

Genomic Analysis of Nasopharyngeal Carcinoma Reveals TME-Based Subtypes



Li Zhang¹, Kenzie D. MacIsaac², Ting Zhou¹, Pei-Yu Huang¹, Chunlin Xin³, Jason R. Dobson², Kun Yu², Derek Y. Chiang², Yue Fan³, Marc Pelletier², Yan Wang², Savina Jaeger⁴, Viveksagar Krishnamurthy Radhakrishnan², Lellean JeBailey², Peter Skewes-Cox⁵, Jing Zhang¹, Wenfeng Fang¹, Yan Huang¹, Hongyun Zhao¹, Yuanyuan Zhao¹, En Li³, Bin Peng³, Alan Huang⁶, Glenn Dranoff², Peter S. Hammerman², Jeffrey Engelman², Hans Bitter², Yi-Xin Zeng¹, and Yao Yao³

Abstract

Nasopharyngeal carcinoma (NPC) is an Epstein–Barr virus (EBV) associated cancer characterized by a poor prognosis and a high level of lymphocyte infiltrate. Genetic hallmarks of NPC are not completely known but include deletion of the *p16* (*CDKN2A*) locus and mutations in NF- κ B pathway components, with a relatively low total mutational load. To better understand the genetic landscape, an integrated genomic analysis was performed using a large clinical cohort of treatment-naïve NPC tumor specimens. This genomic analysis was generally concordant with previous studies; however, three subtypes of NPC were identified by differences in immune cell gene expression, prognosis, tumor cell morphology, and genetic characteristics. A gene expression signature of prolif-

eration was poorly prognostic and associated with either higher mutation load or specific EBV gene expression patterns in a subtype-specific manner. Finally, higher levels of stromal tumor-infiltrating lymphocytes associated with good prognosis and lower expression of a WNT and TGF β pathway activation signature.

Implications: This study represents the first integrated analysis of mutation, copy number, and gene expression data in NPC and suggests how tumor genetics and EBV infection influence the tumor microenvironment in this disease. These insights should be considered for guiding immunotherapy treatment strategies in this disease. *Mol Cancer Res*; 15(12); 1722–32. ©2017 AACR.

Introduction

Nasopharyngeal carcinoma (NPC) is a cancer of the epithelial cells of the nasopharynx and is associated with a high incidence of treatment failure and overall poor prognosis (1, 2). The prevalence of NPC is skewed geographically with South-East Asia, the middle East, and Northern Africa being disproportionately affected (3). Epidemiologic factors associated with NPC include tobacco (4) and alcohol use (5), consumption of salt-preserved foods (6, 7), and smoke inhalation from cooking fires (8, 9). A path-

ologic classification of the disease into keratinizing squamous cell carcinoma, nonkeratinizing (differentiated and undifferentiated), and basaloid squamous cell carcinoma has been proposed (10). Epstein–Barr virus (EBV) infection is strongly associated with the undifferentiated form of nonkeratinizing NPC (11). The EBV genome is maintained in tumor cells, and viral mRNA and proteins are detectable in most tumors, suggesting that the virus plays a fundamental role in disease pathogenesis and development (12). Another hallmark of undifferentiated NPC is a marked and substantial infiltration of immune cells into the tumor microenvironment (13, 14). There is evidence that interactions between these stromal immune cells and tumor cells are important in NPC growth and proliferation (15), and tumor cells may coopt immune escape mechanisms evolved by EBV to evade immunosurveillance (16, 17).

Previous studies have indicated that the frequency of somatic genomic alterations in NPC is low (18–20). However, prior work has identified several recurrent genomic alterations in NPC, including deletion of the *p16*-locus on cytoband 9p21, *CCND1* amplification, *TP53* mutation, inactivating mutations in negative regulators of the NF- κ B pathway, mutations in MAPK and PI3K signaling pathways, and mutations in epigenetic regulators like *MLL3* (18, 19, 21). In this study, we extend this prior work by presenting an integrated analysis of RNA sequencing (RNA-seq), whole-exome sequencing, copy number analysis, histology, and clinical data for 113 patients with undifferentiated NPC to characterize the disease at a molecular level. Our objectives were to identify distinct gene expression-based subtypes of this disease, to define the tumor and microenvironment gene expression

¹Sun Yat-Sen University Cancer Center, Guangzhou, China. ²Novartis Institute for Biomedical Research, Cambridge, Massachusetts. ³Novartis Institute for Biomedical Research, Shanghai, China. ⁴Novartis Oncology Global Development, Cambridge, Massachusetts. ⁵Novartis Institute for Biomedical Research, Emeryville, California. ⁶Third Rock Ventures, LLC, Boston, Massachusetts.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (<http://mcr.aacrjournals.org/>).

L. Zhang, K.D. MacIsaac, T. Zhou, and P.-Y. Huang are co-first authors of this article.

Corresponding Authors: Kenzie D. MacIsaac, Novartis Institute for Biomedical Research, 250 Massachusetts Avenue, Cambridge, MA 02139. Phone: 617-871-3648; Fax: 617-871-4714; E-mail: kenzie.macisaac@novartis.com; Yixin Zeng, Sun Yat-Sen University Cancer Center, 651 Dongfeng Road East, Guangzhou 510060, China. E-mail: zengyx@sysucc.org.cn; and Yao Yao, Novartis Institute for Biomedical Research (China), 898 Halei Road, Building 8, Pudong, Shanghai 201203, China. E-mail: yao.yao@novartis.com

doi: 10.1158/1541-7786.MCR-17-0134

©2017 American Association for Cancer Research.

features of these subtypes, to examine how these features influence prognosis, and to characterize the mutational landscape of NPC and how it relates to subtype.

Materials and Methods

Tissue collection

Treatment-naïve primary NPC biopsies and matching blood samples were collected at Sun Yat-Sen University Cancer Center (SYSUCC, Guangzhou, China) under patients' informed consent that has been approved by the Ethical Committee at SYSUCC (demographic Supplementary Table S1). To avoid stromal dilution of tumor cell-specific rare events, we procured tumors with a tumor cell content of at least 50%.

Progression-free survival

Progression-free survival (PFS) was defined as the time from the date of diagnosis to the date of objective tumor progression (excluding clinical deterioration without evidence of objective progression) according to the RECIST, or death from any cause. Median follow-up time for the enrolled patients is 27.3 months.

Transcriptome sequencing by RNA-seq

The quality of purified RNA sample was measured by Agilent 2111 Bioanalyzer and/or gel electrophoresis and samples with RIN value >7 were considered to pass QC. Transcriptome sequencing was performed at WuXi AppTec. Approximately 50 million pair-end reads (100 bp) were generated for each sample. The RNA-seq data were analyzed using Omicsoft ArraySuite software. Briefly, the analysis steps include QC, alignment to Human B37 genome with UCSC gene models, and gene level expression quantification using FPKM. The mean rate of unique read mapping was 88% and the mean uniquely paired rate was 83%. EBV gene expression was quantitated by aligning unmapped RNA-seq reads to the EBV genome using the Omicsoft ArrayStudio pipeline. Data have been deposited in the Gene Expression Omnibus (accession GSE102349).

Gene expression-based subtypes and analysis

Principal components analysis was performed on log FPKM gene expression data. We then calculated the dissimilarity between pairs of samples as their Euclidean distance in the PCA-projected space. Clustering was performed using the Affinity Propagation algorithm, varying the self-similarity parameter to obtain clustering solutions varying from two to six clusters. For each solution, the WB index (22) was calculated as the product of the number of clusters and the ratio of within cluster to between cluster variance. Three clusters minimized the WB index. To evaluate the robustness of this clustering solution, we generated 1,000 randomly subsampled datasets comprised of two thirds of the total samples and repeated the clustering procedure, using the WB index to identify the best solution. We found that the best number of clusters was between 2 and 4 in the large majority of trials, with three clusters the most frequent number identified (Supplementary Fig. S1). To further evaluate robustness of our clustering, we used the same random sampling and clustering procedure to group patient samples into three clusters. Across 1,000 random samplings, we computed the probability that two patient samples assigned to the same subtype would be grouped into the same cluster in the subsampled dataset. This probability averaged around 80% across the trials, which was dramatically

higher than random expectation estimated using permuted sample labels (Supplementary Fig. S2). Gene signature scores were calculated as the average log₂ FPKM value for all genes in a signature. Gene set enrichment analysis was carried out using Fisher exact test.

Whole-exome data analysis

DNA was extracted from FFPE and whole blood control samples using the Promega Maxwell DNA Extraction Kit. Libraries were constructed using Illumina TruSeq, captured with the Agilent SureSelect Human All Exon V5 and protocol, and sequenced on an Illumina HiSeq 2500 as 100 base pair paired-end reads. Resulting FASTQ files were aligned to the hg19 reference genome using BWA-MEM (23), duplicates were marked with PICARD, and local realignment and base quality recalibration was performed with GATK (24, 25).

The whole-exome sequencing was performed on tumor samples from 111 NPC patients, of which 57 patients have matched blood (hereafter referred to "paired cohort") and 54 have only tumor samples (hereafter referred to "unpaired cohort"). We had access to additional blood samples from 13 NPC patients whose tumor samples were not available for sequencing analysis. We compiled all the blood samples (70) as a reference panel to assist somatic mutation analysis. We achieved between 40 and 60× coverage on normal blood samples and 66 and 118× coverage on tumor samples, with greater than 95% of the samples achieving a percent target base coverage 10× greater than 95.0%, and 95% of samples achieving percent target base coverage 20× greater than 82.9%. Nonsynonymous mutations identified are provided in Supplementary Table S2 for the paired cohort and Supplementary Table S3 for the unpaired cohort.

Detection of somatic mutations

Single-nucleotide variants were called in paired normal mode using MuTect (26), while indels were called with Pindel (27) and filtered against panel of normal. Annotation was performed with SnpEff (28), using dbSNP v141, COSMIC v70 (29), and dbNSFP v2.4 (30). Nonsynonymous protein-coding SNVs or indels with coverage of at least 10× or 30×, respectively, in at least one tumor sample were retained.

Finally, the following SNVs were eliminated: (i) SNVs identified using sequencing data from paired normal DNA; (ii) SNVs identified in the blood reference panel; or (iii) variants registered in either dbSNP141 or the 1000 Genomes project. To further remove private germline variants in the unpaired cohort, we removed variants with allele frequency between 45% and 55% or 95% and 100%. Mutation load was evaluated separately in the paired and unpaired cohorts using the total number of nonsynonymous mutations observed in that sample. A sample was labeled as high or low mutation load based on whether it was above or below the median value for its cohort.

SNP6.0 arraying and copy number analysis

Copy number analysis was performed using Affymetrix genome-wide human SNP 6.0 Array at Shanghai Biotechnology Corp. following protocols from Affymetrix. The raw CEL files were returned. Affymetrix Genotyping Console (v4.2) was used to perform QC. Samples with contrast QC less than 0.4 or median of the absolute values of all pairwise differences greater than 0.35 were excluded from the analysis. Genomic segmentation algorithm of Partek Genomics Suite (v6.6) was run with default

settings in paired mode for samples with matched normals and unpaired mode for tumor-only samples to get segment-level calls. Gene-level calls were obtained using the mean of the segments spanning gene coordinates. Significant, gene-level copy number aberrations were called based on a P value < 0.05 and a \log_2 fold change cutoff of 1 for amplifications and -1 for deletions (Supplementary Table S4).

Statistical analyses

Tukey honest significant difference (HSD) test was used for pairwise comparisons of gene expression signatures or principal components between the three subtypes. For two-group gene expression comparisons, we evaluated differences with a two-sided unpaired t test, and for two-group comparisons of the nonnormally distributed tumor-infiltrating lymphocytes (TIL) scores, we used a Mann–Whitney U test. Group-wise comparisons were performed by ANOVA for gene expression measures and by a Kruskal–Wallis test for TILs.

Morphologic evaluation of TILs

Hematoxylin and eosin (H&E)-stained tumor sections were scanned by Aperio ScanScope XT, and digital images were managed in Spectrum. Morphologic evaluation of TILs was manually performed by a CLIA-certified pathologist following the guidelines set forth by the International TILs Working Group 2014 (31).

Results

Gene expression–based subtypes

We employed RNA-seq to quantify genome-wide transcript levels in 113 fresh, treatment-naïve undifferentiated NPCs. The patient characteristics for the cohort under study are provided in Supplementary Table S1. Using genome-wide mRNA expression profiles, we derived three NPC subtypes using unsupervised clustering (Fig. 1). Sixty-one samples were assigned to subtype I, 31 to subtype II, and 21 to subtype III. Examining PFS, we observed a trend to differences among subtypes (Fig. 1A, $P = 0.08$, log-rank test). Patients in subtype I had the worst prognosis, whereas no progressions were observed in subtype III. Histopathologic examination of tumor samples defined four morphologically distinct groups of NPC tumors (Supplementary Fig. S3), and we observed significant differences in morphology by expression-based subtype ($P = 0.027$, χ^2 test). Subtype I was the only subtype that included samples with a differentiated morphology, had the highest fraction of samples with mixed morphology, and relatively few undifferentiated samples. In contrast, both subtypes II and III were more enriched for an undifferentiated morphology (Fig. 1B). Stromal TILs quantified by histopathologic examination (Fig. 1C) were highest in subtype III, and lowest in subtype II. Group differences did not reach statistical significance by a Kruskal–Wallis test, although the difference in stromal TILs between subtypes II and III was nominally significant ($P = 0.03$, MWU test). Median tumor cellularity was similar across subtypes (Supplementary Fig. S4), although subtype I was observed to have significantly higher tumor content than subtype II ($P = 0.006$, MWU test).

Tumor microenvironment characteristics of subtypes

Principal components analysis of gene expression data revealed that the first four principal components were significantly different across subtypes, and that they explained approximately 36% of variance in gene expression levels (Supplementary Fig. S5). For

each principal component (PC), we identified the most correlated and anticorrelated genes ($|\text{Pearson } R| > 0.6$) and assessed their canonical pathway enrichment in the MSIGDB collection (Supplementary Tables S5 and S6; ref. 32). PC1 was positively correlated with immune system genes and anticorrelated with cell cycle and proliferation markers. PC2 was positively correlated with TGF β signaling. PC3 was positively correlated with integrins, collagens, and genes involved in extracellular matrix receptor interactions. PC4 was negatively correlated with genes annotated as being involved in cancer-specific pathways. The strong association of PC1 and PC3 with immune and stromal cell genes, and the trend to differences by subtype in percent stromal TILs, prompted us to characterize the microenvironment of these tumor samples by examining prespecified signatures comprised of genes with enriched expression in various immune and stromal cell types, genes induced by interferons, and canonical proliferation markers (Supplementary Table S7; Supplementary Fig. S6). Subtype I tumors had the lowest expression of immune and stromal genes, and the highest expression of the proliferation signature (Fig. 1E). Subtype II was characterized by high expression of B-cell and T-cell genes but relatively low expression of cytotoxic markers and markers of IFN γ activation. Subtype III also showed high expression of immune genes, but was distinguished by the highest expression of cytotoxic cell markers, IFN γ activation signature, and macrophage markers. Interestingly, we also observed clear differences in the correlation between immune cell signatures by subtype (Supplementary Fig. S6). In particular, general T-cell gene expression markers correlated strongly with the cytotoxic cell and IFN γ signatures in subtypes I and III, but not in subtype II. In addition, the ratio of cytotoxic gene expression to T-cell gene expression was significantly lower, and the ratio of IgD to IgG heavy chain gene expression was significantly higher in subtype II versus subtype III ($P < 0.05$ Tukey HSD test), perhaps suggesting an enrichment of naïve lymphocytes in these tumors (Supplementary Fig. S7).

EBV gene expression patterns by subtype

All (100%) of the samples expressed characteristic EBV type II latency genes *LMP1*, *LMP2B*, and *EBNA1*. We observed significant differences in the level of EBV gene expression by subtype (Fig. 1D). Subtype II had significantly lower levels of total EBV gene expression, *A73* expression, *RPMS1* expression, and *EBNA1* expression than samples from other subtypes ($P < 0.05$ Tukey HSD), although expression of *LMP1* and *LMP2B* were not significantly different across subtypes. To explore how EBV gene expression programs are associated with tumor gene expression, we identified human genes whose expression was correlated with each of six EBV genes (Supplementary Tables S8 and S9; Supplementary Fig. S8). *LMP1* and *LMP2B* expression were strongly associated with activation of immune pathways such as NF- κ B, cytokine and chemokine signaling, expression of *CD274* (*PD-L1*), IFN γ signaling, and expression of cytotoxic effector molecules like *PRF1* and granzymes. *LMP1*, but not *LMP2B* expression, was negatively correlated with WNT pathway components, including *LEF1*, *WNT3*, and *PRICKLE1*. BART-associated transcripts *RPMS1* and *A73* were associated with transcriptional activation of the p53 signaling pathway and expression of cell cycle–related genes and negatively associated with expression of genes involved in hemostasis and tight junction formation. *EBNA1* was strongly associated with cell cycle and proliferation markers.

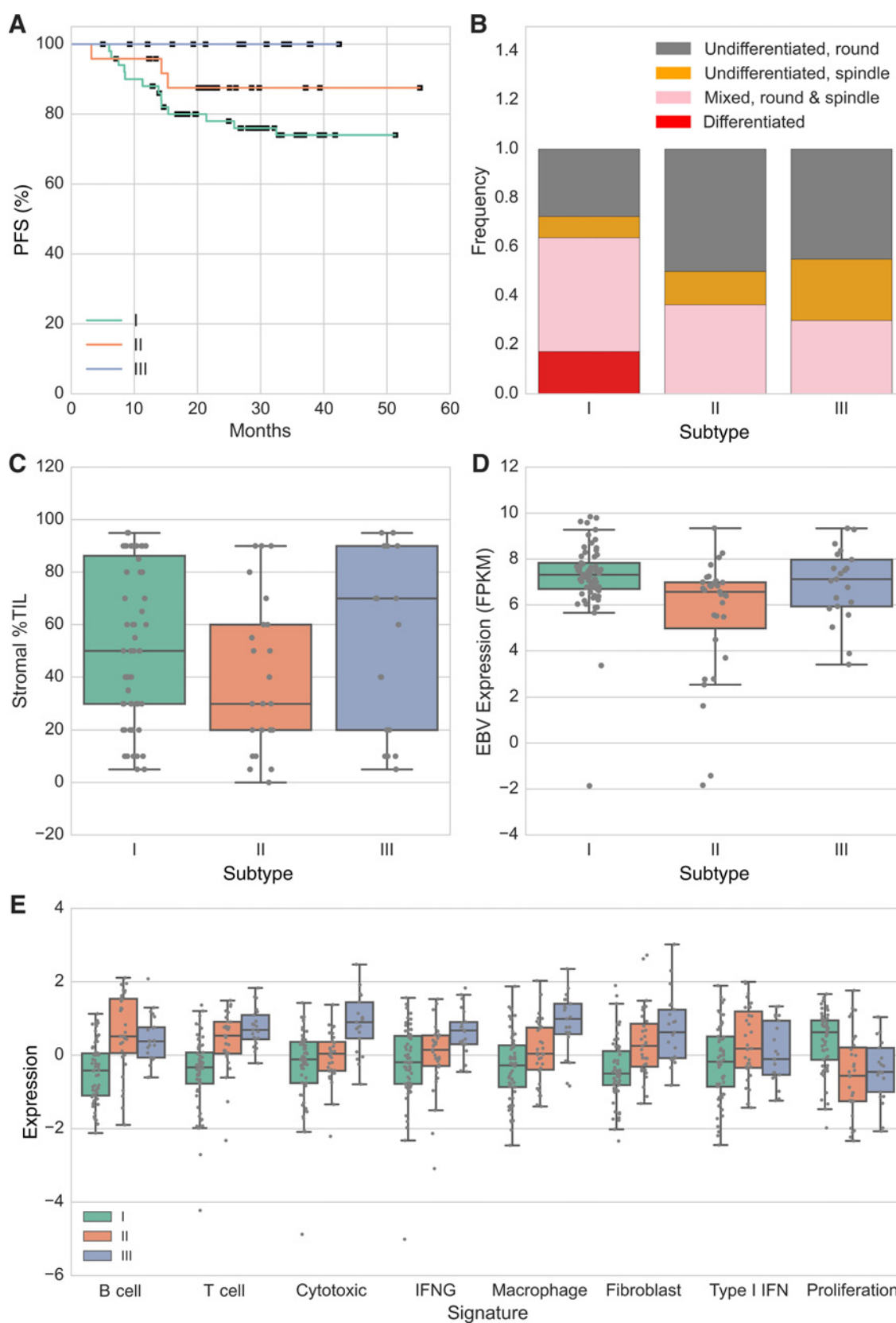


Figure 1. Characteristics of NPC gene expression-based subtypes. **A**, Progression-free survival by subtype. **B**, Distribution of morphologic subtype determined by H&E by gene expression subtype. **C**, Percent stromal TILs by subtype. **D**, Total EBV transcript expression by subtype. **E**, Gene signature expression levels (z-normalized) by subtype.

The mutational landscape of NPC

We characterized the mutational and somatic copy number landscape of NPC by performing whole-exome sequencing and SNP 6.0 array profiling on tumor samples from 94 patients that were also profiled by RNA-seq (Fig. 2). For 51 of these samples, we also sequenced matched normal blood (hereafter referred to as the "paired cohort"), and for the remaining 43 samples, no matched normal reference was sequenced or the blood sample failed QC (unpaired cohort, Supplementary Fig. S9). In the paired cohort, we identified a total of 1,520 nonsilent somatic mutations across 1,380 genes. The median mutation rate was 22 nonsynonymous lesions per tumor, which is comparable with the values reported in previous studies (18–20). There were relatively few genes that were recurrently mutated in the paired cohort. Similar to previous studies (18–20), we observed mutations in *TP53* (3/51; 5.9%), *KMT2D/MLL2* (2/51; 3.9%), and *TSHZ3* (2/51; 3.9%). However, mutations in genes, such as *BAP1*, *ARID1A*, and *RASSF1*, were not observed in this paired cohort. Consistent with previous reports, we observed that the majority of samples had arm-level losses of 3p, 14q, and 16q. We also observed recurrent somatic copy number variants, with most frequent events being deletions in the *CDKN2A* locus (9p21.2; 15/51; 29.4%) and amplifications in the *CCND1* locus (11q13; 2/51; 3.9%). Across both the paired and unpaired cohorts, *CDKN2A* deletion was enriched ($P = 0.01$ Fisher-Freeman-Halton test) in subtype I where 33% samples (19/57) harbored a deletion versus 5% (1/22) for subtype II and 12% (2/16) for subtype III.

Two recently published studies of NPC reported recurrent mutations in genes that inhibit the NF- κ B pathway (19, 20). Looking across both paired and unpaired cohorts, we observed genetic lesions in NF- κ B inhibitors in 37% (30/104) of patients. The most frequently altered genes were *CYLD* (mutated or deleted in 9 patients), *TRAF3* (9 patients), and *FBXW7* (6 patients). *CYLD* is a deubiquitinase that targets several NF- κ B signaling components to negatively regulate the pathway (33). *TRAF3* negatively regulates NF- κ B pathway activation by promoting degradation of the activating kinase NIK (34). *FBXW7* negatively regulates NF- κ B by promoting degradation of p100 (35). We observed an association between NF- κ B mutation status and higher stromal TILs ($P < 0.03$, MWU test), and a trend to lower intratumoral TILs ($P < 0.06$, MWU test). We also observed a strong enrichment ($P = 0.002$, Fisher-Freeman-Halton test) for NF- κ B pathway mutations in subtype I, with 24 of 58 patients (41%) harboring a mutation or deletion in the pathway, versus 2 of 28 (7%) in subtype II and 4 of 18 (22%) in subtype III. Also consistent with previously reported findings, we observed recurrent mutations in PI3K and MAPK pathway genes. In total, 15 patients were observed to have at least one mutation in *PIK3CA*, *KRAS*, *PTEN*, *BRAF*, *NRAS*, *FGFR2*, or *FGFR3*. The most frequently mutated genes were *PIK3CA* (4 patients) and *KRAS* (3 patients). Lesions in these genes were strongly associated with higher PC2 expression ($P = 0.002$, t test), significantly lower expression of all immune signatures, and a trend to lower stromal TILs ($P = 0.07$ MWU test). There were no significant differences in the frequency of mutation in these genes by subtype.

TILs are associated with better PFS

The previously described (36) association between immune cell infiltrate and outcome in NPC prompted us to determine

whether TILs were associated with PFS in this patient cohort. We partitioned samples into high (top 50%) and low (bottom 50%) TIL groups based on histopathologic estimates of %TIL fraction intratumorally and in the tumor stroma (Supplementary Fig. S10). High stromal TILs, but not intratumoral TILs, were significantly associated with better PFS (Fig. 3A; $P = 0.013$, log-rank test). Interestingly, PC2 was found to be significantly higher in the low stromal TIL samples versus high stromal TIL samples (Fig. 3B; $P = 0.02$, t test). To explore which pathways and processes may influence exclusion of TILs from the tumor microenvironment and lead to worse prognosis, we examined the 197 Entrez genes showing a strong ($R > 0.6$) correlation with PC2 across all 113 samples (Supplementary Table S3), and their overlap with canonical pathway annotations from the MSIGDB database (32). We identified nine gene sets significant at an FDR-corrected P value < 0.1 (Supplementary Table S4), with the most significant enrichment observed for TGF β pathway signaling (FDR-corrected $P = 3.2e-6$). Other pathways related to TGF β signaling were also significantly enriched, including the ALK pathway ($P = 1.8e-3$) and BMP signaling ($P = 4.4e-3$). Interestingly, we also observed several WNT pathway members among the 197 genes highly correlated with PC2, including *WNT3*, *WNT9A*, *FZD1*, *DVL3*, *PPP2CB*, and *LRP3*. Average log expression of the 197 genes showed a strong correlation with expression of genes activated by WNT/ β -catenin (Fig. 3C, $R = 0.66$, $P = 3.5e-15$) and genes upregulated by TGF β (Fig. 3D, $R = 0.55$, $P = 4.2e-10$) annotated in the MSIGDB hallmark collection (37).

A proliferation signature is associated with poor PFS

To explore whether any gene expression characteristics were associated with patient outcome, we tested our compendium of signatures and the first five PCs for associations with PFS by fitting Cox proportional hazards models. Controlling FDR at 10%, we found that PC1, the T-cell, and the B-cell signatures were associated with better PFS, whereas the proliferation signature was associated with worse outcome (Table 1). The proliferation signature showed the largest effect of any signature tested, and stratifying samples into high (upper 50%) and low (lower 50%) proliferation groups revealed a clear difference in patient outcome (Fig. 4A). In a multivariate Cox model, including stromal %TILs, PC1, the T-cell, B-cell, and proliferation signatures, the proliferation signature and stromal TIL terms remained independently significant ($P < 0.05$). High stromal TILs in patient samples with a high proliferation score appeared to be protective (Fig. 4B). Separating samples into high mutation load (upper 50%) and low mutation load (bottom 50%) groups, we observed higher proliferation signature score in the high mutation load group ($P = 0.04$, t test). Examining this association more closely, we observed that in subtype I (which was characterized by the highest mean proliferation signature score), there was no association between proliferation signature and mutation load, whereas in subtypes II and III, there was a strong association between mutation burden and proliferation signature (Fig. 4C; $P = 2.1e-3$, t test). Interestingly, this association appears to be a common feature of solid tumors as we observed a similar positive association between the proliferation signature and mutation load in 13 of 24 cancer indications characterized by The Cancer Genome Atlas (Supplementary Table S7). This association did not appear to be an artifact driven by differences in percent tumor composition across samples. A significant association between mutation load and

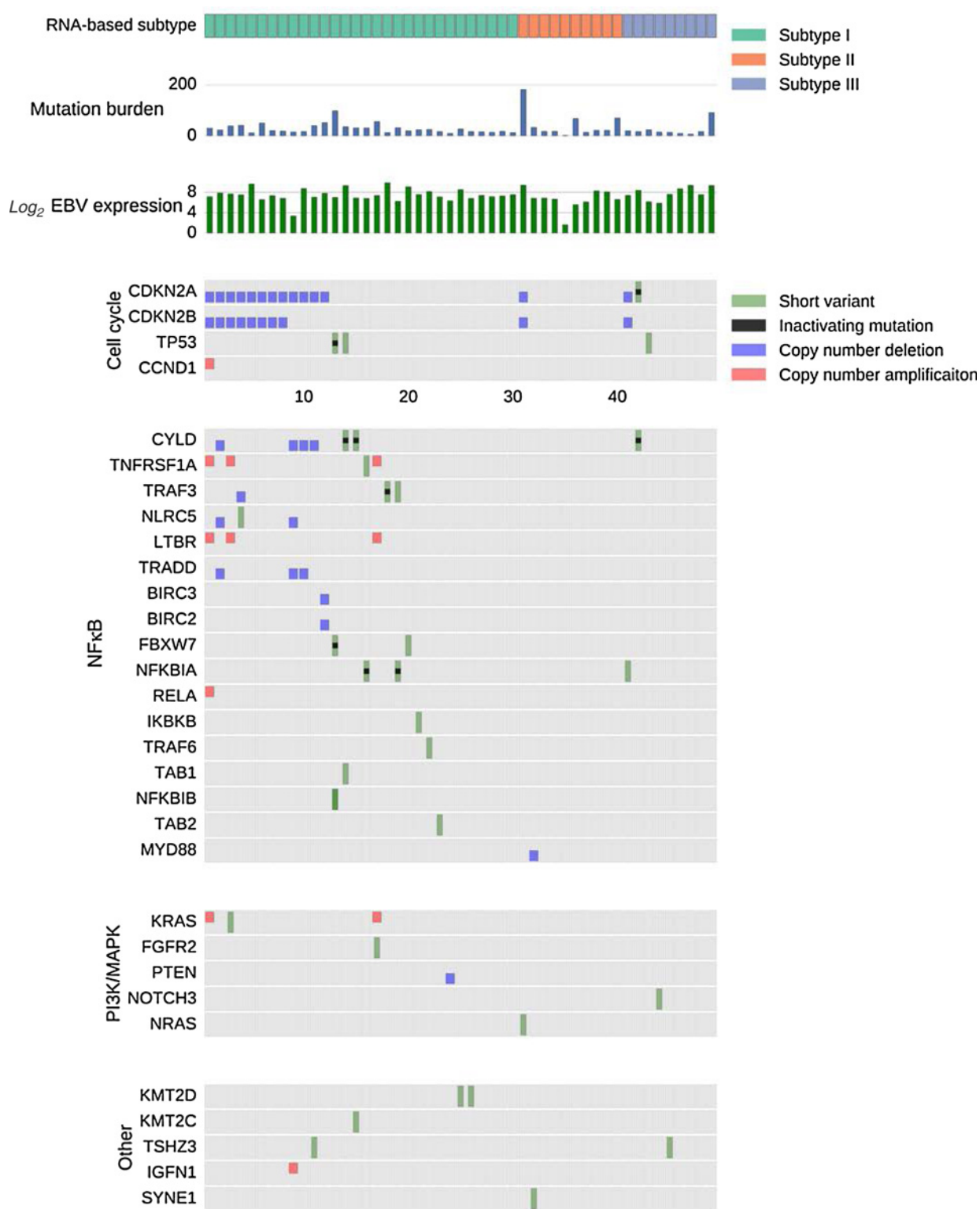
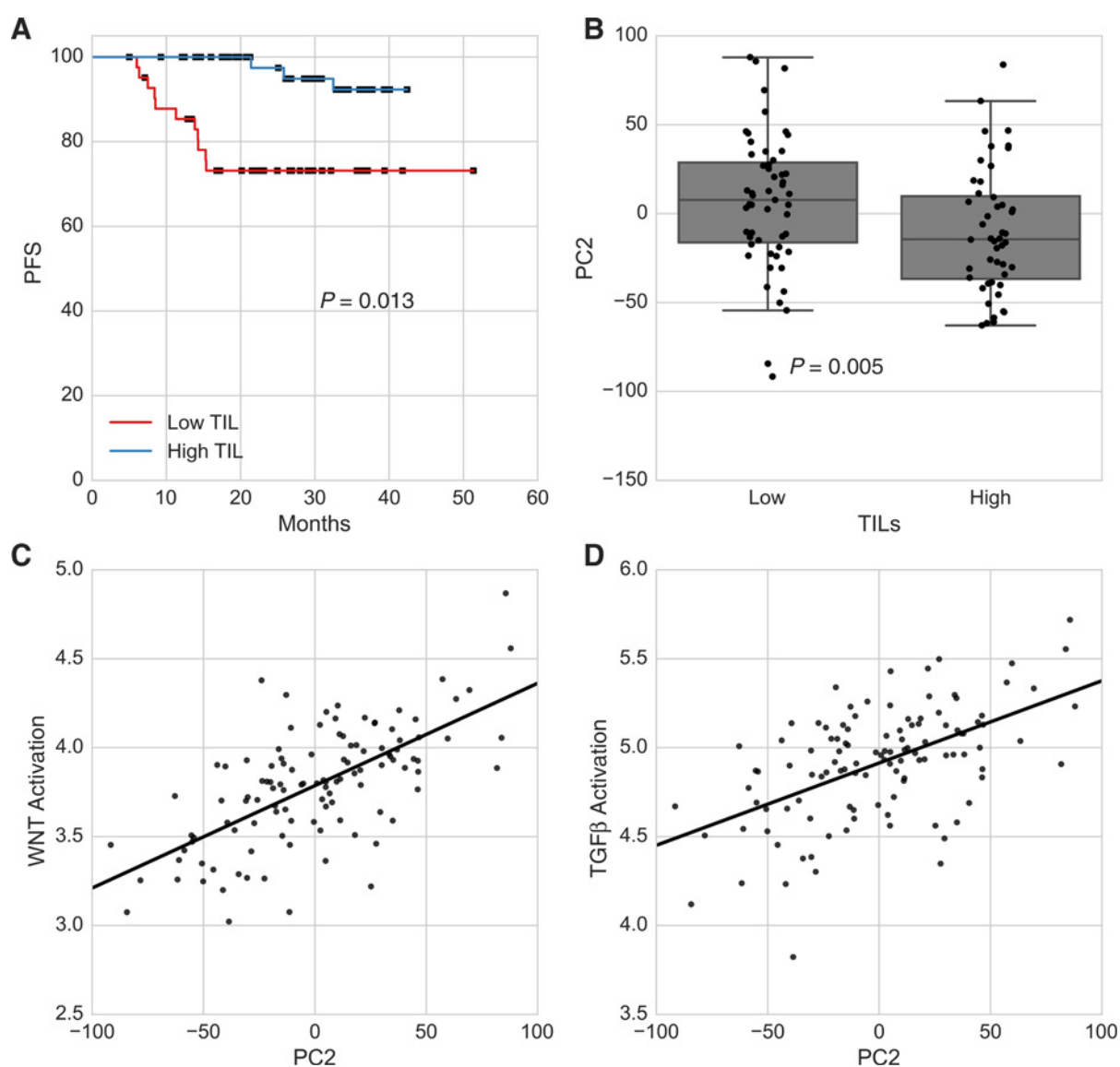


Figure 2. Mutation and somatic copy number variations for the 51 patient "paired" cohort. Deletions in the p16 locus are the most frequently observed genetic lesions. RNA-based subtype classification, total mutation burden, and total EBV gene expression are also shown for each patient.

Downloaded from <http://aacrjournals.org/mcr/article-pdf/15/12/1722/2310997/1722.pdf> by guest on 07 August 2024

**Figure 3.**

Stromal TILs and their relationship with prognosis and gene expression. **A**, High percentage of stromal TILs is associated with better progression-free survival ($P = 0.013$, log-rank test splitting samples on median). **B**, PC2 is significantly higher in samples with low stromal TILs ($P = 0.005$, MWU). **C**, PC2 correlates significantly with a signature of WNT pathway activation (Pearson $R = 0.66$, $P = 3.5e-15$). **D**, PC2 correlates strongly with a transcriptional signature of TGF β pathway activation ($R = 0.55$, $P = 4.2e-10$).

proliferation signature was observed after fitting ordinary least squares models that included sample purity estimate as either a categorical ($P = 0.003$) or real-valued ($P = 0.002$) covariate. We did not observe a strong association between percent tumor estimates and mutation load in this study (Supplementary Fig. S11). Within subtype I, but not subtypes II and III, expression of EBV transcripts *RPMS1* and *A73* ($R = -0.49$, $P = 6.9e-5$) was associated with lower proliferation signature (Fig. 4D).

Discussion

This study highlights the interplay between EBV infection, tumor genetics, and the immune microenvironment in NPC. The

three expression-based subtypes we identified are characterized by notable differences in expression of stromal and immune cell genes. Subtype I, constituting approximately 54% of the samples under study, was found to have the least, but still robustly detectable, immune and stromal cell gene expression. In this subtype, T-cell-specific genes were highly correlated with effector markers like perforin, granzymes, and IFN γ . Subtype II was characterized by very high expression of T-cell and B-cell markers, but with a pattern of expression consistent with naïve lymphocyte-rich infiltrate: IFN γ , perforin, and granzyme expression were low, and were not highly correlated with general T-cell markers. *IGHD* expression (a marker of naïve B cells) was over 4-fold higher than the other subtypes, whereas *IGHG* expression trended lower.

Table 1. Gene signature associations with prognosis. Log Hazard Ratios (HR) and corresponding nominal *P* values, and FDR-corrected *P* values are shown for Cox proportional hazards models fit to overall survival data for each gene signature

Signature	Log HR	<i>P</i>	FDR-corrected <i>P</i>
Proliferation	1.1	3.0e−3	0.049
T cell	−0.61	9.2e−3	0.063
PC1	−0.79	0.015	0.063
B cell	−0.70	0.016	0.063
PC2	0.54	0.040	0.11
Macrophage	−0.56	0.041	0.11
Fibroblast	−0.52	0.059	0.12
PC5	0.53	0.062	0.12
IFNG	−0.35	0.15	0.26
Cytotoxic	−0.35	0.18	0.28
Type I IFN	−0.22	0.40	0.56
NFκB	−0.19	0.42	0.56
PC4	−0.18	0.46	0.57
PC3	−0.10	0.70	0.80
EBV	0.04	0.89	0.95
Cilia and Mucins	−0.01	0.97	0.97

Finally, subtype III also had high expression of immune cell markers, but with the hallmarks of a preexisting antitumor or anti-EBV immune response. The cytotoxic and IFN γ signatures and *PD-L1* gene expression were all highest in this group of samples and were highly correlated with expression of general T-cell genes. A similar association between cytotoxic gene expression and oncogenic virus expression as well as *PD-L1* has been reported previously (38).

The differences in immune cell gene expression across subtypes are striking. We cannot exclude the possibility that these differences are driven in part by sampling variation during biopsy; however, other tumor characteristics that associate with subtype support the hypothesis that these groups represent bona fide molecular categories. NF- κ B pathway mutations were enriched in the less infiltrated subtype I, with 41% of these patients harboring a mutation or somatic copy number loss in a negative regulator of the pathway. We also observed enrichment for *CDKN2A* deletion in these patients, with 34% showing evidence of deletion versus 0% in subtype II and 11% in subtype III. Histopathologic analysis revealed that tumors in subtype I were enriched for a differentiated morphology. PFS appears different across subtypes with no progressions observed in subtype III and the poorest prognosis observed for patients in subtype I. EBV *BART*-associated transcript expression levels (*A73* and *RPMS1*) were substantially lower (although still robustly detectable) in subtype II. Finally, the correlates of the poorly prognostic proliferation gene expression signature were different across subtypes: In the two immune cell-infiltrated groups (II and III), total mutation load showed a strong association with proliferation, whereas in subtype I, there was no association. A previous study reported that higher mutation burden was associated with poor prognosis in NPC (20). We speculate that this may be linked to the association between mutation load and proliferation reported here. In subtype I, expression of *BART*-associated EBV transcripts *A73* and *RPMS1* were strongly anticorrelated with proliferation. This may be explained by the previously described negative regulation of *LMP1* protein expression by *BART*-derived miRNAs (39), with a subsequent reduction in *LMP1*-regulated proliferation.

NPC is known to be associated with a lymphocyte-rich stroma, and it has been suggested that the disease may rely on interactions

with these immune cells for growth and development signals (15). It has previously been reported that CD8⁺ CTL or PD-1⁺CD8⁺ T-cell infiltrate is associated with poor outcome in this disease (14, 36). The influence of the immune microenvironment on patient outcome in this cohort appears quite complex. High stromal %TILs were associated with better PFS in this study. Stromal TILs have also been found to correlate with prognosis following chemotherapy in breast cancer (31). At the gene expression level, general T-cell and B-cell signatures were associated with better prognosis, and *PDCD1* (*PD-1*) expression showed a very high ($R = 0.93$) correlation with the T-cell signature. Furthermore, patients assigned to subtype III, which had the highest expression of T-cell, IFN γ , and cytotoxic effector markers had no progressions. However, the IFN γ and cytotoxic cell signatures were not themselves significantly associated with prognosis, perhaps suggesting that the prognostic influence of CD8⁺ T cells is context or subtype dependent.

The second principal component of gene expression in this study, which correlates with TGF β and WNT transcriptional pathway activation as well as mutations in PI3K and MAPK pathway genes, showed a significant negative association with stromal %TILs. A role for TGF β in immunomodulation has been well described (40), and recent reports in melanoma have proposed a role for WNT signaling in regulating antitumor immune responses (41). These results suggest that similar mechanisms may be relevant in NPC. Intriguingly, *LMP1* gene expression was observed to anticorrelate with the expression of several WNT pathway members, consistent with a previously reported role for *LMP1* in antagonizing WNT signaling (42). Conversely, *LMP1* expression was positively correlated with immune activation markers, including *PD-L1* and the cytotoxic cell signature, but not a general T-cell or B-cell signature, perhaps pointing to a role for *LMP1* in promoting infiltration of specific lymphocyte subsets into the tumor microenvironment in part via WNT pathway downregulation.

There are several important limitations in this work that should be addressed in future studies and influence the interpretation of the results presented here. First, it will be important to validate the subtypes we propose in an independent cohort. Clinical stage and smoking history were not available for all patients, limiting our ability to examine associations between subtypes and these patient characteristics. Second, as some of the primary differences across subtypes appear to be related to stromal and immune cell infiltrate, the subtypes we report could be sensitive to sample collection protocols. Furthermore, the extent to which the gene expression signal we are measuring is tumor or immune/stromal in origin could be elucidated by approaches like laser capture microdissection coupled to expression profiling (43) or single-cell sequencing (44).

Taken together, our results show how EBV-driven expression patterns in tumor cells and tumor genetic characteristics are associated with the immune and stromal compartments in NPC, and how these characteristics may impact patient outcome. These findings may provide a framework for developing immunotherapy and targeted treatment strategies. Subtype III, with high expression of T effector genes and high IFN γ signature, may be enriched for potential responders to checkpoint blockade therapy (e.g., anti-PD-1 antibodies). The association we observed between transcriptional signatures of WNT and TGF β pathway activation and infiltration of stromal TILs perhaps suggests that therapies

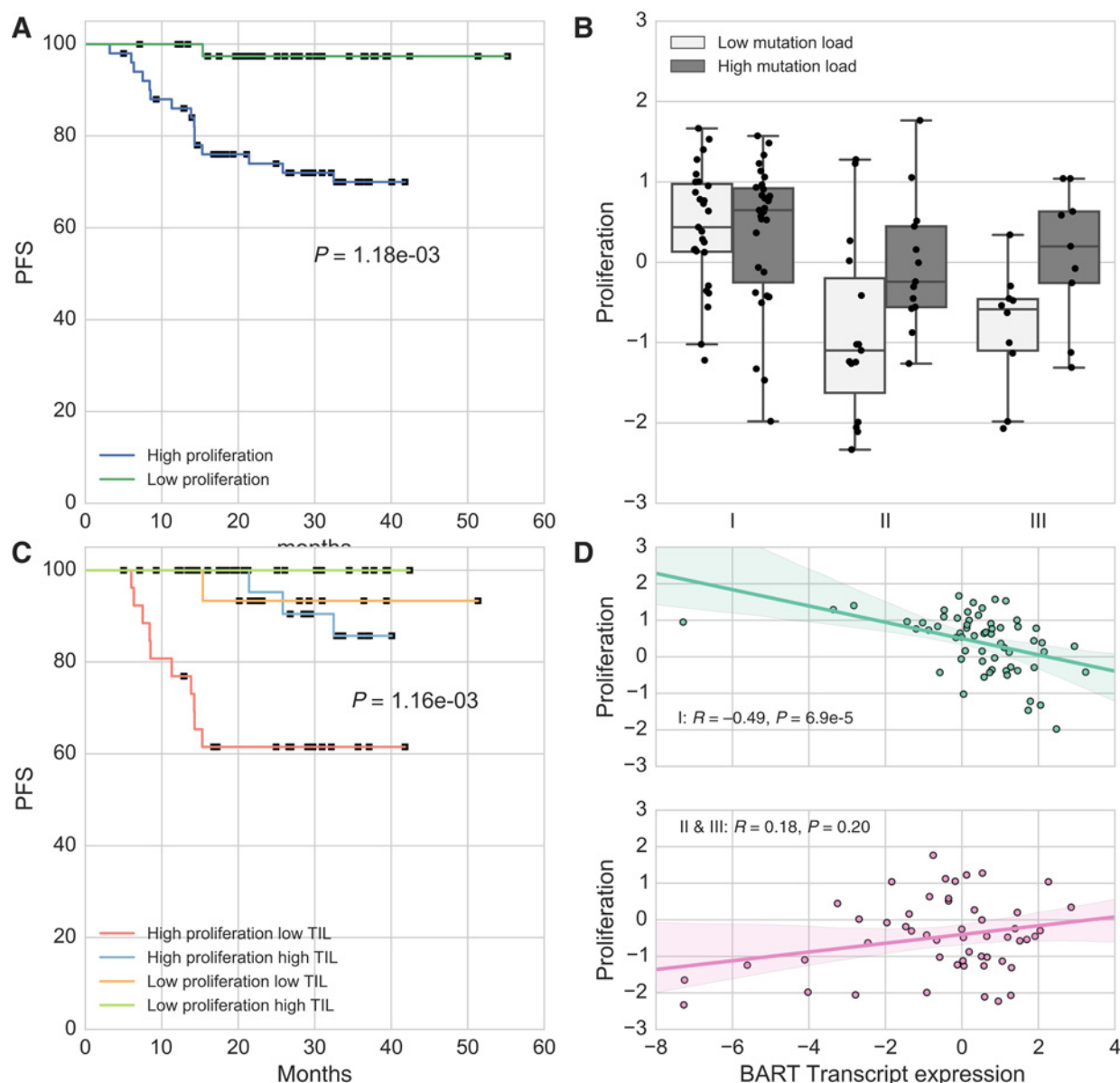


Figure 4.

A, A high proliferation signature is associated with poor PFS ($P = 1.2 \times 10^{-3}$, log-rank test splitting samples on median). **B**, High mutation burden is associated with high proliferation signature in subtypes II and III ($P = 2.1 \times 10^{-3}$, t test). **C**, Progression-free survival in patients stratified on median proliferation signature and median percent stromal TILs. **D**, Expression of BART-associated EBV transcripts is anticorrelated with the proliferation signature in subtype I.

targeting these pathways should be considered to improve infiltration of antitumor T cells into the NPC microenvironment, potentially in combination with checkpoint blockade. The strongest poorly prognostic signal in this cohort was the proliferation signature, which correlated with mutation load, a biomarker of durable response to checkpoint blockade (45–48). Therapies targeting dysregulated cell-cycle progression may therefore be of interest, perhaps in combination with immunotherapy. The large majority of patients in this cohort had no mutations or copy number losses in *RB1*, and so dual CDK4/6 inhibitors (e.g., palbociclib and ribociclib) could be considered as proliferation inhibitors in this disease.

Disclosure of Potential Conflicts of Interest

D.Y. Chiang has ownership interest (including patents) in Novartis. B. Peng is the head of TCO at Novartis. G. Dranoff is the global head (Exploratory Immunotherapies-Oncology) at and reports receiving a commercial research grant from Novartis. J.A. Engelman is the global head (Oncology) at Novartis. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: L. Zhang, Y. Wang, S. Jaeger, L. JeBailey, J. Zhang, A. Huang, Y.-X. Zeng, Y. Yao
Development of methodology: L. Zhang, K.D. Maclsaac, C. Xin, J.R. Dobson, D.Y. Chiang, Y. Wang, L. JeBailey, J. Zhang, Y. Yao

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Zhang, T. Zhou, P.-Y. Huang, C. Xin, Y. Wang, L. JeBailey, J. Zhang, W. Fang, Y. Huang, H. Zhao, Y. Yao

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Zhang, K.D. MacIsaac, J.R. Dobson, K. Yu, D.Y. Chiang, Y. Fan, Y. Wang, S. Jaeger, V. Krishnamurthy Radhakrishnan, L. JeBailey, P. Skewes-Cox, H. Bitter

Writing, review, and/or revision of the manuscript: L. Zhang, K.D. MacIsaac, P.-Y. Huang, J.R. Dobson, K. Yu, D.Y. Chiang, M. Pelletier, Y. Wang, S. Jaeger, V. Krishnamurthy Radhakrishnan, W. Fang, E. Li, A. Huang, G. Dranoff, P.S. Hammerman, J. Engelman, H. Bitter, Y. Yao

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L. Zhang, K.D. MacIsaac, C. Xin, Y. Zhao, Y. Yao

Study supervision: L. Zhang, S. Jaeger, B. Peng, A. Huang, Y. Yao

Other (acquired all genomics data, imaging, and clinical data): Y. Yao

Grant Support

This work was supported by Novartis Institutes for BioMedical Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 13, 2017; revised June 29, 2017; accepted August 23, 2017; published OnlineFirst August 29, 2017.

References

- Lee AW, Ng WT, Chan LL, Hung WM, Chan CC, Sze HCK, et al. Evolution of treatment for nasopharyngeal cancer—success and setback in the intensity-modulated radiotherapy era. *Radiother Oncol* 2014;110:377–84.
- Ma B, Hui EP, King A, To KF, Mo F, Leung SF, et al. A phase II study of patients with metastatic or locoregionally recurrent nasopharyngeal carcinoma and evaluation of plasma Epstein-Barr virus DNA as a biomarker of efficacy. *Cancer Chemother Pharmacol* 2008;62:59–64.
- Cao SM, Simons MJ, Qian CN. The prevalence and prevention of nasopharyngeal carcinoma in China. *Chin J Cancer* 2011;30:114–9.
- Xue WQ, Qin HD, Ruan HL, Shugart YY, Jia WH. Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011. *Am J Epidemiol* 2013;178:325–38.
- Chen L, Gallicchio L, Boyd-Lindsley K, Tao XG, Robinson KA, Lam TK, et al. Alcohol consumption and the risk of nasopharyngeal carcinoma: a systematic review. *Nutr Cancer* 2009;61:1–15.
- Jia WH, Luo XY, Feng BJ, Ruan H-L, Bei JX, Liu WS, et al. Traditional Cantonese diet and nasopharyngeal carcinoma risk: a large-scale case-control study in Guangdong, China. *BMC Cancer* 2010;10:446.
- Yuan JM, Wang XL, Xiang YB, Gao YT, Ross RK, Yu MC. Preserved foods in relation to risk of nasopharyngeal carcinoma in Shanghai, China. *Int J Cancer* 2000;85:358–63.
- Guo X, Johnson RC, Deng H, Liao J, Guan L, Nelson GW, et al. Evaluation of nonviral risk factors for nasopharyngeal carcinoma in a high-risk population of Southern China. *Int J Cancer* 2009;124:2942–7.
- Zheng YM, Tuppin P, Hubert A, Jeannel D, Pan YJ, Zeng Y, et al. Environmental and dietary risk factors for nasopharyngeal carcinoma: a case-control study in Zangwu County, Guangxi, China. *Br J Cancer* 1994;69:508–14.
- Thompson LD. Update on nasopharyngeal carcinoma. *Head Neck Pathol* 2007;1:81–6.
- Raab-Traub N, Flynn K, Pearson G, Huang A, Levine P, Lanier A, et al. The differentiated form of nasopharyngeal carcinoma contains Epstein-Barr virus DNA. *Int J Cancer* 1987;39:25–9.
- Young LS, Dawson CW. Epstein-Barr virus and nasopharyngeal carcinoma. *Chin J Cancer* 2014;33:581–90.
- Gallimore AP. Nasopharyngeal carcinoma. *Clin Oncol* 1995;7:388–93.
- Zhang YL, Li J, Mo HY, Qiu F, Zheng LM, Qian CN, et al. Different subsets of tumor-infiltrating lymphocytes correlate with NPC progression in different ways. *Mol Cancer* 2010;9:4.
- Gourzones C, Barjon C, Busson P. Host–tumor interactions in nasopharyngeal carcinomas. *Semin Cancer Biol* 2012;22:127–36.
- Pai S, O'Sullivan B, Abdul-Jabbar I, Peng J, Connolly G, Khanna R, et al. Nasopharyngeal carcinoma-associated Epstein-Barr virus-encoded onco-gene latent membrane protein 1 potentiates regulatory T-cell function. *Immunol Cell Biol* 2007;85:370–7.
- Wang Y, Guo Z, Shu Y, Zhou H, Wang H, Zhang W. BART miRNAs: an unimaginable force in the development of nasopharyngeal carcinoma. *Eur J Cancer Prev* 2017;26:144–50.
- Lin DC, Meng X, Hazawa M, Nagata Y, Varela AM, Xu L, et al. The genomic landscape of nasopharyngeal carcinoma. *Nat Genet* 2014;46:866–71.
- Zheng H, Dai W, Cheung AK, Ko JM, Kan R, Wong BW, et al. Whole-exome sequencing identifies multiple loss-of-function mutations of NF- κ B pathway regulators in nasopharyngeal carcinoma. *Proc Natl Acad Sci U S A* 2016;113:11283–8.
- Li YY, Chung GT, Lui VW, To KF, Ma BB, Chow C, et al. Exome and genome sequencing of nasopharynx cancer identifies NF- κ B pathway activating mutations. *Nat Commun* 2017;8:14121.
- Bruce JP, Yip K, Bratman SV, Ito E, Liu FF. Nasopharyngeal cancer: molecular landscape. *J Clin Oncol* 2015;33:3346–55.
- Zhao Q, Fránti P. WB-index: A sum-of-squares based index for cluster validity. *Data Knowl Eng* 2014;92:77–89.
- Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio*. Ithaca, NY: Cornell University Library; 2013[cited 2016 Nov 26]. Available from: <http://arxiv.org/abs/1303.3997>.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–9.
- Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009;25:2865–71.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 2012;6:80–92.
- Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015;43:D805–811.
- Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 2013;34:E2393–2402.
- Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs working group 2014. *Ann Oncol* 2015;26:259–71.
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40.
- Sun SC. CYLD: a tumor suppressor deubiquitinase regulating NF- κ B activation and diverse biological processes. *Cell Death Differ* 2009;17:25–34.
- Liao G, Zhang M, Harhaj EW, Sun SC. Regulation of the NF- κ B inducing kinase by TRAF3-induced degradation. *J Biol Chem* 2004;279:26243–50.
- Arabi A, Ullah K, Branca RM, Johansson J, Bandarra D, Haneklaus M, et al. Proteomic screen reveals Fbw7 as a modulator of the NF- κ B pathway. *Nat Commun* 2012;3:976.
- Hsu MC, Hsiao JR, Chang KC, Wu YH, Su IJ, Jin YT, et al. Increase of programmed death-1-expressing intratumoral CD8 T cells predicts a poor prognosis for nasopharyngeal carcinoma. *Mod Pathol* 2010;23:1393–403.

37. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–25.
38. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015;160:48–61.
39. Lo AK, To KF, Lo KW, Lung RW, Hui JW, Liao G, et al. Modulation of LMP1 protein expression by EBV-encoded microRNAs. *Proc Natl Acad Sci U S A* 2007;104:16164–9.
40. Pickup M, Novitskiy S, Moses HL. The roles of TGF β in the tumour microenvironment. *Nat Rev Cancer* 2013;13:788–99.
41. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic β -catenin signalling prevents anti-tumour immunity. *Nature* 2015;523:231–5.
42. QingLing Z, LiNa Y, Li L, Shuang W, YuFang Y, Yi D, et al. LMP1 antagonizes WNT/ β -catenin signalling through inhibition of WIX and promotes nasopharyngeal dysplasia but not tumorigenesis in LMP1(B95-8) transgenic mice. *J Pathol* 2011;223:574–83.
43. Sriuranpong V, Mutirangura A, Gillespie JW, Patel V, Amornphimoltham P, Molinolo AA, et al. Global gene expression profile of nasopharyngeal carcinoma by laser capture microdissection and complementary DNA microarrays. *Clin Cancer Res* 2004;10:4944–58.
44. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016;352:189–96.
45. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 2014;371:2189–99.
46. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 2015;348:124–8.
47. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* 2016;387:1909–20.
48. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 2015;350:207–11.