Integrated multiomics analysis and machine learning refine molecular subtypes and prognosis for muscle-invasive urothelial cancer

Guangdi Chu,[1,3] Xiaoyu Ji,[2,3] Yonghua Wang,[1] and Haitao Niu[1]

[1]Department of Urology, The Affiliated Hospital of Qingdao University, Qingdao 266003, China; [2]Department of Gynecology Minimally Invasive Center, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing, China

Muscle-invasive urothelial cancer (MUC), characterized by high aggressiveness and significant heterogeneity, is currently lacking highly precise individualized treatment options. We used a computational pipeline to synthesize multiomics data from MUC patients using 10 clustering algorithms, which were then combined with 10 machine learning algorithms to identify molecular subgroups of high resolution and develop a robust consensus machine learning-driven signature (CMLS). Through multiomics clustering, we identified three cancer subtypes (CSs) that are related to prognosis, with CS2 exhibiting the most favorable prognostic outcome. Subsequent screening enabled identification of 12 hub genes that constitute a CMLS with robust predictive power for prognosis. The low-CMLS group exhibited a more favorable prognosis and greater responsiveness to immunotherapy and was more likely to exhibit the "hot tumor" phenotype. The high-CMLS group had a poor prognosis and lower likelihood of benefitting from immunotherapy, but dasatinib and romidepsin may serve as promising treatments for them. Comprehensive analysis of multiomics data can offer important insights and further refine the molecular classification of MUC. Identification of CMLS represents a valuable tool for early prediction of patient prognosis and for screening potential candidates likely to benefit from immunotherapy, with broad implications for clinical practice.

INTRODUCTION

Urothelial carcinoma is highly malignant and often leads to a poor prognosis for patients.[1] The most common urothelial carcinoma is bladder cancer, which is the second most common urological malignancy in the world, with over 430,000 people diagnosed and 170,000 deaths worldwide each year.[1,2] Bladder cancers mainly progress along two distinct pathways. Nonmuscle-invasive bladder cancer (NMIBC) accounts for 70% of the incidence of bladder cancer. It progress slowly but has a high probability of recurrence and requires lifelong monitoring. Muscle-invasive bladder cancer (MIBC) tends to progress rapidly and metastasize and has a higher mortality rate.[2–4]

With the improvement of current treatment technology, people are constantly trying to use advanced treatment methods for innovative treatment of urinary tumors.[5] Nonetheless, strategies aimed at treating advanced muscle-invasive urothelial cancer (MUC) have so far made limited progress. For instance, immunotherapy has been the focus of substantial research efforts and has demonstrated favorable outcomes in certain patients. However, a significant proportion of patients still fails to derive clinical benefits from this approach. This may be due to the marked heterogeneity of urothelial cancer, and molecular subtypes may help address this difficulty.[6,7] Given the substantial costs associated with immunotherapy and the potential for severe adverse reactions, there is a compelling need to leverage large-scale multiomics data with advanced machine learning algorithms to identify biomarkers capable of facilitating effective outcome prediction and immunotherapy management in patients with MUC.
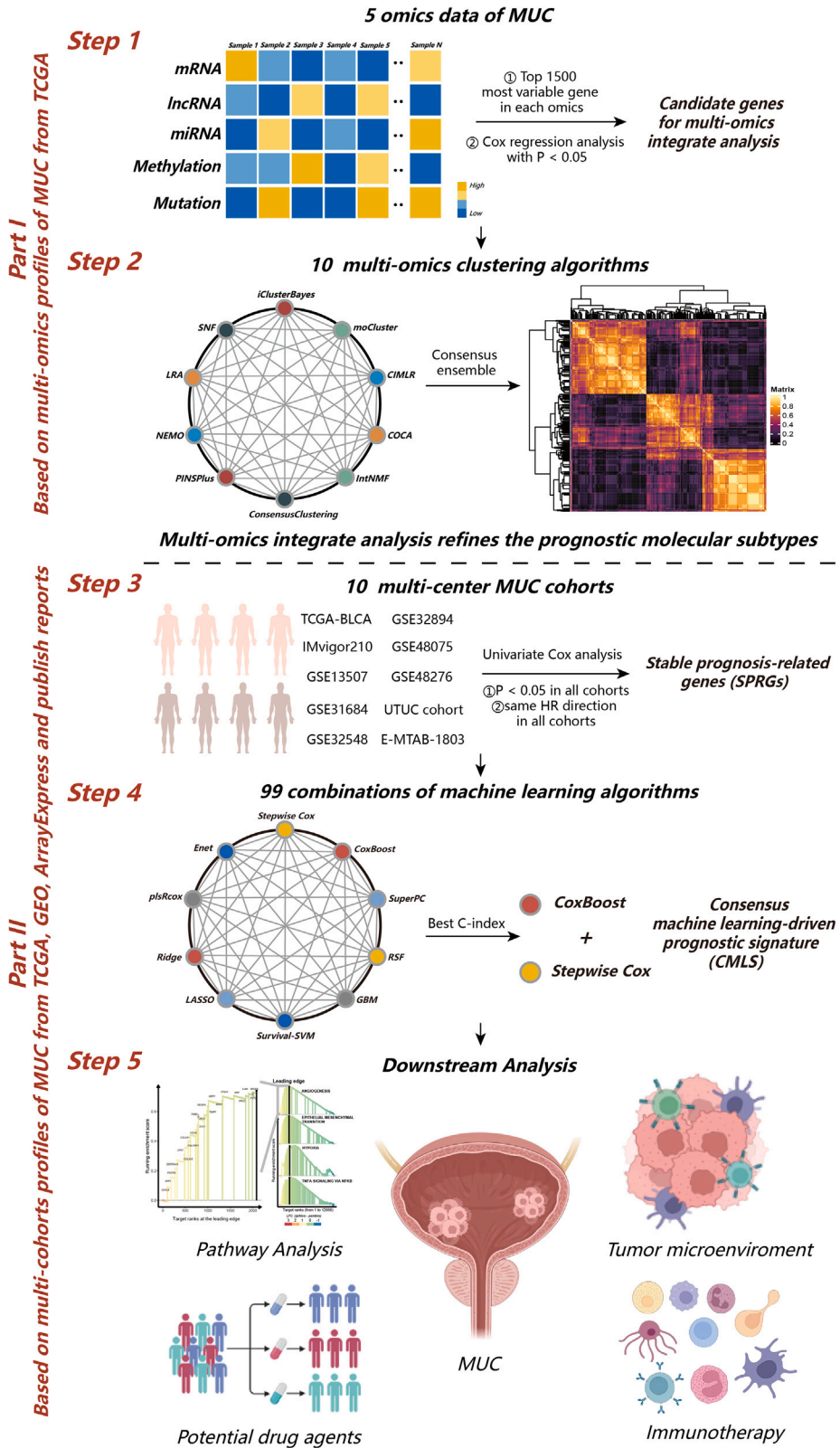
In this study, we combined mRNA, long non-coding RNA (lncRNA), and microRNA (miRNA) expression profiles, genomic mutations, and epigenomic DNA methylation data to develop an integrated consensus subtype of MUC using 10 multiomics integration strategies. Subsequently, we identified 32 stable prognosis-related genes (SPRGs) based on differential expression across subtypes and employed 10 machine learning algorithms to construct the consensus machine learning-driven signature (CMLS). In training and validation cohorts, the CMLS demonstrated significant prognostic value while exhibiting robust performance in predicting response to immunotherapy and drug-based therapies. Our findings represent a critical reference point for refining the molecular subtypes of MUC and enhancing precision stratification and individualized treatment approaches for this malignancy.

5 omics data of MUC

Step 1 — ① Top 1500 most variable gene in each omics → *Candidate genes for multi-omics integrate analysis*

② Cox regression analysis with P < 0.05

mRNA · lncRNA · miRNA · Methylation · Mutation

High / Low

10 multi-omics clustering algorithms

Step 2 — iClusterBayes, moCluster, SNF, CIMLR, LRA, COCA, NEMO, IntNMF, PINSPlus, ConsensusClustering

Consensus ensemble → Matrix 1 / 0.8 / 0.6 / 0.4 / 0.2 / 0

Multi-omics integrate analysis refines the prognostic molecular subtypes

Part I — Based on multi-omics profiles of MUC from TCGA

10 multi-center MUC cohorts

Step 3 — TCGA-BLCA, GSE32894, IMvigor210, GSE48075, GSE13507, GSE48276, GSE31684, UTUC cohort, GSE32548, E-MTAB-1803

Univariate Cox analysis
① P < 0.05 in all cohorts
② same HR direction in all cohorts

→ *Stable prognosis-related genes (SPRGs)*

99 combinations of machine learning algorithms

Step 4 — Stepwise Cox, CoxBoost, Enet, SuperPC, plsRcox, RSF, Ridge, GBM, LASSO, Survival-SVM

Best C-index → *CoxBoost* + *Stepwise Cox* → *Consensus machine learning-driven prognostic signature (CMLS)*

Step 5 — *Downstream Analysis*

Pathway Analysis

MUC

Tumor microenviroment

Potential drug agents

Immunotherapy

Part II — Based on multi-cohorts profiles of MUC from TCGA, GEO, ArrayExpress and publish reports

(legend on next page)

RESULTS

Multiomics consensus prognosis-related molecular subtypes of MUC

The workflow of this research is shown in Figure 1. Following effective preprocessing of all data, we further verified our results by employing a principal-component analysis (PCA) to analyze the data before and after processing the batch effect (Figures S1A and S1B). We independently identified three subtypes from 10 multiomics ensemble clustering algorithms, and the number of subtypes was determined by comprehensively referring to the cluster prediction index, gap statistical analysis, silhouette score, and previous research experience (Figures S2 and S3). Then, the clustering results were further combined through the consensus ensemble approach with distinctive molecular patterns of expression across transcriptomes (mRNA, lncRNA, and miRNA), epigenetic methylation, and somatic mutations (Figures 2A–2C). Our classification system was closely related to overall survival (OS) ($p < 0.001$; Figure 2D). Notably, cancer subtype 2 (CS2) exhibited the most favorable survival outcomes among all subtypes evaluated.

Partitioning of MUC integrative consensus molecular subtypes

At present, most of the molecular subtypes of MUC are classified based on molecular expression levels and may be related to specific biological functions. Therefore, we also tried to explore the different molecular features of these CSs. The single-sample gene set enrichment analysis (ssGSEA) algorithm measured the enrichment of different molecular signatures in the sample. Interestingly, we found that pathways such as luminal differentiation and urothelial differentiation were significantly enriched in CS2, while pathways such as basal differentiation and immune differentiation were significantly enriched in CS3, suggesting that CS2 may be more prone to the currently recognized luminal-like subtypes and that CS3 is more inclined to the basal-like subtypes. Moreover, there were also large differences in response to specific treatments among different subtypes, and pathways such as immune-inhibited oncogenic pathways were significantly enriched in CS2, while CS3 may be more likely to benefit from treatments such as radiotherapy or targeted therapy (Figure 3A).

To further research transcriptome differences, we analyzed potential regulators associated with cancer chromatin remodeling and 23 transcription factors (TFs) of MUC (Figure 3B). The close correlation of regulator activity with CSs confirmed the biological relevance of CSs. Androgen Receptor (AR), ERBB2, Fibroblast Growth Factor Receptor

3 (FGFR3), and FOXA1 regulators were significantly activated in CS1 and CS2, while EGFR, TP63, HIF1A, and STAT3 were specifically enriched in CS3. The regulon activity profiles associated with cancerous chromatin remodeling further highlight potential patterns of differential regulation between CSs, suggesting that epigenetically driven transcriptional networks may be important differentiating factors for these molecular subtypes. Given the pivotal role of tumor immunity in tumorigenesis and progression, we quantified the infiltration levels of microenvironmental cells and showed that immune cell infiltration was significantly increased in CS1 and CS3 but relatively lower in CS2 (Figure 3C). Based on the results of differential expression analysis between subtypes (Table S1), we selected 20 genes specifically upregulated for each subtype as classifiers and validated them in multiple external cohorts to further validate the stability of subtypes. The nearest template prediction (NTP) classifies each sample in the external cohort as one of the identified CSs. Consistent with this, CS2 in the meta-muscle invasive urothelial cancer cohort (META-MUC) cohort (combined with 8 cohorts) had the best prognosis of all subtypes ($p < 0.005$), and similar results were obtained in other external cohorts (Figures 3D and 3E). The consistency of CSs with NTP and partition around medoids (PAM) algorithms was also evaluated ($p < 0.005$; Figures 3F–3J).

Development of the CMLS

We performed univariate Cox regression analysis to screen 32 SPRGs whose expression was significantly related to OS from the consensus genes IMvigor 210 cohort (IMvigor)-MUC, The Cancer Genome Atlas (TCGA)-MUC, and META-MUC. Subsequently, SPRGs were included in the ensemble framework to perform CMLS. In the IMvigor-MUC training cohort, we built consistent models based on 99 algorithmic combinations and calculated the average C-index for each model in all cohorts to assess the predictive power of all models (Figure 4A). As shown in the 99 models, the algorithm consisting of CoxBoost and stepwise Cox (direction = forward) maintained the highest mean C-index to build the final model. The CoxBoost algorithm identified the most valuable SPRGs, and the stepwise Cox algorithm filtered the most valuable model, which was constructed by 12 hub genes (Figures 4B and 4C; Table S2). Then, we calculated the CMLS score per sample for all cohorts. High-CMLS patients had poor clinical outcomes in the TCGA, META, IMvigor, and combined cohorts (Figures 4D–4G).

For the hub genes of CMLS, we further verified the prognostic value of these hub genes in bladder cancer through the Biomarker Exploration

Figure 1. The computational framework for establishing the CMLS

The top 1,500 genes with the greatest degree of variation were screened from the omics data of each dimension, and then the genes associated with prognosis were selected as clustering candidate genes based on Cox regression analysis. For mutation data, we screened candidate genes based on mutation frequency. Stable MUC prognostic subtypes were obtained by consensus clustering using 10 multiomics clustering algorithms: iClusterBayes, moCluster, Cancer Integration via MultIkernel LeaRning (CIMLR), Integrative Clustering of Multiple Genomic Dataset (IntNMF), ConsensusClustering, Cluster-Of-Cluster-Assignments (COCA), NEighborhood based Multi-Omics clustering (NEMO), PINSPlus, Similarity network fusion (SNF), and Landscape Reconstruction Algorithm (LRA). To further identify stable prognostic genes according to subtypes, all combinations of machine learning algorithms (stepwise Cox, CoxBoost, ridge regression, RSF, GBM, Survival-SVM, LASSO, SuperPC, plsRcox, and Enet) were used to screen out the hub signature with the highest C-index. Finally, CMLS was constructed based on the combination of CoxBoost and stepwise Cox. The relationship between CMLS and prognosis, tumor immune microenvironment, immunotherapy response, and potential therapeutic agents was comprehensively determined.
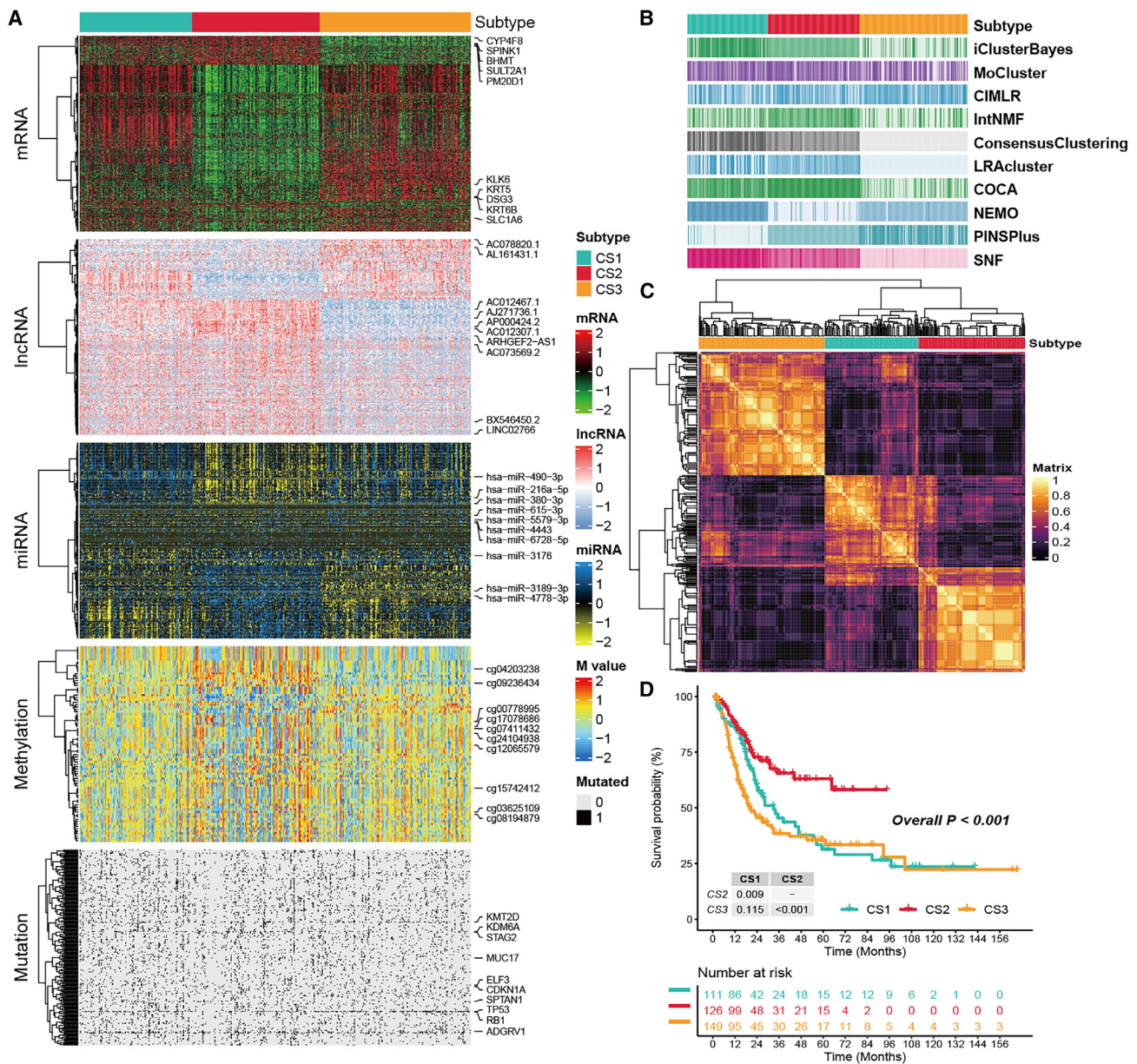
Figure 2. The multiomics integrative consensus subtypes of MUC

(A) Comprehensive heatmap of consensus ensemble subtypes, including mRNA, lncRNA, miRNA, DNA CpG methylation site, and mutant gene. (B) Clustering of MUC patients through 10 cutting-edge multiomics clustering methods. (C) Consensus clustering matrix for three novel prognostic subtypes based on the 10 algorithms. (D) Different survival outcomes among the three subtypes.

for Solid Tumors (BEST) database (https://rookieutopia.com/app_direct/BEST/) using a Kaplan-Meier analysis, and the obtained results were basically consistent with the results we calculated by the Cox algorithm (Figure S4). We also observed that these genes were significantly associated with progression-free survival (PFS) and disease-specific survival (DSS) of bladder cancer, underscoring their close prognostic relevance for patients (Figures S5 and S6). Subsequently, we utilized the GSCALite public server (http://bioinfo.life.

hust.edu.cn/web/GSCALite/) to systematically examine the multiomics phenotypes of CMLS across 33 distinct cancer types in TCGA. The results showed that, in cancer types with more than 10 tumor and normal samples, FJX1, FSCN1, EREG, BNC1, SIRPG, and PTHLH genes were found to be highly expressed in several cancer tissues (Figure S7A). We also found a positive correlation between mRNA expression levels and copy number variations (CNVs) of CMLS genes in most cancer types, especially for FJX1 (Figure S7B).
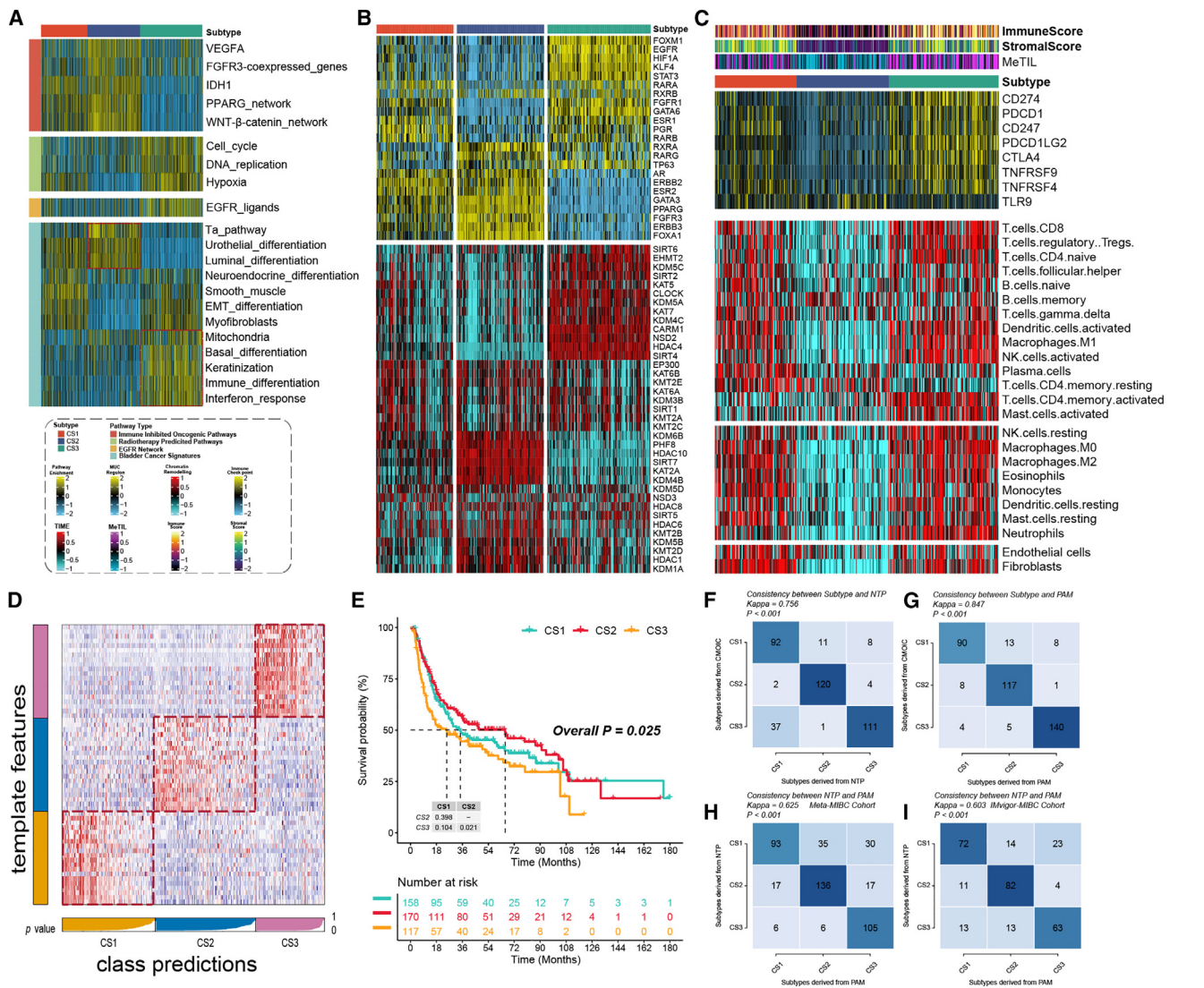
**Figure 3. Molecular landscape and validation of MUC CSs**

(A) The enrichment of three subtypes for different treatment-related signatures and bladder cancer-related signatures. (B) Regulon activity profiles for 23 TFs (top) and potential regulators associated with chromatin remodeling (bottom) of three subtypes. (C) Immune profiles in the TCGA-MUC cohort. The top annotation of the heatmap shows the immune enrichment score, stromal enrichment score, and DNA methylation of tumor-infiltrating lymphocytes. The top panel shows the expression of canonical immune checkpoint genes, and the bottom panel shows the enrichment levels of 24 TME-related immune cells. (D) Validation of MUC CSs in the nearest template of the META-MUC cohort. (E) Survival analysis of MUC CSs in the META-MUC cohort. (F) The consistency of CSs with NTP in the TCGA-MUC cohort. (G) The consistency of CSs with PAM in the TCGA-MUC cohort. (H) The consistency of NTP with PAM in the META-MUC cohort. (I) The consistency of NTP with PAM in the IMvigor-MUC cohort.

Through analysis of CNV frequency changes, the CNVs of CMLS genes showed significant differences in various cancer types, among which SIRPG and FSCN1 had the highest CNV frequencies, mainly copy number heterozygous amplification (Figures S7C and S7D). In addition, we found that the methylation levels of CMLS genes in most cancer specimens were significantly different between tumor and normal samples (Figure S8A). Methylation levels of CMLS genes were negatively correlated with mRNA expression levels of these genes in most cancers (Figure S8B). All of these indicated

CMLS genes may have an influence on the prognosis of patients through epigenetic changes. CMLS genes could activate the pan-cancer epithelial-mesenchymal transition (EMT) pathway and had a significant inhibitory effect on the hormone AR inhibition pathway (Figures S8C and S8D).

Comparison of prognostic signatures in MUC

With the advent of next-generation sequencing technologies, numerous gene expression-based prognostic signatures have been
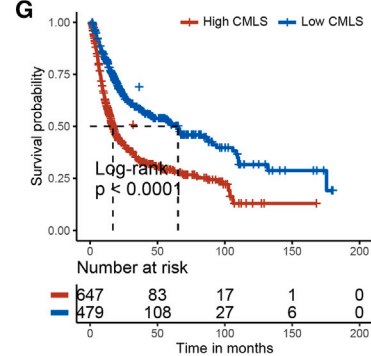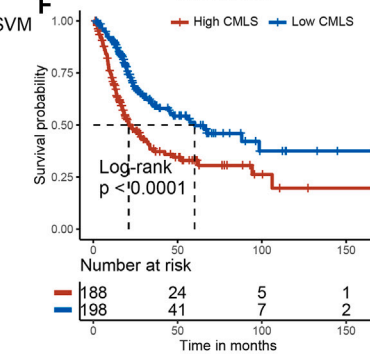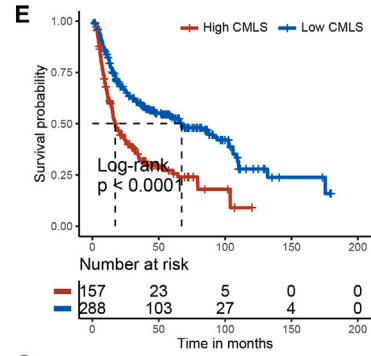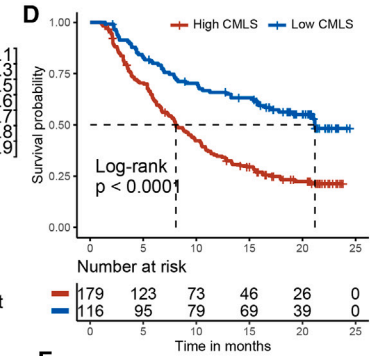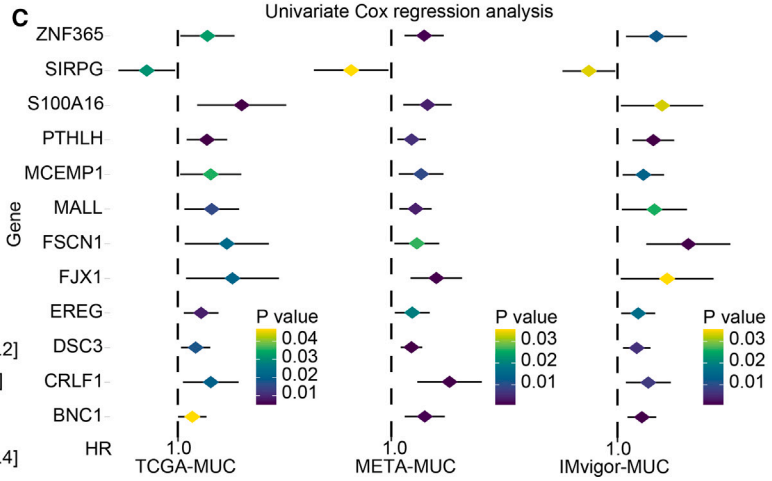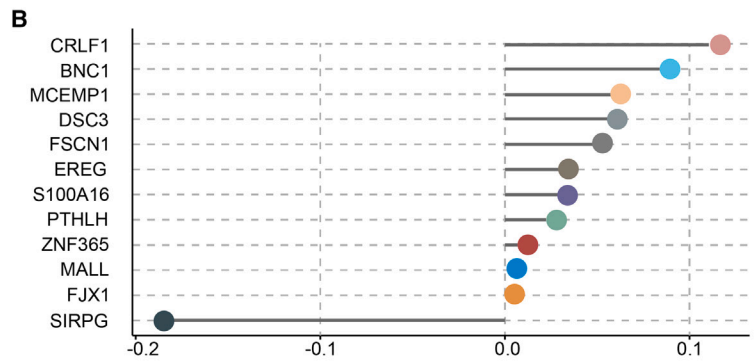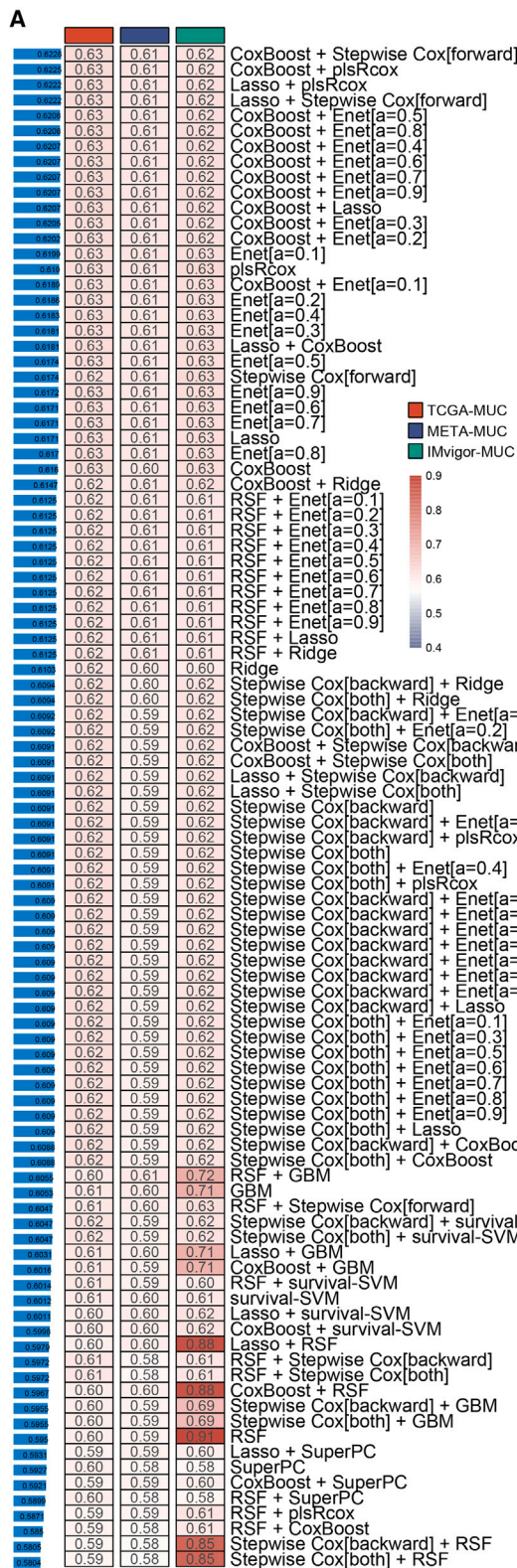
(legend on next page)

reported extensively in recent years. To enable a comprehensive comparison between CMLS and other signatures, we conducted a systematic search for relevant literature published within the past 5 years and ultimately incorporated 22 distinct signatures into our study (Table S3). These features are associated with different biological processes, such as immunotherapy response, immune infiltration, and glycolysis. Notably, CMLS showed better C-index performance than almost all models in the TCGA-MUC, IMvigor-MUC, and META-MUC datasets (Figures 5A–5C). Considering the clinical application prospects of CMLS, we screened potential independent prognostic factors for MUC by independent prognostic analysis (Figure S9) and integrated them to construct a comprehensive nomogram in the form of a web calculator (https://the-nomogram.shinyapps.io/CMLS-DynNomapp/; Figure 5D). The calibration curve proved that the nomogram has an accuracy that is consistent with the actual situation (Figure 5E). Decision Curve Analysis (DCA) indicated that the clinical benefit of the nomogram for patients was significantly higher than that of CMLS alone (Figures 5F and 5G), and the time-dependent C-index further proved that the nomogram had better prediction performance (Figure 5H).

Immune characteristics related to CMLS

Employing the Immuno-Oncology Biological Research (IOBR) R package, we conducted a comprehensive analysis of the tumor microenvironment (TME) of MUC and observed that immune cell infiltration levels, including T cells, B cells, and macrophages, were significantly higher in low-CMLS patients than in high-CMLS patients, indicative of an immune activation state (Figure 6A). These findings suggest that MUCs characterized by low CMLS levels are more likely to be classified as "hot tumors." Fibroblasts and neutrophils were mainly enriched in high-CMLS patients, and molecular markers related to immunosuppression and exclusion, such as the EMT pathway, were also mainly enriched in high groups, showing an immunosuppressive state (Figures 6B and 6C). This means that high-CMLS MUC was more inclined to be a "cold tumor." As we expected, previously reported signatures associated with better immunotherapy were also significantly enriched in the low-CMLS group (Figure 6D). Tumor mutational burden (TMB) and tumor neoantigen burden (TNB) are currently recognized biomarkers for evaluating the response of patients to immunotherapy, and Zeng et al.[8] also proposed a special role of M1 macrophages in bladder cancer immunotherapy. Therefore, we analyzed the differences in the content of these biomarkers between the two groups. The low-CMLS group had higher enrichment of TMB, TNB, and M1 macrophages, which means that the low group may have higher immunogenicity (Figure 6E–6H). Survival analysis also indicated that the CMLS could serve as an effective complementary factor for TMB, TNB, and M1

macrophages to differentiate patient outcomes (Figures 6I–6K). Lower CMLS with higher TMB or TNB or M1 macrophage infiltration tended to have a better survival prognosis for MUC patients.

The CMLS has excellent predictive power for immunotherapy response

To comprehensively assess the role of CMLS in MUC immunotherapy, we conducted a systematic analysis. Initially, we conducted a detailed analysis of the IMvigor-MUC cohort in view of the comprehensive prognostic and treatment-related information available for this patient population. Unlike many prior studies, we accounted for the delayed clinical effects of immunotherapy by comparing the restricted mean survival (RMS) between the two groups at 6 and 12 months while evaluating long-term survival differences among patients after 3 months of treatment (p < 0.05; Figures 7A and 7B). The lower group showed better prognostic outcomes, which indicates that the benefit of immunotherapy is greater. The distribution of CMLS among patients with different response degrees also showed that the CMLS score of the responder group (complete response [CR]/partial response [PR]) was significantly lower than that of the nonresponder group (progressive disease [PD]/stable disease [SD]) (p < 0.05; Figure 7C). Then, we calculated the tracking tumor immunophenotype (TIP) to explore the potential biological mechanisms associated with CMLS, and as we expected, the low-CMLS group showed significant differences mainly at step 4 (tumor immune-infiltrating cell recruitment), step 5 (immune cell infiltration), and step 7 (cancer cell killing), consistent with the results of our above analysis (Figure 7D). In addition, the tumor immune dysfunction and exclusion (TIDE) algorithm was used to assess patient response to immunotherapy and showed better responsiveness in the low CMLS group (P [Fisher's exact test] = 5.38e−06; Figure 7E). The subclass mapping algorithm was performed with another group of melanoma patients receiving immunotherapy, and the results also showed that low CMLS indicated a better response to PD-1 therapy (Bonferroni-corrected p = 0.008' Figure 7F). Finally, we revalidated our conclusions in multiple immunotherapy validation cohorts with prognostic information. Low CMLS tended to have better prognostic outcomes in the post-immunotherapy population (GSE78220, p = 0.015 [Figure 7G]; GSE135222, p = 0.026 [Figure 7H]), and low CMLS tended to be associated with better immunotherapy outcomes (GSE91061, p = 0.032; Figure 7I).

Screening of potential therapeutic drugs

There were significant differences in prognosis between high- and low-CMLS populations, and GSEA also showed that angiogenesis, EMT, hypoxia, and other pathways were significantly activated in high-CMLS patients (Figure 8A). In view of the poor response to

Figure 4. The generation and prognostic value of CMLS
(A) Through a comprehensive computational framework, a combination of 99 machine learning algorithms was generated. The C-index of each model was calculated through the TCGA-MUC, IMvigor-MUC, and META-MUC cohorts and sorted by the average C-index of the validation set. (B) The hub gene selected through the CoxBoost algorithm. (C) The univariate Cox regression analysis results of hub genes in training and validation cohorts. (D–G) Survival analysis of MUC patients with high CMLS and low CMLS in the IMvigor-MUC, META-MUC, TCGA-MUC, and combined-MUC cohorts.
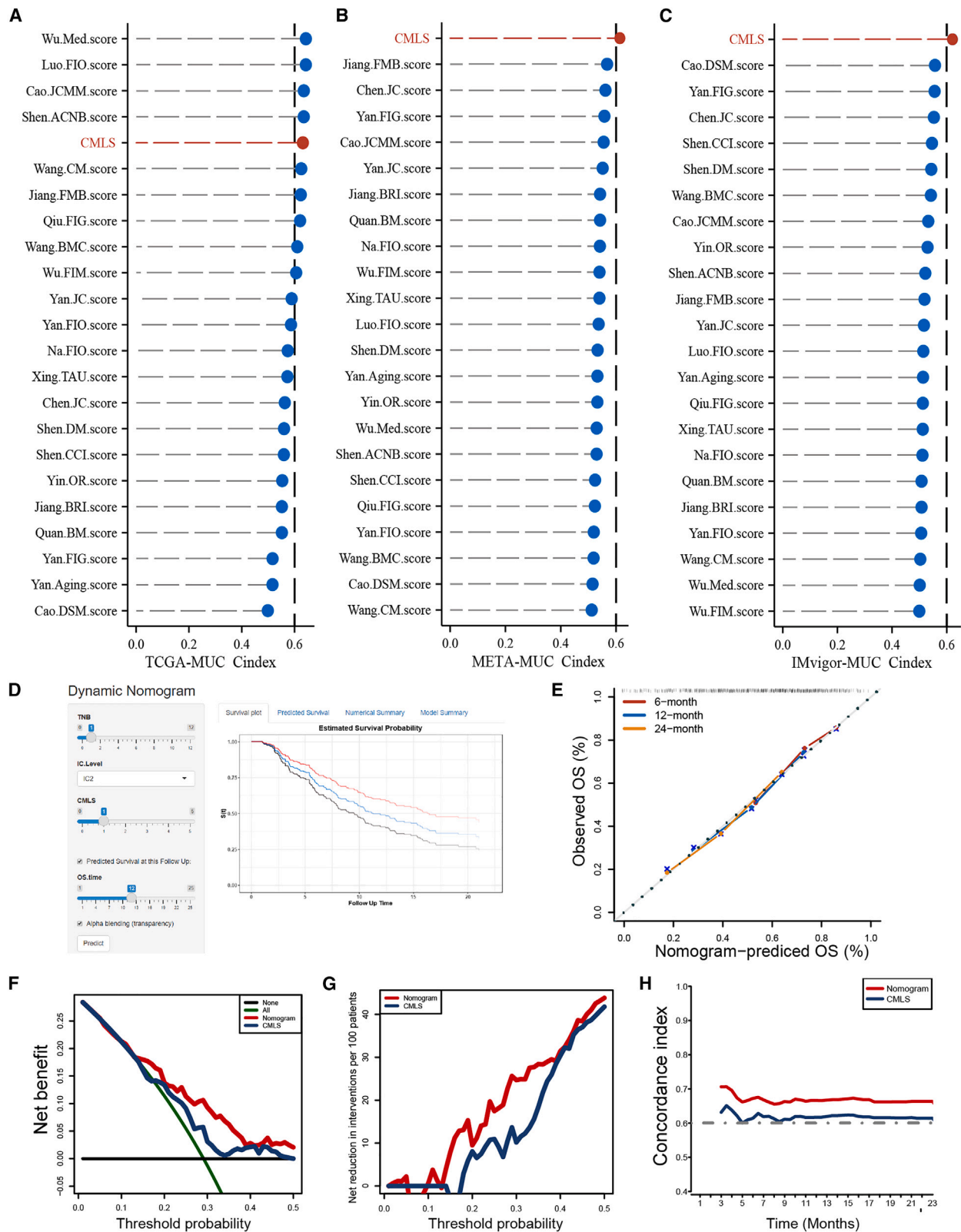
immunotherapy in patients with high CMLS, we used Cancer Therapeutics Response Portal (CTRP) and Profiling Relative Inhibition Simultaneously in Mixtures (PRISM) to screen potential therapeutic drugs for patients with high CMLS. To ensure the robustness of our methodology, we employed cisplatin, a widely utilized treatment for bladder cancer, as a means of validating whether algorithm-derived sensitivities were consistent with established clinical practices. A previous report indicated that ERCC1 is a prognostic biomarker in patients with advanced bladder cancer receiving cisplatin-based chemotherapy.[9] Our algorithm yielded similar findings, demonstrating that patients characterized by low ERCC1 expression levels exhibited a more robust response to cisplatin therapy, thereby conferring potential benefits for patient chemotherapy (Figure 8B). Then, we systematically explored potential drugs for high-CMLS patients according to the previous studies (Figure 8C). Finally, we screened one CTRP-derived agent (dasatinib; Figure 8D) and two PRISM-derived agents (romidepsin and ispinesib; Figure 8E). We then assessed the differences in the expression levels of target genes for drug candidate action in tumor tissues and normal tissues (including paired and unpaired analyses) (Figures 8F and 8G). A higher fold change indicates greater potential for drug candidate therapy (dasatinib: ABL1, FYN Proto-Oncogene, Src Family Tyrosine Kinase [FYN], KIT Proto-Oncogene, Receptor Tyrosine Kinase [KIT], STAT5B; romidepsin: HDAC1). Finally, we searched for evidence of candidate compounds in PubMed (https://www.ncbi.nlm.nih.gov/PubMed/). Overall, dasatinib and romidepsin are considered promising potential drugs for treatment of high-CMLS patients.

DISCUSSION

Gene expression is finely regulated by various genetic/epigenetic processes, such as methylation, mutation, and histone modifications.[10] Therefore, comprehensive analysis of multiomics data from patients will help us gain insights into disease-specific regulatory mechanisms.[10,11] However, most studies to date have focused primarily on individual -omics research.[12] The choice of clustering methods for omics is also largely based on personal preference, which further expands the limitations of a certain method with expansion of the scope of use. Our research tries to make up for this. We integrated the latest 10 clustering algorithms to identify three types of prognostic subtypes with different characteristics, which may have certain value for accurate stratified treatment of MUC patients. The novel subtypes were proven to be stable in multiple cohorts. In addition, we found that CS2 may have a certain similarity to the currently known luminal molecular subtypes, and our classification may further refine the traditional luminal/basal classification method.

Machine learning algorithms are currently an effective method for analyzing multiomics data.[13] To gain an understanding of the differences in molecular characteristics among different prognostic subtypes and improve clinical practical value, we included 10 MUC multicenter cohorts in this study and finally combined them into 1 training set and 2 validation sets and then selected the best CMLS through 99 algorithm combinations to overcome the limitations caused by algorithm selection. At present, when algorithms such as artificial intelligence are deeply integrated with a large amount of biological big data, overfitting is an important problem that cannot be ignored in the process of model construction.[14] It is not uncommon for models that demonstrate favorable performance in the training set to prove challenging to generalize effectively to other validation sets. To avoid the trouble caused by overfitting of the training cohort, we used the average of the C-index of multiple validation cohorts as the sorting criterion. As found in the study, when trained with random survival forest (RSF), it has excellent performance in the training set, but it is difficult to generalize in validation sets. Along similar lines, we observed that CMLS, which had been meticulously screened, exhibited robust prognostic value in each cohort compared with other published signatures. Conversely, while four other signatures demonstrated marginally superior performance than CMLS in the training set, their predictive performance markedly decreased across all validation sets, which was potentially related to overfitting of the model.

Through the GSEA algorithm and the IOBR R package, we analyzed the enrichment of dozens of immune-related signatures between the two groups. We found that various oncogenic pathways were significantly activated in the high-CMLS group, which was more prone to the cold tumor phenotype.[15] The low-CMLS group has higher TMB and TNB, is rich in immune cell types, and may have stronger antitumor immunity.[16] Survival analysis also shows better prognostic outcomes, and this is also validated in multiple immunotherapy cohorts in conclusion. Two widely used predictive tools, TIDE and subclass mapping, also demonstrated a better response to immunotherapy in the low-CMLS group, which is consistent with our analysis and indicates that CMLS may be useful for early identification of immunotherapy-sensitive populations.

In view of the current situation of poor prognosis response to immunotherapy in the high-CMLS group, we systematically screened its potential therapeutic drugs using a comprehensive screening framework that has been proven effective in previous studies,[17,18] and finally screened dasatinib and romidepsin as possible candidates for the high-CMLS group. Previous reports have found that the combination therapy of anti-PD-1 and dasatinib can lead to a reduction in tumor burden in tumor-bearing mice.[19] Dasatinib combined with a selective FGFR inhibitor could be an option to overcome the intrinsic drug resistance of urothelial cancer patients with FGFR3 alterations.[20] In addition, romidepsin has also been shown to effectively and safely control urothelial carcinoma by compensatory combination of gemcitabine and

Figure 5. Clinical practice value of CMLS
(A–C) Comparison between the CMLS and the other 22 published models in the TCGA-MUC, META-MUC, and IMvigor-MUC cohorts. (D) A comprehensive nomogram constructed based on CMLS and presented in the form of a web calculator for enhanced usability. (E) Calibration curve for the comprehensive nomogram. (F and G) Net decision curve analyses demonstrating the benefit for the comprehensive nomogram in clinical practice for MUC patients. (H) Comparison of the time-dependent C-index between the comprehensive nomogram and CMLS.
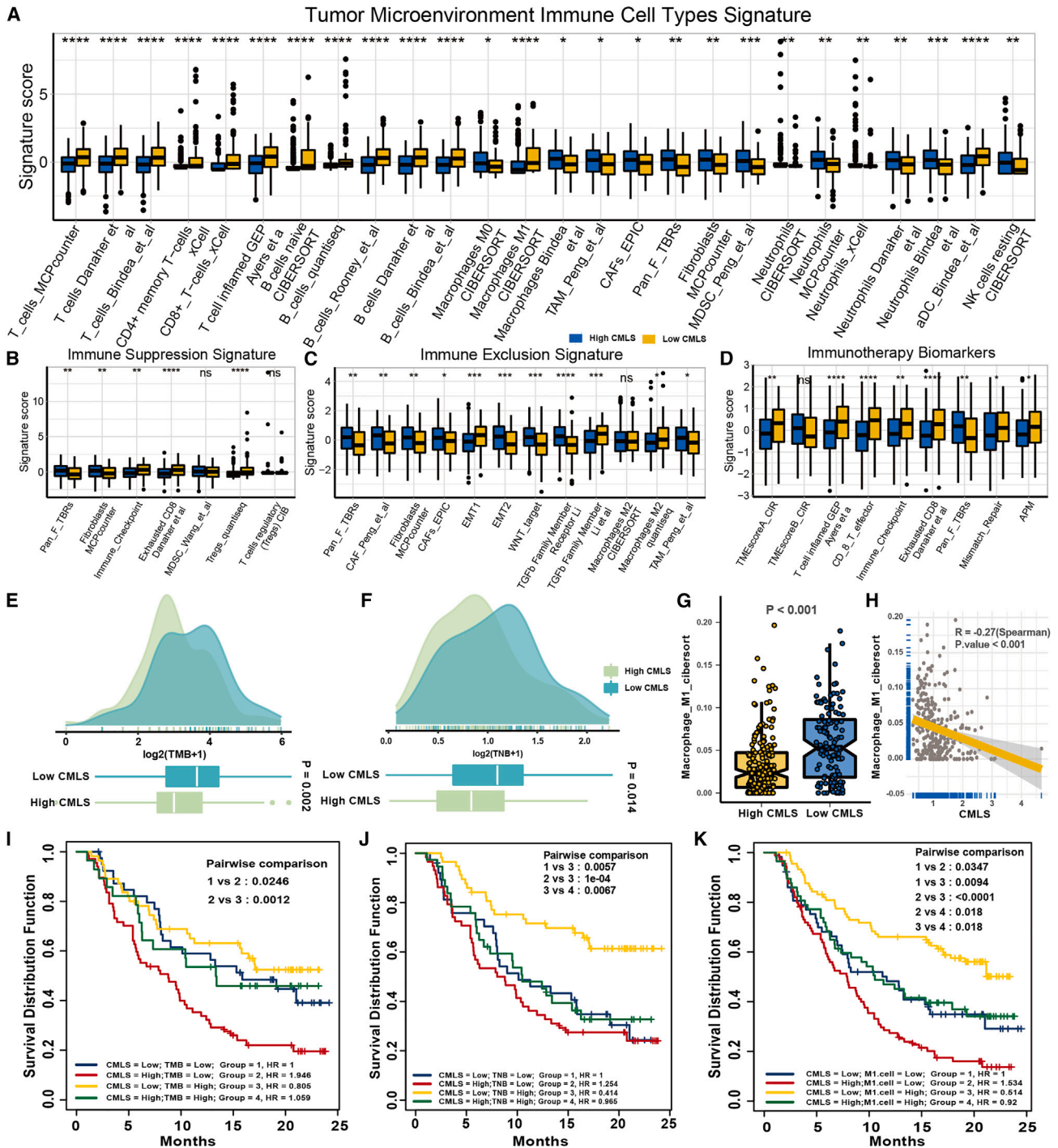
Figure 6. The TME-related molecular characteristics of high- and low-CMLS patients

(A) The distribution of TME immune cell type signatures between high- and low-CMLS patients. (B) The distribution of immune suppression signatures between high- and low-CMLS patients. (C) The distribution of immune exclusion signatures between high- and low-CMLS patients. (D) The distribution of immunotherapy biomarkers between high- and low-CMLS patients. (E) The distribution of TMB between high- and low-CMLS patients. (F) The distribution of TNB between high- and low-CMLS patients. (G) The distribution of M1 macrophages between high- and low-CMLS patients. (H) The relationship between CMLS and M1 macrophages. (I–K) Survival analysis combined CMLS with TMB, TNB, and M1 macrophages.
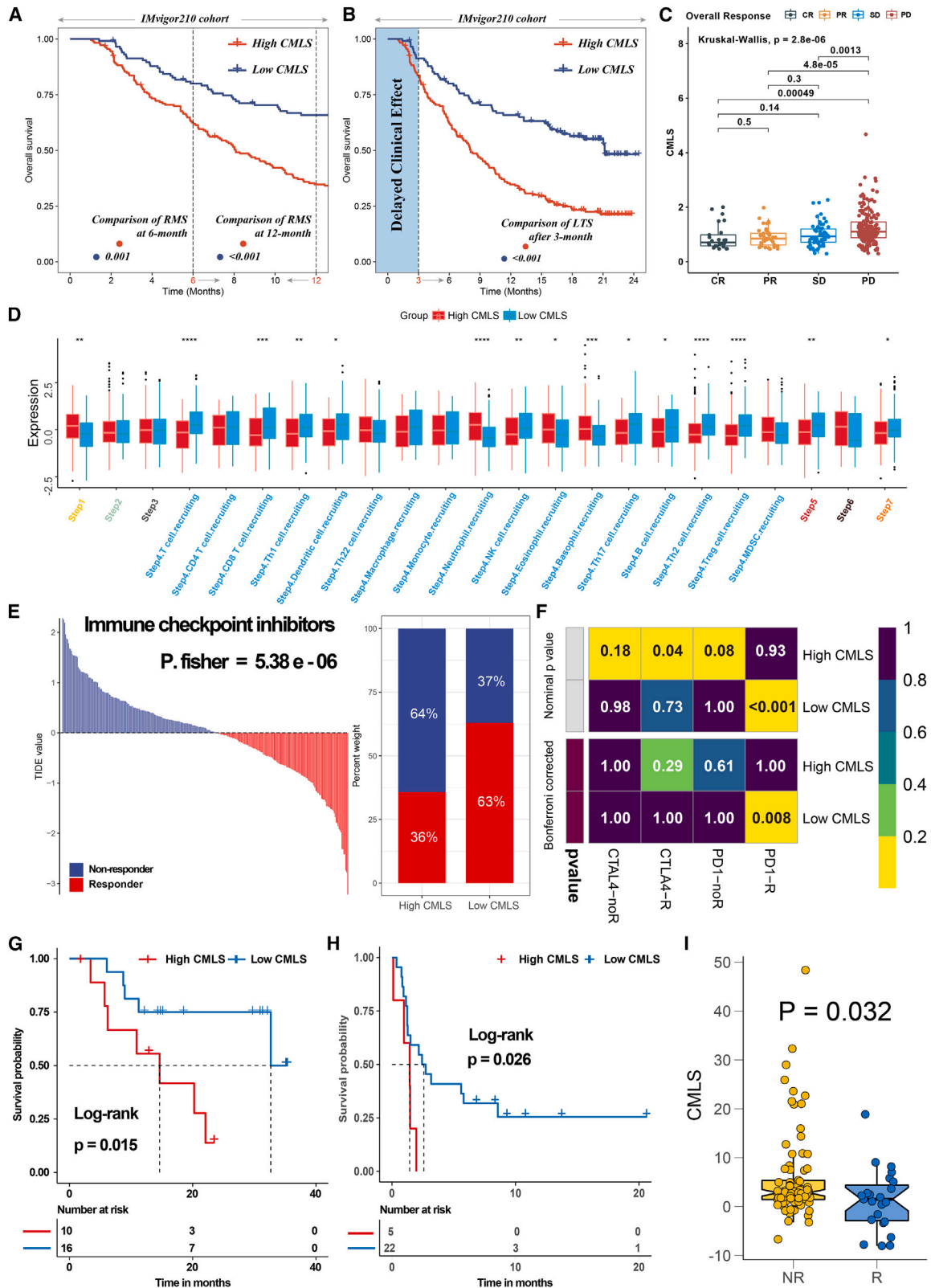
cisplatin,[21] and their combination with 2,3-dimethoxycinnamoyl azide can also enhance the resistance of romidepsin to bladder cancer *in vitro* and *in vivo*.[22]

Compared with earlier published research, our study features several notable distinctions. First, recognizing the marked heterogeneity of urothelial cancer, we conducted our analyses specifically in MUC, affording more precise patient stratification and treatment. Second, we incorporated omics information from all five dimensions of MUC and utilized a comprehensive set of 10 clustering algorithms to fully leverage the informative content of each omics dimension while mitigating the impact of clustering method selection preferences on our analyses. Third, our modeling factors were selected based on SPRGs derived from multiple cohorts, enhancing the stability and prognostic value of our model genes. Fourth, by systematically incorporating data from 10 MUC multicenter cohorts and leveraging 10 widely used machine learning algorithms, we identified the model with the best average C-index performance to establish CMLS, aiming to minimize the potential impact of overfitting on our findings. Nonetheless, we acknowledge that our study still has some limitations. For instance, the cohorts we included differed in size and sequencing platform despite utilizing correction algorithms to mitigate these differences. Additionally, the specific mechanisms underlying the tumorigenesis-associated activity of CMLS genes warrant further exploration. Furthermore, the clinical value of CMLS should be validated more extensively in larger, prospective multicenter cohorts.

Conclusions

This study identified three molecular subtypes of MUC via multiomics consensus clustering, revealing significant differences in prognosis among them and potentially refining the molecular typing of MUC. Leveraging a machine learning algorithm framework, we defined CMLS, which exhibited superior performance across multiple cohorts for robustly predicting patient prognosis while demonstrating close associations with immunotherapy response. Given the observed poor prognoses and low immunotherapy responses among high-CMLS groups, we further explored the potential therapeutic benefits of dasatinib and romidepsin for this population. Through integrating multiomics data and cutting-edge computational algorithms, this study provides a foundation for early diagnosis and precise treatment of MUC patients.

MATERIALS AND METHODS

### Multiomics data of MUC and data preprocessing of multicenter cohorts

We first collected multiomics data of MUC from the TCGA (https:// portal.gdc.cancer.gov) Bladder Cancer (BLCA) cohort, including patients with complete transcriptome expression, DNA methylation, somatic mutations, and clinical information available. The transcriptome profiles of mRNA and lncRNA were obtained by the TCGAbiolinks package.[23] The ID of the mature miRNA of TCGA was noted via the miRBaseVersions.db package. The somatic mutations were also obtained by TCGAbiolinks and were processed through the maftools package. The DNA methylation profile and clinical information were downloaded from UCSC xena (https:// xenabrowser.net/). In addition, we collected full information on MUC from nine other cohorts, including six from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; GEO: GSE13507,[24] GSE31684,[25] GSE32548,[26] GSE32894,[27] GSE48075,[28] and GSE48276[3]), one from ArrayExpress (https://www.ebi.ac.uk/ arrayexpress/; E-MTAB-1803[29]), one from the Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra, Su et al.[30]), and one clinical trial from http://research-pub.gene.com/IMvigor210CoreBiologies, which is available under the Creative Commons 3.0 license.[31] The high-throughput sequencing of the transcriptome was converted to transcripts per kilobase million (TPM), and all expression profiling by array was deduplicated and standardized. In this study, we defined samples with Ta- or T1-stage tumors as NMIBC that were initially removed. Patients with an OS time of less than 1 month were also excluded to enhance the robustness of downstream analyses.[32]

For the data from microarrays, the expression matrix and clinical information were downloaded from the official website. After downloading, the data were processed using the robust limma package for background correction, log 2 transformation, and quantile normalization and visualized using the boxplot function to assess the uniformity of sample expression abundance value distribution.[33] Similar methods have been applied in previous studies on glioma,[34] esophageal cancer,[35] liver cancer,[36] thyroid cancer,[37] and lung adenocarcinoma.[38] When multiple probes were mapped to a single gene symbol, the probe with the highest expression was annotated as gene expression. For RNA sequencing (RNA-seq) data from high-throughput sequencing, we used TPM values because they are more similar to gene expression from microarrays and enhance the comparability between samples.[39] For the cohort of Su et al.,[30] we downloaded the raw data from the BioProject (PRJNA678814), used fastqc to quality control the raw data, and used trim-galore for filtering. Then the data were compared through hisat2 with the gencode27 version of the genome. Next, we used featureCounts to quantify the bam files and filter the low-expressed genes according to the criteria of Su et al.[30] The clinical information was obtained from the supplementary files reported by Su et al.[30] For the merging of different datasets, the "ComBat" function in the sva package was used to adjust for batch effects from non-biological technical biases of each dataset

---

Figure 7. The value of CMLS in predicting immunotherapy response in MUC patients

(A) The restricted mean survival (RMS) time difference by 6 months and 1 year after treatment between high- and low-CMLS groups. (B) The long-term survival (LTS) difference after 3 months of treatment between high- and low-CMLS groups. (C) The distribution of CMLS in different immunotherapy response groups. (D) Differences in the degree of activation between high- and low-CMLS groups at each step of TIP. (E) The TIDE algorithm predicts response to immunotherapy between high- and low-CMLS groups. (F) The subclass mapping algorithm predicts response to immunotherapy between high- and low-CMLS groups. (G) Survival analysis of high- and low-CMLS groups in GSE78220. (H) Survival analysis of high- and low-CMLS group in GSE135222. (I) Distribution of CMLS in different immunotherapy response groups of GSE91061.
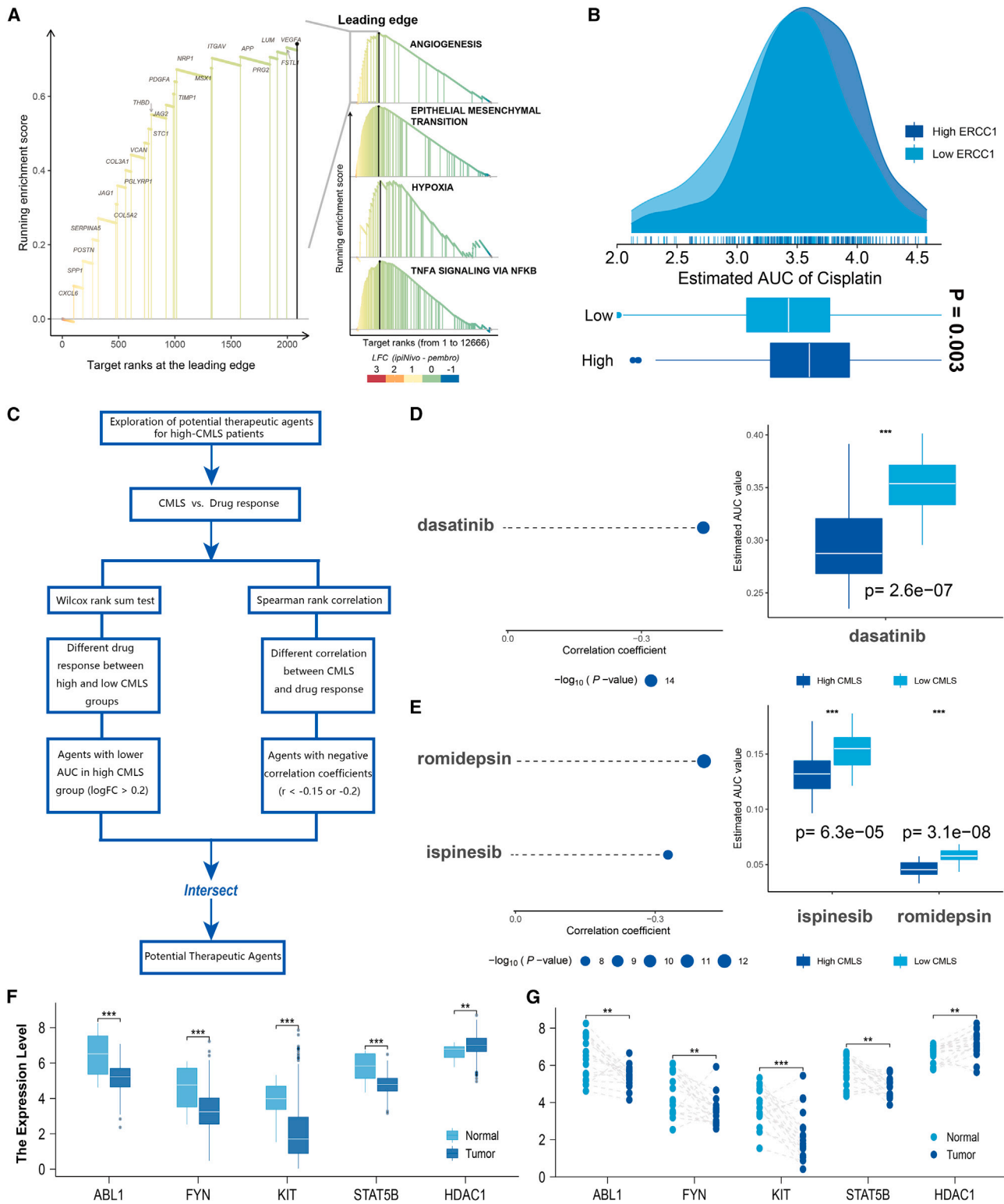
(legend on next page)

using an empirical Bayes framework.[40,41] Similar algorithms have been reported in many high-quality studies, ensuring the scientific validity of our methods.[13,42–46] To further validate the effectiveness of data merging, similar to the study conducted by Liu et al.,[47] we also performed PCA on the merged samples before and after merging.

Multiomics consensus ensemble analysis

To effectively perform a comprehensive analysis, we first matched the omics information of the five dimensions through the sample ID (n = 386). The TPM expression data were transformed by log2. For DNA methylation data, we selected probes of promoter CpG islands. For the gene mutation matrix, we identified the gene mutated when it contained any of the following nonsynonymous variations: frameshift insertion or deletion, in-frame insertion or deletion, nonsense or missense or nonstop mutation, or splice site or translation start site mutation.

In this study, we utilized the "getElites" function of the Multi-Omics integration and VIsualization in Cancer Subtyping (MOVICS) package to screen gene features. For continuous variables (mRNA, lncRNA, miRNA, and methylation), we set the "method" parameter of the "getElites" function to "mad" to screen for the top 1,500 genes with the greatest degree of variation. We then set the "method" parameter to "cox" and combined clinical data to identify prognostic genes with a significance level of $p < 0.05$ in each data dimension. For binary variable gene mutation data, we first used the "oncoPrint" function of the maftools package to screen for the top 5,000 genes with the highest degree of mutation. We then further screened based on mutation frequency by setting the "method" parameter to "freq" to identify the top 5% most frequently mutated genes. The resulting data from these five dimensions were included for further analysis in our study.

After preliminary feature selection, we further determined the optimal number of clusters for our study. As is well known, the optimal number of clusters for data should be small enough to reduce noise but large enough to retain important information. Therefore, we employed the "getClustNum" function from the MOVICS package, which integrates Clustering prediction index (CPI), Gaps-statistics, and Silhouette score to estimate the number of subgroups.[48,49] We also incorporated our prior knowledge of urothelial carcinoma from previous studies and ultimately decided to classify it into three subtypes. Subsequently, we applied the "getMOIC" function for clustering analysis. We used 10 clustering algorithms (CIMLR, ConsensusClustering, SNF, iClusterBayes, PINSPlus, moCluster, NEMO, IntNMF, COCA, and LRA) as input for the "methodslist" parameter and utilized the default parameters provided by the MOVICS package. As a result, we obtained the clustering results

for each method. After computing the clustering results for the 10 methods, we integrated the results from different algorithms using the "getConsensusMOIC" function based on the concept of consensus clustering to improve the robustness of the clustering.; he "distance" parameter was set to "euclidean," and the '"inkage" parameter was set to "average."[50,51] The final clustering result was obtained through this integration process.

Specific molecular characteristics and stability of consensus subtypes

Through gene set variation analysis (GSVA),[52] we calculated the enrichment scores of multitherapy-related signatures, including signatures positively associated with the response to atezolizumab in BLCA from Mariathasan et al.,[53] BLCA molecular subtype-related signatures from the Bladder Cancer Molecular Taxonomy Group, and signatures associated with targeted therapy and radiotherapy. These gene signatures were collected from the research of Hu et al.[54] The transcriptional regulatory networks (regulons) were constructed through the Reconstruction of Transcriptional regulatory Networks and analysis of regulons (RTN) R package, including 23 induced/repressed target-associated TFs and 71 candidate regulators related to cancerous chromatin remodeling collected from Lu et al.[45] Then, the distribution of immune checkpoints among these subgroups was compared, and the ESTIMATE R package was used to estimate the immune/stromal score of tumor tissue. The score of DNA methylation of tumor-infiltrating lymphocytes (MeTIL) was calculated according to published protocols. The enrichment of 24 kinds of tumor immune microenvironment cells was also evaluated through GSVA. For the stability of subtypes, we first validated the clustering results using subtype-specific biomarkers in the validation cohort and then compared the consistency of consensus clustering with the NTP and PAM classifier.[55]

Establishment of a consensus machine learning-driven prognostic signature

To enhance the comparability via different cohorts, we performed Z score processing on all data in advance. Then, to evaluate the relationship among CMLS, immunotherapy, and prognosis, we selected the IMvigor-MUC cohort with relatively complete treatment information as the training set and the other cohorts were used as validation sets. Considering that some cohorts had fewer sample accounts, we combined 8 cohorts into the META-MUC cohort and removed batch effects using the sva package. To construct CMLS with high accuracy and generalizability, we integrated 10 machine learning algorithms, including CoxBoost, stepwise Cox, Lasso, Ridge, elastic net (Enet), survival support vector machines (survival-SVMs), generalized boosted regression models (GBMs), supervised principal

Figure 8. Potential agents for patients with high CMLS

(A) Discovery of pathways significantly activated in the high-CMLS group through the GSEA algorithm. (B) Pre-predicting the sensitivity of cisplatin to validate the feasibility of the computational algorithm. (C) The comprehensive computational pipeline for the screen of potential agents. (D and E) The correlation and differential analysis of drug sensitivity for potential drugs screened from the CTRP and PRISM datasets. (F and G) The unpaired and paired differential expression analyses for potential target gene of screened drugs in normal and tumor tissue. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

components (SuperPC), partial least Cox (plsRcox) and RSF. Notably, algorithms such as stepwise Cox, RSF, CoxBoost, and Lasso possess feature selection capabilities. During the model-building phase, we used the IMvigor210 cohort as the training set for preliminary model construction.

For the CoxBoost model, we first called the "optimCoxBoostPenalty" function to determine the optimal penalty (shrinkage) value. We then combined this with cross-validation to perform 10-fold cross-validation on the CoxBoost model to search for the best number of boosting steps. Finally, the "CoxBoost" function was used to fit the model. The stepwise cox analysis was performed using the survival package, and the complexity of the statistical model was evaluated based on the Akaike information criterion (AIC). We calculated all possible combinations for the direction parameter, including "both," "backward," and "forward." Lasso, Ridge, and Enet models were implemented using the glmnet package and the "cv.glmnet" function. The regularization parameter lambda was determined through 10-fold cross-validation, while the tradeoff parameter alpha was set between 0 and 1 (interval = 0.1). When alpha is equal to 1, Lasso is executed, while Ridge is executed when alpha is equal to 0. For other values of alpha, Enet is executed. The survival-SVM model was implemented using the "survivalsvm" function from the survivalsvm package, which employs support vector analysis on datasets with survival outcomes. The GBM model was implemented using the gbm package. The "gbm" function was used with 10-fold cross-validation to fit a GBM. The SuperPC model was implemented using the superpc package, which is an extension of PCA. The "superpc.cv" function was also used with 10-fold cross-validation. For the plsRcox model, the "cv.plsRcox" function of the plsRcox package was used. For the RSF model, we utilized the randomForestSRC package and used the "rfsrc" function with two important parameters: "ntree" and "nodesize." The parameter "ntree" represents the number of trees in the random forest, while "nodesize" represents the minimum size of the terminal nodes. In this study, we set "ntree" to 1,000 and "nodesize" to 5.

The CMLS development pipeline is as follows.

- We conducted univariate Cox analysis in the IMvigor-MUC, TCGA-MUC, and META-MUC cohorts. Genes with p < 0.05 and the same hazard ratio (HR) orientation in all cohorts were considered SPRGs.
- Ten machine learning algorithms were used. Ninety-nine combinations of these algorithms were used to construct the most predictive CMLS with the best C-index performance.
- After model establishment on the training set, we further tested all validation cohorts. We calculated the average C-index for each model, where the model with the highest value was considered optimal.

Prognostic value of CMLS and potential clinical application

We scored each sample in the training and validation sets according to the resulting model and divided the samples into high- and low-CMLS groups based on the score. The prognostic significance of

CMLS was evaluated through the Kaplan-Meier survival curve. In addition, we systematically retrieved 22 prognostic features associated with urothelial cancer and calculated the score for each sample based on the published coefficients. The ability of all signatures to predict prognosis was assessed in each cohort by the C-index. To enhance the clinical utility value of CMLS, we constructed a nomogram using factors obtained after multivariate Cox regression. The time-dependent C-index curve and calibration curve were drawn to describe the accuracy, and the decision curve was used to calculate the clinical benefit to patients.

Comprehensive analysis of immune-omics molecular characterization and immunotherapy response based on CMLS

Based on the IOBR package,[56] we collected dozens of previously published signatures related to TME cell types, immunotherapy responses, immune suppression, and immune exclusion and used a unified method to calculate the enrichment score for each sample, which comprehensively analyzed the immunological differences between high- and low-CMLS patients. Differences in the distribution of TMB, TNB, and M1 macrophages between these two groups were compared, and patients were regrouped in combination with CMLS. For the immunotherapy response, we first evaluated the patients' delayed response survival to immunotherapy and combined the TIP algorithm, the subclass mapping, and TIDE algorithm to estimate the immunotherapy response.[57–59] This was further verified in GSE78220,[60] GSE135222,[61] and GSE91061.[61]

In silico analysis to screen potential therapy agents for patients with high CMLS

The activation status of oncogenic pathways between high- and low-CMLS patients was analyzed through the GSEA algorithm.[62] Expression data for human cancer cell lines (CCLs) were obtained from the Broad Institute Cancer Cell Line Encyclopedia (CCLE). The CTRP v.2.0 (https://portals.broadinstitute.org/ctrp) and PRISM Repurposing datasets (19Q4; https://depmap.org/portal/prism/) were used to obtain drug sensitivity data for CCLs. The area under the dose-response curve (AUC) values acted as a measure of drug sensitivity.

Statistical analysis

For comparisons of two groups, the normally distributed variables were tested by unpaired Student's t tests, and the non-normally distributed variables were tested by the Wilcoxon rank-sum test. For comparisons of more than two groups, the parametric and nonparametric variables were tested by one-way ANOVA and Kruskal-Wallis tests, respectively. A two-sided Fisher's exact test was performed for the contingency tables. The cutoff value of the CMLS score was determined through the "surv-cutpoint" function of the survminer package. We repeatedly tested all possible cutoff points to identify the maximum rank statistics and then used the two-classification method to classify the CMLS score. According to the maximum selected logarithmic rank statistics, patients could be distinguished into high- and low-score groups in each cohort to reduce the computational batch effect, which is similar to previous research.[63–65] The differential expression analysis was analyzed by the limma package,

and multiomics clustering was completed through the MOVICS package.[55] All statistical analyses were performed in R v.4.1.0.

DATA AVAILABILITY

All data used in this work can be acquired from TCGA (https://portal.gdc.cancer.gov), GEO (http://www.ncbi.nlm.nih.gov/geo), ArrayExpress (https://www.ebi.ac.uk/arrayexpress/), and SRA (https://www.ncbi.nlm.nih.gov/sra). The key code and parameters of this study were uploaded to GitHub (https://github.com/BioUro/MTNA-MUC).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2023.06.001.

AUTHOR CONTRIBUTIONS

G.C. and X.J. designed the whole study, analyzed the data, and drafted the manuscript. H.N. and Y.W. revised the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Patel, V.G., Oh, W.K., and Galsky, M.D. (2020). Treatment of muscle-invasive and advanced bladder cancer in 2020. CA A Cancer J. Clin. 70, 404–423.

2. Compérat, E., Amin, M.B., Cathomas, R., Choudhury, A., De Santis, M., Kamat, A., Stenzl, A., Thoeny, H.C., and Witjes, J.A. (2022). Current best practice for bladder cancer: a narrative review of diagnostics and treatments. Lancet.

3. Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E.R., Hoffman-Censits, J., Roth, B., Cheng, T., Tran, M., Lee, I.L., et al. (2014). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer Cell 25, 152–165.

4. Tran, L., Xiao, J.-F., Agarwal, N., Duex, J.E., and Theodorescu, D. (2021). Advances in bladder cancer biology and therapy. Nat. Rev. Cancer 21, 104–121.

5. Li, J., Yang, X., Chu, G., Feng, W., Ding, X., Yin, X., Zhang, L., Lv, W., Ma, L., Sun, L., et al. (2022). Application of Improved Robot-Assisted Laparoscopic Telesurgery with 5G Technology in Urology (European Urology).

6. Robertson, A.G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A.D., Hinoue, T., Laird, P.W., Hoadley, K.A., Akbani, R., et al. (2017). Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell 171, 540–556.

7. Sjödahl, G., Abrahamsson, J., Holmsten, K., Bernardo, C., Chebil, G., Eriksson, P., Johansson, I., Kollberg, P., Lindh, C., Lövgren, K., et al. (2022). Different responses to neoadjuvant chemotherapy in urothelial carcinoma molecular subtypes. Eur. Urol. 81, 523–532.

8. Zeng, D., Ye, Z., Wu, J., Zhou, R., Fan, X., Wang, G., Huang, Y., Wu, J., Sun, H., Wang, M., et al. (2020). Macrophage correlates with immunophenotype and predicts anti-PD-L1 response of urothelial cancer. Theranostics 10, 7002–7014.

9. Bellmunt, J., Paz-Ares, L., Cuello, M., Cecere, F.L., Albiol, S., Guillem, V., Gallardo, E., Carles, J., Mendez, P., de la Cruz, J.J., et al. (2007). Gene expression of ERCC1 as a novel prognostic marker in advanced bladder cancer patients receiving cisplatin-based chemotherapy. Ann. Oncol. 18, 522–528.

10. Oh, M., Park, S., Kim, S., and Chae, H. (2021). Machine learning-based analysis of multi-omics data on the cloud for investigating gene regulations. Briefings Bioinf. 22, 66–76.

11. Reel, P.S., Reel, S., Pearson, E., Trucco, E., and Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: a review. Biotechnol. Adv. 49, 107739.

12. Ma, C., Wu, M., and Ma, S. (2022). Analysis of cancer omics data: a selective review of statistical techniques. Briefings Bioinf. 23, bbab585.

13. Liu, Z., Liu, L., Weng, S., Guo, C., Dang, Q., Xu, H., Wang, L., Lu, T., Zhang, Y., Sun, Z., and Han, X. (2022). Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. Nat. Commun. 13, 816.

14. Wang, L., Liu, Z., Liang, R., Wang, W., Zhu, R., Li, J., Xing, Z., Weng, S., Han, X., and Sun, Y.-L. (2022). Comprehensive machine-learning survival framework develops a consensus model in large-scale multicenter cohorts for pancreatic cancer. Elife 11, e80150.

15. Galon, J., and Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat. Rev. Drug Discov. 18, 197–218.

16. Borst, J., Ahrends, T., Bąbała, N., Melief, C.J.M., and Kastenmüller, W. (2018). CD4 T cell help in cancer immunology and immunotherapy. Nat. Rev. Immunol. 18, 635–647.

17. Chu, G., Shan, W., Ji, X., Wang, Y., and Niu, H. (2021). Multi-omics analysis of novel signature for immunotherapy response and tumor microenvironment regulation patterns in urothelial cancer. Front. Cell Dev. Biol. 9, 764125.

18. Yang, C., Chen, J., Li, Y., Huang, X., Liu, Z., Wang, J., Jiang, H., Qin, W., Lv, Y., Wang, H., and Wang, C. (2021). Prognosis and personalized treatment prediction in TP53-mutant hepatocellular carcinoma: an in silico strategy towards precision oncology. Briefings Bioinf. 22, bbaa295.

19. Tu, M.M., Lee, F.Y.F., Jones, R.T., Kimball, A.K., Saravia, E., Graziano, R.F., Coleman, B., Menard, K., Yan, J., Michaud, E., et al. (2019). Targeting DDR2 enhances tumor response to anti-PD-1 immunotherapy. Sci. Adv. 5, eaav2437.

20. Lima, N.C., Atkinson, E., Bunney, T.D., Katan, M., and Huang, P.H. (2020). Targeting the src pathway enhances the efficacy of selective FGFR inhibitors in urothelial cancers with FGFR3 alterations. Int. J. Mol. Sci. 21, 3214.

21. Pattarawat, P., Hong, T., Wallace, S., Hu, Y., Donnell, R., Wang, T.-H., Tsai, C.-L., Wang, J., and Wang, H.-C.R. (2020). Compensatory combination of romidepsin with gemcitabine and cisplatin to effectively and safely control urothelial carcinoma. Br. J. Cancer 123, 226–239.

22. Fan, J., Stanfield, J., Guo, Y., Karam, J.A., Frenkel, E., Sun, X., and Hsieh, J.-T. (2008). Effect of trans-2,3-dimethoxycinnamoyl azide on enhancing antitumor activity of romidepsin on human bladder cancer. Clin. Cancer Res. 14, 1200–1207.

23. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 44, e71.

24. Kim, W.-J., Kim, E.-J., Kim, S.-K., Kim, Y.-J., Ha, Y.-S., Jeong, P., Kim, M.-J., Yun, S.-J., Lee, K.M., Moon, S.-K., et al. (2010). Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer. Mol. Cancer 9, 3.

25. Riester, M., Taylor, J.M., Feifer, A., Koppie, T., Rosenberg, J.E., Downey, R.J., Bochner, B.H., and Michor, F. (2012). Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. Clin. Cancer Res. 18, 1323–1333.

26. Lindgren, D., Sjödahl, G., Lauss, M., Staaf, J., Chebil, G., Lövgren, K., Gudjonsson, S., Liedberg, F., Patschan, O., Månsson, W., et al. (2012). Integrated genomic and gene expression profiling identifies two major genomic circuits in urothelial carcinoma. PLoS One 7, e38863.

27. Sjödahl, G., Lauss, M., Lövgren, K., Chebil, G., Gudjonsson, S., Veerla, S., Patschan, O., Aine, M., Fernö, M., Ringnér, M., et al. (2012). A molecular taxonomy for urothelial carcinoma. Clin. Cancer Res. 18, 3377–3386.

28. Guo, C.C., Bondaruk, J., Yao, H., Wang, Z., Zhang, L., Lee, S., Lee, J.-G., Cogdell, D., Zhang, M., Yang, G., et al. (2020). Assessment of luminal and basal phenotypes in bladder cancer. Sci. Rep. 10, 9743.

29. Rebouissou, S., Bernard-Pierrot, I., de Reyniès, A., Lepage, M.-L., Krucker, C., Chapeaublanc, E., Hérault, A., Kamoun, A., Caillault, A., Letouzé, E., et al. (2014). EGFR as a potential therapeutic target for a subset of muscle-invasive bladder cancers presenting a basal-like phenotype. Sci. Transl. Med. 6, 244ra291.

30. Su, X., Lu, X., Bazai, S.K., Compérat, E., Mouawad, R., Yao, H., Rouprêt, M., Spano, J.-P., Khayat, D., Davidson, I., et al. (2021). Comprehensive integrative profiling of upper tract urothelial carcinomas. Genome Biol. 22, 7.

31. Balar, A.V., Galsky, M.D., Rosenberg, J.E., Powles, T., Petrylak, D.P., Bellmunt, J., Loriot, Y., Necchi, A., Hoffman-Censits, J., Perez-Gracia, J.L., et al. (2017). Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet 389, 67–76.

32. Lu, X., Meng, J., Zhu, J., Zhou, Y., Jiang, L., Wang, Y., Wen, W., Liang, C., and Yan, F. (2021). Prognosis stratification and personalized treatment in bladder cancer through a robust immune gene pair-based signature. Clin. Transl. Med. 11, e453.

33. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47.

34. Cong, P., Wu, T., Huang, X., Liang, H., Gao, X., Tian, L., Li, W., Chen, A., Wan, H., He, M., et al. (2021). Identification of the role and clinical prognostic value of target genes of m6A RNA methylation regulators in glioma. Front. Cell Dev. Biol. 9, 709022.

35. Chen, H., Shi, X., Ren, L., Zhuo, H., Zeng, L., Qin, Q., Wan, Y., Sangdan, W., and Zhou, L. (2022). Identification of the miRNA-mRNA regulatory network associated with radiosensitivity in esophageal cancer based on integrative analysis of the TCGA and GEO data. BMC Med. Genom. 15, 249.

36. Wang, Z., Pan, L., Guo, D., Luo, X., Tang, J., Yang, W., Zhang, Y., Luo, A., Gu, Y., and Pan, Y. (2021). A novel five-gene signature predicts overall survival of patients with hepatocellular carcinoma. Cancer Med. 10, 3808–3821.

37. Wang, X., Peng, W., Li, C., Qin, R., Zhong, Z., and Sun, C. (2021). Identification of an immune-related signature indicating the dedifferentiation of thyroid cells. Cancer Cell Int. 21, 231.

38. Liu, C., Li, X., Shao, H., and Li, D. (2020). Identification and validation of two lung adenocarcinoma-development characteristic gene sets for diagnosing lung adenocarcinoma and predicting prognosis. Front. Genet. 11, 565206.

39. Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theor. Biosci. 131, 281–285.

40. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882–883.

41. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127.

42. Zhang, X., Shi, M., Chen, T., and Zhang, B. (2020). Characterization of the immune cell infiltration landscape in head and neck squamous cell carcinoma to aid immunotherapy. Mol. Ther. Nucleic Acids 22, 298–309.

43. Cao, R., Ma, B., Wang, G., Xiong, Y., Tian, Y., and Yuan, L. (2021). Characterization of hypoxia response patterns identified prognosis and immunotherapy response in bladder cancer. Mol. Ther. Oncolytics 22, 277–293.

44. Huang, M., Liu, L., Zhu, J., Jin, T., Chen, Y., Xu, L., Cheng, W., Ruan, X., Su, L., Meng, J., et al. (2021). Identification of immune-related subtypes and characterization of tumor microenvironment infiltration in bladder cancer. Front. Cell Dev. Biol. 9, 723817.

45. Lu, X., Meng, J., Su, L., Jiang, L., Wang, H., Zhu, J., Huang, M., Cheng, W., Xu, L., Ruan, X., et al. (2021). Multi-omics consensus ensemble refines the classification of muscle-invasive bladder cancer with stratified prognosis, tumour microenvironment and distinct sensitivity to frontline therapies. Clin. Transl. Med. 11, e601.

46. Zhang, X., Zhang, Y., Zhao, L., Wang, J., Li, J., Wang, X., Zhang, M., and Hu, X. (2022). Exploitation of tumor antigens and construction of immune subtype classifier for mRNA vaccine development in bladder cancer. Front. Immunol. 13, 1014638.

47. Liu, Z., Guo, C., Dang, Q., Wang, L., Liu, L., Weng, S., Xu, H., Lu, T., Sun, Z., and Han, X. (2022). Integrative analysis from multi-center studies identities a consensus machine learning-derived lncRNA signature for stage II/III colorectal cancer. EBioMedicine 75, 103750.

48. Chalise, P., and Fridley, B.L. (2017). Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. PLoS One 12, e0176278.

49. Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. Roy. Stat. Soc. B 63, 411–423.

50. Strehl, A., and Ghosh, J. (2002). Cluster ensembles - a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. 3, 583–617.

51. Lu, X., Meng, J., Zhou, Y., Jiang, L., and Yan, F. (2021). MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. Bioinformatics 36, 5539–5541.

52. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. 14, 7.

53. Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 554, 544–548.

54. Hu, J., Yu, A., Othmane, B., Qiu, D., Li, H., Li, C., Liu, P., Ren, W., Chen, M., Gong, G., et al. (2021). Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. Theranostics 11, 3089–3108.

55. Lu, X., Meng, J., Zhou, Y., Jiang, L., and Yan, F. (2020). MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. Bioinformatics 36, 5539–5541.

56. Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., Zhou, R., Qiu, W., Huang, N., Sun, L., et al. (2021). IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. Front. Immunol. 12, 687975.

57. Xu, L., Deng, C., Pang, B., Zhang, X., Liu, W., Liao, G., Yuan, H., Cheng, P., Li, F., Long, Z., et al. (2018). TIP: a web server for resolving tumor immunophenotype profiling. Cancer Res. 78, 6575–6580.

58. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat. Med. 24, 1550–1558.

59. Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2007). Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One 2, e1195.

60. Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell 165, 35–44.

61. Jung, H., Kim, H.S., Kim, J.Y., Sun, J.-M., Ahn, J.S., Ahn, M.-J., Park, K., Esteller, M., Lee, S.-H., and Choi, J.K. (2019). DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nat. Commun. 10, 4278.

62. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550.

63. Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., Bin, J., Liao, Y., Rao, J., and Liao, W. (2019). Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. Cancer Immunol. Res. 7, 737–750.

64. Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y.L. (2020). m6A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Mol. Cancer 19, 53.

65. Chong, W., Shang, L., Liu, J., Fang, Z., Du, F., Wu, H., Liu, Y., Wang, Z., Chen, Y., Jia, S., et al. (2021). m6A regulator-based methylation modification patterns characterized by distinct tumor microenvironment immune profiles in colon cancer. Theranostics 11, 2201–2217.

Supplemental information

Integrated multiomics analysis and

machine learning refine molecular subtypes and

prognosis for muscle-invasive urothelial cancer

Guangdi Chu, Xiaoyu Ji, Yonghua Wang, and Haitao Niu

Supplemental Figures：
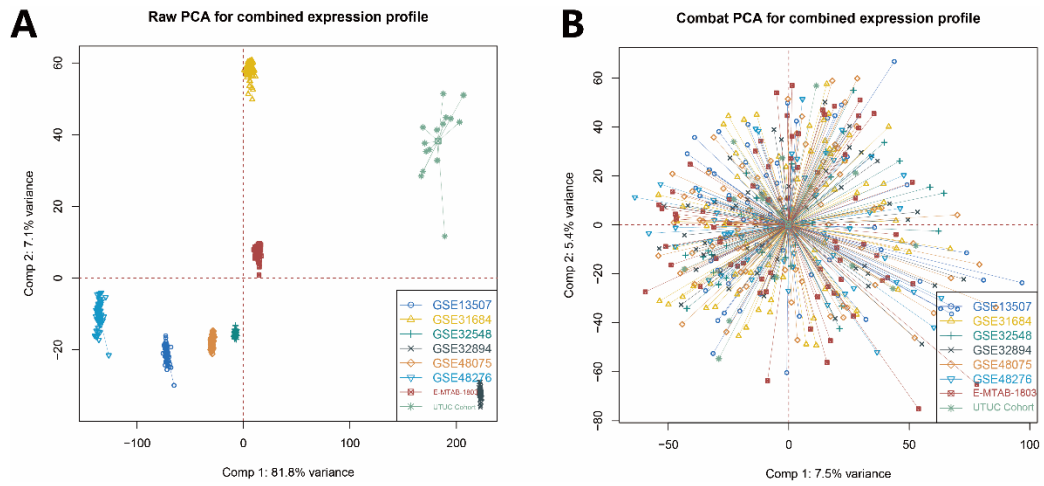
Figure S1: The distribution of each data set before and after batch correction through the PCA algorithm. (A) Before batch correction, (B) After batch correction.
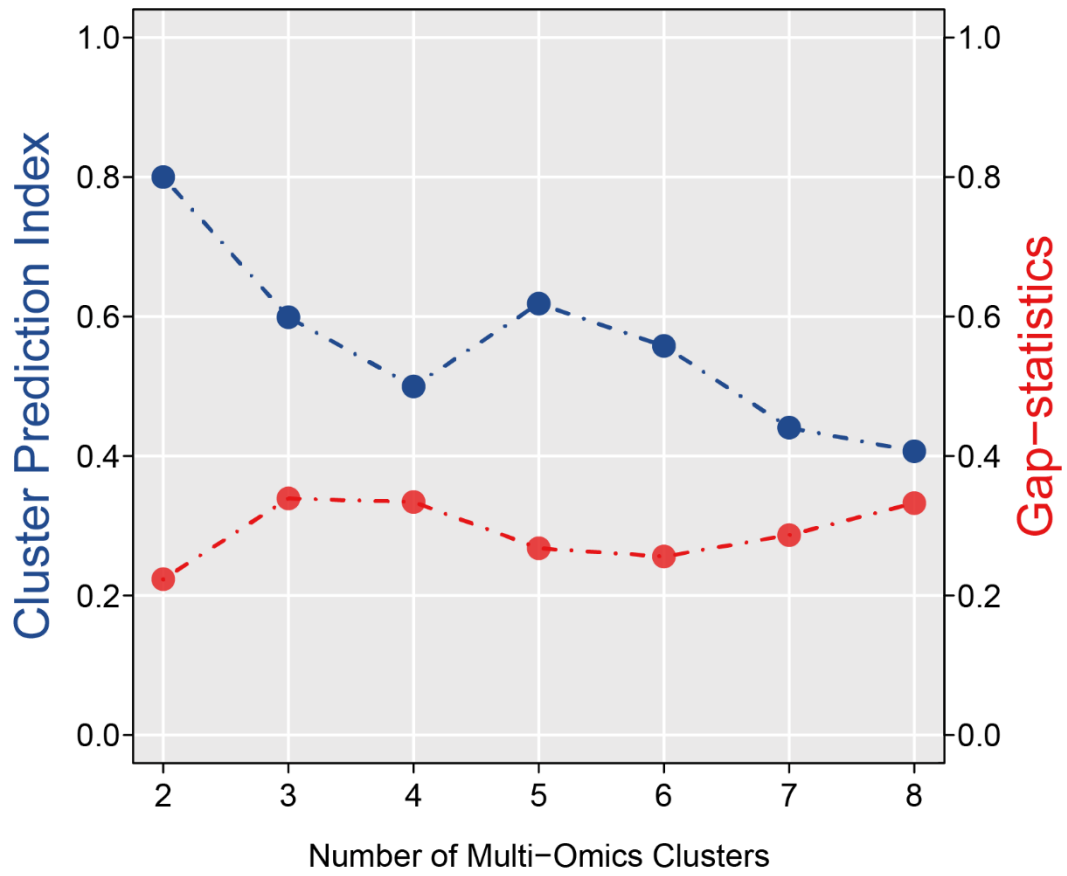
Figure S2: The Cluster Prediction Index and Gap Statistical Analysis of the multiomics clusters.

Silhouette plot
n = 386

3 clusters C_j
j : n_j | $ave_{i \in C_j}$ s_i

1 : 111 | 0.51

2 : 126 | 0.65

3 : 149 | 0.63

Silhouette width s_i

Average silhouette width : 0.6
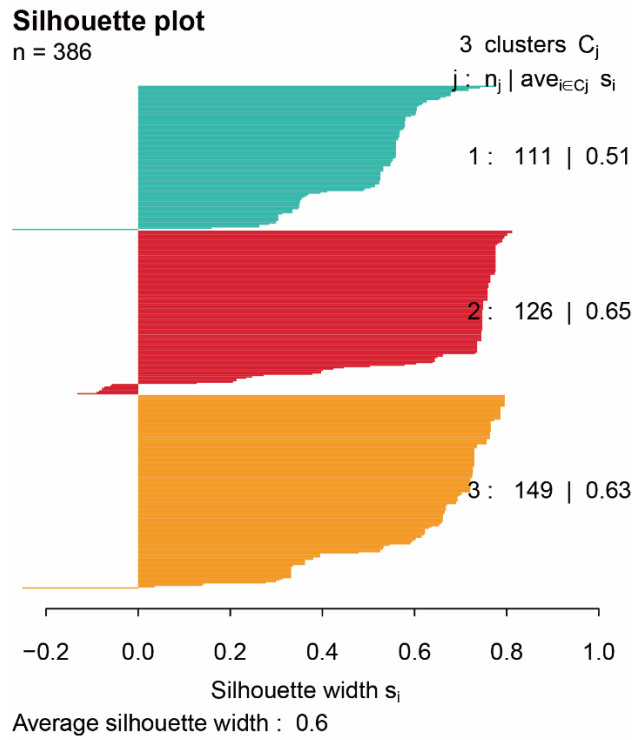
Figure S3: The sample similarity of each subgroup was assessed by calculating the Silhoutte score.

3
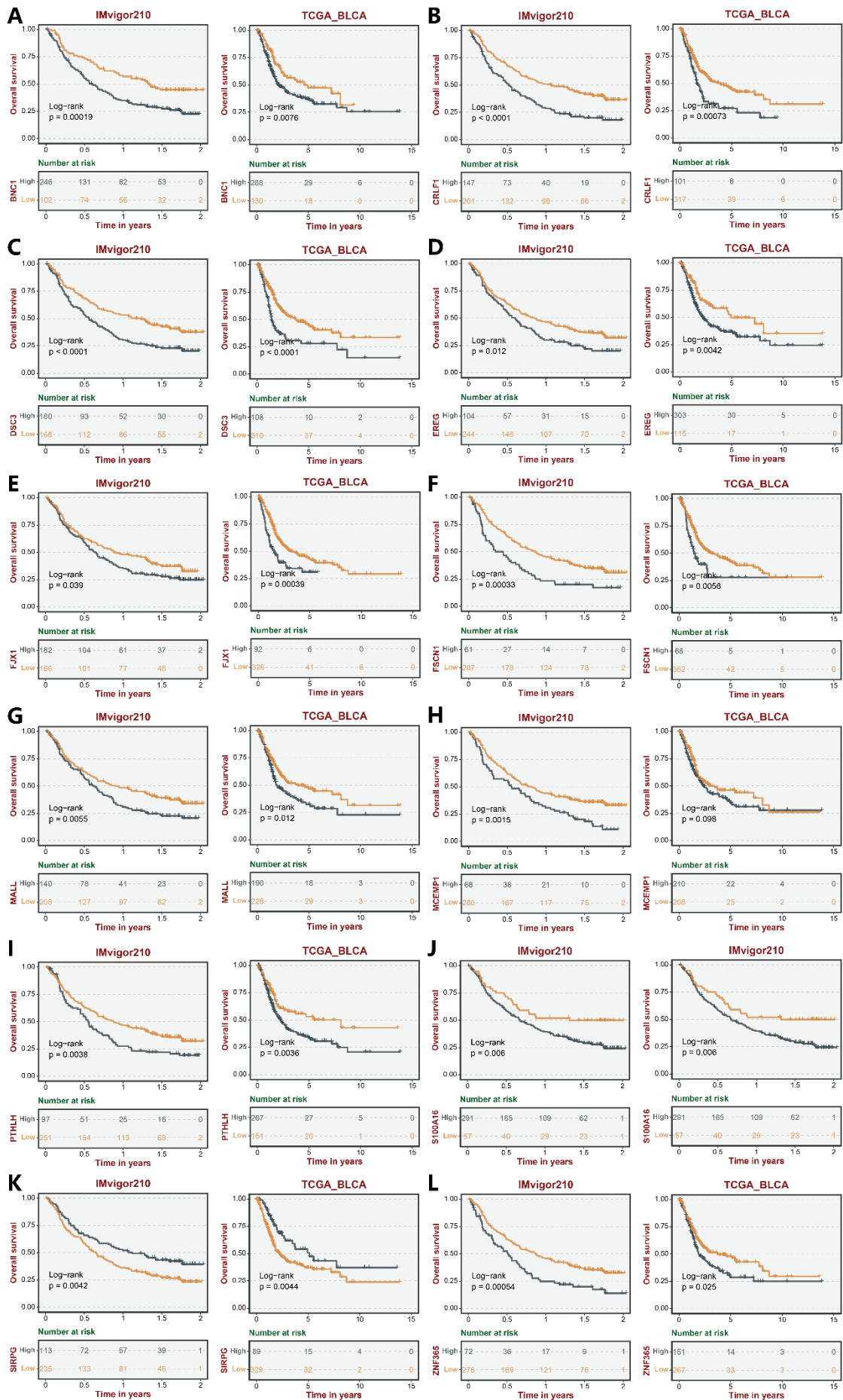
Figure S4: Survival analysis of hub genes in CMLS (OS) in IMvigor cohort and TCGA-BLCA cohort. (A) BNC1, (B) CRLF1, (C) DSC3, (D) EREG, (E) FJX1, (F) FSCN1, (G) MALL, (H) MCEMP1, (I) PTHLH, (J) S100A16, (K) SIRPG, (L) ZNF365.
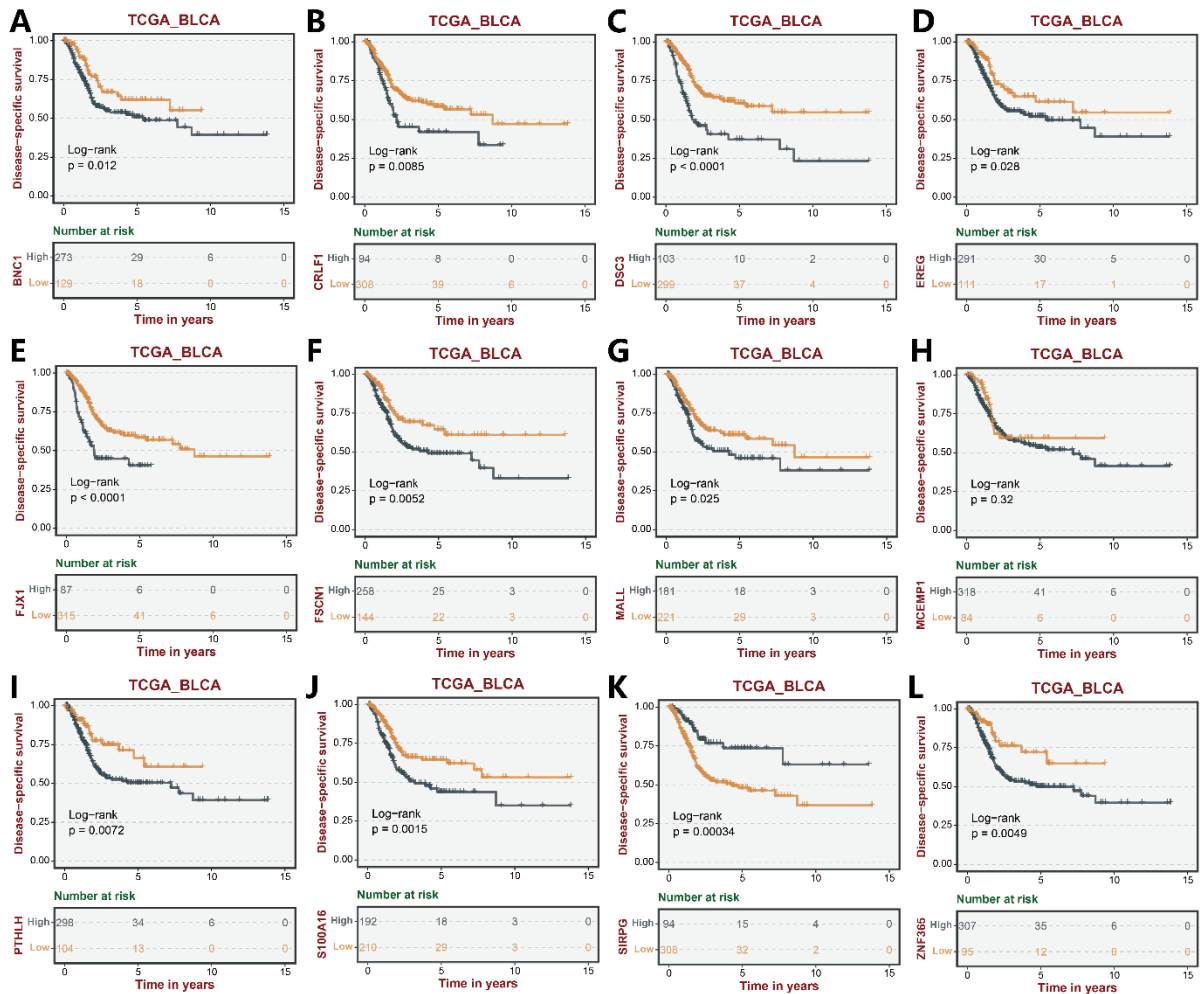
Figure S5: Survival analysis of hub genes in CMLS (DSS) in TCGA-BLCA cohort. (A) BNC1, (B) CRLF1, (C) DSC3, (D) EREG, (E) FJX1, (F) FSCN1, (G) MALL, (H) MCEMP1, (I) PTHLH, (J) S100A16, (K) SIRPG, (L) ZNF365.
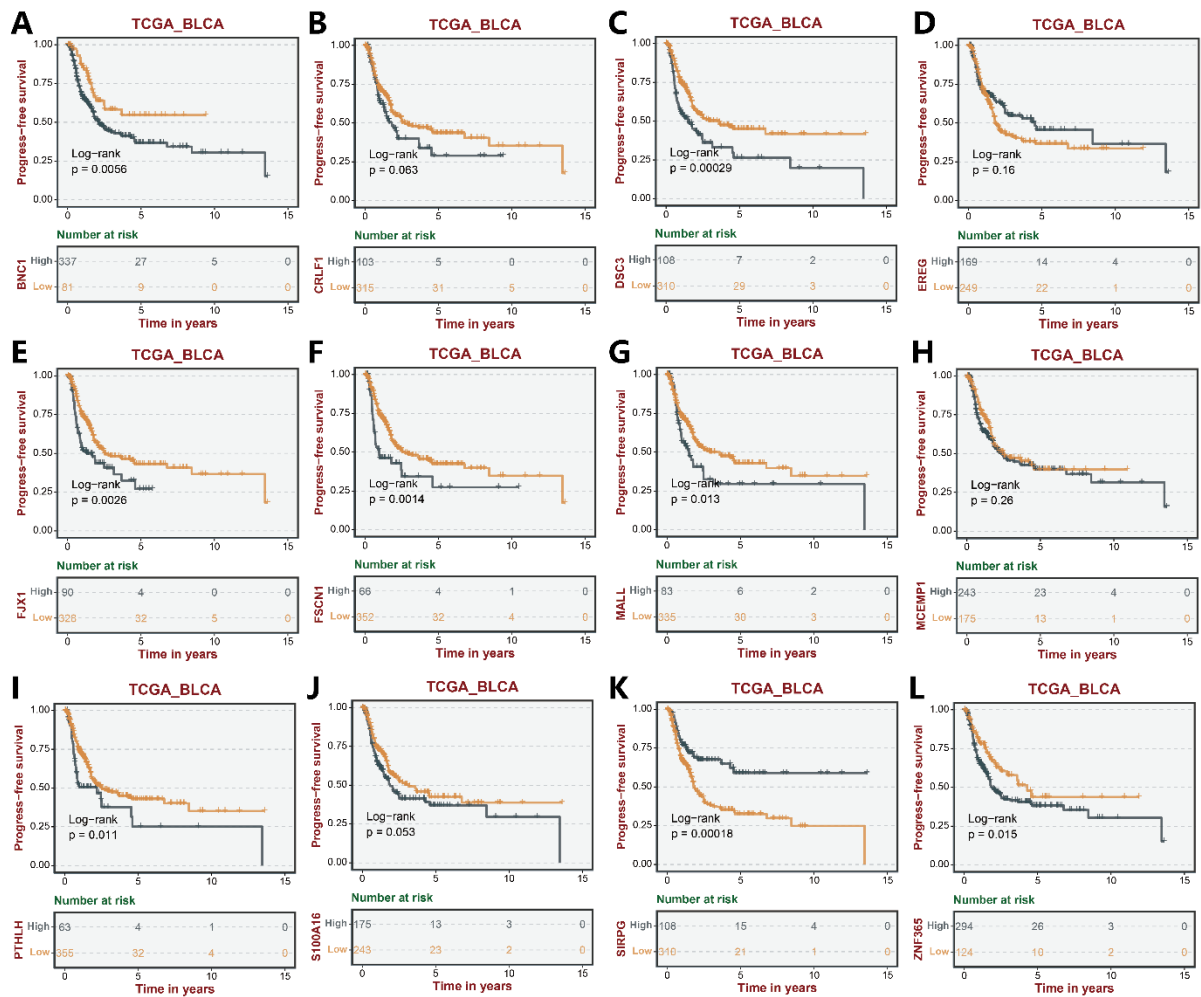
Figure S6: Survival analysis of hub genes in CMLS (PFS) in TCGA-BLCA cohort. (A) BNC1, (B) CRLF1, (C) DSC3, (D) EREG, (E) FJX1, (F) FSCN1, (G) MALL, (H) MCEMP1, (I) PTHLH, (J) S100A16, (K) SIRPG, (L) ZNF365.
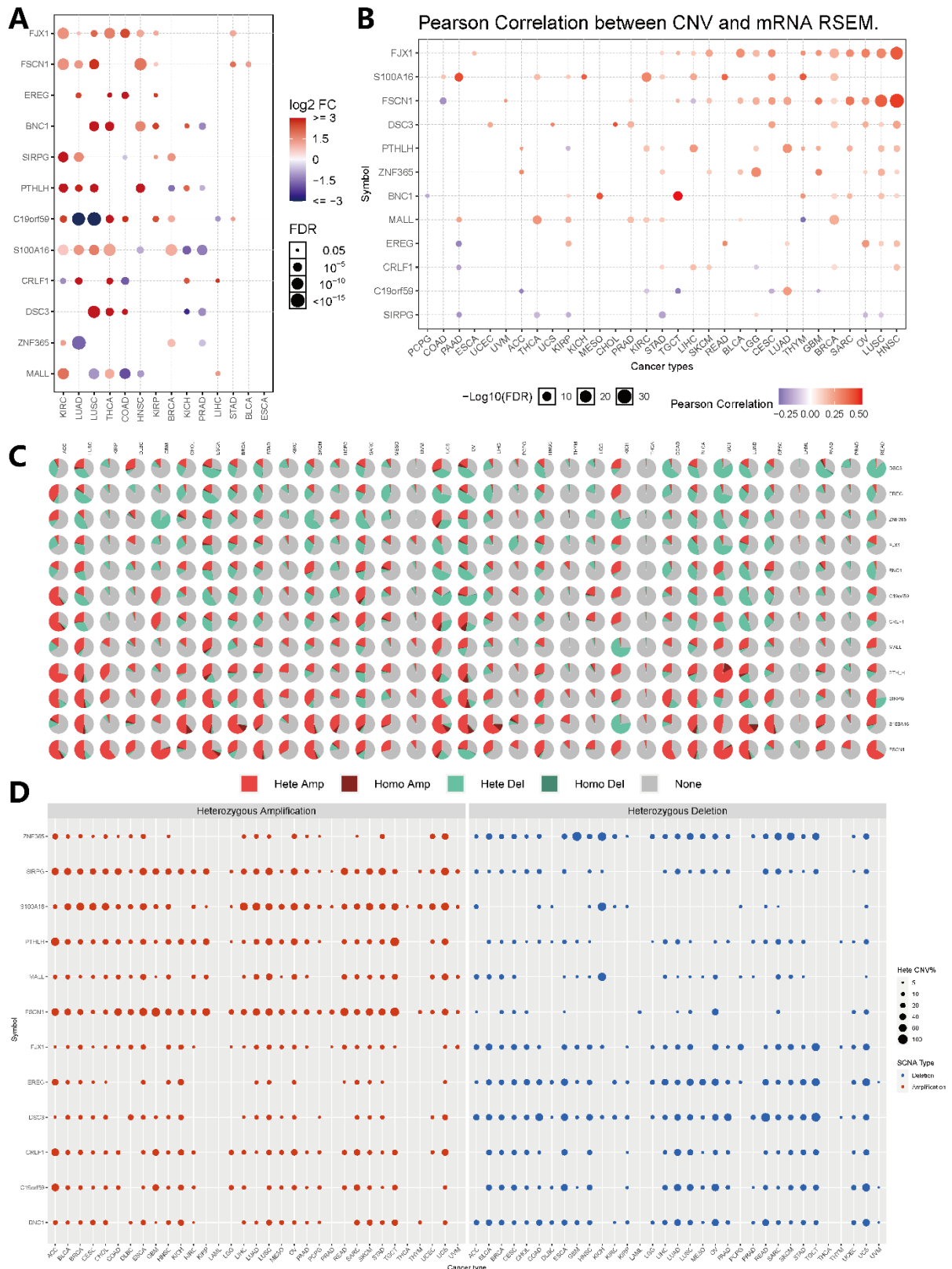
Figure S7: Differential expression analysis and CNV-related analysis of hub genes in CMLS. (A) Differential expression analysis of hub genes in CMLS. (B) Correlation analysis between CNV and mRNA expression level of hub genes in

CMLS. (C) The CNV pie distribution indicates the constitution of Heterozygous/Homozygous CNV of each hub genes in CMLS in each cancer. (D) The distribution of heterozygous and homozygous CNV of hub genes in CMLS in each cancer.
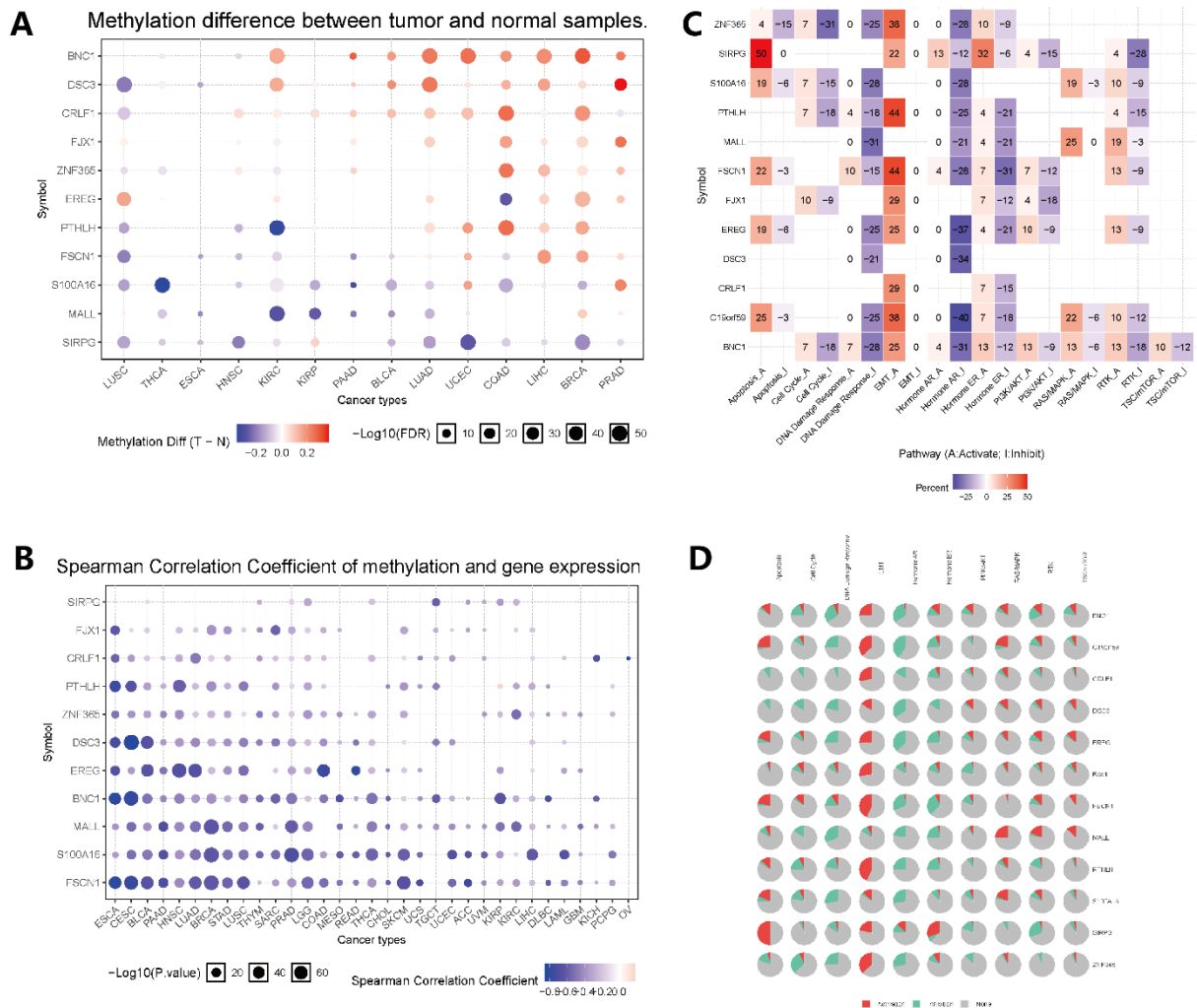
Figure S8: Methylation-related analysis and pathway analysis of hub genes in CMLS. (A) Differential methylation expression analysis of hub genes in CMLS. (B) Correlation analysis between methylation and mRNA expression level of hub genes in CMLS. (C)-(D) The Heatmap and the pie chart shows the

activation and repression status of hub genes in CMLS on related pathways.

A

Characteristics	P−value	HR	95% CI
IC.Level	0.002	0.69	0.546−0.873
TC.Level	0.739	0.959	0.75−1.226
TMB	0.001	0.957	0.931−0.983
Sex	0.801	1.059	0.683−1.639
TNB	1e−04	0.685	0.563−0.833
TCGA.Subtype	0.363	1.081	0.914−1.278
RS	8.872e−08	2.347	1.717−3.208

B

Characteristics	P−value	HR	95% CI
IC.Level	0.023	0.759	0.597−0.963
TMB	0.787	1.005	0.966−1.046
TNB	0.025	0.735	0.562−0.962
RS	2.809e−06	2.144	1.558−2.949
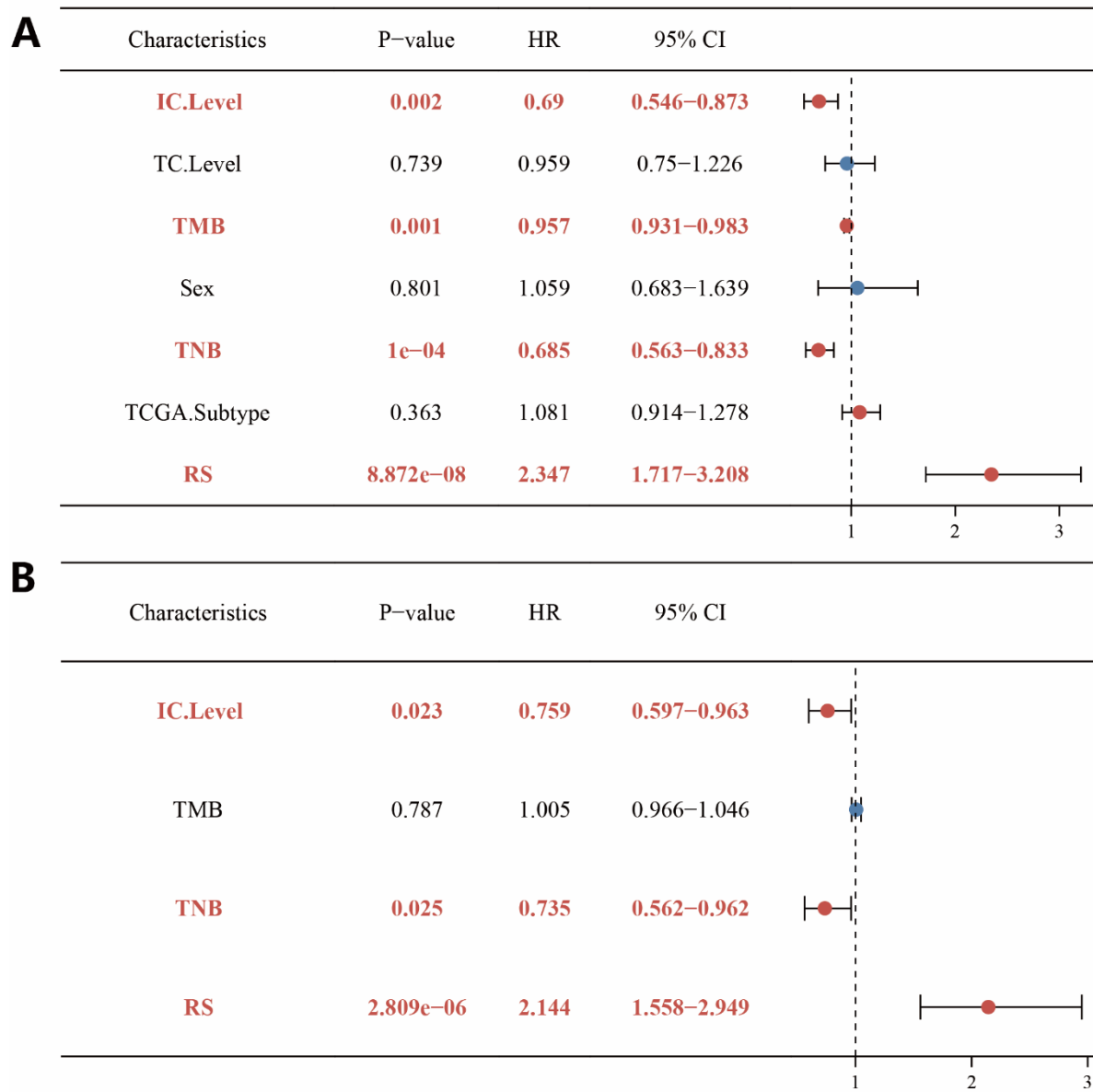
Figure S9: The independent prognostic analysis of hub genes in CMLS. (A) The univariate Cox analysis of hub genes in CMLS. (B) The multivariate Cox analysis of hub genes in CMLS.

Supplemental Table Legends：

Table S1：The result of differential expression analysis for three molecular subgroups.

Table S2：Annotation information of 12 hub genes of CMLS.

SYMBOL	ENSEMBL	Description
BNC1	ENSG00000169594	Basonuclin 1
CRLF1	ENSG00000006016	Cytokine Receptor Like Factor 1
DSC3	ENSG00000134762	Desmocollin 3
EREG	ENSG00000124882	Epiregulin
FJX1	ENSG00000179431	Four-Jointed Box Kinase 1
FSCN1	ENSG00000075618	Fascin Actin-Bundling Protein 1
MALL	ENSG00000144063	Mal, T Cell Differentiation Protein Like
MCEMP1	ENSG00000183019	Mast Cell Expressed Membrane Protein 1
PTHLH	ENSG00000087494	Parathyroid Hormone Like Hormone
S100A16	ENSG00000188643	S100 Calcium Binding Protein A16
SIRPG	ENSG00000089012	Signal Regulatory Protein Gamma

Table S3：The 22 collected signatures.

ID	PMID	Author	Gene	Coef
1	33204364	Cao-DSM	ACTA2	-0.000950773
1	33204364	Cao-DSM	ANPEP	0.000485521
1	33204364	Cao-DSM	COMP	-0.000333665
1	33204364	Cao-DSM	DAB2	0.008354169
1	33204364	Cao-DSM	ELN	0.007837673
1	33204364	Cao-DSM	FOXC2	-0.02707915
1	33204364	Cao-DSM	IGFBP2	-0.003429749
1	33204364	Cao-DSM	MGP	1.66E-06
1	33204364	Cao-DSM	MMP1	-5.03E-05
1	33204364	Cao-DSM	MYL9	-0.000124407
1	33204364	Cao-DSM	OXTR	0.003535212
1	33204364	Cao-DSM	PRSS2	-0.000130369
1	33204364	Cao-DSM	PTX3	0.00323547
1	33204364	Cao-DSM	RGS4	0.03024129
1	33204364	Cao-DSM	SPP1	4.99E-05
1	33204364	Cao-DSM	TAGLN	0.001122562

1	33204364	Cao-DSM	WNT5A	-0.001098872
2	31657881	Cao-JCMM	LAMA2	0.0382
2	31657881	Cao-JCMM	GPC1	0.0726
2	31657881	Cao-JCMM	ECM1	0.0717
2	31657881	Cao-JCMM	FBN2	0.0386
2	31657881	Cao-JCMM	LRP1	0.0685
2	31657881	Cao-JCMM	PVR	0.00903
2	31657881	Cao-JCMM	FUCA1	-0.0475
3	31737111	Chen-JC	TMPRSS11E	1.414
3	31737111	Chen-JC	SCEL	2.471
3	31737111	Chen-JC	KRT78	5.305
3	31737111	Chen-JC	TMEM185A	-2.988
4	32884943	Jiang-BRI	MMP9	0.0003617
4	32884943	Jiang-BRI	RBP7	0.0096326
4	32884943	Jiang-BRI	PDGFRA	0.0399101
4	32884943	Jiang-BRI	AHNAK	0.014035
4	32884943	Jiang-BRI	OAS1	-0.008116
4	32884943	Jiang-BRI	OLR1	0.0072663
4	32884943	Jiang-BRI	RAC3	0.0297334
4	32884943	Jiang-BRI	SLIT2	-0.209438
4	32884943	Jiang-BRI	IGF1	0.2684786
4	32884943	Jiang-BRI	AGTR1	0.1416244
5	34026819	Jiang-FMB	ANXA2	0.7752
5	34026819	Jiang-FMB	GALK1	0.764
5	34026819	Jiang-FMB	COL5A1	0.1973
5	34026819	Jiang-FMB	HS3ST1	-0.3086
6	34150627	Luo-FIO	ALDH1L2	0.096626563
6	34150627	Luo-FIO	ANXA1	0.035594173
6	34150627	Luo-FIO	CERCAM	0.016800805
6	34150627	Luo-FIO	GNA12	0.006655467
6	34150627	Luo-FIO	PLOD1	0.003607814
6	34150627	Luo-FIO	GSDMB	-0.175524875
7	32733809	Na-FIO	MMP9	0.000325672
7	32733809	Na-FIO	PDGFRA	0.029410199
7	32733809	Na-FIO	AHNAK	0.013037691
7	32733809	Na-FIO	OAS1	-0.007175947
7	32733809	Na-FIO	OLR1	0.00516244
7	32733809	Na-FIO	RAC3	0.023591819
7	32733809	Na-FIO	IGF1	0.282607117
7	32733809	Na-FIO	PGF	0.017645622
7	32733809	Na-FIO	SH3BP2	-0.083043743

8	32117435	Qiu-FIG	RBP7	0.0044
8	32117435	Qiu-FIG	PDGFRA	0.0021
8	32117435	Qiu-FIG	AHNAK	0.0054
8	32117435	Qiu-FIG	OAS1	-0.0021
8	32117435	Qiu-FIG	RAC3	0.0043
8	32117435	Qiu-FIG	EDNRA	0.0142
8	32117435	Qiu-FIG	SH3BP2	-0.0019
9	32749145	Quan-BM	MMP9	0.00036
9	32749145	Quan-BM	RBP7	0.00963
9	32749145	Quan-BM	PDGFRA	0.03991
9	32749145	Quan-BM	AHNAK	0.01403
9	32749145	Quan-BM	OAS1	-0.008112
9	32749145	Quan-BM	OLR1	0.00727
9	32749145	Quan-BM	RAC3	0.02973
9	32749145	Quan-BM	SLIT2	-0.20944
9	32749145	Quan-BM	IGF1	0.26848
9	32749145	Quan-BM	AGTR1	0.14162
10	33459039	Shen-ACNB	ANXA6	0.0082
10	33459039	Shen-ACNB	AHNAK	0.0161
10	33459039	Shen-ACNB	CACYBP	0.0252
10	33459039	Shen-ACNB	STAT1	-0.0079
10	33459039	Shen-ACNB	CALR	0.0011
11	33292266	Shen-CCI	PLOD1	0.0049
11	33292266	Shen-CCI	CKB	0.0018
11	33292266	Shen-CCI	PYGB	0.0033
11	33292266	Shen-CCI	AKR1B1	0.0031
11	33292266	Shen-CCI	PDE5A	0.0486
12	33456631	Shen-DM	AHNAK	0.0009
12	33456631	Shen-DM	NR2F6	-0.006
13	34112144	Wang-BMC	ACAP1	-0.0589
13	34112144	Wang-BMC	ADAMTS9	0.0515
13	34112144	Wang-BMC	TAP1	-0.0022
13	34112144	Wang-BMC	IFIT3	-0.001
13	34112144	Wang-BMC	FBN1	0.0184
13	34112144	Wang-BMC	FSTL1	0.0029
13	34112144	Wang-BMC	COL6A2	2.46E-06
14	32841548	Wang-CM	SULT1C2	-0.285729313
14	32841548	Wang-CM	EGFL6	-0.132834959
14	32841548	Wang-CM	DUSP2	-0.047080965
14	32841548	Wang-CM	BLNK	-0.036593582
14	32841548	Wang-CM	BATF	-0.027054952

14	32841548	Wang-CM	PPFIBP2	-0.015812131
14	32841548	Wang-CM	HES1	-0.00830084
14	32841548	Wang-CM	THBS4	0.011476589
14	32841548	Wang-CM	CPA3	0.061002869
14	32841548	Wang-CM	NEFL	0.069959198
14	32841548	Wang-CM	VWF	0.0806877
14	32841548	Wang-CM	LRP8	0.09542848
14	32841548	Wang-CM	PDGFD	0.110454703
15	33391264	Wu-FIM	ORM1	-0.42
15	33391264	Wu-FIM	PTHLH	0.17
15	33391264	Wu-FIM	CTSE	-0.39
16	33545950	Wu-Med	CHPF	0.156689
16	33545950	Wu-Med	AK3	-0.40384
16	33545950	Wu-Med	GALK1	0.350739
16	33545950	Wu-Med	NUP188	0.444476
17	33209669	Xing-TAU	TAP1	-0.000824
17	33209669	Xing-TAU	RBP7	0.00397
17	33209669	Xing-TAU	STAT1	-0.000031
17	33209669	Xing-TAU	PDGFRA	0.005626
17	33209669	Xing-TAU	AHNAK	0.004692
17	33209669	Xing-TAU	OLR1	0.00029
17	33209669	Xing-TAU	RAC3	0.009473
17	33209669	Xing-TAU	EDNRA	0.001882
17	33209669	Xing-TAU	IGF1	0.078175
17	33209669	Xing-TAU	SH3BP2	-0.009884
18	33408272	Yan-Aging	AHNAK	0.2427
18	33408272	Yan-Aging	CALR	0.3584
18	33408272	Yan-Aging	PLA2G2A	-0.0863
18	33408272	Yan-Aging	OAS1	-0.188
18	33408272	Yan-Aging	CDH1	-0.0401
18	33408272	Yan-Aging	PDGFRA	0.2079
18	33408272	Yan-Aging	SEMA3F	-0.0014
18	33408272	Yan-Aging	RAC3	0.2281
18	33408272	Yan-Aging	IRF5	-0.0702
18	33408272	Yan-Aging	CARD11	-0.1192
19	31134129	Yan-FIG	BUB1B	0.347
19	31134129	Yan-FIG	CCNB1	0.324
19	31134129	Yan-FIG	CDK1	0.218
19	31134129	Yan-FIG	ISG15	0.264
19	31134129	Yan-FIG	KIF15	0.268
19	31134129	Yan-FIG	RAD54L	0.386

20	34221960	Yan-FIO	BOC	0.009
20	34221960	Yan-FIO	FGF7	0.064
20	34221960	Yan-FIO	MAP1A	0.178
21	32194788	Yan-JC	ASPM	-0.135
21	32194788	Yan-JC	C4orf46	-0.142
21	32194788	Yan-JC	CCNB1	0.632
21	32194788	Yan-JC	DIAPH3	0.105
21	32194788	Yan-JC	MLF1	0.803
21	32194788	Yan-JC	MTFR2	-0.456
21	32194788	Yan-JC	NSUN6	-0.287
21	32194788	Yan-JC	OIP5	0.05
22	31894276	Yin-OR	S100A8	0.17027
22	31894276	Yin-OR	CTHRC1	0.06678
22	31894276	Yin-OR	C15orf48	0.10548
22	31894276	Yin-OR	MPPED2	-0.16639
22	31894276	Yin-OR	CKAP2L	0.256
22	31894276	Yin-OR	SNCAIP	0.17159
22	31894276	Yin-OR	SALL4	0.06363
22	31894276	Yin-OR	MMP11	0.01503
22	31894276	Yin-OR	COL1A1	0.00057
22	31894276	Yin-OR	AEBP1	0.22674
22	31894276	Yin-OR	C10orf99	-0.08764
22	31894276	Yin-OR	KIF1A	0.22416