Article | Received 17 May 2026; Revised 9 July 2026; Accepted 20 July 2026; Published 10 September 2026
https://doi.org/10.55092/acr20260013

Identification PDF as a biomarker of "cold" CRC using integrated bioinformatics and machine learning

Haoming Chen[1], Ziyue Zhong[1], Jianyu Lin[1] and Xianli Shi[2,*]

[1] Department of Biotechnology, School of Life Sciences and Biopharmaceutics, Guangdong Pharmaceutical University, Guangzhou, China

[2] Department of Biochemistry, School of Basic Medical Science, Guangdong Pharmaceutical University, Guangzhou, China

* Correspondence author; E-mail: shixl@gdpu.edu.cn.

Highlights:

- CRC was classified into immune cold/hot subtypes, with distinct TMB, stemness, drug sensitivity, and OXPHOS metabolism characterized.
- Ensemble machine learning identified PDF as a cold CRC biomarker associated with suppressed immunity, enhanced proliferation, and a worse prognosis.
- PCBP1 was identified as a potential transcriptional regulator of PDF.

Abstract: Colorectal cancer (CRC) is strikingly immunologically heterogeneous but the mechanistic basis for the immune-refractory "cold" feature is unknown. On basis of The Cancer Genome Atlas-Colon Adenocarcinoma (TCGA-COAD) data, we distinguished tumors into "cold" and "hot" subtypes. We discovered peptide deformylase (PDF) as a marker of cold CRC by multi-algorithm differential expression and machine learning (Support Vector Machine (SVM), Random Forest (RF), XGBoost). High expression of PDF correlated with poor prognosis, low immune infiltration and high level of oxidative phosphorylation (OXPHOS). Mechanismwise, according to extant literature evidence and bioinformatic inference, PDF is speculated to deformylate N-terminal of mitochondrial new-peptides, thereby contributing to maturation of ETC and oxidative phosphorylation to afford ATP supplies to accelerate tumor growth.Concurrently, this process may curtail the release of N-formyl peptides—damage-associated molecular patterns that recruit $CD8^+$ T cells and macrophages via Formyl Peptide Receptor 1 (FPR1)—thereby reinforcing an immune-excluded microenvironment. Bioinformatic analyses further nominated Poly(rC)-binding protein 1 (PCBP1) as a potential transcriptional regulator of PDF, a finding supported by an independent Clinical Proteomic Tumor Analysis Consortium—Phase 2 (CPTAC-2) cohort. Our work suggests a PCBP1–PDF–OXPHOS axis associated with immune suppression in CRC.

Keywords: colorectal cancer; hot and cold tumor; PDF gene; machine learning; OXPHOS

1. Introduction

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer-related death globally [1–4] Although immunotherapy has transformed cancer treatment, its efficacy is largely restricted to "hot" tumors with high immune infiltration and Microsatellite Instability–High (MSI-H) status, while "cold" CRCs, characterized by minimal tumor-infiltrating lymphocytes, low Programmed Death-Ligand 1 (PD-L1) expression, and immunosuppressive microenvironments, remain refractory to immune checkpoint inhibitors (ICIs) [5–9].Elucidating the mechanisms driving the cold tumor phenotype is essential for developing strategies to convert cold tumors into hot ones.

In this study, we classified The Cancer Genome Atlas-Colon Adenocarcinoma (TCGA-COAD) samples into "hot" and "cold" subtypes, and identified peptide deformylase (PDF) as a robust biomarker of cold CRC through multi-algorithm differential expression analysis and machine learning screening (Support Vector Machine (SVM), Random Forest (RF), and XGBoost). Mechanistically, existing evidence from non-CRC cancer models suggests that PDF may promote oxidative phosphorylation and adenosine triphosphate (ATP) generation by hydrolyzing mitochondrial N-formylmethionine (fMet) on electron transport chain components, while potentially degrading fMet-peptide DAMPs that recruit CD8+ T cells and macrophages via Formyl Peptide Receptor 1 (FPR1), thereby reinforcing immune exclusion. These mechanistic links remain to be experimentally validated in CRC cells and tissues. Furthermore, Poly(rC)-binding protein 1 (PCBP1) was identified as a potential transcriptional regulator of PDF. These findings suggest that PDF is associated with the cold tumor phenotype, while PCBP1 functions as a potential regulator of PDF. Moreover, targeting PDF may offer a promising strategy to "warm" cold CRCs and enhance immunotherapy responsiveness.

2. Methods

2.1. Data sources and tumor classification

TCGA-COAD data set was used, including the gene expression data (both (Transcripts Per Kilobase of exon model per Million mapped reads) TPM and Raw Read Count values) of 524 samples, including 483 tumor samples and 41 normal tissue samples. Except for differential gene analysis, all our analyses will focus on TPM data types. We performed initial clustering into three distinct subgroups using Consensus Cluster Plus (v3.20) [10] in R 4.4.3, with the Partitioning Around Medoids (PAM) algorithm and Spearman distance.

2.2. Classification of "hot" and "cold" CRC

After initial sample classification, we comprehensively characterized immune infiltration in tumor tissues using the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) R package (version 1.0.13) [11] (R version 4.4.3). This algorithm yields four quantitative indices: ESTIMATE score, immune score, stromal score, and tumor purity, based on which we stratified tumors into immune "cold" and "hot" subtypes according to the cohort-wide distribution of the four metrics. We further applied the Microenvironment Cell Populations-counter (MCP-counter) algorithm [12] embedded in Immuno-Oncology Biological Research (IOBR) (v2.2.3.9000) [13] to the

three predefined clusters, with Dunn's post hoc test for intergroup pairwise comparisons. Finally, we separately validated the cold/hot tumor stratification using the MCP-counter and ESTIMATE.

2.3. Analysis of immune cell infiltration abundance

Analysis of relative immune cell abundance of cold and hot tumor subgroups was generated by the TIMER algorithm incorporated into IOBR (v2.2.3.9000) ourselves. This allows the pairwise comparison of immune infiltration landscape between cold and hot CRC subgroups to determine different TME features. To verify our results further, we also did cross-comparison analysis using the TIMER algorithm [14] and the MCP-counter.

2.4. TMB analysis

Using somatic mutation data obtained from the TCGA database, we tested the mutational landscape of all tumor samples and screened genes with high frequency of mutation. We performed the computational test of these genes using the maftools package (version 2.22.0) [15] in R language version 4.4.3, and calculated the Tumor Mutation Burden (TMB) for each sample. We further made a comparison of TMB difference between "cold" and "hot" tumors formed by the immunologic classification.

2.5. Analysis of the overall drug sensitivity of hot and cold tumors

Comprehensive drug sensitivity analysis was performed on the tumour samples of the 'cold' and 'hot' groups with the oncoPredict package (Version 1.2) [16] performed in R version 4.4.3, which allows systematic calculation of differential drug response of the different types of immunologically 'cold' tumours and 'hot' tumours..

2.6. Tumor stemness index and cell cycle analysis

To reveal the stemness between the hot and cold tumor subgroups, we perform the One-Class Logistic Regression (OCLR) algorithm [17] and trained with the Progenitor Cell Biology Consortium (PCBC) dataset (available on https://www.synapse.org). Furthermore, we integrated the G0 gene set [18] with gene set enrichment analysis algorithm (Gene Set Enrichment Analysis, GSEA) algorithm [19] to analyze the progression pattern of the two tumor subtypes, cell cycle.

2.7. Analysis of differentially expressed genes between hot and cold tumors

Differentially expressed genes (DEGs) between cold and hot tumors were examined by four different methods: DESeq2 (version 1.46.0) [20], Wilcoxon rank-sum test, limma (version 3.62.2) [21], and edgeR (Version 4.4.2) [22], both in R version 4.4.3. Significantly different expressed genes (DEGs) were selected bed using very stringent criteria: significance cutoff is FDR < 0.05 and absolute value of log2 fold change > 1. Subsequently, intersection comparisons were conducted on the significantly upregulated genes of cold tumors identified by each of the four methods. All intersect genes obtained in above test are considered to be a common core of upregulated genes of cold tumors compared with hot tumors.

2.8. KEGG analysis of the upregulated DEGs in "cold" CRC

The upregulated DEGs found from the screening were performed Kyoto Encyclopedia of Genes and Genomes (KEGG) [23] pathway analysis using the Metascape platform [24], Analysis results of Kyoto Encyclopedia of Genes and Genomes were analyzed and visualized for precise interpretation.

2.9. Machine learning algorithms-based screening of "cold/hot" CRC biomarker genes

Intersecting the common DEGs further implicated immune-related genes in the Immport database [25]. Candidate immune-related DEGs were subjected to feature screening and ensemble machine learning modeling based on three base algorithms: random forest (RF) [26], support vector machine (SVM) [27], and XGBoost [28]. before modeling, the whole dataset was randomly divided into a train set (80%) and an inner test set (20%) according to sample labels to ensure that data leakage did not happen when performing feature selection. For the RF algorithm, the R package randomForest [29] was used; a random forest model with 500 trees (ntree = 500) was initially constructed, and the best number of trees was to minimize the total error rate of the model. The importance of candidate genes was calculated according to the mean decrease Gini coefficient, and the top 5 priority genes were taken as preliminary feature candidates. For the SVM algorithm, the R package e1071 [30] was used for SVM with a radial basis kernel and feature scale applied; feature importance was defined as the absolute value of linear regression coefficients calculated from the trained linear SVM model, and the top 5 genes were picked accordingly. For XGBoost analysis, R package xgboost [28]. Reference was used. Normalized and log2-transformed TPM expression matrix was transposed to a sample-specific feature matrix and it was converted to an xgb.DMatrix sparse data object. For XGBoost model, we configured the hyper parameters of XGBoost model, such as max_depth = 3, eta = 0.3, objective = 'binary: logistic', nrounds = 30, and model training with no verbose. Gene importance was calculated on the basis of the Gain metric and the top 5 predictive genes were screened. For robust biomarker set, overlapping genes from top 5 feature genes in the ranking of the top 3 algorithms were intersected as the final consensus gene set of all biomarker genes. When intersection yielded fewer than 3 genes, genes in more than 2 of the 3 ranking lists were supplemented for modelling and stability. According to the optimized consensus genes, 3 independent base models (RF, SVM and XGBoost) were formulated on the training cohort. To enhance predictive robustness and differentiation performance further, 2 methods of ensemble learning based on soft voting were applied to combine the predicted probability results of the 3 base models: average weighted ensemble model and ensemble model for voting by the AUC (area under the receiver operating characteristic curve).For the weighted ensemble strategy, the weight of each base model was determined by its test-set AUC value, where models with higher discriminative performance were assigned greater weights. All base models and ensemble models were comprehensively evaluated on the unseen test cohort using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity in the test cohort, and calibration curves; the AUC-weighted ensemble model was finally determined as the optimal diagnostic model for distinguishing cold and hot tumors.

2.10. Survival analysis

Overall survival (OS) analysis was performed on two independent colorectal cancer cohorts, GSE17538 and GSE17537, retrieved from the Gene Expression Omnibus (GEO) database. The patients of the merged group had 287 unique patients after omitting the repeated samples. The best expression cutoff of each candidate gene was calculated by the maximally selected rank statistics (minimum P-value approach) by the surv_cutpoint function of survminer R package (version 0.4.9), minprop = 0.1, which categorizes patients into a high-expression group and a low-expression group on the basis of the cutoff with the smallest log-rank P value. The survival curves were estimated by the Kaplan–Meier method and compared by the log-rank test. Survival curves were estimated by the Kaplan–Meier method and compared using the log-rank test. Univariate Cox proportional hazards regression was applied to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). Multivariate Cox regression was further performed to adjust for age, gender, AJCC stage, and tumor grade, with dataset (GSE17537 *vs.* GSE17538) included as a stratification variable (strata(dataset)) to control for potential batch effects between cohorts. The proportional hazards (PH) assumption was assessed using Schoenfeld residual tests (cox.zph). As the survival analysis targeted three *a priori* candidate genes identified by machine learning rather than a genome-wide screen, no multiple testing correction was applied. All survival analyses were conducted in R (version 4.4.3) using the survival (version 3.5–5) and survminer packages. A two-sided $P < 0.05$ was considered statistically significant.

2.11. Single gene immune infiltration analysis

To evaluate the impact of the key gene on tumor purity and immune cell infiltration, single-gene immune infiltration analysis was performed using the TIMER 1.0 database [31]. This approach enabled the systematic evaluation of the relationship between the Key gene and the tumor microenvironment.

2.12. Single gene pathway activation analysis

To investigate the functional impact of the identified key genes, Single-gene pathway activation analysis was performed using the Gene Set Cancer Analysis (GSCA) database [32]. This analysis was performed to systematically evaluate the activation effects exerted by these key genes on relevant biological pathways.

2.13. Transcription factor analysis of key gene

The ChIP-Atlas database [33] was utilized to identify the potential transcription factors (TFs) regulating *PDF* expression. Subsequently, the identified TFs were cross-referenced with the DEGs to obtain the transcription factors with significant differential expression between “cold” and “hot” CRC. Finally, the IGA software (version 2.19.4) [34] was employed to visualize the predicted transcription factor binding sites in the *PDF* promoter region. The cBioPortal database was utilized to validate the expression correlation between key transcription factors and *PDF*.

2.14. The validation analysis

The colon cancer dataset from the CPTAC-2 prospective cohort (Cell, 2019) was retrieved from the cBioPortal database [35,36]. Another independent colon cancer dataset (accession number: GSE220148)

was obtained from the GEO database. A total of 213 samples from the two datasets were included in the subsequent analysis. The previously established integrated learning model was applied to stratify all enrolled samples into cold and hot tumor immune subtypes. The ESTIMATE algorithm was utilized to evaluate the accuracy and reliability of the tumor immunophenotype classification. Subsequently, the limma package was employed to perform differential expression analysis between cold and hot tumors to screen for candidate core biomarkers associated with immune landscape heterogeneity in colon cancer.

3. Results

3.1. The overall workflow

An overview of the overall workflow of this study was provided in the Figure 1.

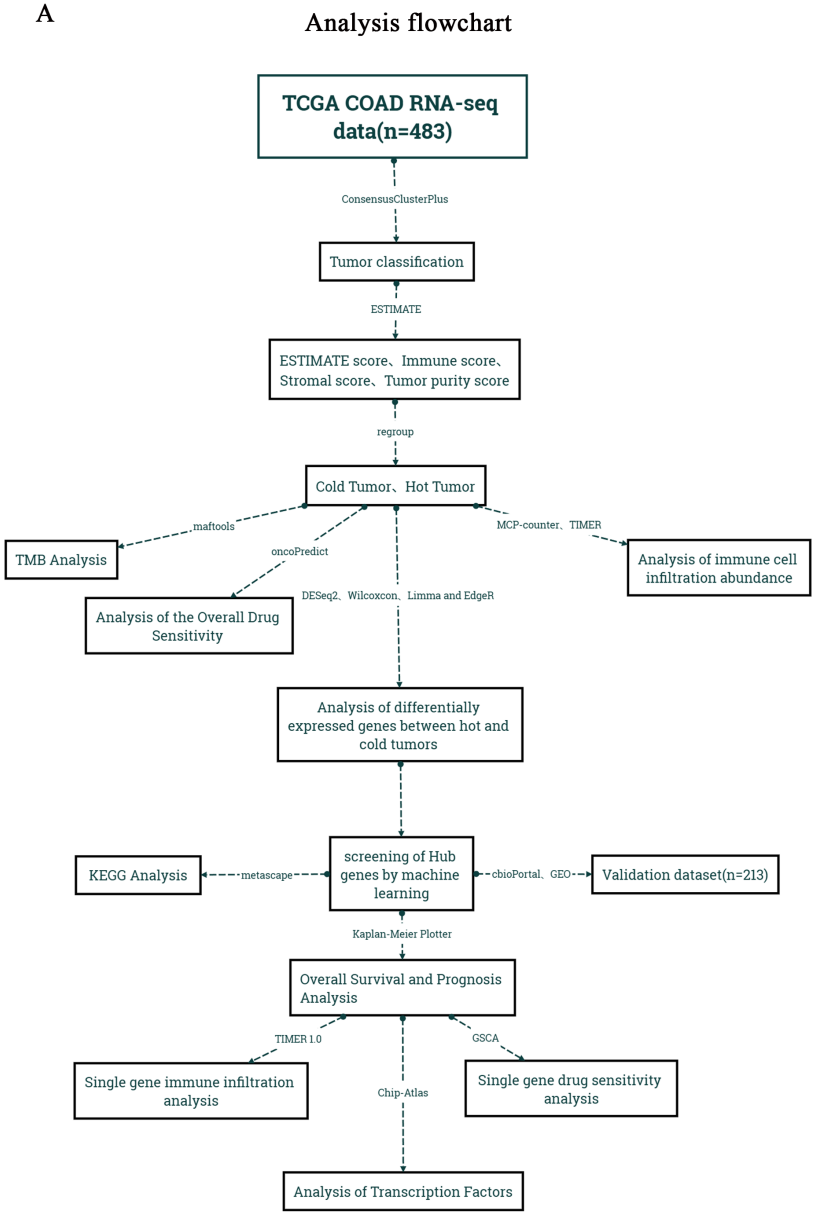

Figure 1. The overall workflow for identifying candidate biomarkers of immune "cold" colorectal cancer. TCGA-COAD samples were stratified into "hot" and "cold" subtypes via unsupervised consensus clustering followed by immune characterization, and candidate biomarkers were screened through multi-algorithm differential expression analysis, immune-related gene filtering, and ensemble machine learning.

3.2. Classification of "hot" and "cold" CRC

The Cancer Genome Atlas-Colon Adenocarcinoma dataset, comprising 483 tumor and 41 normal tissue samples, was analyzed. Unsupervised consensus clustering of the tumor samples was performed using the ConsensusClusterPlus package (version 3.20) in R 4.4.3, identifying three distinct clusters: Cluster 1 (n = 182), Cluster 2 (n = 158), and Cluster 3 (n = 143) (Figure 2A). Comparison of the ESTIMATE scores (immune score, stromal score and tumor purity) of three samples groups was markedly heterogeneous between groups (Figure 2B). Cluster 1 had the largest ESTIMATE score, immune score and stromal score, while the two other clusters (2 and 3) had smaller scores. On the other hand, cluster 1 had the smallest tumor purity, while clusters 2 and 3 had larger tumor purity scores. Another measure of the abundance of immune cells was the MCP-counter algorithm. Cluster 1 had the largest abundance of T cell, CD8+ T cell, cytotoxic lymphocyte, and other immune cell groups. Dunn's post hoc test verified that the abundance of cytotoxic lymphocytes, NK cells, monocyte lineage cells, and dendritic cells of Cluster 1 was higher than that of Cluster 2 or 3 significantly, and there was no significant difference between Cluster 2 and 3 (Figure 2C).

On the basis of this, Cluster 1 had the highest value of immune infiltration and the lowest tumor purity and was further grouped as "Hot" CRC group. Cluster 2 and 3 were with lower immune scores and higher tumor purity, and were hence defined as "Cold" CRC group together. After reclassification into "Hot" and "Cold" CRC groups, the tumor immune infiltration was analysed again (Figure 2C), which revealed significant difference between "Hot" and "Cold" CRC groups in ESTIMATE score, immune, stromal and tumor purity scores (Figure 2D).

3.3. Distinct immunological profiles between "hot" and "cold" CRC subgroups

Comparison between immune cell infiltration in "hot" and "cold" CRC groups demonstrated different immunologic characteristics. By the MCP-counter algorithm, the immune-hot group had higher levels of infiltrated abundances of T cells, $CD8^+$ T cells, cytotoxic T lymphocytes, NK cells, B-lineage cells, monocytic lineage cells, and myeloid DC compared with the immune-cold group. We use another algorithm, TIMER, for cross-validation of immune infiltration. Consistent with the MCP-counter algorithm, hot tumours showed higher levels of $CD4^+$ T cell, $CD8^+$ T cells, B cells, neutrophil, macrophage and DC abundances compared with cold tumours. Integrated analysis using the two deconvolution algorithms consistently revealed significant enrichment of T cells, B cells, and dendritic cells (DCs) within immune-hot tumors, reinforcing the robust, immune-exuberant phenotype of the hot CRC subtype (Figure 2E).

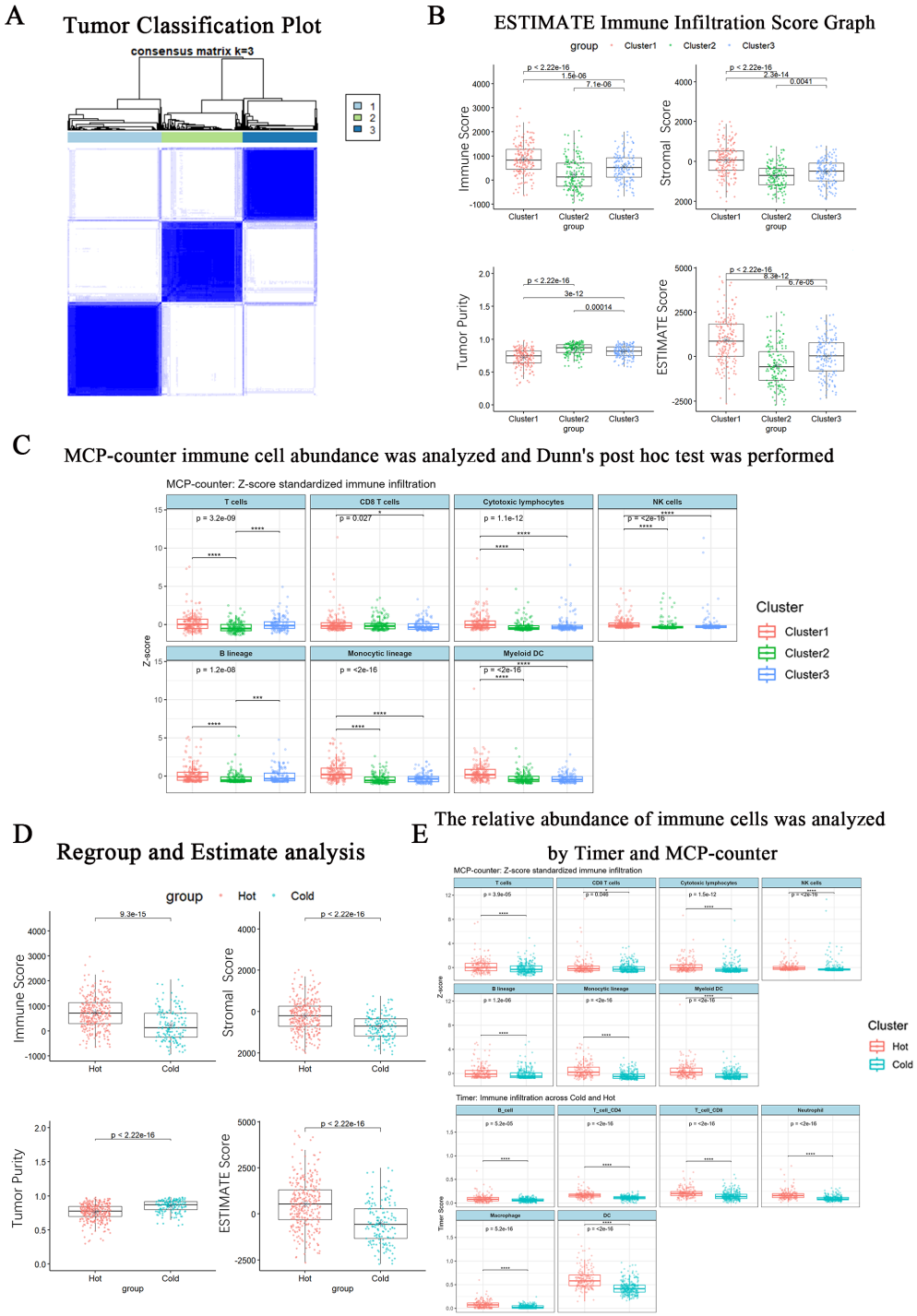

Figure 2. Classification of "cold" and "hot" tumors in colon cancer. **A**, using the ConsensusClusterPlus package, a consistency clustering analysis was conducted on the tumor samples, and the current tumor samples were classified into 3 categories; **B**, using the Estimate package, the tumor immune infiltration analysis was conducted on the divided tumor samples, including ESTIMATE score, immune score, stromal score, and tumor purity score; **C**, MCP-counter immune abundance analysis and Dunn's post hoc test were used to verify the correctness of the groups; **D**, after re-grouping, an "estimate" analysis is conducted to confirm the correctness of the division between cold and hot tumors; **E**, the MCP-counter and Timer analysis of immune cell abundance revealed the differences in the composition of immune cells between cold and hot tumors. Through cross-analysis of the two groups, the most likely different immune cells were identified.

3.4. Lower TMB in "cold" CRC and higher TMB in "hot" CRC

Tumor mutation burden was specifically calculated for "cold" and "hot" CRC subgroups. We found that the TMB of the 'hot' CRC group was markedly higher than that of the 'cold' CRC group (Figure 3A). That is, 'hot' CRC cells have a higher mutational burden of somatic mutations and are more favourable to produce a great many neoantigens—which, in turn, facilitates responsiveness to immunotherapy. On the contrary, the 'cold' CRC tumour cells have lower TMB and lead to fewer neoantigens being produced, decreasing 'cold' CRC's ability to tolerate immunotherapy.

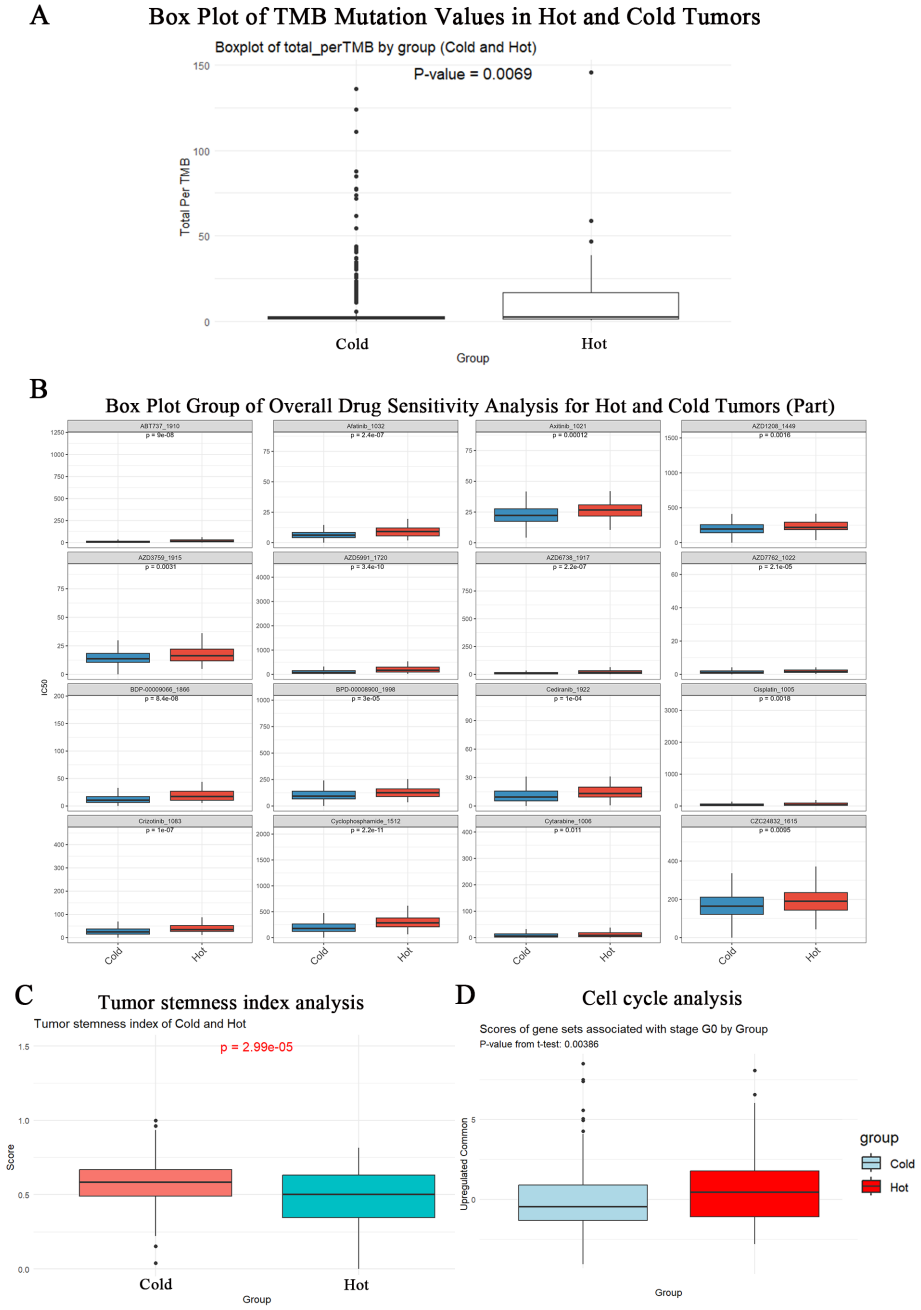

Figure 3. Analysis of Overall Differences in "cold" and "hot" Tumors. **A**, TMB in "cold" and "hot" tumors; **B**, predicted drug sensitivity (IC~50~) from 198 drugs in "cold" and "hot" tumors, used the oncoPredict package to estimate and compared by Wilcoxon rank sum test; **C**, the tumor stemness index (mRNAsi) in "cold" and "hot" tumors calculated by the OCLR algorithm; **D**, GSEA cell cycle analysis (G0 stage gene set) comparing "cold" and "hot" tumors.

3.5. Cold tumors exhibit diminished sensitivity to diverse anticancer drugs

Based on these data, we next analysed our "hot" and "cold" CRC phenotypes comprehensively to evaluate potential drug sensitivity difference between "hot" and "cold" CRC phenotypes. We carried out a pharmacology analysis on 198 compounds, summary of which is listed in Supplementary Tables. Statistical analysis revealed 133 compounds to exhibit significant differences of their sensitivity ($p < 0.05$). Among which, 64 compounds displayed a higher IC50 in cold tumors, including ABT737_1910, AZD1208_1449, AZD3759_1915 and AZD5991_1720 (Figure 3B, Supplementary Tables S1 and S2). We concluded that 'cold' CRC might be insensitive to a large number of anticancer drugs, which renders 'cold' CRC therapeutic challenges. Such in silico prediction, however, needs to be validated by experiments and should not be confused with the directly measured patient drug responses.

3.6. Cold CRC exhibits a significantly elevated stemness index

Regarding biological features of "cold" CRC and "hot" CRC, we did stemness index analysis. As shown in Figure 3C, "cold" CRCs have a much higher stemness index than that of "hot" tumors ($p < 0.05$), indicating that "cold" CRC has stronger ability to maintain self-renewal, which can promote DNA repairing function, restrain apoptotic response pathway and induce drug resistance proteins, promoting tumor cell proliferation and malignant metastasis.

3.7. Cold CRC tumors show enhanced proliferative potential

Cell cycle analysis also showed different proliferative states of "cold" and "hot" CRC (Figure 3D). "Hot" CRC almost completely upregulated gene markers of G0 phase; "cold" CRC almost completely downregulated gene markers of G0 phase. The results suggested that "cold" CRC tumor was mostly in an active state of proliferation, and "hot" CRC tumor was more likely to be in a state of proliferation arrest, possibly promoted by its immune-inflamed microenvironment and other tumor-intrinsic factors.

3.8. Enhanced oxidative phosphorylation pathway activity in cold CRC

Four distinct analytical methodologies (Wilcoxon rank-sum test, limma, DESeq2, and edgeR) were utilized to identify DEGs between "cold" and "hot" CRC tumor specimens. As illustrated in Figure 4A, a total of 295 genes exhibited significant upregulation in "cold" tumors relative to hot tumors. Subsequent KEGG pathway enrichment analysis, conducted via the Metascape database, demonstrated that those upregulated DEGs were predominantly enriched in pathways associated with OXPHOS, aerobic respiration, and related biological processes (Figure 4B), which indicated the increased OXPHOS pathway activity in "cold" CRC.

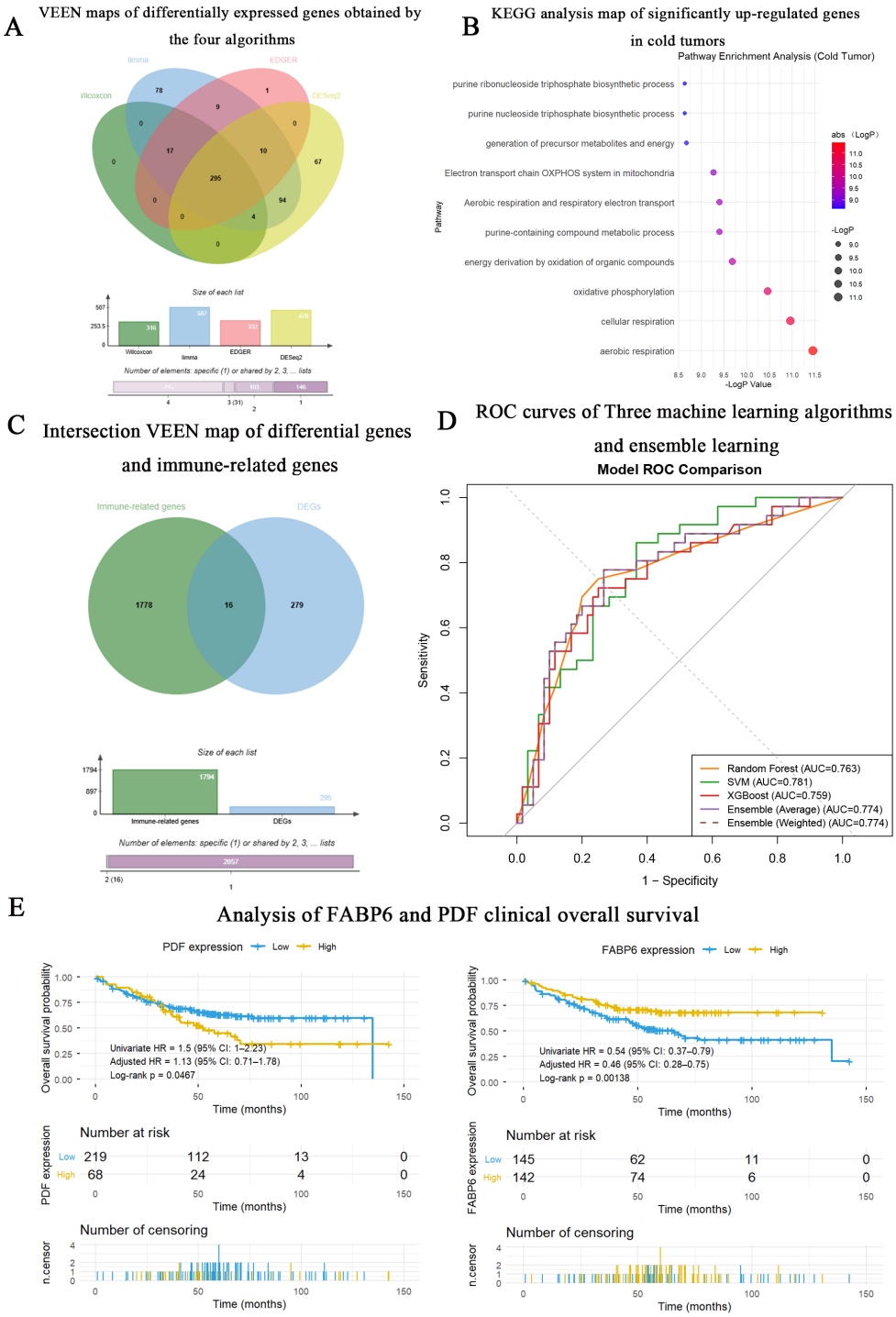

Figure 4. Identifying the key differentially expressed genes that distinguish between "cold" and "hot" tumors. **A**, Upregulated genes in "cold" *versus* "hot" tumors identified by Wilcoxon rank-sum test, limma, DESeq2, and edgeR (|log~2~FC| > 1, FDR < 0.05), with the intersection of four algorithms yielding 295 genes; **B**, KEGG pathway enrichment analysis of the 295 upregulated genes using Metascape; **C**, Intersection of the 295 upregulated genes with the ImmPort immune-related gene set, yielding 16 immune-related DEGs; **D**, Ensemble machine learning feature selection. Left: ROC curves of RF, SVM, XGBoost, and ensemble models in the test cohort. Right: Venn diagram showing the consensus of top 5 feature genes ranked by RF, SVM, and XGBoost, yielding three candidate genes (PDF, FABP6, and IGKJ4); **E**, Overall survival analysis of candidate genes in independent GEO cohorts (GSE17537 and GSE17538).

OXPHOS is a fundamental mitochondrial pathway for energy metabolism and maintains the energy homeostasis of cells by coupling the electron transportation process with the ATP synthesis process. The obvious enrichment of DEGs in OXPHOS pathways in "cold" CRC suggests that "cold" CRC has an increased energy need for "cold" CRC to grow, differentiate or adapt to stress. The connection of the high expression of OXPHOS-associated genes with poor overall patient survival, increased resistance of "cold" CRC to treatment, and increased EMT was also confirmed by Cemile Uslu *et al* [37]. Although pharmacological inhibition of OXPHOS has shown potential to restart the sensitivity of tumor cell to chemotherapy, radiotherapy and immunotherapy, this application is confronted with the obstacle of the importance of OXPHOS in normal cellular functioning, especially in tissues with high energy needs, such as cardiomyocytes and neurons. Moreover, the heterogeneity of OXPHOS activities among different tumors and different regions of the same tumor will prevent a therapeutic strategy being effective across all samples. Hence, pinpointing the most influential regulatory genes within the OXPHOS module will be crucial for the development of precision therapeutic treatments that target and exploit metabolic vulnerabilities, but that avoid off-target effects.

3.9. PDF, a key protein in the OXPHOS pathway, is a candidate biomarker of the "cold" CRC

To get the immune-related DEGs, the DEGs between "cold" and "hot" CRC were further intersected with the immune-gene set from the Immport database, yielding 16 immune-related DEGs, including LCN12, UMODL1, FABP6, RBP2, PDF, and others (Figure 4C). Subsequently, we performed model construction and feature selection on these genes using three machine learning algorithms (SVM, RF, and XGBoost) combined with ensemble learning strategies, with the corresponding results presented in Figure 3D. The AUC values of the RF, XGBoost, and SVM models were 0.763, 0.759 and 0.781, respectively, while both voting-based ensemble learning models yielded an identical AUC of 0.774. Overall, the established machine learning models exhibited favorable predictive performance, enabling accurate discrimination between cold and hot tumors based on the expression profiles of these genes. On the held-out internal test cohort (20% split), the optimal AUC-weighted ensemble model achieved an accuracy of 0.750, a sensitivity of 0.639, a specificity of 0.817, a precision (positive predictive value) of 0.676, an F1-score of 0.657, and a balanced accuracy of 0.728 (Kappa = 0.461). The confusion matrix of this ensemble model on the test set is summarized in Supplementary Table S3 (Cold correctly predicted: 49; Hot correctly predicted: 23; misclassified Cold as Hot: 11; misclassified Hot as Cold: 13).

Three machine learning algorithms exhibited distinct gene ranking results based on feature importance and classification weights. The top five genes prioritized by each algorithm were as follows: PDF, FABP6, MIF, IGKJ4, and RBP2 for RF; IGKJ4, IGKJ1, PDF, CALCA, and FABP6 for SVM; and PDF, FABP6, IGKJ4, MIF, and LCN12 for XGBoost. In the AUC-weighted ensemble model, the three base models showed comparable classification contributions, with contribution weights of 0.331 (RF), 0.339 (SVM), and 0.330 (XGBoost), respectively. Subsequently, we performed intersection analysis on the top 5 ranked genes from each algorithm, which ultimately identified three core feature genes: PDF (peptide deformylase), FABP6 (fatty acid binding protein 6), and IGKJ4 (immunoglobulin kappa joining 4) (as shown in Supplementary Table S3).

To evaluate the clinical relevance of the three consensus genes, OS analysis was performed on the merged independent GEO cohorts (GSE17538 and GSE17537; n = 287) (in Figure 4E). The optimal expression cutoff for each gene was determined by maximally selected rank statistics. The PH

assumption was satisfied for all models (Schoenfeld residual test, $P > 0.05$). For PDF, the optimal cutoff stratified patients into low ($n = 219$) and high ($n = 68$) expression groups. High PDF expression was significantly associated with poor overall survival (log-rank $P = 0.0467$; univariate HR = 1.50, 95% CI: 1.00–2.23). However, after adjusting for age, gender, AJCC stage, tumor grade, and stratifying by dataset, the association was attenuated and no longer statistically significant (adjusted HR = 1.13, 95% CI: 0.71–1.78), suggesting that the prognostic effect of PDF may be partially confounded by clinical covariates. For FABP6, the optimal cutoff stratified patients into low (n = 145) and high (n = 142) expression group. High FABP6 expression predicted significant favourable survival (log-rank P = 0.00138; univariate HR = 0.54, 95% CI: 0.37–0.79), and this favourable impact persisted after adjusting for other variables (adjusted HR = 0.46, 95% CI: 0.28–0.75). IGKJ4 was excluded from survival analysis because no matching probes were available in the queried GEO datasets. Overall, these results identify PDF as a candidate biomarker associated with poor prognosis in CRC and that FABP6 has an unexpectedly good prognostic impact, given its cold-tumour enrichment.

PDF is a very important enzyme in prokaryotic post-translational protein modification and catalyses the removal of the N-terminal formyl group of the nascent polypeptides. For eukaryotic cells, N-formylation of the initiating methionine (fMet) only exists in the nascent proteins in mitochondria. As the initiation amino acid for protein translation in prokaryotes and eukoterotic mitochondria, removal of N-formylated methionine (fMet) by the PDF is an important part of post-translational modification. In the eukaryotic mitochondrial matrix, mitochondrial methionyl-tRNA formyltransferase catalyses formylation of methionine (Met) to get fMet-tRNA [38], which is necessary for mitochondrial translation. PDF can specifically recognize the N-terminal fMet on nascent peptides and hydrolyze the formyl group in its own zinc ion (Zn^{2+})-coordinated active site to get N-Met-modified peptides [39]. As well as for many other mitochondrial proteins, the N-terminal methionine (N-Met) of certain mitochondrial proteins is further hydrolysed by the mitochondrial methionine aminopeptidase (MetAP), to form mature N-terminal peptides for proper folding, localization and function of the proteins in mitochondria. They are mostly subunits of the mitochondrial ETC complexes (for example, components of complexes I, III and IV). Their de-formylation and proper maturation are important for the correct assembly of respiratory chain complexes and proper oxidative phosphorylation.

In “cold” CRC, increased PDF gene expression is observed to be associated with OXPHOS pathway activation, suggesting that PDF overexpression may enhance OXPHOS activity in “cold” CRC by promoting the hydrolysis of fMet, the maturation of mitochondrial nascent proteins in ECT. Thus, PDF is inferred to promote energy generation that sustains the rapid proliferation of ‘cold’ CRC—a role consistent with observations in other malignancies in which PDF inhibition lowers oxygen consumption and ATP production [40]. In addition, the PDF-mediated cleavage of formylmethionine (fMet) from mitochondrial nascent peptides is hypothesized to contribute to the regulation of tumor immunity. The mitochondrial N-fMet-peptide, as a potent “non-self” cue [41], is a damage-associated molecular pattern (DAMP) that promotes chemotaxis of CD8+ T cells and macrophages via binding to its receptor, formyl peptide receptor 1 (FPR1). FPR1 is constitutively expressed not only on myeloid immune cells (neutrophils, monocytes/macrophages, dendritic cells) but also on epithelial, fibroblast, and neuronal populations; therefore, thereby the fMet-FPR1signaling pathway functions as a central player to transduce mitochondrial damage signaling into immune activation [41]. By hydrolyzing fMet, elevated PDF expression may potentially attenuate immune cell recruitment, thereby

reinforcing the immune-excluded or immune-desert state of "cold" tumors. These proposed functions are derived from literature-based inference and require direct experimental validation in CRC models.

3.10. PDF expression correlates with an immune-inhibitory TME and elevated tumor proliferative capacity

The single-gene immune infiltration analysis was performed to evaluate the association between PDF expression and tumor immune infiltration. As demonstrated in Figure 5A, the expression of PDF was significantly negatively correlated with tumor purity and the infiltration of immune cells, including CD8+ T cells, macrophages, and neutrophils. These results suggest that PDF expression is linked to tumor immune modulation and promotes the establishment of a suppressive tumor immune microenvironment.

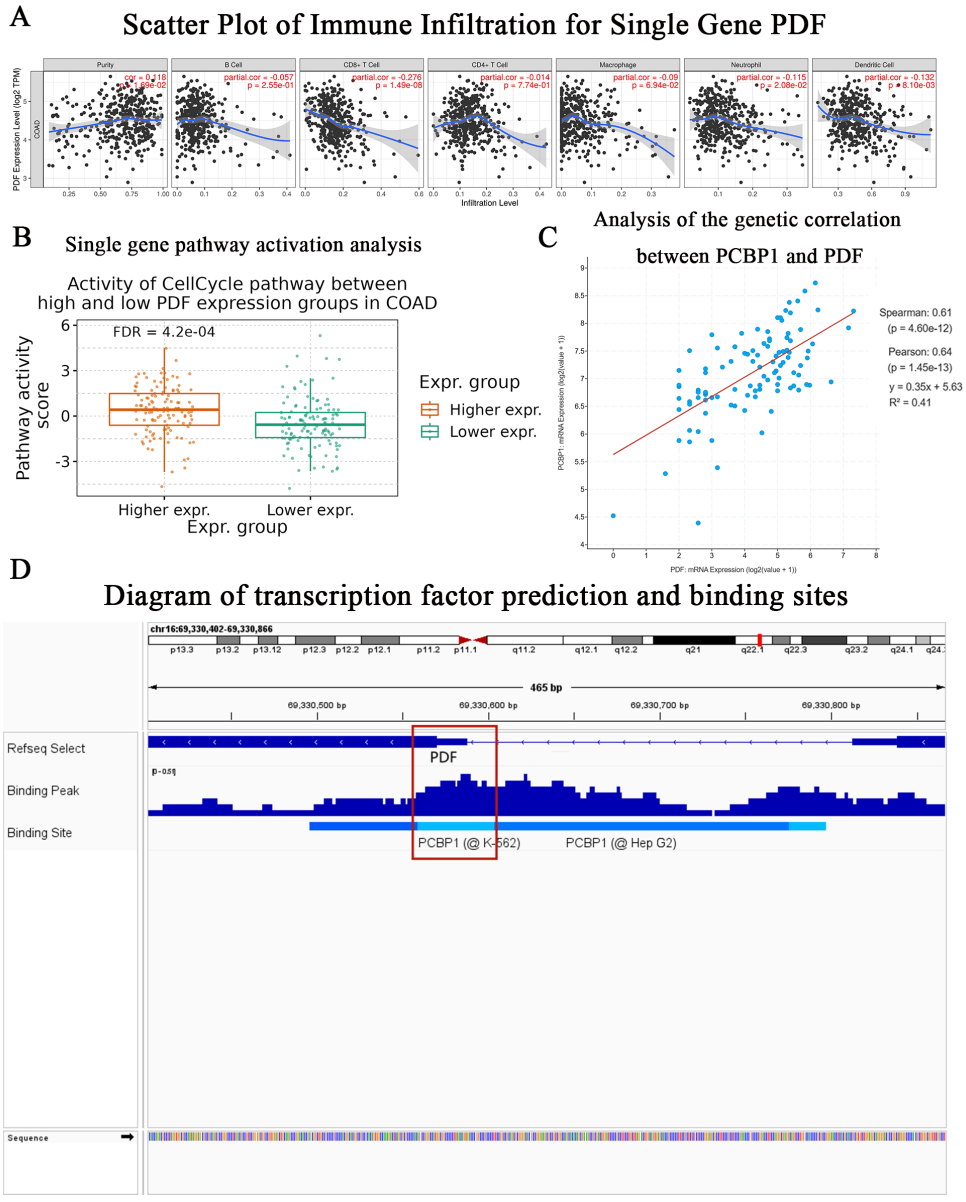

Figure 5. Single-gene analysis of PDF and its transcription factor analysis. **A**, Correlation between PDF expression and immune infiltration levels (CD8+ T cells, macrophages, neutrophils, and tumor purity) via TIMER1.0; **B**, Pathway activation analysis of PDF using the GSCA database; **C**, PDF and PCBP1 gene correlation analysis, spearman = 0.61 (p = 4.60e − 12), pearson = 0.64(1.45e − 13); **D**, ChIP-Atlas peak browser showing PCBP1 binding enrichment within the PDF promoter region (chr16:69,330,495–69,330,730). The y-axis indicates fold enrichment (FE).

Subsequently, the GSCA database was used to perform single-gene pathway activation analysis for the PDF gene. As presented in Figure 5B, the high PDF expression was significantly correlated with the activation of cell cycle-related signaling pathways, suggesting that the increased PDF expression in cold tumors could enhance the tumor proliferative potential.

3.11. PCBP1 is a potential transcription factor for PDF

The transcription factor (TF) enrichment analysis module of the ChIP-Atlas database was used to find the potential TFs regulating the expression of PDF. The resulting candidate TFs were further intersected with the upregulated DEGs in the "cold" CRC tumor. Through those analyses, we identified three potential transcription factors with significant enrichment: PCBP1 (Fold Enrichment (FE) = 4.90328), ZBTB20 (FE = 2.96244), and HOXA4 (FE = 1.93191). PCBP1 was selected as the most promising candidate TF regulating PDF, due to its highest enrichment score. Then, we used the Peak Browser function of ChIP-Atlas database to explore the binding of PCBP1 to PDF promoter. As shown in Figure 5D, PCBP1 binding enrichment around PDF promoter region (chr16:69330495-69330730) of PDF gene was obviously high. Next, we used the cBioPortal database to detect the expression correlation between PCBP1 and PDF. As shown in Figure 5C, the expression of PCBP1 was strongly and positively associated with the expression of PDF. Together, these bioinformatics results nominate PCBP1 as a candidate transcriptional regulator for PDF, but they are not proof of a direct regulatory interaction between the two. Such an interaction would need to be verified in experimental studies—such as by using ChIP-qPCR or a promoter-reporter assay.

Polypyrimidine tract binding protein 1 (PCBP1) is a versatile RNA-binding protein involved in gene transcription, RNA processing, stability and translation control. In head and neck cancer, the expression of PCBP1 was found to have a negative correlation to the rate of survival [42], implying the potential of PCBP1 to promote tumour growth. PCBP1 protein has three conserved K homology domains involved in its specific binding to RNA. To perform its function, PCBP1 needs to interact with other RBP (RNA-binding proteins) or proteins [43]. By using the BioGRID database [44], experimentally annotated 106 PCBP1 interaction proteins can be obtained. Further intersecting these proteins with the upregulated DEGs in the "cold" CRC tumour, we finally inferred that TAF10 (TBP-associated factor 10), a core subunit of the TFIID, a transcription initiation complex, is the potential partner of PCBP1 in the transcriptionally transcriptional expression of PDF. On a transcriptional stimulatory level, TAF10 plays an important role in governing the activation of a broad spectrum of gene expression programmes by participating in the promoter recognition and transcription complex assembly [45]. This in silico intersection gives rise to the speculative idea that a PCBP1–TAF10 complex might regulate PDF transcription, yet this remains completely speculative in CRC cells and even lacks direct supporting evidence.

3.12. Validation confirms PDF as a biomarker of cold CRC

To independently confirm the classification scheme, the colon cancer dataset from the CPTAC-2 prospective cohort (Cell, 2019) and an independent dataset (GSE220148) were downloaded from cBioPortal and GEO database, respectively. A total of 213 samples were used. The samples were classified into cold tumors and hot tumors as 145 and 67 samples by the previously built integrated learning model, respectively. A total of 213 samples were analyzed, among which 145 and 67 were

classified as cold and hot tumors, respectively, using the previously constructed integrated learning model. Immune infiltration analysis by using the ESTIMATE algorithm revealed that hot tumors exhibited significantly higher immune, stromal, and ESTIMATE scores, alongside correspondingly lower tumor purity, relative to cold tumors (all $P < 0.05$) (Figure 6A). Differential expression analysis between the two phenotypes using limma revealed significant upregulation of PDF and PCBP1 in cold tumors (Figure 6B). These independent validation findings further corroborate PDF as a robust candidate biomarker for differentiating cold and hot tumor microenvironment phenotypes in colorectal cancer.

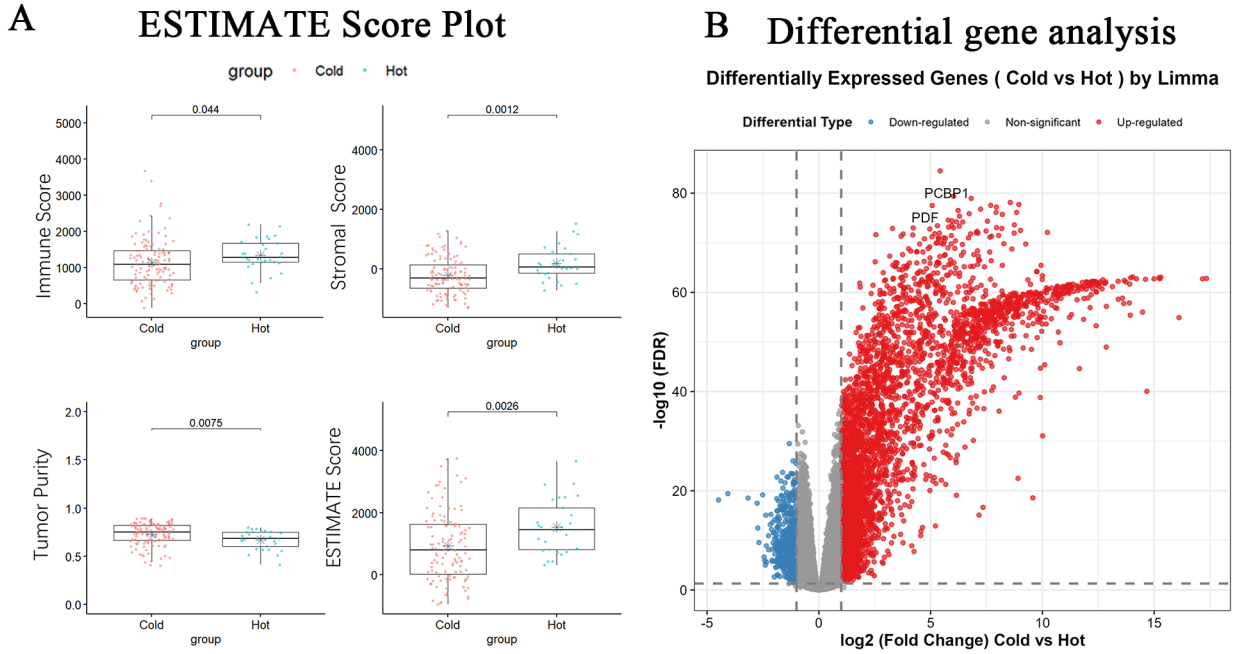

Figure 6. The validation dataset was used to verify the validity of PDF as a key signature gene for "cold" and "hot" tumors. **A**, Use the existing machine learning model to classify the validation set data, and then use the estimate algorithm to conduct immune infiltration analysis on the classification results; **B**, For the pre-grouped validation set of cold and hot tumors, differential gene analysis was conducted using limma to identify whether there were key candidate factors with differential expression after grouping.

4. Discussion and conclusion

This study identifies Peptide Deformylase (PDF) as a candidate biomarker for the immune "cold" phenotype CRC. Specifically, by hydrolyzing mitochondrial N-formylmethionine (fMet), PDF is proposed to exert a dual role in "cold" tumor: (1) PDF may promote the maturation of proteins in ETC. (2) PDF may constrain immune cell recruitment, reinforcing the immune-excluded or immune-desert state of "cold" CRC by reducing the N-fMet-peptide (one of DAMP) release to TME. Furthermore, we identified PCBP1 as a potential transcriptional regulator of PDF.

Our bioinformatic analyses identified PCBP1 as a candidate upstream regulator of PDF based on ChIP-Atlas peak enrichment and expression correlation in public cohorts. While published studies have reported transcription-related roles for PCBP1 in other cellular contexts—such as coactivation of the eIF4E [46] and mu opioid receptor promoters, regulation of CDKN1A via promoter occupancy [47], and upregulation of LIFR via proximal promoter luciferase assays [48]—these findings were not obtained in CRC and do not extend to PDF. Similarly, although TAF10 is a known TFIID/SAGA

subunit that interacts with transcriptional activators such as GATA1 [49], no evidence links it to PDF regulation in colorectal cancer. Therefore, we highlight that our computational suggestion for PCBP1–TAF10–PDF transcriptional cascade is purely a computational hypothesis. Direct experimental evidence—ChIP-qPCR or ChIP-seq for occupancy at the PDF promoter, dual-luciferase reporter assays in CRC cell lines and loss-/gain-of-function perturbation studies—is necessary to demonstrate any transcriptional regulatory relationship. Until such evidence is available, the PCBP1–TAF10 interaction should simply be regarded as an assumption or hypothesis to suggest future studies and not as a validated function.

To establish robust signature genes, we used machine learning algorithms, including Random Forest, Support Vector Machine and XGBoost algorithms, which can process data of a large dimension with the whole genome and can take nonlinear relations into consideration, and overcome the limit of conventional univariate analysis. This strategy established PDF as a gold standard for discriminating “cold” tumours and “hot” tumours.

However, this study also had some limitations. First, our analysis was mainly based on bulk RNA-seq data from the TCGA-COAD cohort. Although bulk RNA-seq data yield important genome-wide expression profiles, the absence of single-cell RNA-seq (scRNA-seq) data restricts our ability to deeply analyze the heterogeneity of the immune cell subsets in the tumor immune microenvironment, as well as reflect detailed cell–cell communications. Second, this study had limitations of data heterogeneity and interpretability of the model. The TCGA-COAD cohort included multi-centre data; despite TPM normalization and four DEG algorithms applied to minimize technical differences, residual technical differences might influence subtype classification. The independent validation of PDF and PCBP1 in the CPTAC-2 and GSE220148 data sets also eliminated this problem, though formal batch correction (for example, ComBat) will be needed in future multi-cohort analysis. Additionally, although RF, SVM and XGBoost were good for identifying PDF as a robust biomarker, they are black-box models and have inherent weaknesses. To improve model interpretation, we tried to limit features to immune-related DEGs and to follow the cross-algorithm consensus; we did not give instance-level explanations (such as SHAP values). In future studies, incorporate explainable AI methods and prospective experimental validation to determine causality between the PCBP1–PDF–OXPHOS axis and the immune-cold phenotype. Third, our bioinformatic study identified PDF as a biomarker candidate, but the physiological roles of PDF in mitochondrial ETC maturation, ATP production and fMet-peptide-inducible immune responses were not directly experimentally tested in CRC cells or tissues. That is, PDF knockdown, OXPHOS assays, ATP assays, fMet-peptide quantifications, fMet-peptide-related FPR1 ligand assays and immune chemotaxis assays are missing to test these mechanisms. Finally, how PDF works mechanistically with other immunosuppressive factors has never been experimentally validated. Lastly, clinical information in public databases is typically incomplete; some of the samples lack detailed treatment history, the precise pathological stage or long-term clinic information, and this may affect the precision of prognostic analysis involving PDF and confound attempts to estimate its direct association with the treatment effect. Moreover, the predicted transcriptional regulation of PDF by PCBP1 (and the potential collaboration with TAF10) is based only on in silico ChIP-Atlas peak information and correlation with expression. These analyses cannot make direct connections to transcriptional causality, and the inferred PCBP1–TAF10 transcription complex remains experimentally unsubstantiated in CRC.

In summary, our findings elucidate a molecular framework centered on the PCBP1–PDF–OXPHOS axis underlying the immune "cold" phenotype in CRC. Future work should integrate single-cell technologies, experimental validation of mechanistic links, and further exploration of the PDF-associated network to advance precision immunotherapy for CRC patients refractory to current treatments.

Data availability statement

The supplementary Table accompanying the article contains the original data after the drug sensitivity analysis (Table S1), Grouped drug sensitivity analysis data and their IC50 preferences (Table S2) and Weighted Integrated Model Performance Summary (Table S3). The data used in this study can be obtained from the following sources: TCGA-COAD transcriptome data (https://portal.gdc.cancer.gov/projects/TCGA-COAD), GSE220148 transcriptome data (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE220148), GSE17538 transcriptome and survival data (https://www.ncbi.nlm.nih gov/geo/query/acc.cgi?acc=GSE17538), GSE17537 transcriptome and survival data (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17537).

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools only to improve language and readability. Specifically, the authors used DeepSeek for language polishing. The authors take full responsibility for the content of the manuscript.

Acknowledgments

This work is supported by the Guangdong Pharmaceutical University Innovation Training Program for Undergraduates: S202610573019; G202604302007; G202604302009. We thank all members of Zhang's lab for their advice and support in preparing this work.

Authors' contribution

Visualization, software, methodology, investigation, formal analysis: H.C.; validation, investigation: Z.Z.; validation, methodology, investigation: J.L.; writing—review and editing, writing—original draft, visualization, supervision, project administration, investigation, conceptualization and funding acquisition: X.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

[1] Matteucci L, Bittoni A, Gallo G, Ridolfi L, Passardi A. Immunocheckpoint inhibitors in microsatellite-stable or proficient mismatch repair metastatic colorectal cancer: are we entering a new era? *Cancers* 2023, 15(21):5189.

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 2021, 71(3):209–249.

[3] Heregger R, Huemer F, Steiner M, Gonzalez-Martinez A, Greil R, *et al*. Unraveling resistance to immunotherapy in msi-high colorectal cancer. *Cancers* 2023, 15(20):5090.

[4] Sur D, Gorzo A, Burz C. Immunotherapy for colorectal cancer in the era of precision medicine. In *Recent Understanding of Colorectal Cancer Treatment*. London: IntechOpen, 2022.

[5] Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, *et al*. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 2018, 24(5):541–550.

[6] Ding K, Mou P, Wang Z, Liu S, Liu J, *et al*. The next bastion to be conquered in immunotherapy: microsatellite stable colorectal cancer. *Front. Immunol*. 2023, 14:1298524.

[7] Kong X, Ou S, Wei Z, Ye X, Chen S, *et al*. Transforming the “cold” tumors to “hot” tumors: strategies for immune activation. *Biochem. Pharmacol.* 2025, 241:117194.

[8] Koustas E, Trifylli EM, Oraiopoulos V, Karamouzis MV, Sarantis P. Microsatellite phenotype as a guide for immunotherapy in colorectal cancer: current status and future perspectives. *Genes* 2026, 17(6):674.

[9] Riley JM, Cross AW, Paulos CM, Rubinstein MP, Wrangle J, *et al*. The clinical implications of immunogenomics in colorectal cancer: A path for precision medicine. *Cancer* 2018, 124(8):1650–1659.

[10] Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010, 26(12):1572–1573.

[11] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, *et al*. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 2013, 4:2612.

[12] Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, *et al*. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016, 17(1):218.

[13] Zeng D, Fang Y, Qiu W, Luo P, Wang S, *et al*. Enhancing immuno-oncology investigations through multidimensional decoding of tumor microenvironment with IOBR 2.0. *Cell Rep. Methods* 2024, 4(12):100910.

[14] Li B, Severson E, Pignon JC, Zhao H, Li T, *et al*. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016, 17(1):174.

[15] Mayakonda A, Lin D, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018, 28(11):1747–1756.

[16] Maeser D, Gruener RF, Huang RS. OncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Briefings Bioinf.* 2021, 22(6):bbab260.

[17] Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, *et al*. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 2018, 173(2):338–354.e15.

[18] Wiecek AJ, Cutty SJ, Kornai D, Parreno-Centeno M, Gourmet LE, *et al*. Genomic hallmarks and therapeutic implications of G0 cell cycle arrest in cancer. *Genome Biol.* 2023, 24(1):128.

[19] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 2005, 102(43):15545–15550.

[20] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014, 15(12):550.
[21] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 2004, 3(1):3.
[22] Robinson MD, Mccarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, 26(1):139–140.
[23] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000, 28(1):27–30.
[24] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 2019, 10(1):1523.
[25] Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, *et al*. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci. Data* 2018, 5:180015.
[26] Breiman L. Random forests. *Mach. Learn.* 2001, 45(1):5–32.
[27] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell. Syst. Appl.* 1998, 13(4):18–28.
[28] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco, USA, August 13–17, 2016, pp. 785–794.
[29] Rigatti SJ. Random forest. *J. Insur. Med.* 2017, 47(1):31–39.
[30] Meyer D. Support vector machines: the interface to libsvm in package e1071. 2001. Available: https://www.semanticscholar.org/paper/Support-Vector-Machines%3A-The-Interface-to-libsvm-in/1c74a17ecead219310059af7e6520d1cab7291dd#extracted (accessed on 18 July 2026).
[31] Li T, Fan J, Wang B, Traugh N, Chen Q, *et al*. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* 2017, 77(21):e108–e110.
[32] Liu C, Hu F, Xie G, Miao Y, Li X, *et al*. GSCA: an integrated platform for gene set cancer analysis at genomic, pharmacogenomic and immunogenomic levels. *Briefings Bioinf.* 2023, 24(1):bbac558.
[33] Zou Z, Ohta T, Oki S. ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. *Nucleic Acids Res.* 2024, 52(W1):W45–W53.
[34] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, *et al*. Integrative genomics viewer. *Nat. Bio.* 2011, 29(1):24–26.
[35] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* 2012, 2(5):401–404.
[36] Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, *et al.* Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 2019, 177(4):1035–1049.e19.
[37] Uslu C, Kapan E, Lyakhovich A. Cancer resistance and metastasis are maintained through oxidative phosphorylation. *Cancer Lett.* 2024, 587:216705.
[38] Kim D, Seok OH, Ju S, Kim SY, Kim JM. *et al*. Detection of Nα-terminally formylated native proteins by a pan-N-formyl methionine-specific antibody. *J. Biol. Chem.* 2023, 299(5):104652.
[39] Hu L, Cai X, Dong S, Zhen Y, Hu J, *et al*. Synthesis and anticancer activity of novel actinonin derivatives as HsPDF inhibitors. *J. Med. Chem.* 2020, 63(13):6959–6978.

[40] Randhawa H, Chikara S, Gehring D, Yildirim T, Menon J, *et al*. Overexpression of peptide deformylase in breast, colon, and lung cancers. *BMC Cancer* 2013, 13:321.

[41] Marchi S, Guilbaud E, Tait SW, Yamazaki T, Galluzzi L. Mitochondrial control of inflammation. *Nat. Rev. Immunol.* 2023, 23(3):159–173.

[42] Lee J, You JH, Roh JL. Poly(rC)-binding protein 1 represses ferritinophagy-mediated ferroptosis in head and neck cancer. *Redox Biol.* 2022, 51:102276.

[43] Rithvik A, Wadhavane S, Rasool M. Decoding poly (RC)-binding protein 1 (PCBP1), the underrated guard at the foothill of ferroptosis. *Pathol. Res. Pract.* 2025, 266:155771.

[44] Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, *et al*. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 2021, 30(1):187–200.

[45] Luna-Arias JP, Castro-Muñozledo F. Participation of the TBP-associated factors (TAFs) in cell differentiation. *J. Cell. Physiol.* 2024, 239(2):e31167.

[46] Choi HS, Hwang CK, Song KY, Law PY, Wei LN, *et al*. Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochem. Biophys. Res. Commun.* 2009, 380(3):431–436.

[47] Karam JA, Fréreux C, Mohanty BK, Law PY, Wei LN, *et al*. The RNA-binding protein PCBP1 modulates transcription by recruiting the G-quadruplex-specific helicase DHX9. *J. Biol. Chem.* 2024, 300(11):107830.

[48] Streitfeld WS, Dalton AC, Howley BV, Howe PH. PCBP1 regulates LIFR through FAM3C to maintain breast cancer stem cell self-renewal and invasiveness. *Cancer Biol. Ther.* 24(1):2271638.

[49] Papadopoulos P, Gutiérrez L, Demmers J, Scheer E, Pourfarzad F, *et al*. TAF10 interacts with the GATA1 transcription factor and controls mouse erythropoiesis. *Mol. Cell. Biol.* 2015, 35(12): 2103–2118.