

RESEARCH

Open Access



ATHENA: an independently validated autophagy-related epigenetic prognostic prediction model of head and neck squamous cell carcinoma

Ziang Xu^{1,2†}, Xinlei Chen^{1,2†}, Xiaomeng Song^{1,3†}, Xinxin Kong^{4†}, Jiajin Chen⁴, Yunjie Song⁴, Maojie Xue⁴, Lin Qiu^{1,2}, Mingzhu Geng^{1,2}, Changyue Xue^{1,5*}, Wei Zhang^{1,2*} and Ruyang Zhang^{4,6,7*}

Abstract

The majority of these existing prognostic models of head and neck squamous cell carcinoma (HNSCC) have unsatisfactory prediction accuracy since they solely utilize demographic and clinical information. Leveraged by autophagy-related epigenetic biomarkers, we aim to develop a better prognostic prediction model of HNSCC incorporating CpG probes with either main effects or gene–gene interactions. Based on DNA methylation data from three independent cohorts, we applied a 3-D analysis strategy to develop An independently validated auTophagy-related epigenetic prognostic prediction model of HEad and Neck squamous cell carcinomA (ATHENA). Compared to prediction models with only demographic and clinical information, ATHENA has substantially improved discriminative ability, prediction accuracy and more clinical net benefits, and shows robustness in different subpopulations, as well as external populations. Besides, epigenetic score of ATHENA is significantly associated with tumor immune microenvironment, tumor-infiltrating immune cell abundances, immune checkpoints, somatic mutation and immunity-related drugs. Taken together these results, ATHENA has the demonstrated feasibility and utility of predicting HNSCC survival (<http://bigdata.njmu.edu.cn/ATHENA/>).

Keywords Head and neck squamous cell carcinoma, DNA methylation, Gene–gene interaction, Autophagy, Prognostic prediction, Immune landscape

[†]Ziang Xu, Xinlei Chen, Xiaomeng Song and Xinxin Kong have contributed equally to this work.

*Correspondence:

Changyue Xue
xuechangyue@njmu.edu.cn
Wei Zhang
zhangwei0508@njmu.edu.cn
Ruyang Zhang
zhangruiyang@njmu.edu.cn

¹ Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, 136 Hanzhong Road, Nanjing 210029, Jiangsu, China

² Department of Oral Special Consultation, Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

³ Department of Oral and Maxillofacial Surgery, Affiliated Hospital of Stomatology, Nanjing Medical University, Nanjing, Jiangsu, China

⁴ Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, SPH Building Room 406, 101 Longmian Avenue, Nanjing 211166, Jiangsu, China

⁵ Department of Implant Dentistry, Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

⁶ China International Cooperation Center for Environment and Human Health, Nanjing Medical University, Nanjing, China

⁷ Changzhou Medical Center, Nanjing Medical University, Changzhou 213164, Jiangsu, China



Introduction

Head and neck squamous cell carcinoma (HNSCC) is an aggressive malignancy that includes a wide range of phenotypes such as cancers of the lip, oral cavity, larynx, nasopharynx, oropharynx and hypopharynx, which results in nearly 900,000 new cases and 450,000 deaths globally in year 2020 [1]. Despite recent breakthroughs in surgery, radiotherapy, chemotherapy, targeted therapy, and immunotherapy, the prognosis of HNSCC is still poor and the 5-year survival rate of HNSCC stagnates at about 50% [2]. In the past decades, great efforts have been made to carry out genetic [3], epigenetic [4], transcriptomic [5] and proteomic [6] studies of HNSCC survival, since effective biomarkers possess the capability to predict prognosis of the disease, and can help to diagnose disease in its early stage, which is essential to improve the overall survival of HNSCC. Therefore, a significant amount of HNSCC associated molecular biomarkers has emerged [7].

Autophagy is the collective term covering a number of catabolic pathways that regulate cellular homeostasis via lysosomal degradation and recycling of cytoplasmic components [8]. The role of autophagy in regulating cancer progression is complex and contradictory; its specific function depends on the cancer type and tumor stage [9, 10]. Autophagy could negatively or positively regulate cancer immunotherapy by degrading immune checkpoint protein, releasing pro-inflammatory cytokines, and generating or degrading antigens [11]. Change of autophagic flux is associated with cancer cell proliferation and metastasis [12], tumor stem cell phenotype [10], tumor malignancy [13], and lymph node metastasis [14] in HNSCC. During HNSCC treatment, regulation of autophagy may modulate cisplatin resistance [15, 16], and help overcome radiotherapy resistance [17].

DNA methylation is a heritable, reversible and one of the most fundamental epigenetic modifications, which regulates gene transcription [18]. Aberrant DNA methylation is also involved in the progression of cancer [19], and tracking the aberrant methylation contributes substantially to the prognostic prediction of cancer survival [20]. However, the effect of DNA methylation of autophagy-related genes (ARGs) on HNSCC survival requires further investigation. Furthermore, almost all prognostic models of HNSCC merely focus on predictors with main effects, but overlook predictors exhibiting gene–gene ($G \times G$) interactions, which may also contribute to discovery of therapeutic targets and boost prognostic prediction accuracy [21, 22].

Hence, we performed a two-step designed study to develop An independently validated auTophagy-related prognostic prediction model of HEad and Neck squamous cell carcinoma (ATHENA) by incorporating

epigenetic biomarkers with main effects and $G \times G$ interactions using all available data in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), and also analyzed the relationships between the epigenetic predictors and immune landscape.

Methods

Study populations with DNA methylation data

The level-3 TCGA-HNSCC DNA methylation and clinical data is obtained from the UCSC XENA browser, whose tumor sites are mostly tongue, larynx or overlapped lesions of lip, oral cavity and pharynx. Two additional independent datasets with clinical and DNA methylation information are downloaded from GEO (GSE75537 [23] and GSE52793 [24]). GSE75537 includes tumor samples from 53 tongue squamous cell carcinomas, while GSE52793 is consisted of oral rinse samples from 82 oral squamous cell carcinoma patients.

Quality control process for DNA methylation data

DNA methylation is assessed by the Illumina Infinium Human Methylation 450 Array. We use R package *CHAMP* to process level-3 data from TCGA and GEO. Ineligible CpG probes are removed if they met any of the quality control (QC) criteria: (i) non-CpG probes, (ii) common single nucleotide polymorphisms (SNPs) located in the position of the CpG probe or 10 bp flanking regions, (iii) cross-reactive probes, (iv) sex chromosome probes, (v) deletion rates >20%, (vi) failed QC in either TCGA or GEO cohorts. Type I and II probe correction is processed using Beta-Mixture Quantile (BMIQ) normalization. Additional file 1: Figure S1 describes the details of the QC process. Subjects without overall survival time are also removed. Finally, 634 subjects (Table 1) and 361,065 CpG probes are remained in the subsequent association analysis.

Study populations with gene expression data and somatic mutation data

In the TCGA cohort, 499 HNSCC patients have complete mRNA sequencing data and 493 patients have complete somatic mutation data. TCGA mRNA sequencing data processing and quality control are performed by the TCGA working group. Level-3 mRNA expression data that downloaded from the UCSC XENA database is composed of fragments per kilobase per million mapped reads (FPKM) values, and is transformed into transcripts per kilobase million (TPM) values for association analysis. The expression value of each gene is also transformed on a \log_2 scale before association analysis.

Table 1 Demographic and clinical characteristics of subjects in different datasets

Biomarker screening Model development Characteristics	Discovery phase	Validation phase		Combined dataset
	Training set	Internal testing set	External testing set	
	TCGA	GSE75537	GSE52793	
Number of samples	499	53	82	634
Age at diagnosis (years)	61.08 ± 11.92	49.36 ± 13.47	–	–
Age at diagnosis, N (%)				
< 40	18 (3.6)	16 (30.2)	–	34 (6.2)
40–49	58 (11.6)	15 (28.3)	–	73 (13.2)
50–59	144 (28.9)	7 (13.2)	–	151 (27.4)
≥ 60	279 (55.9)	15 (28.3)	–	294 (53.2)
Unknown	0	0	82	82
Gender, N (%)				
Male	366 (73.3)	42 (79.3)	–	408 (73.9)
Female	133 (26.7)	11 (20.7)	–	144 (26.1)
Unknown	0	0	82	82
T stage, N (%)				
I	33 (6.8)	13 (24.5)	–	46 (8.6)
II	142 (29.3)	15 (28.3)	–	157 (29.2)
III	130 (26.9)	12 (22.7)	–	142 (26.4)
IV	179 (37.0)	13 (24.5)	–	192 (35.8)
Unknown	15	0	82	97
N stage, N (%)				
0	238 (49.9)	25 (47.2)	–	263 (49.6)
1	80 (16.8)	8 (15.1)	–	88 (16.6)
2	152 (31.9)	20 (37.7)	–	172 (32.5)
3	7 (1.4)	0 (0)	–	7 (1.3)
Unknown	22	0	82	104
M stage, N (%)				
0	469 (99.0)	45 (100.0)	–	514 (99.0)
1	5 (1.0)	0 (0)	–	5 (1.0)
Unknown	25	8	82	115
Smoking status, N (%)				
Yes	378 (77.3)	–	–	378 (77.3)
No	111 (22.7)	–	–	111 (22.7)
Unknown	10	53	82	145
Race, N (%)				
White	426 (87.8)	–	–	426 (87.8)
Black or African American	47 (9.7)	–	–	47 (9.7)
Asian	10 (2.1)	–	–	10 (2.1)
American Indian or Alaska Native	2 (0.4)	–	–	2 (0.4)
Unknown	14	53	82	149

Quality control process for gene expression and somatic mutation data

After quality control, 44 HNSCC subjects with missing overall survival time or clinical information are excluded, yielding a total of 455 HNSCC subjects with complete mRNA sequencing data and 449 subjects with

complete somatic mutation data for subsequent association analysis.

Definition of autophagy-related biomarkers

We focus on a total of 232 ARGs defined by the Human Autophagy Database (HADb, <http://www.autophagy.lu/>),

which is an online database that stores a complete set of human encoded genes related to autophagy as described in the published literature. After QC, there are 4306 CpG probes for association analysis.

Statistical analysis

Model development and validation of ATHENA

We depicted the study design and workflow in Fig. 1. For the development and validation of the ATHENA, we applied a 3-D strategy which was originally proposed in our previous study [21], including Double Types of Effects, Double Steps of Screening, and Double Steps of Modeling.

- (1) *Double Types of Effects* We aimed to include epigenetic predictors with either main effects or $G \times G$ interactions. (i) To test the first type of effect (main effect), we utilized Cox proportional hazards model adjusted for covariates including age, gender, smok-

ing status and TNM stage. (ii) To test the second type of effect ($G \times G$ interaction), we again used Cox proportional hazards model adjusted for covariates aforementioned.

- (2) *Double Steps of Screening* We adopted a two-step design to scan biomarkers with either significant main effects or $G \times G$ interactions on HNSCC overall survival. (i) In the step of biomarker screening, we tested those two types of effects through Cox models aforementioned in TCGA as a discovery phase. Multiple test corrections were performed by controlling the false discovery rate (FDR) at the 5% level. To reduce the impact of outliers, we deleted methylation values out of range $\text{mean} \pm 3 \times \text{standard deviation (SD)}$, and retested these effects as a sensitivity analysis. Hazard ratios (HR) and 95% confidence intervals (CIs) were calculated for incremental methylation per 1% level. (ii) In the step of biomarker validation, we confirmed their signifi-

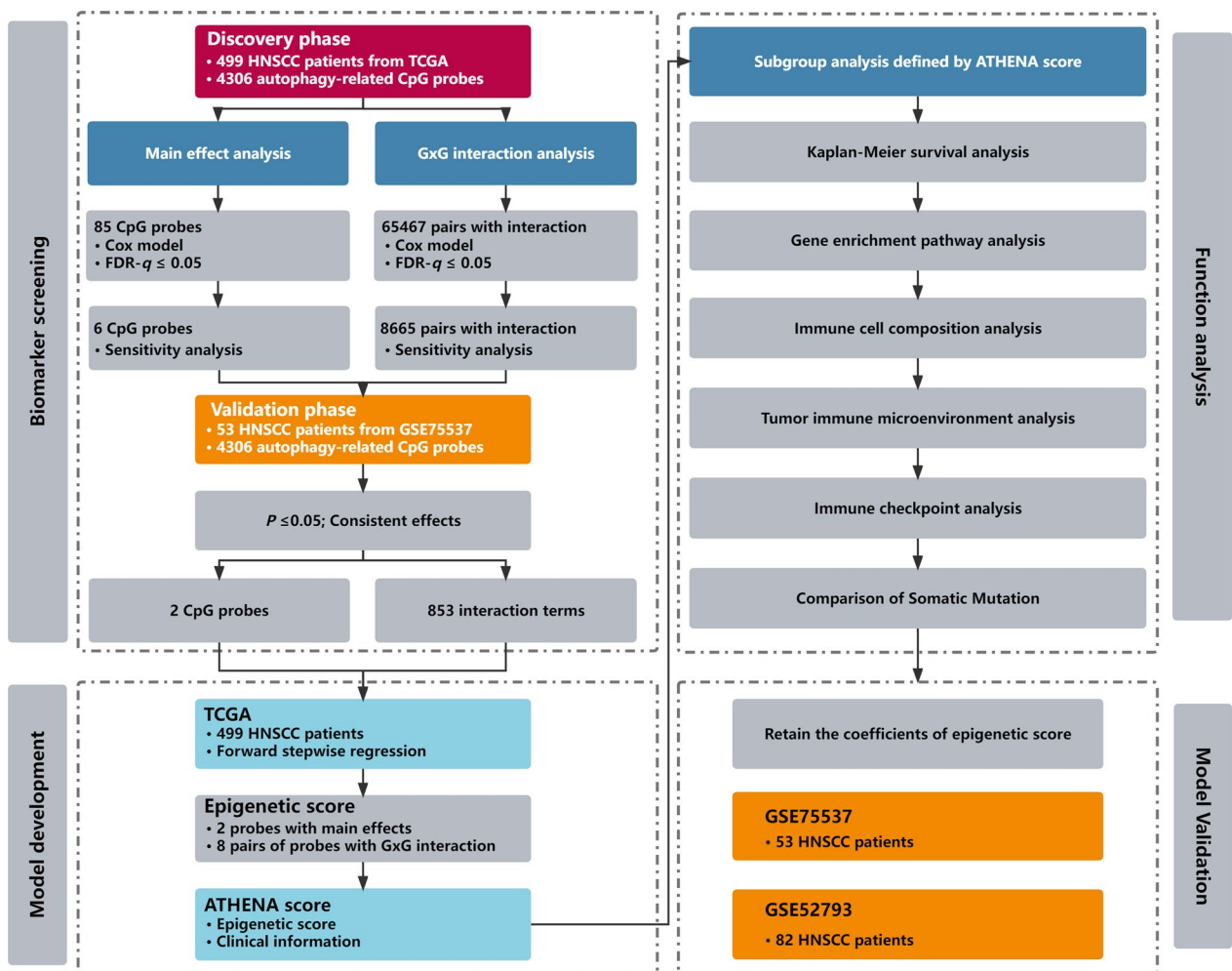


Fig. 1 Flow chart of study design and statistical analyses

cances in GSE75537 as a validation phase. Significant biomarkers were finally retained if they met all following criteria: (i) $FDR-q \leq 0.05$ in the discovery phase; (ii) $P \leq 0.005$ in the sensitivity analysis of discovery phase; (iii) $P \leq 0.05$ in the validation phase; (iv) consistent direction of effects across two phases.

- (3) *Double Steps of Modeling* (i) In the step of model development, we applied forward stepwise regression strategy to select the final predictors for ATHENA from significant biomarkers survived from *Double Steps of Screening* in TCGA data as a training set. Biomarkers retained in the multivariable Cox model were identified by the likelihood ratio test with $P_{\text{entry}} \leq 0.05$ and $P_{\text{removal}} > 0.05$. (ii) In the step of model validation for ATHENA, model performance was assessed in one internal testing set (GSE75537) and another external testing set (GSE52793) with coefficients of all epigenetic predictors the same as those trained in TCGA.

Kaplan–Meier survival curves are drawn to illustrate the survival difference among patients with different risk groups. The accuracy of prediction is represented using the time-dependent receiver operating characteristic (ROC) curve, and is measured by the area under the ROC curve (AUC), which can be obtained from R package *timeROC*. The 95% CI and P value of AUC improvement are calculated by 1000-time bootstrap resampling. The concordance index (C -index), an average accuracy of predictive survival across all follow-up years, is also utilized to estimate predictive performance. Furthermore, we performed decision curve analysis (DCA) to evaluate clinical benefits by using ATHENA to screen out these patients at high risk of death. Stratification analysis of ATHENA scores is displayed within subgroups stratified by age, gender, smoking status, TNM stage and occurrence site using the R package *forestplot*. Finally, one nomogram is generated with R package *regplot*. To facilitate application of ATHENA model, we release an online calculator (<http://bigdata.njmu.edu.cn/ATHENA/>), which can immediately return predicted survival rates and 95% CIs at any time point between 0 and 120 months when inputting values of predictors for a HNSCC patient, based on an interactive web-based Kaplan–Meier survival curve.

Immune landscape analysis of epigenetic predictors of ATHENA

Potential genes *trans*-regulated by epigenetic predictors of ATHENA are identified by genome-wide correlation analysis using linear regression model and Cox model adjusted for the same covariates aforementioned in TCGA cohort. Functional annotation

and gene enrichment pathway analysis based on Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) for potential *trans*-regulated genes are performed using R Package *WebGestaltR*. The ESTIMATE algorithm is used based on gene expression data to explore the pattern of tumor immune microenvironment (TIME) among subgroups [25], and *CIBERSORT*, a linear support vector regression-based deconvolution algorithm [26], is performed to determine the composition of 22 tumor-infiltrating immune cells (TIICs). We explored the difference of immune checkpoint expression in epigenetic score subgroups, and also performed the correlation analysis between immune checkpoints expression and epigenetic score of ATHENA. Then, based on the somatic mutation data from TCGA, we conducted a differential analysis of genomic mutations between high- and low-risk groups using R package *maftools*. Finally, we explored immunity-related drugs targeting epigenetic predictors using the DrugBank database (<https://go.drugbank.com/>) [27].

Continuous variables are summarized as mean \pm SD, while categorized variables are described by frequency (n) and proportion (%) in description analysis. All statistical analyses are performed in R software (version 4.0.3, The R Foundation for Statistical Computing, Vienna, Austria), unless otherwise specified.

Results

ATHENA model development

First, 85 CpG probes with main effects and 65,467 pairs of CpG probes with $G \times G$ interactions were identified ($FDR-q \leq 0.05$) to be possibly associated with overall survival in the discovery phase. Of them, 6 probes with main effects and 8,665 pairs of probes with $G \times G$ interactions passed the sensitivity analysis. Finally, 2 probes and 853 pairs of probes were confirmed with robust significance in the validation phase, and were defined to be candidate epigenetic predictors. By using forward stepwise regression strategy in TCGA cohort as training set, we constructed a Cox model including 2 CpG probes with main effects (Additional file 1: Table S1) and 8 pairs of CpG probes with $G \times G$ interactions (Additional file 1: Table S2), which were used to calculate the epigenetic score (Additional file 1: Table S3). The ATHENA score was defined as a weighted linear combination of epigenetic score and clinical score (Additional file 1: Tables S4, S5), where weights were coefficients derived from Cox model.

ATHENA model evaluation and validation

To evaluate the discriminative ability of biomarkers of ATHENA, patients were categorized into low- and

high-risk groups based on the median of (i) clinical score which was a weighted linear combination of demographic and clinical factors; (ii) main effect score which was a weighted linear combination of epigenetic biomarkers with significant main effects; (iii) $G \times G$ interaction score which was a weighted linear combination of epigenetic biomarkers with significant $G \times G$ interactions; (iv) epigenetic score which was a weighted linear combination of main effect score and $G \times G$ interaction score; and (v) ATHENA score which was a weighted linear combination of clinical and epigenetic scores, respectively. Compared to the low-risk group, the high-risk group was associated with worse survival in TCGA cohort, exhibiting a gradually increasing hazard ratio (HR) ($HR_{\text{Clinical score}} = 1.45$, 95% CI 1.10–1.91, $P = 7.88 \times 10^{-03}$; $HR_{\text{Main effect score}} = 1.58$, 95% CI 1.20–2.09, $P = 1.23 \times 10^{-03}$; $HR_{\text{G} \times \text{G score}} = 2.99$, 95% CI 2.22–4.02, $P = 4.54 \times 10^{-13}$; $HR_{\text{Epigenetic score}} = 3.58$, 95% CI 2.65–4.83, $P < 2.00 \times 10^{-16}$; $HR_{\text{ATHENA score}} = 3.63$, 95% CI 2.68–4.91, $P < 2.00 \times 10^{-16}$) (Fig. 2A–E). To further illustrate the discriminative ability of the ATHENA score, we categorized patients by the quintiles and the 90 percentile of the score in the TCGA cohort. We observed a dose–response association between higher-percentile groups and higher mortality risk ($HR_{\text{Level 5 vs 1}} = 8.69$, 95% CI 5.22–14.47, $P < 2.00 \times 10^{-16}$; $HR_{\text{Level 4 vs 1}} = 5.94$, 95% CI 3.69–9.56, $P = 2.23 \times 10^{-13}$; $HR_{\text{Level 3 vs 1}} = 2.79$, 95% CI 1.78–4.39, $P = 8.34 \times 10^{-06}$; $HR_{\text{Level 2 vs 1}} = 1.46$, 95% CI 0.90–2.37, $P = 1.25 \times 10^{-01}$) (Fig. 2F).

We then evaluated the prediction performance of these biomarkers. The model with only clinical score had a limited prediction ability ($AUC_{36\text{-month}} = 0.61$, $AUC_{60\text{-month}} = 0.53$, $C\text{-index} = 0.60$), and the prediction accuracy increased by adding main effect score ($AUC_{36\text{-month}} = 0.67$, $AUC_{60\text{-month}} = 0.59$, $C\text{-index} = 0.64$) or $G \times G$ interaction score ($AUC_{36\text{-month}} = 0.75$, $AUC_{60\text{-month}} = 0.72$, $C\text{-index} = 0.72$). Further, by adding epigenetic score, the AUC increased by 27.9% (95% CI 27.2–28.5%, $P < 2.00 \times 10^{-16}$) and 37.7% (95% CI 37.2–38.3%, $P < 2.00 \times 10^{-16}$) for 3-year and 5-year survival prediction, respectively (Fig. 3A, B). And, ATHENA achieved an acceptable prediction accuracy ($AUC_{36\text{-month}} = 0.78$, $AUC_{60\text{-month}} = 0.73$, $C\text{-index} = 0.73$). In further subgroup analyses in subpopulations stratified by age, gender, smoking status and TNM stage, ATHENA still presented robust discriminative ability with HR ranging from 2.61 (95% CI 2.21–3.08, $P < 2.00 \times 10^{-16}$) to 3.49 (95% CI 2.45–4.96, $P = 4.51 \times 10^{-12}$), and exhibited reasonable prediction accuracy with AUC ranging from 0.71 (95% CI 0.64–0.78) to 0.90 (95% CI 0.84–0.97) for 36-month survival prediction, and 0.67 (95% CI 0.58–0.77) to 0.86 (95% CI 0.73–0.99) for 60-month survival prediction, respectively (Fig. 4A–C). Considering the potential tissue

heterogeneity, we also evaluated ATHENA model in different subgroups by occurrence sites, and observed its robust performance (Additional file 1: Figure S2A–C). DCA showed that ATHENA presented more clinical net benefits than model with only clinical and demographic indicators (Fig. 5A–D). To facilitate application of ATHENA in clinical practice, we developed a nomogram, which estimated patients' 36- or 60-month survival (Fig. 5E).

Finally, for ATHENA model validation, we retained the coefficients of CpG probes when applying to two extra datasets (GSE75537 as an internal validation and GSE52793 as an external validation). ATHENA showed a satisfactory prediction accuracy in GSE75537 ($AUC_{36\text{-month}} = 0.82$, $AUC_{60\text{-month}} = 0.80$, $C\text{-index} = 0.75$) (Additional file 1: Figure S3A), while epigenetic score was still an independent and significant risk factor for prognosis ($HR = 1.55$, 95% CI 1.16–2.07, $P = 3.09 \times 10^{-03}$) (Additional file 1: Figure S3B). The AUC of ATHENA model in GSE52793 were limited because of the absence of covariates ($AUC_{36\text{-month}} = 0.59$, $AUC_{60\text{-month}} = 0.62$, $C\text{-index} = 0.59$) (Additional file 1: Figure S4A), while epigenetic score was again significantly associated with HNSCC overall survival ($P_{p=0, q=1} = 2.39 \times 10^{-02}$ and $P_{p=1, q=1} = 2.97 \times 10^{-02}$) as shown by Kaplan–Meier survival curves (Additional file 1: Figure S4B), which was confirmed by Harrington–Fleming test that was designed for the late or delayed effect of variable during the follow-up [28].

Trans-regulation analyses of epigenetic predictors of ATHENA and immune landscape analysis

Genome-wide *trans*-regulation analyses by the linear regression model indicated that expressions of 6507 genes were significantly *trans*-regulated by the epigenetic predictors of ATHENA ($FDR\text{-}q \leq 0.05$). Among them, 1564 genes were further significantly associated with HNSCC overall survival ($P \leq 0.05$), which were evaluated by the Cox proportional hazards model adjusted for the same covariates aforementioned. KEGG enrichment analysis categorized gene probes into 21 pathways, including classic autophagy-related pathways such as *PI3K-Akt signaling pathway*, and GO annotation identified 65 biological process pathways, 10 cellular component pathways and 14 molecular function pathways, suggesting potential biological functions (Fig. 6A–D). By extracting CpG probes of autophagy-related genes in *PI3K-Akt signaling pathway*, and testing these biomarkers using the same criteria aforementioned, again, we observed 35 pairs of CpG probes with significant $G \times G$ interactions in discovery and validation phases, which could be potential drug target therapeutics to overcome autophagy in HNSCC (Additional file 1: Table S6).

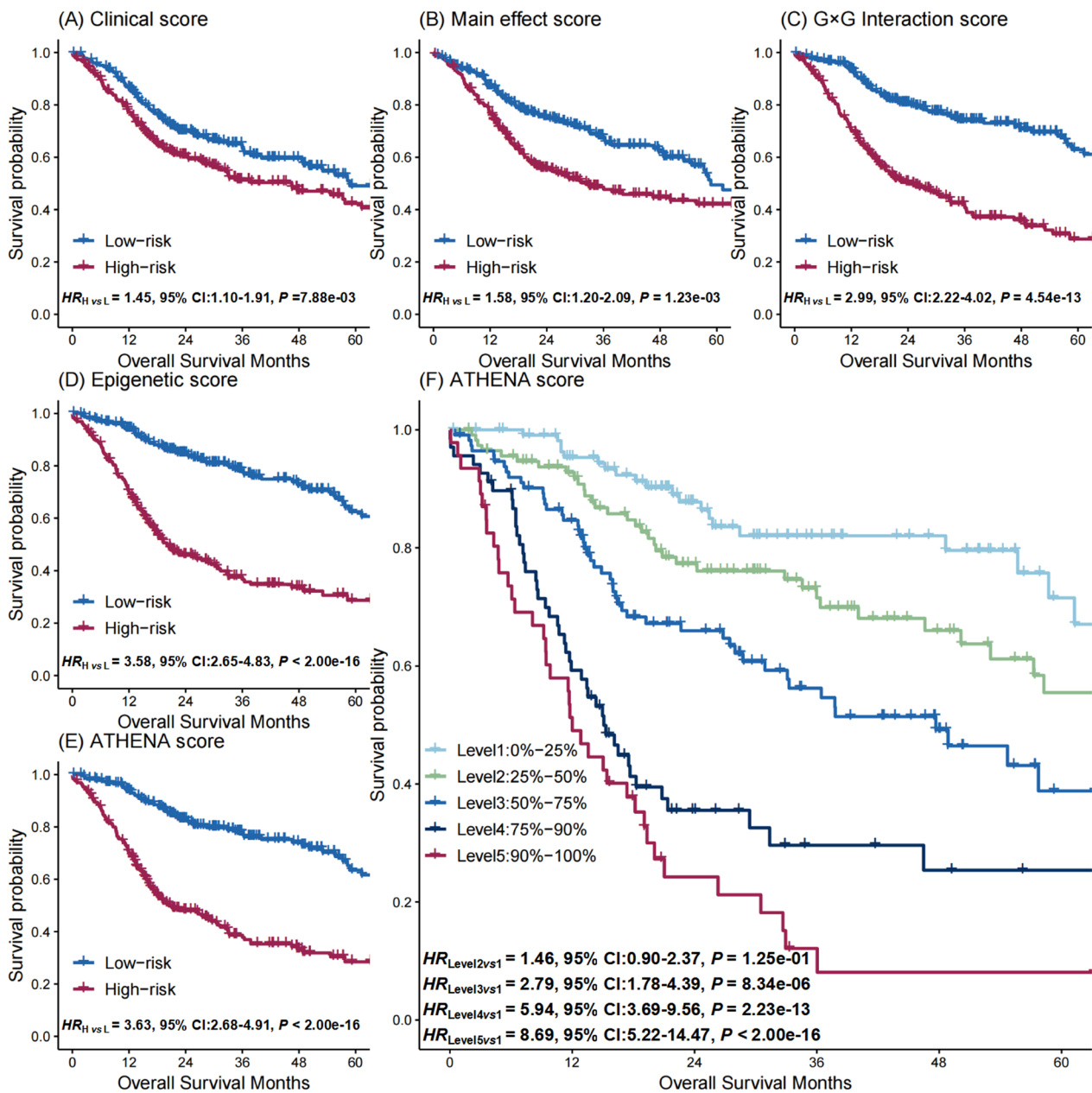


Fig. 2 Kaplan–Meier survival curves for high- and low-risk HNSCC patients. The high- and low risk groups are defined by the median of **A** clinical score, **B** main effect score, **C** G×G interaction score, **D** epigenetic score, and **E** ATHENA score. **F** Discriminative ability of the ATHENA score is illustrated by Kaplan–Meier survival curves of six groups, defined by quantiles at 20%, 40%, 60%, 80% and 90% of ATHENA score

Moreover, the compositions of 12 types of immune cells were differently distributed between low- and high-risk groups determined by median of epigenetic score of ATHENA (Fig. 7A), the correlation of epigenetic score of ATHENA with each immune cell composition varied a lot (Fig. 7B), including positive (e.g., with M0 Macrophages) and negative correlations (e.g., with Plasma cells) (Fig. 7C). Further, epigenetic score of ATHENA

had statistically significant but very weak positive correlation with the stromal score ($r=0.11, P=1.99 \times 10^{-02}$), and showed no significant negative correlation with the immune score ($r=-0.09, P=5.24 \times 10^{-02}$) (Additional file 1: Figure S5). As a result, we checked the connectivity and correlations between epigenetic score of ATHENA and immune checkpoint genes. Almost all of the immune checkpoints genes were lower expressed in the high-risk

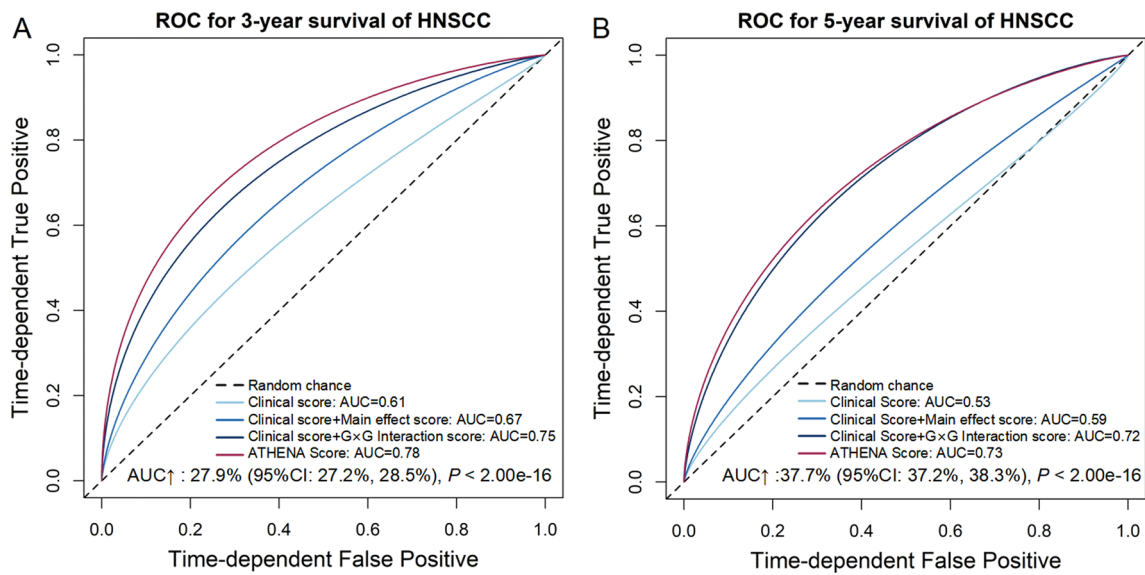


Fig. 3 ROC curves of different prognostic prediction models using different combinations of clinical information, epigenetic predictors with main effects and G×G interactions. ROC curves are presented for both **A** 36-month and **B** 60-month survival prediction. The AUC increase (%) is evaluated by comparing ATHENA model and the model with only covariates. *P* values and 95% CIs are calculated by using 1000 bootstrap samples

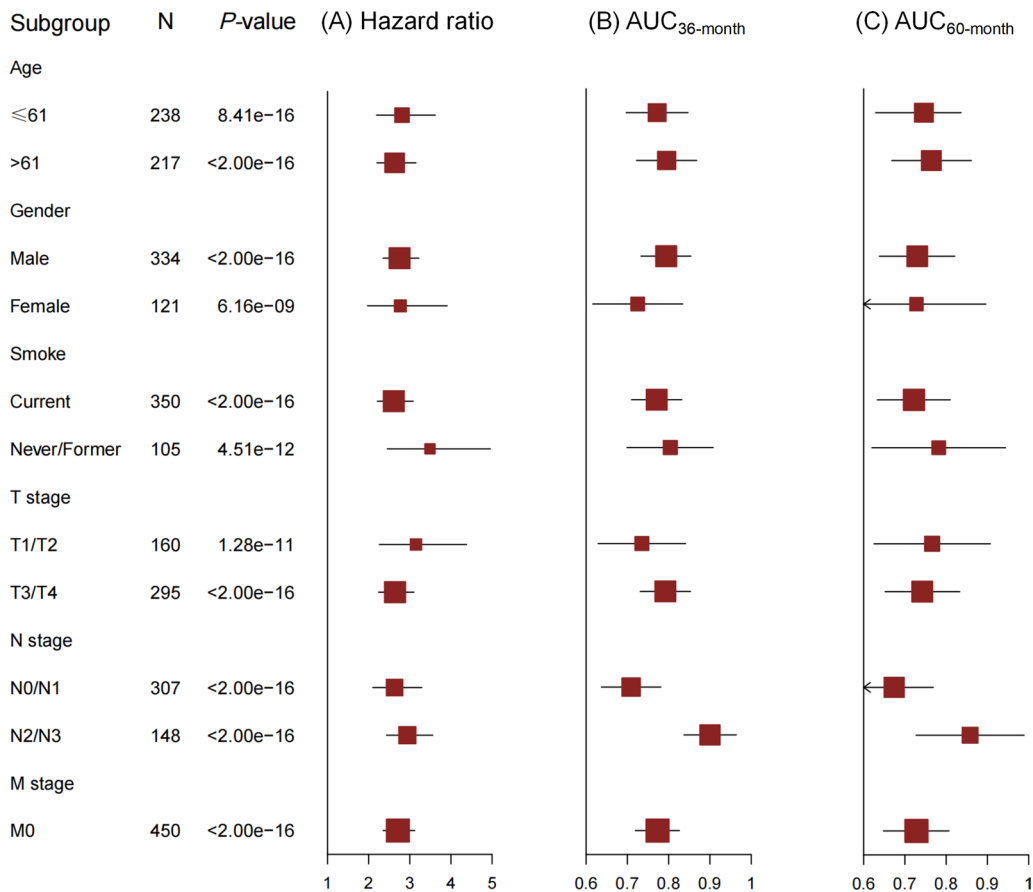


Fig. 4 Subgroup analyses of ATHENA score. **A** Hazard ratio is used to evaluate the association between ATHENA score and HNSCC survival. The AUC is used to measure the prediction accuracy of ATHENA for **B** 36-month and **C** 60-month survival

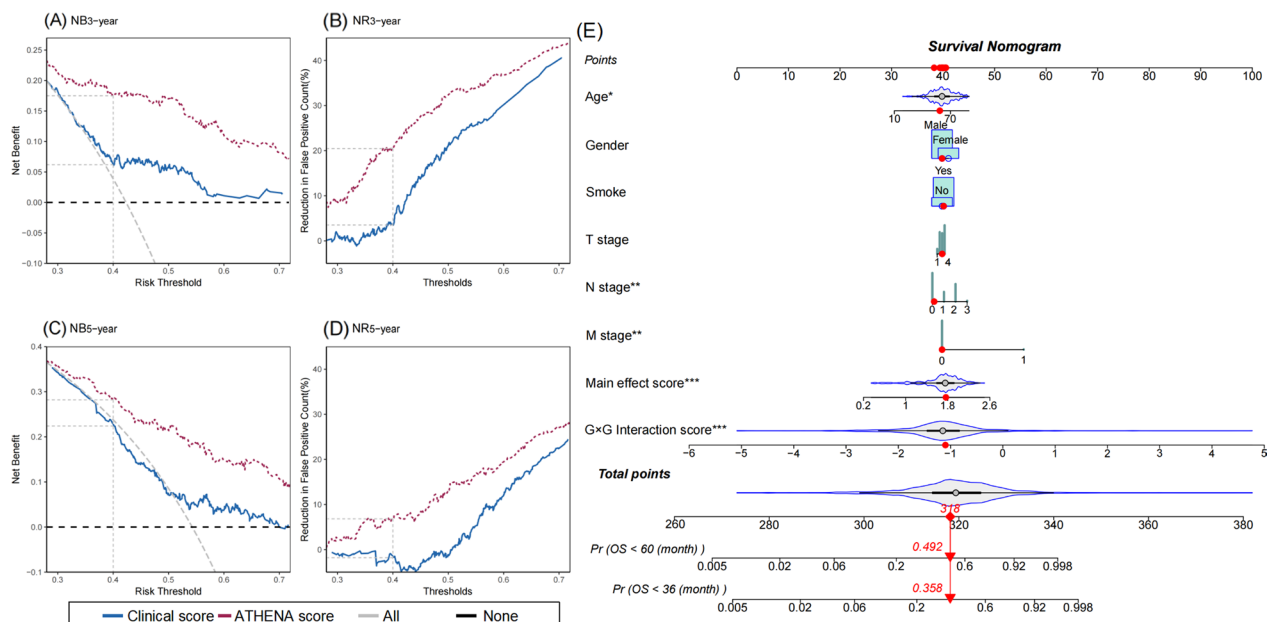


Fig. 5 Decision curve analysis and nomogram of ATHENA. The net benefit (NB) and net reduction (NR) of patients avoided unnecessary interventions are given at threshold (0.4) for both 36-month (A, B) and 60-month (C, D) survival. E For the nomogram of ATHENA model, the value of each predictor can be converted into the corresponding points according to the axis in the top of nomogram. The sum of points for each predictor can correspond to the total points axis at the bottom of the nomogram and further be used to estimate the patient’s 36- and 60-month survival rate

group (Fig. 8A), thus suggested a negative correlation between immune checkpoint gene expressions and epigenetic score of ATHENA (Fig. 8B), especially *ICOSLG* ($r = -0.30$, $P = 1.19 \times 10^{-10}$) (Fig. 8C). Then we used waterfall maps to investigate the differences in genomic mutations between the high- and low-risk groups. *TP53* (79%), *TTN* (36%), *CDKN2A* (26%), *FAT1* (24%) and *LRPIB* (19%) were the top 5 genes with the highest mutation frequencies in the high-risk group, while *TP53* (60%), *TTN* (40%), *CSMD3* (20%), *SYNE1* (19%) and *FAT1* (18%) were the top 5 genes in the low-risk group (Fig. 9A, B). Finally, numerous immunity-related drugs targeting genes, which epigenetic predictors located, have been documented (Additional file 1: Table S7), and, thereby ATHENA model may have potential roles in guiding immunotherapy.

Discussion

One major reason of limited accuracy of the prognostic model is the solely use of demographic and clinical information. These unsatisfactory models cannot accurately indicate high-risk patients who require close follow-up and postoperative adjuvant therapy [29]. Hence, there is an urgent need to develop accurate prognostic prediction models aiding in clinical decisions [30]. Using available public HNSCC epigenetic data from three independent cohorts, we employed a 3-D analysis strategy to screen

biomarkers and established an autophagy-related epigenetic prognostic prediction model, ATHENA. ATHENA showed acceptable prediction accuracy in both training and internal testing sets, and also showed fair accuracy and discrimination in external testing set with oral rinse samples of HNSCC patients, suggesting the robustness and clinical significance of ATHENA.

Gene–environment and gene–gene interactions provided additional insights into the biological mechanisms of complex diseases [31–33]. In our previous study [34], we explored epigenome-wide gene–age interaction and significantly improved the accuracy of prognostic prediction model of oral squamous cell carcinoma. ATHENA is the first attempt of an autophagy-related epigenetic prognostic model in HNSCC patients, and also one of the earliest explorations of G×G interaction of HNSCC overall survival on epigenome-wide scale. Our results showed that biomarkers with G×G interaction significantly improved the prediction accuracy of prognostic model of HNSCC and demonstrated the importance of complex association patterns among multiple factors in the study of complex diseases (e.g., HNSCC) again.

Interestingly, CpG probes located on *ITPR1* appeared several times in the interaction terms of ATHENA model. Inositol 1, 4, 5-trisphosphate receptor type 1 (*ITPR1*), located on chromosome 3, is a pivotal gene for autophagy [35, 36]. Expression of *ITPR1* can be upregulated by

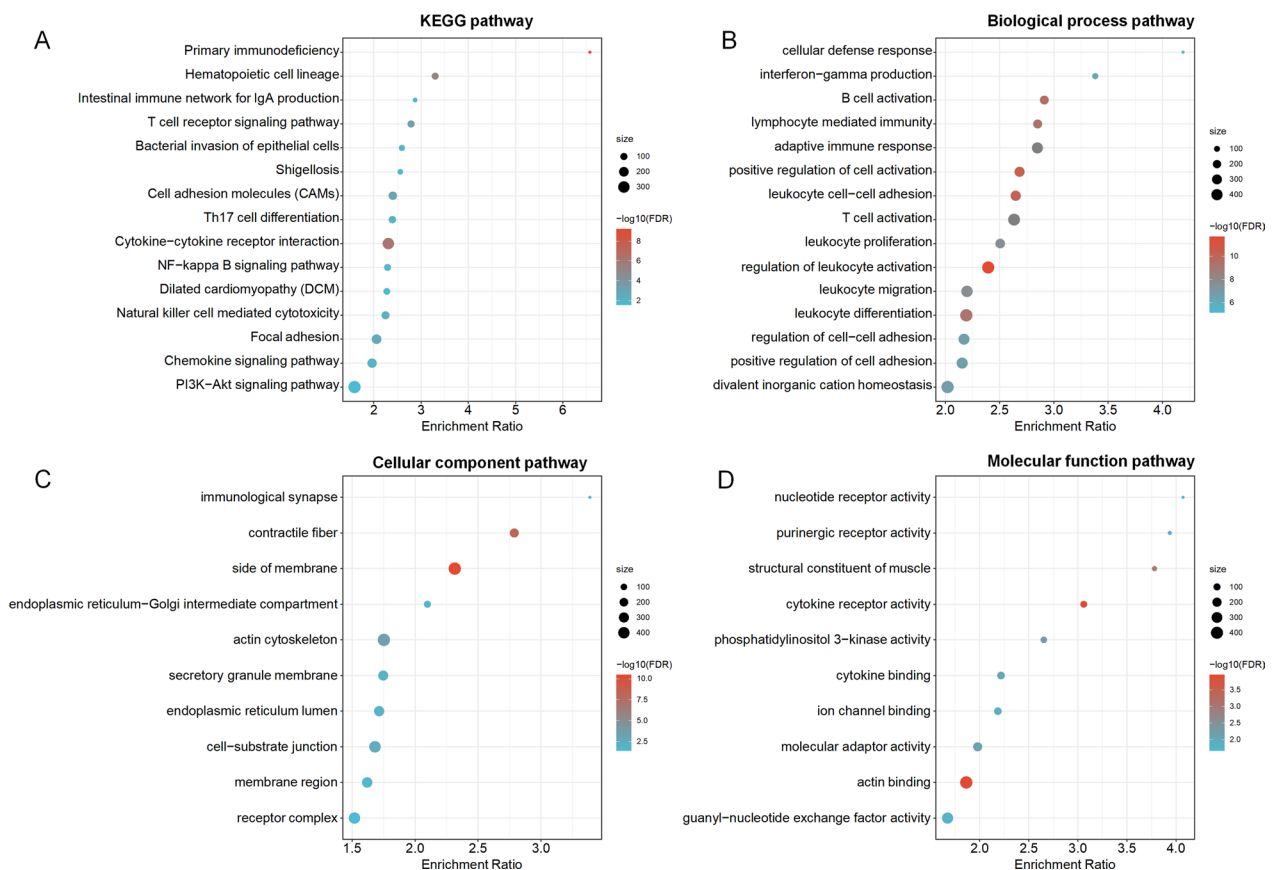


Fig. 6 Significant pathways with genes *trans*-regulated by epigenetic predictors of ATHENA in gene enrichment pathway analysis. **A** The top 15 significant KEGG pathways, **B** the top 15 significant biological process pathways, **C** the top 10 significant cellular component pathways, and **D** the top 10 significant molecular function pathways were sorted by enrichment ratio

EGOT via RNA–RNA and RNA–protein interactions to enhance autophagy [37], meanwhile, showed association with tumorigenesis of cells squamous [38] and prognosis of HNSCC patients [39]. We hypothesize that *ITPR1* may be a hub gene of epigenome-wide, and even transcriptome-wide interaction of autophagy-related genes in HNSCC. In the enrichment pathway analysis, *trans*-regulated genes were significantly enriched in autophagy-related pathways or processes (*PI3K-Akt signaling pathway*; *NF-kappa B signaling pathway*), and classic processes of the extracellular matrix (ECM) (*Focal adhesion*). This suggests that there may also be a cooperative exchange between ECM and autophagy during progression of tumor in HNSCC patients, which is proved in other kinds of neoplasms [40, 41]. Notably, in our study, epigenetic score of ATHENA has statistically significant positive correlation with stromal scores, indicating that higher stromal scores may lead to poorer HNSCC survival and suggesting that higher ECM stiffness may affect autophagy [42].

In addition, the relationship between autophagy and immunity has been widely reported. Autophagy is upregulated in many cancers, which may support the growth, survival and malignancy of neoplasm, may suppress activation of the innate immune response, and may suppress the adaptive immune response and contribute to tumor immune evasion [43], suggesting epigenetic score of ATHENA is negatively correlated with immune score. Also, autophagy may reduce the ability of T cells to kill tumor cells [44], inhibit antigen presentation [45], which explains the negative correlation between epigenetic score of ATHENA and helper follicular T cells. The correlation between the epigenetic score and immune cell composition could be also partially explained by the biological process pathways enriched in *trans*-regulated genes. For example, the negative correlation between CD8 T cells and epigenetic score may be caused by impaired T cell activation due to increased tumor malignancy. Moreover, inhibition of autophagy may enhance response of targeted therapy and blockade of immune checkpoints [46, 47]. In our study, there was

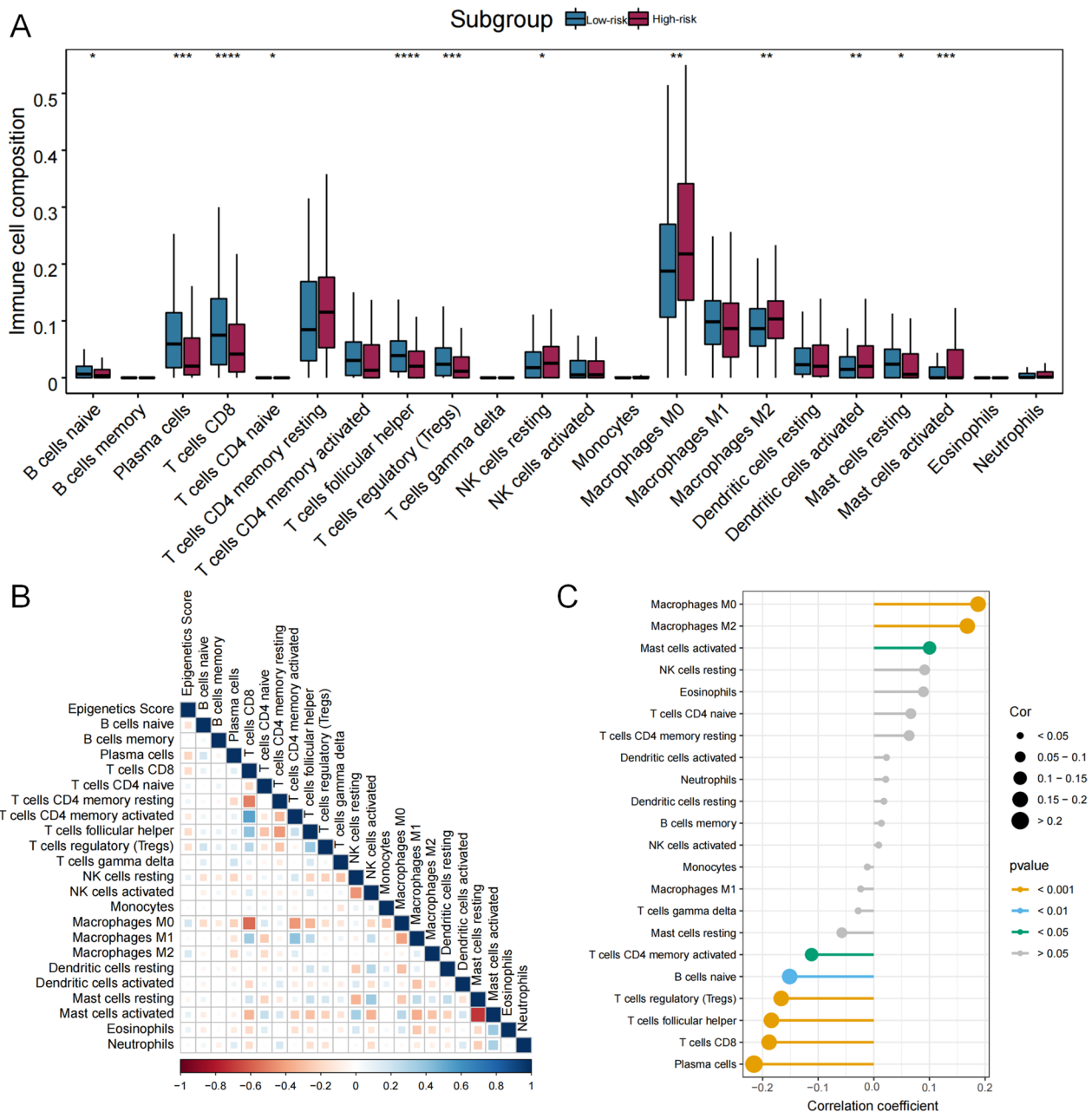


Fig. 7 The association analysis between immune cells and epigenetic score of ATHENA. **A** The abundances of 22 immune cells are compared between high- and low- risk-groups. * means $P < 0.05$, ** means $P < 0.01$, *** means $P < 0.001$ and **** means $P < 0.0001$. **B** The correlation coefficients between immune cells and epigenetic score of ATHENA are derived from Pearson correlation analyses and are presented in a heatmap. **C** The correlation coefficients between immune cells and epigenetic score of ATHENA are derived from Pearson correlation analyses and these pairs are listed in lollipop chart

also a clear inverse correlation between epigenetic score and immune checkpoint expression, which is also consistent with previous researches [48, 49], indicating that HNSCC is an immunosuppressive malignancy. Finally, genes involved in epigenetic score were transcriptional predictors with immune relevance, which can be immunotherapeutic targets.

Our study has several strengths. First, to our knowledge, this may be the first study to investigate G×G interaction on HNSCC survival on epigenome-wide scale, which provides new insights into the prognosis of HNSCC patients. Second, we adopt an effective 3-D strategy for biomarker screening and model construction, and focus on biomarkers with either significant

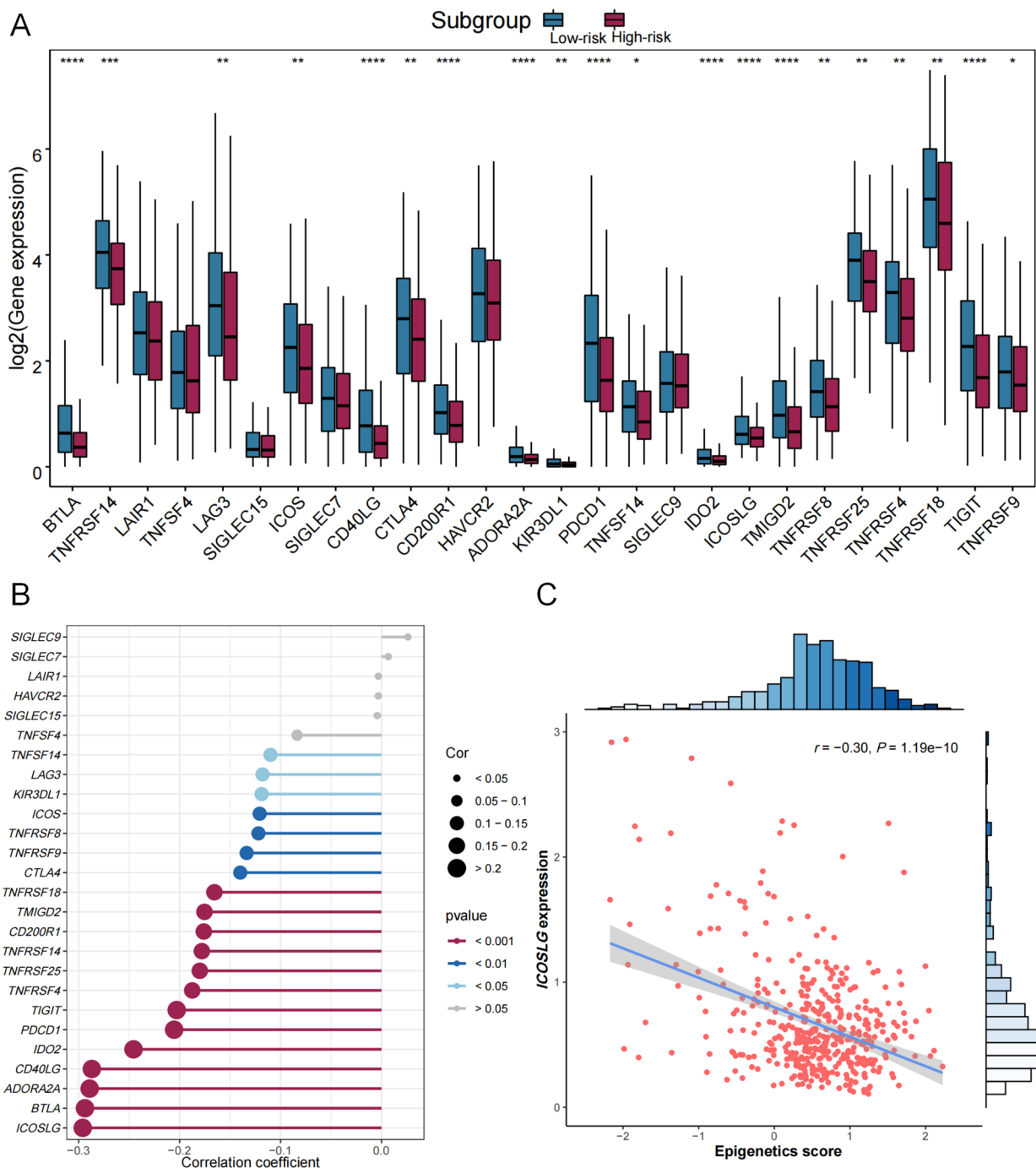


Fig. 8 The association analysis between immune checkpoints and epigenetic score of ATHENA. **A** The gene expressions of 26 immune checkpoints are compared between high- and low-risk-groups. * means $P < 0.05$, ** means $P < 0.01$, *** means $P < 0.001$ and **** means $P < 0.0001$. **B** The correlation coefficients between immune checkpoints and epigenetic score of ATHENA are derived from Pearson correlation analyses and these pairs are listed in lollipop chart. **C** The scatter plot and linear regression analysis between epigenetic score of ATHENA and expression of *ICOSLG* with the strongest association

main effects or $G \times G$ interactions, which can improve the accuracy of the prediction model. Third, we perform internal and external model validation of ATHENA,

therefore, confirm the accuracy and extrapolation of ATHENA in HNSCC patients. Finally, we provide a web-based tool to facilitate the application of ATHENA.

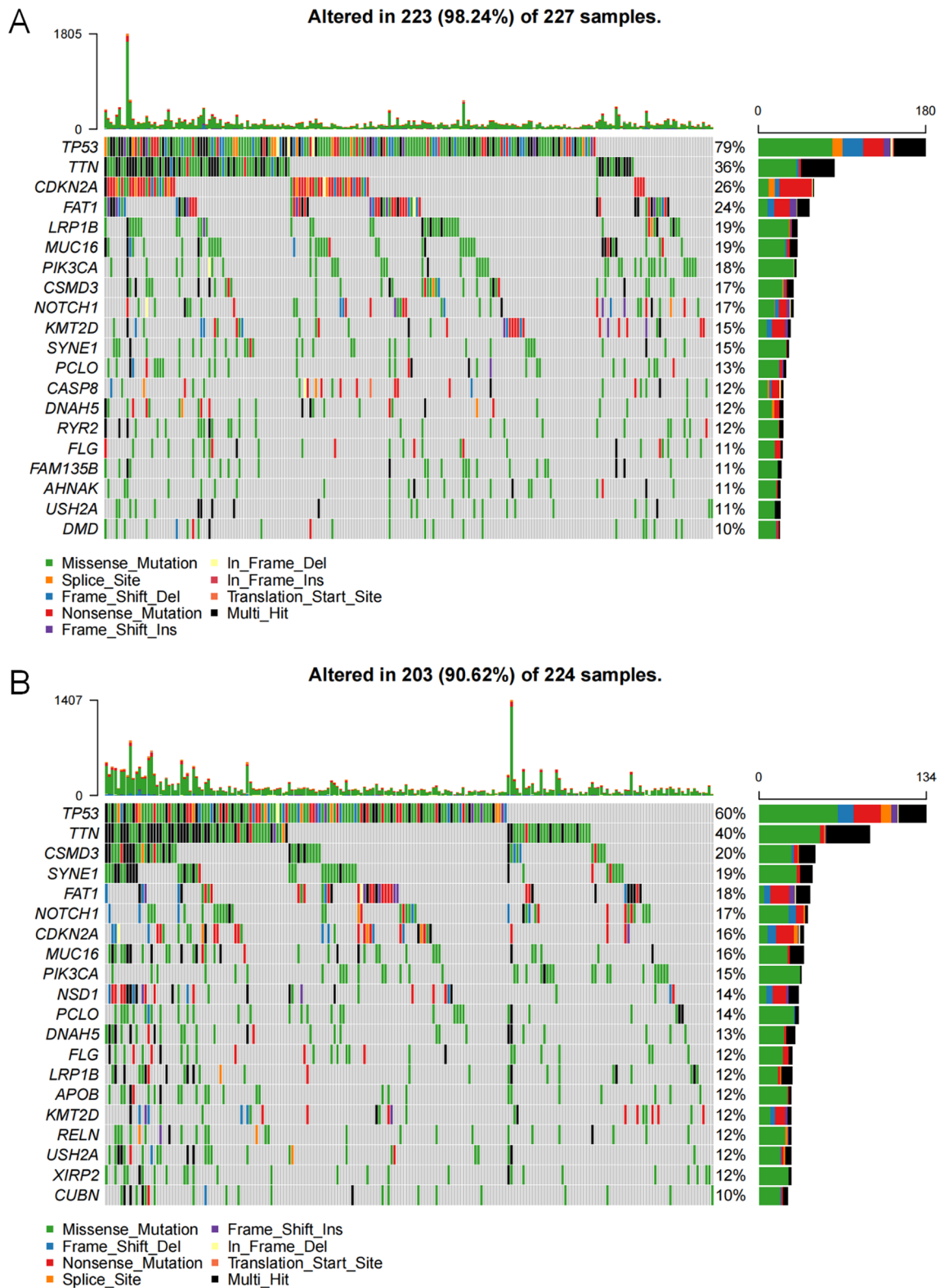


Fig. 9 Waterfall plots of the top 20 somatic mutated genes in high- and low-risk groups. **A** The high-risk and **B** the low-risk group are defined by the median of the epigenetic score of ATHENA

We also acknowledge some limitations. First, the sample sizes of the discovery and validation cohorts are unbalanced, which may affect our results. Though we performed a comprehensive database search, only three available HNSCC DNA methylation datasets with overall survival information were suitable for our study, including TCGA ($n=499$), GSE75537 ($n=53$) and GSE52793 ($n=82$). We performed strict correction of type I error and sensitivity analysis in TCGA cohort, and again validated significant signals in GSE75537 cohort to reduce the false positive probability. Indeed, limited sample size of GSE75537 yields to limited statistical power of confirming the significance of epigenetic predictors. Anyway, we still observed 2 CpG probes with significant main effects and 853 pairs of CpG probes with significant $G \times G$ interactions, which indicating our results are conservative. But, we envision more available database with large sample size and more identified epigenetic predictors in future, which will probably improve the accuracy of ATHENA. Second, sample origins vary across three different cohorts. HNSCC tumor samples in TCGA are composed of 23 types of tissues by occurrence site, including tongue, larynx, overlapped lesion of lip, oral cavity and pharynx, floor of mouth, etc. While, GSE75537 includes merely tongue tumor samples, and GSE52793 is consisted of oral rinse samples. Though sample heterogeneity exists among different types of tissues, under the strict premise of retaining the coefficients of epigenetic biomarkers instead of retraining the model, ATHENA still reflects acceptable prediction accuracy and discrimination ability in all three cohorts, indicating its robustness. Besides, ATHENA still retains statistical significance and prediction accuracy in almost all occurrence site subgroups with sufficient sample size (Additional file 1: Figure S2), which suggests well generalization ability of ATHENA model again. Finally, since the majority of samples in the TCGA cohort are Caucasians (87.8%), generalization of our results to the other ethnicities should be cautioned.

Conclusion

We propose ATHENA, an accurate and independently validated prognostic prediction model of HNSCC incorporating autophagy-related epigenetic biomarkers with either main effects or $G \times G$ interactions. A free and user-friendly online tool is released at <http://bigdata.njmu.edu.cn/ATHENA/>.

Web resources

TCGA: <https://portal.gdc.cancer.gov>
 GEO: <https://www.ncbi.nlm.nih.gov/geo/>
 UCSC Xena browser: <https://xenabrowser.net>

Human Autophagy Database (HADb): <http://www.autophagy.lu/>

The DrugBank database: <https://go.drugbank.com/>

Abbreviations

HNSCC	Head and neck squamous cell carcinoma
ARG	Autophagy-related gene
ATHENA	An independently validated autophagy-related prognostic prediction model of Head and Neck squamous cell carcinoma
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
QC	Quality control
SNP	Single nucleotide polymorphism
BMIQ	Beta-mixture quantile
FPKM	Fragments per kilobase per million mapped reads
TPM	Transcripts per kilobase million
HADb	The Human Autophagy Database
FDR	False discovery rate
SD	Standard deviation
HR	Hazard ratio
CI	Confidence interval
ROC	Receiver operating characteristic
AUC	Area under the receiver operating characteristic curve
C-index	Concordance index
DCA	Decision curve analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
TIME	Tumor immune microenvironment
TIICs	Tumor-infiltrating immune cells

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-023-01501-0>.

Additional file 1. Supplementary File.

Acknowledgements

The authors thank TCGA and GEO for contributing clinical, DNA methylation, and RNA sequencing data, as well as all study subjects who participated in the study cohorts.

Author contributions

ZX, CX, WZ and RZ contributed to the study design. ZX, XC, XS, XK, JC, YS, MX, LQ and MG contributed to data collection, statistical analysis and interpretation. ZX, XK and JC constructed the online tool. ZX, XC, XS, XK, CX, WZ and RZ drafted the manuscript. All authors contributed to critical revision of the manuscript and approved its final version. Financial support and study supervision were provided by CX, WZ and RZ. All authors read and approved the final manuscript.

Funding

This study was funded by the Natural Science Foundation of China (82273737 to R.Z. and 82001099 to C.X.), the Natural Science Foundation of Jiangsu Province (SBK2017043261 to W.Z.) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). R.Z. was partially supported by the Qing Lan Project of the Higher Education Institutions of Jiangsu Province and the Outstanding Young Level Academic Leadership Training Program of Nanjing Medical University.

Availability of data and materials

The datasets analyzed during the current study are available in the TCGA (<https://portal.gdc.cancer.gov>). The independent validation cohorts are available in the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) (GSE75537 and GSE52793).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All participants or their surrogate care providers gave written informed consent. All authors have reviewed the manuscript and consented for publication.

Competing interests

The authors declare that they have no conflict of interest.

Received: 16 November 2022 Accepted: 9 May 2023

Published online: 09 June 2023

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49.
- Zhao X, Chen H, Qiu Y, Cui L. FAM64A promotes HNSCC tumorigenesis by mediating transcriptional autoregulation of FOXM1. *Int J Oral Sci.* 2022;14(1):25.
- Le Naour J, Sztupinszki Z, Carbonnier V, Casiraghi O, Marty V, Galluzzi L, et al. A loss-of-function polymorphism in ATG16L1 compromises therapeutic outcome in head and neck carcinoma patients. *Oncoimmunology.* 2022;11(1):2059878.
- Hsieh YP, Chen KC, Chen MY, Huang LY, Su AY, Chiang WF, et al. Epigenetic deregulation of protein tyrosine kinase 6 promotes carcinogenesis of oral squamous cell carcinoma. *Int J Mol Sci.* 2022;23(9):4495.
- Li J, Yan T, Wu X, Ke X, Li X, Zhu Y, et al. Aberrant overexpression of transcription factor Forkhead box D1 predicts poor prognosis and promotes cancer progression in HNSCC. *BMC Cancer.* 2021;21(1):1205.
- Huang C, Chen L, Savage SR, Eguez RV, Dou Y, Li Y, et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell.* 2021;39(3):361–79 e16.
- van Harten AM, Brakenhoff RH. Targeted treatment of head and neck (pre)cancer: preclinical target identification and development of novel therapeutic applications. *Cancers (Basel).* 2021;13(11):2774.
- Kocak M, Ezazi Erdi S, Jorba G, Maestro I, Farres J, Kirkin V, et al. Targeting autophagy in disease: established and new strategies. *Autophagy.* 2022;18(3):473–95.
- Li X, Dai Z, Wu X, Zhang N, Zhang H, Wang Z, et al. The comprehensive analysis identified an autophagy signature for the prognosis and the immunotherapy efficiency prediction in lung adenocarcinoma. *Front Immunol.* 2022;13:749241.
- Chen Y, Zhao H, Liang W, Jiang E, Zhou X, Shao Z, et al. Autophagy regulates the cancer stem cell phenotype of head and neck squamous cell carcinoma through the noncanonical FOXO3/SOX2 axis. *Oncogene.* 2022;41(5):634–46.
- Duan Y, Tian X, Liu Q, Jin J, Shi J, Hou Y. Role of autophagy on cancer immune escape. *Cell Commun Signal.* 2021;19(1):91.
- Fan T, Wang X, Zhang S, et al. NUPR1 promotes the proliferation and metastasis of oral squamous cell carcinoma cells by activating TFE3-dependent autophagy. *Signal Transduct Target Ther.* 2022;7(1):130.
- Liu PF, Chen CF, Ger LP, et al. MAP3K11 facilitates autophagy activity and is correlated with malignancy of oral squamous cell carcinoma. *J Cell Physiol.* 2022;237(11):4275–91.
- Lu T, Li Y, Pan M, et al. TBC1D14 inhibits autophagy to suppress lymph node metastasis in head and neck squamous cell carcinoma by downregulating macrophage erythroblast attachment. *Int J Biol Sci.* 2022;18(5):1795–812.
- Gao L, Zhang Q, Li S, Zheng J, Ren W, Zhi K. Circ-PKD2 promotes Atg13-mediated autophagy by inhibiting miR-646 to increase the sensitivity of cisplatin in oral squamous cell carcinomas. *Cell Death Dis.* 2022;13(2):192.
- Antonioni M, Pagni B, Vescovo T, et al. HPV sensitizes OPSCC cells to cisplatin-induced apoptosis by inhibiting autophagy through E7-mediated degradation of AMBRA1. *Autophagy.* 2021;17(10):2842–55.
- Lee M, Nam HY, Kang HB, et al. Epigenetic regulation of p62/SQSTM1 overcomes the radioresistance of head and neck cancer cells via autophagy-dependent senescence induction. *Cell Death Dis.* 2021;12(3):250.
- Zhang Q, Wu Y, Xu Q, Ma F, Zhang CY. Recent advances in biosensors for in vitro detection and in vivo imaging of DNA methylation. *Biosens Bioelectron.* 2021;171:112712.
- Zhang R, Lai L, He J, Chen C, You D, Duan W, et al. EGLN2 DNA methylation and expression interact with HIF1A to affect survival of early-stage NSCLC. *Epigenetics.* 2019;14(2):118–29.
- Xu Y, Hong M, Kong D, Deng J, Zhong Z, Liang J. Ferroptosis-associated DNA methylation signature predicts overall survival in patients with head and neck squamous cell carcinoma. *BMC Genom.* 2022;23(1):63.
- Chen J, Shen S, Li Y, Fan J, Xiong S, Xu J, et al. APOLLO: an accurate and independently validated prediction model of lower-grade gliomas overall survival and a comparative study of model performance. *EBioMedicine.* 2022;79:104007.
- Zhang R, Chen C, Dong X, Shen S, Lai L, He J, et al. Independent validation of early-stage non-small cell lung cancer prognostic scores incorporating epigenetic and transcriptional biomarkers with gene–gene interactions and main effects. *Chest.* 2020;158(2):808–19.
- Krishnan NM, Dhas K, Nair J, Palve V, Bagwan J, Siddappa G, et al. A minimal DNA methylation signature in oral tongue squamous cell carcinoma links altered methylation with tumor attributes. *Mol Cancer Res.* 2016;14(9):805–19.
- Langevin SM, Butler RA, Eliot M, Pawlita M, Maccani JZ, McClean MD, et al. Novel DNA methylation targets in oral rinse samples predict survival of patients with oral squamous cell carcinoma. *Oral Oncol.* 2014;50(11):1072–80.
- Yoshihara K, Shahmoradgolgi M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–7.
- Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46(D1):D1074–82.
- Harrington DP. A class of rank test procedures for censored survival data. *Biomedica.* 1982;62:553–66.
- Jiang J, Niu L, Zhang MX, Wang H, Xie JQ, Sun GP. A novel nomogram combining alternative splicing events and clinical factors for prognosis prediction in head and neck squamous cell carcinoma. *J Oncol.* 2022;2022:4552445.
- Chen J, Lu T, Zhong F, Lv Q, Fang M, Tu Z, et al. A signature of N(6)-methyladenosine regulator-related genes predicts prognoses and immune responses for head and neck squamous cell carcinoma. *Front Immunol.* 2022;13:809872.
- Ji X, Lin L, Fan J, Li Y, Wei Y, Shen S, et al. Epigenome-wide three-way interaction study identifies a complex pattern between TRIM27, KIAA0226, and smoking associated with overall survival of early-stage NSCLC. *Mol Oncol.* 2022;16(3):717–31.
- Zhang R, Shen S, Wei Y, Zhu Y, Li Y, Chen J, et al. A large-scale genome-wide gene–gene interaction study of lung cancer susceptibility in Europeans with a trans-ethnic validation in Asians. *J Thorac Oncol.* 2022;17(8):974–90.
- Shen S, Wei Y, Li Y, Duan W, Dong X, Lin L, et al. A multi-omics study links TNS3 and SEPT7 to long-term former smoking NSCLC survival. *NPJ Precise Oncol.* 2021;5(1):39.
- Xu Z, Gu Y, Chen J, Chen X, Song Y, Fan J, et al. Epigenome-wide gene–age interaction study reveals reversed effects of MORN1 DNA methylation on survival between young and elderly oral squamous cell carcinoma patients. *Front Oncol.* 2022;12:941731.
- Gu Y, Li P, Peng F, Zhang M, Zhang Y, Liang H, et al. Autophagy-related prognostic signature for breast cancer. *Mol Carcinog.* 2016;55(3):292–9.
- Messai Y, Noman MZ, Hasmim M, Janji B, Tittarelli A, Boutet M, et al. ITPR1 protects renal cancer cells against natural killer cells by inducing autophagy. *Cancer Res.* 2014;74(23):6820–32.

37. Xu S, Wang P, Zhang J, Wu H, Sui S, Zhang J, et al. Ai-lncRNA EGOT enhancing autophagy sensitizes paclitaxel cytotoxicity via upregulation of ITPR1 expression by RNA–RNA and RNA–protein interactions in human cancer. *Mol Cancer*. 2019;18(1):89.
38. Franco-Salla GB, Prates J, Cardin LT, Dos Santos AR, Silva WA Jr, da Cunha BR, et al. *Euphorbia tirucalli* modulates gene expression in larynx squamous cell carcinoma. *BMC Complement Altern Med*. 2016;16:136.
39. Ren Z, Zhang L, Ding W, Luo Y, Shi Z, Shrestha B, et al. Development and validation of a novel survival model for head and neck squamous cell carcinoma based on autophagy-related genes. *Genomics*. 2021;113(1 Pt 2):1166–75.
40. Chen CG, Iozzo RV. Extracellular matrix guidance of autophagy: a mechanism regulating cancer growth. *Open Biol*. 2022;12(1):210304.
41. Mao X, Zhang X, Zheng X, Chen Y, Xuan Z, Huang P. Curcumin suppresses LGR5(+) colorectal cancer stem cells by inducing autophagy and via repressing TFAP2A-mediated ECM pathway. *J Nat Med*. 2021;75(3):590–601.
42. Hupfer A, Brichkina A, Koeniger A, Keber C, Denkert C, Pfefferle P, et al. Matrix stiffness drives stromal autophagy and promotes formation of a protumorigenic niche. *Proc Natl Acad Sci U S A*. 2021;118(40):e2105367118.
43. Deretic V, Levine B. Autophagy balances inflammation in innate immunity. *Autophagy*. 2018;14(2):243–51.
44. Lawson KA, Sousa CM, Zhang X, Kim E, Akthar R, Caumanns JJ, et al. Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature*. 2020;586(7827):120–6.
45. White E, Lattime EC, Guo JY. Autophagy regulates stress responses, metabolism, and anticancer immunity. *Trends Cancer*. 2021;7(8):778–89.
46. Bryant KL, Stalnecker CA, Zeitouni D, Klomp JE, Peng S, Tikunov AP, et al. Combination of ERK and autophagy inhibition as a treatment approach for pancreatic cancer. *Nat Med*. 2019;25(4):628–40.
47. Kinsey CG, Camolotto SA, Boespflug AM, Guillen KP, Foth M, Truong A, et al. Protective autophagy elicited by RAF→MEK→ERK inhibition suggests a treatment strategy for RAS-driven cancers. *Nat Med*. 2019;25(4):620–7.
48. Liu G, Yuan C, Ma J, Pan Y, Xu H. Influence of immune microenvironment on diagnosis and prognosis of head and neck squamous cell carcinoma. *Front Oncol*. 2021;11: 604784.
49. Tang Y, Li C, Zhang YJ, Wu ZH. Ferroptosis-related long non-coding RNA signature predicts the prognosis of head and neck squamous cell carcinoma. *Int J Biol Sci*. 2021;17(3):702–11.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

