Heliyon 10 (2024) e27357

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification and verification of AK4 as a protective immune-related biomarker in adipose-derived stem cells and breast cancer

Wei Lu [a,1], Zhenyu Yang [a,1], Mengjie Wang [a], Shiqi Li [b], Hui Bi [c], Xiaonan Yang [a,*]

[a] *Department of Hemangioma and Vascular Malformation, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100144, China*
[b] *Chinese Academy of Medical Sciences & Peking Union Medical College, 4+4 M.D. Program, Beijing, 100144, China*
[c] *Department of Internal Medicine, Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100144, China*

ARTICLE INFO

Keywords:
Adipose-derived stem cell
Breast cancer
Fat grafting
Adenylate kinase 4
Immune landscape
Bioinformatics

ABSTRACT

Background: Breast cancer (BC) remains the most common cancer among women, and novel post-surgical reconstruction techniques, including autologous fat transplantation, have emerged. While Adipose-derived stem cells (ADSCs) are known to impact the viability of fat grafts, their influence on breast cancer progression remains unclear. This study aims to elucidate the genetic interplay between ADSCs and breast cancer, focusing on potential therapeutic targets.

Methods: Using the GEO and TCGA databases, we pinpointed differentially expressed (DE) mRNAs, miRNAs, lncRNAs, and pseudogenes of ADSCs and BC. We performed functional enrichment analysis and constructed protein-protein interaction (PPI), RNA binding protein (RBP)-pseudogene-mRNA, and lncRNA-miRNA-transcription factor (TF)-gene networks. Our study delved into the correlation of AK4 expression with 33 different malignancies and examined its impact on prognostic outcomes across a pan-cancer cohort. Additionally, we scrutinized immune infiltration, microsatellite instability, and tumor mutational burden, and conducted single-cell analysis to further understand the implications of AK4 expression. We identified novel sample subtypes based on hub genes using the ConsensusClusterPlus package and examined their association with immune infiltration. The random forest algorithm was used to screen DE mRNAs between subtypes to validate the powerful prognostic prediction ability of the artificial neural network.

Results: Our analysis identified 395 DE mRNAs, 3 DE miRNAs, 84 DE lncRNAs, and 26 DE pseudogenes associated with ADSCs and BC. Of these, 173 mRNAs were commonly regulated in both ADSCs and breast cancer, and 222 exhibited differential regulation. The PPI, RBP-pseudogene-mRNA, and lncRNA-miRNA-TF-gene networks suggested AK4 as a key regulator. Our findings support AK4 as a promising immune-related therapeutic target for a wide range of malignancies. We identified 14 characteristic genes based on the AK4-related cluster using the random forest algorithm. Our artificial neural network yielded excellent diagnostic performance in the testing cohort with AUC values of 0.994, 0.973, and 0.995, indicating its ability to distinguish between breast cancer and non-breast cancer cases.

* Corresponding author.
E-mail address: yxnan@aliyun.com (X. Yang).
[1] Contributed equally.

https://doi.org/10.1016/j.heliyon.2024.e27357
Received 13 August 2023; Received in revised form 5 February 2024; Accepted 28 February 2024
Available online 15 March 2024

Conclusions: Our research sheds light on the dual role of ADSCs in BC at the genetic level and identifies AK4 as a key protective mRNA in breast cancer. We found that AK4 significantly predicts cancer prognosis and immunotherapy, indicating its potential as a therapeutic target.

1. Introduction

Breast cancer is a leading cause of cancer-related morbidity and mortality in women worldwide [1], and treatment options include various surgical and non-surgical interventions. Among these, autologous fat transplantation, also known as fat grafting or lipofilling, has emerged as an innovative method to reconstruct breast shape following oncological surgery [2]. This procedure involves transferring a patient's adipose tissue to the breast area, yielding aesthetically pleasing outcomes and reduced complications at donor sites. Despite its advantages, long-term fat graft survival remains uncertain due to inadequate vascularization, leading to fat apoptosis and necrosis [3].

To address this challenge, the deployment of adipose-derived stem cells (ADSCs) has been proposed. ADSCs, initially characterized as multipotent mesenchymal stem cells, have exhibited potential in various therapeutic contexts, including treating inflammatory and autoimmune diseases, as well as neurodegenerative conditions, by secreting cytokines and growth factors [4]. ADSCs have demonstrated the potential to enhance fat graft retention [5–7], as well as to promote beneficial effects on scars [8] and facilitate the repair of damaged skin resulting from radiation therapy [9]. Surgeons have developed cell-assisted lipotransfer (CAL) by mixing in vitro cultured ADSCs into lipoaspirate, which has been shown to improve fat graft survival in breast augmentation [5]. However, their effects on breast cancer remain a subject of debate. Hence, this study primarily focuses on unveiling potential risks and the inherent mechanisms tied to the utilization of ADSCs in autologous fat transplantation for individuals with breast cancer.

Research on ADSCs' effects on breast cancer has yielded conflicting results, with some studies suggesting that they promote tumor growth [10] while others indicate that they suppress it [11]. ADSCs can stimulate vascularization through cytokine secretion [12], and under hypoxic conditions, they exhibit an increased expression of VEGF [13]. Research has indicated that ADSCs, when subjected to elements within the tumor microenvironment, can evolve into carcinoma-associated fibroblasts [14], thereby enhancing tumor growth. Additionally, it has been established that ADSCs play a role in fostering an inflammatory milieu within tumors [15], which further exacerbates the progression of cancerous activities. Furthermore, there is a theory suggesting that cancer stem cells, potentially originating from indigenous stem cells, can accelerate tumor proliferation and invasiveness by inducing epithelial-mesenchymal transition [16]. Contrarily, studies using xenograft models propose that the synergy between transplanted ADSCs and dormant cancer cells might not trigger cancer resurgence. This is because dormant cancer cells exhibit higher resistance to apoptosis and their survival is independent of stromal support or vascular architecture [17,18]. In fact, when transplanted into a mouse breast cancer model, ADSCs were found to inhibit tumor growth by inducing tumor cell apoptosis through Poly ADP ribose polymerase cleavage [19]. There is currently no evidence of breast malignant transformation in clinical studies examining the safety of fat grafting [6,12].

Considering these contradictory results and the possible hazard of breast cancer re-emergence, more comprehensive studies are essential to elucidate the dangers and underlying processes linked to employing ADSCs in autologous fat grafting. Our research entailed a thorough examination of gene expression data sourced from the GEO and TCGA databases, aiming to pinpoint mRNAs, miRNAs, lncRNAs, and pseudogenes that are co-expressed in ADSCs and breast cancer. To better understand the shared gene expression patterns and mechanisms, we further identified lncRNA-miRNA-TF-gene and mRNA-RBP-pseudogene networks. Immune Infiltration and pan-cancer analysis suggests potential targets for preventing breast cancer recurrence in fat grafting procedures, including AK4, the corresponding transcription factor (TF) ATF3, and RNA binding protein (RBP) MBNL2. Moreover, categorizing breast cancer specimens into clusters according to traits associated with ADSCs enhances our study, offering a solid structure for the corroboration of our results.

2. Materials and methods

2.1. Data acquisition and preprocessing

We compared the RNA expression profiles of ADSCs and adipose tissue using microarray data obtained from ADSCs before and after adipogenic differentiation. Three datasets were sourced from the GEO database: GSE61302 [20] for mRNA profiling, GSE72429 for miRNA profiling, and GSE57593 [21] for ncRNA profiling. Gene expression values for GSE61302 were derived from Affymetrix CEL files by employing the gcrma package in R software. Simultaneously, the normalizeBetweenArrays algorithm within the limma package [22] was employed for background adjustment and data standardization in GSE72429 and GSE57593. We assessed the integrity of the normalized array data utilizing the arrayQualityMetrics package, which bases its evaluation on the mean absolute deviation of the M value across each array pair. Ultimately, principal component analysis (PCA) was employed to scrutinize differences between groups and identify duplications within sample groups [23].

For our analysis of breast cancer, we used the 'TCGAbiolinks' R package [24] to obtain sequence data, including genomic, epigenomic, and transcriptomic profiles, as well as corresponding clinical information [25]. A total of 1226 samples were chosen for in-depth analysis, from which the anticipated counts and normalized expression rates of genes, expressed in transcripts per million (TPM), were extracted. To reduce the numerical span of TPM values, we presented gene expressions as $\log_2(\text{TPM} + 1)$ and merged them into a single matrix table. We downloaded 90, 86 and 121 breast cancer samples from the GSE20711 [26], GSE15852 [27], and

GSE42568 [28] datasets, respectively, from the GEO database. Datasets from GEO database were used as verification sets, while the TCGA dataset served as the training set.

2.2. Differential mRNA, miRNA, lncRNA and pseudogene analysis

We obtained normalized mRNA expression matrix files from GSE61302 and miRNA expression matrix files from GSE72429. Profiles for lncRNA and pseudogenes were sourced from GSE57593, with gene symbols acquired via probe conversion, contingent on the annotation specifics of the platform. In instances where multiple probes corresponded to a single gene symbol, the final expression level was determined by averaging the expression levels of the associated mRNAs or noncoding RNAs [29]. Differential expression analysis was conducted utilizing the limma package, applying a threshold of an adjusted P value < 0.05 and $|\log2-\text{fold change}| > 1$. To identify DE transcripts in BC, we used the raw counts data of transcriptomic profiles (mRNA, miRNA, lncRNA, pseudogene) from TCGA database, and applied DESeq2 package [30] in R. The DE mRNAs, miRNAs, lncRNAs, and pseudogenes were then visualized through volcano and heatmap plots generated with the ggplot2 and pheatmap packages in R.

2.3. Enrichment analysis and PPI analysis of the shared mRNA in BC and ADSCs

We identified the common DE mRNAs, miRNAs, lncRNAs, and pseudogenes by plotting Venn diagrams. To delve into the functionalities of these genes, we executed enrichment analyses using two widely used gene function classification systems, namely Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). GO categorizes genes into three functional domains: biological process, cellular component, and molecular function [31]. KEGG offers an in-depth overview of cellular functionalities and metabolic pathways [32]. Furthermore, we employed Gene-Set Enrichment Analysis (GSEA), a computational approach designed to ascertain whether the distribution of genes significantly diverges across two distinct groups [33]. The clusterProfiler [34] and GOplot [35] R packages and the Enrichment Map plugin [36] in Cytoscape [37] were used to analyze and visualize the results.

To identify key genes involved in both BC and ADSCs, we conducted a protein-protein interaction (PPI) analysis on the intersected mRNAs. The STRING (Search Tool for the Retrieval of Interacting Genes) online database [38] was utilized to assess the interconnectivity between the mRNAs, and only DE mRNAs with a minimum required confidence score of 0.7 were selected for network construction.

2.4. Construction of the shared lncRNA-miRNA-TF-gene and RBP-mRNA-pseudogene networks

Transcription factors are a group of proteins that possess the ability of attaching to distinct DNA sequences, thereby regulating the expression of genes [39]. To predict the interaction between common genes and TFs, we used the TRRUST database [40] and constructed a TF-gene regulatory network using Cytoscape. MicroRNAs, a category of endogenous, short non-coding RNAs, orchestrate the degradation or translation inhibition of target mRNAs. They are instrumental in the dysregulation of gene expression across a spectrum of physiological and pathological states. LncRNAs are longer than 200 bp and play critical roles in gene regulation and post-transcriptional mechanisms. A mRNA-miRNA-lncRNA regulatory network was constructed using the ceRNA hypothesis [41], where miRNAs act as a bridge between lncRNAs and mRNAs, using the multiMiR package [42] which scans predictions from various databases such as TargetScan, miRTarBase, miRanda, and others.

RNA binding proteins encompass a varied collection of proteins that engage with RNA molecules, including messenger RNAs, non-coding RNAs, and other regulatory RNA species [43]. They play a critical role in post-transcriptional modulation of gene expression and are implicated in various biological functions, including cell growth, differentiation, and development. We engaged the StarBase platform [44] to examine the interplay among mRNA, RBP, and pseudogene expression, leveraging its integration of CLIP-seq, degradome-seq, and RNA-RNA interactome data. Using this platform, we developed an RBP-pseudogene-mRNA network and visualized the results using Cytoscape.

2.5. Pan-cancer analysis of AK4 expression

To identify the most significant gene in BC and ADSCs, we employed an algorithm that compared the nodes of the RBP-pseudogene-mRNA and lncRNA-miRNA-TF-gene networks. AK4 emerged as the sole RNA shared between the two networks, indicating its potential as a core regulator in breast cancer and ADSCs. To investigate AK4 expression in pan cancer, we utilized the TCGA database and applied Wilcoxon rank-sum test for unpaired groups and paired Student's t-test for matched samples. AK4 protein expression in normal tissues and tumor was examined using immunohistochemical images from the HPA database [45]. Survival analysis, encompassing Kaplan-Meier curves and univariate Cox regression, was undertaken to delve into the prognostic impact of AK4's variable expression across 33 cancer varieties, with a concentrated focus on two principal metrics: overall survival (OS) and disease-specific survival (DSS).

Total mutational burden (TMB) represents the total number of somatic coding anomalies, and an elevated TMB is typically associated with increased responsiveness to immune checkpoint inhibitors (ICI) [46]. Microsatellite instability (MSI) data were quantified. Notably, high levels of MSI, referred to as MSI-H, are frequently linked with more favorable treatment outcomes [47]. The correlation between AK4 gene expression and TMB or MSI was analyzed through Spearman's method, with the P value indicated by the number of '*' on the graph.

2.6. Correlation between immune cell infiltration and AK4 expression

The tumor microenvironment (TME) represents a complex ecosystem, encompassing a variety of cell types like immune, stromal, and endothelial cells, in addition to non-cellular elements such as the extracellular matrix and soluble factors [48]. These components interact with each other and with cancer cells, significantly influencing cancer progression and treatment responses. To assess the expression of AK4 at the single-cell level within breast and adipose tissue, we utilized single-cell RNA-seq data available in the HPA database. We then employed the ESTIMATE algorithm [49], a gene expression-based method that uses single-sample gene set enrichment analysis, to appraise the abundance of stromal and immune cells within the TME across 33 cancer types. Utilizing the ESTIMATE package with standard settings, three distinct scores were generated for each of the 33 cancer types: the stromal score, immune score, and ESTIMATE score. Post-generation, the connection between AK4 gene expression and these scores was investigated using Spearman's correlation test.

To delve into the link between AK4 expression and the infiltration of immune cells, we applied CIBERSORT deconvolution analysis to approximate the proportions of various immune cell types [50]. Normalized expression data from 33 cancer types were matched with 22 immune cell types using the CIBERSORT algorithm. We generated heatmaps of Spearman's correlations between AK4 expression and diverse immune cell types. To delve deeper into the connection between AK4 and the immune system, we examined the relationship between AK4 expression and immune regulatory genes, encompassing major histocompatibility complexes, immunostimulators, chemokines, and their receptors.

2.7. Construction and validation of the sample cluster

AK4 and its corresponding RBP and TF, referred to as the hub genes, were determined to be of utmost importance. Unsupervised clustering analysis was performed to discern unique groupings predicated on AK4 expression levels, employing the ConsensusClusterPlus package [51] and focusing on the expression profile data of hub genes. We employed the PAM algorithm with 1-Spearman correlation distance metric in the analysis, utilizing 1000 bootstrap replicates with each replicate comprising 80% of patients in the training set. The K value, indicative of the number of clusters, was varied from 2 to 9 to ascertain the most effective classification. To determine this optimal classification, both the consistency matrix and the consistency cumulative distribution function were computed. Our analysis revealed the presence of 2 distinct clusters based on the expression levels of AK4. We assessed the overall survival of the two clusters using the Kaplan-Meier method. The "KEGG biological process" gene set, obtained from the molecular signature database, was utilized to conduct Gene Set Variation Analysis (GSVA) [52]. This analysis was conducted to elucidate the alterations in signaling pathways between the two identified clusters. In order to gain deeper insights into the inherent immune distinctions between the two molecular subtypes, we conducted a thorough immune analysis. This involved profiling the relative abundance of different immune cell types using the CIBERSORT tool.

2.8. Construction of a random forest model by using DE mRNAs between two clusters

DE mRNAs between the two clusters were pinpointed using the limma R package, employing a threshold of an adjusted *P* value cutoff of less than 0.05 and a $|\log2-\text{fold change}|$ greater than 1. Enrichment analysis was performed as previously described. To screen for signature genes among the identified DE mRNAs, we employed the randomForest package to construct a model with DE mRNAs as response variables and tumor/normal tissues as outcome variables, using the TCGA-BRCA dataset [53]. The model was trained using 5000 decision trees, and the optimal "mtry" parameter, which signifies the number of variables considered at each split, was determined through recursive random forest classification across all feasible values. The selection was based on the value with the lowest average error rate. The training set underwent sampling with replacement, and the significance of response variables was gauged using metrics like mean decrease Gini index, which aided in ranking their relative importance. Genes with an importance score greater than 5 were considered signature genes for breast cancer. To assess the significance of the signature genes, we generated a gene importance plot.

2.9. Development of ANN models and validation of the prediction model

To ensure batch effects were removed between the train and test groups, we scored the signature genes identified by the random forest model before building the artificial neural networks (ANN) prediction model [54]. We developed an ANN model using the neuralnet package, consisting of an input layer, a hidden layer, and an output layer. In ANN, the input layer encompassed the scores of all signature genes identified by the random forest model. Conversely, the output layer was comprised of two nodes, signifying membership in either the tumor or normal sample group. The ANN model classified a sample as belonging to the tumor or normal group based on its score, with higher scores indicating membership in the corresponding group. We validated the effectiveness of the ANN model by using data from GSE20711, GSE42568, and GSE15852 to verify the results obtained from the training dataset. The model's proficiency in distinguishing between tumor and normal samples was assessed using the receiver operating characteristic (ROC) curve. The model's discriminative capability was evaluated by the Area under the ROC curve (AUC) values, which span from 0 to 1, with values exceeding 0.9 denoting exceptional performance.

3. Results

3.1. Data preprocessing and identification of DE transcripts

We retrieved three datasets from the GEO database: GSE61302 for mRNA profiling, GSE72429 for miRNA profiling, and GSE57593 for ncRNA profiling. Relevant information was summarized in Table S1. We excluded abnormal samples from GSE61302 using the arrayQualityMetrics package, based on the mean absolute deviation of the M value for each array pair before (Fig. 1A) and after normalization (Fig. 1B). To assess the effects of normalization on the undifferentiated ADSCs and control groups, we used PCA to visualize the disparity before (Fig. S1A) and after normalization (Fig. S1B). Our preprocessing methods effectively reduced systematic and dataset-specific biases in the data.

The results for all DE transcripts analyzed are shown in Table S2. We identified 1304 DE mRNAs, 7 DE miRNAs, and 580 DE lncRNAs using a significance threshold of adjusted P value < 0.05 and $|\log2-\text{fold change}| > 0.5$, which are visualized in the heatmap (Figs. S1C, S1D, S1E) and volcano plot (Fig. 1C, D, 1E). Additionally, we present the results of pseudogene differential analysis in the volcano plot (Fig. 1F) and heatmap (Fig. S2F), which identified 282 DE pseudogenes under the shared cut-off value. To explore the biological processes linked to the DE mRNAs, we conducted GO annotation analysis. The results showed that the DE mRNAs were involved in various processes, such as the response to hypoxia, cell chemotaxis, and defense response to viruses (Fig. 1G). KEGG enrichment analysis further highlighted notable enrichment of DE mRNAs across various pathways, encompassing cytokine-cytokine receptor interaction, lysine degradation, and the TNF signaling pathway (Fig. 1H). Moreover, by performing differential expression analysis on the TCGA-BRCA training dataset, we identified 3854 DE mRNAs, 186 DE miRNAs, 1167 DE lncRNAs, and 305 pseudogenes between breast cancer tissues and normal tissues (Fig. S2 A–D).

3.2. Common mRNA signatures in BC and ADSC

We classified mRNAs by identifying the intersection between the GSE61302 and TCGA-BRCA databases and evaluating whether their expression changes were in the same direction in both datasets. In the context of ADSCs and BC, a mRNA was deemed concordant if it displayed changes in the same direction (either upregulation or downregulation) in both conditions. Conversely, a mRNA was deemed discordant if it was upregulated in ADSCs but downregulated in breast cancer. A total of 395 overlapping DE mRNAs were detected in both ADSC and BC datasets (Fig. 2A), consisting of 173 concordant mRNAs (104 upregulated and 69 downregulated in both ADSCs and BC) and 222 discordant mRNAs (77 upregulated in BC but downregulated in ADSCs, and 145 downregulated in BC but upregulated in ADSCs).

GO enrichment analysis revealed over 500 significantly enriched gene functions (Table S3), with discordant mRNAs enriched in endothelium development, regulation of vasculature development, and extracellular matrix structural constituent (Fig. 2D). Concordant mRNAs were enriched in mitotic nuclear division and chromosomal region. To facilitate interpretation of functional enrichment, we employed EnrichmentMap to create a network of GO terms. Concordant mRNAs were significantly enriched in 'anaphase checkpoint', 'cell cycle', and 'chromosome segregation' (Fig. 2G), while discordant mRNAs were enriched in 'angiogenesis', 'BMP signaling pathway', and 'viral genome replication' (Fig. 2H). Similarly, KEGG analysis identified significant pathways, with discordant mRNAs enriched in 'Toll-like receptor pathway', 'NOD-like receptor pathway', and 'MAPK pathway', and concordant

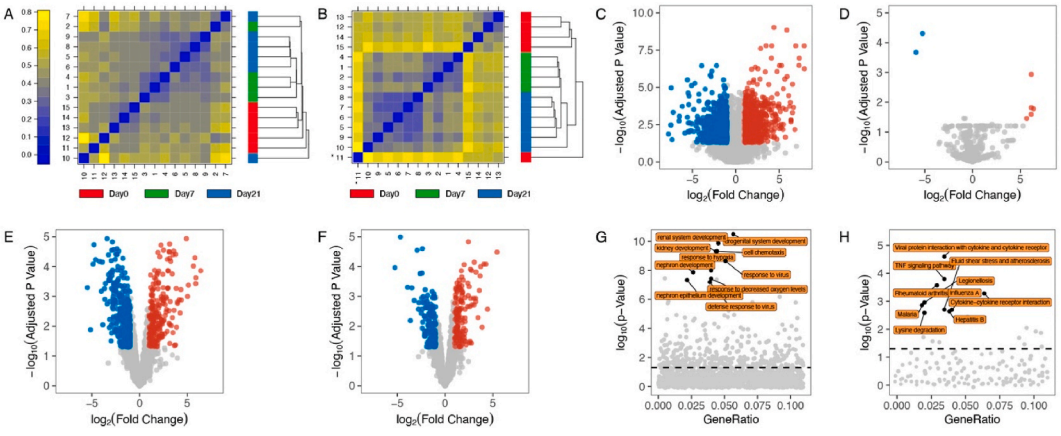


Fig. 1. Data preprocessing and identification of DE transcripts between undifferentiated ADSCs and differentiated ADSCs. **A–B** A false color heatmap of the distances between arrays. Patterns in this plot can indicate clustering of the arrays either because of intended biological or unintended experimental factors (batch effects) before (A) and after normalization (B). **C–F** Volcano plot showing the distribution of log2−fold change and − log10 adjusted P value of all quantified transcripts between the undifferentiated ADSC group and the differentiated ADSC for mRNA (C), miRNA (D), lncRNA (E) and pseudogene (F). Blue circles: adjusted P value < 0.05, log2−fold change < -1; red circles: adjusted P value < 0.05, log2−fold change >1. **G–H** Scatterplots displaying enriched GO (G) and KEGG (H) terms of DE mRNAs between undifferentiated ADSCs and the differentiated ADSCs. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

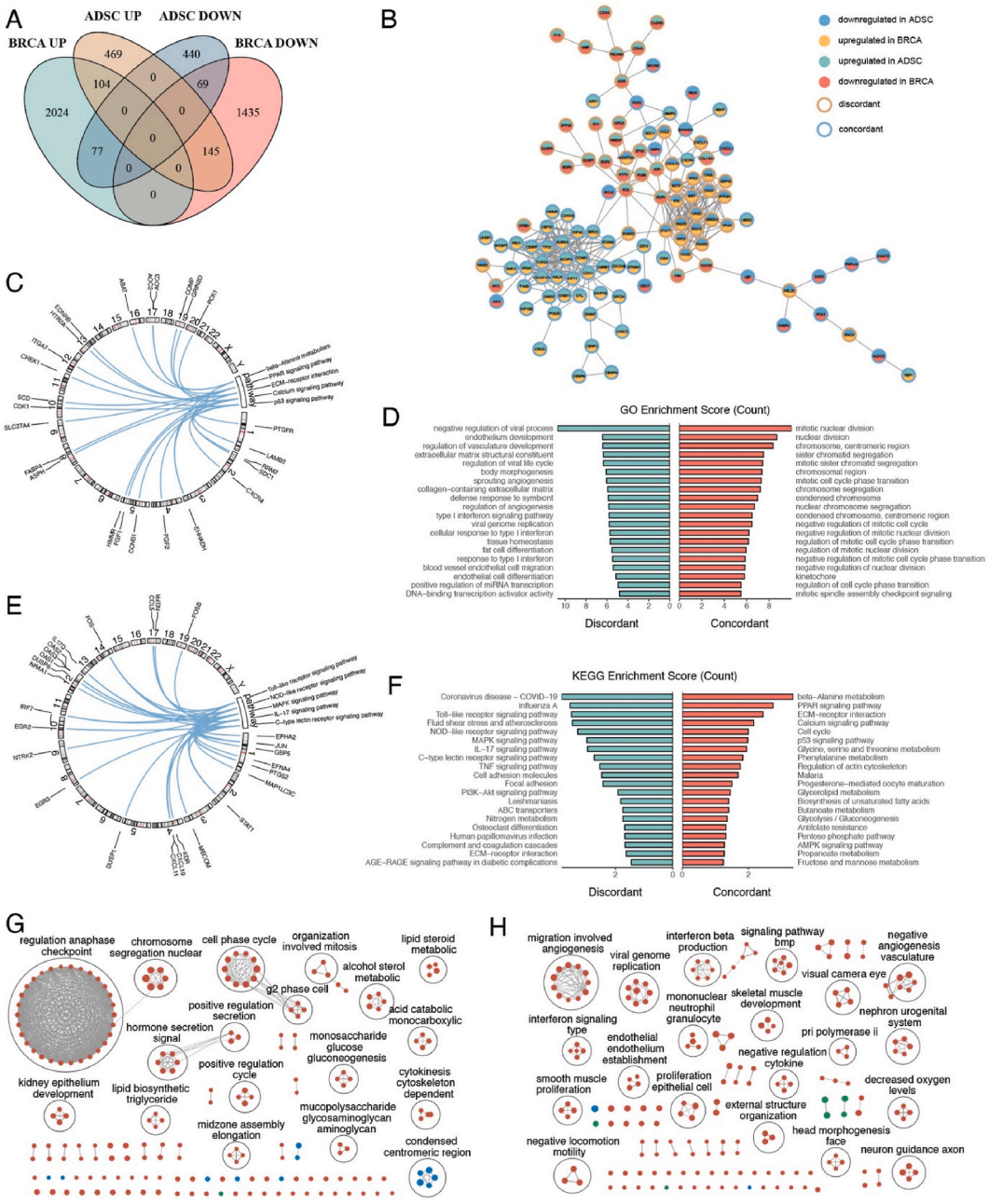


Fig. 2. Common gene signatures and enrichment analysis and PPI analysis in BC and ADSCs. **A** Overlapping mRNA of GSE61302 and TCGA−BRCA databases. **B** Protein-protein interaction network analysis of the overlapping mRNAs, with green and blue circle nodes representing upregulated and downregulated mRNAs in ADSCs, while yellow and red circle nodes represent upregulated and downregulated mRNAs in BC. The blue outer ring indicates concordant genes, while the yellow outer ring represents discordant genes. **C** Chromosome distribution of mRNAs in enriched KEGG pathways of concordants. **D** GO enrichment analysis for common significantly regulated mRNAs in ADSCs and BC. **E** Chromosome distribution of mRNAs in enriched KEGG pathways of discordants. **F** KEGG enrichment analysis for common significantly regulated mRNAs in ADSCs and BC. **G−H** Enrichment analysis identifying over 500 significantly enriched gene functions clustered using EnrichmentMap and AutoAnnotate in Cytoscape to identify the key gene functions among concordants (G) and discordants (H). Nodes represent individual GO terms, with size relating to the log2−fold change value in each term and the color indicating the functional category (Red: biological process, Blue: cellular component, Green: molecular function). Edges represent connections between nodes that share genes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

mRNAs involved in 'beta-Alanine metabolism', 'PPAR pathway', and 'ECM-receptor interaction' (Fig. 2F). We also annotated the genomic positions of common DE mRNAs related to important KEGG terms on human chromosomes (Fig. 2C and E). The PPI network analysis demonstrated that 115 proteins, translated from the overlapping mRNAs, interacted (Fig. 2B).

3.3. The common lncRNA-miRNA-TF-gene and pseudogene-RBP-mRNA network in BC and ADSCs

To identify regulatory pairs, we intersected the differentially expressed mRNAs with the TRRUST database, resulting in 60 TF-mRNA pairs and 47 nodes. Using Cytoscape software, we constructed an mRNA-TF network (Fig. S3D) from these pairs. We then analyzed the TCGA-BRCA database and the GSE73429 profile to identify shared DE miRNAs, resulting in the selection of hsa-miR-130b-3p, hsa-miR-377-3p, and hsa-miR-503-5p (Fig. S3A). 84 common DE lncRNAs (Fig. S3B) and 26 pseudogenes (Fig. S3C) were also identified. Based on the common DE miRNAs, target lncRNA and mRNA were predicted using the multiMiR package. The predicted genes were intersected with the previously generated differential TF-mRNA network to identify 3 DE miRNAs and their corresponding target mRNAs (including TF) and lncRNAs. Based on this information, we constructed a lncRNA-miRNA-TF-gene network consisting of 23 nodes, including 3 miRNAs, 7 mRNAs (without TF), 7 TFs, and 6 lncRNAs (Fig. 3A).

We used the StarBase database to download RBP-pseudogene pairs and mRNA-RBP pairs for five RBPs (VIM, MBNL2, ZFP36, CELF2, and CBX7) identified in our bioinformatics analysis of BC and ADSCs. As RBPs bind to mRNA and pseudogenes may act as competing endogenous RNAs that bind to RBPs, we analyzed the interactions between these RBPs and the DE pseudogenes identified in our study. We established an RBP-mRNA-pseudogene network, utilizing the target genes forecasted by the online dataset. The network included 57 nodes, consisting of 3 RBPs, 3 pseudogenes, and 51 mRNAs, with 57 edges (Fig. 3B). The network's core RBP was muscleblind like splicing regulator 2 (MBNL2), which exhibited an experimental correlation with 53 significant mRNAs (containing RBPs). The results showed that "synapse organization" and "Calcium signaling pathway" were the most significantly enriched GO and

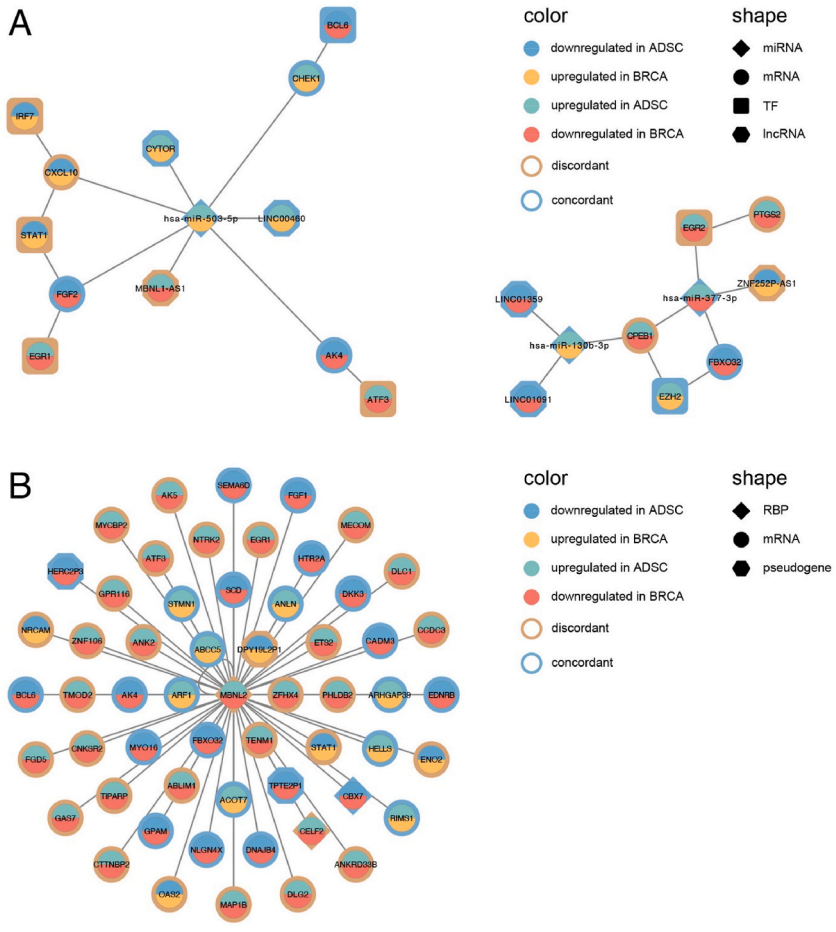
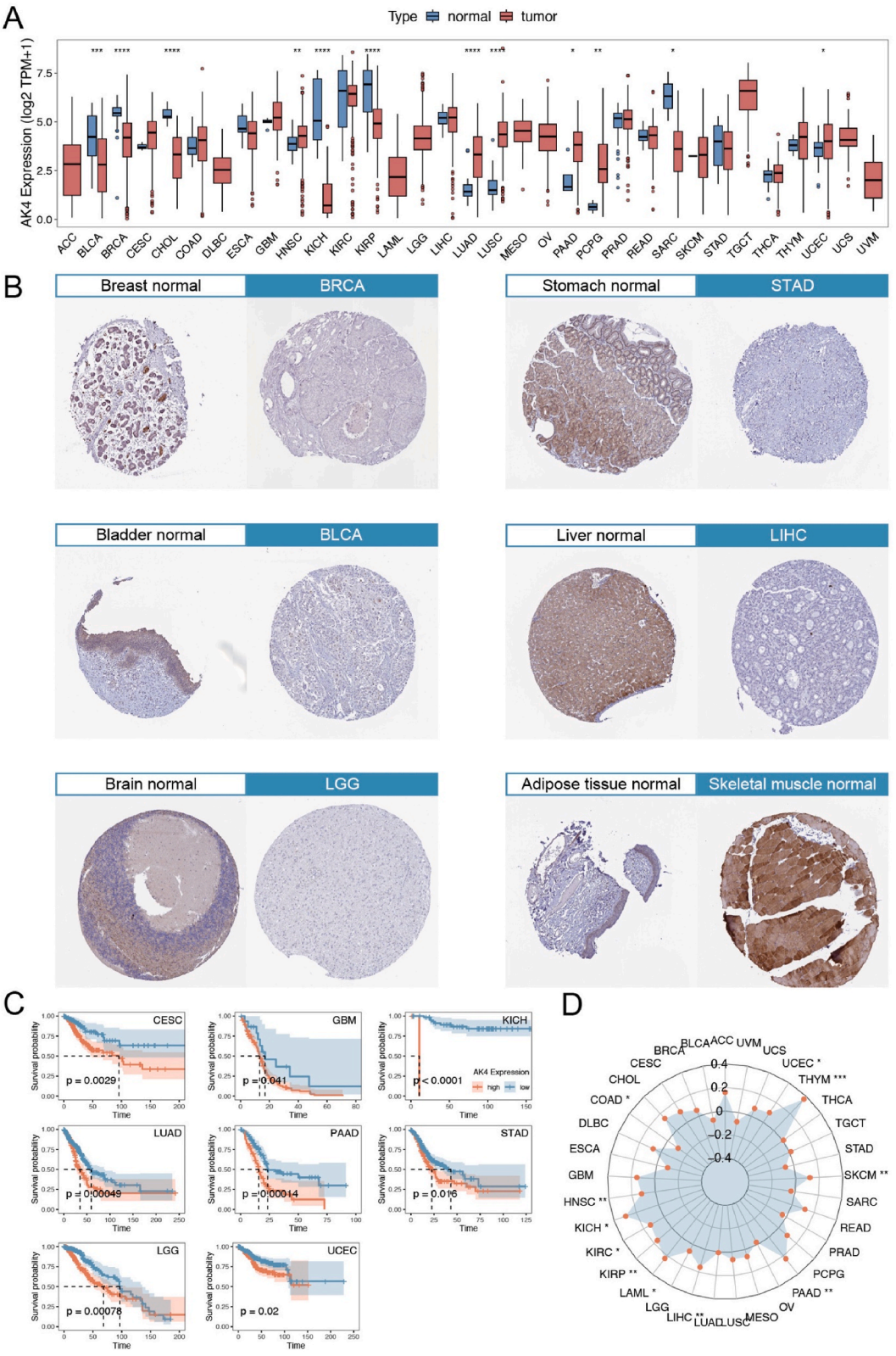


Fig. 3. Construction of mRNA regulatory network of the overlapping transcripts. **A** lncRNA-miRNA-TF-gene regulatory network. **B** RBP-mRNA-pseudogene regulatory network, with green and blue circle nodes representing upregulated and downregulated mRNAs in ADSCs, while yellow and red circle nodes represent upregulated and downregulated mRNAs in BC. The blue outer ring represents concordant genes, while the yellow outer ring represents discordant genes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



(caption on next page)

Fig. 4. AK4 expression and survival analysis in pan-cancer. **A** AK4 mRNA expression in pan-cancer in TCGA database. **B** AK4 protein expression in BRCA, STAD, BLCA, LIHC, LGG, adipose tissue and skeletal muscle from the HPA database. **C** The Kaplan-Meier curves for OS between AK4 expression (red: high AK4 expression; blue: low AK4 expression) in pan-cancer. **D** The correlation of AK4 expression with TMB (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

KEGG terms (Figs. S3E and S3F).

3.4. Pan-cancer analysis of AK4

AK4, or adenylate kinase 4, is the exclusive mRNA shared among the PPI, lncRNA-miRNA-TF-gene and RBP-mRNA-pseudogene networks, and its role in cancer was investigated through pan-cancer analysis. Notable disparities in AK4 mRNA expression levels were detected when comparing cancer patients and controls within the TCGA database (Fig. 4A). Specifically, AK4 mRNA expression was significantly higher in six tumor tissues, including head and neck squamous cell carcinoma (HNSC), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), and uterine corpus endometrial carcinoma (UCEC). Conversely, AK4 mRNA expression was significantly lower in six tumor tissues, including breast invasive carcinoma (BRCA), bladder urothelial carcinoma (BLCA), cholangiocarcinoma (CHOL), kidney chromophobe (KICH), kidney renal papillary cell carcinoma (KIRP), and sarcoma (SARC). Table S4 provides the full names of the abbreviations used. Furthermore, we examined AK4 protein expression across a range of tumor and normal tissues, utilizing the HPA database, with the findings concisely presented in Table 1. Our analysis revealed variations in AK4 expression levels across different cancer types, with lower expression observed in KIRC, LIHC, LGG, and STAD (Fig. 4B), and higher expression noted in skeletal muscle tissues. Immunofluorescence localization studies in Hep-G2 and A-431 cells confirmed that AK4 was primarily located in mitochondria (Fig. 5A).

OS and DSS were employed to evaluate the prognostic significance of AK4. Kaplan-Meier analysis indicated that low AK4 expression was associated with favorable OS in CESC, GBM, KICH, LUAD, LGG, PAAD, STAD, and UCEC (Fig. 4C). Cox regression analysis identified AK4 as a protective factor for BLCA, CESC, HNSC, KICH, LIHC, LGG, LUAD, PAAD, PRAD, UCEC, and UVM, while AK4 expression was a risk factor in KIRC (Fig. 5B). As for DSS, we observed that low expression of AK4 was associated with higher DSS in BLCA, CESC, HNSC, KICH, LUAD, LGG, PAAD, STAD, and UCEC (Fig. 5C). Cox regression analysis indicated that AK4 expression served as a protective factor in several cancers, including BLCA, CESC, HNSC, KICH, LIHC, LGG, PRAD, PAAD, UVM, and UCEC. The results suggest that AK4 could act as an independent prognostic indicator for specific cancers such as CESC, KICH, LGG, STAD, PAAD, and UCEC.

Additionally, we conducted a correlation analysis to ascertain the connection between AK4 expression and both TMB and MSI. In cancers such as HNSC, KICH, KIRC, KIRP, LIHC, LAML, PAAD, SKCM, THYM, and UCEC, AK4 expression demonstrated a significantly positive correlation with TMB. Conversely, in COAD, the correlation between AK4 expression and TMB was negative (Fig. 4D). Moreover, AK4 showed a positive correlation with MSI in UCEC, while a negative correlation was observed in COAD, THCA, and PRAD (Fig. 5D). These findings suggest that AK4 may have a role in the response to ICIs in specific cancer types.

3.5. Tumor microenvironment analysis of AK4 in pan-cancer

To gain a deeper insight into AK4's function within the tumor microenvironment, we conducted a comprehensive analysis from four perspectives: single-cell sequencing, tumor microenvironment analysis, immune infiltration analysis, and correlation with immune-related genes. Using the single-cell RNA-seq data from the HPA database, we investigated AK4 mRNA expression at the

Table 1
AK4 protein expression in various cancers.

Tissue	Number of samples	High expression	Low expression	Medium expression	Not detected
Brain glioma	11	0	2	2	7
Breast cancer	11	0	5	3	3
Cervix cancer	11	0	1	0	10
Colorectal cancer	13	0	4	5	4
Endometrial cancer	11	0	4	0	7
Liver cancer	9	0	3	1	5
Lung cancer	13	0	0	0	13
lymphoma	11	0	0	0	11
Melanoma	9	0	0	1	8
Ovarian cancer	10	0	2	2	6
Pancreatic cancer	13	0	3	2	8
Prostate cancer	12	0	0	11	1
Renal cancer	12	1	5	6	0
Skin cancer	10	0	0	0	10
Stomach cancer	9	0	3	1	5
Testis cancer	10	4	1	5	0
Urothelial cancer	6	0	3	0	3

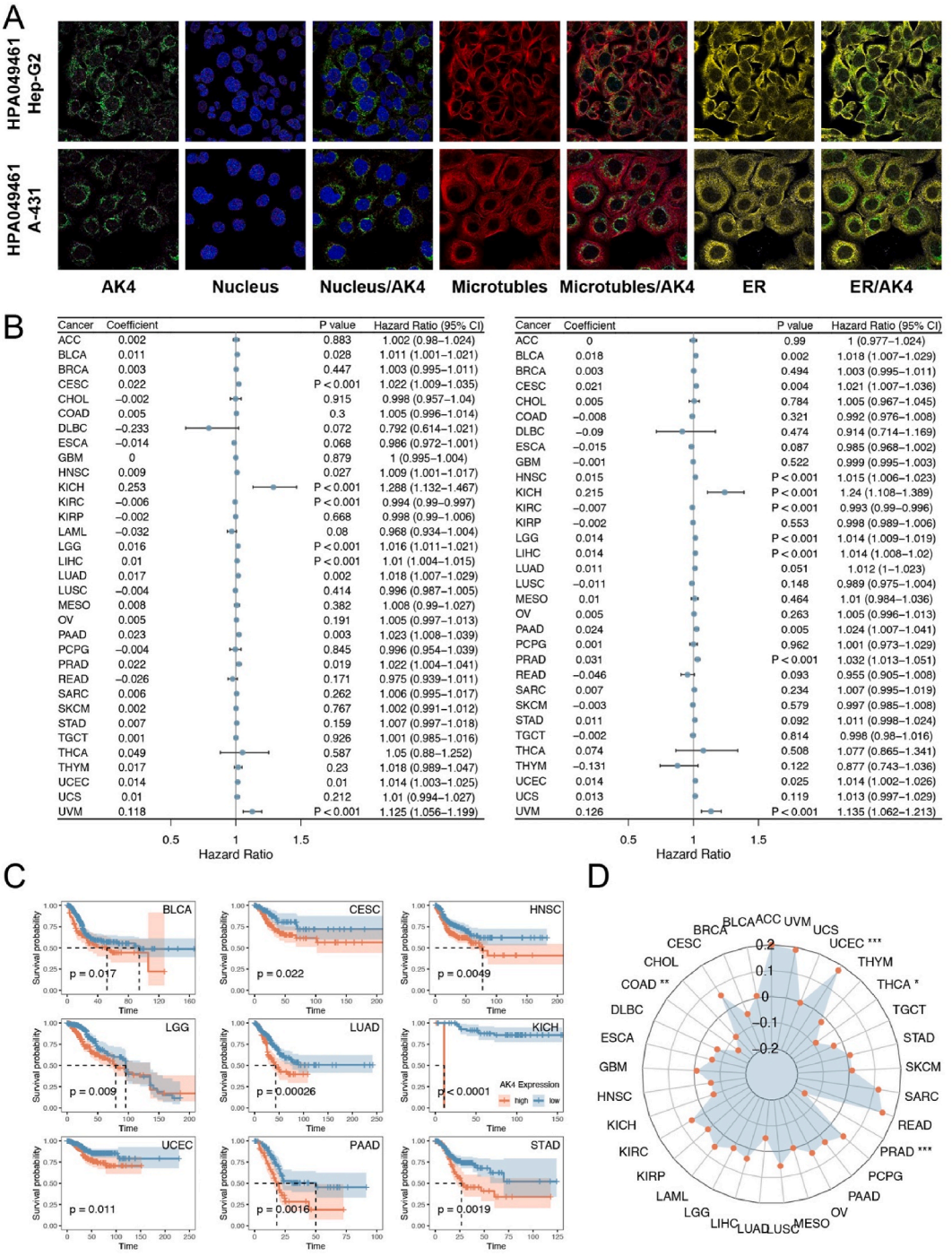


Fig. 5. AK4 expression and survival analysis in pan-cancer. **A** Immunofluorescence staining of the subcellular localization of AK4 in HPA database. **B** The univariate regression for OS in pan-cancer. **C** The univariate regression for DSS in pan-cancer. **D** The Kaplan- Meier curves for DSS between AK4 expression (red: high AK4 expression; blue: low AK4 expression) in pan-cancer. **E** The correlation of AK4 expression with MSI (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

cellular level in normal breast (Fig. 6A) and adipose tissue (Fig. 6B). Our analysis demonstrated that AK4 exhibited high expression levels in fibroblasts, macrophages, and adipocytes, whereas its expression was lower in most other individual cell types. We utilized the ESTIMATE algorithm to probe the relationship between AK4 expression and both stromal and immune scores across 33 types of cancer. Our analysis showed that AK4 expression was negatively correlated with stromal and immune scores in ACC, CESC, LUSC, SARC,

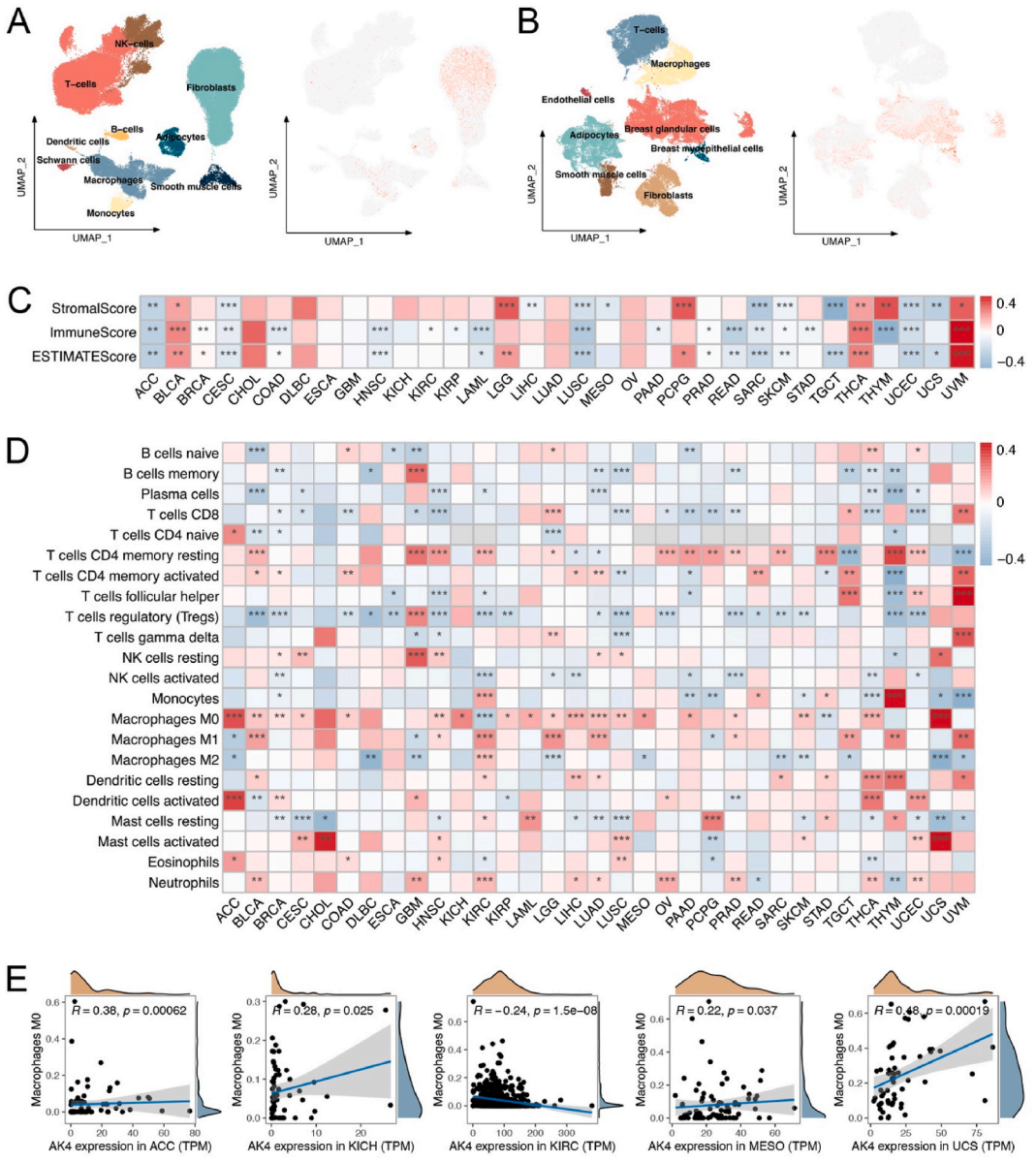


Fig. 6. Tumor microenvironment analysis of AK4 in pan-cancer. **A**—**B** Single-cell analysis of AK4 in adipose tissue (A). and breast tissue (B). **C** The heatmap of the correlation between AK4 expression and the stromal score, immune score. **D** Heatmap of AK4 expression correlation with 24 tumor infiltrating cells (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). **E** The correlation between AK4 expression and Macrophages M0 level in ACC, KICH, KIRC, MESO and UCS.

SKCM, UCEC, and TGCT, while it was positively correlated in BLCA, THCA, and UVM. Moreover, AK4 expression displayed either a positive or negative correlation with stromal scores and immune scores in specific cancer types (Fig. 6C). Subsequently, we explored the relationship between AK4 expression and the infiltration of Macrophages M0, employing Spearman correlation analyses. Our results showed that AK4 was positively related to Macrophages M0 in 21 cancers, with the correlation coefficient greater than 0.2 in 5 cancers (Fig. 6D and E). Furthermore, AK4 exhibited either a positive or negative correlation with the infiltration of 21 other immune cell types in the majority of the analyzed cancers.

In alignment with our objective of comprehending the role of AK4 in the TME, we explored the correlation between AK4 expression and immune regulatory genes across 33 different types of tumors. Major histocompatibility complex (Fig. S4A), chemokine receptors (Fig. S4B), and chemokines (Fig. S4C) were among the genes of interest. Our findings indicate that in certain cancers, AK4 expression exhibited a significant and positive correlation with immune regulatory genes. Additionally, we observed that AK4 was positively correlated with immune checkpoint genes, including CTLA4, TIGIT, PD-1, PD-L1 (CD274), LAG3, SIGLEC15, HAVCR2, and PDCD1LG2, in the majority of the analyzed cancers (Fig. S4D).

3.6. Identification of two molecular subtypes based on AK4, ATF3, and MBNL2 expression

We identified AK4, its corresponding RNA-binding protein MBNL2, and transcription factor ATF3 as hub genes of utmost importance. By employing the ConsensusClusterPlus package in R software and a consistent clustering methodology grounded on the expression of these three signature genes, we delineated two hub gene-based patterns, denoted as cluster 1 and cluster 2. We visualized the clustered data in heatmaps based on the multiplied relative expression levels in the two clusters, with 417 samples in Cluster 1 and 696 samples in Cluster 2 (Fig. 7A). Additionally, t-SNE analysis confirmed that the clusters were well-distributed (Fig. 7B). Further analysis was undertaken to contrast the overall survival times between the two clusters, revealing that Cluster 1 exhibited a notably shorter survival duration relative to Cluster 2 ($P = 0.01$) (Fig. 7C). To delve into the connection between the enriched pathways and the prognosis of hub gene-based clusters, we executed GSVA to assess the relative expression disparities of pathways in the two clusters. The GSVA analysis unveiled a multitude of differentially expressed pathways, which were depicted in a heatmap (Fig. 7D). Our analysis showed that cluster 1 had higher expression levels of pathways related to autophagy, apoptosis, ErbB signaling pathway, and ECM receptor interaction compared to cluster 2. Conversely, cluster 2 had lower expression levels of energy metabolism-associated pathways such as oxidative phosphorylation and cardiac muscle contraction. In addition, our statistical analysis showed significant differences in the immune status of the two molecular subtypes, particularly in NK cells and T cells (Fig. 7E).

3.7. Screening of BC signature genes based on DEG between two distinct clusters and development of ANN models

To compare the differences between the two hub gene-based patterns, we identified 162 DE mRNAs between cluster 1 and cluster 2 (Table S5). Subsequently, we examined the impact of DE mRNAs between the two clusters on the biologically relevant functions of patients (Fig. 8A). KEGG analysis revealed that these DE mRNAs were enriched in pathways related to ECM-receptor interaction, the MAPK pathway, and regulation of actin cytoskeleton (Fig. 8B).

To assess the significance of the hub genes and validate their diagnostic utility, we constructed a random forest classifier using the DEGs between the two hub gene-based patterns as response variables in the TCGA database. After optimizing the parameters, we found that the optimal mtry value for random forest selection was 2 and the number of decision trees was 5000 (Fig. 8C). We then ranked the importance of the hub genes. The genes that ranked highest in significance included CACHD1, CHRDL1, SRPX, MME, SPRY2, NTRK2, FMO2, MSRB3, ADIPOQ, EGFR, DST, ECRG4, MYLK, and PROS1, each with a MeanDecreaseGini index exceeding 5, indicating their significant contribution to the accuracy of the model (Fig. 8D).

Subsequently, we applied the ANN algorithm to optimize the weight values of each gene, converting the expression data of the 14 DE mRNAs into a Gene Score. The control and experimental samples were aggregated in the model, enabling the distinction of pathogenic gene expression between normal and BC samples (Fig. 8E). We calculated the AUC of our diagnostic prediction model using the pROC package, which was 0.938, indicating the excellent predictive power of our model (Fig. 8F). We further validated our model on three independent datasets (GSE20711, GSE42568, GSE15852) and found that the AUCs of these datasets were 0.994, 0.973, and 0.995, respectively (Fig. 8G). These findings indicate that our BC diagnostic prediction model can function as an independent predictor of BC.

4. Discussion

Recent research has shown that breast cancer is not simply a disease of malignant cells, but rather a complex ecosystem comprising a diverse range of non-cancerous cells and their interactions within the tumor microenvironment. These non-cancerous cells include

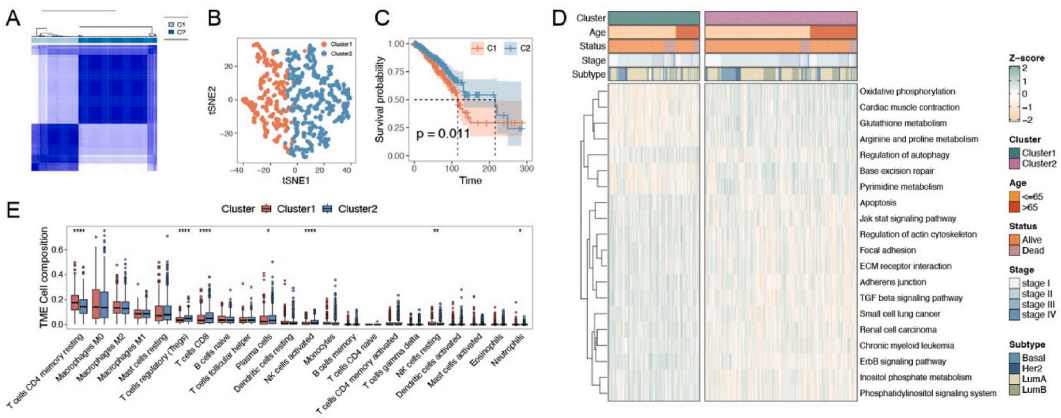


Fig. 7. Consensus clustering analysis of hub gene in the TCGA-BRCA cohort. **A** Heatmap of consensus matrix k = 2 in the TCGA-BRCA cohort. **B** T-SNE analysis of clusters. **C** Kaplan-Meier curve of overall survival in two clusters of the TCGA-BRCA cohort. **D** Heatmap illustrating the result of GSVA analysis between two clusters of the TCGA-BRCA cohort. **E** Boxplot depicting the enriching level of 22 immune-related cells evaluated by CIBERSORT algorithm between two clusters.

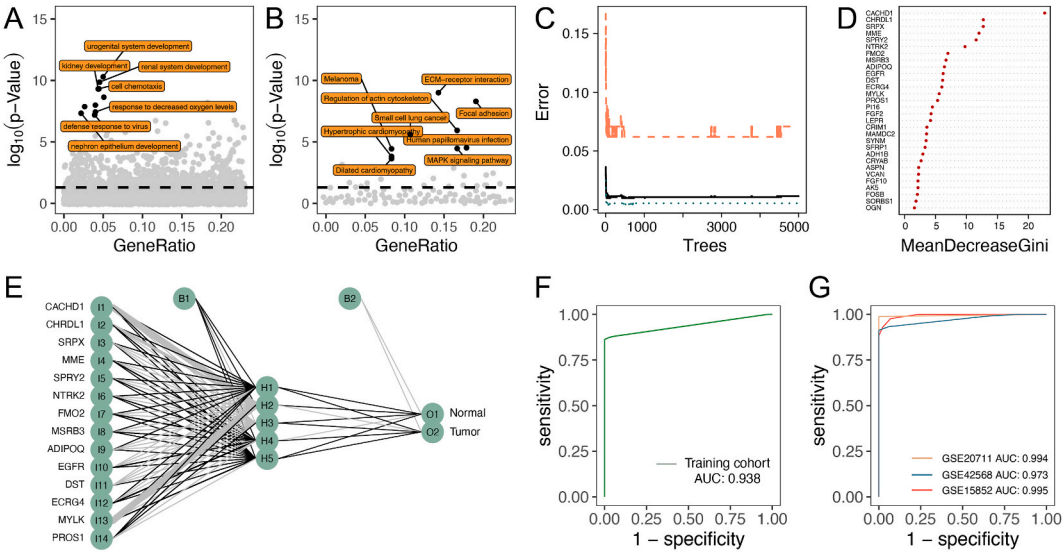


Fig. 8. Differential analysis between two hub genes-based patterns. **A-B** Scatterplots displaying enriched GO (A) and KEGG (B) terms of DE mRNAs between two clusters. **C** The relationship between the number of decision tree and the model error. **D** Random Forest algorithm showed the most important candidate genes. **E** artificial neural network which is divided into three layers. On the left is the input layer, which is the score of the fourteen signature genes. In the middle is a hidden layer containing 5 knots. On the right are an output layer containing normal and tumor sample junctions. **F-G** Receiver operating characteristic (ROC) curve of the ANN in training dataset (F) and validation datasets (G).

stem cells, immune cells, fibroblasts, and endothelial cells, among others. ADSCs are particularly crucial for the survival of autologous fat transplantation, a popular technique for breast reconstruction after cancer surgery. However, the relationship between ADSCs and breast cancer is not fully understood, and it is unclear whether ADSCs promote or inhibit tumorigenesis. To address this gap, we utilized various bioinformatics analyses to identify shared molecular pathways between ADSCs and breast cancer.

In this study, we downloaded ADSC datasets (GSE61302, GSE57593, GSE72429) and BC datasets. Our results present compelling genetic proof that ADSCs have a dual role in breast cancer, exhibiting both promoting and inhibitory effects on tumorigenesis. Specifically, 173 mRNAs were commonly upregulated or downregulated in both breast cancer and ADSCs, while 222 mRNAs exhibited differential regulation in breast cancer and ADSCs, further highlighting the complex interplay between ADSCs and breast cancer. In GO and KEGG enrichment analyses, the concordant mRNAs were primarily associated with cell proliferation, suggesting a promoting effect on tumorigenesis. The discordant mRNAs were related to angiogenesis, extracellular matrix constituent, and fat cell differentiation, which align with the known stromal behavior of ADSCs and further support their dual role in breast cancer.

Previous studies have emphasized the significance of ceRNA interactions in various diseases. However, these studies have predominantly concentrated on the competitive interplays between two molecules, overlooking the existence of numerous competitive connections inherent within cancer. To overcome this limitation, we conducted an extensive investigation of RBP-pseudogene-mRNA and lncRNA-miRNA-TF-gene networks. This approach allowed us to gain a more comprehensive understanding of the complex interplay among multiple molecules involved in cancer. In the mRNA-RBP-pseudogene network, MBNL2 was the sole RBP detected and displayed known regulatory associations with several pseudogenes and mRNAs, highlighting its potential role as a crucial regulatory gene within the ceRNA network. Research has shown that RNA-binding protein MBNL2 modulates tumor cell proliferation and DNA damage response by post-transcriptional regulation of target genes, including the modulation of the PI3K/AKT pathway and the protein levels of p21 [55]. These regulatory mechanisms affect DNA damage response and repair mechanisms, leading to changes in tumor cell fate after DNA damage. Juan Zhang et al. [56] reported that Neobractatin can efficiently inhibit metastasis of breast and lung cancer cells, partly by upregulating the expression of MBNL2. They found that MBNL2 was significantly downregulated in breast and lung carcinoma tumor tissues. Collectively, these discoveries imply that MBNL2 may hold a pivotal role in the regulation of gene expression and could potentially emerge as a therapeutic target not only for breast cancer but also for other forms of cancer.

Transcription factors are different from RBPs in that they exert control over gene expression at the transcriptional level. Several studies have demonstrated the regulatory role of various transcription factors in breast cancer pathogenesis [57]. To identify key transcription factors in BC and ADSCs, we obtained lncRNA-miRNA-TF-gene pairs by combining DE miRNAs and DE lncRNAs. In the ceRNA network, AK4 is regulated by the ATF3, a stress-induced transcription factor that plays a critical role in the cellular adaptive-response network by regulating metabolism, immunity, and oncogenesis [58]. ATF3 possesses the capacity to serve as both a transcriptional activator and a transcriptional repressor. In mammary tumors, it has been demonstrated to activate the canonical Wnt/β-catenin pathway, leading to the upregulation of genes associated with cancer cell metastasis [59]. Furthermore, ATF3 has been demonstrated to regulate adipose tissue through several mechanisms, including the downregulation of adipocyte differentiation, promotion of beige and brown cell differentiation, and inhibition of adipogenesis and lipogenesis [58]. These findings further support the notion that ATF3 may hold a significant role in both BC and ADSCs. Furthermore, ATF3 serves as a vital regulator of immune

responses and the preservation of host defense mechanisms. This is exemplified by its induction in response to various Toll-like receptors (TLRs), encompassing TLR4, 2/6, 3, 5, 7, and 9 [60]. Therefore, investigating the immunological role of ATF3 in breast cancer and adipose-derived stem cells may provide new insights into its function in these contexts.

Our study has identified AK4 as the common mRNA most closely related to both ADSCs and BC. AK4 belongs to the adenylate kinase family, which holds a pivotal role in cellular energy metabolism and the maintenance of adenine nucleotide homeostasis within various subcellular compartments [61]. Specifically, adenylate kinases facilitate the reversible transfer of a phosphate group from ATP to AMP, resulting in the formation of two molecules of ADP. AK4 has been shown to interact with ADP/ATP translocase and indirectly modulate mitochondrial membrane permeability, indicating a potential role in regulating mitochondrial function. To further explore the relationship between AK4 and various cancers, we conducted a pan-cancer analysis, which unveiled a downregulation of AK4 in both ADSCs and BC. Our results suggest that AK4 downregulation in both ADSCs and BC is a protective factor against breast cancer, providing further evidence for the anti-tumor effect of ADSCs. Consistent with our results, several studies have provided evidence that elevated AK4 expression is strongly linked to tumor metastasis and drug resistance, primarily through the regulation of mitochondrial activity and oxidative stress [62]. Specifically, AK4 has been demonstrated to contribute to tamoxifen resistance in BC through mechanisms such as the attenuation of mitochondrial apoptosis, heightened production of reactive oxygen species, and increased activation of p38, all mediated by an m6A-mediated epitranscriptomic mechanism [63]. Considering the evidence presented, it is reasonable to posit that AK4 holds promise as a potential novel molecular biomarker for both cancer diagnosis and prognosis. It is prudent to advocate for further exploration and research in this direction to fully realize its potential.

To delve deeper into the potential significance of AK4, we conducted an exploration of its correlation with immune cell infiltration. Given the significant contribution of mitochondria function to the control of inflammation, it is clear that the adenylate kinase family present in mitochondrial compartments are integrated into the network of inflammatory modulators [64]. A retrospective study has demonstrated a connection between AK1 and the season of conception and fetal sex through metabolic adaptability and immune modulation to environmental changes during conception [65]. Our research revealed a positive correlation between AK4 expression and M0-type macrophage expression in various tumors, including ACC, KICH, MESO, and UCS, based on CIBERSORT analysis. Furthermore, single-cell analysis in our research showed that AK4 was substantially upregulated in fibroblasts and macrophages. A more in-depth investigation into AK4 expression in fibroblasts and their role in interactions within the immune microenvironment, particularly the intricate mechanisms connecting fibroblasts with immune cells, has the potential to yield innovative strategies for BC immunotherapies.

In the last decade, cancer immunotherapy, including ICIs such as PD-1 inhibitors, has demonstrated significant clinical progress [66]. Our study revealed a significant positive correlation between AK4 expression and PD-L1 in 16 tumors, as well as a high correlation with TMB in 10 tumors, suggesting the potential of AK4 as a target in immunotherapy. The adenylate kinase family plays a crucial role in the eADO-generating pathway, which regulates both innate and adaptive immune responses [67]. Ecto-AK, a secreted or membrane-associated form of adenylate kinase, can phosphorylate extracellular AMP to generate eATP. Furthermore, CD39 and CD73 hydrolyze eAMP and eATP to produce immunosuppressive adenosine(eADO) [68]. Inhibitors of eADO-generating enzymes, such as ecto-AK, have been shown to promote antitumor immunity [67]. Thus, adenylate kinase might have a role in cancer immunotherapy by modulating the levels of eADO and affecting the immunosuppressive adenosinergic axis. Additionally, MSC-derived extracellular vesicles (EVs) have been shown to contain enzymatically active glycolytic enzymes, ATPases, and ATP-generating enzymes, including adenylate kinase [69]. This suggests a potential role for ADSC-derived EVs in regulating energy metabolism and anti-cancer immunotherapy. Nevertheless, the precise function of AK4 in the context of cancer immunotherapy remains incompletely understood, and additional research is required to unravel its underlying mechanisms and implications.

A previous study reported an association between aberrant expression of AK4 and heightened malignancy in lung cancers through downregulation of the transcription factor ATF3 [70]. Given the significant role of AK4 in tumor drug resistance and metastasis, which may be related to its corresponding transcription factor ATF3 and RNA-binding protein MBNL2, we performed clustering analysis based on the expression levels of AK4, ATF3, and MBNL2. We used random tree machine learning classifiers based on these DE mRNAs between two clusters and selected fourteen mRNAs with a MeanDecreaseGini index greater than 5 to build our machine learning models. Our artificial neural network yielded the best diagnostic performance with AUC values of 0.994, 0.973, and 0.995, indicating its excellent ability to distinguish between breast cancer and non-breast cancer cases. These results provide new insights into breast cancer diagnosis and highlight the potential of machine learning for improving disease management and developing innovative therapeutic options.

Several limitations of our study need to be addressed. Firstly, mRNA validation was conducted using online and GEO datasets instead of experimental validation. Additionally, the information on miRNA, lncRNA, RBP, TF, pseudogene and mRNA targets was collected from TRRUST, STRING and StarBase database, which may not be exhaustive due to data recording limitations. Secondly, it is imperative to conduct both clinical and experimental studies to validate the differential expression levels of the 14 DE mRNAs in breast cancer patients. Additionally, collecting a more comprehensive set of clinical features is essential to further validate the performance of the predictive model. Despite these limitations, our study offers valuable insights into the molecular mechanisms driving the development and progression of breast cancer, particularly in relation to the protective relationship between AK4 and breast cancer. Future research can build upon these findings to validate the predictive model's performance and elucidate the complex pathways involved in adenosine metabolism and shared pathological processes in ADSCs and breast cancer.

5. Conclusions

In conclusion, our study uncovered the dual role of ADSCs in breast cancer at the genetic level, with both promoting and inhibitory

effects on tumorigenesis. Through various bioinformatics analyses, we constructed PPI, lncRNA-miRNA-TF-gene and RBP-mRNA-pseudogene networks to comprehensively investigate the shared mechanisms of BC and ADSCs and identified AK4 as a key protective mRNA in pan-cancer. Our findings suggest that immune cells and AK4 may be critical susceptibility factors for BC and have potential as biomarkers or therapeutic targets for immunotherapy. Additionally, we identified 14 genes based on the AK4 cluster and developed a robust diagnostic prediction model for breast cancer using an artificial neural network, which may aid in the development of new diagnostic tools. Despite the limitations of our study, these results provide valuable insights into the molecular mechanisms of breast cancer and its relationship with ADSCs and offer potential targets for further research and the development of novel therapeutic options.

Funding

This study was financially supported by the National Key Clinical Specialty Project (No. 23003) and National Major Disease Multidisciplinary Diagnosis and Treatment Cooperation Project (No.1112320139). The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Ethics statement

Not applicable.

Availability of data and materials

Datasets analyzed in the current study are available in the GEO (https://www. ncbi.nlm.nih.gov/geo/) and TCGA databases. The GEO database accession numbers for the datasets include: GSE61302, GSE72429, GSE57593, GSE20711, GSE15852, GSE42568.

CRediT authorship contribution statement

Wei Lu: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhenyu Yang:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mengjie Wang:** Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Shiqi Li:** Writing – review & editing, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hui Bi:** Writing – review & editing, Visualization, Validation, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Xiaonan Yang:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We acknowledge the Gene Expression Omnibus (GEO) and TCGA databases for data sharing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27357.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al., Global cancer Statistics 2020: GLOBOCAN Estimates of Incidence and mortality worldwide for 36 cancers in 185 Countries, CA Cancer J Clin 71 (3) (2021) 209–249, https://doi.org/10.3322/caac.21660.
[2] S.R. Coleman, A.P. Saboeiro, Fat grafting to the breast revisited: safety and efficacy, Plast. Reconstr. Surg. 119 (3) (2007) 775–787, https://doi.org/10.1097/01.prs.0000252001.59162.c9.
[3] H. Eto, H. Kato, H. Suga, N. Aoi, K. Doi, S. Kuno, et al., The fate of adipocytes after nonvascularized fat grafting: evidence of early death and replacement of adipocytes, Plast. Reconstr. Surg. 129 (5) (2012) 1081–1092, https://doi.org/10.1097/PRS.0b013e31824a2b19.
[4] L.E. Kokai, K. Marra, J.P. Rubin, Adipose stem cells: biology and clinical applications for tissue repair and regeneration, Transl. Res. 163 (4) (2014) 399–408, https://doi.org/10.1016/j.trsl.2013.11.009.
[5] K. Yoshimura, K. Sato, N. Aoi, M. Kurita, T. Hirohi, K. Harii, Cell-assisted lipotransfer for cosmetic breast augmentation: supportive use of adipose-derived stem/stromal cells, Aesthetic Plast. Surg. 32 (1) (2008) 48–57, https://doi.org/10.1007/s00266-007-9019-4.

[6] S.F. Kølle, A. Fischer-Nielsen, A.B. Mathiasen, J.J. Elberg, R.S. Oliveri, P.V. Glovinski, et al., Enrichment of autologous fat grafts with ex-vivo expanded adipose tissue-derived stem cells for graft survival: a randomised placebo-controlled trial, Lancet 382 (9898) (2013) 1113–1120, https://doi.org/10.1016/s0140-6736(13)61410-5.

[7] B.J. Philips, T.L. Grahovac, J.E. Valentin, C.W. Chung, J.M. Bliley, M.E. Pfeifer, et al., Prevalence of endogenous CD34+ adipose stem cells predicts human fat graft retention in a xenograft model, Plast. Reconstr. Surg. 132 (4) (2013) 845–858, https://doi.org/10.1097/PRS.0b013e31829fe5b1.

[8] V.L. Negenborn, J.W. Groen, J.M. Smit, F.B. Niessen, M.G. Mullender, The Use of autologous fat grafting for treatment of scar tissue and Scar-related conditions: a systematic review, Plast. Reconstr. Surg. 137 (1) (2016) 31e–43e, https://doi.org/10.1097/prs.0000000000001850.

[9] S.M. Sultan, C.S. Stern, R.J. Allen Jr., V.D. Thanik, C.C. Chang, P.D. Nguyen, et al., Human fat grafting alleviates radiation skin damage in a murine model, Plast. Reconstr. Surg. 128 (2) (2011) 363–372, https://doi.org/10.1097/PRS.0b013e31821e6e90.

[10] A.E. Karnoub, A.B. Dash, A.P. Vo, A. Sullivan, M.W. Brooks, G.W. Bell, et al., Mesenchymal stem cells within tumour stroma promote breast cancer metastasis, Nature 449 (7162) (2007) 557–563, https://doi.org/10.1038/nature06188.

[11] J. Fang, F. Chen, D. Liu, F. Gu, Y. Wang, Adipose tissue-derived stem cells in breast reconstruction: a brief review on biology and translation, Stem Cell Res. Ther. 12 (1) (2021) 8, https://doi.org/10.1186/s13287-020-01955-6.

[12] K.E. Freese, L. Kokai, R.P. Edwards, B.J. Philips, M.A. Sheikh, J. Kelley, et al., Adipose-derived stems cells and their role in human cancer development, growth, progression, and metastasis: a systematic review, Cancer Res. 75 (7) (2015) 1161–1168, https://doi.org/10.1158/0008-5472.Can-14-2744.

[13] S. Gehmert, S. Gehmert, L. Prantl, J. Vykoukal, E. Alt, Y.H. Song, Breast cancer cells attract the migration of adipose tissue-derived stem cells via the PDGF-BB/PDGFR-beta signaling pathway, Biochem. Biophys. Res. Commun. 398 (3) (2010) 601–605, https://doi.org/10.1016/j.bbrc.2010.06.132.

[14] C. Jotzu, E. Alt, G. Welte, J. Li, B.T. Hennessy, E. Devarajan, et al., Adipose tissue derived stem cells differentiate into carcinoma-associated fibroblast-like cells under the influence of tumor derived factors, Cell. Oncol. 34 (1) (2011) 55–67, https://doi.org/10.1007/s13402-011-0012-1.

[15] Z. Sun, S. Wang, R.C. Zhao, The roles of mesenchymal stem cells in tumor inflammatory microenvironment, J. Hematol. Oncol. 7 (2014) 14, https://doi.org/10.1186/1756-8722-7-14.

[16] L. Yang, P. Shi, G. Zhao, J. Xu, W. Peng, J. Zhang, et al., Targeting cancer stem cell pathways for cancer therapy, Signal Transduct. Targeted Ther. 5 (1) (2020) 8, https://doi.org/10.1038/s41392-020-0110-5.

[17] V.S. Donnenberg, L. Zimmerlin, J.P. Rubin, A.D. Donnenberg, Regenerative therapy after cancer: what are the risks? Tissue Eng., Part B 16 (6) (2010) 567–575, https://doi.org/10.1089/ten.TEB.2010.0352.

[18] L. Zimmerlin, A.D. Donnenberg, J.P. Rubin, P. Basse, R.J. Landreneau, V.S. Donnenberg, Regenerative therapy and cancer: in vitro and in vivo studies of the interaction between adipose-derived stem cells and breast cancer cells from clinical isolates, Tissue Eng. 17 (1–2) (2011) 93–106, https://doi.org/10.1089/ten.TEA.2010.0248.

[19] B. Sun, K.H. Roh, J.R. Park, S.R. Lee, S.B. Park, J.W. Jung, et al., Therapeutic potential of mesenchymal stromal cells in a mouse breast cancer metastasis model, Cytotherapy 11 (3) (2009) 289–298, https://doi.org/10.1080/14653240902807026, 281 pp. following 298.

[20] L. Satish, J.M. Krill-Burger, P.H. Gallo, S.D. Etages, F. Liu, B.J. Philips, et al., Expression analysis of human adipose-derived stem cells during in vitro differentiation to an adipocyte lineage, BMC Med. Genom. 8 (2015) 41, https://doi.org/10.1186/s12920-015-0119-8.

[21] T. Xiao, L. Liu, H. Li, Y. Sun, H. Luo, T. Li, et al., Long noncoding RNA ADINR regulates adipogenesis by transcriptionally activating C/EBPα, Stem Cell Rep. 5 (5) (2015) 856–865, https://doi.org/10.1016/j.stemcr.2015.09.007.

[22] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47, https://doi.org/10.1093/nar/gkv007, e47.

[23] J. Yang, D. Zhang, A.F. Frangi, J-y Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.

[24] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, et al., TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (8) (2016) e71, https://doi.org/10.1093/nar/gkv1507.

[25] K. Tomczak, P. Czerwińska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge, Contemp. Oncol. 19 (1a) (2015) A68–A77, https://doi.org/10.5114/wo.2014.47136.

[26] S. Dedeurwaerder, C. Desmedt, E. Calonne, S.K. Singhal, B. Haibe-Kains, M. Defrance, et al., DNA methylation profiling reveals a predominant immune component in breast cancers, EMBO Mol. Med. 3 (12) (2011) 726–741, https://doi.org/10.1002/emmm.201100801.

[27] I.B. Pau Ni, Z. Zakaria, R. Muhammad, N. Abdullah, N. Ibrahim, N. Aina Emran, et al., Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context, Pathol. Res. Pract. 206 (4) (2010) 223–228, https://doi.org/10.1016/j.prp.2009.11.006.

[28] C. Clarke, S.F. Madden, P. Doolan, S.T. Aherne, H. Joyce, L. O'Driscoll, et al., Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis, Carcinogenesis 34 (10) (2013) 2300–2308, https://doi.org/10.1093/carcin/bgt208.

[29] R.A. Irizarry, B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, et al., Exploration, normalization, and summaries of high density oligonucleotide array probe level data, Biostatistics 4 (2) (2003) 249–264, https://doi.org/10.1093/biostatistics/4.2.249.

[30] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (12) (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.

[31] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (1) (2000) 25–29, https://doi.org/10.1038/75556.

[32] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30, https://doi.org/10.1093/nar/28.1.27.

[33] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. USA 102 (43) (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.

[34] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287, https://doi.org/10.1089/omi.2011.0118.

[35] W. Walter, F. Sánchez-Cabo, M. Ricote, GOplot: an R package for visually combining expression data with functional analysis, Bioinformatics 31 (17) (2015) 2912–2914, https://doi.org/10.1093/bioinformatics/btv300.

[36] D. Merico, R. Isserlin, O. Stueker, A. Emili, G.D. Bader, Enrichment map: a network-based method for gene-set enrichment visualization and interpretation, PLoS One 5 (11) (2010) e13984, https://doi.org/10.1371/journal.pone.0013984.

[37] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.

[38] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, et al., STRING v10: protein-protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (Database issue) (2015) D447–D452, https://doi.org/10.1093/nar/gku1003.

[39] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, et al., The human transcription factors, Cell 172 (4) (2018) 650–665, https://doi.org/10.1016/j.cell.2018.01.029.

[40] H. Han, J.W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, et al., TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions, Nucleic Acids Res. 46 (D1) (2018) D380–d386, https://doi.org/10.1093/nar/gkx1013.

[41] L. Salmena, L. Poliseno, Y. Tay, L. Kats, P.P. Pandolfi, A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146 (3) (2011) 353–358, https://doi.org/10.1016/j.cell.2011.07.014.

[42] Y. Ru, K.J. Kechris, B. Tabakoff, P. Hoffman, R.A. Radcliffe, R. Bowler, et al., The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations, Nucleic Acids Res. 42 (17) (2014) e133, https://doi.org/10.1093/nar/gku631.

[43] M.W. Hentze, A. Castello, T. Schwarzl, T. Preiss, A brave new world of RNA-binding proteins, Nat. Rev. Mol. Cell Biol. 19 (5) (2018) 327–341, https://doi.org/10.1038/nrm.2017.130.

[44] J.H. Li, S. Liu, H. Zhou, L.H. Qu, J.H. Yang, starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Res. 42 (Database issue) (2014) D92–D97, https://doi.org/10.1093/nar/gkt1248.

[45] P.J. Thul, C. Lindskog, The human protein atlas: a spatial map of the human proteome, Protein Sci. 27 (1) (2018) 233–244, https://doi.org/10.1002/pro.3307.
[46] D. Sha, Z. Jin, J. Budczies, K. Kluck, A. Stenzinger, F.A. Sinicrope, Tumor mutational burden as a predictive biomarker in solid tumors, Cancer Discov. 10 (12) (2020) 1808–1825, https://doi.org/10.1158/2159-8290.Cd-20-0522.
[47] M. Baretti, D.T. Le, DNA mismatch repair in cancer, Pharmacol. Ther. 189 (2018) 45–62, https://doi.org/10.1016/j.pharmthera.2018.04.004.
[48] D.C. Hinshaw, L.A. Shevde, The tumor microenvironment innately modulates cancer progression, Cancer Res. 79 (18) (2019) 4557–4566, https://doi.org/10.1158/0008-5472.Can-18-3962.
[49] K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612, https://doi.org/10.1038/ncomms3612.
[50] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, Methods Mol. Biol. 1711 (2018) 243–259, https://doi.org/10.1007/978-1-4939-7493-1_12.
[51] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (12) (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170.
[52] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7, https://doi.org/10.1186/1471-2105-14-7.
[53] A.L. Boulesteix, S. Janitza, J. Kruppa, I.R. König, Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics, Wiley Interdisciplinary Reviews: Data Min. Knowl. Discov. 2 (6) (2012) 493–507.
[54] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, Briefings Bioinf. 18 (5) (2017) 851–869, https://doi.org/10.1093/bib/bbw068.
[55] J. Cai, N. Wang, G. Lin, H. Zhang, W. Xie, Y. Zhang, et al., MBNL2 regulates DNA damage response via Stabilizing p21, Int. J. Mol. Sci. 22 (2) (2021), https://doi.org/10.3390/ijms22020783.
[56] J. Zhang, Z. Zheng, M. Wu, L. Zhang, J. Wang, W. Fu, et al., The natural compound neobractatin inhibits tumor metastasis by upregulating the RNA-binding-protein MBNL2, Cell Death Dis. 10 (8) (2019) 554, https://doi.org/10.1038/s41419-019-1789-5.
[57] M. Marqués, M.A. Sorolla, I. Urdanibia, E. Parisi, I. Hidalgo, S. Morales, et al., Are transcription factors plausible oncotargets for triple negative breast cancers? Cancers 14 (5) (2022) 1101, https://doi.org/10.3390/cancers14051101.
[58] H.C. Ku, C.F. Cheng, Master regulator activating transcription factor 3 (ATF3) in metabolic homeostasis and cancer, Front. Endocrinol. 11 (2020) 556, https://doi.org/10.3389/fendo.2020.00556.
[59] L. Yan, L. Della Coletta, K.L. Powell, J. Shen, H. Thames, C.M. Aldaz, et al., Activation of the canonical Wnt/β-catenin pathway in ATF3-induced mammary tumors, PLoS One 6 (1) (2011) e16515, https://doi.org/10.1371/journal.pone.0016515.
[60] M.M. Whitmore, A. Iparraguirre, L. Kubelka, W. Weninger, T. Hai, B.R. Williams, Negative regulation of TLR-signaling pathways by activating transcription factor-3, J. Immunol. 179 (6) (2007) 3622–3630, https://doi.org/10.4049/jimmunol.179.6.3622.
[61] C. Panayiotou, N. Solaroli, A. Karlsson, The many isoforms of human adenylate kinases, Int. J. Biochem. Cell Biol. 49 (2014) 75–83, https://doi.org/10.1016/j.biocel.2014.01.014.
[62] K. Fujisawa, S. Terai, T. Takami, N. Yamamoto, T. Yamasaki, T. Matsumoto, et al., Modulation of anti-cancer drug sensitivity through the regulation of mitochondrial activity by adenylate kinase 4, J. Exp. Clin. Cancer Res. 35 (2016) 48, https://doi.org/10.1186/s13046-016-0322-2.
[63] X. Liu, G. Gonzalez, X. Dai, W. Miao, J. Yuan, M. Huang, et al., Adenylate kinase 4 modulates the resistance of breast cancer cells to tamoxifen through an m(6)a-based epitranscriptomic mechanism, Mol. Ther. 28 (12) (2020) 2593–2604, https://doi.org/10.1016/j.ymthe.2020.09.007.
[64] A. Meyer, G. Laverny, L. Bernardi, A.L. Charles, G. Alsaleh, J. Pottecher, et al., Mitochondria: an organelle of bacterial origin controlling inflammation, Front. Immunol. 9 (2018) 536, https://doi.org/10.3389/fimmu.2018.00536.
[65] F. Gloria-Bottini, A. Neri, A. Pietropolli, E. Bottini, A. Magrini, Ak(1) genetic polymorphism and season of conception, Eur. J. Obstet. Gynecol. Reprod. Biol. 166 (2) (2013) 161–163, https://doi.org/10.1016/j.ejogrb.2012.09.019.
[66] J.A. Marin-Acevedo, E.O. Kimbrough, Y. Lou, Next generation of immune checkpoint inhibitors and beyond, J. Hematol. Oncol. 14 (1) (2021) 45, https://doi.org/10.1186/s13045-021-01056-8.
[67] B. Allard, D. Allard, L. Buisseret, J. Stagg, The adenosine pathway in immuno-oncology, Nat. Rev. Clin. Oncol. 17 (10) (2020) 611–629, https://doi.org/10.1038/s41571-020-0382-2.
[68] S.H. Donaldson, M. Picher, R.C. Boucher, Secreted and cell-associated adenylate kinase and nucleoside diphosphokinase contribute to extracellular nucleotide metabolism on human airway surfaces, Am. J. Respir. Cell Mol. Biol. 26 (2) (2002) 209–215, https://doi.org/10.1165/ajrcmb.26.2.4650.
[69] F.D.C. Gonçalves, F. Luk, S.S. Korevaar, R. Bouzid, A.H. Paz, C. López-Iglesias, et al., Membrane particles generated from mesenchymal stromal cells modulate immune responses by selective targeting of pro-inflammatory monocytes, Sci. Rep. 7 (1) (2017) 12100, https://doi.org/10.1038/s41598-017-12121-z.
[70] Y.H. Jan, H.Y. Tsai, C.J. Yang, M.S. Huang, Y.F. Yang, T.C. Lai, et al., Adenylate kinase-4 is a marker of poor clinical outcomes that promotes metastasis of lung cancer by downregulating the transcription factor ATF3, Cancer Res. 72 (19) (2012) 5119–5129, https://doi.org/10.1158/0008-5472.Can-12-1842.