, serving as a negative control. Serial diluted mimics should be run along with the tested samples to generate the standard curve.

### **EBV antibody test**

The EBV antibody test was performed according to the protocol established previously.<sup>7</sup> Briefly, the samples were stored in -80°C before use. VCA-IgA (EUROIMMUN, Germany) and EBNA1-IgA (Zhongshan Bio-tech, Zhongshan) commercialized ELISA kits were used and ELISA assay was performed according to the user manual.

### **Statistical analysis**

The experimental data should be normalized with a median normalization procedure.<sup>25</sup> We used

Chi-square tests to compare clinicopathological categorical variables in different subgroups. For mean age of different groups, t-test was used. The diagnostic value of each BART microRNA was evaluated by sensitivity, specificity and the areas under curve (AUC) with 95%CI. Z-test was used to compare AUC in different groups.

## Results

### Profiling of EBV BART microRNAs in discovery stage

Comprehensive profiling of EBV microRNAs in NPC tissues has been finished by QPCR or RNA sequencing in 2009 and 2010.<sup>19, 24</sup> In 2012, our group found that 12 BART microRNAs were also present in sera of patients and the copy number of which was positively correlated with that in the NPC tissue, suggesting these microRNAs might be used as biomarkers.<sup>23</sup> Based on these discoveries, we selected 17 BART microRNAs as potential biomarkers and compared their abundance in serum between 24 NPC patients and 24 non-cancer controls in discovery stage. Among them, 12 microRNAs were selected according to our previous finding and 5 more microRNAs (BART 3-3p, BART4-5p, BART8-5p, BART10-3p and BART19-5p) were also included according to RNA sequencing result reported.<sup>24</sup> The mean age of participants was comparable between patients and non-cancer controls ( $52.0 \pm 9.4$  vs  $58.7 \pm 9.3$ ,  $p < 0.12$ ). (Table.1) To evaluate the diagnostic performance of each BART microRNA, we used ROC analysis and found that BART 2-5p (AUC=0.921, 95%CI: 0.835-1.007) was the only one that met the criterion. The AUC of other BART microRNAs were listed in Supplemental Table 3. Among them, BART 7-3p (AUC=0.727, 95%CI: 0.626-0.909) was also promising and reported by another group.<sup>20</sup>

### Characterization of BART 2-5p in serum samples from validation cohorts

As the AUC of BART 2-5p was above 0.80 in discovery stage, we proceeded to evaluate it in validation cohort 1, which was composed of 148 NPC patients and 118 non-cancer controls from Hong Kong. For patient group, the mean age was  $50.6 \pm 11.5$  years of old and the male proportion was 75.7%. The number of patients in different stages was 11( I ), 58( II ), 50( III ), 29(IV), respectively. The mean age for non-cancer controls was  $51.3 \pm 14.5$  years of old and male participants accounted for 53.4%. Serologic EBV antibody test identified 106 low risk controls, 8 medium risk controls and 4 high risk controls. The cutoff value (Ct value) was set to achieve the largest area under curve in ROC. To calculate the absolute copy number in serum, the standard curve was established:  $y = -1.2029x + 45.452$ ,  $R^2 = 0.9963$ .  $y$  stands for the  $\log_2$  value of copy number in per  $80 \mu\text{L}$  dilution,  $x$  is the Ct value in QPCR. After calculation, the corresponding copy number was set as 29 copies/10ul and the overall sensitivity and specificity was 93.9% and 89.8% respectively. The copy number distribution of BART2-5p was shown in Fig.2A. The AUC was 0.972 (95%CI, 0.954-0.989). (Table.2) To evaluate the capability of BART 2-5p to identify NPC patients at early stages, we calculated sensitivity, specificity and AUC for each stage. For early stages (Stage1&2), the sensitivity, specificity of BART 2-5p was 94.2%, 89.8%, respectively, with an AUC of 0.966 (95%CI 0.939-0.994), suggesting this biomarker was promising for early detection of NPC. The sensitivity for diagnosing late stage patients (Stage 3-4) was 93.7%, with an AUC of 0.976 (95% CI 0.959-0.994), which was not statistically significant with that of early stage patients. ( $P > 0.05$ ) (Table.2)



In order to examine if BART2-5p is applicable in patients in different areas, we further determined the diagnostic value of BART 2-5p in validation cohort 2 consisting of samples from Guangzhou which is another city in the endemic area of NPC. The mean age for patients and healthy controls in this cohort was  $46.4 \pm 10.6$  and  $49.0 \pm 8.4$  years of old respectively. The 103 NPC patients included 9 stage I patients, 44 stage II patients, 30 stage III patients, 20 stage IV patients. The copy number of BART2-5p for each individual was shown in Fig.2B. The overall sensitivity and specificity in this cohort was 94.2% and 83.5%, respectively. (Table.2) The AUC was 0.959 (95%CI: 0.939-0.980). The sensitivity for patients in early stage was 94.3%, for advanced stages, it was 94.0%. The AUC for early and late stage patients were 0.951 (95%CI: 0.926-0.977) and 0.968 (95%CI: 0.937-0.999) respectively, the difference of which was also not statistically significant. Overall, the diagnostic performance of BART 2-5p in validation cohort 1 and 2 was consistent.

#### **Evaluation of BART 2-5p in a Nested Case-control Cohort**

To evaluate the capability of BART 2-5p in identifying preclinical patients from high risk individuals, we constructed a nested case-control cohort (validation cohort3). The mean age of patients and matched controls in this cohort was  $50.1 \pm 8.4$  and  $49.5 \pm 8.3$  years of old, respectively. Male proportion was 72.7% and 76.1% in NPC and high-risk group, respectively. (Table.1) Here, 20/22 patients were tested positive at least 2 months prior to their clinical diagnosis, showing a satisfying sensitivity for preclinical patients. The two serum samples tested negative was taken more than 5 years before the diagnosis. 48 out of 88 healthy high-risk individuals were tested negative, indicating a specificity of 54.5%. It means these high-risk individuals can be exempt from unnecessary check without sacrificing the sensitivity. Fig.3A showed the copy number for each participant in this cohort. The AUC for this cohort was 0.858 (95%CI: 0.765-0.951), suggesting BART2-5p is a valuable biomarker for early detection of preclinical patients. (Table.2&Fig.3B)

#### **Comparison of BART 2-5p and EBV antibody test in early detection of NPC patients**

To compare the diagnostic ability of BART2-5p and serologic antibody test for early stage patients, all early stage samples (69 stage I & II patients in validation cohort1 and 53 stage I & II patients in validation cohort 2) were also tested by EBV antibody ELISA. For NPC diagnosis, no statistical difference was found between the two methods in both validation cohorts: the sensitivity, specificity and AUC of ELISA test were 92.8%,96.6%, 0.986(95%CI: 0.972-1.001) in validation cohort 1 and 90.6%,97.2%,0.981 ( 95%CI:0.966-0.995 ) in validation cohort 2, respectively. (Fig.4) However, EBV antibody test could not precisely identify the preclinical patients from the high-risk population because they were all tested positive. In this case, BART2-5p was able to distinguish those preclinical patients with a much higher specificity (54.5%).

#### **Comparison of BART 2-5p and EBV DNA test in early stage NPC patients**

Previously, the sensitivity of EBV DNA test in diagnosing early stage patients was 81.5% by using 0 copies/mL as the cutoff value.<sup>13</sup> To better compare the performance of BART 2-5p and EBV DNA test in early diagnosis, the EBV DNA load in circulation was also quantified in 50 early stage patients from validation cohort 1. The quantifying methods and cut-off point setting was the same as previously described.<sup>13</sup> The sensitivity for EBV DNA test was 76%,

which was consistent with the previous report and a systematic review.<sup>26</sup>

### **Evaluation of data reliability**

To evaluate data reliability, we performed stratified random sampling to obtain 76 participants from all the cohorts and retested the circulating level of BART2-5p by re-extraction of total RNA from the same serum samples. 66 cases showed consistent result (Positive or Negative) with that of previous test, with an accordance rate of 0.868 (95%CI: 0.791-0.946). If the repeated QPCR was performed with the RNA extracted previously, 74 cases had consistent results, indicating the accordance rate of 0.974 (95%CI: 0.937-1.011).

### **Discussion**

In this study, we have selected 17 EBV BART microRNAs according to previous reports and evaluated their diagnostic performance in NPC patients, especially for early stage ones. At discovery stage, we profiled them in 24 NPC patients and 24 non-cancer controls to select candidate biomarkers for further validation. The criterion for selection was that the AUC of BART microRNA was higher than 0.8. Among them, only BART2-5p met this criterion and was therefore chosen to be tested in validation stage. In this stage, we evaluated the diagnostic value of BART 2-5p in two case-control cohorts and one nested case-control cohort from three different cities in South China and paid particular attention to its capability in diagnosing early stage patients or preclinical patients. To our knowledge, this is the first systematic evaluation of the diagnostic performance of EBV BART microRNAs in the early diagnosis of NPC patients.

Here, we demonstrated that BART 2-5p was a sensitive and specific biomarker for NPC diagnosis and could help distinguish those preclinical patients in the high-risk population. Compared with a sensitivity of 76% for using EBV DNA to identify early stage patients, BART 2-5p showed a sensitivity of 94.2% for early stage patients from the same cohort. More importantly, while only half the preclinical patients (7/14) could be identified by EBV DNA at least one year before their diagnosis, the detectable rate of BART 2-5p was 90.9% (20/22), indicating that the level of BART2-5p began to rise before clinical diagnosis in most cases. Meanwhile, we found that BART2-5p was negative for about 54.5% of healthy high-risk individuals, making it possible to relieve them from unnecessary endoscopy check. Consequently, BART 2-5p might be valuable for screening preclinical patients.

Here, we noted that two preclinical NPC cases tested negative for BART2-5p in validation cohort 3 were taken blood more than 5 years prior to their clinical diagnosis, indicating that this method may not be suitable for identifying preclinical patients taken blood 5 years before their clinical diagnosis. The reason is probably that EBV activation and release of viral products usually occurs closely to the initiation of NPC, as suggested by many previous studies. In 2001, it was reported that the hazard ratios associated seropositivity of both EBV VCA IgA and EBV DNase antibody were higher for those NPC developed <5 years than that developed >5 years.<sup>27</sup> In 2007, higher positive rate of EBV VCA IgA at diagnosis than that before diagnosis was also reported.<sup>28</sup> Our previous result was consistent with these two studies.<sup>29</sup> Taken together, our method is significant to evaluate the risk of developing NPC in 5 years but not for more than 5 years.



It was also interesting to find that BART 2-5p was abundantly presented in circulation even several years before clinical diagnosis, and therefore, the functional role and the reason why this microRNA is highly abundant in circulation are worth for investigation. Yet, the evidence on these factors is quite limited at present. One of the possibilities is that the microRNA may assist EBV infected cells in escaping from killing of NK cells by degrading MICB.<sup>30</sup> We don't know if the EBV positive NPC cells may also benefit from this process.

We should point out certain limitations of the study. Firstly, the 17 BART microRNAs were selected based on our microarray data and sequencing data, and since these data reported for gene expression profile in NPC tissue and non-tumor tissue, some microRNAs which were enriched in circulation might not be included. In addition, the NPC cases in the nested case-control were limited, which might account for the low specificity of 54.5%. Finally, we noted that the performance of EBV antibody test seemed to be better than BART2-5p test in the ROC curve shown in Fig 4, although no significant statistical difference was found between the two methods. Considering that EBV antibody test costs less time and money than microRNA quantification, it is an ideal method for mass screening. However, even the mass screening was conducted in endemic area, the positive prediction value was still too low. Therefore, we were more focused on developing a novel biomarker that can increase the positive prediction value in EBV antibody-based mass screening. BART 2-5p is a valuable biomarker to facilitate mass screening as it could accurately identify the preclinical patients from the high-risk population. We hope the healthy high-risk individuals with little possibility to develop NPC may benefit from this finding.

In summary, we have profiled EBV BART microRNAs in the serum samples of NPC patients. The signatures of the early detection of circulating BART 2-5p before their clinical diagnosis suggested that BART 2-5p was a valuable biomarker for NPC early diagnosis. The circulating level of BART 2-5p went up even before their clinical diagnosis. In the near future, elucidating the pathophysiologic role of BART 2-5p in NPC development in order to find out why it was abundant in the blood of preclinical patients would be one of the promising fields of biomarker research.

### Acknowledgement

This work was supported by National Key R&D Program of China (2017YFC1309000), (2016YFC0902000) from Ministry of Science and Technology of the People's Republic of China and AoE/M-06/08 from Research Grants Council of the Hong Kong Special Administrative Region, People's Republic of China.

### References

1. Wei WI, Sham JS. Nasopharyngeal carcinoma. *Lancet* 2005;**365**: 2041-54.
2. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA: a cancer journal for clinicians* 2005;**55**: 74-108.
3. Hara W, Loo BW, Jr., Goffinet DR, Chang SD, Adler JR, Pinto HA, Fee WE, Kaplan MJ, Fischbein NJ, Le QT. Excellent local control with stereotactic radiotherapy boost after external beam radiotherapy in patients with nasopharyngeal carcinoma. *International journal of radiation oncology, biology, physics* 2008;**71**: 393-400.

4. Lee AW, Sze WM, Au JS, Leung SF, Leung TW, Chua DT, Zee BC, Law SC, Teo PM, Tung SY, Kwong DL, Lau WH. Treatment results for nasopharyngeal carcinoma in the modern era: the Hong Kong experience. *International journal of radiation oncology, biology, physics* 2005;**61**: 1107-16.
5. Henle W, Henle G. Evidence for a relation of Epstein-Barr virus to Burkitt's lymphoma and nasopharyngeal carcinoma. *Bibl Haematol* 1970: 706-13.
6. Sam CK, Brooks LA, Niedobitek G, Young LS, Prasad U, Rickinson AB. Analysis of Epstein-Barr virus infection in nasopharyngeal biopsies from a group at high risk of nasopharyngeal carcinoma. *International journal of cancer Journal internationale du cancer* 1993;**53**: 957-62.
7. Liu Y, Huang Q, Liu W, Liu Q, Jia W, Chang E, Chen F, Liu Z, Guo X, Mo H, Chen J, Rao D, et al. Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China. *International journal of cancer Journal internationale du cancer* 2012;**131**: 406-16.
8. Liu Z, Ji MF, Huang QH, Fang F, Liu Q, Jia WH, Guo X, Xie SH, Chen F, Liu Y, Mo HY, Liu WL, et al. Two Epstein-Barr virus-related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China. *American journal of epidemiology* 2013;**177**: 242-50.
9. Chan KC, Lo YM. Clinical applications of plasma Epstein-Barr virus DNA analysis and protocols for the quantitative analysis of the size of circulating Epstein-Barr virus DNA. *Methods in molecular biology* 2006;**336**: 111-21.
10. Lin JC, Wang WY, Chen KY, Wei YH, Liang WM, Jan JS, Jiang RS. Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma. *The New England journal of medicine* 2004;**350**: 2461-70.
11. Lo YM, Leung SF, Chan LY, Chan AT, Lo KW, Johnson PJ, Huang DP. Kinetics of plasma Epstein-Barr virus DNA during radiation therapy for nasopharyngeal carcinoma. *Cancer research* 2000;**60**: 2351-5.
12. Lo YM, Chan AT, Chan LY, Leung SF, Lam CW, Huang DP, Johnson PJ. Molecular prognostication of nasopharyngeal carcinoma by quantitative analysis of circulating Epstein-Barr virus DNA. *Cancer research* 2000;**60**: 6878-81.
13. Ji MF, Huang QH, Yu X, Liu Z, Li X, Zhang LF, Wang P, Xie SH, Rao HL, Fang F, Guo X, Liu Q, et al. Evaluation of plasma Epstein-Barr virus DNA load to distinguish nasopharyngeal carcinoma patients from healthy high-risk populations in Southern China. *Cancer* 2014;**120**: 1353-60.
14. Chan KC, Hung EC, Woo JK, Chan PK, Leung SF, Lai FP, Cheng AS, Yeung SW, Chan YW, Tsui TK, Kwok JS, King AD, et al. Early detection of nasopharyngeal carcinoma by plasma Epstein-Barr virus DNA analysis in a surveillance program. *Cancer* 2013;**119**: 1838-44.
15. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;**116**: 281-97.
16. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nature reviews Cancer* 2006;**6**: 259-69.
17. Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T. Identification of virus-encoded microRNAs. *Science* 2004;**304**: 734-6.
18. Cai X, Schafer A, Lu S, Bilello JP, Desrosiers RC, Edwards R, Raab-Traub N, Cullen BR. Epstein-Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PLoS pathogens* 2006;**2**: e23.

19. Cosmopoulos K, Pegtel M, Hawkins J, Moffett H, Novina C, Middeldorp J, Thorley-Lawson DA. Comprehensive profiling of Epstein-Barr virus microRNAs in nasopharyngeal carcinoma. *Journal of virology* 2009;**83**: 2357-67.
20. Zhang G, Zong J, Lin S, Verhoeven RJ, Tong S, Chen Y, Ji M, Cheng W, Tsao SW, Lung M, Pan J, Chen H. Circulating Epstein-Barr virus microRNAs miR-BART7 and miR-BART13 as biomarkers for nasopharyngeal carcinoma diagnosis and treatment. *International journal of cancer Journal international du cancer* 2015;**136**: E301-12.
21. Cai LM, Lyu XM, Luo WR, Cui XF, Ye YF, Yuan CC, Peng QX, Wu DH, Liu TF, Wang E, Marincola FM, Yao KT, et al. EBV-miR-BART7-3p promotes the EMT and metastasis of nasopharyngeal carcinoma cells by suppressing the tumor suppressor PTEN. *Oncogene* 2015;**34**: 2156-66.
22. Cai L, Ye Y, Jiang Q, Chen Y, Lyu X, Li J, Wang S, Liu T, Cai H, Yao K, Li JL, Li X. Epstein-Barr virus-encoded microRNA BART1 induces tumour metastasis by regulating PTEN-dependent pathways in nasopharyngeal carcinoma. *Nature communications* 2015;**6**: 7353.
23. Wong AM, Kong KL, Tsang JW, Kwong DL, Guan XY. Profiling of Epstein-Barr virus-encoded microRNAs in nasopharyngeal carcinoma reveals potential biomarkers and oncomirs. *Cancer* 2012;**118**: 698-710.
24. Chen SJ, Chen GH, Chen YH, Liu CY, Chang KP, Chang YS, Chen HC. Characterization of Epstein-Barr virus miRNAome in nasopharyngeal carcinoma by deep sequencing. *PloS one* 2010;**5**.
25. 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.
26. Han BL, Xu XY, Zhang CZ, Wu JJ, Han CF, Wang H, Wang X, Wang GS, Yang SJ, Xie Y. Systematic review on Epstein-Barr virus (EBV) DNA in diagnosis of nasopharyngeal carcinoma in Asian populations. *Asian Pacific journal of cancer prevention : APJCP* 2012;**13**: 2577-81.
27. Chien YC, Chen JY, Liu MY, Yang HI, Hsu MM, Chen CJ, Yang CS. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *The New England journal of medicine* 2001;**345**: 1877-82.
28. Ji MF, Wang DK, Yu YL, Guo YQ, Liang JS, Cheng WM, Zong YS, Chan KH, Ng SP, Wei WI, Chua DT, Sham JS, et al. Sustained elevation of Epstein-Barr virus antibody levels preceding clinical onset of nasopharyngeal carcinoma. *British journal of cancer* 2007;**96**: 623-30.
29. Cao SM, Liu Z, Jia WH, Huang QH, Liu Q, Guo X, Huang TB, Ye W, Hong MH. Fluctuations of Epstein-Barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PloS one* 2011;**6**: e19100.
30. Nachmani D, Stern-Ginossar N, Sarid R, Mandelboim O. Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. *Cell host & microbe* 2009;**5**: 376-85.

#### Figure legends:

**Figure.1 Overview of the study design.** This is a two stages study. Profiling of candidate microRNAs was performed in discovery stage. MicroRNAs with AUC>0.8 would be examined in validation stage, which consisted of two case-control cohorts and one nested case-control cohort.

**Figure.2 The copy number distribution of BART 2-5p for each individual in Validation Cohort 1(A) and 2(B).** (A) Control: 118 non-cancer controls from QMH in Hong Kong. NPC: 148 NPC

patients from QMH in Hong Kong. (B) Control: 284 healthy controls from Cancer Center of SYSU. NPC: 103 NPC patients from Cancer Center of SYSU. Y axis was shown in  $\text{Log}_{10}$  scale.

**Figure.3 The diagnostic performance of BART 2-5p in Validation Cohort 3. (Nested case control cohort)** (A) Circulating level of BART2-5p for each individual in this cohort. Y axis was shown in  $\text{Log}_{10}$  scale. (B) ROC analysis showed the capability of BART 2-5p to discriminate preclinical patients from high risk population. (AUC: 0.858, 95%CI: 0.765-0.951, sensitivity 90.9%, specificity 54.5%).

**Figure.4 Comparison of EBV antibody and BART 2-5p in detecting early stage NPC ROC analysis from validation cohort 1(A) and validation cohort 2(B).** This figure indicated that BART2-5p had a comparable diagnostic ability with EBV antibody test in early stage patients in the two validation cohorts. 69 stage I & II patients and 118 non-cancer controls in validation cohort1, 53 stage I & II patients and 284 healthy controls in validation cohort 2 were analyzed. Blue line for BART 2-5p test, green line for EBV antibody test. The AUC,95%CI of AUC, sensitivity, specificity of BART 2-5p in validation cohort1 was 0.966,0.939-0.994,94.2%,89.8%, respectively, compared with 0.986,0.972-1.001, 92.8%,96.6% for EBV antibody test. In validation cohort2, the AUC,95%CI of AUC, sensitivity, specificity of BART 2-5p was 0.951, 0.926-0.977, 94.3%,83.5%, respectively, compared with 0.981, 0.966-0.995, 90.6%,97.2% for EBV antibody test.

## Clinicopathological Characteristics of Individuals by Subgroup

	Discovery Stage (n=48)	Validation Cohort 1 (n=266)	Validation Cohort 2 (n=387)	Validation Cohort 3 (n=110)	p Value*
<b>Patients with NPC</b>					
N	24	148	103	22	-
Age (mean±SD)	52.0±9.4	50.6±11.5	46.4±10.6	50.1±8.4	0.004
Men	12 (50.0%)	112 (75.7%)	78 (75.7%)	16 (72.7%)	0.992
City	Hong Kong	Hong Kong	Guangzhou	Sihui -	
<b>Clinical Stage</b>					
I	3	11	9		0.866
II	7	58	44		
III	8	50	30		
IV	6	29	20		
<b>Non-cancer Controls</b>					
N	24	118			
Age (mean±SD)	58.7±9.3	51.3±14.5			0.120
Men	8 (33.3%)	63 (53.4%)			<0.01
City	Hong Kong	Hong Kong			
<b>Risk Assessment</b>					
Low Risk	21	106			
Medium Risk	3	8			
High Risk	0	4			
<b>Healthy Controls</b>					

Accepted Article

N	284	88
Age (mean±SD)	49.0±8.4	49.5±8.3
Men	227 (79.9%)	67(76.1%)
City	Guangzhou	Sihui
Risk Assessment		
Low Risk	183	0
Medium Risk	93	0
High Risk	8	88

---

\*p values for the comparisons between validation cohort1 and validation cohort2



### Characterization of BART 2-5p in Serum Samples from Validation Cohorts

	Sensitivity (%)	Specificity (%)	AUC (95%CI)	p Value
<b>Cohort 1</b>				
<b>Clinical Stage</b>				
I	90.9(10/11)	89.8(106/118)	0.901(0.768-1.000)	<0.001
II	94.8(55/58)	89.8(106/118)	0.979(0.961-0.996)	<0.001
III	92.0(46/50)	89.8(106/118)	0.971(0.946-0.995)	<0.001
IV	96.6(28/29)	89.8(106/118)	0.986(0.971-1.000)	<0.001
Early Stage( I + II )	94.2(65/69)	89.8(106/118)	0.966(0.939-0.994)	<0.001
Advanced Stage(III+IV)	93.7(74/79)	89.8(106/118)	0.976(0.959-0.994)	<0.001
All Stages	93.9(139/148)	89.8(106/118)	0.972(0.954-0.989)	<0.001
<b>Cohort 2</b>				
<b>Clinical Stage</b>				
I	88.9(8/9)	83.5(237/284)	0.901(0.830-0.973)	<0.001
II	95.5(42/44)	83.5(237/284)	0.961(0.937-0.985)	<0.001
III	93.3(28/30)	83.5(237/284)	0.974(0.952-0.995)	<0.001
IV	95.0(19/20)	83.5(237/284)	0.959(0.889-1.000)	<0.001
Early Stage( I + II )	94.3(50/53)	83.5(237/284)	0.951(0.926-0.977)	<0.001
Advanced Stage(III+IV)	94.0(47/50)	83.5(237/284)	0.968(0.937-0.999)	<0.001
All Stages	94.2 (97/103)	83.5(237/284)	0.959(0.939-0.980)	<0.001
<b>Cohort 3</b>				
	90.9 (20/22)	54.5(48/88)	0.858(0.765-0.951)	<0.001

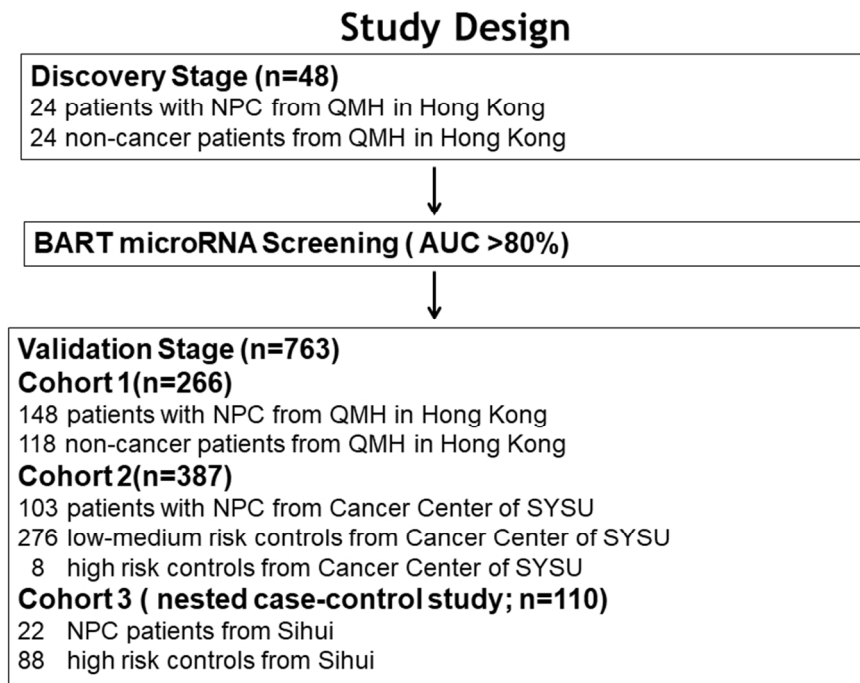
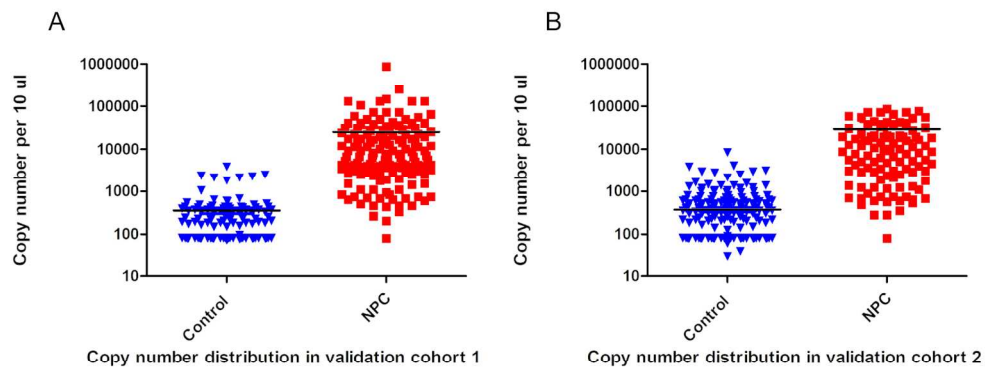


Figure.1 Overview of the study design. This is a two stages study. Profiling of candidate microRNAs was performed in discovery stage. MicroRNAs with AUC>0.8 would be examined in validation stage, which consisted of two case-control cohorts and one nested case-control cohort.

254x190mm (96 x 96 DPI)

Accep

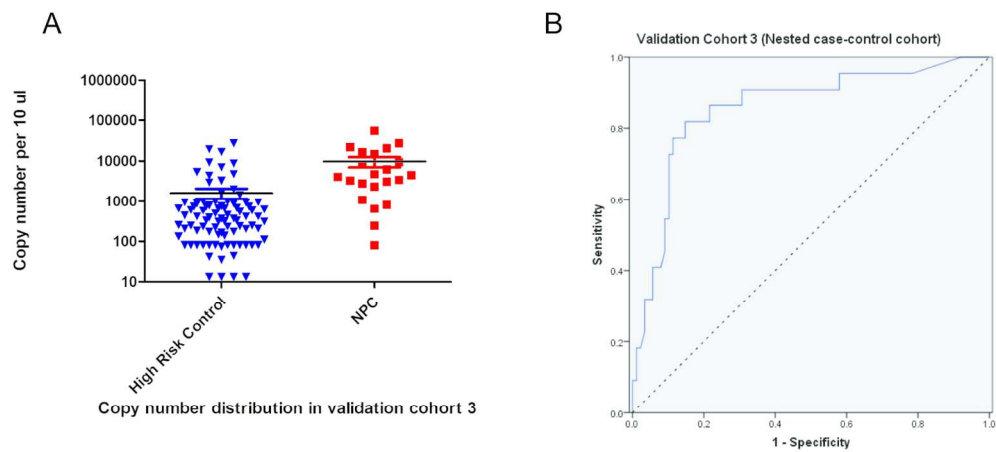


Jiang et.al Figure.2

Figure.2 The copy number distribution of BART 2-5p for each individual in Validation Cohort 1(A) and 2(B). (A) Control: 118 non-cancer controls from QMH in Hong Kong. NPC: 148 NPC patients from QMH in Hong Kong. (B) Control: 284 healthy controls from Cancer Center of SYSU. NPC: 103 NPC patients from Cancer Center of SYSU. Y axis was shown in Log<sub>10</sub> scale.

152x69mm (300 x 300 DPI)

Accepted

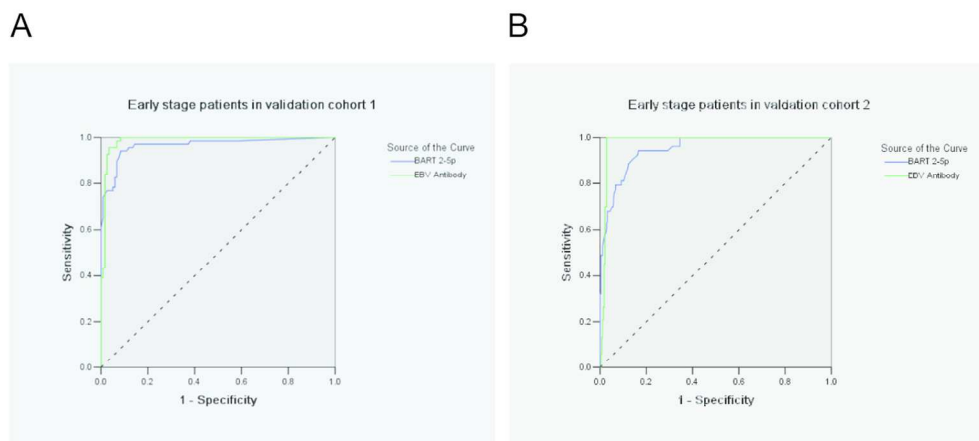


Jiang et.al Figure.3

Figure.3 The diagnostic performance of BART 2-5p in Validation Cohort 3. (Nested case control cohort) (A) Circulating level of BART2-5p for each individual in this cohort. Y axis was shown in Log10 scale. (B) ROC analysis showed the capability of BART 2-5p to discriminate preclinical patients from high risk population. (AUC: 0.858, 95%CI: 0.765-0.951, sensitivity 90.9%, specificity 54.5%).

133x69mm (300 x 300 DPI)

Accepted



Jiang et.al Figure.4

Figure.4 Comparison of EBV antibody and BART 2-5p in detecting early stage NPC ROC analysis from validation cohort 1(A) and validation cohort 2(B). This figure indicated that BART2-5p had a comparable diagnostic ability with EBV antibody test in early stage patients in the two validation cohorts. 69 stage I & II patients and 118 non-cancer controls in validation cohort1, 53 stage I & II patients and 284 healthy controls in validation cohort 2 were analyzed. Blue line for BART 2-5p test, green line for EBV antibody test. The AUC,95%CI of AUC, sensitivity, specificity of BART 2-5p in validation cohort1 was 0.966,0.939-0.994,94.2%,89.8%, respectively, compared with 0.986,0.972-1.001, 92.8%,96.6% for EBV antibody test. In validation cohort2, the AUC,95%CI of AUC, sensitivity, specificity of BART 2-5p was 0.951, 0.926-0.977, 94.3%,83.5%, respectively, compared with 0.981, 0.966-0.995, 90.6%,97.2% for EBV antibody test.

118x65mm (300 x 300 DPI)

Accept