



OPEN Integrating multiomics analysis and machine learning to refine the molecular subtyping and prognostic analysis of stomach adenocarcinoma

Miaodong Wang^{1,3}, Qin He^{1,3}, Zeshan Chen^{2✉} & Yijue Qin²

Stomach adenocarcinoma (STAD) is a common malignancy with high heterogeneity and a lack of highly precise treatment options. We downloaded the multiomics data of STAD patients in The Cancer Genome Atlas (TCGA)-STAD cohort, which included mRNA, microRNA, long non-coding RNA, somatic mutation, and DNA methylation data, from the sxdyc website. We synthesized the multiomics data of patients with STAD using 10 clustering methods, construct a consensus machine learning-driven signature (CMLS)-related prognostic models by combining 10 machine learning methods, and evaluated the prognosis models using the C-index. The prognostic relationship between CMLS and STAD was assessed using Kaplan-Meier curves, and the independent prognostic value of CMLS was determined by univariate and multivariate regression analyses. We also evaluated the immune characteristics, immunotherapy response, and drug sensitivity of different CMLS groups. The results of the multiomics analysis classified STAD into three subtypes, with CS1 resulting in the best survival outcome. In total, 10 hub genes (CES3, AHCYL2, APOD, EFEMP1, CYP11B1, ASPN, CPE, CLIP3, MAP1B, and DKK1) were screened and constructed the CMLS was significantly correlated with prognosis in patients with STAD and was an independent prognostic factor for patients with STAD. Using the CMLS risk score, all patients were divided into a high CMLS group and a low CMLS group. Patients in the low-CMLS group had better survival, more enriched immune cells, and higher tumor mutation load scores, suggesting better immunotherapy responsiveness and a possible “hot tumor” phenotype. Patients in the high-CMLS group had a significantly poorer prognosis and were less sensitive to immunotherapy but were likely to benefit more from chemotherapy and targeted therapy. In this study, 10 clustering methods and 10 machine learning methods were combined to analyze the multiomics of STAD, classify STAD into three subtypes, and constructed CMLS-related prognostic model features, which are important for accurate management and effective treatment of STAD.

Keywords Stomach adenocarcinoma, Multiomics, Molecular subtyping, Prognosis, Immunity

Abbreviations

STAD	Stomach adenocarcinoma
CMLS	Consensus machine learning-driven signature
PD-1	Programmed death receptor 1
miRNA	MicroRNA
lncRNA	Long chain non-coding RNA
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
TPM	Transcript per million

¹Department of Traditional Chinese Medicine, Jinhua Central Hospital, Jinhua 321000, Zhejiang, People's Republic of China. ²Department of Traditional Chinese Medicine, People's Hospital of Guangxi Zhuang Autonomous Region, 6 Taoyuan Road, Qingxiu District, Nanning City, Guangxi Zhuang Autonomous Region, People's Republic of China.

³Miaodong Wang and Qin He have contributed equally to this work and should be considered co-first authors. ✉email: chen-zeshan12334@126.com

ssGSEA	Single-sample gene set enrichment analysis
TME	Tumor microenvironment
Enet	Elastic Networks
SuperPC	Supervisory principal component
Survival-SVM	Survival Support Vector Machines
RSF	Random Survival Forest
GBM	Generalized boosted regression modeling
OS	Overall survival
PFS	Progression-free survival
PFI	Progression-free interval
IOBR	Immunotumor Biology Research
TIDE	Tumor Immune Dysfunction and Exclusion
NTP	Nearest Template Prediction
EMT	Epithelial mesenchymal transition
TGF	Transforming growth factor
ITH	Intra-Tumor Heterogeneity

Stomach cancer is highly malignant. According to epidemiologic surveys, there will about 1.1 million new cases of stomach cancer worldwide in 2020; it is the fourth leading cause of death due to cancers worldwide¹. Among them, Asia has the highest incidence and mortality rate of stomach cancer, and the five-year survival rate is less than 20%^{1–3}. The most common type of stomach cancer is stomach adenocarcinoma (STAD), which accounts for more than 95% of all types of stomach cancer^{4,5}. Patients with early STAD may have a better prognosis through radical surgery. However, most patients have no prominent early symptoms, and when diagnosed, they are already in the middle or late stage with metastasis; at this point, they have probably lost the chance of surgery, and thus have a high mortality rate and poor prognosis^{5,6}. The treatments for advanced stomach cancer include chemotherapy, radiotherapy, targeted therapy, and antiangiogenic drugs. However, the efficacy of these methods is limited, the adverse effects are serious, and overall survival is significantly low^{7,8}.

Immunotherapy has made considerable progress in the treatment of stomach cancer. Programmed death receptor 1 (PD-1) monoclonal antibodies have been approved for third-line treatment of advanced stomach cancer, and PD-1 monoclonal antibody in combination with chemotherapy have become the new standard for first-line treatment of advanced stomach cancer⁹, providing new strategies for first-line and backline treatment of advanced STAD^{10,11}. However, a large proportion of patients with STAD do not benefit from immunotherapy, which is limited by tumor heterogeneity and the lack of efficacy predictive biomarkers¹². Therefore, accurate screening of immunodominant populations is necessary, and molecular subtyping can help address this issue. Zhao et al. determined via immunohistochemistry that breast cancers can be divided into five subtypes, of which CD8+ T cells in the IHC-IM subtype infiltrate a greater number of tumor cells and exhibit more potent immunogenicity, implying that this subtype may be more effective for immunotherapy¹³. Hong et al. classified thyroid cancer into four subtypes based on transcriptome sequencing, genomic analysis, and clinicopathological information; among them, there is an immune-enriched subtype in which tumors exhibit an increase in immune infiltration and the overexpression of immune checkpoints¹⁴. These studies are important for accurately screening immunodominant populations for cancer treatment.

High-throughput sequencing technology has developed rapidly in recent times¹⁵. While single-omics studies can unidirectionally provide a large amount of information on a tumors, cancer-host interactions, molecular interactions within cancers, and associations between different histologies characteristics require a multidimensional approach for representation. The use of integrated multiomics studies in conjunction with advanced machine learning algorithms is a highly promising tool for gaining insights into the cancer pathogenesis and heterogeneity^{16,17}. The combination of the two facilitates the development of precise diagnostic strategies for cancer.

In this study, mRNA, microRNA (miRNA) and long non-coding RNA (lncRNA) expression profiles, and genomic mutation and epigenomic DNA methylation data were combined, and three subtypes of STAD were constructed using 10 multiomics integration methods. In total, 60 prognosis-related genes were identified based on differential expression of different subtypes. We developed a Consensus machine learning-driven signature (CMLS) based on prognostically relevant genes via 10 machine learning methods. Finally, the prognosis, immune profile, immune response and drug treatment response of patients with STAD were predicted based on the CMLS score, which served as a framework for accurate stratification of patients with STAD patients and selection of personalized treatment strategies.

Materials and methods

Data preprocessing for multiomics data and multihub cohorts in STAD

We downloaded multiomics data for patients with STAD from The Cancer Genome Atlas (TCGA)-STAD cohort, which include mRNA, miRNA, lncRNA, somatic mutation data, DNA methylation data, and clinical data, from the sxdyc website (<http://www.sxdyc.com/>). Moreover, the sxdyc website performed ID along with the summarization and organization of these multiomics data to facilitate the subsequent clustering analysis. Complete information on STAD in seven cohorts, GSE15459, GSE26253, GSE84437, GSE31210, GSE78220, GSE91061, and GSE135222, was obtained from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>), and information on a clinical trial was obtained from <http://research-pub.gene.com/IMvigor210CoreBiologies>, which is available under a Creative Commons 3.0 license¹⁸. We converted high-throughput sequencing of the transcriptome into transcripts per kilobase million (TPM). All expression profiles from the array were subsequently replicated and normalized. The limma software package was used for differential

analysis of tumor and normal samples¹⁹. Additionally, TPM expression was used to represent RNA sequencing (RNA-seq) data²⁰.

Multiomics consensus integration analysis

Five dimensions of information on the TCGA-STAD cohort samples were matched ($n=348$). The TPM data were log2-transformed. The DNA methylation data were selected to probe the promoter CpG islands²¹. Gene mutation matrix was used to determine whether the had a shifted code insertion, deletion or in-frame insertion, deletion, nonsense or missense or uninterrupted mutation, and splice site or translation initiation site mutation.

The “getElites” function of the MOVICS software package was used to screen the genetic features of the STAD subtype multicombinatorial data²². To screen the top 1,500 genes with the largest variations in mRNAs, lncRNAs, miRNAs and methylation, we set the “method” parameter of the “getElites” function to “mad”. The “method” parameter was set to “cox”, and the clinical data were combined to screen statistically significant prognostic genes ($P < 0.05$). For mutation data, we used the Oncoprint function of the maftools package to obtain the 5000 genes with the highest mutation frequencies. The “method” parameter was subsequently set to “freq” to obtain the top 5% most frequently mutated genes in STAD. Data from the five dimensions were included in the next study.

After the above genetic characterization, we obtained the optimal number of clusters. We used the “getClustNum” function in the MOVICS package to estimate the number of subgroups and finally categorized them into three subtypes. We applied the “getMOIC” function for clustering analysis, which contains 10 clustering algorithms (SNF, COCA, CIMLR, NEMO, Consensus clustering, LRAcluster, IntNMF, iClusterBayes, PINSPPlus, MoCluster). These 10 clustering algorithms were used as inputs to the “methodlist” parameter, and the other parameters were default parameters. Then, we obtained the results of the 10 clustering methods, and use the “getConsensusMOIC” function to integrate the results of the 10 clustering methods, with the “distance” parameter set to “Euclidean”, and the “inkage” parameter was set to “average”²². The integrated clustering result is presented here.

Specific molecular characterization and stability of the consensus subtypes

To determine the differences in the activation status of biological pathways among the three subtypes, we performed single-sample gene set enrichment analysis (ssGSEA) using the R software “GSVA” package to assess the enrichment of relevant pathways in the three subtypes²³. Transcriptional regulatory networks (regulons) were constructed using the Reconstruction of Transcriptional regulatory Networks and analysis of regulons (RTN) R package. Tumor microenvironment (TME) scores, including the stromal score, immune score, and tumor purity, were calculated using the R language “estimate” R package²⁴, and the enrichment of 22 immune cells in the three subtypes was determined by GSVA. For subtype stability, we clustering results were validated using subtype-specific biomarkers from the validation cohort and then the consistency of the consensus clustering was compared with nearest template prediction (NTP)²².

Establishment of a consensus machine learning-driven signature

To obtain prognosis-related genes, we extracted differential genes in three subgroups in the TCGA-STAD dataset. Prognosis-related genes were identified by conducting univariate Cox regression analysis, and $P < 0.05$ was considered to be statistically significant. We constructed a CMLS-related prognostic models based on prognosis-related genes using 10 machine learning 99 combination methods, in which the TCGA-STAD cohort was used as the training set and the GSE15459 cohort was used as the validation set; the 10 machine learning methods included Elastic Network (Enet), Lasso, CoxBoost, Stepwise Cox, Partial Least Squares regression of cox (plsRcox), Supervised Principal Component (SuperPC), Ridge, Survival Support Vector Machine (survival-SVM), Random Survival Forest (RSF), and Generalized Boosted Regression Modeling (GBM). The average C-index of each model was calculated to evaluate the predictive ability of the model. Finally, the best combination of algorithms was identified with robustness and clinical translational significance based on the C-index of the training and validation sets.

Prognostic value of CMLS and prospects for clinical applications

We used the “surv-cutpoint” function of the survminer package to score each sample in the training and validation sets of model results for CMLS and divided the samples into high and low CMLS groups based on the median risk score. The prognostic relationship between CMLS and STAD was assessed using Kaplan-Meier curves. To compare the value of our prognostic features with other \prognostic-related features of STAD in published studies, we systematically searched and included 10 publications and 10 prognostic features and calculated the score for each sample based on the published coefficients. Their prognostic predictive ability was ultimately shown by the C-index. Univariate and multivariate Cox regression analyses of risk scores and associated clinical parameters were performed in the TCGA-STAD cohort to determine whether risk scores and associated clinical parameters could be used as independent predictors of overall survival (OS) in patients with STAD. Based on the obtained predictors, column line plots were constructed using the R software “regplot” package to predict the one-year and three-year OS of patients with STAD in TCGA, and calibration curves were used to confirm that the predicted probability of survival agreed with the actual observed results. Decision curves were used to assess the clinical benefit.

Immune characteristics and immunotherapy response in different risk groups based on CMLS scores

We used the Immuno-Oncology Biology Research (IOBR) R package to analyze TME, immunotherapy response, and immunosuppression in different CMLS risk groups²⁵. The tumor mutational load (TMB) was also determined and compared between the two groups²⁶ and the prediction of patients with STAD was predicted

by TMB combined with CMLS. Subclass mapping and the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm were used in the GSE78220 and GSE135222 datasets to estimate immunotherapy response^{27–29}.

Significance of CMLS scores for drug sensitivity in different CMLS groups

The “oncoPredict” R package was used to predict drug sensitivity in patients with STAD with different CMLS risk scores³⁰, and $P < 0.05$ was considered to be statistically significant.

Statistical analyses

All data were processed and analyzed using the R4.3.0 software through appropriate packages. Unpaired Student's t-test performed to determine the differences between two normally distributed variables. The Wilcoxon rank sum test was performed to estimate the difference between two non-normally distributed variables. The Kruskal–Wallis test was conducted for between-group comparisons of more than two groups. TIDE pairs of lists were tested by conducting the two-sided Fisher exact test. All results were considered to be statistically significant at $P < 0.05$.

Result

Multi-omics consensus prognostic-related molecular subtypes in STAD

After preprocessing all data, we independently identified three isoforms using one of 10 multiomics integrated clustering algorithms (Fig. 1A and B). We combined the clustering results with the unique molecular expression patterns of mRNAs, lncRNAs, miRNAs, epigenetic methylation, and somatic mutations of STAD through a consensus clustering approach (Fig. 1C and E). Our classification system revealed three subtypes, which were strongly associated with overall survival (OS) ($p < 0.001$; Fig. 1F). Among the three subtypes, cancer subtype 1 (CS1) had the best survival outcome (Fig. 1F).

Delineation of STAD integration consensus molecular subtypes

The enrichment of different molecular features in the samples using the ssGSEA algorithm. The results revealed that the three subtypes had different responses to specific treatments, with the immunosuppressive oncogenic pathway significantly enriched in CS2, while CS1 and CS3 were significantly enriched in the radiotherapy pathway. This finding suggested that patients with the CS2 subtype may benefit more from immunotherapy and that CS1 and CS3 patients may benefit more from radiotherapy (Fig. 2A). To investigate the differences in the transcriptome, we further analyzed 23 transcription factors of STAD and potential regulators associated with cancer chromatin remodeling, and found that ERBB2, FOXA1, FOXM1, GATA6, and ERBB3 regulators were significantly activated in CS1 and CS3, whereas androgen receptor (AR), PGR, ESR1, STAT3, FGFR1, RARA,

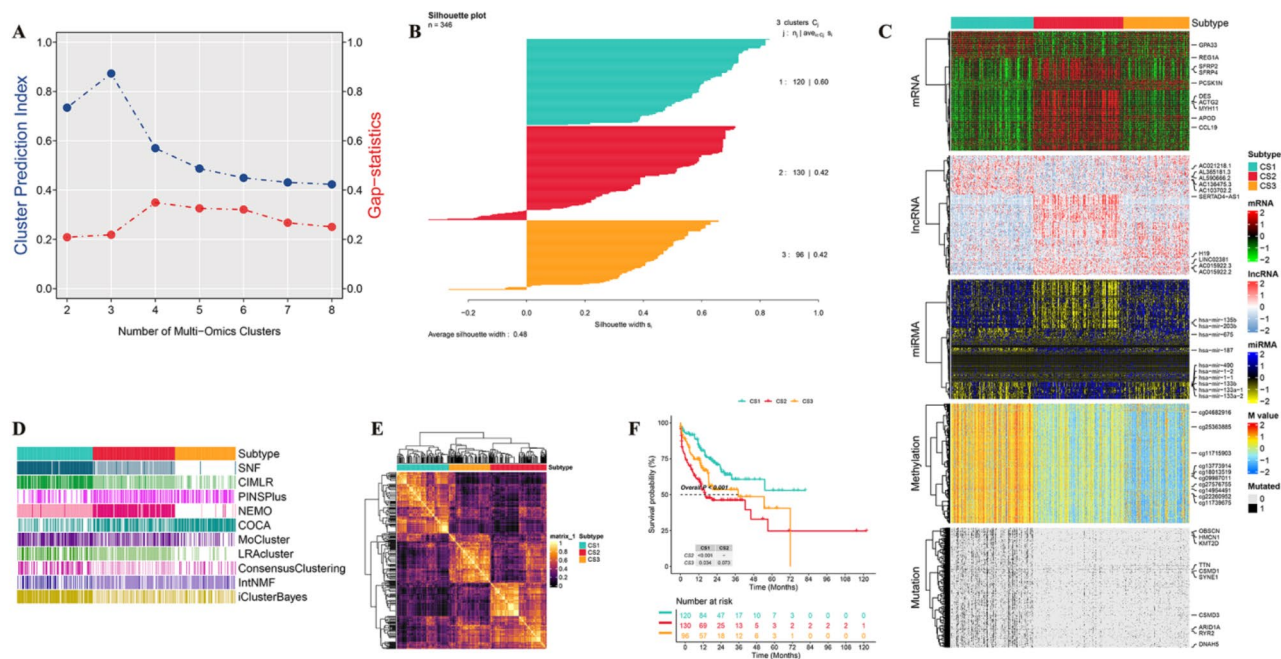


Fig. 1. Multi-omics integration of consensus subtypes of STAD. (A) Cluster prediction indices and gap statistics were analyzed for multiomics clustering. (B) Sample similarity for each subgroup was assessed by calculating Silhouette scores. (C) Integrated heatmap of STAD multiomics data sharing pooled subtypes. (D) 10 multiomics clustering methods for clustering STAD. (E) Consensus clustering matrix of three novel prognostic subtypes obtained by 10 multiomics clustering methods. (F) Survival outcomes of the three subtypes, with CS1 having the best survival outcome.

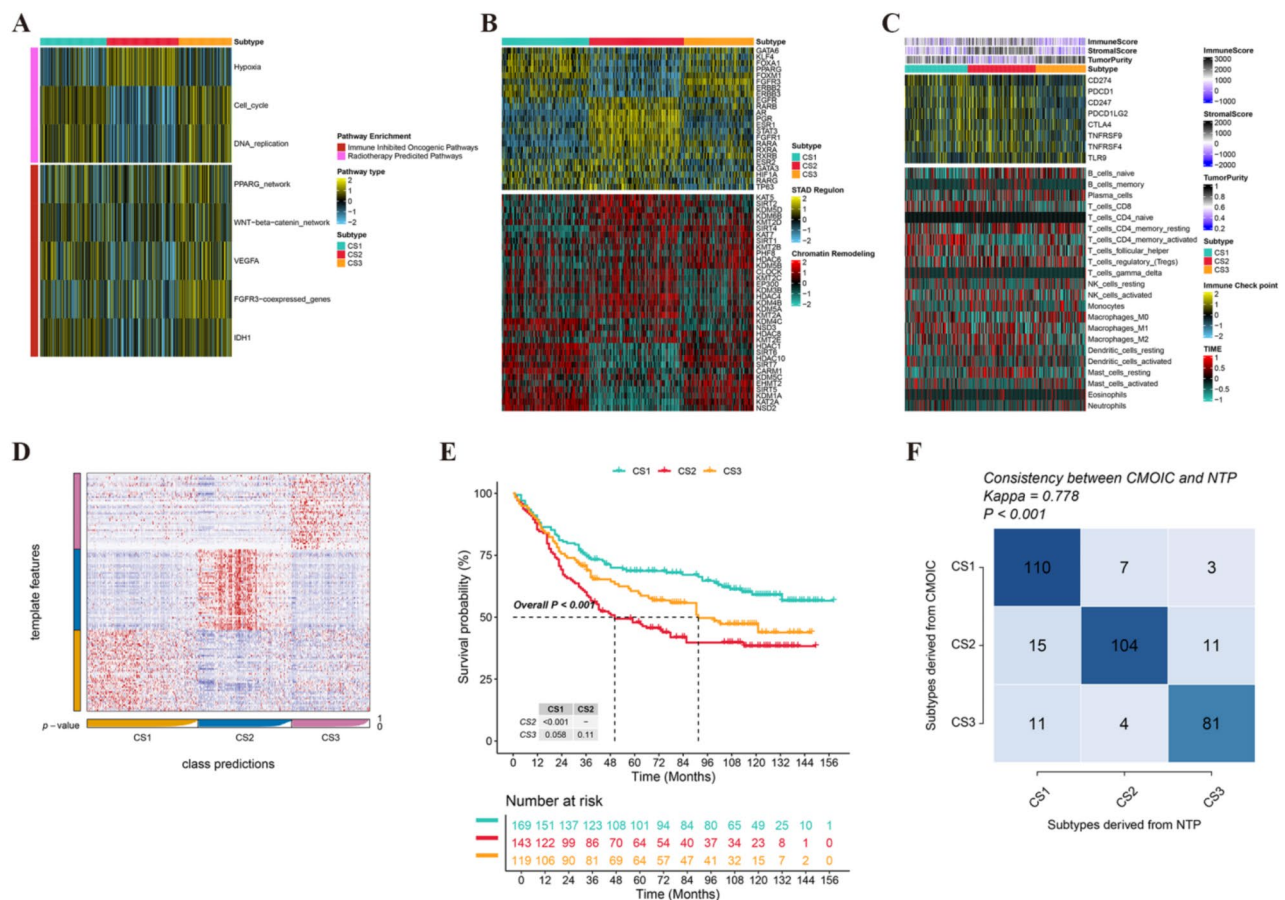


Fig. 2. Molecular landscape and validation of STAD CSs. **(A)** Enrichment profiles of different molecular features of the three isoforms. **(B)** The 23 transcription factor profiles of the three subtypes (top panel) with potential regulators associated with chromatin remodeling of the three subtypes (bottom panel). **(C)** Immunological profiles of the TCGA-STAD cohort, with the top of the heatmap representing the immunity score, stroma score, and tumor purity; the expression of typical immune checkpoint genes is shown in the top panel, and the enriched level of immune cells in the TME is shown in the bottom panel. **(D)** Validation of STAD CS in a recent template of the GSE15459 cohort. **(E)** Survival analysis of patients with STAD CS in the GSE15459 cohort. **(F)** Consistency of CS with NTP in the GSE15459 cohort.

and RXRA were significantly activated in CS2. Similarly, the pattern of differential regulation between different CS subgroups was illustrated in the activity profiles of cancer chromatin remodeling-associated regulators (Fig. 2B). The results suggested that these different transcriptional networks may be important differentiators factors for the three molecular subtypes. The tumor immune microenvironment directly influences tumor immune efficacy, and based on this fact, we quantified the level of enrichment of microenvironmental cells in different subgroups. The results showed that immune cell infiltration increased significantly in CS2 but was low in CS1 and CS3 (Fig. 2C). To confirm the stability of the subtypes, we selected 20 specific upregulated genes associated with the three subtypes, used the nearest template prediction (NTP) algorithm as a classifier, and validated it in the GSE15459 cohort. The results indicated that CS1 in GSE15459 had the best prognosis among all subtypes ($p < 0.005$) (Fig. 2D and E). We also assessed the concordance of CS with NTP ($p < 0.005$; Fig. 2F).

CMLS construction via integrated machine learning

We performed univariate Cox regression analysis in the GSE15459 and TCGA cohorts to obtain 60 prognosis-related genes, which were incorporated into the integrated machine learning framework to construct CMLS. The TCGA dataset was used as a training set and the GSE15459 dataset was used as a validation set; subsequently, consistent models were constructed based on 99 algorithm combinations and the average C-index of each model was calculated to evaluate the predictive ability of the model. As shown in Fig. 3A, CoxBoost + GBM was used to construct the highest average C-index of 0.656, but since the C-index of its training set was only 0.618, considering the insufficient fit of the model of this combined algorithm, we selected Lasso + StepCox[forward] as the prediction model with high accuracy and translation relevance; the training set C-index was 0.666, the validation set C-index was 0.623, and the average C-index was 0.645. The Lasso + StepCox algorithm model consisted of 10 pivotal genes (CES3, AHCYL2, APOD, EFEMP1, CYP1B1, ASPN, CPE, CLIP3, MAP1B, DKK1) (Fig. 3B and C). We subsequently calculated the CMLS scores for each sample in the training and test groups,

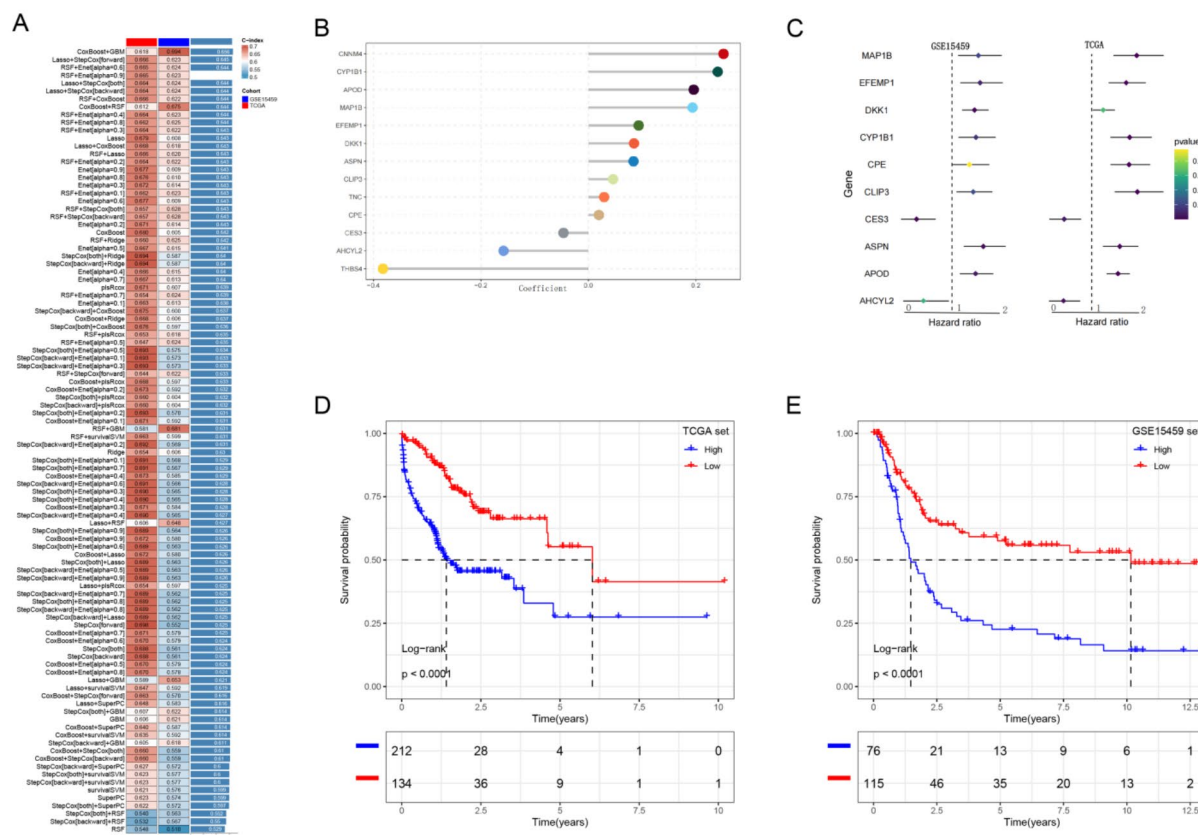


Fig. 3. CMLS generation and prognostic values. **(A)** The C-index for the training and validation sets of the prediction model for the 99 sets of algorithm combinations. **(B)** Hub genes were selected using the Lasso algorithm. **(C)** The Results of one-way Cox regression analysis of pivotal genes in the training and validation cohorts. **(D)** Survival probability of patients with high and low CMLS risk in the TCGA training set. **(E)** Survival probability of patients with high and low CMLS risk in the GSE15459 validation set.

and the results revealed that patients with low CMLS were more likely to survive and have a better prognosis in the TCGA and GSE15459 cohorts (Fig. 3D and E).

To show the prognostic value of the pivotal genes in STAD, we analyzed the Biomarker Exploration for Solid Tumors (BEST) database (https://rookieutopia.com/app_direct/BEST/) via Kaplan-Meier analysis. These 10 genes were significantly associated with overall survival for each patient in the TCGA and GSE15459 cohorts and these 10 genes were significantly associated with progression-free survival (PFS) and progression-free interval (PFI) for each patient in the TCGA cohort. The results revealed that these 10 genes were significantly associated with OS, PFS and PFI in patients with STAD (Figure S1–S3), which implied that the 10 pivotal genes screened by the Lasso + StepCox algorithm were closely related to the prognosis of patients with STAD.

Comparison of the prognostic characteristics of patients with STAD

Many gene expression-based prognostic features have been reported in many diseases because of advances in sequencing technology in recent years. TO compare our prognostic model features with those reported by other researchers, we searched the literature for different prognostic features in patients with STAD over the past five years and ultimately included 10 different features in our subsequent study (Table S1). The incorporated features were related to different biological processes, which included processes such as immune infiltration, apoptosis, and angiogenesis. CMLS showed excellent C-index performance in the TCGA and GSE15459 datasets, which ranked first among all models (Fig. 4A and B). As CMLS has promising clinical applications, we screened independent prognostic factors of STAD by univariate and multivariate Cox analysis (Fig. 4C and D) and integrated them to form a comprehensive column-line diagram (Fig. 4E). The calibration curve of the comprehensive column-line diagram matched the actual situation (Fig. 4F). Decision curve analysis (DCA) showed that the clinical benefit of column-line diagrams for patients was significantly greater than that of CMLS alone (Fig. 4G).

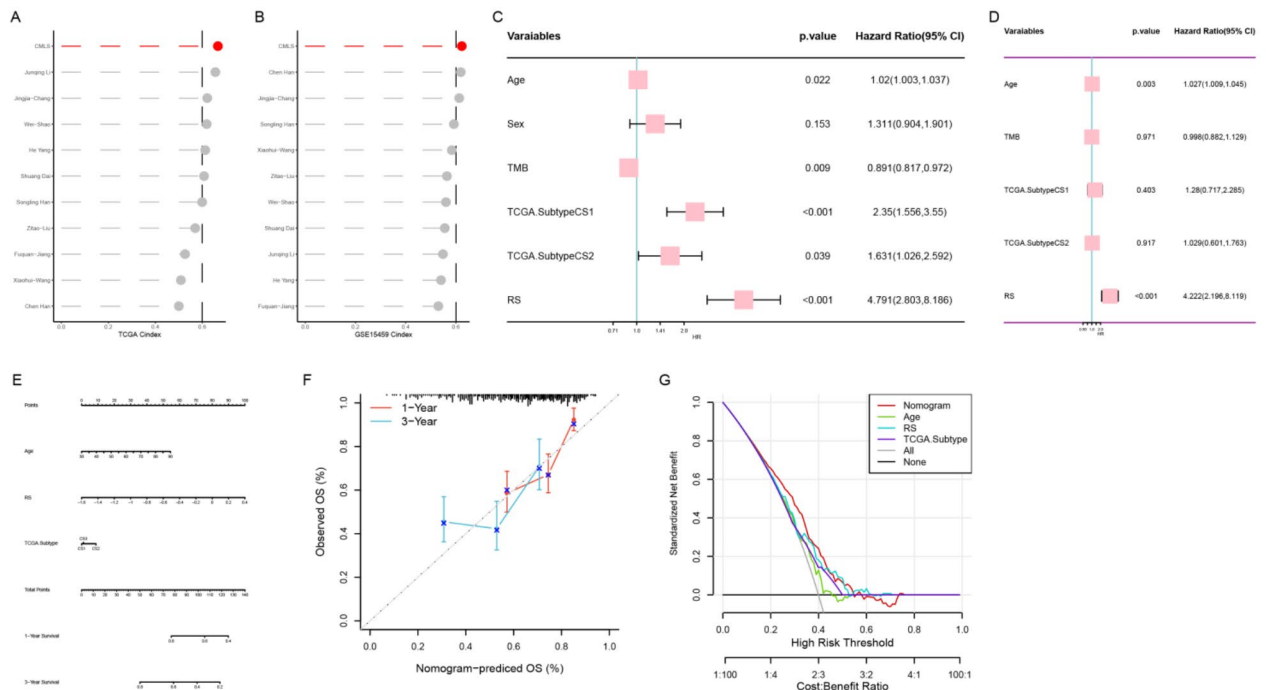


Fig. 4. Clinical practice value of CMLS. (A–B) Comparison of CMLS with 10 other published models in the TCGA-STAD and GSE15459 cohorts. (C) Univariate Cox analysis of hub genes in CMLS. (D) Multivariate Cox analysis of hub genes in CMLS. (E) Column-line diagram of CMLS combined with clinical features. (F) Calibration curves of the combined column-line plots. (G) DCA revealed that the combined column-line diagram was more beneficial than CMLS alone for patients with STAD.

Immunologic characteristics of different CMLS subgroups

We analyzed the TME of STAD using the R package of IOBR and found that the infiltration levels of B cells, T cells, natural killer cells, and M1 macrophages were significantly greater in patients with low CMLS than in patients with high CMLS. This suggesting that immune-infiltrating cells were predominantly enriched in patients with low CMLS and that patients with low CMLS were in a state of immune activation (Fig. 5A). Additionally, the expression levels of fibroblasts and M2 macrophages were significantly lower in patients with high CMLS than in patients with low CMLS, and some immunosuppressive and suppression-related pathways, such as epithelial mesenchymal transition (EMT) and transforming growth factor (TGF)- β pathways, were also enriched in patients with high CMLS than in patients with low CMLS, suggesting that patients with high CMLS were in an immunosuppressive state (Fig. 5B and C). In contrast immunotherapy may be more effective in patients with low CMLS (Fig. 5D). TMB is a biological marker of the degree of tumor mutations, and a positive correlation exists between TMB and antigen recognition by T lymphocytes and the effectiveness of immunotherapy, which can be used to predict the efficacy of immune checkpoint inhibitors^{31,32}. We analyzed the differences in TMB between the two groups of patients with high and low CMLS. The results revealed that TMB was greater in patients with low CMLS (Fig. 5E), indicating a lower CMLS pair with greater immunogenicity. We also combined the CMLS score with TMB for prognostic analysis. The results revealed that patients with lower CMLS and higher TMB had the best prognosis (Fig. 5F).

The ability of CMLS to predict the response to immunotherapy

To determine the role of CMLS in immunotherapy in patients with STAD, we evaluated the distribution of CMLS in patients with different remission levels after administering immunotherapy to patients with STAD and revealed that patients with complete response (CR) and partial response (PR) had significantly lower CMLS scores than patients with progressive disease (PD) versus stable disease (SD) (Fig. 6A). A subclass mapping algorithm for a group of patients with STAD treated with immune checkpoint inhibitors revealed that patients in the low-CMLS subgroup responded better to immune checkpoint inhibitors (Bonferroni corrected $p = 0.000$ (Fig. 6B). Moreover, in the GSE78220 and GSE135222 datasets, that STAD patients with low CMLS who were

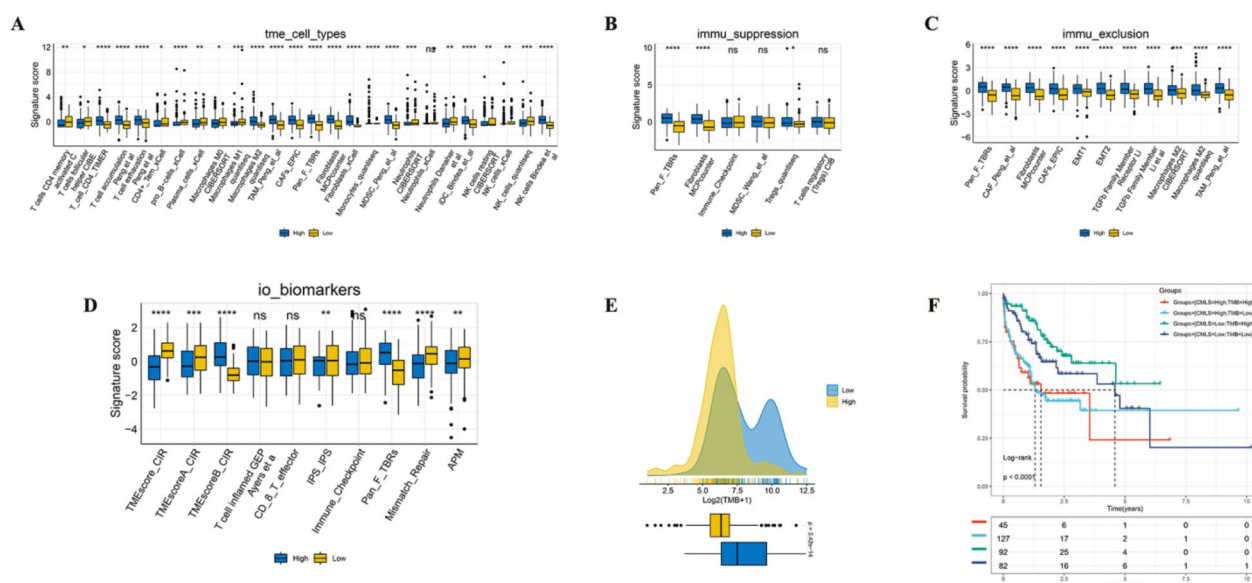


Fig. 5. TME-associated molecular features in patients with high and low CMLS. (A) Distribution of TME immune cell types in the high-CMLS and low-CMLS groups. (B) Distribution of immunosuppressive features in the high-CMLS and low-CMLS groups. (C) Distribution of immune rejection features in the high-CMLS and low CMLS groups. (D) Distribution of immunotherapy biomarkers in the high-CMLS and low-CMLS groups. (E) Distribution of TMB in the high-CMLS and low-CMLS groups. (F) Survival analysis of CMLS combined with TMB; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

treated with immunotherapy had better prognostic outcomes (GSE78220, $p < 0.001$ [Fig. 6C]; GSE135222, $p = 0.017$ [Fig. 6D]).

Analysis of the correlation between CMLS and drug sensitivity

To personalize treatment, we used the R package ‘pRRophetic’ to predict drug sensitivity in STAD patients with different CMLS risk scores. The results were considered to be statistically significant at $p < 0.05$. The high-CMLS group benefited more from drugs such as 5-fluorouracil, cisplatin, gemcitabine, epirubicin, afatinib, and darafenib, and had higher drug sensitivity than the low-CMLS group (Fig. 7A and F), suggesting that the high-risk group may benefit from chemotherapy and targeted therapies. In contrast, the low-CMLS group benefited from only a few drugs, such as dasatinib, BMS-754,807 (IGF-1RIR inhibitor), and WEHI-539Bcl-XL (an inhibitor) (Fig. 7G-I).

Discussion

Stomach adenocarcinoma is a highly aggressive malignancy, and biomarkers associated with cancer stratification and prognosis at the molecular level are urgently needed to develop individualized treatment plans³³. Single-omics studies are limited by their inability to characterize organisms in multiple dimensions. Multiomics allows for more comprehensive and systematic information through epigenomics, as well as enhance ambiguous patterns in gene expression data. Complementary information between multiomics methods can be used to interpret classification results more accurately and improve predictive performance^{34,35}. Wang et al. established 35 primary cell models of prostate cancer by integrating multiomics data of lung cancer and accurately captured the molecular features of prostate cancer and drug responses, providing a basis for the precise diagnosis and treatment of prostate cancer³⁶. Similarly, Gao et al. analyzed multiomics data from patients with breast cancer and found that CXCL12 plays an important role in predicting the response to immunotherapy and the prognosis of patients with breast cancer³⁷. In this study, we used 10 advanced multiomics clustering methods to identify three prognostic subtypes with different characteristics, among which CS1 had the best survival outcome and effectively identified high-risk patients with STAD, which provides some value for stratified management and precision treatment of STAD.

Advanced machine learning algorithms models can assist in cancer diagnosis and prognosis³⁸ and are also effective tools for analyzing and understanding multiomics data of diseases³⁹. Ma et al.⁴⁰ used 10 machine learning algorithms to construct a prognostic model of mitochondrial function and developed a mitochondria-related score, which greatly assisted in the diagnosis and prognosis of stomach cancer patients. Wang et al.⁴¹ constructed a prognostic model based on cancer-associated fibroblast genes and reported that three genes, CDH6, EGFLAM, and RASGRF2, were significantly associated with immunotherapy, drug sensitivity, and prognosis in stomach cancer. To understand the differences in molecular features between different prognostic subtypes. In this study, we used a combination of 99 algorithms from 10 machine learning methods to select the best CMLS to predict the prognosis and immunotherapeutic response in patients with STAD. After multidimensional validation, 10

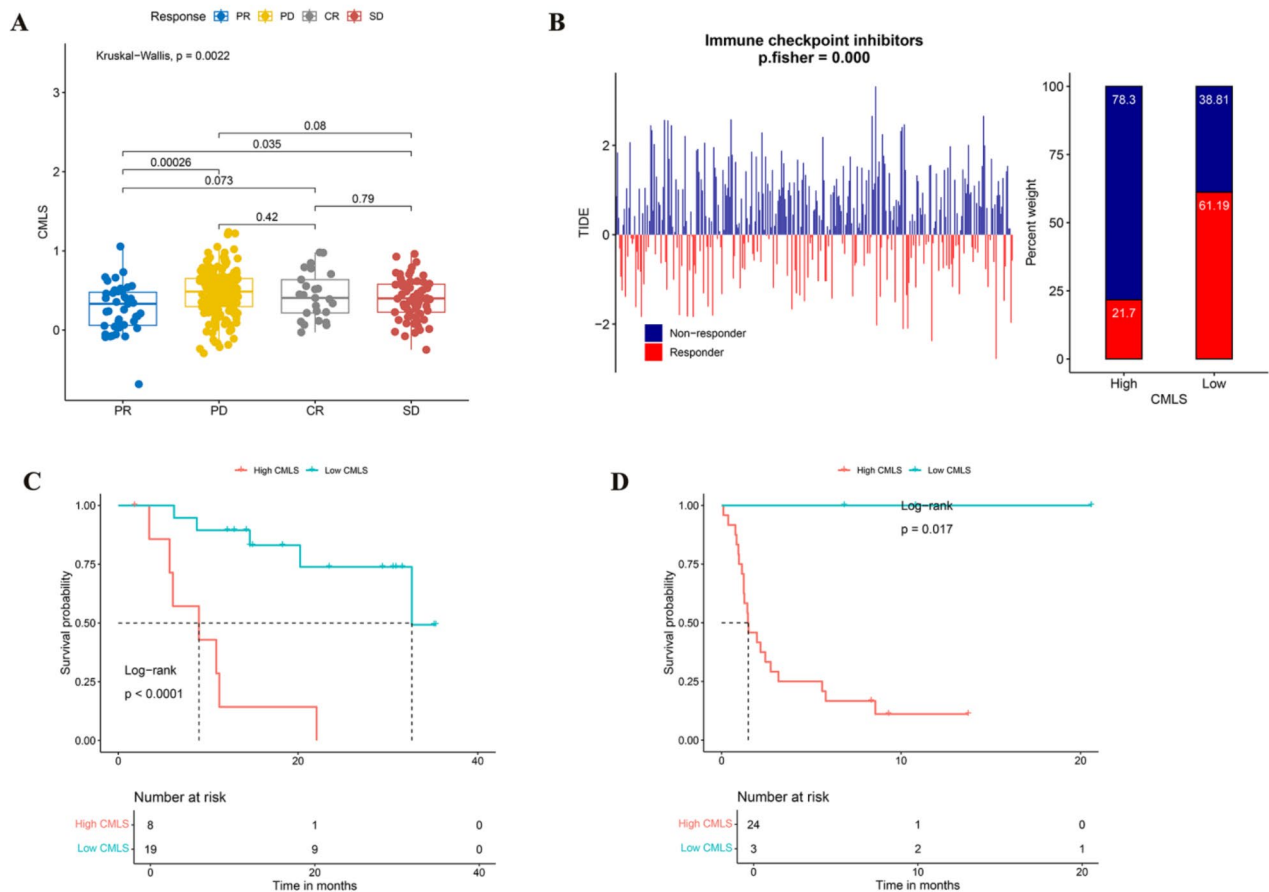


Fig. 6. Value of CMLS in predicting the response to immunotherapy in patients with STAD. **(A)** Distribution of CMLS in patients with STAD who achieved CR, PR, PD, and SD after immunotherapy. **(B)** A subclass mapping algorithm was used to predict the response to immunotherapy in the high-CMLS and low-CMLS groups. **(C)** Survival analysis of patients in the high-STAD and low-STAD groups in the GSE78220 dataset after immunotherapy was administered. **(D)** Survival analysis of patients in both the high-STAD group and low STAD group in the GSE135222 dataset after immunotherapy was administered.

pivotal genes (including CES3, AHCYL2, APOD, EFEMP1, CYP1B1, ASPN, CPE, CLIP3, MAP1B, and DKK1) were identified and used to construct a prognostic model for patients with STAD via Lasso + StepCox [forward]. We further analyzed these 10 genes with OS, PFS and PFI in patients with STAD and found that 10 genes were significantly associated with the prognosis of patients with STAD. Based on the CMLS risk score, we divided all patients into high-CMLS and low-CMLS groups. The results of Kaplan-Meier (KM) analysis and univariate and multivariate Cox regression analyses revealed that patients in the low-CMLS subgroup had better survival and that CMLS could risk stratify patients with STAD according to OS and served as an independent prognostic factor for STAD. CMLS also demonstrated optimal prognostic value in every cohort compared to 10 previously published prognostic characteristics.

Ten hub genes have been investigated in the context of cancer prognosis. For example, CES3 plays a protective role in cancer progression⁴² and serves as an immune-related prognostic marker for colon cancer⁴³. Bioinformatics data suggested that APOD is a component of the stomach cancer risk model, which is associated with immune cell infiltration and cellular senescence in stomach cancer^{44,45}. In a multicenter retrospective study, high APOD expression was an independent prognostic risk factor for stomach cancer patients⁴⁶. AHCYL2 is expressed at low levels in STAD cells, which is associated with copper death in stomach cancer and is a prognostic biomarker for stomach cancer⁴⁷. High expression of EFEMP1 is associated with a good prognosis in stomach cancer patients, and its low expression is associated with stomach cancer differentiation, depth of tumor infiltration, and lymph node metastasis⁴⁸. CYP1B1 is a drug-metabolizing enzyme that is highly expressed in diffuse stomach cancers and is associated with lymphatic invasion and tumor TNM stage⁴⁹. Additionally, high expression of CYP1B1 reduces the sensitivity of stomach cancer cells to cisplatin, which in turn promotes the progression of stomach cancer⁵⁰. CPE is a regulator of growth and metastasis in multiple cancer types and has good prognostic value⁵¹. CPE promotes epithelial-mesenchymal transition in stomach cancer cells, leading to stomach cancer cell proliferation, invasion, and metastasis; high expression of CPE is also associated with low survival and a poor prognosis in stomach cancer patients, which is associated with poor survival and a poor prognosis⁵². A decrease in the expression of CLIP3 promotes glycolysis and induces radioresistance in

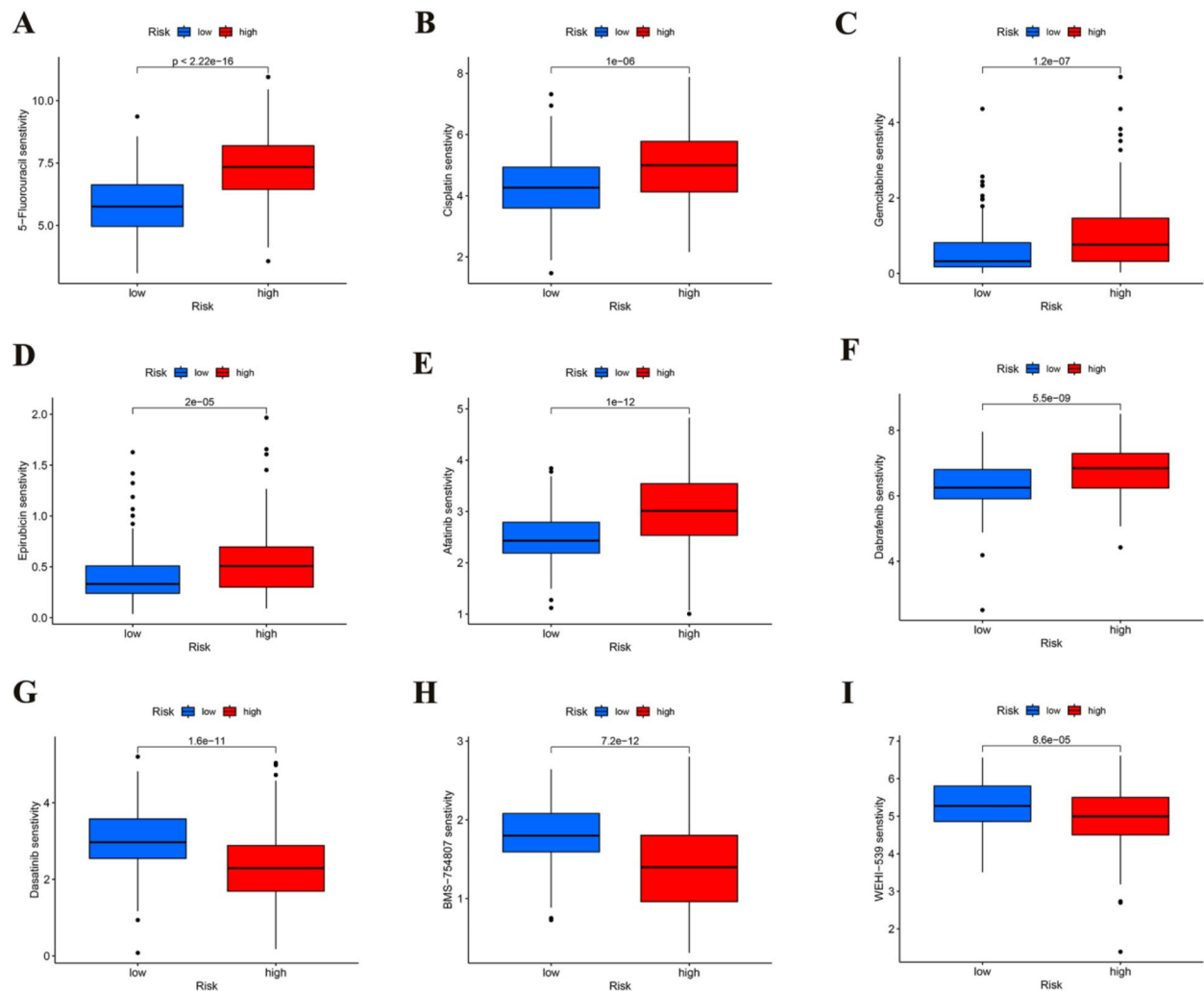


Fig. 7. Potential drugs for patients in the high-CML and low-CML groups. Comparison of the sensitivity of A-I high-risk and low-risk patients to 5-fluorouracil (A), cisplatin (B), gemcitabine (C), epirubicin (D), afatinib (E), darafenib (F), dasatinib (G), BMS-754,807 (H), and WEHI-539Bcl-XL (I).

cancer, and CLIP3 may be a protective prognostic biomarker⁵³. The overexpression of MAP1B, a microtubule protein, is associated with poor cancer prognosis and is an independent prognostic marker⁵⁴. Additionally, the downregulation of DDK1 expression is associated with the proliferation, invasion, and chemosensitivity of stomach cancer⁵⁵. ASPN is highly expressed in stomach cancer, and its low expression inhibits stomach cancer proliferation, migration, and invasion⁵⁶. ASPN can also reprogram the glucose metabolism pathway, thus allowing stomach cancer cells to resist oxidative stress and promoting migration and invasion. These functions of ASPN suggest that it is associated with the poor prognosis of stomach cancer patients⁵⁷.

Although immunotherapy for malignant tumors has shown positive results, many patients do not benefit from it because of intratumor heterogeneity (ITH)⁵⁸. ITH is a key factor in tumor lethality, the failure of immune and targeted therapies, and drug resistance⁵⁹. Using TCGA and GEO databases, some studies have also reported gene expression-based immune landscapes in stomach cancer. For example, Deng et al. developed a degradome-based prognostic signature (DPS) and reported that patients with low DPS STADs were associated with an increase in immune infiltration abundance and TMB scores and a decrease in immune evasion⁶⁰. Huang et al., on the other hand, developed prognostic models using three prognostic genes, and patients in the low-risk group had better potential for immunotherapeutic outcomes⁶¹. In our study, mRNA, miRNA, and lncRNA expression profiles, genomic mutation data, and epigenomic DNA methylation data were combined to reveal immune profiles and genomic alterations in different histological variants and risk cohorts, along with bioinformatics and machine learning approaches. We analyzed the enrichment of dozens of immune-related features in the high-CML and low-CML groups and found that patients with low CML had more enriched immune cells and higher TMBs, suggesting greater antitumor immunity; the results also indicated that patients with low CML may be categorized as having a 'hot tumor' phenotype. However, patients with high CML have more

fibroblasts and M2 macrophages enriched and multiple cancer pathways activated, such as the EMT and TGF- β pathways, implying that these patients may be more consistent with a “cold tumor” phenotype^{62,63}. We also evaluated CMLS scores in patients with different clinical outcomes after immunotherapy and found that patients in complete and partial remission had lower CMLS scores, suggesting that patients with low CMLS scores have better survival outcomes. We performed subclass mapping to demonstrate a better response to immunotherapy in the low-CMLS group, suggesting that the population sensitive to immunotherapy can be screened based on CMLS scores, which has important implications for the precision treatment of patients with STAD.

As patients in the high-CMLS group were insensitive to immunotherapy, we predicted drug sensitivity in STAD patients with high CMLS scores and found that the high-risk group was sensitive to 5-fluorouracil, cisplatin, gemcitabine, epirubicin, and afatinib. Patients in the high CMLS group may benefit from chemotherapy and targeted therapy. Some studies have shown that escitalopram oxalate in combination with 5-fluorouracil has a synergistic inhibitory effects on stomach cancer⁶⁴. Li et al. found that 5-fluorouracil enhanced the ability of tumor necrosis factor-associated apoptosis-inducing ligands to induced apoptosis in stomach cancer cells by inhibiting the MAPK pathway⁶⁵. Cisplatin is also effective in combination with capecitabine for treating human epidermal growth factor receptor 2-negative advanced stomach cancer and is well-tolerated⁶⁶. Moreover, afatinib induces G1 cell cycle arrest and apoptosis for treating trastuzumab-sensitive and trastuzumab-resistant HER2 gene-amplified stomach cancer⁶⁷.

Compared to previous studies, our study has several advantages. First, for more precise and personalized treatment of stomach cancer, we integrated five dimensions of histological data of STAD and used 10 advanced clustering methods to classify STAD patients into three subtypes, which assisted in the stratified management and treatment of STAD. Second, the 10 advanced clustering methods not only yielded more robust typing results but also improved the accuracy of determining the survival time of patients with STAD. Third, we used 10 machine learning methods to construct an STAD prognostic model after clustering. Finally, we screened 10 pivotal genes based on the C-indexes of the training and validation sets and created CMLS, which was significantly associated with aspects of STAD prognosis and the immune response. However, this study needs to be further combined with molecular biology experiments to confirm its biological significance. The clinical value of CMLS also needs to be validated by conducting a multicenter, prospective study in the clinic.

Conclusions

In this study, three molecular subtypes of STAD were identified using 10 clustering methods, which revealed the prognostic differences among the three molecular subtypes and refined the molecular subtypes of STAD, which is important for the stratified management of this disease. Using 10 machine learning algorithms, CMLS was constructed, which exhibited superior performance in predicting the prognosis of patients with STAD. CMLS was significantly correlated with immune efficacy and drug sensitivity in patients with STAD at different risks levels. This finding indicating that CMLS is a promising biomarker. This study provided a strong foundation for the early diagnosis and precise treatment of patients with STAD through multiple combinations and multiple machine learning algorithms.

Data availability

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Received: 9 October 2024; Accepted: 20 January 2025

Published online: 30 January 2025

References

- Sung, H. et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. *Ca-a Cancer J. Clin.* **71**(3), 209–249 (2021).
- Fitzmaurice, C. et al. Global, Regional, and National Cancer incidence, mortality, years of Life Lost, Years lived with disability, and disability-adjusted life-years for 29 Cancer groups, 1990 to 2017: A systematic analysis for the global burden of Disease Study. *JAMA Oncol.* **5**(12), 1749–1768 (2019).
- Ilic, M. & Ilic, I. Epidemiology of stomach cancer. *World J. Gastroenterol.* **28**(12), 1187–1203 (2022).
- Long Parma, D. et al. Gastric adenocarcinoma burden and late-stage diagnosis in latino and non-latino populations in the United States and Texas, during 2004–2016: A multilevel analysis. *Cancer Med.* **10**(18), 6468–6479 (2021).
- Amrani Hassani Joutei, H. et al. Étude des caractéristiques épidémiologiques cliniques et anatomopathologiques de l'adénocarcinome gastrique chez une population marocaine [Study of epidemiological clinical and pathological characteristics of gastric adenocarcinoma in a Moroccan population]. *Ann. Pathol.* **40**(6), 442–446 (2020).
- Bhattarai, S., Gyawali, M. & Regmi, S. Prevalence of gastric cancers among patients undergoing Upper gastrointestinal endoscopies in a Tertiary Care Hospital in Nepal: a descriptive cross-sectional study. *J. Nepal. Med. Association.* **59**(233), 65–68 (2021).
- Yu, Y. et al. Oxaliplatin plus Capecitabine in the Perioperative treatment of locally advanced gastric adenocarcinoma in combination with D2 Gastrectomy: NEO-CLASSIC study. *Oncologist* **24**(10), 1311–e989 (2019).
- Xu, R. H. et al. Efficacy and safety of weekly paclitaxel with or without ramucirumab as second-line therapy for the treatment of advanced gastric or gastroesophageal junction adenocarcinoma (RAINBOW-Asia): A randomised, multicentre, double-blind, phase 3 trial. *Lancet Gastroenterol. Hepatol.* **6**(12), 1015–1024 (2021).
- Wang, F. H. et al. The Chinese Society of Clinical Oncology (CSCO): Clinical guidelines for the diagnosis and treatment of gastric cancer, 2021. (2021). Cancer Communications (London, England). **41**(8), 747–795
- Li, S. et al. Neoadjuvant therapy with immune checkpoint blockade, antiangiogenesis, and chemotherapy for locally advanced gastric cancer. *Nat. Commun.* **14**(1), 8 (2023).
- Jin, X. et al. Recent progress and future perspectives of immunotherapy in advanced gastric cancer. *Front. Immunol.* **13**, 948647 (2022).

12. Shitara, K. et al. Association between gene expression signatures and clinical outcomes of pembrolizumab versus paclitaxel in advanced gastric cancer: Exploratory analysis from the randomized, controlled, phase III KEYNOTE-061 trial. *J. Immunother. Cancer* **11**(6), e006920 (2023).
13. Zhao, S. et al. Molecular Subtyping of Triple-negative breast cancers by immunohistochemistry: Molecular basis and clinical relevance. *Oncologist* **25**(10), e1481–e1491 (2020).
14. Hong, S. et al. Distinct molecular subtypes of papillary thyroid carcinoma and gene signature with diagnostic capability. *Oncogene* **41**(47), 5121–5132 (2022).
15. Machlowska, J. et al. High-throughput sequencing of gastric Cancer patients: Unravelling genetic predispositions towards an early-onset subtype. *Cancers (Basel)* **12**(7), 1981 (2020).
16. Chu, G. et al. Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer. *Mol. Therapy-Nucleic Acids* **33**, 110–126 (2023).
17. Shi, R. et al. APOBEC-mediated mutagenesis is a favorable predictor of prognosis and immunotherapy for bladder cancer patients: evidence from pan-cancer analysis and multiple databases. *Theranostics* **12**(9), 4181–4199 (2022).
18. Yuan, X. et al. Apoptosis-related gene-mediated cell death pattern induces immunosuppression and Immunotherapy Resistance in Gastric Cancer. *Front. Genet.* **13**, 921163 (2022).
19. Sui, Y. et al. Bioinformatics analyses of combined databases identify shared differentially expressed genes in cancer and autoimmune disease. *J. Translational Med.* **21**(1), 109 (2023).
20. Vera Alvarez, R. et al. TPMCalculator: One-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**(11), 1960–1962 (2019).
21. Martisova, A. et al. DNA methylation in solid tumors: Functions and methods of detection. *Int. J. Mol. Sci.* **22**(8), 4247 (2021).
22. Lu, X. et al. MOVICS: An R package for multi-omics integration and visualization in cancer subtyping. *Bioinformatics* **36**(22–23), 5539–5541 (2021).
23. Chen, Z. et al. A novel prognostic signature based on four glycolysis-related genes predicts survival and clinical risk of hepatocellular carcinoma. *J. Clin. Lab. Anal.* **35**(11), e24005 (2021).
24. Zheng, X. et al. Identification and validation of immunotherapy for four novel clusters of colorectal cancer based on the tumor microenvironment. *Front. Immunol.* **13**, 984480 (2022).
25. Zeng, D. et al. IOBR: Multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. *Front. Immunol.* **12**, 687975 (2021).
26. Zhang, Z. et al. Integrated analysis of single-cell and bulk RNA sequencing data reveals a pan-cancer stemness signature predicting immunotherapy response. *Genome Med.* **14**(1), 45 (2022).
27. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**(10), 1550–1558 (2018).
28. Hoshida, Y. et al. Subclass mapping: Identifying common subtypes in independent disease data sets. *PLoS One* **2**(11), e1195 (2007).
29. Li, K. et al. EIF4G1 is a potential prognostic biomarker of breast Cancer. *Biomolecules* **12**(12), 1756 (2022).
30. Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: An R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* **22**(6), bbab260 (2021).
31. Ai, L., Xu, A. & Xu, J. Roles of PD-1/PD-L1 pathway: Signaling, Cancer, and Beyond. *Adv. Exp. Med. Biol.* **1248**, 33–59 (2020).
32. Sha, D. et al. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* **10**(12), 1808–1825 (2020).
33. Wu, B. et al. Multi-omics profiling and digital image analysis reveal the potential prognostic and immunotherapeutic properties of CD93 in stomach adenocarcinoma. *Front. Immunol.* **14**, 984816 (2023).
34. He, X. et al. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. *Sem. Cancer Biol.* **88**, 187–200 (2023).
35. Liu, Z. et al. Integrated multi-omics profiling yields a clinically relevant molecular classification for esophageal squamous cell carcinoma. *Cancer Cell.* **41**(1), 181–195e9 (2023).
36. Wang, Z. et al. Integrative multi-omics and drug-response characterization of patient-derived prostate cancer primary cells. *Signal. Transduct. Target. Therapy.* **8**(1), 175 (2023).
37. Gao, Z. J. et al. Integrative multi-omics analyses unravel the immunological implication and prognostic significance of CXCL12 in breast cancer. *Front. Immunol.* **14**, 1188351 (2023).
38. Sun, J. et al. Characterization of immune landscape in papillary thyroid cancer reveals distinct tumor immunogenicity and implications for immunotherapy. *Oncoimmunology* **10**(1), e1964189 (2021).
39. Cai, Z. et al. Machine learning for multi-omics data integration in cancer. *iScience* **25**(2), 103798 (2022).
40. Ma, Y. et al. Integrated multi-omics analysis and machine learning developed a prognostic model based on mitochondrial function in a large multicenter cohort for gastric Cancer. *J. Transl. Med.* **22**(1), 381 (2024).
41. Wang, J. et al. Integrated multi-omics analysis and machine learning identify hub genes and potential mechanisms of resistance to immunotherapy in gastric cancer. *Aging (Albany NY)*. **16**(8), 7331–7356 (2024).
42. Quiroga, A. D. et al. Hepatic carboxylesterase 3 (Ces3/Tgh) is downregulated in the early stages of liver cancer development in the rat. *Biochim. Biophys. Acta* **1862** (11), 2043–2053 (2016).
43. He, L. et al. A potential novel biomarker: Comprehensive analysis of prognostic value and immune implication of CES3 in colonic adenocarcinoma. *J. Cancer Res. Clin. Oncol.* **149**(14), 13239–13255 (2023).
44. Zhou, R. et al. Identifying an immunosenescence-associated gene signature in gastric cancer by integrating bulk and single-cell sequencing data. *Sci. Rep.* **14**(1), 17055 (2024).
45. Khan, M. et al. A novel necroptosis-related gene index for predicting prognosis and a cold tumor immune microenvironment in stomach adenocarcinoma. *Front. Immunol.* **13**, 968165 (2022).
46. Wang, Z. et al. Uncovering the potential of APOD as a biomarker in gastric cancer: A retrospective and multi-center study. *Comput. Struct. Biotechnol. J.* **23**, 1051–1064 (2024).
47. Chen, X. et al. Machine learning developed an intratumor heterogeneity signature for predicting prognosis and immunotherapy benefits in cholangiocarcinoma. *Translational Oncol.* **43**, 101905 (2024).
48. Zhu, X. J. et al. Novel tumor-suppressor gene epidermal growth factor-containing fibulin-like extracellular matrix protein 1 is epigenetically silenced and associated with invasion and metastasis in human gastric cancer. *Mol. Med. Rep.* **9**(6), 2283–2292 (2014).
49. Perrot-Applanat, M. et al. High expression of AhR and Environmental Pollution as AhR-Linked ligands Impact on Oncogenic Signaling pathways in Western patients with gastric Cancer-A pilot study. *Biomedicine* **12**(8), 1905 (2024).
50. Cheng, H. et al. LncRNA UCA1 enhances Cisplatin Resistance by regulating CYP1B1-mediated apoptosis via miR-513a-3p in human gastric Cancer. *Cancer Manag Res.* **13**, 367–377 (2021).
51. Hareendran, S. et al. Carboxypeptidase E and its splice variants: key regulators of growth and metastasis in multiple cancer types. *Cancer Lett.* **548**, 215882 (2022).
52. Lin, J. et al. CPE correlates with poor prognosis in gastric cancer by promoting tumorigenesis. *Heliyon* **10**(9), e29901 (2024).
53. Kang, H. et al. Downregulated CLIP3 induces radioresistance by enhancing stemness and glycolytic flux in glioblastoma. *J. Exp. Clin. Cancer Res.* **40**(1), 282 (2021).
54. Chien, T. M. et al. Role of microtubule-associated protein 1b in urothelial carcinoma: Overexpression predicts poor prognosis. *Cancers (Basel)* **12**(3), 630 (2020).

55. Jia, X. et al. miR-493 mediated DKK1 down-regulation confers proliferation, invasion and chemo-resistance in gastric cancer cells. *Oncotarget* **7**(6), 7044–7054 (2016).
56. Yin, L. et al. Identification of a five m6A-relevant mRNAs signature and risk score for the prognostication of gastric cancer. *J. Gastrointest. Oncol.* **13**(5), 2234–2248 (2022).
57. Sasaki, Y. et al. Expression of asporin reprograms cancer cells to acquire resistance to oxidative stress. *Cancer Sci.* **112**(3), 1251–1261 (2021).
58. Chen, S. et al. Multiomic analysis reveals comprehensive tumor heterogeneity and distinct immune subtypes in multifocal intrahepatic cholangiocarcinoma. *Clin. Cancer Res.* **28**(9), 1896–1910 (2022).
59. Chao, S. et al. Targeting intratumor heterogeneity suppresses colorectal cancer chemoresistance and metastasis. *EMBO Rep.* **24**(8), e56416 (2023).
60. Deng, Z. et al. A degradome-related signature for predicting the prognosis and immunotherapy benefit in stomach adenocarcinoma based on machine learning procedure. *Med. (Baltim).* **103**(15), e37728 (2024).
61. Huang, L. et al. Bioinformatics-based analysis of programmed cell death pathway and key prognostic genes in gastric cancer: implications for the development of therapeutics. *J. Gene Med.* **26** (1), e3590 (2024).
62. Lin, W. et al. Tumor-intrinsic YTHDF1 drives immune evasion and resistance to immune checkpoint inhibitors via promoting MHC-I degradation. *Nat. Commun.* **14** (1), 265 (2023).
63. Zhang, J. et al. Turning cold tumors hot: From molecular mechanisms to clinical applications. *Trends Immunol.* **43**(7), 523–545 (2022).
64. Chen, V. C. et al. Synergistic effects of the combinational use of escitalopram oxalate and 5-Fluorouracil on the inhibition of gastric Cancer SNU-1 cells. *Int. J. Mol. Sci.* **23**(24), 16179 (2022).
65. Li, H. et al. 5-Fluorouracil enhances the chemosensitivity of gastric cancer to TRAIL via inhibition of the MAPK pathway. *Biochem. Biophys. Res. Commun.* **540**, 108–115 (2021).
66. Sato, K. et al. Multicenter phase II study of capecitabine plus cisplatin as first-line therapy for human epidermal growth factor receptor 2-negative advanced gastric cancer: Yokohama Clinical Oncology Group Study YCOG1107. *Cancer Chemother. Pharmacol.* **80**(5), 939–943 (2017).
67. Nakata, S., Fujita, M. & Nakanishi, H. Efficacy of Afatinib and Lapatinib Against HER2 gene-amplified trastuzumab-sensitive and -resistant human gastric Cancer cells. *Anticancer Res.* **39**(11), 5927–5932 (2019).

Acknowledgements

Not applicable.

Author contributions

MW and QH wrote the manuscript and prepared the figures and tables, YJ revised the manuscript, and ZC conceptualized the study and oversaw the process. All authors reviewed the manuscript.

Funding

This research received no funding.

Declarations

Competing interests

The authors declare no competing interests.

Consent to publish

All authors agree to publication.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-87444-3>.

Correspondence and requests for materials should be addressed to Z.C.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025