

Integrative transcriptomic and machine learning analyses identify ABCD1 as a diagnostic and prognostic biomarker in cutaneous melanoma

Hongyang Zhang

General Hospital of Ningxia Medical University

Yang Rui

Ningxia Medical University

Huan Liu

Ningxia Medical University

Le Wang

General Hospital of Ningxia Medical University

Jing Yang

General Hospital of Ningxia Medical University

Jinping Li

General Hospital of Ningxia Medical University

Nan Yu

13895095009@163.com

General Hospital of Ningxia Medical University

Research Article

Keywords: Skin malignant melanoma (SKCM), Multi-omics integration, Machine learning, ABCD1, Biomarker

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11115022/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

1 **Integrative transcriptomic and machine learning analyses identify ABCD1 as a**
2 **diagnostic and prognostic biomarker in cutaneous melanoma**

3 Hongyang Zhang^{1 †}, Rui Yang^{2 †}, Huan Liu², Le Wang¹, Jing Yang¹, Jinping Li^{3*}, Nan Yu^{1*}

4 1. Department of Dermatology, General Hospital of Ningxia Medical University, Yinchuan,
5 Ningxia 750004, China.

6 2. The first clinical medical college, Ningxia Medical University, Yinchuan, Ningxia 750004,
7 China;

8 3. Department of Surgical Oncology, Ningxia Medical University, Yinchuan, Ningxia
9 750004, China;

10 † Hongyang Zhang and Rui Yang contributed equally to this work and share first authorship

11 * Jinping Li and Nan Yu contributed equally to this work and share senior authorship

12 *Correspondence: Nan Yu, 13895095009@163.com; Jinping Li, 13995105618@163.com

13
14 **Abstract**

15 **Background:** Skin cutaneous melanoma (SKCM) is highly aggressive and heterogeneous, and
16 robust biomarkers remain limited. We integrated bulk and single-cell transcriptomic data with
17 machine learning to identify key driver genes and prognostic markers in SKCM.

18 **Methods:** This study integrated transcriptomic data from two GEO cohorts, GSE15605 and
19 GSE46517, together with the TCGA-SKCM cohort to identify consistently differentially
20 expressed genes. Four machine learning algorithms, including LASSO, SVM-RFE, XGBoost,
21 and Boruta, were then applied to systematically screen core candidate genes. Publicly available
22 single-cell RNA-seq data from GSE125121 were further analyzed to determine the cellular
23 origin of these core genes. Survival analysis and clinicopathological correlation analysis were
24 performed to evaluate the clinical relevance and potential therapeutic value of ABCD1. Finally,
25 the protein expression of ATP-binding cassette subfamily D member 1 (ABCD1) was validated
26 by Western blot.

27 **Results:** We identified 866 consistently differentially expressed genes and five robust
28 candidates: ABCD1, KIF3C, CNNM1, TRIB1, and NT5DC3. Single-cell analysis showed that
29 these genes were mainly enriched in malignant melanoma cells and displayed marked
30 intratumoral heterogeneity. Among them, ABCD1 exhibited the broadest and highest
31 expression across malignant subpopulations. Pseudotime analysis indicated
32 that ABCD1 and NT5DC3 were predominantly associated with terminal malignant states,

whereas TRIB1 spanned transitional and terminal states. In TCGA-SKCM, high ABCD1 expression was significantly associated with worse overall survival, while the other genes showed limited prognostic value. Western blot further showed that ABCD1 protein expression was low in normal skin, nevus, and stage I SKCM tissues, but markedly increased in stage II–III SKCM tissues

Conclusions: ABCD1 is a promising prognostic biomarker and potential therapeutic target in SKCM. This multi-omics framework provides a robust strategy for biomarker discovery in melanoma.

Keywords: Skin malignant melanoma (SKCM), Multi-omics integration, Machine learning, ABCD1, Biomarker

1 Introduction

Skin cutaneous melanoma (SKCM) is a highly invasive cutaneous malignancy originating from melanocytes[1, 2]. Although its incidence is lower than that of many common solid tumors, melanoma remains one of the leading causes of skin cancer-related mortality[3]. According to GLOBOCAN 2022 estimates, approximately 331,722 new cases of melanoma and 58,667 melanoma-related deaths occur worldwide each year[4]. The disease burden is particularly pronounced in Australia, New Zealand, North America, and several European regions, and its incidence continues to increase in certain countries and populations[5]. In recent years, advances in surgical resection, immunotherapy, and targeted therapy have substantially improved clinical outcomes for a subset of patients. However, advanced or metastatic melanoma remains highly aggressive and is characterized by a high risk of recurrence and metastasis, as well as marked heterogeneity in therapeutic response. Currently, clinical assessment of SKCM mainly relies on histopathological features and the American Joint Committee on TNM staging system. Nevertheless, these conventional approaches have limitations in accurately predicting prognosis, evaluating therapeutic benefit, and guiding individualized management. Therefore, the identification of molecular biomarkers with high sensitivity, specificity, and reproducibility is of great significance for improving early detection, risk stratification, prognostic evaluation, and precision treatment of SKCM.

With the widespread application of high-throughput sequencing technologies, research into the molecular mechanisms underlying tumor development and progression has advanced substantially[6]. Public databases, such as the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), have accumulated extensive transcriptomic datasets, providing valuable resources for cross-cohort integrative analyses and improving the robustness of research findings[7]. However, bulk transcriptomic analyses remain limited by technical noise, platform-related bias, and the inability to resolve cellular heterogeneity, as bulk samples

represent averaged signals from malignant, stromal, and immune components. Although previous SKCM studies have made important progress, many rely on single datasets or single machine learning algorithms, which may reduce reproducibility and increase the risk of model bias or overfitting. More importantly, candidate biomarkers identified from bulk data often lack clear information regarding their cellular origins and spatial distribution within the tumor microenvironment. This gap limits mechanistic interpretation and weakens their potential for clinical translation.

Considering the multi-level limitations, this study established a systematic multi-cohort analytical framework to identify and validate molecular biomarkers with both diagnostic and prognostic potential in SKCM. First, transcriptomic data from multiple public databases were integrated, and four complementary machine learning algorithms were jointly applied to perform robust screening of core candidate genes. To further address the limited cellular resolution of bulk transcriptomics, single-cell transcriptomic analysis was incorporated to characterize the cellular origins and expression heterogeneity of candidate genes[8]. The overall workflow is shown in Fig.1. By integrating computational screening with single-cell evidence, this study aims to identify reliable candidate biomarkers and provide insights into their cellular characteristics and clinical relevance in SKCM.

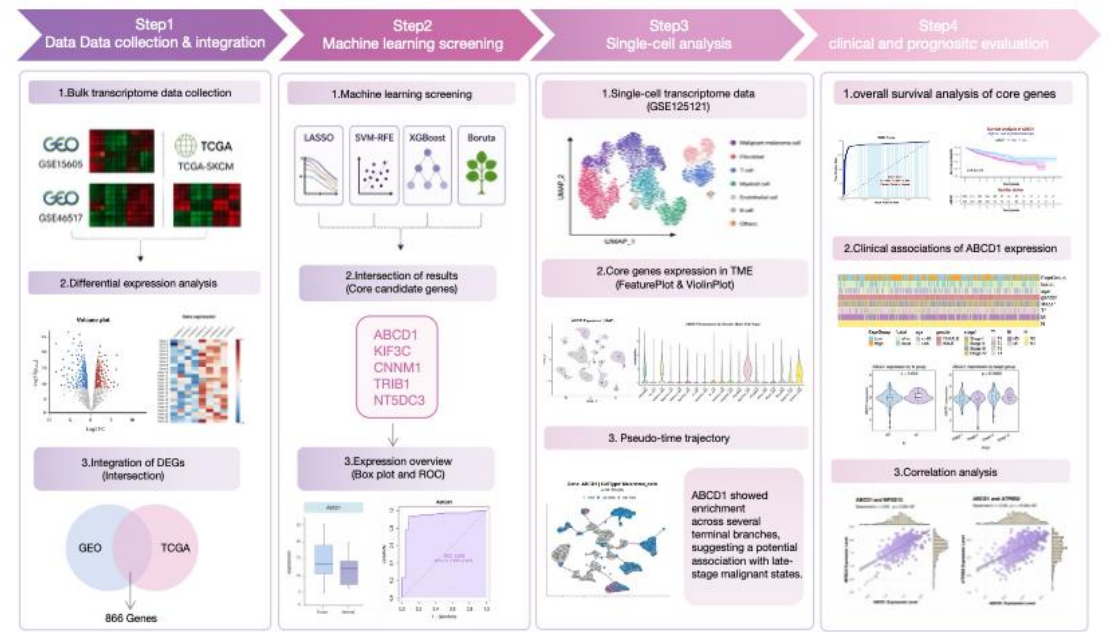


Fig.1 Study design overview.

2 Materials and methods

2.1 Data acquisition and preprocessing

Two GEO microarray cohorts, GSE15605 and GSE46517, were used as the training datasets, while GSE114445 served as an independent validation cohort. RNA-seq data from TCGA-

SKCM (<https://portal.gdc.cancer.gov/>) and normal skin samples from GTEx (<https://gtexportal.org/>) were obtained from the UCSC Xena platform. All analyses were performed in R. GEO datasets were downloaded using the GEOquery package, and probe IDs were annotated according to the corresponding GPL platform files. When multiple probes mapped to the same gene symbol, their average expression value was used to generate gene-level expression matrices. TCGA and GTEx expression data were transformed as $\log_2(\text{FPKM} + 1)$ and normalized using the `normalizeBetweenArrays` function in the `limma` package to reduce technical variability across datasets.

After data preprocessing, tumor and normal expression matrices were extracted according to sample annotations, and design and contrast matrices were constructed. Differential expression analysis was performed using the `limma` package, with linear model fitting by `lmFit` and empirical Bayes moderation by `eBayes`. DEGs were defined as genes with $|\log_2\text{FC}| > 1$ and adjusted $P < 0.05$. Heatmaps and volcano plots were generated using `pheatmap` and `ggplot2`, respectively. The mean expression level, fold change, and standard error of each DEG were further calculated for downstream functional annotation and predictive model construction.

2.2 Batch effect correction and data integration

To integrate the GEO datasets GSE15605 and GSE46517, the corresponding expression matrices were merged based on shared gene symbols. Considering that differences in data sources and detection platforms may introduce systematic batch effects, batch correction was performed using the `removeBatchEffect` function in the `limma` package, in conjunction with `sva`-related procedures. Batch information was extracted from the original dataset annotations and incorporated into the model as a batch factor to minimize non-biological variability arising from technical differences. To assess the effectiveness of batch correction, boxplots and principal component analysis (PCA) scatter plots were generated before and after correction. Boxplots were used to visualize the distribution of overall expression values across samples, thereby evaluating normalization consistency. PCA plots based on the first two principal components were employed to examine sample clustering patterns, providing a visual assessment of batch effects and their correction.

2.3 Differential expression analysis

For the GSE15605 and GSE46517 datasets, background correction and normalization were first performed to generate gene-level expression matrices for downstream analysis. Differential expression analysis was then conducted using the `limma` pipeline, applying the same criteria as described for GEO datasets, with thresholds set at $|\log_2\text{FC}| > 1$ and adjusted P value ($\text{adj. } P.\text{Val}$) < 0.05 . Subsequently, differential expression analyses were also performed for the combined GEO dataset and the TCGA-SKCM cohort using consistent parameters. The intersection of

differentially expressed genes across GSE15605, GSE46517, the integrated dataset, and TCGA-SKCM yielded a total of 866 consistently differentially expressed genes, which were retained for subsequent functional analyses and model construction.

2.4 Functional enrichment analysis

Functional enrichment analysis was conducted to characterize the biological relevance of the 866 consistently differentially expressed genes. GO and KEGG enrichment analyses were performed using the clusterProfiler, org.Hs.eg.db, and enrichplot packages in R. Gene symbols were converted to Entrez IDs before analysis. GO enrichment was conducted for biological process, cellular component, and molecular function categories using enrichGO, while KEGG pathway analysis was performed using enrichKEGG with the human genome as the reference. Terms or pathways with $P < 0.05$ were considered statistically significant, and the top enriched results were visualized using bubble plots.

2.5 Machine learning–based feature selection

Machine learning–based feature selection was performed to identify key predictive genes from the 866 consistently differentially expressed genes. Four complementary algorithms, including LASSO regression, SVM-RFE, XGBoost, and Boruta, were applied to improve the robustness of feature screening. For LASSO regression, the optimal penalty parameter was determined by cross-validation. SVM-RFE was used to recursively rank and eliminate features according to their contribution to classification performance. XGBoost and Boruta were applied to evaluate gene importance based on gradient boosting and random forest frameworks, respectively. The candidate genes selected by each algorithm and their importance scores were recorded. Genes commonly identified by all four methods were defined as the final key feature genes for subsequent analyses.

2.6 Validation in an independent dataset

To validate the differential expression of the core genes, GSE114445 was used as an independent validation cohort. A total of four core genes were included, and their expression levels were compared between tumor and control groups. The standardized expression matrix was processed using the limma package to extract the expression profiles of these genes, followed by Student's t-test to evaluate differences between the two groups, with $P < 0.05$ considered statistically significant. Boxplots were then generated to visualize the validation results and illustrate the expression patterns of the core genes in the independent dataset.

2.7 Single-cell RNA-seq data processing and analysis

Single-cell RNA sequencing analysis was performed to characterize the cellular expression

patterns and dynamic changes of the core genes in SKCM. The publicly available scRNA-seq dataset GSE125121 was obtained from the GEO database and analyzed using the Seurat package in R. After quality control based on detected genes, UMI counts, and mitochondrial gene proportion, cells were normalized, clustered, and visualized using PCA and UMAP. Cell types, including malignant melanoma cells, fibroblasts, and immune cell subsets, were annotated according to canonical marker genes. The cell type-specific expression patterns of the core genes were visualized using DotPlot, FeaturePlot, and VlnPlot. Malignant cell subclusters were further subjected to pseudotime trajectory analysis using Monocle3, and the dynamic distribution of the hub genes along malignant cell evolution was visualized with plot cells.

2.8 Survival and clinical correlation analysis

Survival and clinicopathological correlation analyses were conducted to assess the prognostic and clinical relevance of core genes in SKCM. Patients in the TCGA-SKCM cohort were divided into high- and low-expression groups according to the median gene expression level. Kaplan–Meier survival analysis was performed using the survival package in R, and differences between groups were evaluated using the log-rank test. Associations between gene expression and clinicopathological variables, including sex, age, tumor stage, and survival status, were further analyzed. Chi-square tests were used for categorical variables, while Student's t-tests were applied for continuous variables. A value of $P < 0.05$ was considered statistically significant.

2.9 Functional enrichment analysis of the candidate gene

Single-gene functional annotation was conducted to explore the potential biological roles of hub genes in SKCM. Based on TCGA-SKCM expression data, Spearman correlation analysis was performed to identify high-related genes, with $|r| > 0.5$ and $P < 0.00001$ used as significance thresholds. The correlated genes were then subjected to GO and KEGG enrichment analyses using the clusterProfiler package in R. GO terms included biological process, cellular component, and molecular function categories. Enriched terms or pathways with adjusted $P < 0.05$ were considered statistically significant. This analysis was used to characterize biological processes and pathways potentially associated with key gene in SKCM.

2.10 Western blot analysis

Human tissue specimens, including normal skin, nevus, and SKCM tissues, were obtained from the General Hospital of Ningxia Medical University. A total of 18 tissue specimens were included, comprising normal skin ($n = 3$), nevus ($n = 3$), stage I SKCM ($n = 3$), stage II ($n = 3$), stage III ($n = 3$), and stage IV ($n = 3$). All specimens were histopathologically confirmed before

experimental analysis. The study was approved by the Medical Research Ethics Review Committee of the General Hospital of Ningxia Medical University (Approval No. KYLL-2025-1773), and written informed consent was obtained from all participants prior to the collection and use of their tissue samples.

Total protein was extracted from normal skin, nevus, and SKCM tissue samples using RIPA lysis buffer supplemented with protease and phosphatase inhibitors. After centrifugation, the supernatant was collected, and protein concentrations were determined using a BCA protein assay kit. Equal amounts of protein from each sample were separated by SDS-PAGE and transferred onto PVDF membranes. The membranes were blocked with 5% non-fat milk or bovine serum albumin at room temperature and then incubated overnight at 4°C with a rabbit polyclonal anti-ABCD1 antibody (Proteintech, Cat. No. 18138-1-AP). After washing with TBST, the membranes were incubated with an HRP-conjugated secondary antibody at room temperature. Protein bands were visualized using an enhanced chemiluminescence detection system. GAPDH was used as the internal loading control. The relative protein expression level of ABCD1 was quantified by densitometric analysis using ImageJ software and normalized to GAPDH.

2.11 Statistical analysis

All statistical analyses were performed using R software (version 4.3). For differential expression analysis, *P* values were adjusted for multiple testing using the Benjamini–Hochberg method to control the false discovery rate (FDR), and an adjusted *P* value (adj. *P*.Val) < 0.05 was considered the threshold for identifying differentially expressed genes. In survival analysis, the Kaplan–Meier method was used to estimate survival curves for different groups, and differences between groups were evaluated using the log-rank test, with *P* < 0.05 defined as statistically significant. For correlation analysis, Pearson or Spearman correlation coefficients were applied according to the data distribution and the nature of variable relationships; Pearson correlation was used for linear relationships, whereas Spearman correlation was applied for monotonic associations. Correlation results were visualized using heatmaps, scatter plots, and correlation matrices. Unless otherwise specified, all statistical tests were two-sided, and *P* < 0.05 was considered statistically significant.

3 Results

3.1 Differential expression analysis and batch effect correction of GEO datasets

After differential expression analysis of the GSE15605 dataset, a total of 2,582 differentially expressed genes (DEGs) were identified, including 1,542 downregulated and 1,039 upregulated

genes (Fig. 2A,2B). In the GSE46517 dataset, 817 DEGs were detected, comprising 568 downregulated and 249 upregulated genes (Fig. 2C, 2D). Considering the heterogeneity in data sources and platforms, potential batch effects were addressed prior to data integration. Batch effect correction and dataset integration were performed to enhance the stability and comparability of subsequent analyses. Boxplots of gene expression distributions before and after correction (Fig. 2E,2F) indicated that inter-sample variability was reduced, and overall expression distributions became more consistent following normalization. Principal component analysis (PCA) further confirmed the effectiveness of batch correction, as samples exhibited clear separation before correction but showed improved clustering and mixing across datasets after correction (Fig. 2G,2H). Based on the batch-corrected integrated dataset, differential expression analysis was repeated, yielding 1,147 DEGs, including 747 downregulated and 400 upregulated genes (Fig.2I,2J). These results provide a robust foundation for subsequent functional enrichment analyses and mechanistic investigations.

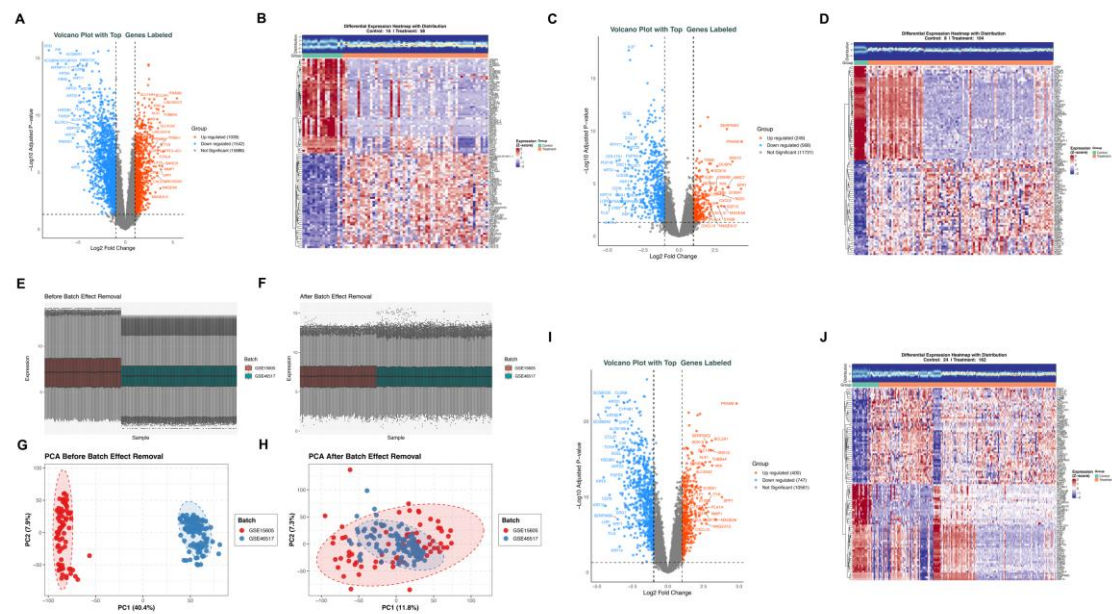


Fig.2 Differential expression analysis and batch effect correction of GEO datasets. (A, B) Volcano plot and heatmap of DEGs identified in the GSE15605 dataset. (C, D) Volcano plot and heatmap of DEGs identified in the GSE46517 dataset. (E, F) Boxplots showing the distribution of gene expression values before and after batch effect correction. (G, H) PCA plots illustrating sample clustering patterns before and after batch effect removal. (I, J) Volcano plot and heatmap of DEGs derived from the batch-corrected integrated dataset.

3.2 Functional enrichment analysis of differential expression genes

To explore the potential biological functions of differentially expressed genes associated with SKCM, a total of 7,570 DEGs were identified from the TCGA-SKCM dataset, including 3,860

upregulated and 3,710 downregulated genes, as visualized by a volcano plot and ranked gene scatter plot (Fig.3A, 3B). Intersection analysis with the 1,147 DEGs from the batch-corrected integrated GEO dataset yielded 866 overlapping genes (Fig.3C). GO enrichment analysis revealed that these genes were predominantly involved in epidermis development, keratinocyte differentiation, and skin morphogenesis, highlighting their critical roles in epidermal structure and barrier function (Fig.3D). The overlapping DEGs were mainly enriched in extracellular matrix- and membrane-related components, with molecular functions involving actin binding, transcriptional regulation, serine hydrolase activity, and cytokine activity. KEGG analysis further showed significant enrichment in cornified envelope formation, cadherin signaling, cytoskeletal regulation, and cytokine–cytokine receptor interaction pathways (Fig. 3E, F). Pathway correlation analysis indicated close functional links between epidermal differentiation, cell–cell adhesion, and inflammation-related processes. These findings suggest that the identified DEGs may contribute to SKCM progression by regulating epidermal differentiation, intercellular adhesion, cytoskeletal remodeling, and immune–inflammatory signaling.

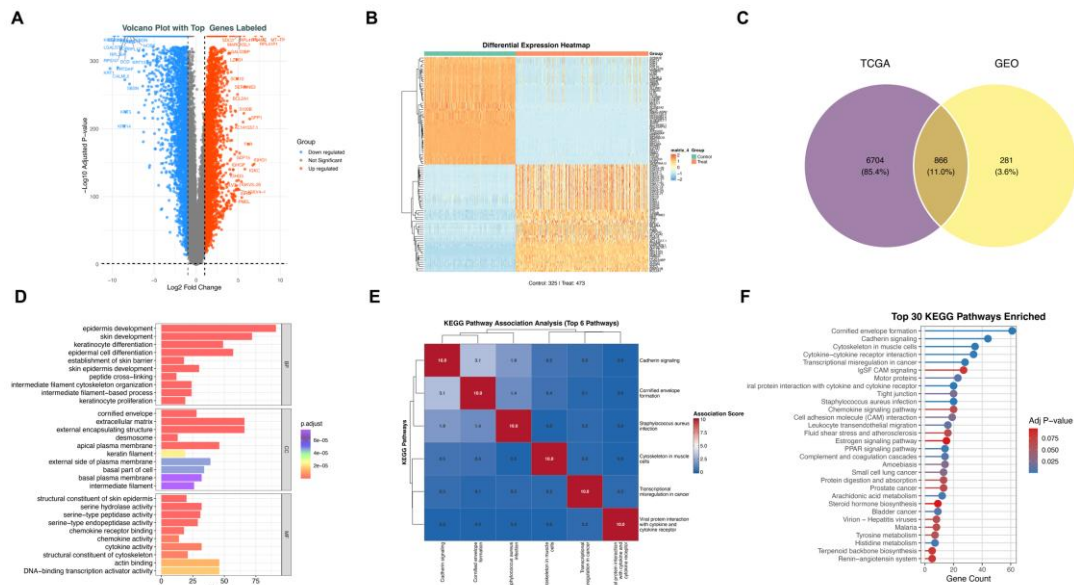


Fig.3 Integrated DEGs analysis and functional enrichment in SKCM. (A) Volcano plot representing DEGs in TCGA-SKCM. (B) Heatmap of DEG expression profiles. (C) Venn diagram of overlapping DEGs between TCGA-SKCM and the integrated GEO dataset. (D) GO enrichment for the overlapping DEGs. (E) Correlation heatmap of enriched KEGG pathways. (F) Top-ranked KEGG pathways according to adjusted *P*-value significance.

3.3 Integration of multiple machine learning algorithms and independent validation of key genes

Machine learning–based feature selection was performed to identify key predictive genes from

the 866 SKCM-associated DEGs. Four complementary algorithms, including LASSO regression, SVM-RFE, XGBoost, and Boruta, were applied to improve the robustness of candidate gene screening. In the LASSO model, 10-fold cross-validation identified the optimal penalty parameter, and 18 genes retained non-zero coefficients under the optimal λ value (Fig. 4A). SVM-RFE analysis achieved the best classification performance with 18 features, with a cross-validation accuracy of 0.905 and an error rate of 0.0952 (Fig. 4B). XGBoost analysis identified 20 genes with high feature importance, among which PPP1R13L, PRAME, ABCD1, LAMB4, and KIF3C showed the highest gain values (Fig. 4C). Boruta analysis further selected 144 statistically significant genes with importance scores higher than those of shadow features (Fig. 4D, E). By intersecting the results of the four algorithms, five core genes, ABCD1, KIF3C, CNNM1, TRIB1, and NT5DC3, were consistently identified (Fig. 4F). To evaluate their robustness, these genes were further validated in the independent cohort GSE114445. Boxplot analysis showed that all five genes were significantly differentially expressed in the validation cohort (all $P < 0.001$) (Fig. 4J). ROC curve analysis further demonstrated their strong diagnostic potential for distinguishing SKCM from normal samples (Fig. 4H).

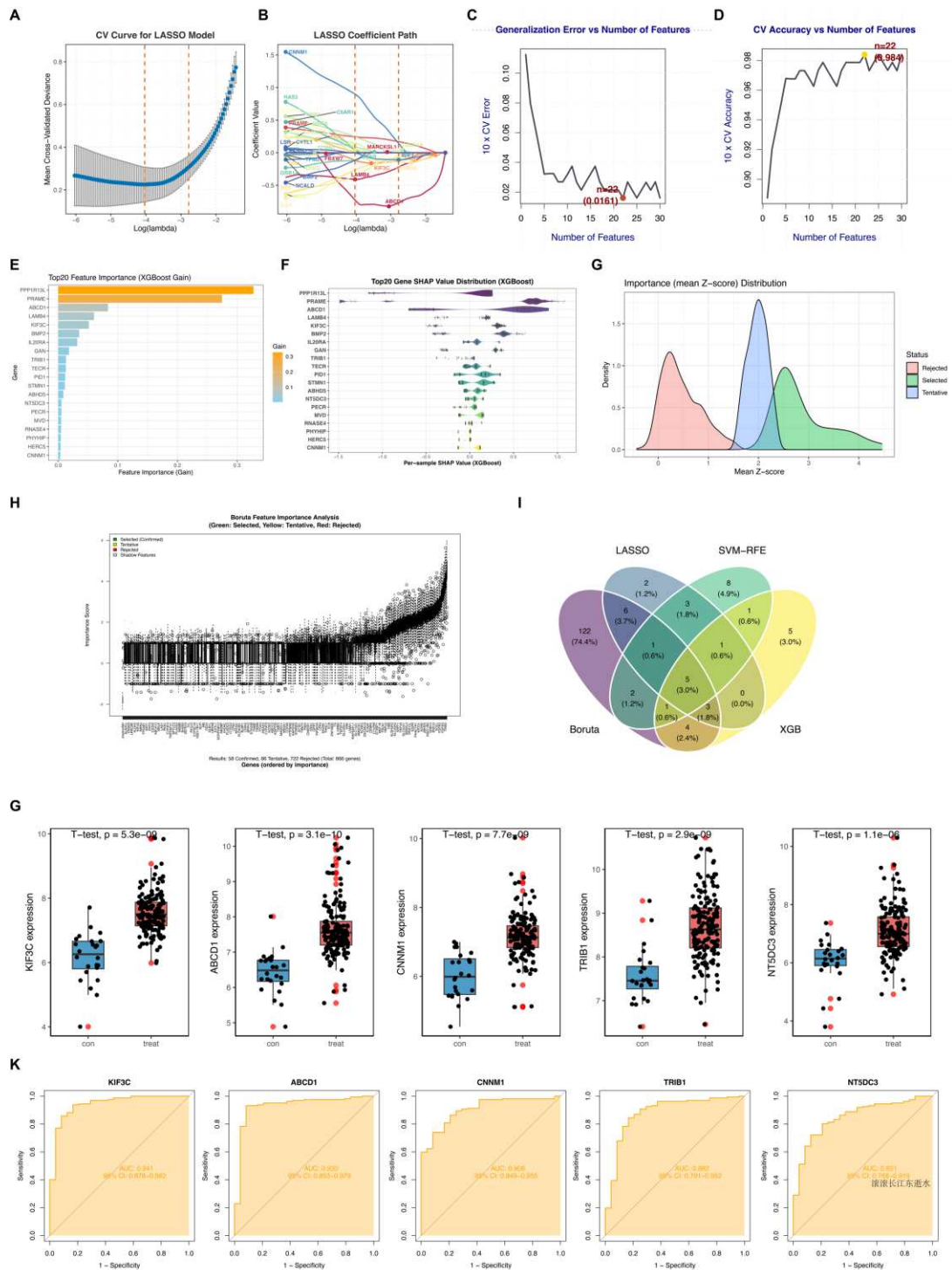


Fig.4 Identification and validation of core genes using multiple machine learning approaches. (A) LASSO regression showing cross-validation and coefficient profiles. (B) SVM-RFE analysis of model performance with increasing feature number. (C) XGBoost feature importance and SHAP value distribution. (D) Distribution of feature importance (mean Z-score). (E) Boruta feature selection results. (F) Venn diagram showing overlap of candidate genes from four algorithms. (G) Differential expression of the five core genes (KIF3C, ABCD1,

CNNM1, TRIB1, and NT5DC3) between groups. (H) ROC curve analysis evaluating diagnostic performance of the five genes.

3.4 Single-cell analysis reveals cellular heterogeneity of core genes within the SKCM tumor microenvironment

Single-cell transcriptomic analysis revealed distinct cell type-specific expression patterns of the five core genes within the SKCM tumor microenvironment, with predominant enrichment in malignant melanoma cells. ABCD1, KIF3C, and CNNM1 showed similar distribution patterns and were mainly expressed in multiple malignant cell subclusters, with limited expression in non-tumor cells. Among them, ABCD1 exhibited the highest expression level and broadest distribution, suggesting strong tumor specificity. NT5DC3 was also primarily expressed in malignant cell subsets, although detectable expression was observed in fibroblasts. In contrast, TRIB1 displayed a broader expression pattern, with substantial expression in fibroblasts and mononuclear cells in addition to melanoma cells, indicating a closer association with the tumor microenvironment. Overall, these findings demonstrate functional heterogeneity among the core genes, with ABCD1 showing the most prominent tumor-specific expression, whereas TRIB1 exhibited dual tumor-microenvironment characteristics. These patterns were further supported by FeaturePlot and VlnPlot analyses (Fig.4C-F).

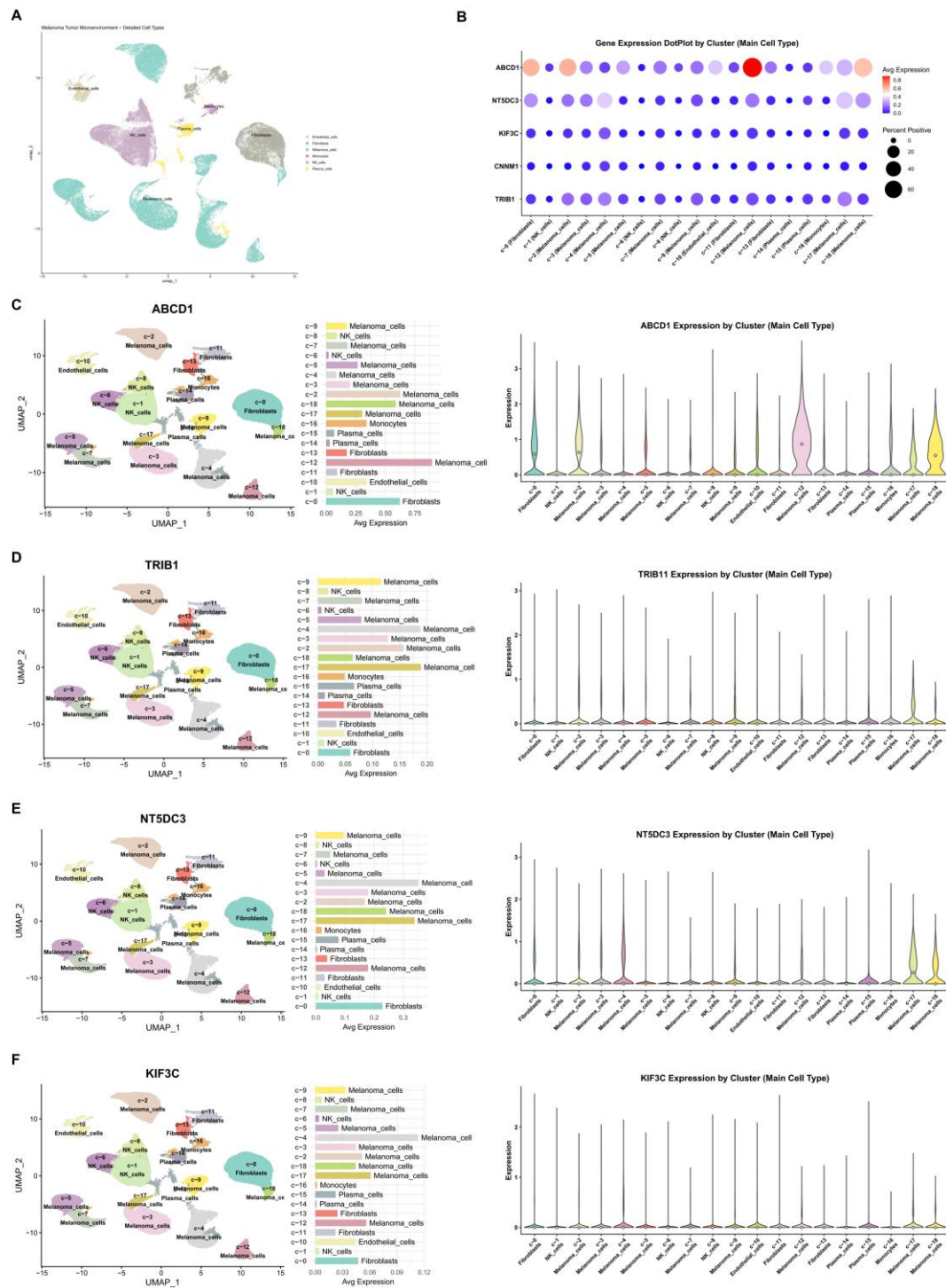


Fig.5 Single-cell sequencing reveals the expression profiles of ABCD1, TRIB1, NT5DC3, and KIF3C across different cell subpopulations. (A) UMAP plot showing annotated cell populations in the melanoma tumor microenvironment, including melanoma cells, fibroblasts, endothelial cells, monocytes, NK cells, and plasma cells. (B) Dot plot illustrating the average expression levels and percentage of expression for ABCD1, TRIB1, NT5DC3, and KIF3C in various cell types. (C) Feature plot and violin plot of ABCD1 expression on the UMAP, with

predominant expression in melanoma cells and additional expression observed in fibroblasts and plasma cells. (D) Feature plot and violin plot of TRIB1 expression, indicating minimal baseline expression across nearly all cell types. (E) Feature plot and violin plot of NT5DC3 expression, demonstrating its expression in multiple cell types, with notable presence in Fibroblasts and melanoma cells. (F) Feature plot and violin plot of KIF3C expression, demonstrating its expression in multiple cell types.

Pseudotime trajectory analysis revealed dynamic expression heterogeneity of the core genes during malignant melanoma cell evolution (Fig.6A). Malignant cells exhibited a continuous progression along the pseudotime axis, with distinct branches representing different cellular states. Among the five core genes, ABCD1 was broadly enriched in several terminal branches, suggesting its potential association with late-stage malignant states (Fig.6B). NT5DC3 showed a similar but more restricted terminal distribution (Fig.6D), whereas KIF3C and CNNM1 exhibited relatively low and sporadic expression signals (Fig.6E, F). In contrast, TRIB1 was detected in both transitional and terminal regions, indicating its potential involvement across multiple stages of malignant cell progression (Fig.6C). These results suggest that the core genes exhibit distinct dynamic expression patterns along malignant melanoma cell evolution.

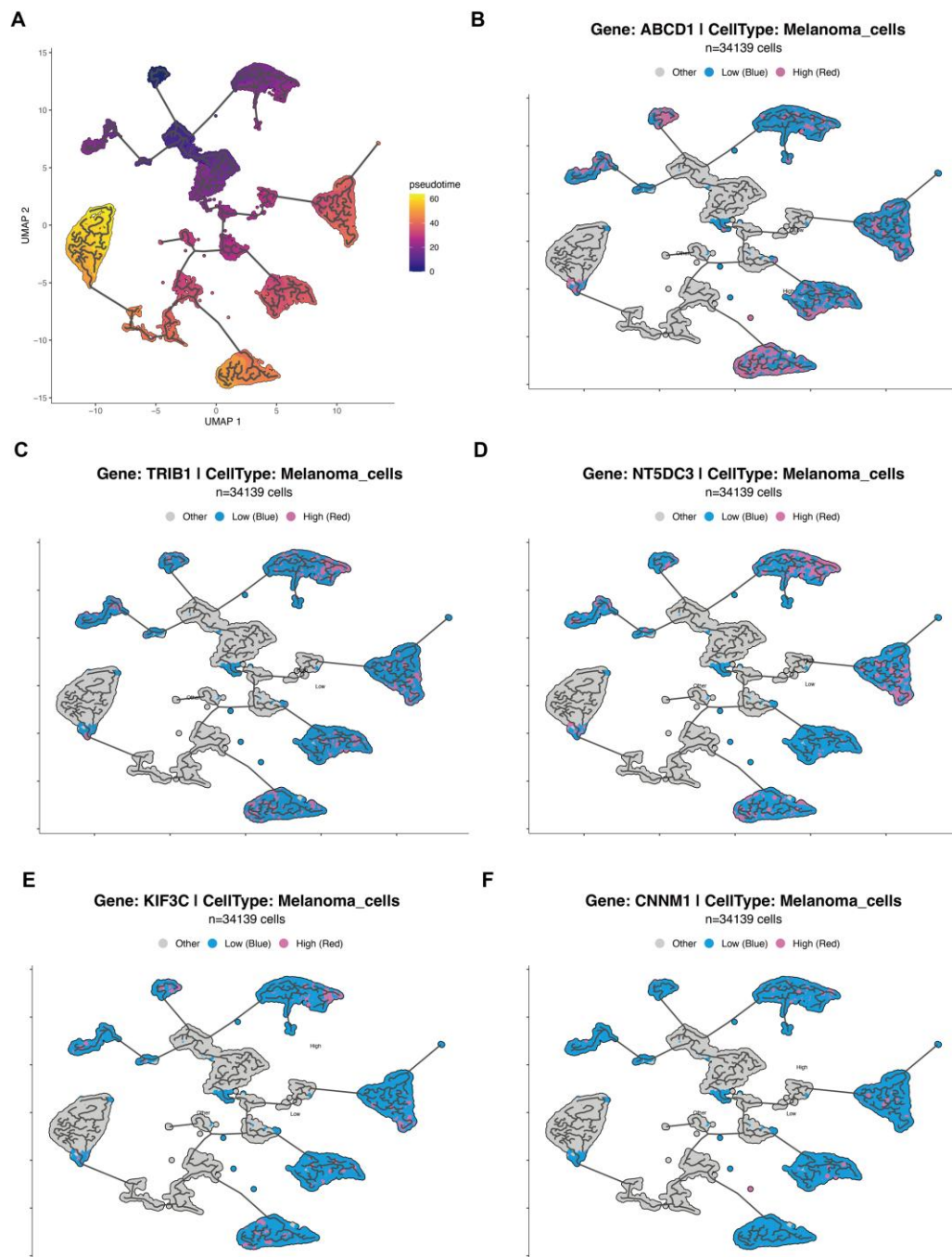


Fig.6 Pseudotime trajectory analysis of Melanoma cells. (A)UMAP plot of the cellular differentiation trajectory for malignant cells in SKCM, with colors representing the pseudotime progression. (B) Expression pattern of ABCD1 along the malignant cell trajectory, showing the vast majority of malignant cells maintain a low-expression state (blue). (C) Expression pattern of TRIB1 along the malignant cell trajectory. (D) Expression pattern of NT5DC3 along the malignant cell trajectory. (E) Expression pattern of KIF3C along the malignant cell trajectory. (F) Expression pattern of CNNM1 along the malignant cell trajectory. Cells are colored based

on expression levels: low (blue) and high (pink), with non-malignant cells shown as a gray background.

3.5 Analysis of expression differences and clinical characteristics of key genes in SKCM

Analysis of expression differences and clinical relevance of key genes in SKCM was performed using the TCGA-SKCM cohort (Fig.7A–D). All five genes showed significant differential expression between tumor and normal tissues (all $P < 0.05$). Patients were then stratified into high- and low-expression groups based on the median expression of each gene, and Kaplan–Meier analysis was conducted to assess prognostic associations (Fig.7E–H). ABCD1 expression was significantly associated with overall survival ($P = 4.4 \times 10^{-4}$), with the high-expression group exhibiting poorer survival, supporting its potential as a prognostic biomarker. KIF3C showed a weaker but significant association ($P = 0.0162$), which requires validation in larger or independent cohorts. In contrast, NT5DC3 ($P = 0.3082$) and TRIB1 ($P = 0.1114$) were not significantly associated with survival, indicating limited or context-dependent prognostic relevance.

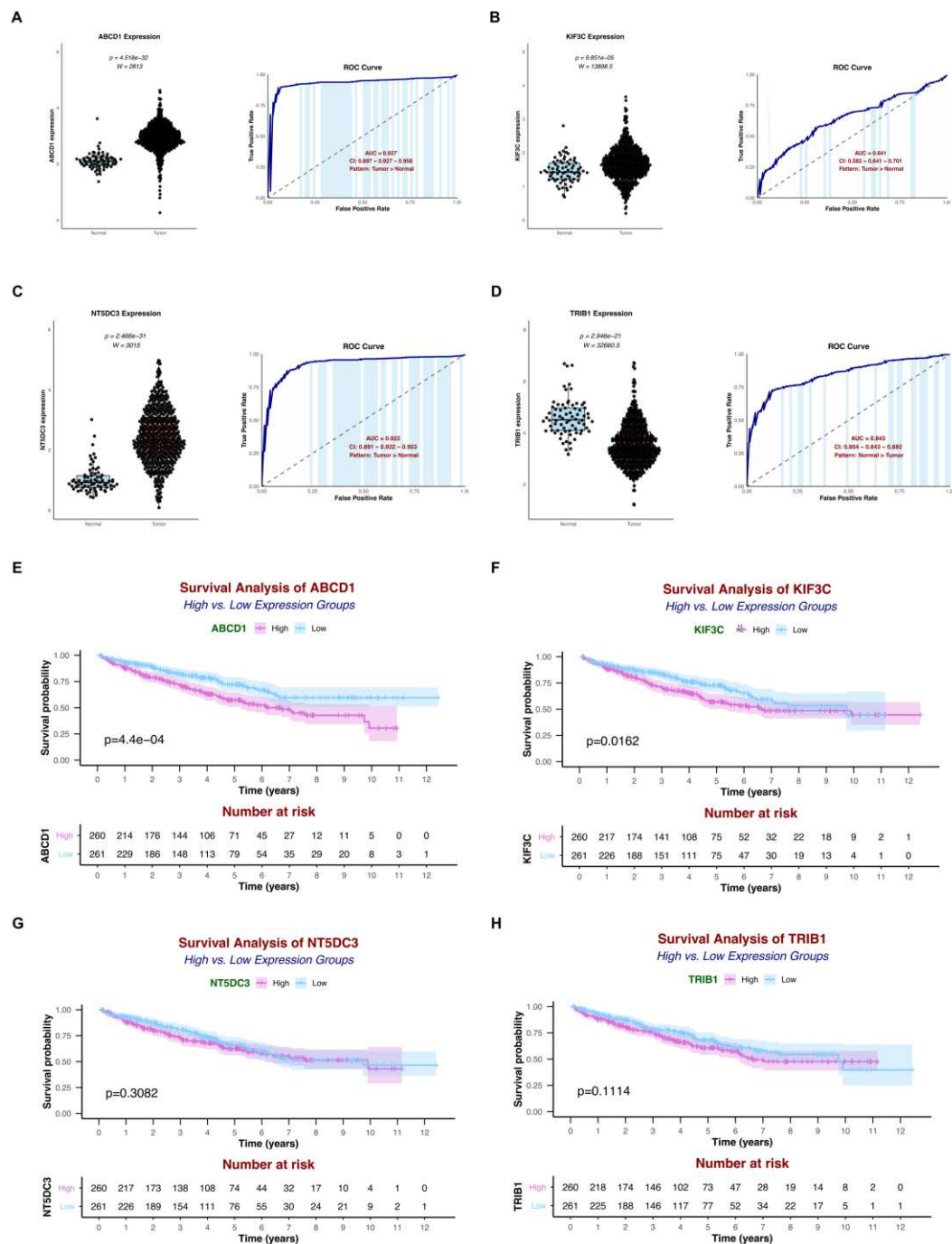


Fig.7 Expression profiles and overall survival analysis of four core genes in the TCGA-SKCM dataset. (A–D) Violin plots showing log2-transformed expression levels of ABCD1, KIF3C, NT5DC3, and TRIB1 in tumor and normal tissues. Differences between groups were assessed using the Wilcoxon rank-sum test. (E–H) Kaplan–Meier survival curves comparing overall survival between high- and low-expression groups for each gene. Patients were stratified based on the median expression level of each gene, and statistical significance was

evaluated using the log-rank test.

To assess the clinical relevance of ABCD1, its association with multiple clinicopathological features was analyzed (Fig.8A–F). Heatmaps and boxplots indicated no significant differences in ABCD1 expression across age groups, survival status, N stage, or M stage. In contrast, significant differences were observed by sex ($P = 0.017$), T stage ($P = 0.033$), and overall stage ($P = 0.0082$), suggesting potential associations with tumor progression and sex-related variation.

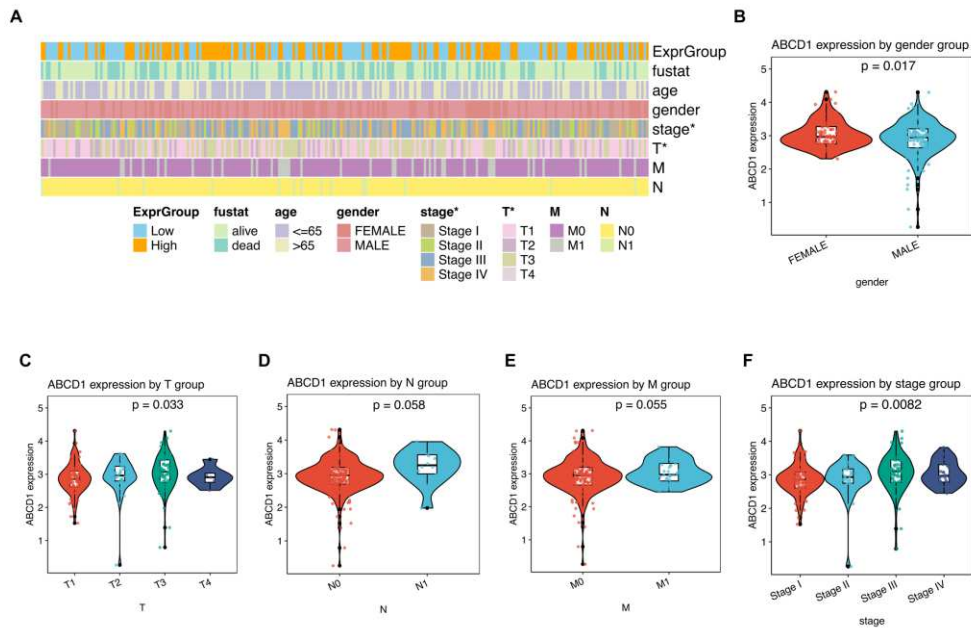


Fig.8 Clinical associations of ABCD1 expression in the TCGA-SKCM cohort.

(A) Heatmap showing the distribution of ABCD1 expression across clinical variables, including expression group, survival status, age, gender, stage, and TNM classification. (B–F) Violin plots showing ABCD1 expression according to gender (B), T stage (C), N stage (D), M stage (E), and overall stage (F). P values are indicated in each panel. Differences among multiple groups were evaluated using the Kruskal–Wallis test.

3.6 Functional annotation and pathway enrichment analysis of ABCD1 related genes

Functional annotation and pathway enrichment analysis of ABCD1-related genes were performed to explore its potential biological roles in SKCM. Spearman correlation analysis identified strong positive correlations between ABCD1 and MFSD12 ($r = 0.64$, $P = 2.82 \times 10^{-62}$), ATP8B2 ($r = 0.64$, $P = 6.02 \times 10^{-62}$), and RENBP ($r = 0.61$, $P < 1 \times 10^{-200}$) (Fig. 9A). GO enrichment analysis indicated that these genes were primarily involved in endothelial cell migration, angiogenesis regulation, and vascular development, with molecular functions related to extracellular matrix structural components, extracellular matrix binding, and growth factor binding (Fig.9B). KEGG pathway analysis showed that the gene set was enriched in pathways

including melanoma, autophagy, PI3K–Akt signaling, and metabolic processes such as amino sugar and nucleotide sugar metabolism (Fig.9C). Collectively, these findings suggest that ABCD1-associated genes may participate in tumor microenvironment remodeling, angiogenesis-related processes, and key signaling pathways in SKCM.

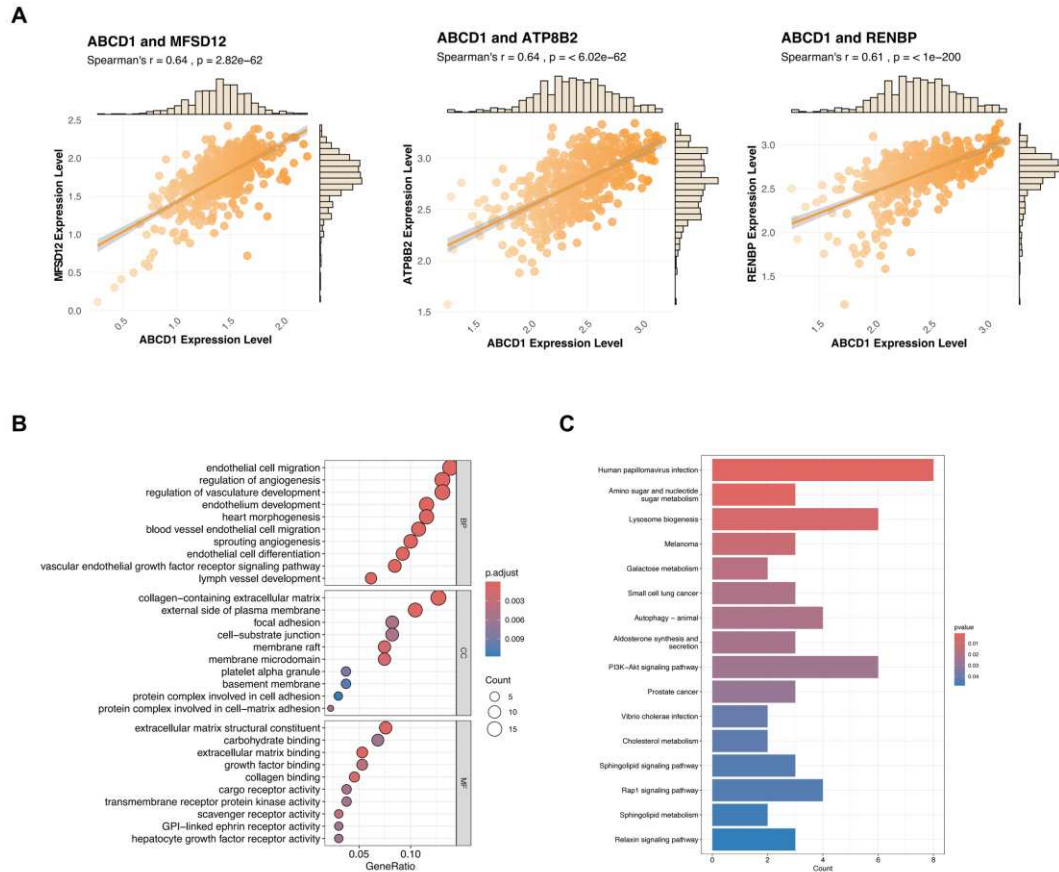


Fig.9 Correlation and functional enrichment analysis of ABCD1 in SKCM.

(A) Scatter plots showing positive correlations between ABCD1 and MFSD12, ATP8B2, and RENBP based on Spearman correlation analysis. (B) GO enrichment analysis of ABCD1-related genes, including biological process, cellular component, and molecular function categories. (C) KEGG pathway enrichment analysis highlighting significantly enriched signaling and metabolic pathways.

3.7 Western blot validation of ABCD1 expression in SKCM

Western blot analysis was conducted to validate the protein expression of ABCD1 in normal skin, nevus, and SKCM tissues at different clinical stages. ABCD1 protein expression was low in normal skin, nevus, and stage I SKCM tissues, whereas a pronounced increase was observed in stage II and stage III SKCM tissues (Fig.10A). Densitometric quantification showed that ABCD1 expression was significantly elevated in stage II and stage III SKCM compared with the low-expression groups (Fig.10B). Notably, stage III SKCM exhibited the highest ABCD1

protein level. Although ABCD1 expression declined in stage IV SKCM relative to stage III, it remained higher than that in normal skin, nevus, and stage I SKCM tissues. These results indicate that ABCD1 protein upregulation may be associated with SKCM progression, particularly during intermediate stages of the disease. The detailed Western blot images are provided in Supplementary Figure S1.

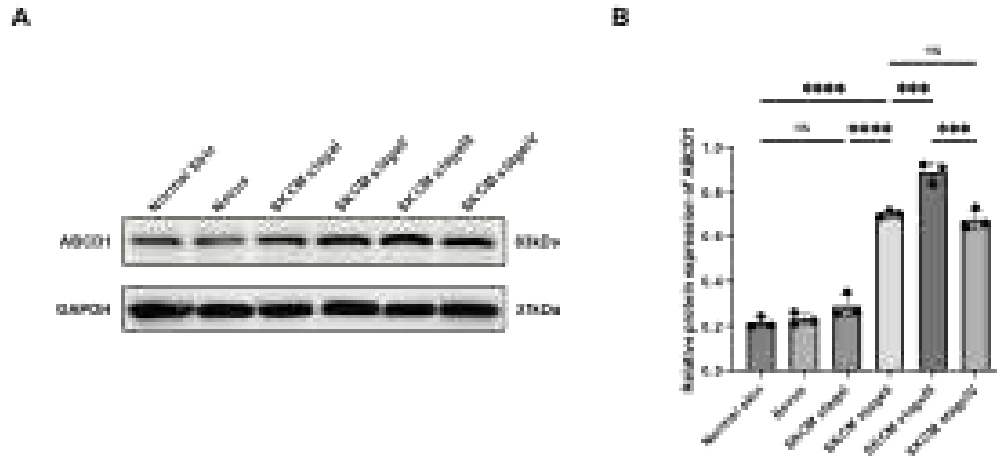


Fig.10 Western blot validation of ABCD1 protein expression in normal skin, nevus, and SKCM tissues. (A) Representative Western blot bands showing ABCD1 protein expression in normal skin, nevus, and SKCM tissues at different clinical stages. (B) Quantitative analysis of ABCD1 protein expression normalized to GAPDH. ABCD1 expression was relatively low in normal skin, nevus, and stage I SKCM tissues, but was markedly increased in stage II and stage III SKCM tissues. Stage IV SKCM showed a partial decrease compared with stage III but maintained a relatively elevated expression level. Data are presented as mean $\pm$ SD. ns, not significant; *** $P < 0.001$; **** $P < 0.0001$

4 Discussion

Cutaneous malignant melanoma (SKCM) is a highly aggressive malignancy originating from melanocytes[9]. Characterized by rapid progression, a strong metastatic potential, and pronounced molecular heterogeneity, it remains one of the leading causes of skin cancer-related mortality[3]. Despite substantial advances in early detection, surgical management, targeted therapies, and immune checkpoint blockade, patients with advanced or metastatic melanoma continue to face significant clinical challenges, including recurrence, distant metastasis, therapeutic resistance, and poor outcomes[10, 11]. At present, risk stratification and prognostic evaluation of SKCM largely rely on conventional clinicopathological parameters, such as Breslow thickness, ulceration status, and TNM staging, with Breslow thickness and ulceration recognized as key prognostic indicators[12]. However, growing evidence suggests that, although clinically valuable, these factors are insufficient to fully capture the biological

complexity and molecular heterogeneity of melanoma[13, 14]. As a result, their ability to accurately predict prognosis, assess therapeutic response, and guide individualized treatment remains limited. Therefore, the identification of robust and reliable molecular biomarkers is critical for improving early detection, refining risk stratification, and advancing precision medicine in SKCM.

Conventional bulk transcriptomic analyses, while effective for identifying differentially expressed genes, primarily capture averaged expression signals across heterogeneous tissue samples[15]. As a result, they are limited in resolving cellular heterogeneity and cannot reliably determine the true cellular origins of candidate genes[16]. This constraint hinders precise interpretation of the biological roles of key molecules in tumor initiation and progression. To overcome this limitation, our study integrated multi-cohort transcriptomic screening, machine learning-based feature selection, and single-cell analysis to identify and characterize robust molecular biomarkers in SKCM. Through this framework, five core genes, ABCD1, KIF3C, CNNM1, TRIB1, and NT5DC3, were identified as candidate molecular features. Among them, ABCD1 showed predominant enrichment in malignant melanoma cell populations, suggesting that its dysregulation may mainly reflect tumor-intrinsic alterations rather than changes in tumor microenvironmental composition. Clinically, high ABCD1 expression was associated with poorer overall survival and showed significant correlations with T stage and histological grade, with an increasing trend across N and M stages. These findings suggest that ABCD1 may be related to tumor progression and invasive potential. Since current prognostic assessment of SKCM largely relies on conventional clinicopathological parameters, such as Breslow thickness, ulceration, and AJCC staging, ABCD1 may provide complementary molecular information for risk stratification and prognostic evaluation.

From a functional perspective, ABCD1 encodes a peroxisomal ATP-binding cassette (ABC) half-transporter responsible for the import of coenzyme A-activated very long-chain fatty acids (VLCFAs) into peroxisomes for β -oxidation[17]. Loss or dysfunction of ABCD1 leads to VLCFA accumulation, which disrupts cholesterol esterification, lipid droplet formation, and overall lipid homeostasis, placing this gene at a critical intersection of lipid metabolism, membrane remodeling, and redox regulation[18]. Increasing evidence indicates that peroxisomes function as dynamic metabolic hubs rather than passive organelles, supporting tumor cell survival under hypoxia, nutrient deprivation, and therapeutic stress through lipid catabolism, lipid-derived signaling, and reactive oxygen species (ROS) homeostasis[19]. Consistently, melanoma is characterized by pronounced lipid metabolic reprogramming, with tumor progression, phenotypic plasticity, invasion, and therapeutic resistance closely linked to fatty acid uptake, fatty acid oxidation, and membrane lipid remodeling[20-22]. Notably, MAPK inhibitor treatment has been shown to induce PPAR α /CPT1A-dependent fatty acid oxidation,

and a subset of resistant melanoma cells relies on peroxisome-driven metabolic adaptation for survival[23, 24]. In parallel, melanoma cells are exposed to sustained oxidative stress, and their capacity to modulate ROS levels and resist lipid peroxidation and ferroptosis plays a key role in disease progression and treatment response[25].

Integrating current evidence with our findings, we propose that the elevated expression of ABCD1 in SKCM is unlikely to be incidental. On one hand, increased ABCD1 activity may facilitate the transport and β -oxidation of VLCFAs within peroxisomes, thereby supporting membrane lipid turnover, mitigating lipid toxicity, and enhancing cellular adaptation to oxidative stress[26]. On the other hand, ABCD1-mediated peroxisomal lipid metabolism may confer metabolic flexibility, enabling tumor cells to better withstand microenvironmental and therapeutic stress[27]. Previous studies have shown that abnormal expression or activation of ATP-binding cassette (ABC) transporters is a common mechanism underlying tumor multidrug resistance, as these transporters can actively efflux anticancer agents, thereby reducing intracellular drug accumulation and compromising therapeutic efficacy. Notably, melanoma cells express multiple ABC transporters, including ABCD1, ABCB1, ABCB5, ABCB8, ABCC1, ABCC2, and ABCA9, suggesting that these molecules may contribute not only to the intrinsic or acquired resistance of melanoma to diverse anticancer therapies, but also to the protective clearance of toxic melanin intermediates and metabolites in melanocytic cells[28, 29]. These discrepancies likely reflect context-dependent regulation influenced by tumor type, disease stage, cellular composition, and treatment history. In highly heterogeneous tumors such as melanoma, bulk transcriptomic data are particularly susceptible to confounding by stromal and immune components[30]; however, our single-cell analysis demonstrates that ABCD1 is predominantly enriched in malignant melanoma cells, supporting a tumor-intrinsic role. Collectively, these findings suggest that ABCD1 may contribute to SKCM progression by linking peroxisomal function, lipid metabolic reprogramming, oxidative stress adaptation, and therapeutic resistance. Nevertheless, whether ABCD1 directly regulates melanoma cell proliferation, invasion, ferroptosis sensitivity, or resistance to MAPK inhibitors warrants further investigation through functional and metabolic assays, including knockdown or overexpression studies and metabolic flux analysis.

In addition to ABCD1, the three other genes identified in this study—KIF3C, NT5DC3, and TRIB1—also merit attention. KIF3C is a member of the kinesin family and is primarily involved in microtubule-dependent intracellular transport; in several tumor types, it has been linked to cell proliferation, migration, and invasion[31-33]. NT5DC3 belongs to a class of proteins containing a 5'-nucleotidase domain[34]. Although current evidence remains limited, it may be associated with mitochondrial-related processes and tumor metabolic regulation, while its specific role in melanoma has yet to be clarified. TRIB1, a member of the Tribbles

pseudokinase family, mainly functions as a signaling scaffold protein involved in MAPK signaling, COP1-mediated degradation, and the regulation of cell growth and differentiation[35]. It has been implicated in tumor progression, therapeutic resistance, and immune microenvironment remodeling in multiple cancers, and recent studies further suggest that TRIB1 may participate in relevant signaling networks in BRAFV600E melanoma[36]. Overall, although the precise roles of these three genes in SKCM remain insufficiently defined, they point to distinct biological processes—including cytoskeletal transport, metabolic homeostasis, and signal transduction regulation—suggesting that, together with ABCD1, they may form a multidimensional molecular network contributing to melanoma progression.

Overall, the integration of multi-cohort transcriptomic analysis, machine learning, and single-cell validation enhanced the robustness and biological interpretability of ABCD1 as a candidate biomarker in SKCM. Nevertheless, several limitations should be noted. First, this study was mainly based on public retrospective datasets, and heterogeneity in sample sources, sequencing platforms, and clinical annotations may affect the robustness of the findings. Second, although single-cell analysis supported the enrichment of ABCD1 in malignant melanoma cells, its spatial organization and microenvironmental context remain unclear. Third, the direct functional role of ABCD1 in SKCM progression has not yet been experimentally validated. Therefore, further validation in large independent cohorts, spatial omics analyses, and mechanistic studies is warranted to clarify the clinical significance, biological role, and potential therapeutic relevance of ABCD1 in SKCM. Overall, ABCD1 represents a promising biomarker in SKCM, although its clinical utility and underlying mechanisms remain to be fully defined.

5 Conclusions

This study identified ABCD1 as a promising biomarker for cutaneous melanoma through integrated transcriptomic and machine learning analyses. Single-cell analysis showed preferential ABCD1 expression in malignant melanoma cells, while clinical analyses linked elevated ABCD1 expression to poorer overall survival and progression-related features. Western blotting further confirmed increased ABCD1 protein expression in melanoma tissues. Collectively, these findings support the potential diagnostic and prognostic value of ABCD1 and warrant further validation in larger independent cohorts and functional studies.

Acknowledgments

We thank all contributors to the GEO and TCGA databases and developers of the related databases.

Author contributions

R.Y and H.Z. designed the research. H.L and L.W performed data collection, bioinformatics analysis. J.Y revised the manuscript and provided overall study supervision. N.Y and J.L contributed to manuscript revision. All authors reviewed and approved the final manuscript.

Funding

Not applicable.

Data availability

The datasets analyzed in this study are publicly available. The GEO datasets GSE15605, GSE46517, GSE114445, and GSE125121 are available from the NCBI Gene Expression Omnibus (GEO) through the following direct links:

GSE15605: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15605>

GSE46517: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46517>

GSE114445: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114445>

GSE125121: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125121>

TCGA-SKCM data are available from the Genomic Data Commons (GDC) through the TCGA-SKCM project page: <https://portal.gdc.cancer.gov/projects/TCGA-SKCM>. Normal skin transcriptomic data from GTEx are publicly available through the GTEx datasets page: <https://gtexportal.org/home/datasets/>. TCGA-SKCM and GTEx expression data used in this study were accessed through the UCSC Xena platform.

Declarations

Ethics approval

This study was approved by the Medical Research Ethics Review Committee of the General Hospital of Ningxia Medical University (Approval No. KYLL-2025-1773) and was conducted in accordance with the Declaration of Helsinki and relevant institutional guidelines and regulations.

Consent to participate

Written informed consent was obtained from all participants.

Consent to publish

Not applicable. No identifiable personal information or images are included in this manuscript.

Competing interests

The authors declare no competing interests

References

1. Eggermont AM, Spatz A, Robert C. Cutaneous melanoma. *Lancet*. 2014;383(9919):816-27.
2. Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross MI, et al. Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin*. 2017;67(6):472-92.
3. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. *CA Cancer J Clin*. 2025;75(1):10-45.
4. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023;72(2):338-44.
5. Tímár J, Ladányi A. Molecular Pathology of Skin Melanoma: Epidemiology, Differential Diagnostics, Prognosis and Therapy Prediction. *Int J Mol Sci*. 2022;23(10).
6. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11(10):733-9.
7. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113-20.
8. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*. 2018;36(5):421-7.
9. Joshi UM, Kashani-Sabet M, Kirkwood JM. Cutaneous Melanoma: A Review. *Jama*. 2025;334(23):2113-25.
10. Knight A, Karapetyan L, Kirkwood JM. Immunotherapy in Melanoma: Recent Advances and Future Directions. *Cancers (Basel)*. 2023;15(4).
11. Fenton SE, Sosman JA, Chandra S. Resistance mechanisms in melanoma to immuneoncologic therapy with checkpoint inhibitors. *Cancer Drug Resist*. 2019;2(3):744-61.
12. Garbe C, Peris K, Hauschild A, Saiag P, Middleton M, Bastholt L, et al. Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline - Update 2016. *Eur J Cancer*. 2016;63:201-17.
13. Karras P, Bordeu I, Pozniak J, Nowosad A, Pazzi C, Van Raemdonck N, et al. A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature*. 2022;610(7930):190-8.
14. Tasdogan A, Faubert B, Ramesh V, Ubellacker JM, Shen B, Solmonson A, et al. Metabolic heterogeneity confers differences in melanoma metastatic potential. *Nature*. 2020;577(7788):115-20.
15. Nesari AM, MotieGhader H, Ghorbian S. Advances and challenges in single-cell RNA sequencing data analysis: a comprehensive review. *Brief Bioinform*. 2026;27(1).
16. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*. 2019;20(11):631-56.
17. Mosser J, Douar AM, Sarde CO, Kioschis P, Feil R, Moser H, et al. Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. *Nature*. 1993;361(6414):726-30.
18. Hillebrand M, Verrier SE, Ohlenbusch A, Schäfer A, Söling HD, Wouters FS, et al. Live

cell FRET microscopy: homo- and heterodimerization of two human peroxisomal ABC transporters, the adrenoleukodystrophy protein (ALDP, ABCD1) and PMP70 (ABCD3). *J Biol Chem.* 2007;282(37):26997-7005.

19. Schaaf MB, Houbaert D, Meçe O, Agostinis P. Autophagy in endothelial cells and tumor angiogenesis. *Cell Death Differ.* 2019;26(4):665-79.

20. Gurung S, Budden T, Mallela K, Jenkins B, von Kriegsheim A, Manrique E, et al. Stromal lipid species dictate melanoma metastasis and tropism. *Cancer Cell.* 2025;43(6):1108-24.e11.

21. Zhang M, Di Martino JS, Bowman RL, Campbell NR, Baksh SC, Simon-Vermot T, et al. Adipocyte-Derived Lipids Mediate Melanoma Progression via FATP Proteins. *Cancer Discov.* 2018;8(8):1006-25.

22. Alicea GM, Rebecca VW, Goldman AR, Fane ME, Douglass SM, Behera R, et al. Changes in Aged Fibroblast Lipid Metabolism Induce Age-Dependent Melanoma Cell Resistance to Targeted Therapy via the Fatty Acid Transporter FATP2. *Cancer Discov.* 2020;10(9):1282-95.

23. Turner JA, Paton EL, Van Gulick R, Stefanoni D, Cendali F, Reisz J, et al. BRAF Modulates Lipid Use and Accumulation. *Cancers (Basel).* 2022;14(9).

24. Huang F, Cai F, Dahabieh MS, Gunawardena K, Talebi A, Dehairs J, et al. Peroxisome disruption alters lipid metabolism and potentiates antitumor response with MAPK-targeted therapy in melanoma. *J Clin Invest.* 2023;133(20).

25. Wang Y, Li C, Zhuo J, Hui H, Zhou B, Tian J. The Detection of Divalent Iron and Reactive Oxygen Species During Ferroptosis with the Use of a Dual-Reaction Turn-On Fluorescent Probe. *Mol Imaging Biol.* 2023;25(2):423-34.

26. Baarine M, Beeson C, Singh A, Singh I. ABCD1 deletion-induced mitochondrial dysfunction is corrected by SAHA: implication for adrenoleukodystrophy. *J Neurochem.* 2015;133(3):380-96.

27. Kunz HH, Scharnewski M, Feussner K, Feussner I, Flügge UI, Fulda M, et al. The ABC transporter PXA1 and peroxisomal beta-oxidation are vital for metabolism in mature leaves of *Arabidopsis* during extended darkness. *Plant Cell.* 2009;21(9):2733-49.

28. Chen KG, Valencia JC, Gillet JP, Hearing VJ, Gottesman MM. Involvement of ABC transporters in melanogenesis and the development of multidrug resistance of melanoma. *Pigment Cell Melanoma Res.* 2009;22(6):740-9.

29. Szakács G, Annereau JP, Lababidi S, Shankavaram U, Arciello A, Bussey KJ, et al. Predicting drug sensitivity and resistance: profiling ABC transporter genes in cancer cells. *Cancer Cell.* 2004;6(2):129-37.

30. Pompella L, Tirino G, Pappalardo A, Caterino M, Ventriglia A, Nacca V, et al. Pancreatic Cancer Molecular Classifications: From Bulk Genomics to Single Cell Analysis. *Int J Mol Sci.* 2020;21(8).

31. Wang C, Wang C, Wei Z, Li Y, Wang W, Li X, et al. Suppression of motor protein KIF3C expression inhibits tumor growth and metastasis in breast cancer by inhibiting TGF- β signaling. *Cancer Lett.* 2015;368(1):105-14.

32. Wang X, Shi Y, Shi H, Liu X, Liao A, Liu Z, et al. MUC20 regulated by extrachromosomal circular DNA attenuates proteasome inhibitor resistance of multiple myeloma by modulating cuproptosis. *J Exp Clin Cancer Res.* 2024;43(1):68.

- 652 33. Gao Y, Zheng H, Li L, Zhou C, Chen X, Zhou X, et al. KIF3C Promotes Proliferation,
653 Migration, and Invasion of Glioma Cells by Activating the PI3K/AKT Pathway and Inducing
654 EMT. *Biomed Res Int.* 2020;2020:6349312.
- 655 34. Yu X, Sun R, Yang X, He X, Guo H, Ou C. The NT5DC family: expression profile and
656 prognostic value in pancreatic adenocarcinoma. *J Cancer.* 2023;14(12):2274-88.
- 657 35. Iwamoto S, Boonvisut S, Makishima S, Ishizuka Y, Watanabe K, Nakayama K. The role
658 of TRIB1 in lipid metabolism; from genetics to pathways. *Biochem Soc Trans.*
659 2015;43(5):1063-8.
- 660 36. Dobens LL, Nauman C, Fischer Z, Yao X. Control of Cell Growth and Proliferation by the
661 Tribbles Pseudokinase: Lessons from *Drosophila*. *Cancers (Basel).* 2021;13(4).

662

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryFigure10.pdf](#)