Supplementary materials

Clinical specimens

All patients received radical intensity-modulated radiation therapy (IMRT). The cumulative radiation doses delivered to the primary tumor were 66 Gy or greater and that delivered to the involved neck area were 60–70 Gy. The daily fraction dose was 2.00 Gy to 2.34 Gy. Platinum-based chemotherapy was administered to 91.2% (454 of 498) of the patients, including induction chemotherapy (IC), concomitant chemoradiotherapy (CCRT), and adjuvant chemotherapy (AC). 40.4% (201/498) of patients received IC+CCRT and 37.6% (187/498) received CCRT in the combined cohort. IC consisted of cisplatin with 5-fluorouracil or taxanes, or both every 3 weeks for two or three cycles. Concurrent cisplatin-based chemotherapy (every three weeks or weekly) was administered during radiotherapy.

DNA Extraction from Plasma Samples

Samples of peripheral blood (5 mL) were collected in an ethylenediamine tetraacetic acid tube from all participants and were centrifuged at ×1600g for isolation of plasma. Plasma samples were transferred carefully into plain polypropylene tubes for storage at －80 ° C until further processing. DNA from plasma samples was extracted with the QIAamp Blood Kit (Qiagen, Hilden, Germany) using the blood and body fluid protocol, as recommended by the manufacturer. In total, 500 －1000μL of each plasma sample were used for DNA extraction per column, with a final elution volume of 50μL from the extraction column. The exact amounts were documented for the calculation of the target DNA concentration.

Liu S-L, *et al. J Immunother Cancer* 2020; 8:e000205. doi: 10.1136/jitc-2019-000205

Real-Time Quantitative EBV DNA PCR

The real-time quantitative PCR system was developed for EBV DNA detection toward the BamHI-W region. The system consisted of the amplification primers W-44F (5'-AGT CTC TGC CTC CAG GCA-3') and W-119R (5'-ACA GAG GGC CTG TCC ACC G-3') and the dual-labeled fluorescent probe W-67T (5'- [FAM] CAC TGT CTG TAA AGT CCA GCC TCC [TAMRA]-3'). In this study, real-time quantitative PCR for the β-actin gene was used as a control for the amplifiability of plasma DNA. Theβ-actin gene primer sequence was forward, 5'-ACA GGC ACCA GGG CGT GA TGG-3'; and reverse, 5'-CTC CAT GTC GTC CCA GTT GGT-3'; and the dual-labeled fluorescent probe sequence (5'-[FAM] CAT CCT CAC CCT GAA GTA CCC CAT C [TAMRA]-3'). In addition, all samples in our hospital were analysed in the same clinical lab of our institute.

Antibody validation and standardization

1. Typical positive samples (cases with high expression) were selected to verify the antibody, mainly to test the specificity, pattern and appropriate concentration for Multiplex immunohistochemistry (mIHC). 2.The verified antibodies were applied for the mIHC. The optimization was conducted in the aspects of antibody concentration, dye collocation, dyeing sequence and repair conditions. Among them, antibody concentration and dye collocation were primarily tested to ensure success following the criteria that no string color interference exists and the pattern of multiple dyeing is consistent with that of single dyeing. 3.The stability of mIHC was verified with 5-10 randomly selected clinical samples. Individual antibody concentrations were

fine-tuned and the final system was used for formal experiment. 4. Before the whole experiment began, a mIHC platform was constructed and the stability of the platform was verified by a variety of antibodies. All the reagents except the primary antibody remained consistent during the experiment. During the experiment, positive samples were used for parallel staining to ensure the reliability of staining results.

Identification of intratumor and stromal tissue

1. Initial classification of samples: The site for sampling and the division of tumor and stromal was identified by experienced pathologists first, then the tumor chip and stromal chip were prepared respectively with the help of HE staining. 2. Accurate classification of tissue: The algorithm was constructed by Inform following the steps of check-train-confirm. 2 categories of tissue, TUMOR and STROMA, were constructed. In the process of choose components for training, all markers were selected as reference. Large Pattern Scale, and Medium Segmentation Resolution were selected. The optimization was determined while the results could meet the conclusion of manual interpretation to the greatest extent. 3. Repeatability test: two independent algorithms constructed by two analysts were used for parallel comparison of two chips, and the organizational split results were basically the same. 4. Manual secondary review: the constructed algorithm was used to batch all samples, and the partition results were checked one by one. For samples with large deviations, manual secondary division was conducted.

Multispectral imaging and scoring for immune markers

The Polaris System (PerkinElmer, Waltham, Massachusetts, US) was used to image

the multiplex-stained sections: In each section, we randomly acquired 20 fields of view at 200 × in the multispectral images. Then, for quantitative digital analysis, a group of raw multispectral images were generated. The image were captured at 20-nm wavelength intervals from 420 to 720 nm and then combined to build each image cube. For each marker, images of a single stained section and those of an unstained section were used to extract the spectrum of each fluorophore and that of the tissue autofluorescence to create a spectral library. This library was then applied to separate each multispectral image cube into its individual components (termed spectral unmixing), which enabled the color-based identification of markers of interest in a single image using the inForm image analysis software (Version 2.4, PerkinElmer, Waltham, Massachusetts, US).[1] The inForm software is capable of actively learning the phenotyping algorithm from all spectrally unmixed images. In addition, each DAPI-stained cell was identified individually using a combination of its fluorophore characteristics and cell morphology features associated with segmented nuclei (DAPI signal). Therefore, we obtained two sets of intratumoral and interstitial data. Two researchers, who had no previous knowledge of the clinical and follow-up data, performed all phenotyping and subsequent quantification. The inForm software was used to assess the absolute percentage of cells demonstrating with single marker or co-expression markers staining. The combination of the percentage and intensity of positively stained cells was used to score immune markers to generate an H-score.[2, 3]. This was calculated using the following equation: H-score = $\sum$ Pi (i + 1), where i was the intensity of the stained cells (0–4+), and Pi was the percentage of stained cells for

each intensity, which varied from 0 to 100%. The mean positivity value across images acquired from 20 fields of view was calculated.

Predictive model construction

The least absolute shrinkage and selection operator (LASSO) has been widely applied for logistic modeling in survival analysis using high-dimensional data; therefore, we adopted a penalized logistic model with a LASSO penalty to select prognostic factors. The optimal values of λ was determined using 10-fold cross-validation.[4] We chose λ via the minimum (standard error) criteria. Herein, we plotted the mean-squared error versus log (λ), where λ was the tuning parameter. We then constructed the prediction model in which the coefficients were weighted by the logistic model in the training cohort. The X-tile software (version 3·6·1, Yale University, CT) was used in the training cohort to select the optimal cut-off value. To separate the patients into high and low risk groups, we used the thresholds for the score outputted from the predictive model, which was determined using the score that produced the largest Chi-squared value in the Mantel-Cox test.[5] The final model was a linear equation that comprised the H-score or percentage of the four markers multiplied by their respective regression coefficients (Table S5). Patients with a risk score < 0.8 were classified as low-risk, while those with a score ≥ 0.8 were classified as high-risk.

References

1. Wang X, Zhang T, Song Z, et al. Tumor CD73/A2aR adenosine immunosuppressive axis and tumor-infiltrating lymphocytes in diffuse large B-cell lymphoma: correlations with clinicopathological characteristics and clinical outcome. *International journal of cancer* 2019; **145**(5): 1414-22.

2. Budwit-Novotny DA, McCarty KS, Cox EB, et al. Immunohistochemical analyses of estrogen receptor in endometrial adenocarcinoma using a monoclonal antibody. *Cancer research* 1986; **46**(10): 5419-25.

3. McClelland RA, Finlay P, Walker KJ, et al. Automated quantitation of immunocytochemically localized estrogen receptors in human breast cancer. *Cancer research* 1990; **50**(12): 3545-50.

4. Tang XR, Li YQ, Liang SB, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. *The Lancet Oncology* 2018; **19**(3): 382-93.

5. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2004; **10**(21): 7252-9.