



Published in final edited form as:

Clin Cancer Res. 2018 March 15; 24(6): 1305–1314. doi:10.1158/1078-0432.CCR-17-1929.

Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan

Anna E. Coghill¹, Ruth M. Pfeiffer¹, Carla Proietti^{2,3}, Wan-Lun Hsu^{4,5}, Yin-Chu Chien^{4,6}, Lea Lekieffre², Lutz Krause², Andy Teng⁷, Jocelyn Pablo⁷, Kelly J Yu, PhD¹, Pei-Jen Lou, MD⁸, Cheng-Ping Wang⁸, Zhiwei Liu¹, Chien-Jen Chen^{4,5}, Jaap Middeldorp⁹, Jason Mulvenna², Jeff Bethony¹⁰, Allan Hildesheim^{1,*}, and Denise L. Doolan^{2,3,*}

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

²Queensland Institute of Medical Research, Brisbane, Australia ³Australian Institute of Tropical Health & Medicine, James Cook University, Cairns, Australia ⁴Genomics Research Center, Academia Sinica, Taipei, Taiwan ⁵Graduate Institute of Epidemiology and Prevention Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan ⁶National Institute of Cancer Research, National Health Research Institute, Miaoli, Taiwan ⁷Antigen Discovery Inc., Irvine, California, USA ⁸Department of Otolaryngology, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan ⁹Vrije University Medical Center, Amsterdam, Netherlands

¹⁰Department of Microbiology, Immunology, and Tropical Medicine, George Washington University Medical Center, Washington DC

Abstract

Background—Epstein-Barr virus (EBV) is necessary for the development of nasopharyngeal carcinoma (NPC). By adulthood, ~90% of individuals test EBV-positive, but only a fraction develop cancer. Factors that identify which individuals are most likely to develop disease, including differential antibody response to the virus, could facilitate detection at early stages when treatment is most effective.

Methods—We measured anti-EBV IgG and IgA antibody responses in 607 Taiwanese individuals. Antibodies were measured using a custom protein microarray targeting 199 sequences from 86 EBV proteins. Variation in response patterns between NPC cases and controls was used to develop an antibody-based risk score for predicting NPC. The overall accuracy (area under the curve[AUC]) of this risk score, and its performance relative to currently-used biomarkers, was evaluated in two independent Taiwanese cohorts.

Findings—Levels of 60 IgA and 73 IgG anti-EBV antibodies differed between Stage I/IIa NPC cases and controls ($P < 0.0002$). Risk prediction analyses identified antibody targets that best discriminated NPC status—BXL1F1, LF2, BZLF1, BRLF1, EA_d, BGLF2, BPLF1, BFRF1, and

Corresponding Author: Anna E. Coghill, PhD, Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20892, +0012402767184, anna.coghill@nih.gov.

*authors contributed equally to this work

Declaration of Interests: No authors report any conflicts of interest.

BORF1. When combined with currently-used VCA/EBNA1 IgA biomarkers, the resulting risk score predicted NPC with 93% accuracy (95%CI 87–98%) in the general Taiwanese population, a significant improvement beyond current biomarkers alone (82%;95%CI 75–90%,*P* 0.01). This EBV-based risk score also improved NPC prediction in genetically high-risk families (89%;95%CI 82–96%) compared to current biomarkers (78%;95%CI 66–90%,*P* 0.03).

Interpretation—We identified NPC-related differences in 133 anti-EBV antibodies and developed a risk score using this microarray dataset that targeted immune responses against EBV proteins from all stages of the viral life cycle, significantly improving the ability to predict NPC.

Keywords

EBV; NPC; EBV antibodies; cancer prediction; protein microarray

INTRODUCTION

Epstein-Barr virus (EBV) is necessary for the development of nasopharyngeal carcinoma (NPC).(1) By adulthood, approximately 90% of individuals globally are EBV-infected.(2–4) The virus establishes lifelong latency in memory B-cells, periodically expressing lytic proteins necessary for replication in plasma B-cells and epithelial cells.(5, 6) Despite the fact that EBV is a ubiquitous pathogen, EBV-associated cancers develop in only a small fraction of infected individuals. Identification of differences present in the subset of EBV-positive individuals who develop cancer, including differences in the immune response to the virus, could guide biomarker development. Biomarkers to aid early detection are particularly important for NPC since clinical symptoms associated with the disease are non-specific, and no known pre-cursors exist. This often results in diagnosis being delayed to advanced stages of disease, when five-year survival is less than 50%, as opposed to identifying NPC at early stages when five-year disease-free survival can reach 90%.(7)

Evidence suggests that antibodies against EBV may be well suited for NPC early detection. Research focusing on select antigens has consistently documented higher antibody titers against viral capsid antigen (VCA) and early antigen (EA) in NPC patients.(8–10) Prospective data from Southeast Asia, where NPC is 50- to 100-fold more common than in the US/Europe, indicates that IgA antibodies against defined VCA sequences (VCAp18) and the EBV nuclear antigen 1 protein (EBNA1) are elevated years prior to NPC diagnosis.(11–14) Accordingly, anti-VCA and EBNA1 IgA antibodies are currently being evaluated in Southeast China as biomarkers to identify disease-free individuals harboring the highest risk for developing NPC.(15)

However, these select antibodies represent a very small fraction of the immune response against EBV's nearly 100 proteins.(16) Whether antibodies targeting additional EBV proteins can improve the stratification of individuals according to cancer risk is unknown. To address this question, we applied a custom protein microarray to measure both IgG and IgA antibody responses against a comprehensive set of 199 EBV protein sequences within three independent studies from Taiwan, one of the highest-incidence regions for NPC globally.

METHODS

We initially probed blood collected from 175 histologically confirmed NPC cases diagnosed in Taiwan between 1991–1994 (67 Stage I/IIa, 16 Stage IIb, 88 Stage III/IV, 4 unknown) and 175 community controls recruited during the same time period, frequency matched to cases on age, sex, and region. (17) Tumor stage was defined as: stage I/IIa (T1/T2, N0, M0); stage IIb (T1/2, N1, M0); stage III (T3, N0-2, M0 or T1/2, N2, M0); and stage IV (T4, N0-2, M0; TN, N3, M0; or TN, MN, M1). Blood was drawn at the time of NPC diagnosis but prior to therapy for cases and at study enrollment for controls.

Studies nested within prospective cohorts

The Cancer Screening Project (CSP) is comprised of 23,943 Taiwanese residents who participated in a population-based screening project between 1991–1992. (18) Blood was drawn at cohort enrollment (baseline). We probed blood collected from 37 individuals in the cohort who were disease-free at the time of blood draw but developed NPC during follow-up (incident cases), as ascertained by linkage with the Taiwanese National Cancer Registry through October 2002. We also probed blood collected from 117 individuals who remained disease-free during follow-up, matched to incident cases on age, sex, township, and cohort enrollment/blood collection date.

The Taiwan Family Study (TFS) is comprised of 2,557 individuals recruited between 1996–2005 from 358 NPC multiplex families (2 first- or second-degree family members affected by NPC). (19) Blood was drawn at cohort enrollment. We probed blood collected from 26 individuals in the cohort who were disease-free at the time of blood draw but developed NPC during follow-up (incident cases), as ascertained by clinical evaluation and linkage with the Taiwanese National Cancer Registry through December 2014. We also probed blood collected from 77 individuals who remained disease-free during follow-up, frequency matched to incident cases on age and sex.

EBV protein microarray testing and quality control

Predicted protein sequences were generated using five EBV strains (AG876, Akata, B95-8, Mutu and Raji), representing both African (Mutu, Raji and AG876) and Asian (Akata) variants. Annotated genomic coordinates in GenBank for each strain (<http://www.ncbi.nlm.nih.gov/>) were supplemented with annotations from identified splice variants in the literature to generate a list of predicted protein sequences. 632 predicted sequences from 86 EBV proteins were identified across the five strains, of which 199 represented non-redundant open reading frames (ORFs). (20–22) High coverage was achieved across the five EBV strains, with 97% of the predicted sequences from each strain represented on the microarray at 99% homology. Comparison of the sequence identity of Chinese EBV strains (i.e., strains likely present in our Taiwanese study populations) to sequences printed on the array at the protein level revealed a mean percent identify of 99% (Supplemental Figure 1).

After PCR amplification of the 199 non-redundant ORFs from purified EBV cell line DNA, each sequence was cloned into a linearized, proprietary T7 expression vector (Antigen Discovery Inc., Irvine, CA), expressed using the *E. coli* cell-free protein system, and printed

onto nitrocellulose slides (i.e., microarrays). Each participant was tested on a single microarray, with 16 arrays per slide/batch. Personnel were blinded to case status, and within each slide of 16 microarrays, we tested a proportional number of cases and controls to minimize batch effects. After testing, air-dried slides were scanned on an Axon GenePix 4300A (Molecular Devices, CA); the raw fluorescence intensities were corrected for spot-specific background using the Axon GenePix Pro; and the corrected data were transformed using variance stabilizing normalization (VSN) in R (www.bioconductor.org).

Four “no DNA” (no translated protein) spots were printed onto the microarray as a participant-specific measure of non-EBV background reactivity (e.g., *E.coli* reactivity). Case-control differences in the mean antibody response were compared using standardized signal intensity, a continuous measure defined as the output for each antibody response divided by the participant-specific background (mean +1.5 standard deviation[SD] of the four “no DNA” spots). Output was further categorized into positive and negative responses, with a positive (i.e., elevated) response defined as ≥ 1.0 .

Comparison to enzyme-linked immunosorbent assay (ELISA) standards

To increase confidence in the first application of this multiplex technology to EBV, we compared IgA data generated from the array to IgA data generated from ELISAs. Specifically, we previously measured IgA antibodies against synthetic VCAp18, EBNA1, and EAdp47 peptides in a subset of participants (N=207) using an ELISA; we compared that data to the IgA antibody output for those participants against these same synthetic peptides printed onto the microarray.(11, 23) Correlations between the microarray IgA output and previously generated ELISA data for IgA antibodies against VCAp18 and EBNA1 were strong and significant (Spearman=0.76 and 0.79, respectively, $P<0.01$). The intra-assay correlation was significant but moderate for EAdp47 (Spearman=0.53). However, the majority (68.6%) of individuals tested negative for anti-EAdp47 IgA; among positive individuals, we observed strong intra-assay correlation (Spearman=0.71; $P<0.01$). Additional details can be found in Supplemental Methods.

We also assessed the reproducibility of this multiplex assay by testing blinded duplicates from 80 participants (50 cross-sectional, 18 CSP, 12 TFS) and using SAS PROC GLM to calculate the coefficient of variation (CV) and the intra-class correlation coefficient (ICC), the latter being a measure of the percent of overall variation due to inter-participant differences rather than assay noise. CVs describing inter-assay variability were $<20\%$ for 199 IgA and 197 IgG markers; average IgA and IgG CVs were 9.7% and 12.7%, respectively (Supplemental Figure 2).

Analyzing cross-sectional differences in the EBV antibody repertoire

In the 50 duplicate cross-sectional samples, 186 of 199 IgA antibodies (93%) and 166 of 199 IgG antibodies (83%) met a ICC ≥ 0.70 threshold (i.e., 70% of antibody variability due to inter-participant differences rather than assay noise) and were carried forward to examine NPC-related differences.

We examined the average level of response targeting groups of proteins defined by EBV life cycle stage (e.g., immediate early, Supplemental Table 1) between 175 community controls

and 175 NPC cases. The mean response level for each individual antibody was further compared between controls and 67 Stage I/IIa (early-stage) NPC cases using unpaired t-tests. For antibodies with early-stage NPC differences meeting a stringent Bonferroni-corrected p-value ($P < 0.0002$), we calculated a delta statistic. The delta statistic normalizes disease-related differences in mean antibody response to account for the distribution of the antibody in the study population and can be used to generate a receiver operative curve (ROC) describing a potential biomarker's predictive ability (e.g., sensitivity and specificity). (24) In our previous research, EBNA1 IgA achieved a delta statistic of ~ 1.0 , (11) so antibodies with deltas $< 70\%$ of this benchmark were not included in further NPC prediction analyses.

Identifying a risk stratification signature

Only antibodies that met the following criteria were carried forward for risk prediction analyses: ICC > 0.70 , $P < 0.0002$, and delta-statistic > 0.70 . Among these potential risk stratification markers, we evaluated which antibodies demonstrated the greatest ability to accurately discriminate between 67 Stage I/IIa NPC cases and 175 community controls by mutually entering them into a stepwise logistic regression model both continuously and as categorical variables (quartiles of the standardized output in controls). We required $P < 0.15$ as the model entry criterion and $P < 0.05$ for an antibody to remain in the model (SAS PROC LOGISTIC). This first approach identified 14 antibodies. In parallel, the continuous output was analyzed using the R statistical package randomForest, which used a recursive-partitioning algorithm to build and average findings across multiple decision trees from randomly sampled subsets of the data. We listed the 14 antibodies with the highest values for MeanDecreaseGini and MeanDecreaseAccuracy, two randomForest prediction metrics. Finally, we considered these two randomForest metrics alongside stepwise logistic regression output and included in our risk stratification signature the 12 antibodies that overlapped as the most promising markers for at least 2 of the 3 metrics.

Independently validating the risk stratification signature in two prospective cohorts

The final risk stratification signature was comprised of the 12 array-identified antibodies described above, in addition to currently-studied VCAp18/EBNA1 IgA biomarkers (14 antibodies in total). Our objective was to determine whether disease-free individuals with elevated levels of these 14 antibodies at baseline were those most likely to develop NPC during follow-up.

The EBV-based NPC risk score for a given individual was calculated as the sum of the product of each antibody's level (quartile level as calculated among controls specific to each study population) multiplied by that antibody's corresponding log-odds ratio from a regression model including each of the potential risk stratifiers. We evaluated the predictive accuracy of this EBV-based risk score in each study population using the area under the curve (AUC), a measure of how well this score classified individuals based on who did or did not develop NPC during follow-up. To obtain unbiased estimates of the AUC, we used leave-one-out cross-validation. In addition to calculating the AUC of our 14-antibody risk score, we considered the accuracy of a model including only the currently-used VCAp18/EBNA1 IgA biomarkers.

Finally, we estimated the specificity (1-false positivity rate in disease-free individuals) corresponding to NPC detection rates of 75–85% (75–85% sensitivity) for each risk model. This information, along with estimated NPC incidence rates in each study population, was used to calculate the number of individuals that tested positive per NPC detected using the formula below.(11)

$$1/\left\{\frac{\text{Sensitivity} \times \text{NPC incidence rate}}{[(\text{Sensitivity} \times \text{NPC incidence rate}) + ((1 - \text{Specificity}) \times (1 - \text{NPC incidence rate}))]}\right\}$$

RESULTS

We measured IgA and IgG antibody responses against 199 EBV protein sequences in a cross-sectional study of NPC patients and community controls, as well as two independent studies nested within prospective Taiwanese cohorts (Table 1). Within the general population cohort (Cancer Screening Program[CSP]), the average time between baseline blood draw (i.e., antibody measurement) and NPC diagnosis for incident cases was 4.2 years (SD=2.2), with nearly three-fifths (59.4%) diagnosed within five years of baseline. The average lag time in families with a high underlying genetic risk of NPC (Taiwan Family Study[TFS]) was 5.8 years (SD=3.7), with two-fifths (42.3%) diagnosed within five years.

Cross-sectional differences in the EBV antibody repertoire

Levels of 119 IgA and 139 IgG antibodies were different ($P<0.05$) between NPC cases and community controls; this included significant case-control variation in the IgA and IgG responses to groups of proteins representing all stages of the EBV life cycle ($P<0.001$; Figure 1). This pattern of differential response persisted in those diagnosed with early-stage NPC; after Bonferroni correction, 133 anti-EBV antibody levels (60 IgA and 73 IgG) were elevated in Stage I/IIa NPC patients ($P<0.002$; Figure 2 and Supplemental Table 2). The strongest early-stage NPC differences included IgA responses against two predicted BMRF1 (early antigen[EA]) sequences, LF2, and BXL1 (thymidine kinase[TK]), as well as IgG responses against the same LF2 and BXL1(TK) sequences, in addition to BRLF1(Rta), BZLF1(Zta), BGLF2, and BFRF1. Table 2 summarizes output for the ten IgA and IgG antibodies with the highest delta statistic, a metric that highlights a potential biomarker's ability to discriminate between cases and controls. Notably, five of these top ten responses overlapped for IgA and IgG–BXL1,LF2,BMRF1, BRLF1, and BALF2-p138(ssDNA binding protein).

Among the 199 anti-EBV antibodies evaluated, only 12 IgA and 11 IgG markers displayed differences ($P<0.05$) when comparing late-stage (Stage III/IV) to early-stage NPC (Supplemental Figure 3), although none of these stage-related differences reached the Bonferroni correction threshold of $P<0.0002$. The most pronounced elevations in antibody positivity for late-stage compared to early-stage disease were observed for BALF2-p138 ssDNA binding protein (IgA OR=9.3;95%CI 4.9–17; IgG OR=19;95%CI 9.4–39) and BRF2 IgG (OR=12;95%CI 5.1,27).

Identifying a NPC risk stratification signature

A total of 103 antibodies (45 IgA and 58 IgG) evaluated in cross-sectional data met all analytical filtering criteria (see Methods). Within this subset, prediction analyses identified the 12 most promising antibody targets from the microarray (Table 3)—BXL1F1 (IgG and IgA), LF2 (IgG and IgA), BRLF1 IgA, BZLF1 IgG, BGLF2 IgG, BPLF1 IgA, BFRF1 IgG, BORF1 IgG, and 2 distinct BMRF1 IgA sequences.

We first assessed whether each of these 12 array-identified antibodies was differentially present at baseline between disease-free individuals who either did or did not develop NPC during prospective follow-up. When considered individually, each of the 12 antibodies was elevated ($P < 0.01$) at baseline in persons who developed NPC during follow-up in the general population (CSP) cohort. All except BORF1 IgG were elevated in patients diagnosed with NPC within 5 years of blood draw, and all but BFRF1 IgG were also elevated in patients diagnosed 5 years after blood draw. In the genetically high-risk (TFS) cohort, we observed no NPC-related IgA differences, but IgG antibodies against five EBV proteins were elevated at baseline ($P < 0.05$) in those who developed NPC during follow-up—BXL1F1, LF2, BGLF2, BPLF1, and BORF1 (Supplemental Tables 3 and 4).

Independently validating the risk stratification signature in the general population (CSP cohort)

The final risk stratification signature was comprised of the 12 array-identified antibodies, in addition to the currently-studied VCAp18/EBNA1 IgA biomarkers (14 antibodies in total). We calculated a risk score for disease-free individuals based on their level for each of the 14 antibodies to determine whether those with elevated responses were more likely to develop NPC.

Including the 12 array-identified antibodies in the risk signature improved the NPC prediction accuracy of currently-used VCAp18/EBNA1 IgA biomarkers. The 14-antibody risk score measured at baseline predicted NPC in disease-free persons with 93% accuracy (AUC=92.6%; 95%CI 87.0–98.1%), a significant improvement ($P < 0.01$) compared to VCAp18/EBNA1 IgA alone (AUC=82.3%; 95%CI 74.9–89.7%, Figure 3a). In exploratory analyses, marker selection identified four antibodies that contributed the most information to the signature: EBNA1 IgA, VCAp18 IgA, BZLF1 IgG, and BRLF1 IgA. Compared to the VCAp18/EBNA1 IgA biomarkers alone, this four-antibody subset improved NPC prediction (AUC=89.9%; 95%CI 83.2–96.7%, $P = 0.03$), a finding that warrants replication.

To illustrate the potential clinical impact of applying our 14-antibody risk stratification signature in this cohort, we estimated the specificity corresponding to points on the ROC curve in Figure 3a that achieved 75–85% sensitivity for NPC detection. For that sensitivity range, our risk score achieved specificity ranging from 61–83%, compared to 60–67% for VCAp18/EBNA1 IgA biomarkers alone. This improvement in specificity translated into potential reductions in the number of test-positive persons per detected NPC of 5–49%; specifically, the estimated number of individuals testing positive per detected NPC was 268–548 for our 14-antibody risk signature compared to 527–575 for VCAp18/EBNA1 IgA.

Independently validating the risk stratification signature in high-risk families (TFS cohort)

Inclusion of the 12 array-identified antibodies also improved NPC prediction among high-risk family members ($P=0.03$). Whereas VCAp18/EBNA1 IgA biomarkers predicted NPC in disease-free family members with 78% accuracy (AUC=78.1%; 95% CI 66.2–89.9%), the 14-antibody risk score achieved a ~10% improvement (AUC=88.7%; 95% CI 82.0–95.5%, Figure 3b). Exploratory marker selection identified two antibodies contributing the most information to the signature in high-risk families: EBNA1 IgA and LF2 IgG. This two-antibody subset alone, however, did not improve upon the accuracy of VCAp18/EBNA1 IgA (AUC=83.0%; 95% CI 74.6–91.4, $P=0.29$).

In the 75–85% sensitivity range for NPC detection in these genetically-susceptible family members, our 14-antibody risk signature and VCA/EBNA1 IgA biomarkers translated into similar numbers of test-positive individuals per detected NPC. However, at slightly higher sensitivities on the ROC curve in Figure 3b (89–92%), the improved specificity achieved by our 14-antibody risk score resulted in a 24% reduction in the number of test-positive individuals per detected NPC (277 vs. 210).

DISCUSSION

We report the discovery and validation of an EBV antibody-based signature which accurately stratifies individuals according to their risk of developing NPC. Our effort targeted immune responses to 199 distinct peptide sequences, representing the first evaluation of antibodies against the full spectrum of EBV proteins for cancer biomarker discovery. We report NPC case-control differences in 60 IgA and 73 IgG EBV-directed antibodies. A parsimonious subset of targets—TK, LF2, Zta, Rta, EAd, BGLF2, BPLF1, BFRF1, and BORF1 – significantly improved upon the NPC prediction ability of current VCAp18/EBNA1 IgA biomarkers. When applied in a prospective setting in Taiwan, our antibody-based risk score measured at baseline accurately identified 93% of disease-free persons in the general population and 89% of high-risk family members who developed NPC during follow-up.

The protein microarray technology applied here has been used to describe the humoral immune response to multiple pathogens,(21, 22) but this represents to our knowledge the first application of this multiplex technology to explore EBV-related immune responses and the first application for cancer biomarker discovery. Although previous EBV-related disease studies demonstrated aberrant antibody patterns in patients, including elevations in VCA and EBNA IgA prior to NPC, these studies generally utilized simplex assays to examine a small number of EBV targets.(13, 14, 25, 26) The 133 case-control antibody differences we report here represent a marked increase beyond previously reported NPC associations and could not have been efficiently evaluated using simplex assays. We believe this work can serve as a road map for future infection and cancer investigations, particularly for research into how inter-individual differences in the immune response to viruses may impact disease risk in regions most affected by infection-related cancers.

Importantly, the selection of a parsimonious subset of antibodies that successfully discriminated between Stage I/IIa NPC cases and controls was extended to two prospective

cohorts, validating that our antibody-based signature could risk stratify individuals based on their likelihood of future disease. Although both our 14-antibody risk score and the currently-studied VCAp18/EBNA1 IgA biomarkers had the potential to achieve high sensitivity for identifying NPC, our signature achieved higher overall accuracy (AUC) in both cohorts and higher specificity for the defined sensitivity range in the general population CSP cohort. Improvement in this important test characteristic (specificity) could practically translate into a lower false positivity rate and reductions in the number of individuals undergoing unnecessary, potentially invasive clinical follow-up.

We observed more moderate prediction accuracy for our antibody-based risk score in high-risk families (AUC=89%) compared to the general population (AUC=93%), despite improvement in the AUC compared to VCAp18/EBNA1 IgA alone. This may have been due, in part, to the narrower range of array-based IgA antibody responses observed in multiplex family members. Although the assay performed well across study populations (low CVs), the smaller range of IgA response in family members could have translated into a greater proportion of IgA variation being attributable to assay noise (Supplemental Figure 2). Nonetheless, because high-risk family members in TFS have a 10-fold higher underlying NPC risk compared to the general population, when considering a similar case detection rate (e.g., 75–85% sensitivity on the ROC curve), the number of individuals needing to be screened to detect a case (83–169) was considerably lower than in CSP (268–575), making application of the signature more efficient in practice despite lower absolute accuracy.

The 12 array-identified antibodies included in our NPC risk score were directed against a set of nine proteins disproportionately representing early actors in the EBV life cycle. The BGLF2 viral tegument protein promotes EBV reactivation through regulation of Zta (BZLF1),(27) whereas the LF2 protein influences Rta (BRLF1) activity.(28, 29) Both Rta and Zta are crucial, immediate early proteins directly responsible for EBV latent to lytic cycle switching.(30) In addition, the score included antibody responses against three early antigen complex proteins, two peptide variations of the BMRF1 DNA polymerase and BXLF1 (TK).(31, 32) In addition to these early-acting proteins, our risk stratification signature included antibody responses against BPLF1, a deubiquitinating tegument protein hypothesized to contribute to virion production and innate immune evasion,(33, 34) and two late-acting proteins involved in viral egress (BFRF1) and capsid structure (BORF1).(35–37) Expression of five of our nine targeted EBV proteins—LF2, BZLF1,BRLF1,BGLF2, and BMRF1 was also recently detected as part of the lytic EBV transcriptome in NPC biopsies and the NPC C666-1 cell line.(38) Considered alongside a report of increased (abortive) lytic EBV gene expression in nasopharyngeal brushings of NPC patients,(39) our findings point to lytic proteins as a promising focus for EBV-related cancer biomarker research and suggest that current IgA biomarkers may be limited by exclusive targeting of latent (EBNA1) or late-acting (VCA) proteins.

Another notable finding was the importance of not only IgA but also IgG anti-EBV antibodies for NPC risk stratification. Previous NPC studies have focused primarily on the association with IgA, a shorter-lived antibody produced in response to recent pathogen exposure at mucosal surfaces. IgG is a more abundant, systemic antibody reflecting longer-term pathogen exposure. Although the global ubiquity of EBV translates into nearly uniform

IgG positivity against certain proteins (e.g., EBNA1 and VCA), limiting their utility for screening, our comprehensive evaluation revealed several IgG antibodies reflecting longer-term exposure to lytic viral activity that were differentially present in those EBV-positive adults more likely to develop NPC.

Encouragingly, IgA responses to three synthetic EBV peptides were similar when measured using this microarray and a simplex ELISA, increasing confidence in the validity of this newer technology for EBV research (see Supplemental Methods). This assay does have limitations worth noting; EBV proteins were printed onto the microarray as cell-free translated sequences, so antibody responses specific to conformational structure or post-translational processing such as glycosylation may have been missed. Although this phenomenon would not be likely to introduce false associations, it may have precluded our ability to detect certain case-control differences.

Future research is warranted to develop clinical assays that quantitate exact amounts of the most promising antibodies identified here in the circulation of at-risk individuals. Our data allowed us to: (1) report strong associations between NPC risk and the EBV-directed immune response and (2) illustrate the potential of probing additional antibodies, including our 14-antibody signature, to improve upon current EBV-based NPC biomarkers. However, the antibody output from our research-based protein microarray does not directly translate to an amount of antibody in the blood, and our risk score and associated cutoff values should not be interpreted as tests ready for clinical application. Further assay development should also include calibration of any new clinical tests in prospective cohorts to allow for the generation of NPC prediction curves specific to the target population.

We utilized a custom microarray to characterize antibody responses directed against a comprehensive set of 199 EBV protein sequences in three independent study populations at risk of NPC. This multiplex technology facilitated the discovery of NPC case-control differences in 133 anti-EBV antibodies targeting proteins across all stages of the EBV life cycle. We further leveraged this high-dimensional dataset for cancer biomarker discovery, identifying a 14-antibody risk stratification signature that significantly improved NPC prediction compared to current biomarkers. This improved performance has the potential to translate into substantial public health benefits in global regions most affected by this cancer. Future studies should focus on replication of findings in other NPC-endemic populations and work towards the development of highly reproducible clinical tests targeting the identified anti-EBV antibodies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Joseph Campo and Chris Hung for their roles in the management and generation of the EBV protein microarrays. This research was supported by the National Cancer Institute Intramural Research Program, the National Health and Medical Research Council of Australia, and George Washington University.

References

1. zur Hausen H, Schulte-Holthausen H, Klein G, Henle W, Henle G, Clifford P, et al. EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. *Nature*. 1970; 228(5276):1056–8. [PubMed: 4320657]
2. Dowd JB, Palermo T, Brite J, McDade TW, Aiello A. Seroprevalence of Epstein-Barr virus infection in U.S. children ages 6–19, 2003–2010. *PLoS One*. 2013; 8(5):e64921. [PubMed: 23717674]
3. de-The G, Day NE, Geser A, Lavoue MF, Ho JH, Simons MJ, et al. Sero-epidemiology of the Epstein-Barr virus: preliminary analysis of an international study - a review. *IARC Sci Publ*. 1975; (11 Pt 2):3–16.
4. Chen CY, Huang KY, Shen JH, Tsao KC, Huang YC. A large-scale seroprevalence of Epstein-Barr virus in Taiwan. *PLoS One*. 2015; 10(1):e0115836. [PubMed: 25615611]
5. Thorley-Lawson DA. EBV Persistence--Introducing the Virus. *Curr Top Microbiol Immunol*. 2015; 390(Pt 1):151–209. [PubMed: 26424647]
6. Young LS, Yap LF, Murray PG. Epstein-Barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer*. 2016; 16(12):789–802. [PubMed: 27687982]
7. Chua ML, Wee JT, Hui EP, Chan AT. Nasopharyngeal carcinoma. *Lancet*. 2016; 387(10022):1012–24. [PubMed: 26321262]
8. Klein G, Geering G, Old LJ, Henle G, Henle W, Clifford P. Comparison of the anti-EBV titer and the EBV-associated membrane reactive and precipitating antibody levels in the sera of Burkitt lymphoma and nasopharyngeal carcinoma patients and controls. *Int J Cancer*. 1970; 5(2):185–94. [PubMed: 4915449]
9. Henle G, Henle W. Epstein-Barr virus-specific IgA serum antibodies as an outstanding feature of nasopharyngeal carcinoma. *Int J Cancer*. 1976; 17(1):1–7. [PubMed: 175020]
10. de-The G, Lavoue MF, Muenz L. Differences in EBV antibody titres of patients with nasopharyngeal carcinoma originating from high, intermediate and low incidence areas. *IARC Sci Publ*. 1978; (20):471–81. [PubMed: 215529]
11. Coghill AE, Hsu WL, Pfeiffer RM, Juwana H, Yu KJ, Lou PJ, et al. Epstein-Barr virus serology as a potential screening marker for nasopharyngeal carcinoma among high-risk individuals from multiplex families in Taiwan. *Cancer Epidemiol Biomarkers Prev*. 2014; 23(7):1213–9. [PubMed: 24769890]
12. Ji MF, Wang DK, Yu YL, Guo YQ, Liang JS, Cheng WM, et al. Sustained elevation of Epstein-Barr virus antibody levels preceding clinical onset of nasopharyngeal carcinoma. *Br J Cancer*. 2007; 96(4):623–30. [PubMed: 17285127]
13. Cao SM, Liu Z, Jia WH, Huang QH, Liu Q, Guo X, et al. 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(4):e19100. [PubMed: 21544243]
14. Chien YC, Chen JY, Liu MY, Yang HI, Hsu MM, Chen CJ, et al. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med*. 2001; 345(26):1877–82. [PubMed: 11756578]
15. Liu Z, Ji MF, Huang QH, Fang F, Liu Q, Jia WH, 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. *Am J Epidemiol*. 2013; 177(3):242–50. [PubMed: 23255783]
16. Middeldorp JM. Epstein-Barr Virus-Specific Humoral Immune Responses in Health and Disease. *Curr Top Microbiol Immunol*. 2015; 391:289–323. [PubMed: 26428379]
17. Hildesheim A, Apple RJ, Chen CJ, Wang SS, Cheng YJ, Klitz W, et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst*. 2002; 94(23):1780–9. [PubMed: 12464650]
18. Chen CJ, Yang HI, Su J, Jen CL, You SL, Lu SN, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA*. 2006; 295(1):65–73. [PubMed: 16391218]

19. Pickard A, Chen CJ, Diehl SR, Liu MY, Cheng YJ, Hsu WL, et al. Epstein-Barr virus seroreactivity among unaffected individuals within high-risk nasopharyngeal carcinoma families in Taiwan. *Int J Cancer*. 2004; 111(1):117–23. [PubMed: 15185352]
20. Vigil A, Davies DH, Felgner PL. Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future Microbiol*. 2010; 5(2):241–51. [PubMed: 20143947]
21. Doolan DL. Plasmodium immunomics. *Int J Parasitol*. 2011; 41(1):3–20. [PubMed: 20816843]
22. Davies DH, Duffy P, Bodmer JL, Felgner PL, Doolan DL. Large screen approaches to identify novel malaria vaccine candidates. *Vaccine*. 2015; 33(52):7496–505. [PubMed: 26428458]
23. Chang C, Middeldorp J, Yu KJ, Juwana H, Hsu WL, Lou PJ, et al. Characterization of ELISA detection of broad-spectrum anti-Epstein-Barr virus antibodies associated with nasopharyngeal carcinoma. *J Med Virol*. 2013; 85(3):524–9. [PubMed: 23280934]
24. Wentzensen N, Wacholder S. From differences in means between cases and controls to risk stratification: a business plan for biomarker development. *Cancer discovery*. 2013; 3(2):148–57. [PubMed: 23299199]
25. Gu AD, Lu LX, Xie YB, Chen LZ, Feng QS, Kang T, et al. Clinical values of multiple Epstein-Barr virus (EBV) serological biomarkers detected by xMAP technology. *Journal of translational medicine*. 2009; 7:73. [PubMed: 19698162]
26. Fachiroh J, Schouten T, Hariwiyanto B, Paramita DK, Harijadi A, Haryana SM, et al. Molecular diversity of Epstein-Barr virus IgG and IgA antibody responses in nasopharyngeal carcinoma: a comparison of Indonesian, Chinese, and European subjects. *J Infect Dis*. 2004; 190(1):53–62. [PubMed: 15195243]
27. Liu X, Cohen JI. Epstein-Barr Virus (EBV) Tegument Protein BGLF2 Promotes EBV Reactivation through Activation of the p38 Mitogen-Activated Protein Kinase. *J Virol*. 2015; 90(2):1129–38. [PubMed: 26559845]
28. Calderwood MA, Holthaus AM, Johannsen E. The Epstein-Barr virus LF2 protein inhibits viral replication. *J Virol*. 2008; 82(17):8509–19. [PubMed: 18562535]
29. Heilmann AM, Calderwood MA, Johannsen E. Epstein-Barr virus LF2 protein regulates viral replication by altering Rta subcellular localization. *J Virol*. 2010; 84(19):9920–31. [PubMed: 20631124]
30. Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L. Lytic cycle switches of oncogenic human gammaherpesviruses. *Adv Cancer Res*. 2007; 97:81–109. [PubMed: 17419942]
31. Littler E, Baylis SA, Zeng Y, Conway MJ, Mackett M, Arrand JR. Diagnosis of nasopharyngeal carcinoma by means of recombinant Epstein-Barr virus proteins. *Lancet*. 1991; 337(8743):685–9. [PubMed: 1672175]
32. Connolly Y, Littler E, Sun N, Chen X, Huang PC, Stacey SN, et al. Antibodies to Epstein-Barr virus thymidine kinase: a characteristic marker for the serological detection of nasopharyngeal carcinoma. *Int J Cancer*. 2001; 91(5):692–7. [PubMed: 11267982]
33. Kumar R, Whitehurst CB, Pagano JS. The Rad6/18 ubiquitin complex interacts with the Epstein-Barr virus deubiquitinating enzyme, BPLF1, and contributes to virus infectivity. *J Virol*. 2014; 88(11):6411–22. [PubMed: 24672041]
34. van Gent M, Braem SG, de Jong A, Delagic N, Peeters JG, Boer IG, et al. Epstein-Barr virus large tegument protein BPLF1 contributes to innate immune evasion through interference with toll-like receptor signaling. *PLoS pathogens*. 2014; 10(2):e1003960. [PubMed: 24586164]
35. Farina A, Feederle R, Raffa S, Gonnella R, Santarelli R, Frati L, et al. BFRF1 of Epstein-Barr virus is essential for efficient primary viral envelopment and egress. *J Virol*. 2005; 79(6):3703–12. [PubMed: 15731264]
36. Henson BW, Perkins EM, Cothran JE, Desai P. Self-assembly of Epstein-Barr virus capsids. *J Virol*. 2009; 83(8):3877–90. [PubMed: 19158247]
37. Lee CP, Liu GT, Kung HN, Liu PT, Liao YT, Chow LP, et al. The Ubiquitin Ligase Itch and Ubiquitination Regulate BFRF1-Mediated Nuclear Envelope Modification for Epstein-Barr Virus Maturation. *J Virol*. 2016; 90(20):8994–9007. [PubMed: 27466427]
38. Hu L, Lin Z, Wu Y, Dong J, Zhao B, Cheng Y, et al. Comprehensive profiling of EBV gene expression in nasopharyngeal carcinoma through paired-end transcriptome sequencing. *Front Med*. 2016; 10(1):61–75. [PubMed: 26969667]

39. Ramayanti O, Juwana H, Verkuijlen SA, Adham M, Pegtel MD, Greijer AE, et al. Epstein-Barr virus mRNA profiles and viral DNA methylation status in nasopharyngeal brushings from nasopharyngeal carcinoma patients reflect tumor origin. *Int J Cancer*. 2017; 140(1):149–62. [PubMed: 27600027]

Author Manuscript

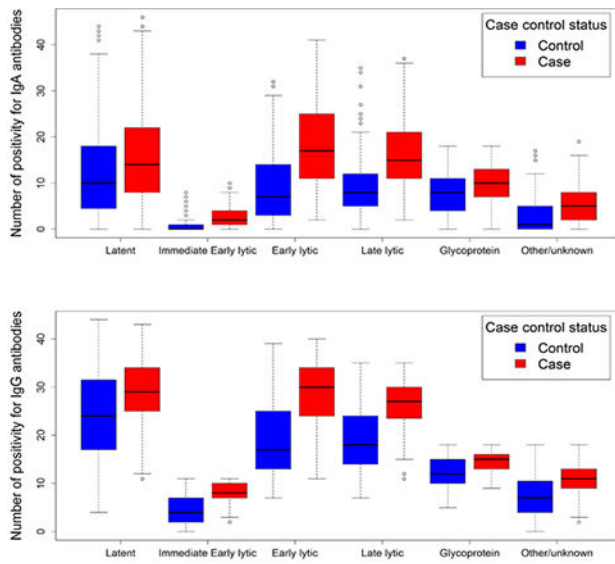
Author Manuscript

Author Manuscript

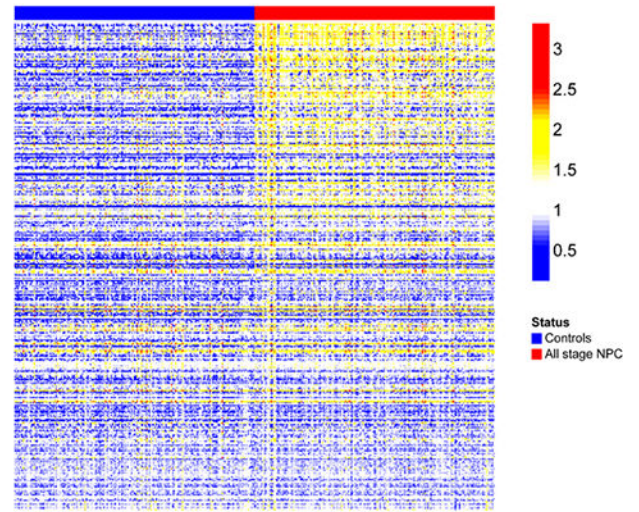
Author Manuscript

TRANSLATIONAL RELEVANCE

Although it has been demonstrated that antibodies against Epstein-Barr virus (EBV) are elevated in patients who develop nasopharyngeal carcinoma (NPC), antibodies targeting only ~10% of EBV's nearly 100 proteins have been investigated as potential biomarkers for this EBV-related tumor. In fact, IgA antibodies against viral capsid antigen (VCA) and EBV nuclear antigen 1 (EBNA1) are the only EBV-based markers currently being evaluated for population-level NPC detection. This study represents the first comprehensive evaluation of the EBV antibody repertoire for cancer biomarker discovery, measuring both IgG and IgA antibody responses against 199 EBV protein sequences. Using this high-dimensional dataset, we observed 133 antibody elevations in Stage I/IIa NPC patients and selected a 14-antibody subset that predicted 5-year NPC risk with 89–93% accuracy in two independent Taiwanese cohorts. This represented a significant improvement beyond current VCAp18/EBNA1 IgA biomarkers alone, suggesting that targeting additional EBV proteins can improve EBV-related cancer risk stratification.

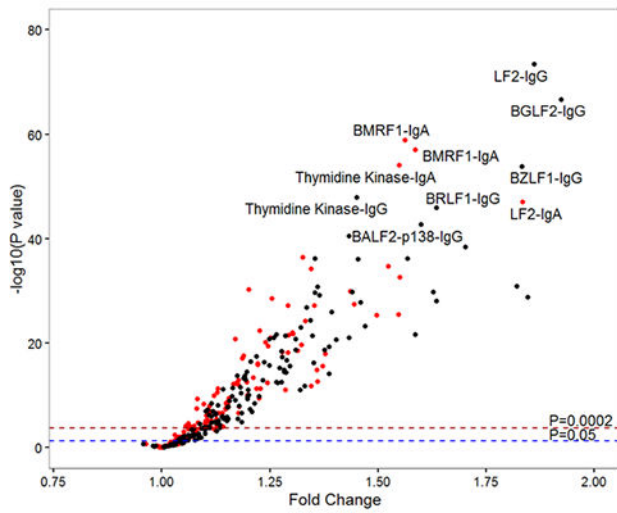


Case-control differences in the antibody response according to antibody type and stage of EBV life cycle

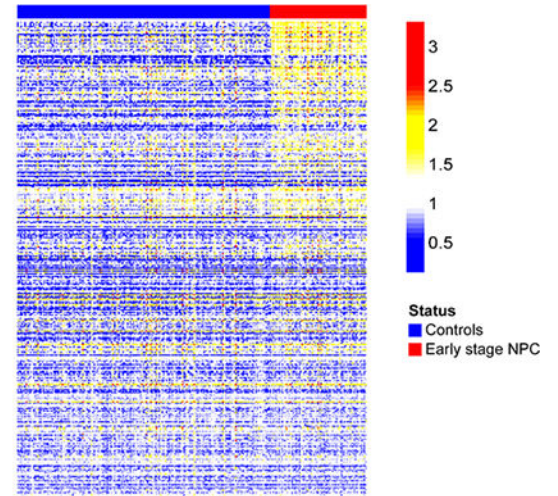


Heat map illustrating strength of antibody response to all microarray sequences, according to case status

Figure 1. Antibody responses in NPC cases (all stages) and controls, overall and by EBV life cycle grouping



Case-control differences in the mean antibody response, including 10 strongest antibody elevations in cases



Heat map illustrating strength of antibody response to all microarray sequences, according to case status

Figure 2.
Differential anti-EBV antibody responses between Stage I/IIa NPC cases and controls

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

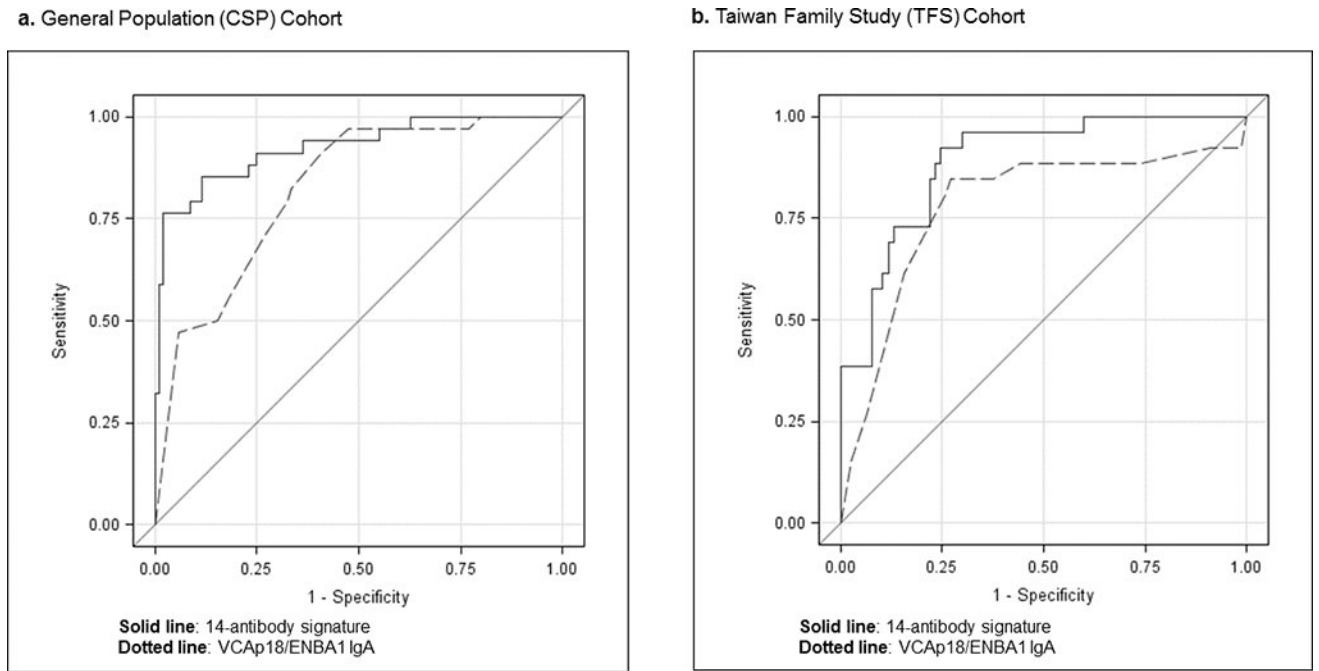


Figure 3. Receiver operating curves (ROC) for NPC prediction in disease-free individuals, by EBV-based antibody risk signature and Taiwanese cohort

Table 1

Participant characteristics according to NPC status, by study population

	Case-control Study (N=350) (1991-1994)		Prospective CSP Cohort (N=23,943) (1991-2002)		Prospective TFS Cohort (N=2,557) (1996-2014)	
	Prevalent NPC (N=175)	Community Controls (N=175)	Incident NPC (N=37)	Disease-Free Individuals (N=117)	Incident NPC (N=26)	Unaffected Family members (N=77)
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Age at enrollment						
<=40	62 (35.4)	60 (34.3)	7 (18.9)	21 (18.0)	11 (42.3)	33 (42.9)
41-50	36 (20.6)	48 (27.4)	7 (18.9)	26 (22.2)	9 (34.6)	27 (35.1)
>50	77 (44.0)	67 (38.3)	23 (62.2)	70 (59.8)	6 (23.1)	17 (22.1)
Sex						
Female	55 (31.4)	59 (33.7)	8 (21.6)	25 (21.4)	6 (23.1)	18 (23.4)
Male	120 (68.6)	116 (66.3)	29 (78.4)	92 (78.6)	20 (76.9)	59 (76.6)
Smoking						
Never	86 (49.1)	100 (57.1)	--	--	10 (38.5)	42 (54.0)
Former	30 (17.1)	15 (8.6)	--	--	3 (11.5)	6 (7.9)
Current	59 (33.7)	60 (34.3)	--	--	13 (50.0)	29 (38.2)
Lag time*						
<2 years			4 (10.8)		4 (15.4)	
2-5 years			18 (48.6)		7 (26.9)	
>5 years			11 (29.7)		12 (46.2)	
Not assigned			4 (10.8)		3 (11.5)	

* Lag time defined as the number of years between study enrollment blood draw and NPC diagnosis for incident cases from the prospective cohorts. In the case-control study, all cases had blood drawn after NPC diagnosis but prior to cancer treatment.

EBV proteins (name and microarray sequence) for the IgA and IgG antibody responses with the highest standardized mean difference between cases and controls

Table 2

	IgA Antibody Responses					IgG Antibody Responses						
	Early stage NPC (N=67)		Disease-free controls (N=175)		Association and prediction	Early stage NPC (N=67)		Disease-free controls (N=175)		Association and prediction		
	Mean	Positive	Mean	Positive		Mean	Positive	Mean	Positive			
				OR* (95%CI)	Delta± (Specificity)				OR* (95%CI)	Delta± (Specificity)		
Overlapping IgA and IgG Markers												
BXLF1 (thymidine kinase): YP_001129497.1-133399-131576	1.32	85.1%	0.87	14.3%	34(16,78)	2.51(89%)	1.85	100%	1.30	91.4%	>999	1.74(68%)
LF2 protein: YP_001129504.1-151808-150519	1.10	61.2%	0.61	1.1%	136(31,598)	2.39(87%)	1.71	97.0%	0.92	32.6%	67(16,28)	2.57(90%)
BRLF1 (Rta): YP_001129468.1-93725-91908	1.23	79.1%	0.83	15.4%	21(10,43)	1.64(65%)	1.78	95.5%	1.10	49.7%	22(6.5,71)	1.83(71%)
BALF2(ss DNA binding protein): YP_001129510.1-165796-162410-1	0.82	22.4%	0.56	1.1%	25(5.5,113)	1.47(58%)	1.20	79.1%	0.79	16.6%	19(9.4,39)	1.57(62%)
Markers in Top 10 for IgA but not IgG												
BMRF1 (EAd): AFY97929.1-67486-68700	1.47	97.0%	0.96	42.3%	44(11,187)	1.97(76%)	1.90	100%	1.43	88.6%	>999	1.21(48%)
BMRF1 (EAd): YP_001129454.1-67745-68959	1.46	97.0%	0.94	40.6%	48(11,201)	1.90(74%)	1.91	100%	1.41	85.1%	>999	1.21(48%)
BPLF1 (tegument protein): CAA24839.1-71527-62078-2	1.33	89.6%	1.01	46.3%	10(4.3,23)	1.66(66%)	1.86	100%	1.57	97.1%	>999	0.79(32%)
BKRF4 (tegument protein): AFY97946.1-98716-99369	0.84	22.4%	0.61	1.7%	17(4.6,59)	1.41(56%)	1.18	70.1%	0.90	31.4%	5.1(2.8,10)	0.97(39%)
BVRF1: YP_001129499.1-133954-135666	1.14	79.1%	0.87	24.0%	12(6.1,24)	1.41(56%)	1.50	95.5%	1.13	69.1%	10(2.9,32)	1.33(53%)
LMP-1: AFY97906.1-168167-168081	1.16	83.6%	0.98	40.6%	8(3.7,15)	1.34(53%)	1.37	92.5%	1.00	51.4%	12(4.5,31)	1.14(45%)
Markers in Top 10 for IgG but not IgA												
BGLF2: YP_001129486.1-115415-114405	0.91	37.3%	0.60	0.8%	6.8(3.3,14)	1.31(52%)	1.36	86.6	0.71	12.6	45(20,100)	2.58(90%)
BZLF1 (Zta): YP_001129467.1-90855-90724	0.91	34.3%	0.71	0%	>999	1.19(47%)	1.43	86.6	0.82	26.9	18(8.1,38)	1.82(71%)
BALF3: CAA24807.1-161678-159312	0.95	32.8%	0.79	6.9%	6.6(3.1,14)	0.88(35%)	1.25	70.1	0.79	14.3	14(7.2,28)	1.57(62%)
BFRF1: YP_001129446.1-46719-47729	1.28	86.6%	1.07	65.1%	3.5(1.6,7.4)	0.98(39%)	1.74	100.0	1.25	83.4	>999	1.57(62%)
BBLF2: CAA24824.1-119080-117515	0.84	26.9%	0.66	5.1%	6.8(2.9,16)	0.78(31%)	1.13	55.2	0.65	9.7	12(5.7,23)	1.46(58%)
BALF2(ss DNA binding protein): YP_001129510.1-165796-162410-2	1.12	59.7%	0.89	13.7%	9.3(4.9,18)	1.25(49%)	1.56	97.0	1.13	64.6	18(4.2,75)	1.45(58%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

* Odds Ratio (OR) describes the association between positivity for selected protein sequence and early-stage NPC..

± Delta statistics captured the potential predictive ability of each biomarker; specificity (1 - false positivity rate in controls) calculated for cutoff value corresponding to 90% sensitivity (positivity in early-stage NPC cases identified cross-sectionally)

Table 3

Most promising EBV-related antibody targets for NPC prediction, identified from custom protein microarray

EBV Protein and Microarray Sequence	Common Description for EBV Protein	Antibody Type in Risk Score
BZLF1: YP_001129467.1-90855-90724	Zebra (Zta) lytic switch protein	IgG
BORF1: YP_001129451.1-63084-64178	Viral capsid subunit	IgG
BFRF1: YP_001129446.1-46719-47729	Capsid egress protein	IgG
BGLF2: YP_001129486.1-115415-114405	Tegument protein	IgG
BXLF1: YP_001129497.1-133399-131576	Thymidine kinase	IgG, IgA
BRLF1: YP_001129468.1-93725-91908	Rta lytic switch protein	IgA
LF2: YP_001129504.1-151808-150519	LF2 protein	IgG, IgA
BMRF1: YP_001129454.1-67745-68959	Early antigen	IgA
BMRF1: AFY97929.1-67486-68700	Early antigen	IgA
BPLF1: CAA24839.1-71527-62078-2	Tegument protein	IgA

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript