IET Systems Biology

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Utilising Machine Learning and Single-Cell Analysis to Uncover SKCM Metastasis–Related Genes

Zhiwei Liao | Weiming Chen | Yingdi He | Yichen Zheng | Xiaonan Chen | Han Shen

Guangdong Pharmaceutical University, Guangzhou University Town, Guangzhou, China

Correspondence: Han Shen (shenhanbc@163.com)

Received: 20 October 2025 | **Revised:** 28 February 2026 | **Accepted:** 2 March 2026

Keywords: biocommunications | biocomputers | genomics | particle swarm optimisation

ABSTRACT

The high mortality rate of metastatic cutaneous melanoma (SKCM) remains a major challenge in clinical treatment. This study used single-cell RNA sequencing (scRNA-Seq) technology to compare the differences between metastatic and primary tumour cells. By manually annotating cell types, significant disparities in cell communication patterns and functional pathways between the two groups were identified. Combined with transcriptomic data, differential gene analysis was performed to screen out a core gene set associated with tumour metastasis. To achieve accurate prediction of tumour metastasis, this study innovatively constructed a binary classification algorithm (PSO–SVM) integrating particle swarm optimisation (PSO) and support vector machines (SVMs). This model optimises SVM parameters via the PSO algorithm, addressing the limitations of traditional machine learning models such as insufficient accuracy and poor generalization ability in tumour metastasis prediction. Verified by comparison with mainstream machine learning methods, the PSO–SVM model exhibited superior classification performance and successfully identified five key metastasis-related genes: SFN, S100A8, KLF5, ARL4D and TINCR. Furthermore, the expression differences of these genes in the metastatic group were verified at the single-cell level, clarifying their regulatory roles in different cell types and states. Through an innovative analytical strategy integrating single-cell and transcriptomic data, this study elucidated the core molecular mechanisms of SKCM metastasis and key regulatory pathways in the tumour microenvironment, providing potential biomarkers and therapeutic targets for the early diagnosis and targeted treatment of SKCM metastasis. This PSO–SVM–integrated analysis method also offers new insights for research on metastasis mechanisms of other cancers.

1 | Introduction

Melanoma is a highly aggressive malignancy arising from melanocytes, most commonly affecting the skin. Although early detection and surgical resection of primary cutaneous melanoma confer favourable clinical outcomes, the disease is marked by insidious onset, high metastatic potential and delayed diagnosis in the majority of patients, culminating in poor prognosis [1]. Moreover, primary and metastatic melanoma exhibit distinct therapeutic paradigms, underscoring the critical need for timely and accurate identification of metastatic traits to inform clinical decision-making.

In recent years, the proliferation of next-generation high-throughput sequencing technologies has led to the accumulation of vast tumour-associated sequencing datasets, most notably in repositories such as The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) [2]. These resources have profoundly advanced cancer research; however, conventional analytical approaches are increasingly inadequate to interrogate the complexity of large-scale biological datasets. The advent of big data in biology has positioned machine learning as an indispensable tool for extracting actionable insights from bioinformatics data, with transformative success across diverse applications [3]. This interdisciplinary integration

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *IET Systems Biology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

has yielded numerous impactful findings, particularly in binary classification tasks involving biological features such as gene expression profiles, where traditional methods often fail to capture the underlying data patterns, necessitating more sophisticated and efficient analytical frameworks.

In this study, we leveraged transcriptomic and clinical data from cutaneous melanoma patients in the TCGA and GEO databases, complemented by single-cell RNA sequencing (scRNA-Seq) profiling, to develop a predictive model for metastatic status using an integrated machine learning approach combining particle swarm optimisation (PSO) and support vector machines (SVMs). We systematically benchmarked this model against alternative machine learning algorithms, with the aim of facilitating accurate identification of metastatic cutaneous melanoma and thereby providing a robust basis for clinical stratification and staging [4]. Furthermore, through single-cell resolution analysis, we delineated the expression dynamics and functional contributions of prioritized key genes across distinct tumour cell subsets and states, illuminating their mechanistic roles within the tumour microenvironment. Collectively, our findings identify novel potential biomarkers and therapeutic targets to enable early detection and precision intervention for metastatic cutaneous melanoma.

2 | Methods

The overall workflow of this study (Figure 1) encompasses four core stages: data sourcing, feature selection, single-cell profiling and model development. Briefly, we integrated single-cell and bulk RNA-sequencing datasets from public repositories to perform differential gene screening and feature characterization. We then constructed and validated a binary classification model (combining PSO and SVM) for phenotype prediction, followed by single-cell resolution analysis to delineate the expression dynamics of key genes and their functional roles in tumour metastasis.

2.1 | Acquisition and Preprocessing of Transcriptomic Data

Bulk RNA-sequencing data for 470 cutaneous melanoma (SKCM) samples were retrieved from The Cancer Genome Atlas (TCGA; https://portal.gdc.cancer.gov/), including raw count matrices, FPKM (Fragments Per Kilobase of exon model per Million mapped fragments)–normalized expression profiles and corresponding clinical metadata. Among these samples, 103 were primary tumours (TP) and 367 were metastatic tumours (TM). To ensure consistency across downstream analyses, FPKM values were converted to TPM (transcripts per kilobase of exon model per million mapped reads). Additionally, the GSE190113 dataset (103 SKCM samples with TPM-normalized expression and clinical data) was obtained from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/gds/). All datasets underwent standardised preprocessing (including normalization and batch correction) to ensure inter-dataset comparability.

2.2 | Acquisition and Preprocessing of Single-Cell RNA Sequencing Data

The single-cell RNA-sequencing (scRNA-Seq) dataset GSE189889 (GEO) was used, which includes 9 SKCM samples (4 metastatic

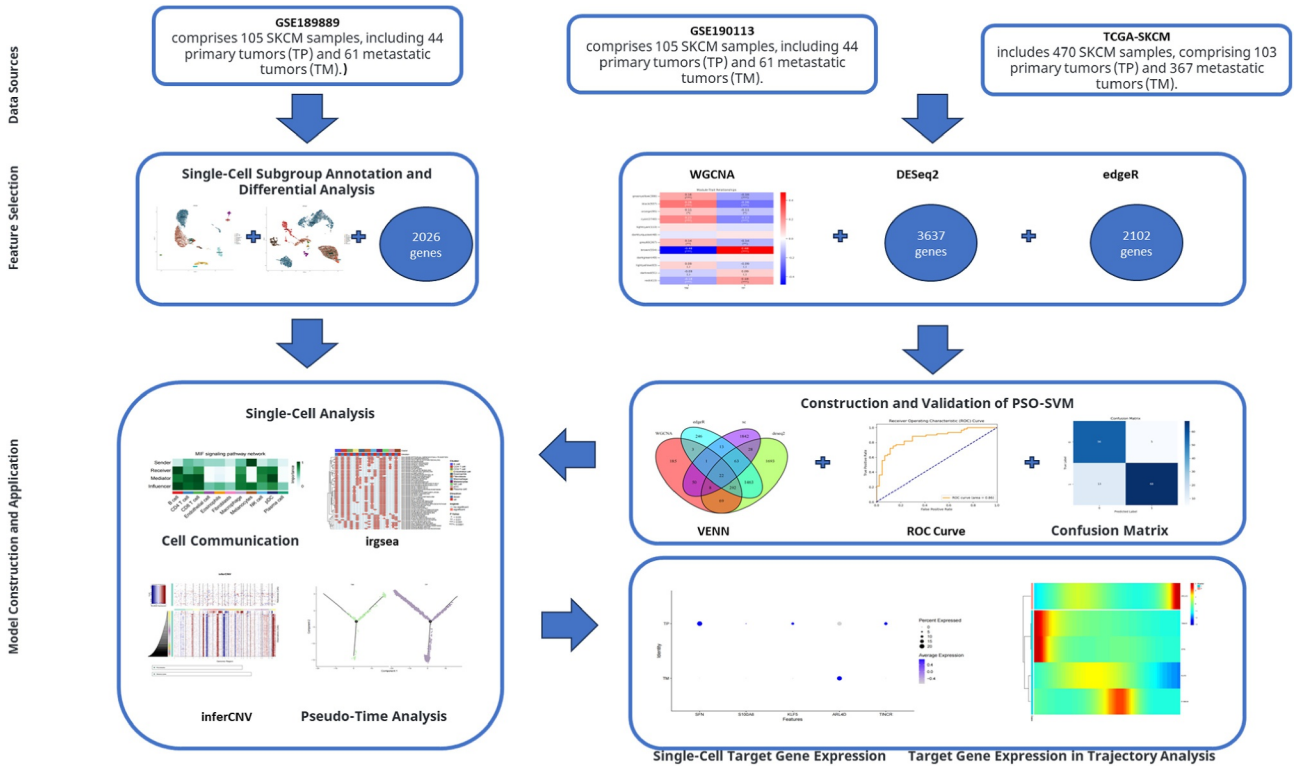

FIGURE 1 | Research flow chart.

and 5 primary). Preprocessing (quality control, normalization and differential expression analysis) was performed using Seurat and Scanpy pipelines. Cell types were annotated manually (distinguishing malignant vs. nonmalignant populations), and the malignant cell identity was validated via inferCNV (copy number variation inference).

2.3 | Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA was employed to identify co-expressed gene modules and their associations with metastatic phenotypes, thereby prioritizing modules and core genes linked to metastasis. Using the R package WGCNA, we input the filtered gene expression matrix and metastasis status (TM vs. TP). An optimal soft-thresholding power (β) was selected to construct a topological overlap matrix (TOM) and a scale-free co-expression network. Similar modules were merged via dynamic tree cutting to complete the WGCNA workflow [5].

2.4 | Particle Swarm Optimisation (PSO) Algorithm

Particle swarm optimisation (PSO) is an efficient, versatile and easy-to-implement global heuristic optimisation algorithm inspired by the collective foraging behaviour of biological swarms (e.g., bird flocks and fish schools). Characterized by simplicity, low computational overhead and minimal tunable parameters [6], PSO is widely utilised for optimisation tasks—particularly hyperparameter search in the current model—owing to its ability to enable information sharing among particles traversing the solution space to search for (near-)optimal solutions [7]. Compared with Bayesian optimisation, PSO features simpler implementation; it is more adept at exploring a broader range of hyperparameter combinations than grid search, and it outperforms random search in efficiency by leveraging historical search information. The particle update process is governed by the following equations:

$$v_{id}^{t+1} = \omega \cdot v_{id}^{t} + c_1 \cdot r_1 \cdot \left(p_{id} - x_{id}^{t}\right) + c_2 \cdot r_2 \cdot \left(p_{gd} - x_{id}^{t}\right) \quad (1)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (2)$$

Here, v_{id}^{t} represents the velocity of a particle in a given dimension, x_{id}^{t} is its position and is the individual best position, p_{gd} is the global best position, ω is the inertia weight, c_1 and c_2 are acceleration coefficients, and r_1 and r_2 are random numbers [8].

2.5 | Construction of the PSO–SVM Model

When optimising support vector machines (SVM) with the particle swarm optimisation (PSO) algorithm, PSO is typically employed to fine-tune SVM hyperparameters, thereby enhancing the model's performance and generalizability [9]. As a supervised learning algorithm suitable for both classification and regression tasks, SVM's performance is highly dependent on the selected hyperparameters [10]. As a metaheuristic approach, PSO simulates the collective foraging behaviour of biological swarms (e.g., bird flocks) to search for optimal solutions in complex spaces. In the present study, the PSO–SVM framework adopted standard default parameters (with the core focus on integrating multi-omics data for melanoma metastasis prediction rather than optimizing the algorithm itself), and the detailed implementation parameters and protocols are specified below.

First, SVM hyperparameters—including the penalty factor (C), kernel type and kernel-specific parameters—must be defined. Herein, the radial basis function (RBF) kernel was selected, with the penalty factor (C) and kernel parameter (γ) optimised via PSO within the ranges of $C \in [0.01, 100]$ and $\gamma \in [0.001, 10]$. The selection of these hyperparameters is critical to SVM performance; however, the vast number of potential combinations necessitates systematic optimisation to identify the optimal set [11].

The PSO workflow involves initializing and updating a swarm of particles, where each particle explores the solution space to find the optimal solution. For SVM optimisation, each particle represents a unique combination of SVM hyperparameters (i.e., specific C and γ values within the predefined ranges). The core PSO parameters used in this study are as follows: number of particles = 30, maximum iterations = 100, inertia weight (ω) = 0.5 and acceleration coefficients $c1 = 1.5$ and $c2 = 1.5$. Each particle is characterized by two key attributes: a position (corresponding to hyperparameter values) and a velocity (governing the manner of position changes during iterations). The fitness of each particle was evaluated using cross-validation metrics (e.g., accuracy and F1 score) of the SVM model trained with its associated hyperparameter combination [12]. Notably, the dataset was split into a training set (70%) and a test set (30%) using stratified sampling to maintain the class distribution balance (primary vs. metastatic melanoma). Fivefold cross-validation was performed on the training set to ensure reliable assessment of model stability during the fitness evaluation process.

The iterative optimisation process of PSO comprises the following key steps:

1. Fitness evaluation: Train an SVM model using the hyperparameter combination encoded by each particle, and assess the model's performance via fivefold cross-validation on the stratified training set.
2. Update particle velocity and position: Adjust each particle's velocity and position based on four factors—its current position, current velocity, personal best position (the optimal solution found by the particle itself) and global best position (the optimal solution found by the entire swarm)—using the predefined PSO parameters ($\omega = 0.5$, $c1 = 1.5$ and $c2 = 1.5$).
3. Update optimal positions: Track and update the personal best and global best positions to reflect the most promising solutions identified in the current iteration.
4. Termination of iterations: Repeat the above steps until either the predefined maximum number of iterations (100) is reached or the swarm converges to a stable solution (meeting the convergence criteria).

During each PSO iteration, changes in particle positions translate to adjustments in SVM hyperparameters, aiming to improve the model's data fitting ability and performance metrics. Ultimately, the PSO algorithm identifies the optimal (or near-optimal) SVM hyperparameter combination, enabling the construction of an SVM model with enhanced generalizability [13]. For fair comparison, benchmark models including random forest (RF) and XGBoost were also trained with their default parameters (consistent with the 'standard parameter' approach of PSO–SVM) and evaluated using the same data splitting strategy and performance metrics, ensuring the consistency and impartiality of the comparative analysis.

2.6 | Single-Cell Research Analysis

To gain mechanistic insights into cutaneous melanoma (SKCM) metastasis, we employed single-cell RNA sequencing (scRNA-Seq) to systematically dissect molecular differences between metastatic and primary tumour cohorts. Detailed methodologies for each analytical component are described below:

Cell type annotation and validation: Following initial scRNA-Seq data processing, we performed comprehensive cell type annotation and validation. Using inferCNV, we verified the purity and consistency of melanoma cell populations extracted from the single-cell dataset. This step is critical to ensure that subsequent differential expression analyses and machine learning model construction are anchored in accurate cell type classification, thereby safeguarding the reliability and reproducibility of our findings.

Data integration and batch effect correction: We integrated tumour cell datasets from metastatic and primary tumour groups, followed by batch effect correction to mitigate biases arising from technical variations. These preprocessing steps ensured data uniformity, laying a robust foundation for downstream statistical analyses and model development [14].

Single-cell–level differential expression analysis in tumour cells: After confirming the accuracy of cell type annotations, we performed differential expression analysis specifically on tumour cells using the FindMarkers function to compare gene expression profiles between metastatic and primary tumour groups. Standard thresholds were applied for DEG calling: (1) min.pct = 0.1 (requiring genes to be expressed in at least 10% of the cells to filter lowly expressed genes), (2) absolute log fold change ($|\log FC| > 0.25$ (to identify biologically meaningful expression changes) and (3) adjusted p-value < 0.05 (Benjamini–Hochberg method for multiple test correction). This analysis identified a set of significant differentially expressed genes (DEGs) associated with the metastatic process, which may serve as potential biomarkers regulating the growth and dissemination of metastatic melanoma.

IRGSEA analysis [15]: We applied IRGSEA (gene set enrichment analysis) to scRNA-Seq data to identify functional gene sets specific to metastatic versus primary tumour groups. The following statistical criteria were used to determine significant enrichment: (1) normalized enrichment score (NES) with $|NES| > 1.5$ (to judge the direction and magnitude of enrichment) and (2) false discovery rate (FDR) < 0.25 (a conventional threshold for GSEA-related analyses). By comparing the enrichment levels and expression patterns of these gene sets, we pinpointed key functional pathways and regulatory mechanisms that may drive melanoma metastasis. This analysis provided critical biological insights to inform feature selection for subsequent machine learning models.

Cell–cell communication network analysis: Cell–cell communication network analysis of scRNA-Seq data was performed using CellChat to unravel intercellular interaction patterns. Statistical testing of communication strength was conducted via permutation tests to calculate p-values for communication probabilities, ensuring the significance of intercellular interactions. Only significant communication events with $p < 0.05$ were retained to guarantee result reliability. By quantifying intercellular co-expression profiles and signalling networks, we identified distinct cell communication signatures between metastatic and primary tumour groups. These signatures included direct cell–cell interactions as well as signalling pathway–mediated regulation, highlighting key signalling cascades that may play pivotal roles in metastatic progression [16].

Pseudotime analysis: Pseudotime analysis was conducted to simulate and compare cellular state transitions across different metastatic stages. Single cells were ordered along a pseudotemporal trajectory corresponding to metastatic progression, and dynamic changes in cell type composition and cellular states were analysed at each stage. This approach captured the temporal dynamics of cellular phenotypes and functions during metastasis, revealing stage-specific characteristics that distinguish metastatic from primary tumour groups [17].

Collectively, these single-cell analyses delineated the key biological features and molecular mechanisms underlying melanoma metastasis at unprecedented cellular resolution. These findings not only advance our understanding of tumour metastatic biology but also provide theoretical and experimental support for precision medicine and personalized therapeutic strategies for SKCM.

3 | Results

3.1 | TCGA Data Differential Analysis

Using DESeq2 [18], we performed differential expression analysis with the filtering criteria set at $p < 0.05$ and $|log2\ Fold\ Change| \geq 1$. Through this analysis, we identified a total of 3637 significantly differentially expressed genes. Subsequently, we conducted differential expression analysis using edgeR [19] with the same filtering criteria, identifying 2102 significantly differentially expressed genes. The cross-validation of these two methods further ensured the reliability of the screening results.

Subsequently, we performed weighted gene co-expression network analysis (WGCNA) following rigorous preprocessing and parameter calibration. First, outlier sample detection was conducted via hierarchical clustering prior to network construction; samples with excessively high topological overlap dissimilarity (> 0.25) were identified and removed to avoid confounding effects on network structure and module detection. Second, the soft-thresholding power (β) was determined based on the scale-free topology criterion: We calculated the scale-free topology fitting index for β values ranging from 1 to 20 using the pickSoftThreshold function and selected the smallest β value that achieved a fitting index of 0.9 (a commonly used threshold for scale-free topology). Ultimately, $\beta = 10$ was adopted, ensuring that the constructed gene co-expression network approximated a scale-free distribution and satisfied the core assumption of WGCNA.

The WGCNA results revealed that the brown module was the gene module most strongly associated with metastatic characteristics, with a correlation coefficient of 0.48. This correlation was highly statistically significant ($p < 0.001$); the significance was calculated via standard Pearson correlation test between the module eigengenes (MEs) of the brown module and the metastatic trait (tumour primary [TP] vs. tumour metastatic [TM]) values across all samples, with Benjamini–Hochberg correction applied to control the false discovery rate (FDR) in multiple testing. This module contained 554 genes, which showed significant expression differences between primary and metastatic melanoma. The identified gene module provided important evidence for feature selection in subsequent machine learning analyses and laid the foundation for exploring the mechanisms of melanoma metastasis.

3.2 | Single-Cell RNA Sequencing Data Differences

We performed rigorous quality control (QC) on melanoma single-cell RNA sequencing (scRNA-Seq) data with the following thresholds to filter out low-quality cells: minimum number of detected features (nFeature_RNA) = 500, minimum number of reads (nCount_RNA) = 1000, maximum mitochondrial gene percentage = 30% (0.3) and doublet removal using the DoubletFinder tool. After filtering for high-quality cells, we standardised the dataset (transcripts per kilobase of exon model per million mapped reads [TPM] conversion was performed using the formula: TPM = (read count/gene length [kb])/(total read count of sample/10^6), followed by $\log_2$ transformation) and conducted dimensionality reduction analyses (principal component analysis [PCA] and uniform manifold approximation and projection [UMAP]). To eliminate technical biases, batch-effect correction was implemented using Harmony software. These analyses revealed substantial heterogeneity in tumour cells between primary and metastatic cohorts. As illustrated in Figure 2, clustering analysis combined with manual annotation enabled precise classification of major cell types, including tumour cells, immune cells and fibroblasts. In both metastatic and primary melanoma groups, melanoma cells were clearly distinguishable from other cell populations (e.g., $CD4^+$ T cells, natural killer [NK] cells and B cells), with their malignant identity validated via inferCNV. Subsequent differential expression analysis using the FindMarkers tool [20] compared gene expression profiles between tumour cell populations from primary and metastatic groups, leading to the identification of 2026 differentially expressed genes (DEGs). These DEGs exhibited significant expression discrepancies between metastatic and primary tumour cells and are

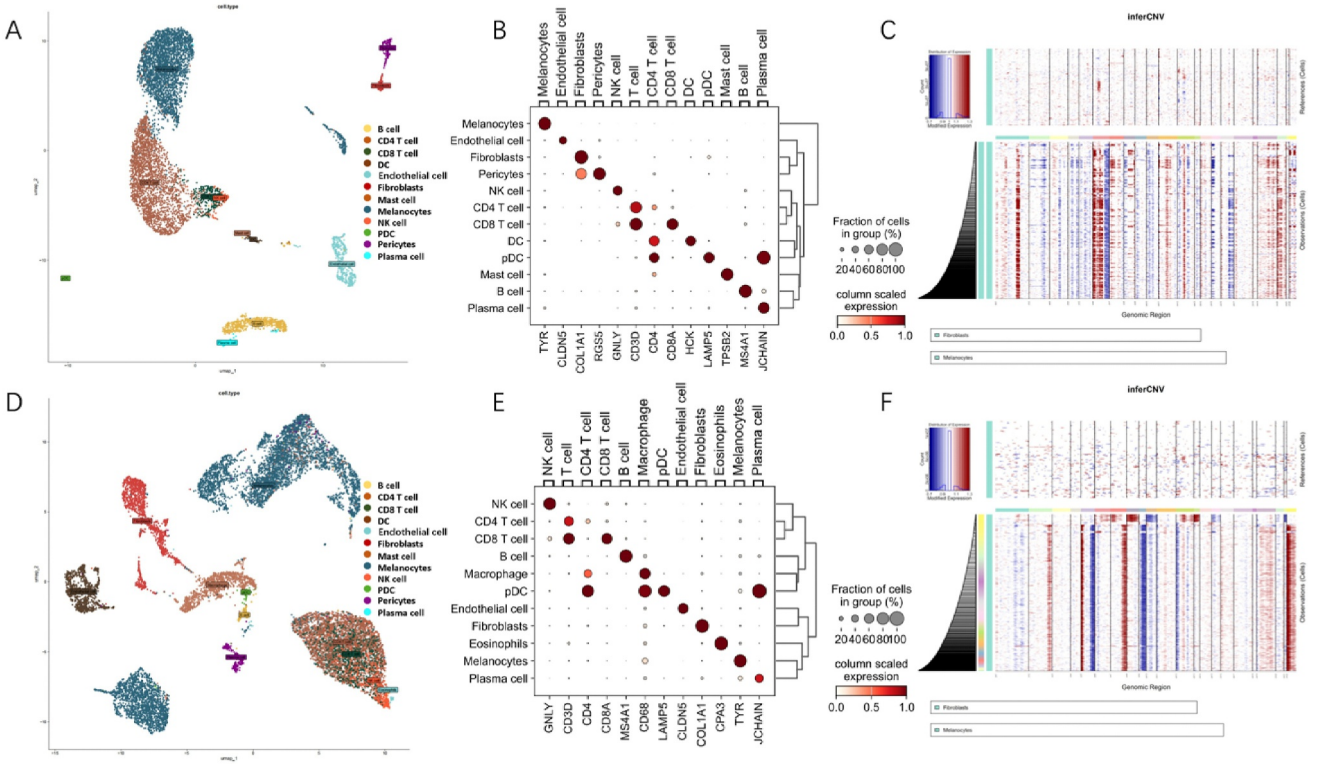

FIGURE 2 | Melanoma cell type annotation and inferCNV validation.

potentially implicated in key biological processes driving tumour progression, such as metastasis, invasion and immune evasion.

3.3 | Integrated Analysis of Functional Pathways, Regulatory Mechanisms and Cell–Cell Interaction Patterns

To systematically dissect the gene set enrichment characteristics and dynamic regulatory rules of intercellular communication networks during cutaneous melanoma metastasis, this study performed IRGSEA analysis on samples from metastatic and primary groups and systematically compared the differences in cell interaction patterns between the two pathological stages by integrating cell communication sequencing data (Figures 3–5). The IRGSEA results (including ssGSEA and RRA dual validation, Figure 3) demonstrated that the enrichment of key signalling pathways in different cell subsets exhibited significant stage specificity and functional relevance: the hallmark epithelial–mesenchymal transition (EMT) pathway was significantly enriched in both primary and metastatic lesions, confirming its core regulatory role throughout the entire process of melanoma initiation, invasion and metastasis; in contrast, the HALLMARK-UV-RESPONSE-DN pathway and the hallmark interferon–gamma response pathway were specifically highly expressed only in the metastatic group, suggesting that these pathways may be involved in mediating biological behaviours unique to the metastatic process (e.g., immune escape and distant colonization). Notably, the enrichment level of CD8+ T cells in the metastatic group was significantly higher than that in the primary group, indicating that this cell subset plays a critical role in regulating the immune microenvironment during tumour metastasis.

To further clarify the intrinsic association between the aforementioned pathway enrichment and cell functions, complementary validation was conducted through intercellular communication network analysis (Figures 4,5): Compared with the primary group, the activity of the MIF signalling pathway in CD8+ T cells, macrophages and fibroblasts was significantly upregulated in the metastatic group, and the intensity of intercellular crosstalk among immune cell subsets (NK cells, CD4+ T cells and CD8+ T cells) was remarkably enhanced, suggesting a more complex interaction network within the immune microenvironment during the metastatic stage. Meanwhile, the key downstream ligand–receptor (L-R) pairs of the MIF pathway (MIF-(CD74+CXCR4) and MIF-(CD74+CD44)) exhibited characteristics of high active expression in the metastatic group, which may regulate tumour invasion and metastasis-related phenotypes by mediating transcellular signal transduction among tumour cells, immune cells and stromal cells. Collectively, these results indicate that the MIF signalling pathway can act as a core regulatory node driving melanoma metastatic progression by coordinating the functional interactions between immune cells and fibroblasts.

In summary, through the integration of gene set enrichment analysis and intercellular communication network dissection, this section systematically reveals the cell type–specific functional characteristics, stage-specific molecular pathway regulatory networks and abnormally activated intercellular interaction patterns during melanoma metastasis. It provides a novel perspective for in-depth elucidation of the molecular mechanisms underlying melanoma metastasis and simultaneously offers important experimental basis and theoretical support for the screening of precise diagnostic biomarkers and the development of targeted therapeutic targets for metastatic cutaneous melanomas on both the metastatic and primary groups.

3.4 | Multisource Integration to Refine Candidate Genes for Metastasis Prediction

We used multiple analytical methods to screen for key gene sets associated with tumour metastasis and constructed a PSO–SVM model to predict melanoma metastasis. During the screening process, DESeq2 and edgeR analyses identified 3637 and 2102 differentially expressed genes, respectively. WGCNA analysis (Figure 6A) yielded 554 core genes, and single-cell data analysis identified 2026 significantly differentially expressed genes. Finally, we used a Venn diagram (Figure 6B) to identify the intersection of these datasets, narrowing down to 22 candidate genes as the gene dataset for the model.

3.5 | Model Training and Validation Results

The training and validation results demonstrated that the PSO–SVM model achieved optimal predictive performance when selecting five specific genes (*SFN, S100A8, KLF5, ARL4D* and *TINCR*). The ROC–AUC value was 0.8651, as shown in Figure 7A, significantly outperforming other mainstream machine learning methods such as random forest and XGBoost, as confirmed in Figure 7C. These genes exhibited significant differential expression in the metastatic group, suggesting their potential critical roles in the tumour metastasis process.

Figure 8D–F illustrate that these five genes (*SFN, S100A8, KLF5, ARL4D* and *TINCR*) showed high expression at the single-cell level, particularly in specific subpopulations of the metastatic group, such as CD8+ T cells and fibroblasts. These genes may play pivotal roles in regulating the tumour microenvironment and promoting cell migration and invasion. These results not only validated the efficiency and robustness of the PSO–SVM model but also indicated that the genes it identified hold potential clinical value, providing significant evidence for the early diagnosis and targeted treatment of melanoma.

3.6 | Dynamic Changes in Cellular Phenotype and Function

As shown in Figure 7A and B, the trajectory of tumour cells evolving from the primary to the metastatic state exhibited a 'Y' shaped branching structure, revealing potential key turning points during the metastasis process. The analysis of target genes in Figure 7C indicated that *ARL4D* showed a significant increase in the late metastatic phase, suggesting it may be a

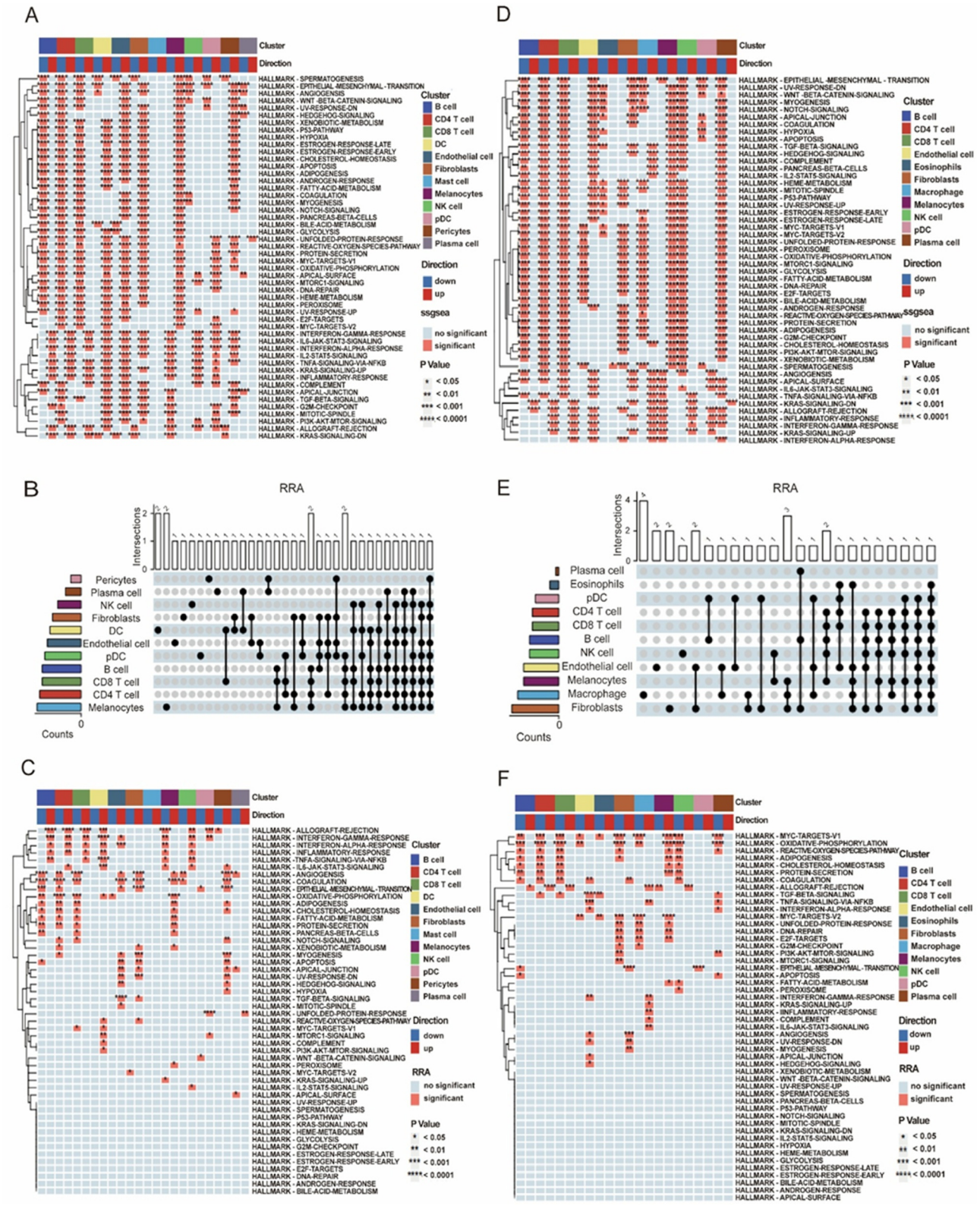

FIGURE 3 | IRGSEA analysis results of the single-cell metastasis group and the primary group.

key gene. Meanwhile, *KLF5* and *S100A8* displayed dynamic fluctuations, indicating their association with metastasis, whereas SFN and TINCR showed weaker changes but slight upregulation in the late phase. These findings provide valuable insights into the mechanisms of tumour metastasis and potential clinical applications.

4 | Conclusions

By integrating machine learning with single-cell RNA sequencing (scRNA-Seq) technology, this study elucidated the core molecular mechanisms driving cutaneous melanoma (SKCM) metastasis and identified five key regulatory genes, namely SFN, S100A8,

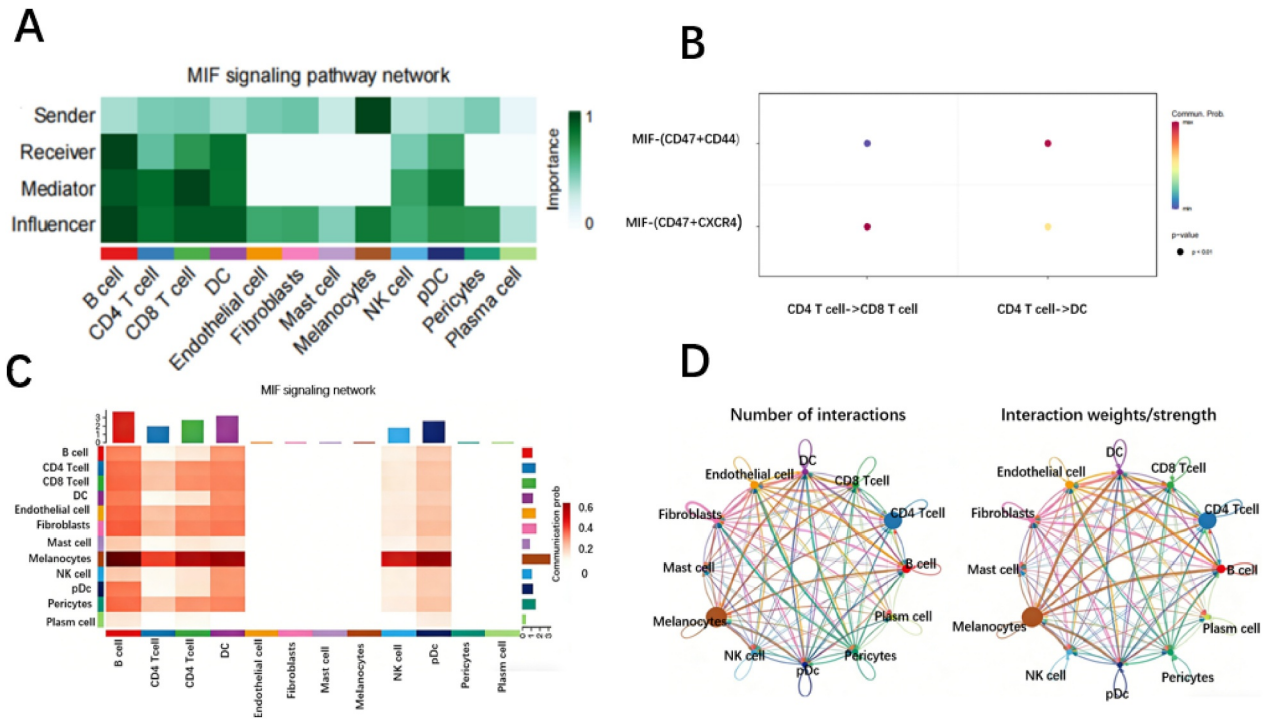

FIGURE 4 | Cell-cell communication analysis of the primary group.

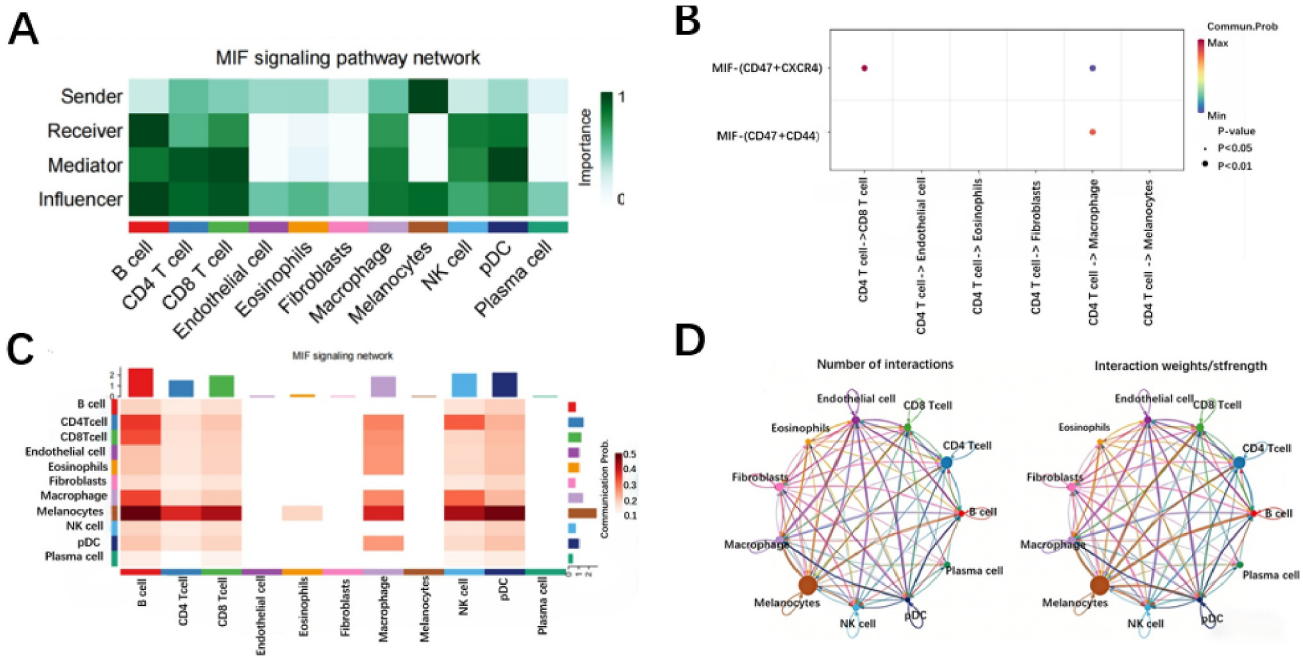

FIGURE 5 | Cell-cell communication analysis results: Primary tumor group.

KLF5, ARL4D and TINCR. The significant differential expression of these genes in metastatic lesions implies their potential critical regulatory roles in SKCM progression, which is consistent with the prediction results of the PSO–SVM model established in this study. Notably, the PSO–SVM model outperformed other mainstream machine learning methods in tumour metastasis prediction [21], further confirming the reliability of these five genes as metastatic signature biomarkers. Moreover, single-cell resolution analyses of cell communication and trajectory uncovered the remodelling of intercellular communication networks among key cell subsets in the metastatic group, with special emphasis on the central regulatory role of the MIF signalling pathway in the tumour microenvironment. Collectively, this study validated the clinical potential of these key genes from multiple perspectives, laying a solid theoretical foundation for subsequent experimental investigations and clinical translation.

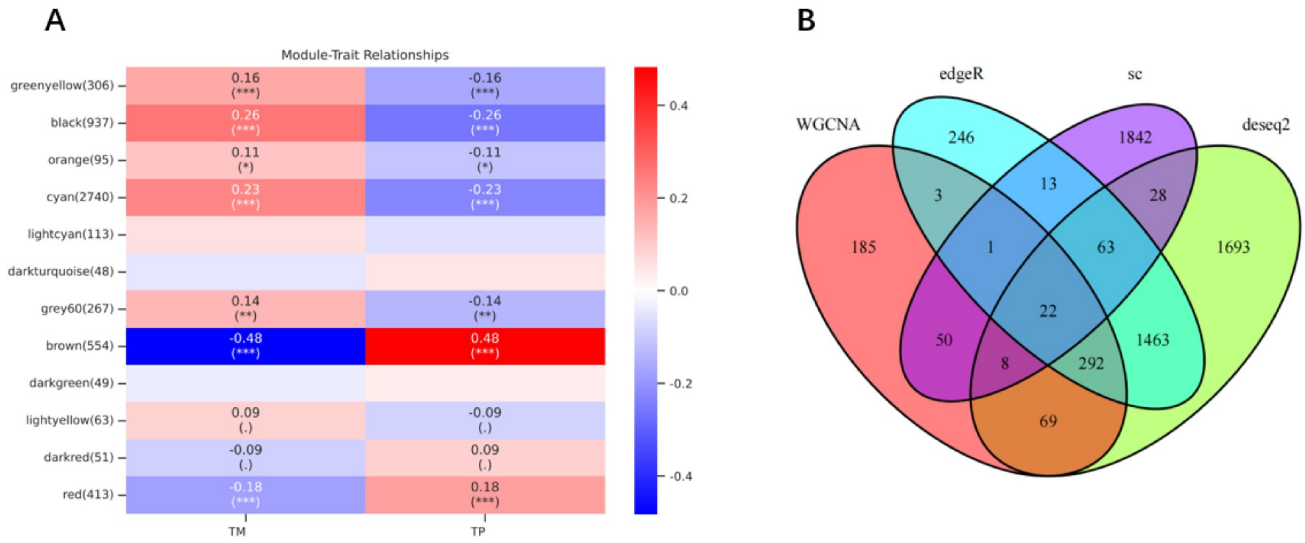

FIGURE 6 | Integration of multi-method analyses to identify candidate genes.

Despite substantial progress in deciphering the metastatic mechanisms of SKCM, this study has several limitations that warrant attention. First, the smaller sample size of the utilised datasets may impair the generalizability of our findings. Future studies should integrate larger scale-independent cohorts to enhance the robustness and applicability of the proposed model. Second, the parameter optimisation and performance validation of the PSO–SVM model were exclusively based on the current dataset; therefore, its scalability and generalization capacity need further verification in additional independent cohorts. Finally, the functional roles of the identified key genes lack in-depth biological validation. Targeted mechanistic studies using in vitro cell lines and in vivo animal models are thus urgently required to clarify their underlying regulatory pathways.

Five SKCM metastasis–related genes (SFN, S100A8, KLF5, ARL4D and TINCR) were screened via machine learning. To reinforce the reliability of our conclusions, we conducted a direct comparative analysis of their functions and associations with SKCM metastasis, incorporating findings from existing literature as follows:

As a member of the calcium-binding protein family, S100A8 has been well documented to promote tumour metastasis, which is highly consistent with our results. Sun et al. [22] demonstrated that S100A8 facilitates melanoma metastasis by regulating the immunosuppressive function of myeloid-derived suppressor cells (MDSCs) to reshape the tumour immune microenvironment. This aligns with our observation that S100A8 is significantly upregulated in the SKCM metastatic group and involved in tumour microenvironment regulatory pathways, further solidifying its role as a core metastasis-promoting factor in SKCM.

ARL4D, a member of the small GTPase family, has primarily been investigated in the context of breast cancer, and this study is the first to establish its close association with SKCM metastasis. Yagin, B et al. [23] confirmed that ARL4D interacts with ARF6 to regulate cytoskeletal rearrangement, thereby markedly enhancing the migration and invasion capabilities of breast cancer cells. Our findings show that ARL4D is abnormally upregulated in SKCM metastatic lesions, suggesting it may promote SKCM metastasis through a conserved cytoskeletal regulatory mechanism. This provides a novel direction for exploring the function of ARL4D in melanoma.

The role of SFN (also known as 14-3-3σ) in tumours is tissue-specific, and our results complement previous research on its function in breast and prostate cancers. Kanematsu et al. [24] reported that SFN exerts anti-metastatic effects in breast and prostate cancers by inhibiting epithelial–mesenchymal transition (EMT) and cancer cell invasiveness. However, our study detected significant abnormal expression of SFN in the SKCM metastatic group, indicating that it may adopt a distinct regulatory mode in melanoma—potentially participating in metastatic progression by targeting specific downstream molecules. This discrepancy offers an important clue for investigating the functional heterogeneity of SFN across different tumour types.

As a zinc finger transcription factor, KLF5 has been validated as a metastasis-promoting factor in various tumours, which is consistent with our findings in SKCM. Zhang et al. [25] showed that KLF5 accelerates the metastasis of breast and colorectal cancers by promoting cell proliferation and EMT. Our study further confirms that KLF5 is highly expressed in SKCM metastatic lesions and participates in core metastatic regulatory pathways, suggesting it may drive SKCM progression through a conserved EMT-mediated mechanism. This provides a universal basis for cross-tumour metastatic mechanism research.

In contrast to the other four genes, research linking TINCR to tumour metastasis remains limited, and our study provides preliminary evidence for its association with SKCM metastasis. Existing literature indicates that TINCR is primarily involved in tumour initiation and progression by regulating epigenetic modifications and promoting cell differentiation [26], but direct evidence for its role in metastasis is scarce. Our findings reveal significant differential expression of TINCR in the SKCM metastatic group, where it acts synergistically with other metastasis-related genes, implying it may indirectly regulate SKCM metastasis by modulating epigenetic factors. This discovery fills

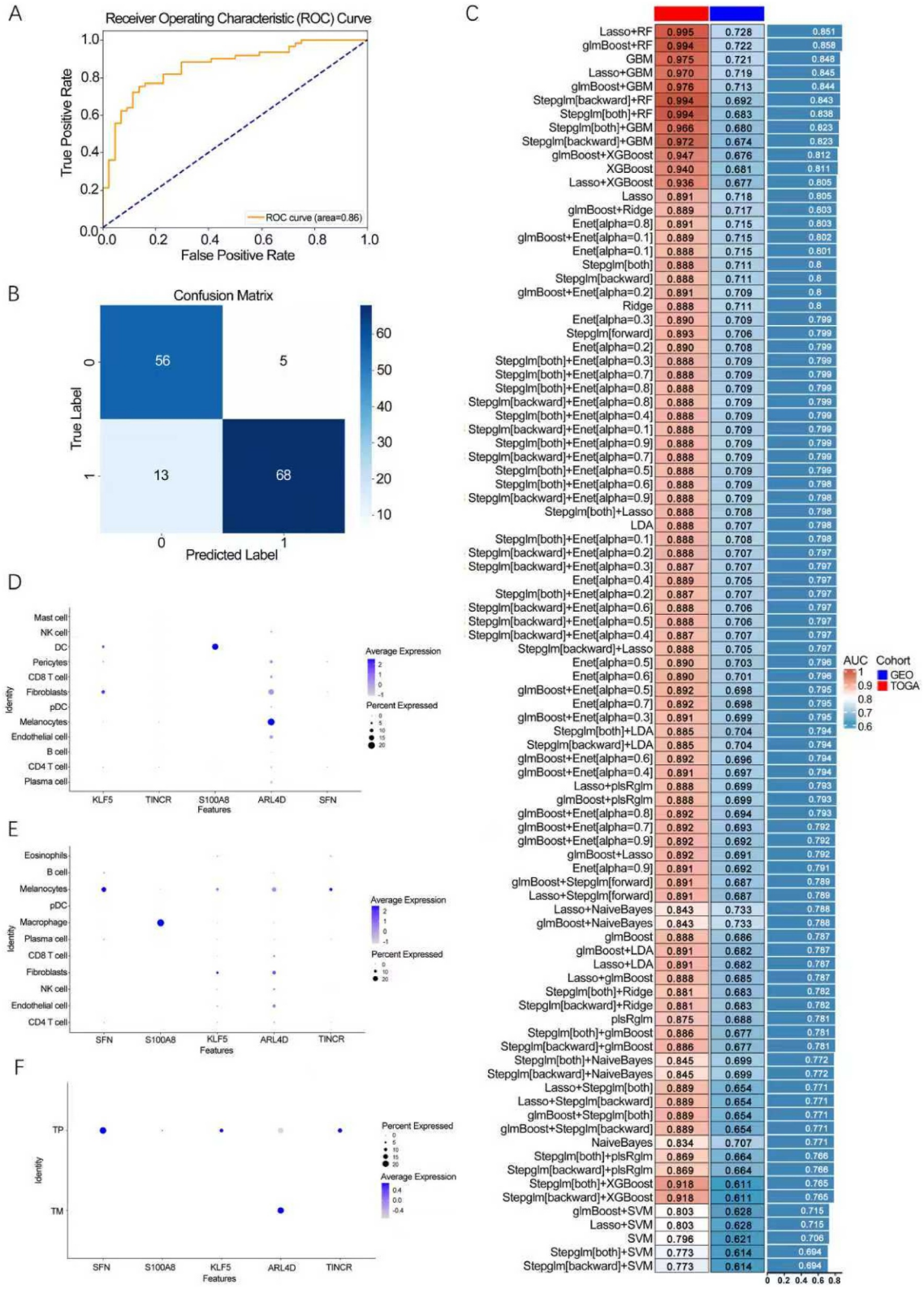

FIGURE 7 | PSO–SVM model construction, validation, comparison and the expression of target genes.

a gap in TINCR research in melanoma metastasis and identifies it as a key target for future mechanistic investigations.

Future studies should further explore the universality of these genes across other tumour types and their roles in immune regulation to refine the molecular network underlying tumour immune escape and invasive metastasis. Additionally, developing targeted therapies against these key genes and performing integrated multi-omics analyses will provide stronger support for personalized tumour diagnosis and treatment. These efforts

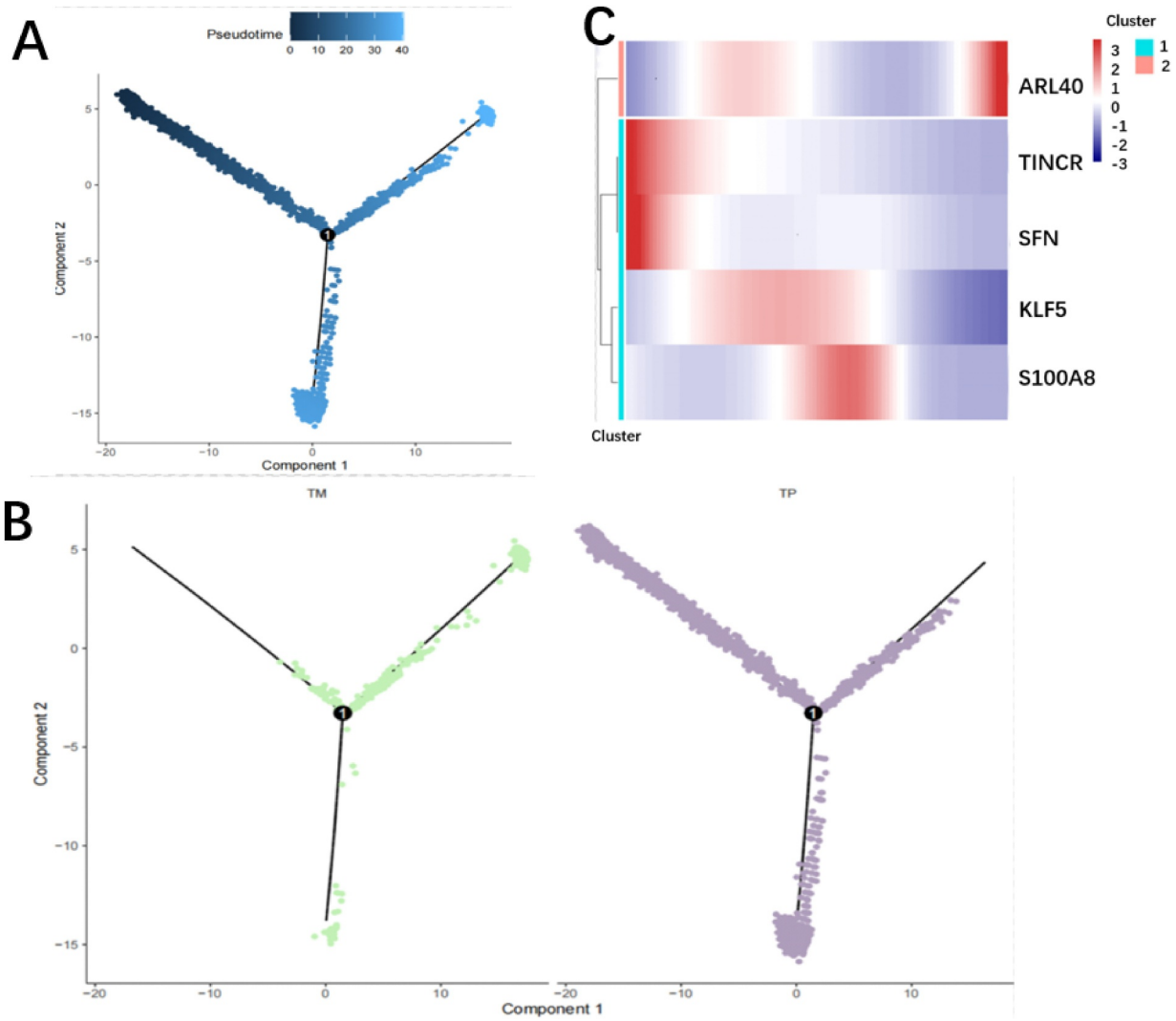

FIGURE 8 | Pseudo-time analysis results and the expression of target genes.

will not only deepen our understanding of tumour biology but also offer critical directions for the clinical translation of precision medicine in melanoma, bringing new hope to patients.

Author Contributions

Zhiwei Liao: conceptualization, formal analysis, writing – original draft. **Weiming Chen:** data curation, investigation. **Yingdi He:** visualization, validation. **Yichen Zheng:** software, data curation. **Xiaonan Chen:** writing – review and editing, investigation. **Han Shen:** supervision, project administration, funding acquisition, writing – review and editing.

Funding

This study was supported by the Natural Science Foundation of Guangdong Province (Grant 2018A030313114), the Special Project in Key Areas of Guangdong Province's General Colleges and Universities (Grant 2024ZDZX2082), and the National Natural Science Foundation of China (Grant 31300737).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. The analysis scripts will be made publicly available after the publication of follow-up studies.

References

1. G. V. Long, S. M. Swetter, A. M. Menzies, J. E. Gershenwald, and R. A. Scolyer, "Cutaneous Melanoma," *Lancet.* 402, no. 10400 (August 2023): 485–502, https://doi.org/10.1016/s0140-6736(23)00821-8.

2. T. Hu, N. Chitnis, D. Monos, and A. Dinh, "Next-Generation Sequencing Technologies: An Overview," *Human Immunology* 82, no. 11 (November 2021): 801–811, https://doi.org/10.1016/j.humimm.2021.02.012.

3. J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, "A Guide to Machine Learning for Biologists," *Nature Reviews Molecular Cell Biology* 23, no. 1 (January 2022): 40–55, https://doi.org/10.1038/s41580-021-00407-0.

4. A. AlMohimeed, M. Shehata, N. El-Rashidy, S. Mostafa, A. Samy Talaat, and H. Saleh, "ViT-PSO-SVM: Cervical Cancer Predication Based on Integrating Vision Transformer With Particle Swarm Optimisation and Support Vector Machine," *Bioengineering (Basel)* 11, no. 7 (July 2024): 729, https://doi.org/10.3390/bioengineering11070729.

5. Q. Sun, Z. Wang, H. Xiu, N. He, M. Liu, and L. Yin, "Identification of Candidate Biomarkers for GBM Based on WGCNA," *Scientific Reports* 14, no. 1 (May 2024): 10692, https://doi.org/10.1038/s41598-024-61515-3.

6. S. Pervaiz, Z. Ul-Qayyum, W. H. Bangyal, L. Gao, and J. Ahmad, "A Systematic Literature Review on Particle Swarm Optimisation Techniques for Medical Diseases Detection," *Computational and Mathematical Methods in Medicine*, Retraction in: Comput Math Methods Med. 2023 Aug 2;2023:9825640 2021 (September 2021): 5990999, https://doi.org/10.1155/2021/5990999.

7. Y. Shi and R. Eberhart, "A Modified Particle Swarm Optimizer," in *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation Proceedings* (IEEE World Congress on Computational Intelligence, 1998).

8. B. Borowska, "Learning Competitive Swarm Optimisation," *Entropy* 24, no. 2 (February 2022): 283, https://doi.org/10.3390/e24020283.

9. G. F. S. Silva, T. P. Fagundes, B. C. Teixeira, and A. D. P. Chiavegatto Filho, "Machine Learning for Hypertension Prediction: A Systematic Review," *Current Hypertension Reports* 24, no. 11 (November 2022): 523–533, https://doi.org/10.1007/s11906-022-01212-6.

10. M. Radak, H. Y. Lafta, and H. Fallahi, "Machine Learning and Deep Learning Techniques for Breast Cancer Diagnosis and Classification: A Comprehensive Review of Medical Imaging Studies," *Journal of Cancer Research and Clinical Oncology* 149, no. 12 (September 2023): 10473–10491, https://doi.org/10.1007/s00432-023-04956-z.

11. S. Jin, D. Qin, B. S. Liang, et al., "Machine Learning Predicts cancer-associated Deep Vein Thrombosis Using Clinically Available Variables," *International Journal of Medical Informatics* 161 (May 2022): 104733, https://doi.org/10.1016/j.ijmedinf.2022.104733.

12. C. Y. Lin, S. M. Guo, J. J. Lien, et al., "Combined Model Integrating Deep Learning, Radiomics, and Clinical Data to Classify Lung Nodules at Chest CT," *Radiologia Medica, La* 129, no. 1 (January 2024): 56–69, https://doi.org/10.1007/s11547-023-01730-6.

13. L. K. Alsaykhan and M. S. Maashi, "A Hybrid Detection Model for Acute Lymphocytic Leukemia Using Support Vector Machine and Particle Swarm Optimization (SVM–PSO)," *Scientific Reports* 14, no. 1 (October 2024): 23483, https://doi.org/10.1038/s41598-024-74889-1.

14. L. Qin, G. Zhang, S. Zhang, and Y. Chen, "Deep Batch Integration and Denoise of Single-Cell RNA-Seq Data," *Advanced Science (Weinheim)* 11, no. 29 (August 2024): e2308934, https://doi.org/10.1002/advs.202308934.

15. C. Fan, F. Chen, Y. Chen, et al., "irGSEA: The Integration of Single-cell Rank-Based Gene Set Enrichment Analysis," *Briefings in Bioinformatics* 25, no. 4 (May 2024): bbae243, https://doi.org/10.1093/bib/bbae243.

16. S. Jin, C. F. Guerrero-Juarez, L. Zhang, et al., "Inference and Analysis of Cell-Cell Communication Using CellChat," *Nature Communications* 12, no. 1 (February 2021): 1088, https://doi.org/10.1038/s41467-021-21246-9.

17. X. Qiu, Q. Mao, Y. Tang, et al., "Reversed Graph Embedding Resolves Complex Single-Cell Trajectories," *Nature Methods* 14, no. 10 (October 2017): 979–982: Epub 2017 Aug 21. PMID: 28825705; PMCID: PMC5764547, https://doi.org/10.1038/nmeth.4402.

18. M. I. Love, W. Huber, and S. Anders, "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq2," *Genome Biology* 15, no. 12 (2014): 550, https://doi.org/10.1186/s13059-014-0550-8.

19. M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data," *Bioinformatics* 26, no. 1 (2010): 139–140, https://doi.org/10.1093/bioinformatics/btp616.

20. T. Stuart, A. Butler, P. Hoffman, et al., "Comprehensive Integration of Single-Cell Data," *Cell* 177, no. 7 (2019): 1888–1902.e21, https://doi.org/10.1016/j.cell.2019.05.031.

21. S. Sarkar and K. Mali, "Firefly-SVM Predictive Model for Breast Cancer Subgroup Classification With Clinicopathological Parameters," *Digital Health* 9 (October 2023): 20552076231207203, https://doi.org/10.1177/20552076231207203.

22. Z. Sun, B. Zeng, D. Liu, Q. Zhao, J. Wang, and H. Rosie Xing, "S100A8 Transported by SEC23A Inhibits Metastatic Colonization via Autocrine Activation of Autophagy," *Cell Death & Disease* 11, no. 8 (2020): 650, https://doi.org/10.1038/s41419-020-02835-w.

23. B. Yagin, F. H. Yagin, C. Colak, F. Inceoglu, S. Kadry, and J. Kim, "Cancer Metastasis Prediction and Genomic Biomarker Identification Through Machine Learning and Explainable Artificial Intelligence in Breast Cancer Research," *Diagnostics* 13, no. 21 (2023): 3314, https://doi.org/10.3390/diagnostics13213314.

24. S. Kanematsu, K. Yoshizawa, N. Uehara, et al., "Sulforaphane Inhibits the Growth of KPL-1 Human Breast Cancer Cells in Vitro and Suppresses the Growth and Metastasis of Orthotopically Transplanted KPL-1 Cells in Female Athymic Mice," *Oncology Reports* 26, no. 3 (2011): 603–608, https://doi.org/10.3892/or.2011.1311.

25. J. Qin, Z. Zhou, W. Chen, et al., "BAP1 Promotes Breast Cancer Cell Proliferation and Metastasis by Deubiquitinating KLF5," *Nature Communications* 6, no. 1 (2015): 8471, https://doi.org/10.1038/ncomms9471.

26. Z. Y. Zhang, Y. X. Lu, Z. Y. Zhang, et al., "Loss of TINCR Expression Promotes Proliferation, Metastasis Through Activating EpCAM Cleavage in Colorectal Cancer," *Oncotarget* 7, no. 16 (2016): 22639–22649, https://doi.org/10.18632/oncotarget.8141.