npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-025-01068-4

Impact of HMGA1 on tumorigenesis, prognosis and immune microenvironment in HNSCC: a multi-omics study

Ziang Xu[1,2,9], Moxu Wang[1,3,9], Jilin Cai[1,3,9], Tianxiao Wang[1,3], Nan Ni[1,3], Tao Liu[1,3], Maojie Xue[4], Xiang Wang[4], Zhixuan Liu[1,3], Hua Yuan[1,3] ✉, Wei Zhang[1,5] ✉ & Ruyang Zhang[4,6,7,8] ✉

HMGA1 plays an important role in a variety of biological processes. However, it is still unclear what role HMGA1 plays in HNSCC. By integrating multi-omics data from public and private cohorts, we conducted a comprehensive analysis. The results showed that *HMGA1* expression was significantly higher in HNSCC tumor tissue, correlated with poor prognosis. Increased *HMGA1* expression indicated distinct somatic mutations and heavier tumor mutation burden, meanwhile, had significant interaction with several immune checkpoints. Single-cell analysis suggested that *HMGA1* was highly enriched in malignant cells. In-vitro and in-vivo experiments also suggested that HMGA1 promoted the proliferation, migration and activation of HNSCC cells. Upstream analysis showed that $cg25207224_{HMGA1}$ tested by oral rinse specimen maybe a non-invasive in-vitro predictive marker for prognostic prediction of HNSCC. Our study revealed the impact of HMGA1 on tumorigenesis, prognosis and immune microenvironment in HNSCC from a multi-omics perspective and provided a therapeutic target for HNSCC patients.

Head and neck squamous cell carcinoma (HNSCC), developing from the mucosal epithelium in the oral cavity, pharynx, and larynx, is one of the top ten most common cancer types in the world[1]. In 2022, there were 389,485 cases of cancer in the lips and oral cavity alone, of which 188,230 died, accounting for 1.9% of all cancer deaths[2]. Despite the use of multiple treatment modalities, the 5-year survival rate of HNSCC remains stagnant at around 50%[3]. There is increasing evidence that biomarkers at the levels of genomics[4], epigenomics[5], transcriptomics[6] and proteomics[7] have the ability to predict disease occurrence and prognosis, and help with the diagnosis of disease at an early stage, which may affect the survival rate of HNSCC.

The high Mobility Group A (HMGA) family, encoded by two related members (HMGA1 and HMGA2), is one of the most abundant and ubiquitous nonhistone chromosomal protein family[8]. HMGA1 play important roles in several biological processes, including autophagy, cellular proliferation, cell apoptosis, inflammation and pluripotency of embryonic stem cells[9]. In addition, by changing the conformational state and accessibility of chromatin through a variety of regulatory factors, HMGA1 participate in the modulation of numerous gene expression[10]. Gene expression of *HMGA1* is high in almost all malignant cancers[11], and is related to the prognosis of multiple cancers, including lung cancer, breast cancer, colorectal cancer, gastric cancer, prostate cancer and liver cancer[12–17]. However, the role of HMGA1 in HNSCC still remains largely unclear.

In addition, majority of these HNSCC studies are only transcriptomic studies, focusing on single type of molecular biomarkers, which may ignore the ability of other types of biomarkers in revealing the biological mechanisms of complex diseases. Furthermore, these existing studies merely focus on the main effects of genes, while neglecting to explore potential gene-gene interactions[18].

Hence, leveraging multi-omics (genomics, epigenetics, transcriptomics, proteomics, etc.) public database from the Cancer Genome

[1]State Key Laboratory Cultivation Base of Research, Prevention and Treatment for Oral Diseases, Nanjing Medical University, Nanjing, Jiangsu, China. [2]Department of the Ninth Outpatient, 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, Nanjing, Jiangsu, China. [5]Department of Oral Special Consultation, Affiliated Stomatological Hospital of Nanjing Medical University, Nanjing, Jiangsu, China. [6]China International Cooperation Center for Environment and Human Health, Nanjing Medical University, Nanjing, Jiangsu, China. [7]Changzhou Medical Center, Nanjing Medical University, Changzhou, Jiangsu, China. [8]Information Center, The Second People's Hospital of Changzhou, the Third Affiliated Hospital of Nanjing Medical University, Changzhou, Jiangsu, China. [9]These authors contributed equally: Ziang Xu, Moxu Wang, Jilin Cai. ✉e-mail: yuanhua@njmu.edu.cn; zhangwei0508@njmu.edu.cn; zhangruyang@njmu.edu.cn

THE HORMEL INSTITUTE
UNIVERSITY OF MINNESOTA

Atlas (TCGA), the Gene Expression Omnibus (GEO) the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and private database from Nanjing Head and Neck Cancer Cohort (NJHNCC), etc., we performed a comprehensive study of HMGA1 on the occurrence, development and immunity of HNSCC, and conducted an experimental verification.

Results

HMGA1 is associated with occurrence and prognosis of multiple cancers, especially HNSCC

The analysis flow was depicted in Fig. 1, showing the multi-omics analysis for the effect of HMGA1 in HNSCC. Basic concepts of the multi-omics data was presented in Table S1. We performed a comprehensive analysis of the specific expression patterns of *HMGA1* in 33 different cancer tissues, and observed that *HMGA1* was differentially expressed between tumor tissues and matched normal tissues among 18 types of tumors, including HNSCC. It is noteworthy that no significant difference of *HMGA1* expression was observed between HPV^- and HPV^+ HNSCC individuals (Fig. 2A). In the TCGA-HNSCC cohort, *HMGA1* showed significantly high expression in tumor samples, both in paired (Fig. 2B) and unpaired samples (Fig. 2C). Survival analysis suggested that *HMGA1* expression was significantly correlated with the prognosis of various tumors, high *HMGA1* expression mostly indicated poor prognosis, except in uveal melanoma (UVM) (Fig. 2D). In the TCGA-HNSCC cohort, high *HMGA1* expression was significantly associated with poor prognosis of patients, both in HNSCC ($HR = 1.51$, 95% CI: 1.15–1.98, $P = 3.0 \times 10^{-03}$) (Fig. 2E) and its subgroup OSCC ($HR = 1.63$, 95% CI: 1.17–2.27, $P = 4.0 \times 10^{-03}$) (Fig. 2F). As validation, we confirmed the high expression of *HMGA1* in HNSCC tissues in our own samples from NJHNCC cohort (Fig. 2G). Also, HMGA1 showed significant high protein abundance in the tumor tissues in CPTAC cohort (Fig. 2H). For intuitive display, we downloaded the immunohistochemistry of HMGA1 in HNSCC patients and normal tissues from the HPA database. The

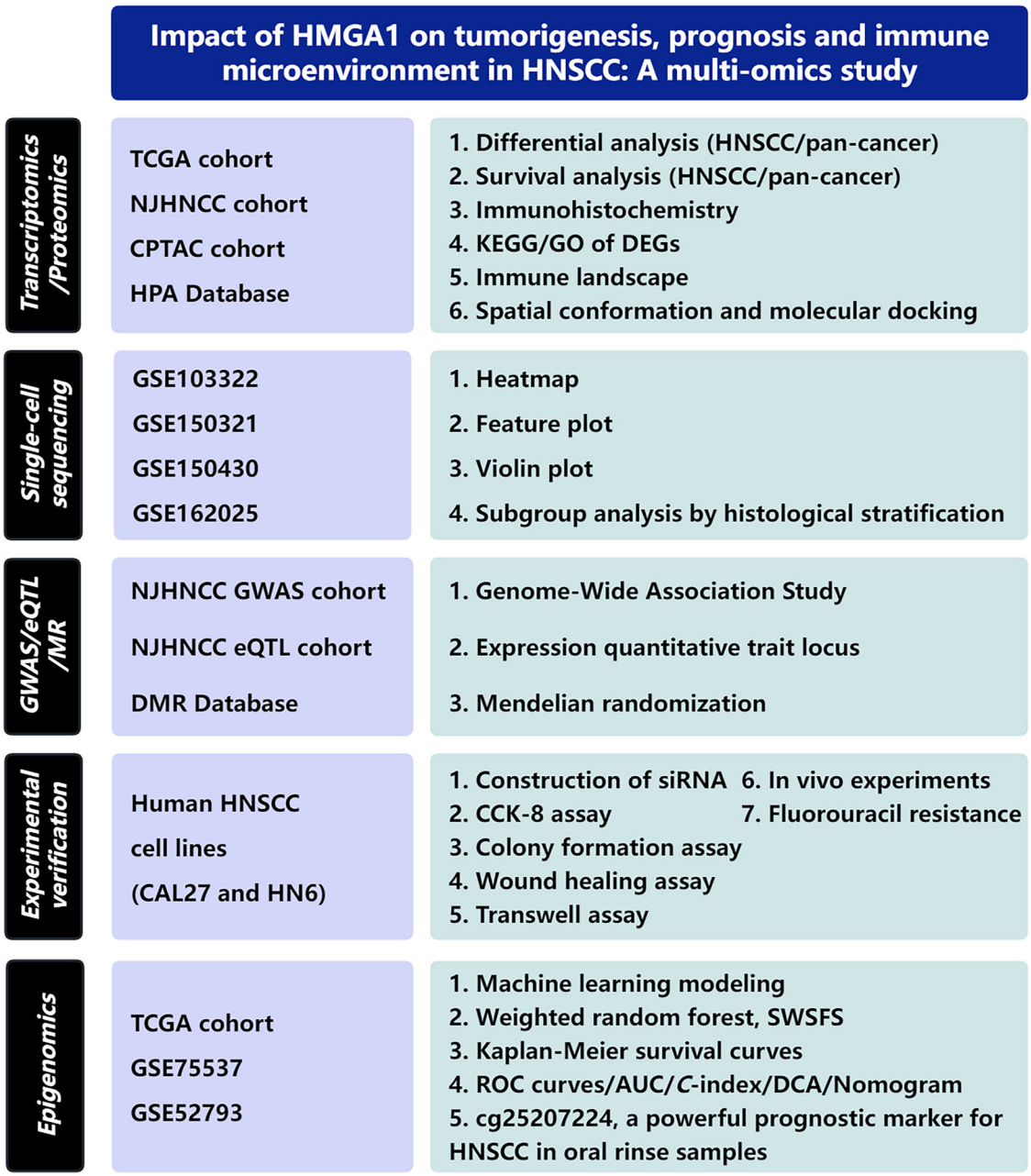

Fig. 1 | Flow chart of study design and statistical analysis.

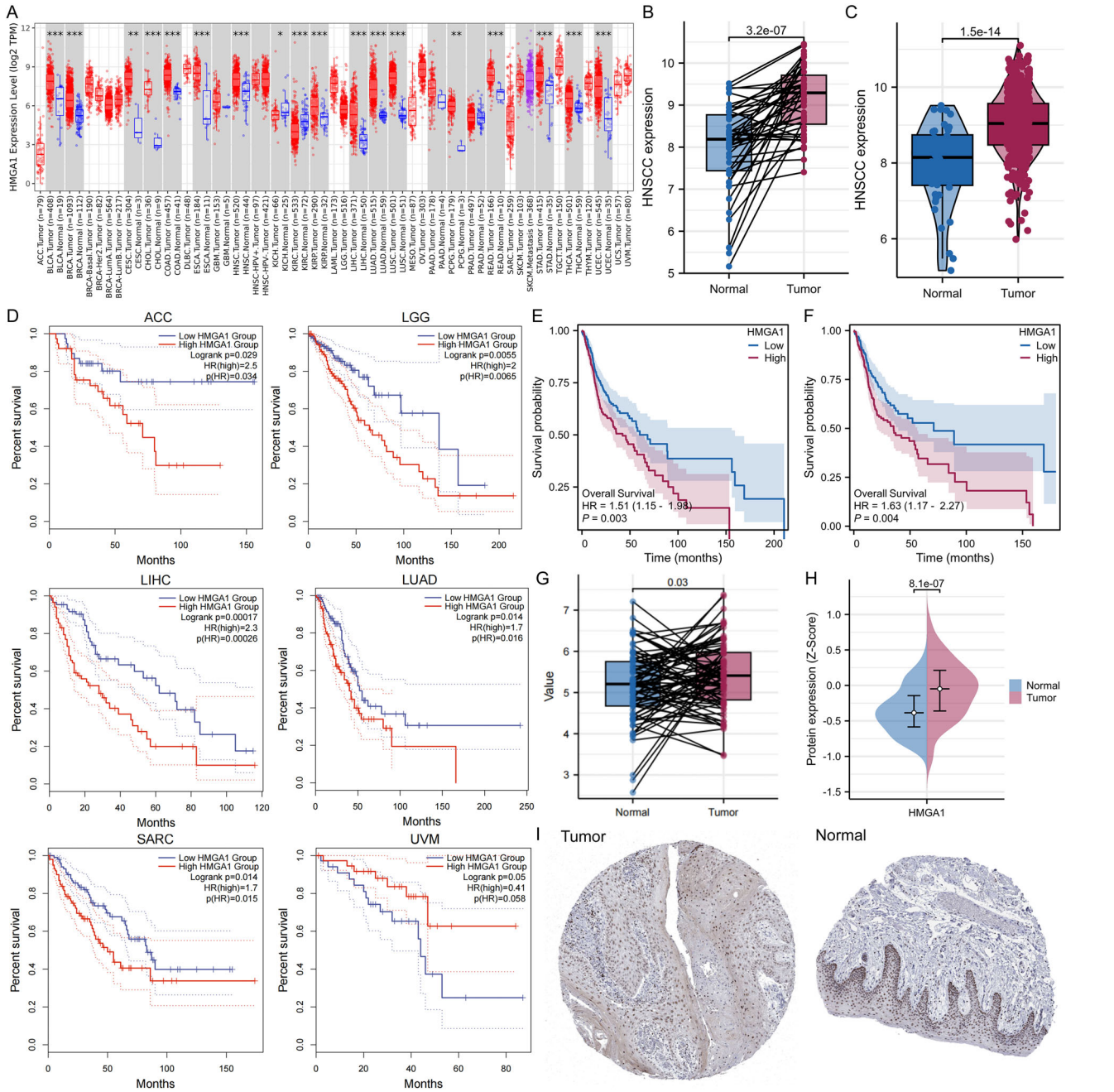

Fig. 2 | Differential analysis of gene and protein expression, and prognostic analysis of *HMGA1*. Differential analysis of *HMGA1* in **A** Pan-cancer, **B** paired HNSCC and **C** unpaired HNSCC patients. Prognostic survival analysis of *HMGA1* in **D** Pan-cancer, **E** HNSCC and **F** OSCC patients. **G** Differential analysis of *HMGA1* of 81 paired HNSCC patients in NJHNCC. **H** Differential analysis of HMGA1 protein expression. **I** HNSCC immunohistochemical image with corresponding normal tissue from HPA.

results showed that HMGA1 showed moderate to strong nuclear positivity in almost all cancer tissues (Fig. S1), and was highly expressed in HNSCC (Fig. 2I).

Immune landscape analysis of *HMGA1* in HNSCC

There were 68 significantly differentially expressed genes in the four subgroups divided according to *HMGA1* expression, which were significantly enriched in 92 biological process pathways, 44 cellular component pathways, 8 molecular function pathways using GO (Gene Ontology) and 2 pathways using KEGG (Kyoto Encyclopedia of Genes and Genomes) (Fig. 3A), which focused primarily on the biogenesis and structure of ribonucleoproteins and ribosomes. Among the differentially expressed genes, 5 genes (*ATIC*, *SNRPE*, *PHB*, *CCT7* and *CCT2*) were identified as the most highly connected hub genes (Fig. 3B). We then plotted waterfall plots to investigate the differences in gene mutations among *HMGA1* expression subgroups. Except for *TP53* and *TTN*, other gene mutations were different among subgroups (Fig. 3C). The per capita tumor mutation burden (TMB) varied statistically significantly among *HMGA1* subgroups, and the magnitude of change was greater in the comparison between subgroup 1 (0–25%) and subgroup 4 (75–100%) (Fig. 3D). Obviously, there was a positive correlation between *HMGA1* expression and per capita TMB ($r = 0.21$, $P = 6.4 \times 10^{-6}$).

HMGA1 expression showed statistically significant strong correlations with various tumor indicators, such as stromal score ($r = -0.30$,

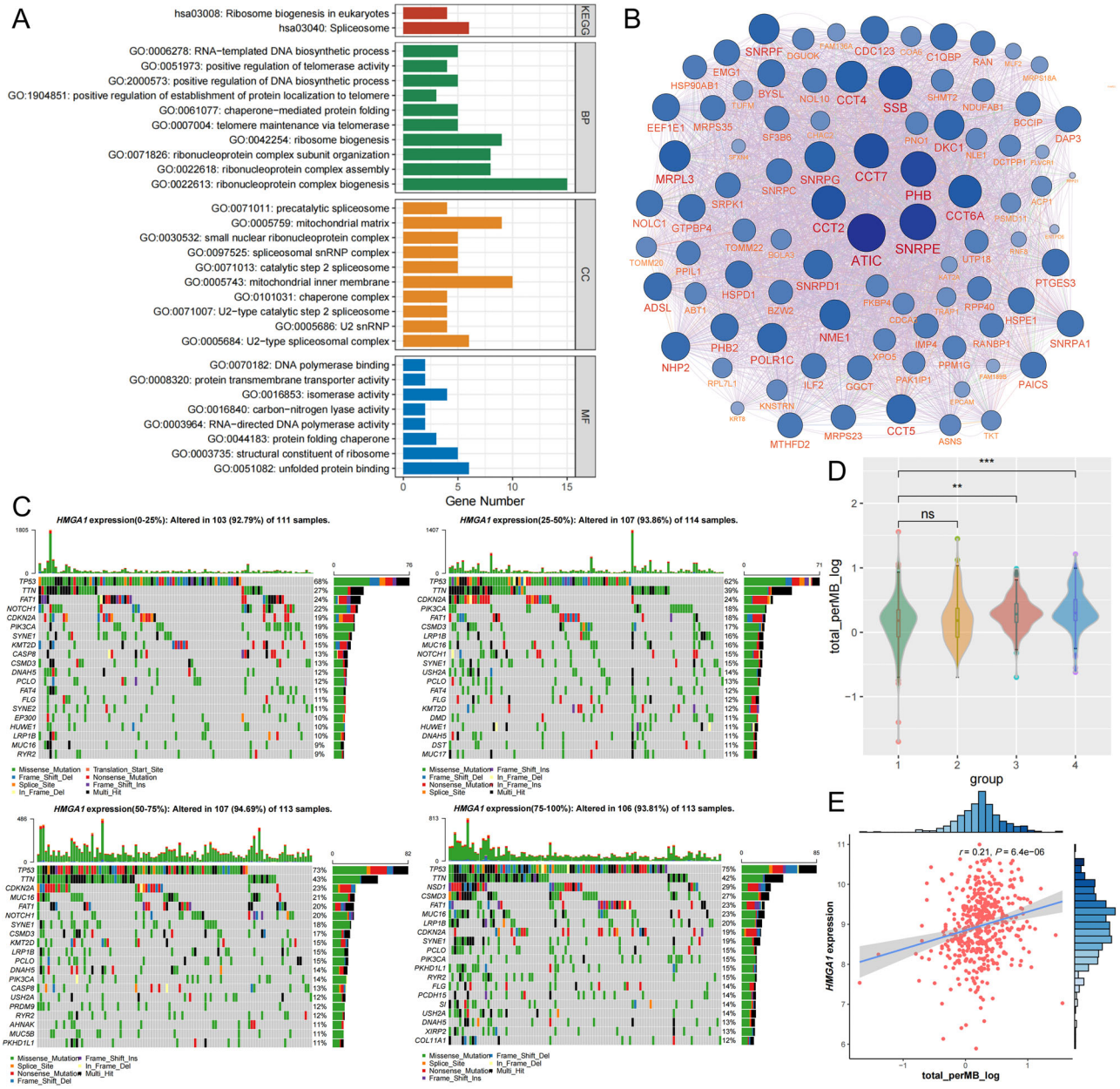

Fig. 3 | Enrichment analysis of differentially expressed genes in *HMGA1*-expressing subgroups, as well as differences in somatic mutations and tumor mutation burden in *HMGA1*-expressing subgroups. **A** KEGG and GO enrichment analysis of differentially expressed genes in *HMGA1*-expressing subgroups. **B** Gene Network of differentially expressed genes in *HMGA1*-expressing subgroups. **C** Somatic mutations among *HMGA1*-expressing subgroups. **D** Differential analysis in per capita tumor mutational burden (TMB) among *HMGA1*-expressing subgroups, * means $P < 0.05$, ** means $P < 0.01$, *** means $P < 0.001$ and **** means $P < 0.0001$. **E** Correlation analysis of per capita TMB and *HMGA1* expression.

$P = 5.1 \times 10^{-11}$), immune score ($r = -0.26$, $P = 1.1 \times 10^{-8}$), estimated score ($r = -0.32$, $P = 5.5 \times 10^{-12}$), and tumor purity ($r = 0.31$, $P = 8.8 \times 10^{-12}$) (Fig. 4A). The composition of 10 immune cells was distributed differently in the *HMGA1* subgroups (Fig. 4B), and the composition of 9 immune cells was statistically significantly correlated with *HMGA1* (Fig. 4C). NK cell activated and mast cells resting were significantly negatively correlated with *HMGA1*, on the contrary, NK cell resting and mast cells activated were significantly positively correlated with *HMGA1*. In addition, we explored the relationship between immune checkpoint expression and *HMGA1* expression. Almost all types of immune checkpoint expression were distributed differently in *HMGA1* subgroups (Fig. 4D). Totally, 13 immune checkpoint genes showed statistically significant, but slight correlation with *HMGA1* (Fig. 4E). Therefore, we tried to find potential interactions between *HMGA1* and immune checkpoints. Gene-gene interaction analysis showed that *HMGA1* had significant interactions with *SIGLEC15* ($HR_{interaction} = 1.490$, $P = 1.29 \times 10^{-2}$) (Fig. 4F) and *TNFRSF18* ($HR_{interaction} = 0.762$, $P = 3.91 \times 10^{-2}$) (Fig. 4G) for the prognosis of HNSCC. Using *AlphaFold 3*, we generated the spatial conformation of the protein binding of HMGA1, SIGLEC15 and TNFRSF18, respectively (Fig. 4H, I). Finally, we molecularly docked *HMGA1* with some common chemotherapeutic agents (e.g., docetaxel and fluorouracil). Figure 4J, K demonstrated the docking conformation and interaction force analysis between *HMGA1* and two common chemotherapeutic agents. The results showed that the binding energies formed by docetaxel, fluorouracil with *HMGA1* amino acid residues were −2.900 kcal/mol and −4.733 kcal/mol, respectively (Table S2).

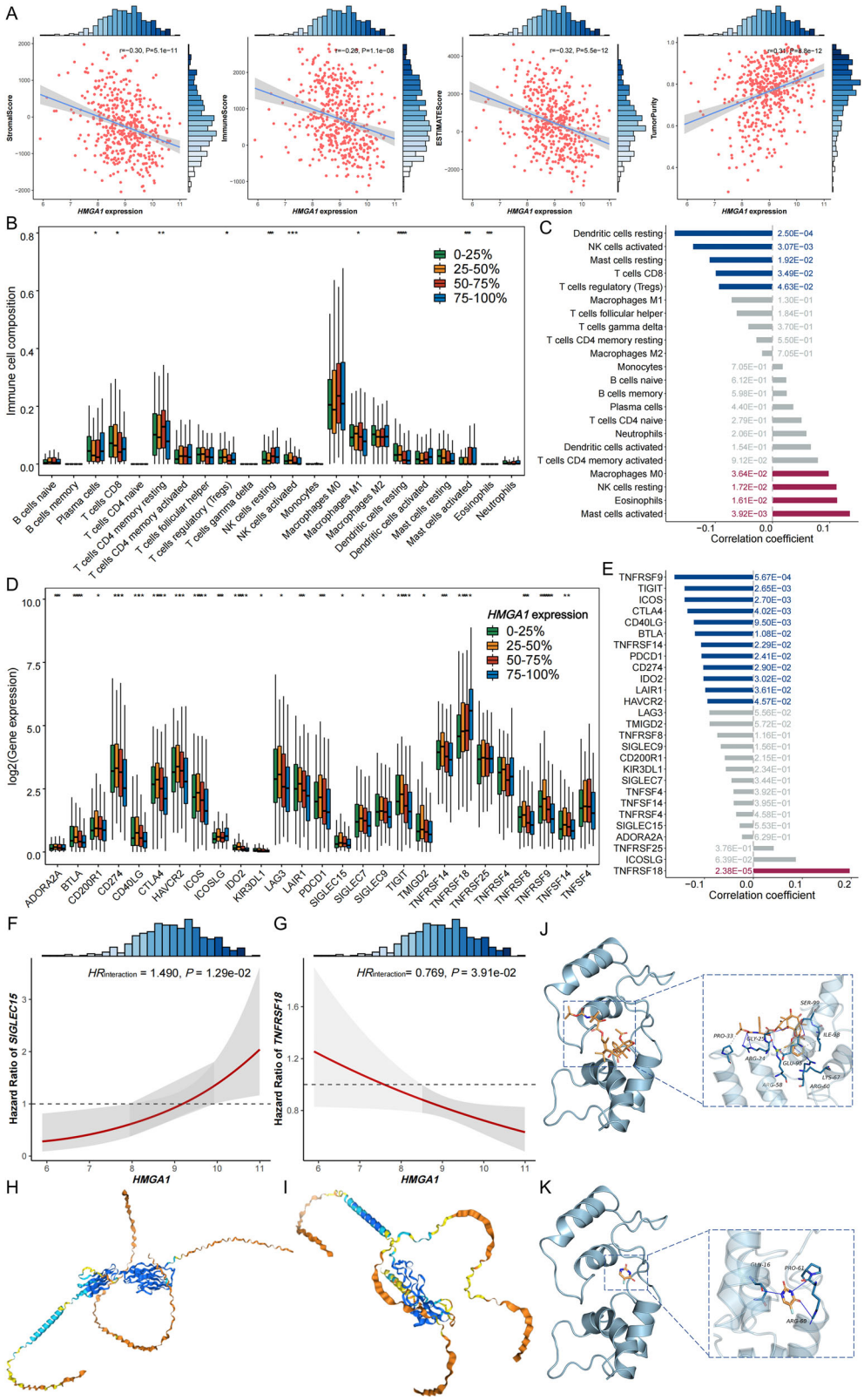

Fig. 4 | Correlation analysis of immune cell composition, immune checkpoint expression, common chemotherapy drugs and *HMGA1*. **A** Correlation analysis of immune microenvironment and *HMGA1* expression. **B** Differential analysis in abundances of 22 immune cells among *HMGA1*-expressing subgroups. **C** The lollipop chart of correlation coefficients between immune cells and *HMGA1* expression. **D** Differential analysis in immune checkpoint expression among *HMGA1*-expressing subgroups. **E** The lollipop chart of correlation coefficients between immune checkpoint expression and *HMGA1* expression. The interaction analysis between *HMGA1*, **F** *SIGLEC15* and **G** *TNFRSF18* in the prognosis of HNSCC, respectively. The spatial conception of the combination of between *HMGA1*, **H** *SIGLEC15* and **I** *TNFRSF18*, respectively. The docking conformation analysis between *HMGA1* and two common chemotherapeutic agents, **J** docetaxel and **k** fluorouracil.

HMGA1 is associated with malignancy, cell proliferation, migration, invasion and drug resistance of HNSCC

We obtained four HNSCC single-cell datasets using TISCH 2.0. Our goal was to examine the cellular distribution of *HMGA1* expression in HNSCC tissues in the single-cell level. These datasets showed that *HMGA1* was highly expressed in malignant cells (Fig. 5A). Feature plot and violin plot showed *HMGA1* was significantly overexpressed in malignant cells compared to other cell types in GSE103322 (Fig. 5B, C). Subgroup analysis of histological stratification showed that the distribution of *HMGA1* in malignant cells was significantly increased in male individuals compared with female individuals, and raised with the increase of clinical stage (Fig. 5D, E). Similar trends were observed in various HNSCC subgroups, including NPC (GSE162025 (Fig. 5F, G) and GSE150430 (Fig. 5H, Fig. S2A)) and LSCC (GSE150321) (Fig. 5I, Fig. S2B).

We carried out with in-vitro cytological experiments and in-vivo experiments in order to explore how HMGA1 promoted the occurrence and development of HNSCC. First, RT-qPCR and Western blot assay proved that *HMGA1* showed significant high expression trend in all four HNSCC cell lines (HSC3, HN4, HN6 and CAL27), especially in HN6 and CAL27 (Fig. 6A). We then established stable HMGA1 knockdown cell lines by siRNAs of HN6 and CAL27 cells, and verified the efficiency of HMGA1 knockdown by si-RNA in these two cell lines (Fig. 6B). The following CCK-8 experiment demonstrated that HMGA1 knockdown could significantly inhibit the proliferation of HNSCC (Fig. 6C). Wound healing and Transwell experiments suggested that the abilities of HNSCC cells to migrate and invade was inhibited when HMGA1 was knocked down in HN6 and CAL27 cells (Fig. 6D, E). The colony formation experiment confirmed that the significant effect of HMGA1 knockdown on inhibiting cell invasion and colony formation ability of HNSCC (Fig. 6F). Molecular docking suggested that HMGA1 binds to fluorouracil. Following *HMGA1* knockdown, HNSCC cells were treated with increasing concentrations of fluorouracil to assess the role of HMGA1 in chemoresistance. The results demonstrated that HMGA1 promotes fluorouracil resistance by enhancing cellular survival capabilities (Fig. S3). Finally, in-vivo xenograft experiments using nude mice demonstrated that HMGA1 knockdown significantly suppressed tumor growth (Fig. 6G), as evidenced by marked reductions in both tumor volume and weight compared to control groups (Fig. 6H, I). Uncropped and unprocessed scans of Western blot were in Fig. S4.

Exploration of SNP sites potentially regulate *HMGA1* expression

Univariate association analysis suggested that among the 53 SNPs in *HMGA1* and its upstream and downstream 10 kb range, rs1759628 may potentially regulate the occurrence of HNSCC ($OR = 1.16$, $P = 4.62 \times 10^{-2}$) (Table S3). Unfortunately, due to the limitation of sample size and cross-ethnic validation (all samples of NJHNCC were from Chinese Han population), we did not find any SNPs significantly associated with *HMGA1* expression at the genome-wide level (Table S4). With the help of DMRdb, no significant correlation between *HMGA1* and HNSCC was found in existing MR studies.

Prognostic model based on *HMGA1* methylation suggested $cg25207224_{HMGA1}$ maybe reliable site for non-invasive prediction of HNSCC prognosis

Finally, 19 CpG probes were remained in *HMGA1* after QC (Table S5). The C-index values of the prognostic models established by 101 machine learning strategies using these 19 CpG probes were not ideal (Fig. S5), The best model among which was built using the ENET (alpha = 0.7) strategy ($\text{C-index}_{TCGA} = 0.591$, $\text{C-index}_{GSE75537} = 0.605$), so we turned to the weighted random forest strategy to find the CpG probes on *HMGA1* which really played an important role. In the biomarker screening stage, the SWSFS algorithm identified 8 and 1 top CpG probes in the discovery and the validation phase, respectively (Fig. 7A). Only one probe, cg25207224 in *HMGA1*, was simultaneously identified in both discovery and validation phases (Fig. 7B). In the biomarker testing stage, $cg25207224_{HMGA1}$, was significantly associated with the HNSCC survival across two phases through Cox regression model adjusted for covariates ($HR = 0.97$, 95% CI: 0.96–0.99, $P = 5.12 \times 10^{-5}$, FDR-$q = 3.58 \times 10^{-4}$ in the discovery phase; $HR = 0.88$, 95% CI: 0.81–0.97, $P = 7.88 \times 10^{-3}$ in the validation phase). In order to prevent false positives caused by overfitting of the model, we randomly divided the TCGA dataset into a discovery set and a validation set at a ratio of 5:5 and 7:3, and repeated the above strategy. The results showed that the importance of $cg25207224_{HMGA1}$ was still significant (Fig. S6). Further, we found there was a significant negative association between $cg25207224_{HMGA1}$ methylation and *HMGA1* expression ($r = -0.11$, $P = 1.2 \times 10^{-2}$, Fig. 7C), and dual luciferase reporter gene assay revealed significant differences in relative luciferase activity between the WT group and the MUT group in the chromosomal region where $cg25207224_{HMGA1}$ located (Fig. S7), indicating *HMGA1* expression was cis-regulated by $cg25207224_{HMGA1}$. To better illustrate the effect of $cg25207224_{HMGA1}$ on HNSCC overall survival, patients were categorized into two groups (low and high) based on the median value of $cg25207224_{HMGA1}$. Kaplan–Meier curves showed significant survival differences between two groups in discovery phase (TCGA, Fig. 7D), validation phase (GSE75537, Fig. 7E) and another oral rinse sample dataset (GSE52793, Fig. 7F).

We then evaluated the predictive performance of $cg25207224_{HMGA1}$. In the discovery phase, the model using only clinical covariates had limited predictive ability ($\text{AUC}_{36\text{-month}} = 0.607$, $\text{AUC}_{60\text{-month}} = 0.531$, C-index = 0.598), and the addition of $cg25207224_{HMGA1}$ increased the AUC for 3-year and 5-year survival prediction by 10.21% and 13.18%, respectively ($\text{AUC}_{36\text{-month}} = 0.669$, $\text{AUC}_{60\text{-month}} = 0.601$, C-index = 0.643) (Fig. 8A, B). DCA analysis of discovery set was in Fig. S8. In the validation phase, the model adding $cg25207224_{HMGA1}$ still showed great prognostic accuracy ($\text{AUC}_{36\text{-month}} = 0.783$, $\text{AUC}_{60\text{-month}} = 0.886$) (Fig. 8C) even if the coefficients were fixed to be consistent with the discovery phase. To facilitate its application in clinical practice, we developed a nomogram to estimate the 36-month or 60-month survival rate of HNSCC patients (Fig. 8D).

Surprisingly, although GSE52793 was short of covariates, and the dataset was all oral rinse samples, which may just the surface shedding of cancer tissue to some extent, the prognostic accuracy of $cg25207224_{HMGA1}$ was still verified ($\text{AUC}_{36\text{-month}} = 0.702$, $\text{AUC}_{60\text{-month}} = 0.808$) (Fig. 8E), proved that $cg25207224_{HMGA1}$ had the potential to predict the prognosis of HNSCC non-invasively and with high accuracy. DCA showed that as a biomarker for non-invasive in-vitro testing (only use oral rinses of HNSCC patients), $cg25207224_{HMGA1}$ showed more clinical net benefits (Fig. 8F, G). Specifically, at a reasonable threshold probability (e.g., $P_t = 0.4$), $cg25207224_{HMGA1}$ showed a higher net benefit (NB). In other words, with the 60-month survival rate as the endpoint, the marker identified 28.0 true positive patients out of every 100 patients who should receive intervention. On the other hand, $cg25207224_{HMGA1}$ can reduce the number of unnecessary clinical interventions by 65.83% and will not miss any intervention measures for patients with a truly high risk of death.

Discussion

In this study, we conducted a comprehensive multi-omics analysis of the role of *HMGA1* in the occurrence, development and immunity of HNSCC using multi-omics (genomics, epigenomics, transcriptomics, proteomics, single-cell sequencing, etc.) data. The results showed that *HMGA1* indicated a poor prognosis for HNSCC patients, may be associated with increased TMB and tumor purity, was significantly associated with the occurrence, development, migration and proliferation of HNSCC, and was enriched in malignant cells. Upstream analysis and dual-luciferase reporter assay showed that rs1759628 and $cg25207224_{HMGA1}$ may potentially regulate *HMGA1*, and $cg25207224_{HMGA1}$ maybe a non-invasive in-vitro predictive marker for prognostic prediction of HNSCC.

The HMGA1, located on chromosome 6p.21, binds to the minor groove of the DNA chain rich in adenine and thymidine (AT) through three AT hooks[19], and interacts with the transcription mechanism to control the transcriptional activity of multiple genes. As mentioned above, the role of HMGA1 in HNSCC has not yet been elucidated. Differential analysis and

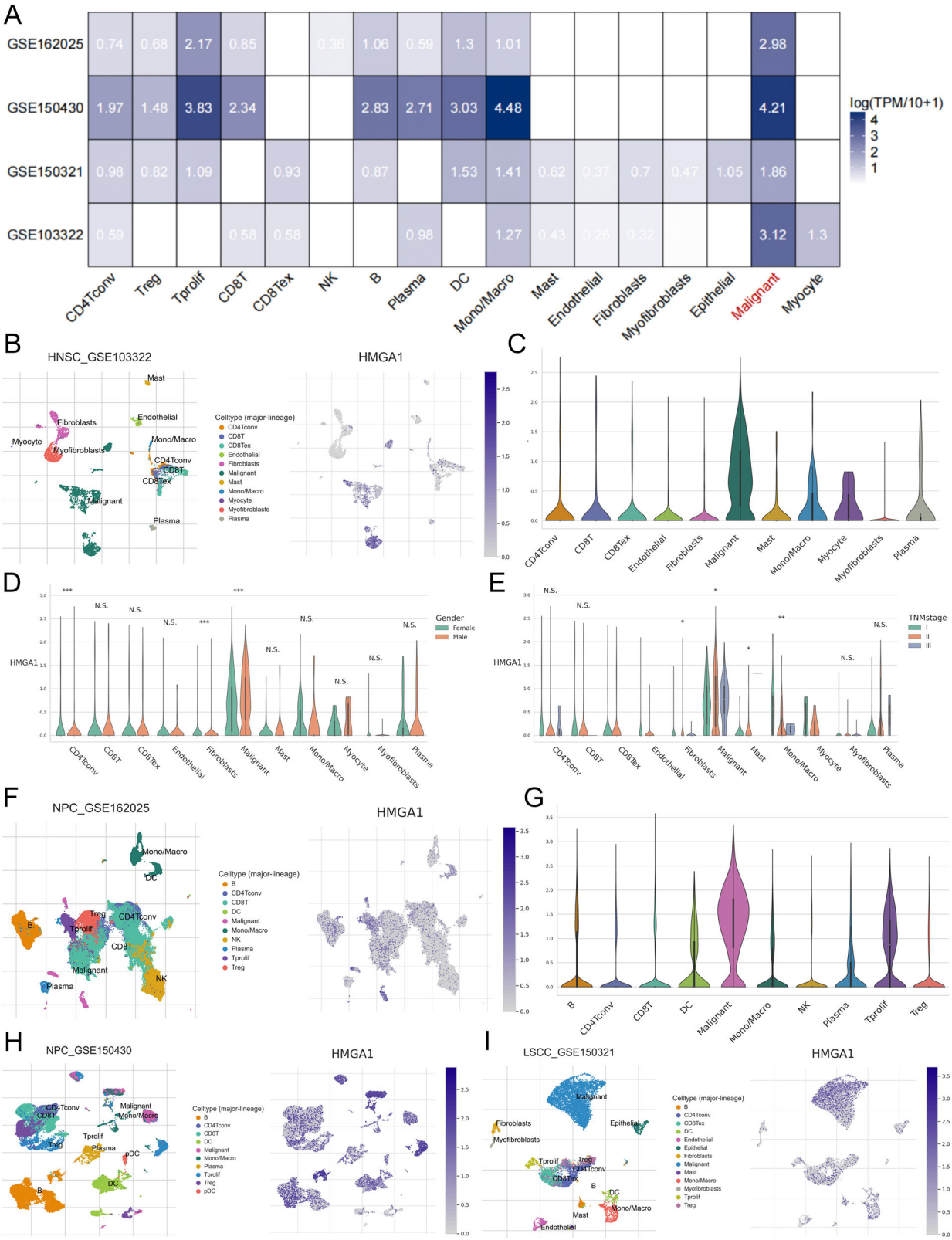

Fig. 5 | Single cell sequencing analysis of *HMGA1* expression in various immune cells of HNSCC. A Heatmap displaying *HMGA1* expression in different cell types from four GEO datasets. **B** Feature plot and **C** violin plot of the distribution of *HMGA1* among various cell types in dataset GSE103322. Subgroup analysis of the distribution of *HMGA1* stratified by **D** gender and **E** clinical stage. **F** Feature plot and **G** violin plot of the distribution of *HMGA1* among various cell types in dataset GSE162025. The distribution of *HMGA1* among various cell types in dataset **H** GSE150430 and **I** GSE150321.

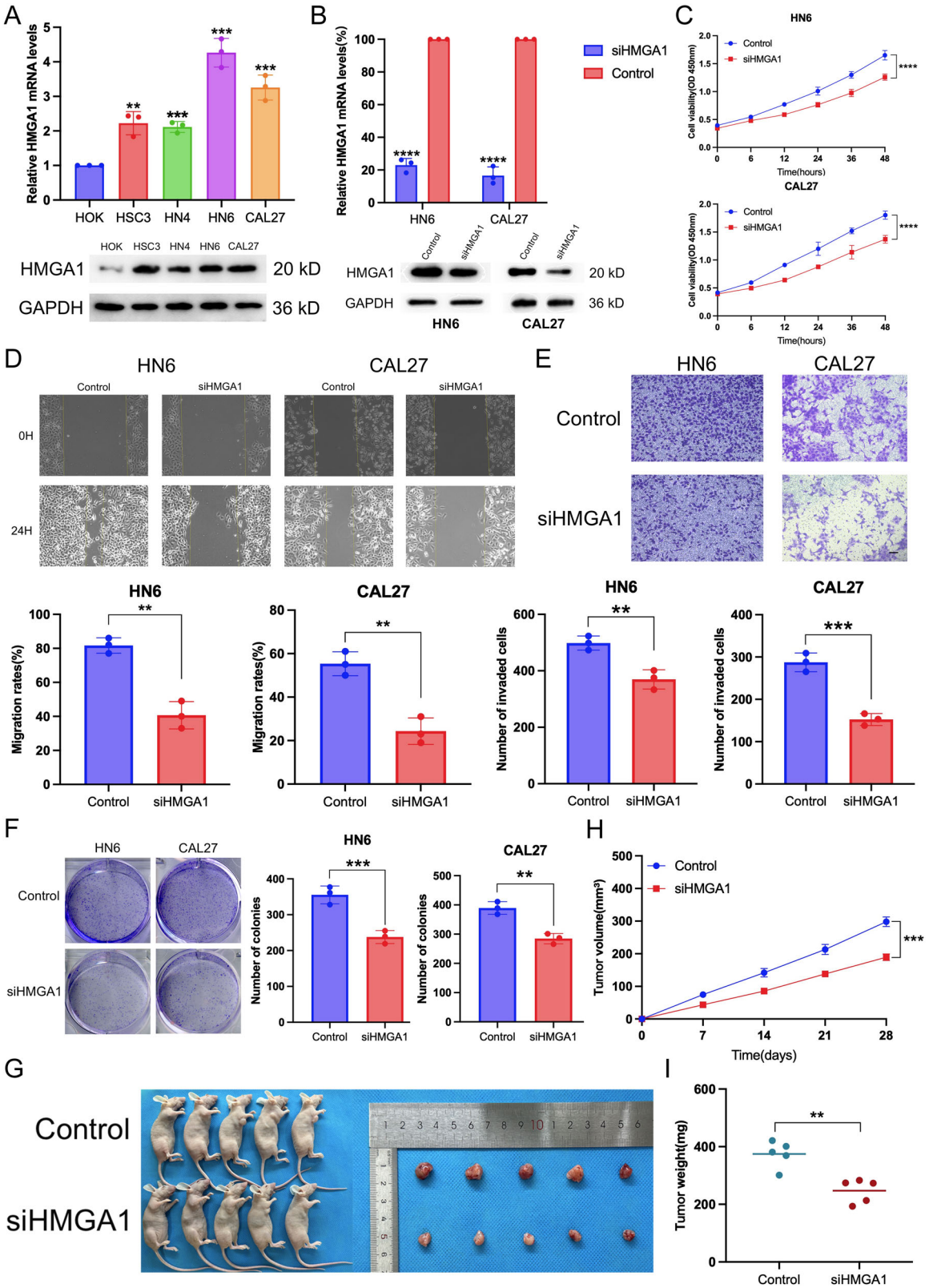

Fig. 6 | HMGA1 accelerated HNSCC cell proliferation, migration, and invasion. **A** PCR and WB analysis demonstrated significantly elevated HMGA1 in HNSCC cells compared to HOK cells. **B** The transfection efficiency of siHMGA1 detected by PCR and WB analysis. **C** The effect of siHMGA1 on the proliferation of HN6 and CAL27 cells in-vitro was detected by CCK-8 assay. **D** The effect of siHMGA1 on HNSCC migration was verified by wound healing experiment. **E** Transwell assay and **F** colony formation experiments demonstrated the invasion and colony formation ability of siHMGA1 on HNSCC cells. The data was presented as the means ± SDs from three independent experiments. Scale bars, 100 μm. **G** Nude mouse and tumor images derived from CAL27 cells transfected with Control and siHMGA1. Statistical graph of tumor volumes (**H**) and weights (**I**) of CAL27 cells transfected with Control and siHMGA1.

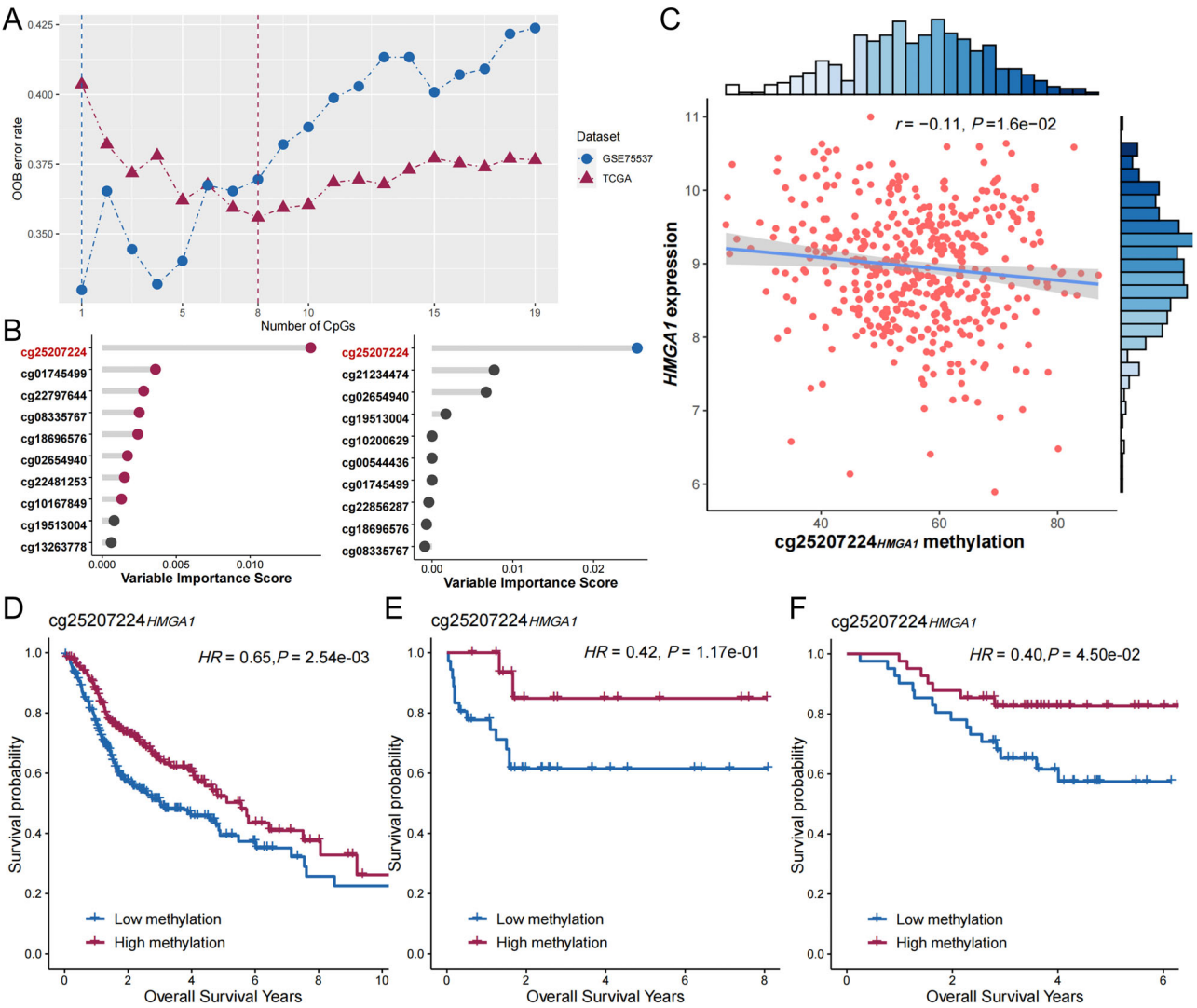

Fig. 7 | Weighted random forest analysis suggested that cg25207224$_{HMGA1}$ could potentially regulate *HMGA1* expression, and was significantly associated with the prognosis of HNSCC. A The results derived from weighted random forest for top 8 CpG probes in the discovery phase reaching the lowest out-of-bag (OOB) error rate, for top 1 CpG probes in the validation phase reaching the lowest error rate. **B** Top CpG probes with their variable important score in the discovery phase and validation phase. **C** Correlation analysis of cg25207224$_{HMGA1}$ methylation and *HMGA1* expression. Kaplan–Meier survival curves of HNSCC patients with high and low DNA methylation level of cg25207224$_{HMGA1}$ in **D** TCGA, **E** GSE75537 and **F** GSE52793, respectively.

survival analysis suggested that the high expression of *HMGA1* gene and protein in HNSCC tumor tissues, and the association with poor prognosis of HNSCC suggested the carcinogenic effect of *HMGA1*. There was no significant difference in *HMGA1* expression between HPV± patients, which may indicate that the effect of *HMGA1* on HNSCC patients was not specifically affected by HPV[20]. A series of immune landscape analyses showed that with the increase of *HMGA1* expression, the tumor burden on the body increased, accompanied by a decrease in immune score and stromal score, and an increase in tumor purity, which may mean difficulty of immunotherapy and the malignant nature of the tumor. It is worth noting that increased *HMGA1* expression simultaneously indicated NK cell dormancy and mast cell activation. During tumor progression, decreased NK cell function indicates poor prognosis[21]. Mast cell aggregation secreted a variety of mediators that promote angiogenesis and inhibit immune responses, including vascular endothelial growth factor (VEGF), histamine, tumor necrosis factor (TNF-α), and interleukin 18 (IL-18)[22], and was involved in tumor lymph node metastasis[23].

Since the expression of *HMGA1* and immune checkpoint genes showed a statistically significant but weak correlation, we turned to explore the potential interaction between *HMGA1* and immune checkpoints. The results showed that *HMGA1* had significant interactions with *SIGLEC15* and *TNFRSF18*. Some studies have shown that *SIGLEC15* may be the most significantly altered immune checkpoint molecule in HNSCC patients, which can directly inhibit $CD8^+$ T cell function to promote immune escape[2–4]. The gene expression and methylation sites of *TNFRSF18* have been shown to be associated with the interferon-γ signature, mutation burden and overall survival of HNSCC[25]. Therefore, we constructed the spatial conformation of the interaction terms. It can be seen that the spatial conformation of the binding part presented high credibility (pLDDT > 90). The two genes may have a certain antagonistic or synergistic interaction through protein binding.

High expression of *HMGA1* promoted the occurrence, development, proliferation and migration of multiple cancers, including myeloproliferative neoplasms[26], endometrial cancer[27], breast cancer[13], colorectal cancer[14] and prostate cancer[16], which was consistent with our results of single-cell analysis and basic experiments. *HMGA1* was enriched in malignant cells, leading to cell migration and proliferation, and increasing the malignancy of HNSCC.

Despite progress in understanding the mechanisms by which *HMGA1* driven malignancy, the upstream regulators that control its activation

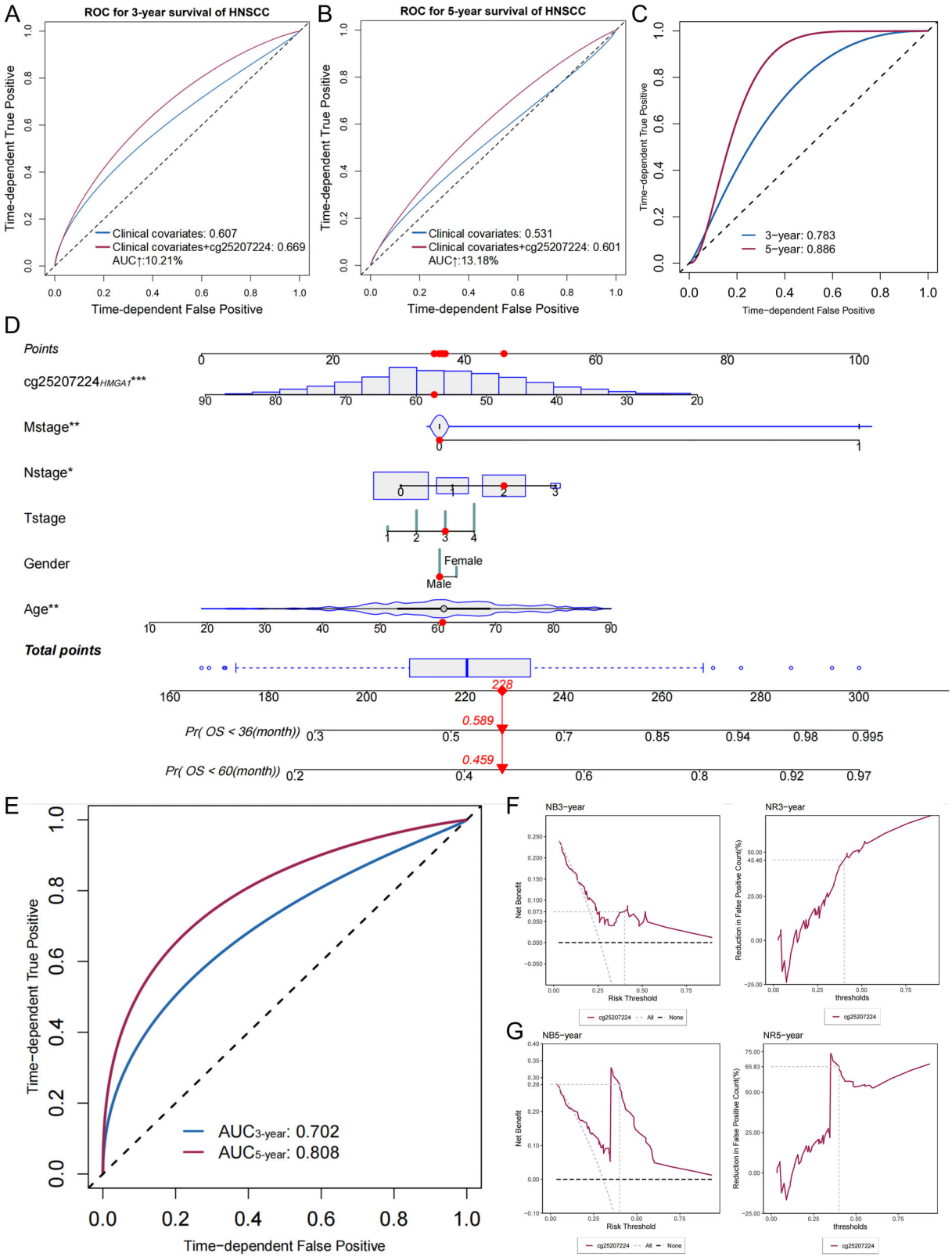

Fig. 8 | Cg25207224$_{HMGA1}$ improved the prognostic accuracy of HNSCC patients, and maybe reliable site for non-invasive prediction of HNSCC prognosis. ROC curves were presented for both **A** 36-month and **B** 60-month survival prediction. The AUC increase (%) was evaluated by comparing model adding cg25207224$_{HMGA1}$ and the model with only covariates. **C** Prediction accuracy of model adding cg25207224$_{HMGA1}$ in the internal testing set (GSE75537). **D** Nomogram of the model adding cg25207224$_{HMGA1}$, the value of each predictor could 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. **E** Prediction accuracy of cg25207224$_{HMGA1}$ in the external validation set (GSE52793). **F**, **G** Decision curve analysis of cg25207224$_{HMGA1}$ as site for non-invasive prediction of HNSCC prognosis. The net benefit (NB) and net reduction (NR) of patients avoided unnecessary interventions were given at threshold (0.4) for both **F** 36-month and **G** 60-month survival.

remained largely unknown. Unfortunately, the eQTL analysis did not find any SNP site significantly associated with *HMGA1* expression, and the existing MR studies did not suggest a significant association between *HMGA1* and HNSCC. This may be due to the small sample size of our self-tested samples and the difficulty of cross-population verification (GWAS and eQTL samples were both Chinse Han populations, but TCGA samples were mostly Caucasian population). Therefore, we retreated and sought whether there were significant epigenetic loci regulating *HMGA1* expression and were significantly associated with the prognosis of HNSCC. By applying the weighted random forest algorithm (*Ranger*), we screened out a DNA methylation probe $cg25207224_{HMGA1}$, which was significantly negatively correlated with *HMGA1*, and after multi-population verification, its low methylation was significantly associated with poor prognosis of HNSCC. The HNSCC prognostic model combining $cg25207224_{HMGA1}$ and clinical covariates showed great prognostic accuracy, both in the training phase and the validation phase.

There is growing interest in developing salivary biomarker tests for HNSCC due to the potential for early diagnosis as well as cost and availability advantages[28]. Surprisingly, even without combining clinical variables, just using oral rinse samples, DNA methylation of $cg25207224_{HMGA1}$ can accurately predict the prognosis of HNSCC patients. According to the global burden of disease, there were nearly 900,000 new HNSCC patients and 450,000 deaths worldwide each year. Assuming that HNSCC patients with a probability of death ≥0.4 need clinical intervention[29], our results were net reduction $(NR)_{36\text{-month}} = 45.46\%$ and $NR_{60\text{-month}} = 65.83\%$, which indicated that compared with the most extreme strategy of intervening in every HNSCC patient, our model can help reduce 409.14 thousand (900 × 45.46%) and 592.47 thousand (900 × 65.83%) unnecessary interventions, respectively, to achieve short-term and long-term survival results. In the future, we may be able to provide patients with the maximum benefit and provide cost-effective precision medicine by customizing $cg25207224_{HMGA1}$-based saliva test strips. Therefore, our manuscript may be a proof of concept.

Our study had several strengths. First, to our knowledge, this may be the first study combining multi-omics data to reveal the impact of *HMGA1* in HNSCC, which provided new insights into the prevention and treatment of HNSCC patients. Second, focusing on the main effect and potential interaction effects of genes, we revealed the complex role of HMGA1 in the immune landscape of HNSCC. Third, through rigorous screening process, we found $cg25207224_{HMGA1}$, which can improve the prognostic accuracy of HNSCC patients and maybe a non-invasive in-vitro predictive marker.

In summary, this study revealed the role of *HMGA1* in the tumorigenesis, prognosis and immune microenvironment of HNSCC at multiple omics levels. The results provide potential therapeutic targets for HNSCC patients and may help to identify high-risk HNSCC populations.

Methods

Study population with transcriptomic or proteomic data

The level-3 TCGA-HNSCC gene expression data (including 44 paired tumor and adjacent tissues) were obtained from the University of California-Santa Cruz Xena browser (http://xena.ucsc.edu/), whose tumor sites were mostly tongue, larynx or overlapped lesions of lip, oral cavity, and pharynx. TCGA mRNA sequencing data processing and quality control (QC) were performed by the TCGA working group. Level-3 mRNA expression data was composed of fragments per kilobase per million mapped reads (FPKM) values, and was transformed into transcripts per kilobase million (TPM) values for association analysis. All gene expression levels were $\log_2$-transformed before analyses. Of these patients with transcriptomics data, 499 patients had epigenetic data and 493 patients had somatic mutation data.

Totally, 81 newly diagnosed HNSCC patients without any treatment before the operation, were enrolled from the Affiliated Stomatology Hospital of Nanjing Medical University (ASHNJMU) into NJHNCC cohort. Surgically resected tumors and adjacent normal oral epithelial tissues were obtained from these patients. Clinical information was obtained through medical records while demographic information including age, gender, smoking, alcohol and Betel nut chewing history was collected from each patient using a standard questionnaire. Frozen tumors and matched normal specimens were confirmed by two independent pathologists. Samples were frozen in liquid nitrogen. HNSCC tissues with the malignant cell purity of over 70% were selected for RNA extraction and sequencing. Of these, 67 patients had peripheral blood samples available for eQTL (expression quantitative trait locus) analysis.

The CPTAC cohort (https://proteomic.datacommons.cancer.gov/) performed a comprehensive integrated proteogenomic characterization of 108 HNSCC patients using global proteome and phosphoproteome analysis from an optimized workflow for mass spectrometry of tissues using isobaric tags tandem mass tags-11[30]. Clinical and demographic information of transcriptomics and proteomics data were provided in Table S6.

Study population with single-cell sequencing data

To provide detailed cell type annotations at the single-cell level, enabling further exploration of *HMGA1* expression on development of HNSCC, we downloaded four single-cell RNA sequencing (scRNA-seq) datasets, including one HNSCC datasets (GSE103322[31]), one laryngeal squamous cell carcinoma dataset (GSE150321[32]) and two nasopharyngeal carcinoma datasets (GSE150430[33] and GSE162025[34]) (Table S7). We utilized the Tumor Immune Single-cell Hub 2.0 database for scRNA-seq analysis of HNSCC in HNSCC[35].

Study population with genomic data

We enrolled patients with HNSCC from four hospitals in Jiangsu Province, including the ASHNJMU, the First Affiliated Hospital of Nanjing Medical University, Nanjing Drum Tower Hospital, and Jiangsu Provincial Cancer Hospital from January 2009 to May 2020 in the genome-wide association study (GWAS) as NJHNCC genomic cohort. Two pathologists confirmed the diagnosis. Patients with tumors in other organs or a history of chemotherapy or radiotherapy, and those refused to participate the study were excluded. Clinical information, including age, gender, smoking status and clinical stage, was obtained from medical records or a standard questionnaire filled out in a face-to-face interview. The control subjects were taken from a community-based screening program. This program was carried out in the Jiangsu Province of China which included more than 30,000 participants with no family history. The general characteristics of the control subjects, such as age and gender, were frequency matched with cases. In order to extract the DNA, about 5 ml of venous blood was extracted from every participant. We performed QC procedures before imputation by excluding individuals with call rates <95%, gender discrepancies, familial relationships (PI_HAT > 0.25) or extreme heterozygosity rates (more than 6 standard deviations from the mean). Single nucleotide polymorphisms (SNPs) with call rates <95%, minor allele frequencies (MAFs) < 0.05 or $P < 1 \times 10^{-6}$ for Hardy-Weinberg equilibrium (HWE) were excluded. Finally, we recruited 1316 HNSCC cases and 3199 frequency-matched controls (Table S8), and filtrated all SNPs located on the region including *HMGA1* and 10 kb of its upstream and downstream by using UCSC hg19 database (http://genome.ucsc.edu/). The 67 samples used for eQTL analysis were also included in the case group of this association study. According to the Declaration of Helsinki, the NJHNCC study has been approved by Ethics Committee of Nanjing Medical University ((2024)144). Written informed consent was obtained from all participating individuals.

Study population with epigenomic data

The level-3 TCGA-HNSCC DNA methylation and two additional independent GEO DNA methylation datasets (GSE75537[36] and GSE52793[37]) were downloaded. 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. We use *R* package *CHAMP* to process level-3 DNA methylation data. Ineligible CpG probes are removed if they met any of the QC criteria: (i) non-CpG probes, (ii) common 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. Figure S9 describes the details of the QC process. Subjects without overall survival time are also removed. Finally, 634 subjects (Table S9) and 361,065 CpG probes are remained in the subsequent association analysis.

Transcription and protein analysis of HMGA1 in pan-cancer level

HMGA1 gene expression differential analysis was performed using *TIMER2.0*[38] in pan-cancer level, as well as survival analysis was performed using *GEPIA2*[39]. Gene expression and protein abundance of HMGA1 in HNSCC patients was analyzed based on paired or unpaired *t*-test differential analysis. To further analyze the differential expression of HMGA1 at the protein level, we downloaded HNSCC immunohistochemical images with their corresponding normal tissues from the Human Proteome Atlas (HPA) database (https://www.proteinatlas.org/). Kaplan-Meier survival curves were drawn to illustrate the survival difference among patients based on the median of *HMGA1* expression.

According to the gene expression level of *HMGA1* (0–25%, 25–50%, 50–75% and 75–100%), we divided the HNSCC patients in TCGA into four groups. We performed differential gene expression analysis using analysis of variance (*aov* function in *R* package *multcomp*) and Tukey's honestly significant difference (HSD) test (*TukeyHSD* function in *R* package *multcomp*) to explore the molecular characteristics of these four subgroups. Genes with false discovery rate (FDR)-$q \le 0.05$ and HSD-$q \le 0.05$ were considered as differentially expressed genes (DEGs). We performed functional annotation and gene enrichment pathway analysis of the DEGs using the *WebGestaltR* package in *R*. Based on the somatic mutation data from TCGA, we conducted differential analysis of genomic mutations among four subgroups using R package *maftools*. We applied the *ESTIMATE* algorithm based on gene expression data to explore the pattern of tumor immune microenvironment (TIME)[40], accomplished differential analysis of immune checkpoint expression using *CIBERSORT*, a linear support vector regression-based deconvolution algorithm[41] to determine the composition of 22 tumor-infiltrating immune cells (TIICs), and explored the differences in immune checkpoint expression. We also performed the correlation analysis between immune checkpoints expression[42], composition of TIICs, TIME, and *HMGA1* expression, respectively, revealed potential interactions, and constructed the spatial conformation of potential interaction terms using *AlphaFold 3*[43]. Finally, we attempted to analyze the molecular docking of HMGA1 and common chemotherapeutic drugs using the Glide module in Schrödinger.

Genetic analysis, eQTL analysis and MR analysis of *HMGA1*

We used logistic regression models to test the effects of SNPs located in *HMGA1* and its upstream and downstream 10 kb range on the occurrence of HNSCC. *FastQTL* was used to calculate the association between genetic variants and *HMGA1* gene expression using a linear regression model adjusted for age, sex, the top 5 principal components and the top 15 probabilistic estimation of expression residuals factors. The eQTLs were defined as their $P < 5 \times 10^{-8}$. Finally, we used *DMRdb*, a disease-centric Mendelian randomization (MR) study database, to explore the causal association of *HMGA1* with occurrence of HNSCC[44].

Cell lines

Human oral squamous cell carcinoma lines (HSC3, HN4, HN6 and CAL27) and Human Oral Keratinocytes (HOK) cell line were obtained from the cell library of the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. The cell lines were cultured in dulbecco's modified eagle medium (DMEM) supplemented with 1% fetal bovine serum and 1% antibiotic (penicillin and streptomycin) at 37 °C and 5% CO_2. All cell lines tested negative for mycoplasma contamination.

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

Total RNA extraction was performed using TRIzol® Reagent (Sigma-Aldrich, St. Louis, MO, USA) from cultured HNSCC cell lines following the manufacturer's protocol. RNA integrity was verified by 1.5% agarose gel electrophoresis with clear 18S/28S ribosomal RNA band resolution. First-strand cDNA synthesis was carried out with 1 μg total RNA using the PrimeScript™ RT Reagent Kit (TaKaRa Bio, Shiga, Japan) in a 20 μL reaction volume. Quantitative PCR amplification was performed in triplicate using SYBR® Premix Ex Taq™II (Tli RNaseH Plus, TaKaRa Bio) on a QuantStudio™ 7 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA), with cycling conditions: 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. The primer sequences are as follows:

HMGA1: Forward 5′-CGAGAAAAGGACGGCACTGAC-3′,
Reverse 5′-TTGTGGTGGTTTTCCGGGTC-3′;
GAPDH: Forward 5′-GGAGCGAGATCCCTCCAAAAT-3′,
Reverse 5′-GGCTGTTGTCATACTTCTCATGG-3′.

Western blot

Western blot analysis was performed using standard procedures. Total proteins were extracted with RIPA lysis buffer (Beyotime) and separated by 8–12% SDS-PAGE. Electrophoretically transferred PVDF membranes (Millipore) were blocked with 10% skim milk for 2 h at room temperature, followed by overnight incubation at 4 °C with rabbit monoclonal anti-HMGA1(1:5000, Proteintech#29895-1-AP, Rosemont, IL, USA) and rabbit monoclonal anti-GAPDH (1:2000, Proteintech#80570-1-RR, Rosemont, IL, USA). After washing, membranes were probed with HRP-conjugated secondary antibodies for 1 h at room temperature. Protein bands were detected using High-sig ECL substrate (Tanon) and visualized through chemiluminescent imaging.

Construction of siRNA

Small interfering RNAs (siRNAs) directed against HMGA1, along with a nontargeting control siRNA, were obtained from GeneChem (China). For the transient transfection procedure, jetPRIME® DNA & siRNA Transfection Reagent (Polyplus, France) was employed according to the manufacturer's guidelines. The siRNA sequences were CTGCTACCAGCG CCAAATGTT.

CCK-8 assay

For CCK-8 assays, transfected cells were cultured in 96-well plates at a density of approximately 3000 cells per well for 6, 12, 24, 36, and 48 h. At each point in time, a solution of CCK-8 (GLPBIO, USA) was added to each hole and the absorbance at 450 nm was measured by an enzyme-labeler. Finally, the average of the three repeated data points is calculated.

Colony formation assay

Transfected cells were cultured at a density of 1000 cells/well in a 6-well cell culture plate and incubated at 37 °C in a complete medium. After 7–10 days, the cells were washed three times with PBS, fixed with 100% methanol, and stained with crystal violet (Solarbio, China) at room temperature for 30 min. Count the number of colonies in each well. Finally, the average of the three duplicate data points is calculated.

Wound healing assay

The transfected cells were cultured in six-well plates until the density reached 90–100%. Gently scrape the sterile 200 μL yellow pipette tip vertically against the center of the hole. Images were taken with the same microscope at 0 and 24 h after trauma. Wound healing rate was analyzed by Image J.

Transwell assay

For the invasion assay, transfected cells were fully resuspended in serum-free DMEM. A Transwell chamber (Corning, USA) containing 100 μL Matrigel on the bottom was placed on a 24-well plate for invasion assay. Matrigel is

composed of the Matrigel substrate (Corning) and DMEM in a ratio of 1 to 9. DMEM containing 15% fetal bovine serum was added to the lower chamber, and transfected cell suspension (density 1×10^5) was inoculated in the upper chamber. After incubation for 24–48 h, cells on the substrate were fixed with paraformaldehyde and then stained with 0.1% crystal violet for 20 min. The images of 5 fields of view were randomly collected for further analysis.

In-vivo experiments

Female BALB/c nude mice (4–6 weeks old, SPF grade) were obtained and raised at the Animal Core Facility of Nanjing Medical University. Mice were randomly divided into control and siHMGA1 groups ($n = 5$/group) using stratified randomization. OSCC cells (5×10^6 cells in 100 μL PBS) were subcutaneously injected into the right axillary fossa. Tumors were measured twice weekly with calipers, and volume calculated as (length × width^2)/2. After 4 weeks, mice were euthanized by CO_2 inhalation, and tumors were excised and weighed.

Fluorouracil resistance assay

HMGA1-knockdown and control HNSCC cells were treated with a fluorouracil gradient (0, 20, 40, 80, and 160 μM) in serum-free medium for 24 h. To quantify drug resistance, cell viability was assessed using the CCK-8 assay immediately post-treatment. Five technical replicates per concentration were analyzed across three biological repeats.

Dual-luciferase reporter assay

Wild-type (WT) and mutant (MUT) reporter plasmids were obtained from GeneChem (China). For the dual-luciferase reporter assay, cells in 24-well plates were washed twice with PBS and lysed with 100 μL of freshly prepared 1 × Passive Lysis Buffer (PLB) on an orbital shaker (60 rpm, 15 min, room temperature). Following gentle pipetting, lysates were centrifuged (12,000 rpm, 10 min), and supernatants were collected. Aliquots (20 μL) of supernatant were transferred to a 96-well luminometer plate. Firefly luciferase activity was measured immediately after adding 100 μL of Luciferase Assay Reagent II (LARII, prepared according to manufacturer's instructions). Subsequently, 100 μL of Stop & Glo® Reagent (prepared by diluting 50× substrate with Stop & Glo® Buffer) was added to quench Firefly luminescence and activate Renilla luciferase activity. The Firefly/Renilla luciferase activity ratio was calculated to represent relative transcriptional activity.

Epigenetic analysis of HMGA1

We used an extensive computational framework integrating 10 machine learning methods: stepwise Cox regression analysis, least absolute shrinkage and selection operator (LASSO), ridge algorithm, partial least squares regression for Cox models and related techniques (plsRcox) method, CoxBoost, random survival forest (RSF), gradient boosting machine (GBM), efficient neural network (ENET), supervised principal component (SuperPC), and survival support vector machine (SVM) to establish prognostic model using CpG probes locating in *HMGA1*. To identify significant and robust DNA methylation CpG probes of *HMGA1* associated with HNSCC survival, we utilized a *Screening before Testing* analysis strategy and a two-phase study design, which was composed of the discovery (TCGA) and the validation (GSE75537) phases. In the biomarker screening stage, we applied Ranger, a fast implementation of survival random forest algorithm which adjusted for covariates using *R* package *ranger*[45], to evaluate the importance of each CpG probe in *HMGA1*, in the discovery and validation phase, respectively. Covariates were age, gender, T stage, N stage and M stage. Variable importance score (VIS) for each CpG probe of *HMGA1* was estimated and then ranked in descending order. To identify the top-drawer CpG probes, we employed the sliding windows sequential forward feature selection (SWSFS) method, which added the CpG probes into Ranger one by one in the order of VIS. Then, we plotted the number of top CpG probes against the out of bagging (OOB) error, which measured the model performance. These top CpG probes corresponding to the lowest OOB error were screened out in the discovery and validation phase, respectively. Finally, the overlapped top CpG probes across two phases were defined as candidates for further model construction. In the biomarker testing stage, these candidate CpG probes were further tested using Cox regression models adjusted for covariates aforementioned in the discovery and validation phase, respectively. Hazard ratio (*HR*) and 95% confidence interval (CI) were described as per 1% methylation increment. Significant probes were finally retained if they met all following criteria: (i) FDR-$q \le 0.05$ in the discovery phase; (ii) $P \le 0.05$ in the validation phase; (iii) consistent effect direction across two phases.

Kaplan-Meier survival curves were plotted to illustrate the survival differences among patients. Prediction accuracy was presented using time-dependent receiver operating characteristic (ROC) curves and measured using the area under the ROC curve (AUC), which was obtained from the *R* package *timeROC*. The consistency index (*C*-index), the average accuracy of predicted survival across all follow-up years, was also used to estimate the prediction performance. In addition, we performed decision curve analysis (DCA) to evaluate the clinical benefit of using methylation sites to screen out patients at high risk of death. Finally, a nomogram was generated using the *R* package *regplot*, and potential DNA methylation markers were validated in an independent oral rinse dataset (GSE52793).

Continuous variables are summarized as mean and standard deviation (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.

Data availability

The datasets analyzed during the current study are available in the TCGA (https://portal.gdc.cancer.gov). The scRNA-seq and epigenetic datasets are available in the GEO (https://www.ncbi.nlm.nih.gov/geo/). The protein dataset is available in the CPTAC (https://proteomic.datacommons.cancer.gov/). Transcriptome, GWAS and eQTL data of NJHNCC are not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Code availability

The *R* software codes that support our findings are not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Abbreviations

HNSCC	head and neck squamous cell carcinoma
HMGA	the high Mobility Group A
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
CPTAC	the Clinical Proteomic Tumor Analysis Consortium
NJHNCC	Nanjing Head and Neck Cancer Cohort
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
NB	net benefit
NR	net reduction
QC	quality control
FPKM	Fragments per kilobase per million mapped reads
TPM	Transcripts per kilobase million
ASHNJMU	the Affiliated Stomatology Hospital of Nanjing Medical University
eQTL	expression quantitative trait locus
scRNA-seq	single-cell RNA sequencing
GWAS	the genome-wide association study
SNPs	single nucleotide polymorphisms
MAFs	minor allele frequencies
HWE	Hardy-Weinberg equilibrium
BMIQ	Beta-Mixture Quantile
HPA	the Human Proteome Atlas

HSD	honestly significant difference
FDR	False discovery rate
DEGs	differentially expressed genes
TIME	Tumor immune microenvironment
TIICs	Tumor-infiltrating immune cells
MR	Mendelian randomization
HOK	Human Oral Keratinocytes
DMEM	dulbecco's modified eagle medium
RT-qPCR	reverse transcription-quantitative polymerase chain reaction
VIS	variable importance score
SWSFS	the sliding windows sequential forward feature selection
OOB	the out of bagging
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
SD	standard deviation

Received: 26 February 2025; Accepted: 23 July 2025;
Published online: 02 August 2025

References

1. Lu, Y. et al. Establishment and validation of a tumor-infiltrating gammadeltaT cell related prognostic gene signature in head and neck squamous cell carcinoma. *Int. Immunopharmacol.* **132**, 112054 (2024).
2. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
3. Zhao, X., Chen, H., Qiu, Y. & Cui, L. FAM64A promotes HNSCC tumorigenesis by mediating transcriptional autoregulation of FOXM1. *Int. J. Oral. Sci.* **14**, 25 (2022).
4. He, Y. et al. Genetic variants were associated with the prognosis of head and neck squamous carcinoma. *Front. Oncol.* **10**, 372 (2020).
5. Xu, Z. et al. ATHENA: an independently validated autophagy-related epigenetic prognostic prediction model of head and neck squamous cell carcinoma. *Clin. Epigenetics* **15**, 97 (2023).
6. Li, Y. et al. Comprehensive analysis of circRNA expression pattern and circRNA-miRNA-mRNA network in oral squamous cell carcinoma. *Oral. Oncol.* **121**, 105437 (2021).
7. Hu, X. et al. Targeting CXCL8 signaling sensitizes HNSCC to anlotinib by reducing tumor-associated macrophage-derived CLU. *J. Exp. Clin. Cancer Res.* **44**, 39 (2025).
8. Gonzalez-Romero, R., Eirin-Lopez, J. M. & Ausio, J. Evolution of high mobility group nucleosome-binding proteins and its implications for vertebrate chromatin specialization. *Mol. Biol. Evol.* **32**, 121–131 (2015).
9. Chia, L. et al. HMGA1 induces FGF19 to drive pancreatic carcinogenesis and stroma formation. *J. Clin. Investig.* **133**, e151601 (2023).
10. Vignali, R. & Marracci, S. HMGA genes and proteins in development and evolution. *Int. J. Mol. Sci.* **21**, 654 (2020).
11. Wang, L. et al. High Mobility Group A1 (HMGA1): structure, biological function, and therapeutic potential. *Int. J. Biol. Sci.* **18**, 4414–4431 (2022).
12. Zeng, L. et al. STMN1 promotes tumor metastasis in non-small cell lung cancer through microtubule-dependent and nonmicrotubule-dependent pathways. *Int. J. Biol. Sci.* **20**, 1509–1527 (2024).
13. Chang, X. et al. Targeting HMGA1 contributes to immunotherapy in aggressive breast cancer while suppressing EMT. *Biochem. Pharm.* **212**, 115582 (2023).
14. Zhao, Y. et al. High mobility group A1 (HMGA1) promotes the tumorigenesis of colorectal cancer by increasing lipid synthesis. *Nat. Commun.* **15**, 9909 (2024).
15. Tang, C., Lei, X., Xiong, L., Hu, Z. & Tang, B. HMGA1B/2 transcriptionally activated-POU1F1 facilitates gastric carcinoma metastasis via CXCL12/CXCR4 axis-mediated macrophage polarization. *Cell Death Dis.* **12**, 422 (2021).
16. Li, K. J. et al. NAT10 promotes prostate cancer growth and metastasis by acetylating mRNAs of HMGA1 and KRT8. *Adv. Sci.***11**, e2310131 (2024).
17. Yan, B. et al. Reversal of HMGA1-mediated immunosuppression synergizes with immunogenic magnetothermodynamic for improved hepatocellular carcinoma therapy. *ACS Nano* **17**, 9209–9223 (2023).
18. Xue, M. et al. ARTEMIS: an independently validated prognostic prediction model of breast cancer incorporating epigenetic biomarkers with main effects and gene-gene interactions. *J. Adv. Res.* **73**, 561-573 (2025).
19. Liau, S. S. & Whang, E. HMGA1 is a molecular determinant of chemoresistance to gemcitabine in pancreatic adenocarcinoma. *Clin. Cancer Res.* **14**, 1470–1477 (2008).
20. Johnson, D. E. et al. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Prim.* **6**, 92 (2020).
21. Wang, J. et al. Differential impact of TIM-3 ligands on NK cell function. *J. Immunother. Cancer* **13**, e010618 (2025).
22. Hemmerlein, B. et al. Is CCL2 an important mediator of mast cell-tumor cell interactions in oral squamous cell carcinoma? *Int. J. Mol. Sci.* **24**, 3641 (2023).
23. Ammendola, M. et al. Mast cell positivity to tryptase correlates with metastatic lymph nodes in gastrointestinal cancer patients treated surgically. *Oncology* **85**, 111–116 (2013).
24. Zhang, X. Y. et al. Metabolic landscape of head and neck squamous cell carcinoma informs a novel kynurenine/Siglec-15 axis in immune escape. *Cancer Commun.* **44**, 670–694 (2024).
25. Loick, S. M. et al. DNA methylation and mRNA expression of OX40 (TNFRSF4) and GITR (TNFRSF18, AITR) in head and neck squamous cell carcinoma correlates with HPV status, mutational load, an interferon-γ signature, signatures of immune infiltrates, and survival. *J. Immunother.* **45**, 194–206 (2022).
26. Schwaller, J. High-mobility-group protein A1 in MPN progression. *Blood J. Am. Soc. Hematol.* **139**, 2730–2732 (2022).
27. Cai, Y., Hao, M., Chang, Y. & Liu, Y. LINC00665 enhances tumorigenicity of endometrial carcinoma by interacting with high mobility group AT-hook 1. *Cancer Cell Int.* **21**, 8 (2021).
28. Balakittnen, J. et al. A novel saliva-based miRNA profile to diagnose and predict oral cancer. *Int. J. Oral. Sci.* **16**, 14 (2024).
29. Chen, 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* **79**, 104007 (2022).
30. Huang, C. et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* **39**, 361–379.e16 (2021).
31. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
32. Song, L. et al. Cellular heterogeneity landscape in laryngeal squamous cell carcinoma. *Int. J. Cancer* **147**, 2879–2890 (2020).
33. Chen, Y. P. et al. Single-cell transcriptomics reveals regulators underlying immune cell diversity and immune subtypes associated with prognosis in nasopharyngeal carcinoma. *Cell Res.* **30**, 1024–1042 (2020).
34. Liu, Y. et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat. Commun.* **12**, 741 (2021).

35. Han, Y. et al. TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res.* **51**, D1425–D1431 (2023).
36. Krishnan, N. M. et al. A Minimal DNA methylation signature in oral tongue squamous cell carcinoma links altered methylation with tumor attributes. *Mol. Cancer Res.* **14**, 805–819 (2016).
37. Langevin, S. M. et al. Novel DNA methylation targets in oral rinse samples predict survival of patients with oral squamous cell carcinoma. *Oral. Oncol.* **50**, 1072–1080 (2014).
38. Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514 (2020).
39. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
40. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
41. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
42. Tang, Y., Li, C., Zhang, Y. J. & Wu, Z. H. Ferroptosis-related long non-coding RNA signature predicts the prognosis of Head and neck squamous cell carcinoma. *Int. J. Biol. Sci.* **17**, 702–711 (2021).
43. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
44. Zheng, X. et al. DMRdb: a disease-centric Mendelian randomization database for systematically assessing causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases. *Nucleic Acids Res.* **53**, D1363–D1371 (2025).
45. Wei, Y. et al. Epigenetic modifications in KDM lysine demethylases associate with survival of early-stage NSCLC. *Clin. Epigenetics* **10**, 41 (2018).

Acknowledgements

This study was funded by the Natural Science Foundation of China (82273737 to R.Z. and 81672678 to H.Y.), Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0520000 and 2024ZD0520003 to R.Z.) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). H.Y. was partially supported by the 2024 Ministry-Provincial Joint Construction of the Tumor Personalized Medicine Collaborative Innovation Center-Hengrui Medicine Clinical Research Fund General Project (2024CICCPMHR019). R.Z. was partially supported by the Outstanding Young Level Academic Leadership Training Program of Nanjing Medical University. The authors thank TCGA, GEO and CPTAC for contributing clinical and multi-omics data, as well as all study subjects who participated in the study cohorts.

Author contributions

Z.X., H.Y., W.Z., and R.Z. contributed to the study design. Z.X., M.W., and J.C. contributed to data collection, statistical analysis, laboratory experiments and interpretation. Z.X., M.W., and J.C. 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 H.Y., W.Z., and R.Z.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-025-01068-4.

Correspondence and requests for materials should be addressed to Hua Yuan, Wei Zhang or Ruyang Zhang.