BMC Musculoskeletal Disorders

RESEARCH

Open Access

Bioinformatics and machine learning approaches to explore key biomarkers in muscle aging linked to adipogenesis

Yumin Zhang[1*], Li Qin[1] and Juan Liu[1*]

Abstract

Adipogenesis is intricately linked to the onset and progression of muscle aging; however, the relevant biomarkers remain unclear. This study sought to identify key genes associated with adipogenesis in the context of muscle aging. Firstly, gene expression profiles from biopsies of the vastus lateralis muscle in both young and elderly population were retrieved from the GEO database. After intersecting with the results of differential gene analysis, weighted gene co-expression network analysis, and sets of adipogenesis-related genes, 29 adipogenesis-related differential expressed genes (ARDEGs) were selected. Connectivity Map (cMAP) analysis identified tamsulosin, fraxidin, and alaproclate as key target compounds. In further, using three machine learning algorithms and the friends analysis, four hub ARDEGs, ESRRA, RXRG, GADD45A, and CEBPB were identified and verified in vivo aged mice muscles. Immune infiltration analysis showed a strong link between several immune cells and hub ARDEGs. In all, these findings suggested that ESRRA, RXRG, GADD45A, and CEBPB could serve as adipogenesis related biomarkers in muscle aging.

Keywords Muscle aging, Adipogenesis, Bioinformatics, WGCNA analysis, Machine learning, Feature gene, Immune cell infiltration

Introduction

As with aging, muscles undergo several changes, including a reduction in mass and quantity. When this reduction progresses to clinical symptoms, it is termed primary sarcopenia [1]. Recent studies indicate that sarcopenia affects 10–27% of individuals over 60, depending on the classification and cut-off point used [2]. It was projected that by 2050, the number of patients would reach 500 million globally [3]. In clinical settings, sarcopenia has resulted in serious outcomes, including an elevated risk of falls, disability, and mortality among the elderly [4, 5]. As a result, muscle-aging-related sarcopenia greatly reduces the quality of life and health of affected individuals, placing a heavy burden on healthcare systems and society.

The pathogenesis of muscle aging remains unclear. However, it is believed to involve several contributing factors, including intermuscular fat infiltration, a reduction in the size and number of muscle fibers, neuromuscular junction dysfunction, alterations in mitochondrial integrity and function, increased oxidative stress due to cellular senescence, insulin resistance, inflammation, and hormonal changes such as fluctuations in testosterone, estrogen, growth hormone, insulin-like growth factor 1, and myostatin, among others [6, 7]. The initial

*Correspondence:
Yumin Zhang
zhangym11889@njmu.edu.cn
Juan Liu
lioujane@njmu.edu.cn
[1]Division of Geriatric Endocrinology, The First Affiliated Hospital with Nanjing Medical University, Nanjing, China

manifestations of muscle aging are often subtle, and there is still a paucity of effective early diagnostic biomarkers. Therefore, it is imperative to investigate the underlying pathophysiological mechanisms underlying muscle aging to identify diagnostic and therapeutic markers that can be clinically implemented.

Adipogenesis denotes the biological process through which precursor or stem cells undergo differentiation to become mature adipocytes [8, 9]. Previous studies showed that adipogenesis was crucial in the fat infiltration which was highly involved in the pathogenesis of muscle aging [6]. With advancing age, the reduction in skeletal muscle strength was observed to be more pronounced than the concomitant loss in body weight, which implied that muscle fiber atrophy alone didn't account for the diminished muscle strength. Thus, the accumulation of fat within the interstitial components might significantly contribute to the accelerated decline in muscle strength compared to weight loss [10, 11]. Fat infiltration had been observed to alter the microenvironment of skeletal muscle via endocrine and paracrine mechanisms, resulting in local inflammation and insulin resistance within the muscle tissue, which contributed to a decline in both absolute and specific muscle strength, adversely affecting muscle mass and function, and ultimately diminishing overall mobility [12]. Consequently, investigating adipogenesis and its associated markers is of significant importance for elucidating the pathogenesis of muscle aging.

Bioinformatics, a subfield of computer science, has been extensively employed to derive gene expression profiles associated with various diseases, identify genes related to these diseases, and elucidate the mechanisms underlying complex disease pathogenesis. Nevertheless, the findings generated through bioinformatics analyses have occasionally exhibited inconsistencies. Recently, the integration of machine learning techniques with bioinformatics methodologies has yielded more reliable and robust results [13, 14]. In this study, we employed bioinformatics techniques to identify adipogenesis-related differential-expressed genes (ARDEGs) from datasets containing both biopsies of young and elderly human muscles and subsequently integrated machine learning algorithms to select hub feature genes. The findings were further verified using independent datasets, single-cell sequencing datasets, and in vivo aged animal experiments. Additionally, we screened the relationships between hub feature genes and immune infiltration. For selecting active compounds targeting genes associated with adipogenesis in muscle aging, we conducted connectivity Map (cMAP) analysis. Our findings aimed to elucidate the pathogenesis of muscle aging from the perspective of adipogenesis and provide new practical ideas for the treatment of muscle aging.

Materials and methods

Data collection and preprocessing

The flow chart of this study was shown in Fig. 1. Generally, the human datasets containing young and aged muscles' biopsies (GSE136344 [15], GSE9676 [16], GSE1428 [17], GSE8479 [18], GSE28422 [19], GSE9103 [20], and GSE164471 [21]) were sourced from the publicly accessible Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) by using the key words "((muscles biopsies) and (aging)) OR ((muscles biopsies) and (sarcopenia))"or retrieved by searching the related paper in PubMed (http://www.ncbi.nlm.nih.gov/pubmed). To mitigate potential tissue heterog eneity, only datasets comprising human vastus lateralis muscle biopsy samples were selected. The other exclusion criteria included missing young or female groups, unclear context of the study, small sample size, intervention influencing the results, and mRNA samples pooled from multiple subjects. Generally, GSE136344, GSE9676, GSE1428, and GSE8479 were randomly selected as training datasets. GSE28422, GSE9103 and GSE164471 were used as verification datasets. The details of datasets were listed in Table 1. Samples from the young cohorts served as the control group, while samples from the elderly cohorts were designated as the aged group. The dataset GSE136344 contained muscle biopsies of participants with or without metabolic syndrome, and only the samples from the participants without metabolic syndrome were selected. In the dataset GSE9676, there were muscle biopsies before and after antibody enhancement, and only samples before the treatment were selected. The dataset GSE8479 contained muscle biopsies before and after exercise, and only samples before exercises were selected. The samples in GSE28422 which annotated with untrained state and at basal time point were included. For the dataset GSE9103, only samples annotated sedentary were selected. For the dataset GSE164471, ages below 35 were used as young control group, and ages over 65 were used as aged group.

The R package "oligo" was used to process the raw CEL files downloaded from the GEO website. Briefly, we annotated the genes using the annotation provided by the microarray vendors. For multiple probes mapping to the same gene, we used the mean aggregation by calculating the average expression value of all probes mapping to the same gene. Following this, we employed "insilicomering" package to integrate the datasets. For merging the datasets in different platforms, we used the genes that were consistently measured across all platforms. To mitigate the batch effects, we applied Combat function in the R package "sva" as described by Johnson et al. [22]. For acquisition of adipogenesis related genes (ARGs), we obtained the genes from the genecards database [23](https://www.genecards.org/)and Molecular Signatures Data

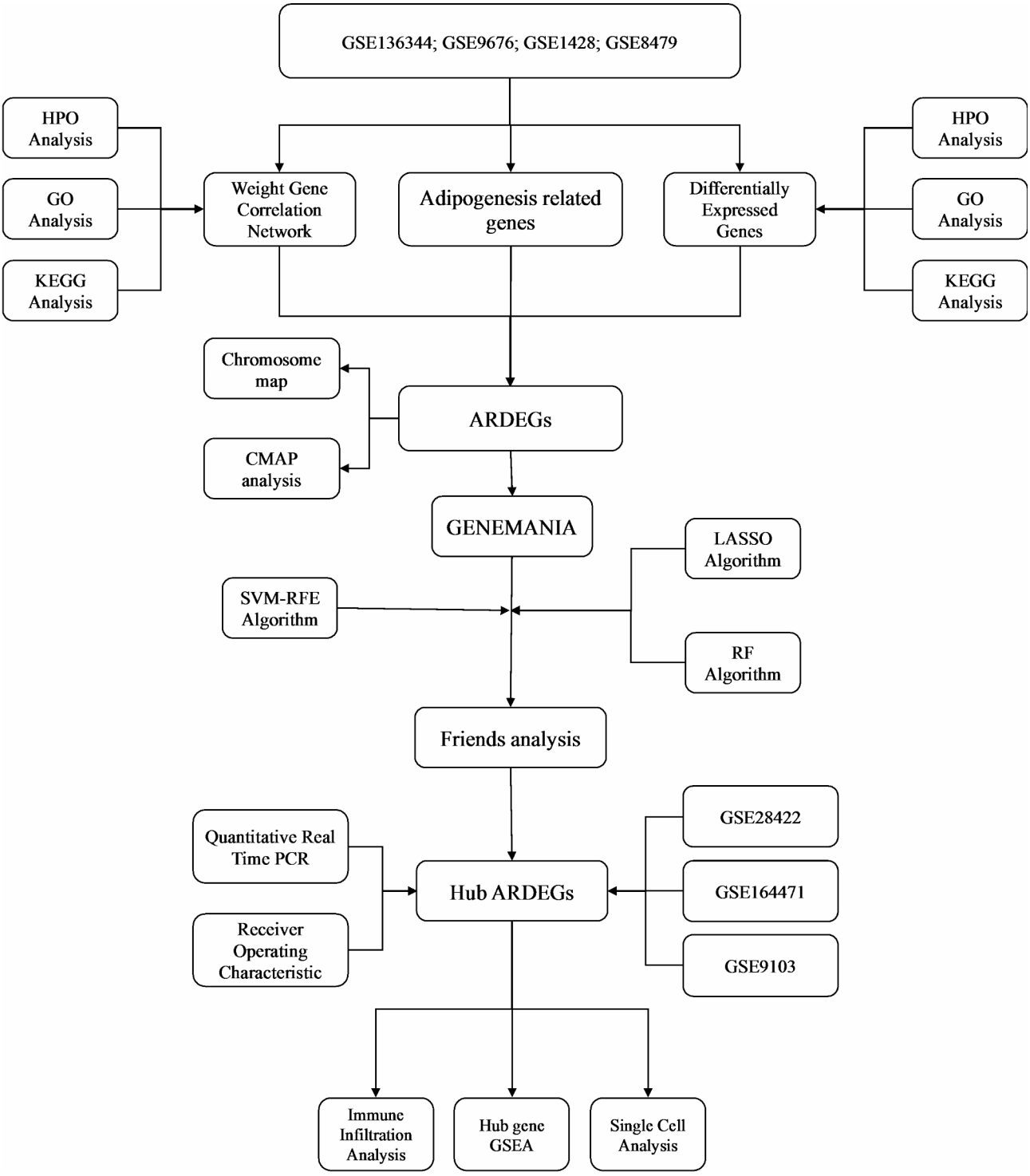

Fig. 1 Flow chart of this study

Table 1 Overview of datasets

	Accession	Platform	Control	Aged	Organism	Biopsies
Training sets	GSE136344 [15]	GPL5175	11	12	Homo sapiens	Vastus lateralis
	GSE9676 [16]	GPL96	14	16	Homo sapiens	Vastus lateralis
	GSE1428 [17]	GPL96	10	12	Homo sapiens	Vastus lateralis
	GSE8479 [18]	GPL2700	26	25	Homo sapiens	Vastus lateralis
Verification sets	GSE28422 [19]	GPL570	15	12	Homo sapiens	Vastus lateralis
	GSE9103 [20]	GPL570	10	10	Homo sapiens	Vastus lateralis
	GSE166427 [21]	GPL16791	12	18	Homo sapiens	Vastus lateralis

base (MSigDB) database (https://www.gsea-msigdb.org/gsea/msigdb), and subsequently merged them. The details of datasets were listed in Table S1.

Identification of differentially expressed genes (DEGs)

In this study, we utilized the "limma" package [24] of R software to conduct differential expression analysis, aiming to identify DEGs between the young groups and aged group. As there were only 37 DEGs under the criteria of|Fold change| ≥ 1.5 and a false discovery rate (FDR) < 0.05 which might miss many important DEGs, the screening criteria ultimately adjusted to|Fold change| ≥ 1.2 and FDR < 0.05. For the visual representation of the data, we employed the "ggplot2" package to generate volcano plots and the "pheatmap" package to construct heat maps.

Gene enrichment analysis

The Human Phenotype Ontology (HPO) offers a standardized vocabulary for the phenotypes associated with human diseases [25]. The Gene Ontology (GO) comprises an extensive collection of gene annotation terms that delineate and describe the functions of genes and proteins, encompassing biological processes (BP), molecular functions (MF), and cellular components (CC) [26]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) serves as a comprehensive bioinformatics database, primarily concentrating on biological and metabolic pathways [27]. The hallmark gene sets from the MSigDB represent specific and well-defined biological states or processes. These sets include high-quality, classic functional gene sets, which are composed of multiple known gene sets [28]. Utilizing the MSigDB databases and David databases, we obtained gene annotations for the HPO, GO, KEGG, and hallmark gene set to investigate gene-related functions and signaling pathways. The R package "cluster profiler" was used for enrichment analysis [29], with a significance threshold set at $p < 0.05$.

Weighted gene co expression network analysis (WGCNA)

In this study, we used the R package "WGCNA" [30] to perform co-expression network analysis on the merged dataset. Initially, the "hclust" function was employed for sample clustering and the removal of outliers. Subsequently, the "pick Soft Threshold" function was utilized to determine appropriate soft threshold weights, thereby achieving a scale-free network topology and transforming the correlation matrix into a weighted adjacency matrix. Subsequently, the adjacency matrix was transformed into a topological overlap matrix, facilitating the division of gene clustering into co-expression modules. To elucidate the relationship between these co-expressed modules and clinical characteristics, a dynamic tree cutting method with a minimum module size of 30 was employed to identify co-expressed gene modules. Ultimately, the correlation between gene modules and aged was computed to ascertain muscle aging related significant modules.

Gene interaction network analysis

Genemania (https://genemania.org/) is an online website which uses functional association datasets to find other genes related to input gene and can help predict the function of genes and genomes [31]. Here we used the Genemania to find interactions between the ARDEGs.

Machine learning analysis

We used three machine learning algorithms to obtain the best characteristic gene including the Least absolute shrinkage and selection operator (Lasso), Support Vector Machine-Recursive Feature Elimination (SVM-RFE) and Random Forest (RF). The aged group was set as predicted outcome. And 29 ARDEGs were used as the initial feature set. Specifically, Lasso, a machine learning algorithm grounded in logistic regression [32], enhances prediction accuracy by identifying variables through the optimization of the λ value that minimizes classification error. In this study, we used the "glmnet" package to perform Lasso regression analysis, which was followed by ten-fold cross-validation. SVM-RFE is a prominent machine learning technique extensively utilized in both classification and regression analyses [33]. This method identifies the optimal variables by iteratively removing features based on the feature weights derived from a support vector machine (SVM). In our study, we applied "e1071" package to systematically eliminate features in a recursive manner, thereby refining the selection of differential genes. The "svmrfe" function was employed to perform

the necessary calculations on the dataset. RF analysis utilizes an ensemble of classification trees to partition data into multiple nodes, thereby maximizing the homogeneity within each group. The RF algorithm aggregates numerous nodes to randomly select the most relevant classification tree [34]. In our study, we set the parameter "nTree" to 100 by using the "randomforest" package, found the best number of trees, and screened out the top 10 characteristic genes. Lastly, the "Venn" package [35] was used to intersect the results of three machine learning analysis to obtain the best characteristic gene, verified the expression of best feature genes with GSE9103 datasets and GSE22284 datasets, and drew the receiver operator characteristic curve (ROC) judgment accuracy using "pROC" package [36].

Friends analysis

GOSemSim is an R package designed for the computation of semantic similarity among GO terms, sets of GO terms, gene products, and gene clusters [37]. In this study, we conducted a Friends analysis utilizing the GOSemSim package to construct a gene interaction network. The significance of each gene was determined through the application of network topology parameters. Subsequently, we analyzed and predicted the functions and regulatory mechanisms of various genes within relevant biological processes. Core genes were identified from a diverse set of genes and were visualized using cloud and rain maps.

Single-cell sequencing expression verification

On the basis of the Human Muscle Ageing Cell Atlas (HLMA) (https://db.cngb.org/cdcp/hlma/) [38] which we accessed on 12 August 2024, single-cell sequencing data were utilized to assess the expression of hub ARDEGs in different cell subpopulations of skeletal muscle.

Immune infiltration analysis

The immune characteristics of young and aged groups were analyzed by single sample gene set enrichment analysis (ssGSEA), using R package "GSVA" [39]. To investigate the relationship between immune cells and characteristic genes, spearman correlation analysis was used between the immune infiltration scores and hub genes' expressions to reveal the regulatory relationship between the two. P values < 0.05 in the correlation analysis was considered statistically significant.

cMAP analysis

The cMAP database (https://clue.io/query) [40] was ut ilized to identify potential small molecule drugs for the treatment of muscle aging via the adipogenesis pathway. The common targets associated with DEGs were categorized into "up-regulated" and "down-regulated" datasets. Subsequently, a systematic high-throughput analysis was conducted to screen for small molecule compounds with potential therapeutic efficacy. The sources of potential small molecule drugs are compounds that have been approved by the United States Food and Drug Administration (FDA). The selection criteria for these compounds are based on their scores obtained through high-throughput screening. The scores are ranked from – 1.

Animal experiments

The experimental protocols of the present study were approved by the Animal Care and Use Committee of Nanjing Medical University (Nanjing, China), and carried out in compliance with the ARRIVE guidelines. Generally, C57BL/6J male mice of different ages were obtained from SPF Biotechnology (Beijing, China) and maintained at 20–25° C with free access to water and food. Young (6 months old, 6 m; $n = 6$), and old (23 months old, 23 m; $n = 6$) mice were included. Mice were euthanized with an intraperitoneal injection of 50 mg/kg sodium pentobarbital, then sacrificed by neck removal. After all animals were euthanized, gastrocnemius muscle tissues were harvested and immediately stored at – 80° C. Total RNA was extracted subsequently using Trizol reagent. The concentration and purity of the extracted RNA were detected by UV spectrophotometer. The real-time PCR reactions were conducted using SYBR Green PCR system (Q341-02/03, Vazyme, China). The quantitative real-time PCR (qRT-PCR) program was set as: 95 °C, 1 min; 95 °C for 10 s; 60 °C for 30 s; for 40 cycles; 95 °C for 15 s; 60 °C for 60 s; 95 °C for 15 s. Relative quantities were calculated using the $2 - \Delta\Delta Ct$ method with β-actin as inner control [41]. All primers were synthesized by Invitrogen (Thermo Fisher Scientific). The primer sequences of each gene are shown in Table S2.

Statistical analysis

Data analysis was conducted using SPSS statistical software, version 26.0. Each experiment was replicated three times, and the data conformed to a normal distribution, represented as mean ± standard deviation ($\bar{x} \pm s$). For intergroup comparisons, the independent sample t-test was employed. P values < 0.05 was considered statistically significant.

Results

Identification and function enrichment of DEGs in aged muscles

Firstly, to obtain gene expression profiles of aged muscles, we mitigated batch effects in the gene expression matrix following the integration of datasets GSE136344, GSE9676, GSE1428, and GSE8479 (Figure S1). The box plots in Figures S2A illustrated significant disparities in sample distributions across the original datasets.

However, post-correction for batch effects, the median distributions of the samples across the datasets converged, indicating successful normalization (Figures S2B). Furthermore, the UMAP results presented in Figure S2C demonstrated that the four datasets were independent and non-overlapping. However, upon the removal of variance, the sample distributions exhibited a tendency towards convergence (Figure S2D).

A comprehensive screening identified 446 differentially expressed genes (DEGs). Among these, 261 genes were up-regulated, while 185 genes were down-regulated (Fig. 2A). A cluster heatmap illustrating the top 20 up-regulated and down-regulated genes is presented in Fig. 2B. Subsequent analyses of the biological functions and pathways associated with the 446 DEGs were conducted using HPO, GO, KEGG enrichment analyses and GSEA. As illustrated in the Fig. 2C and Table S3, the top 10 GO terms in BP were primarily in the cellular respiration, and generation of precursor metabolites specifically the glucose. The top 10 GO terms in CC were mainly focused in the mitochondria, extracellular matrix, and secretory granule lumen. The top 10 GO terms in MF were specifically enriched in the NAD binding, electron transfer activity, and oxidoreductase activity. In terms of KEGG (Fig. 2D and Table S4), the top 10 items were citrate cycle (TCA cycle), carbon metabolism, non-alcoholic fatty liver disease, oxidative phosphorylation, glycolysis/gluconeogenesis, biosynthesis of amino acids, diabetic cardiomyopathy, pyruvate metabolism, thermogenesis, and 2-oxocarboxylic acid metabolism. The HPO results showed that abnormal placental size, small placenta, decreased activity of the pyruvate dehydrogenase complex, increased serum lactate, membranoproliferative glomerulonephritis, abnormality of acid base homeostasis, abnormality of the mitochondrion, and neoplasm of the adrenal gland were significantly enriched (Figure S3). In addition, we performed a GSEA of the DEGs (Fig. 2E and Table S5), and the results showed that the top 10 pathways were oxidative phosphorylation, myogenesis, adipogenesis, apoptosis, fatty acid metabolism, hypoxia, mtorc1 signaling, epithelial mesenchymal transition, and glycolysis.

Identification and function enrichment of genes found in key modules based on WGCNA analysis

In biological systems, the regulation of specific functions is frequently mediated by one or more genes exhibiting similar expression patterns. Consequently, gene co-expression analysis serves as a valuable tool for identifying gene sets implicated in distinct biological functions. To identify the critical modules most closely associated with muscle aging, the WGCNA was performed using the aggregated gene expression profile. The soft threshold was determined to be 9, based on the scale-free topology fitting exponential curve, as illustrated in Figure S4A&B. A hierarchical clustering tree diagram was constructed, as shown in Figure S4C. A total of 9 modules were identified, as shown in Figure S4D. The clustering of module feature vectors was also explored as illustrated in Figure S4E.

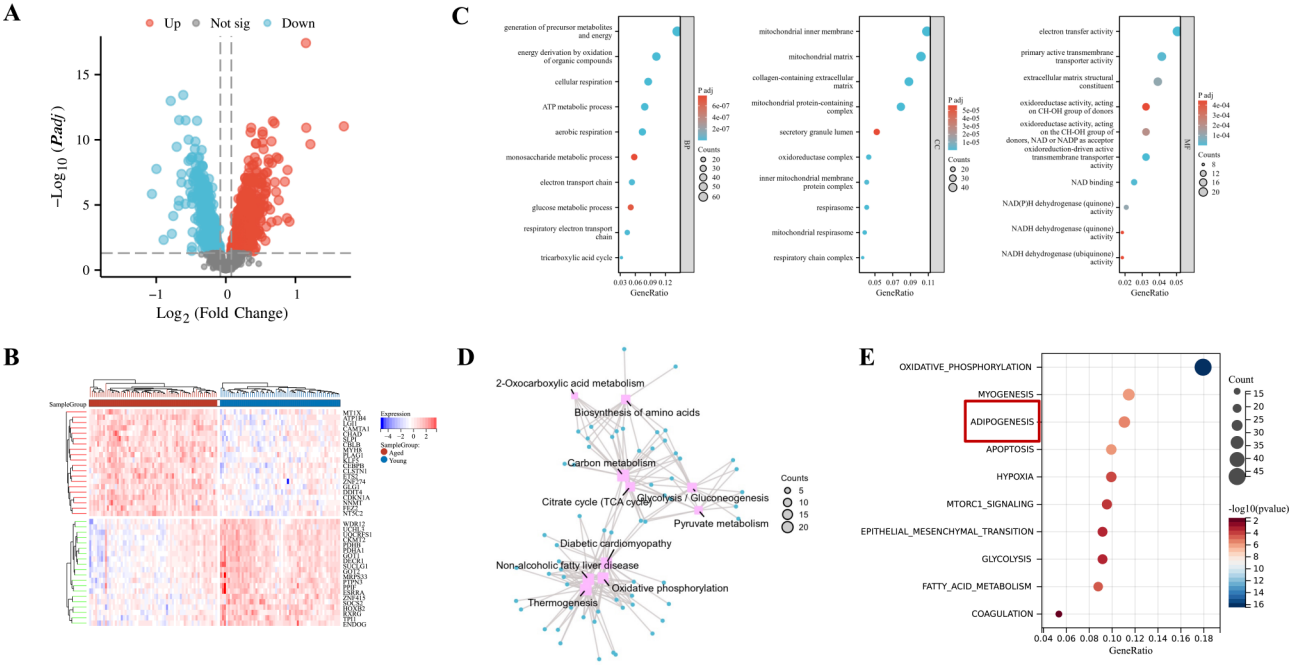

Fig. 2 Identification and function enrichment of DEGs in muscle aging. (**A**) Volcano plot of DEGs. (**B**) Cluster heatmap of DEGs in young and aged group. (**C**) GO analysis of DEGs. (**D**) KEGG pathway analysis of DEGs. (**E**) Hallmark gene set analysis of DEGs

As seen in the Fig. 3A&S5, the green module gene set had the highest positive correlation ($r = 0.57$, 240 genes), while the turquoise gene set had the highest negative correlation (r=-0.53, 323 genes). Then, the above two modules were selected as key modules significantly related to intervertebral disc degeneration, which contained 563 genes in total.

The biological functions and pathways related to genes in key modules were then examined using HPO, GO, KEGG enrichment analyses. The GO analysis (Fig. 3B & Table S6) showed that the top 5 GO terms in BP were generation of generation of precursor metabolites and energy, energy derivation by oxidation of organic compounds, cellular respiration, aerobic respiration, and respiratory electron transport chain. The top 5 GO terms in CC were mitochondrial protein-containing complex, mitochondrial inner membrane, mitochondrial matrix, oxidoreductase complex, and mitochondrial respirasome. The top 5 GO terms in MF were electron transfer activity, oxidoreduction-driven active transmembrane transporter activity, NADH dehydrogenase (ubiquinone) activity, NADH dehydrogenase (quinone) activity, and NADH dehydrogenase activity. In terms of KEGG (Fig. 3C & Table S7), the top 10 items were non-alcoholic fatty liver disease, oxidative phosphorylation, diabetic cardiomyopathy, citrate cycle (TCA cycle), chemical carcinogenesis-reactive oxygen species, thermogenesis, Parkinson disease, carbon metabolism, Huntington disease,

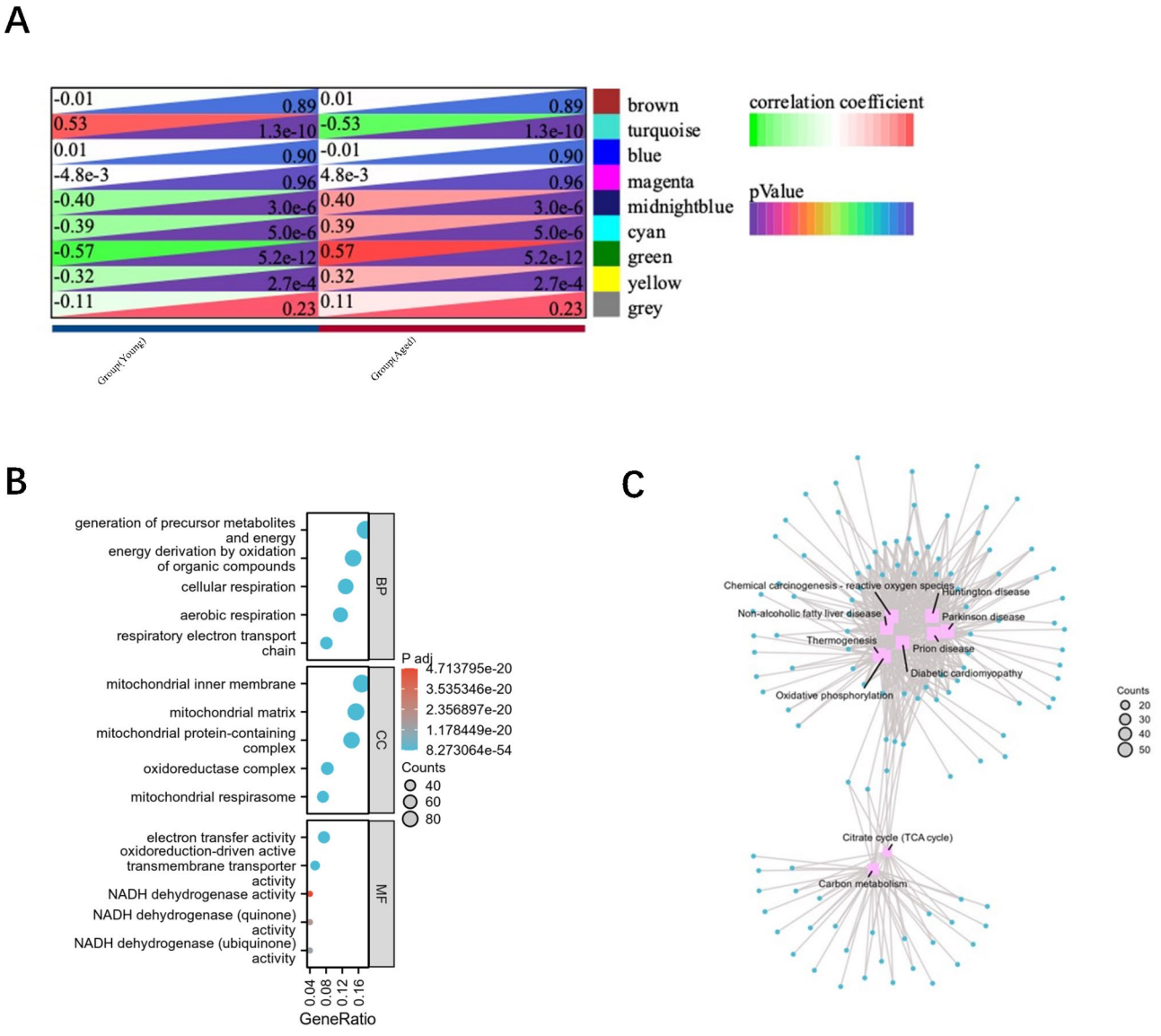

Fig. 3 Function enrichment of genes in key modules in WCGNA. (**A**) Correlation map between modules and muscle aging. The abscissa is the phenotypic trait. The ordinate is the corresponding module. Different colors represent different modules, with red indicating positive correlation and green indicating negative correlation. (**B**) GO biological processes analysis for genes in key modules. (**C**) KEGG pathway analysis for genes in key modules

and prion disease. The top 10 HPO results showed that abnormality of acid base homeostasis, increased serum lactate, abnormality of the mitochondrion, lactic acidosis, abnormal activity of mitochondrial respiratory chain, metabolic acidosis, hypertrophic cardiomyopathy, and decreased activity of the pyruvate dehydrogenase complex were significantly enriched (Figure S6 & Table S8).

Identification of ARDEGs in aged muscles

As shown in Fig. 4A, a total 29 ARDEGs genes were found after intersecting the DEGs, genes found in key modules by WCGNA and the adipogenesis related genes acquired from public database. The expression of these 29 ARDEGs in the training datasets were shown in Fig. 4B. Among the 29 genes, 6 genes were upregulated including CEBPA, GADD45A, CDKN1A, APOE, CYP4B1, and CEBPB while 23 genes were downregulated including ACO2, MTCH2, IDHA, SDHC, ESRRA, NDUFS3, SUCG1, MDH2, UQCRC1, CS, RXRG, COX7B, DLAT, MRPL15, DLD, SDHB, IMMT, COQ3, DECR1, GBE1, IDH1, HIBCH, and PPARGC1A.

Chromosome localization analysis is the process of annotating the positions of gene family members on chromosomes, observing whether gene family members are clustered and distributed on chromosomes, which helps in elucidating the potential heredity connections between ARDEGs. In this study, we used the R package "circlize" [42] to initialize the genome and mark the location of the target genes in the chromosome as illustrated in Figure S7. To further explore the interactions of

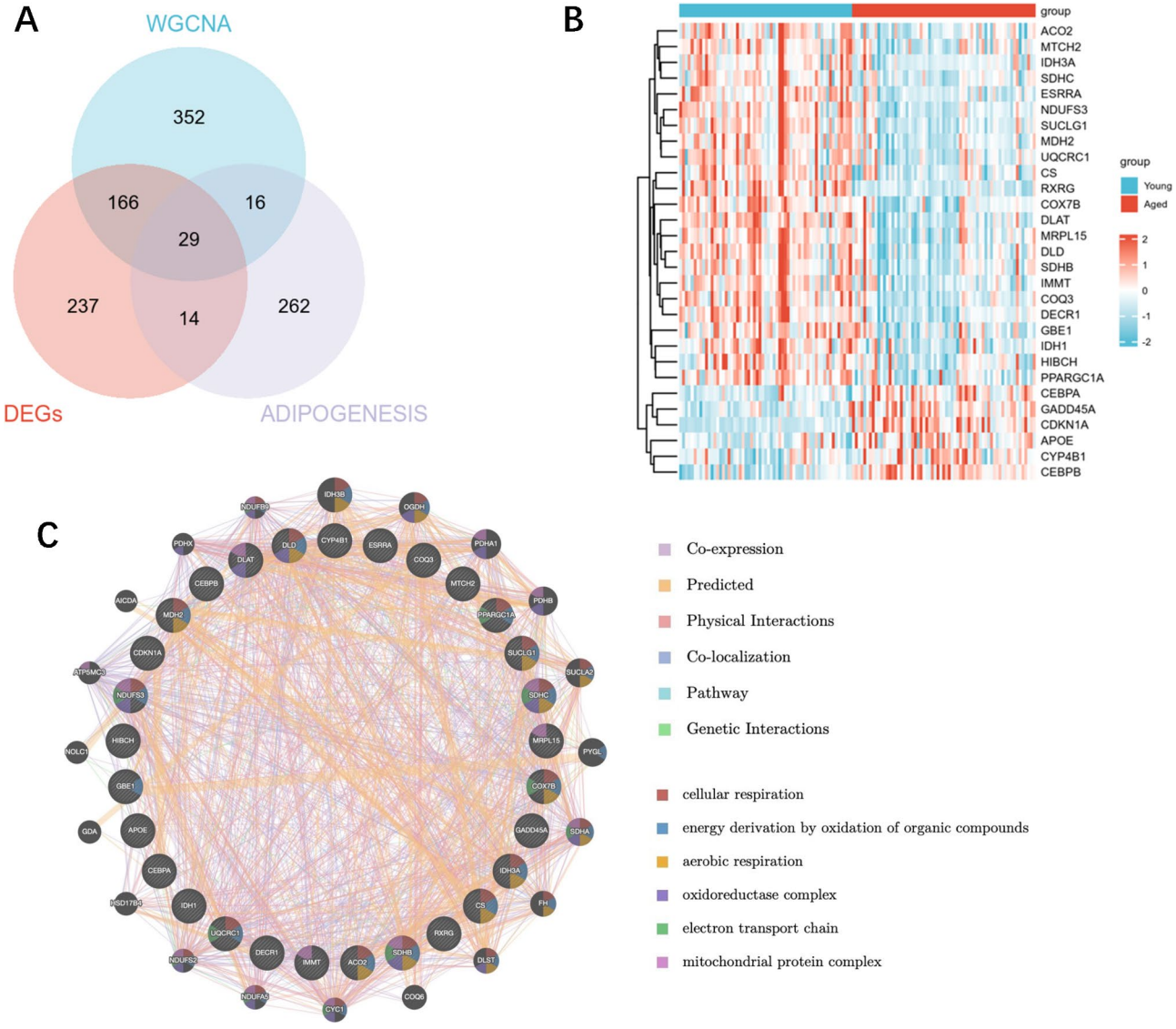

Fig. 4 Identification and characteristics of adipogenesis-related hub genes in muscle aging. (**A**) Venn plot of adipogenesis-related hub genes. (**B**) Heatmap of adipogenesis-related hub genes expression in merged training sets. (**C**) Interaction and their co-expression network of adipogenesis-related hub genes

these ARDEGs, we conducted a gene interaction network analysis using the online website Genemania. As shown in Fig. 4C, we discovered a complex network with 42.70% co-expression, 36.75% predicted interactions, 4.32% physical interactions, and 6.24% co-localization. According to a function enrichment analysis, ARDEGs were mostly linked to cellular respiration, energy derivation by oxidation of organic compounds, aerobic respiration, oxidoreductase complex, electron transport chain and mitochondrial protein complex.

Drug–Gene interaction of ARDEGs

Developing possible therapeutic medicines that target ARDEGs offers a unique therapy strategy. The relevant ARDEGs obtained from the analysis were imported into the cMAP database platform according to the format files of "up-regulated" genes and "down regulated" genes respectively, so as to screen the small molecule compound drugs with potential therapeutic effects (starting from the candidate ingredients with negative scores. The top 3 ingredients were selected as shown in Table 2 for information.

Identification of best feature genes using machine learning analysis

Using the lasso algorithm, 10 characteristic genes including ESRRA, GADD45A, GBE1, IDH1, CDKN1A, CEBPB, PPARGC1A, RXRG, APOE and CYP4B1 were finally screened out, as shown in Figure S8A&B. The SVM-RFE algorithm results showed that when the number of features was 22, the classifier error was the smallest. So 22 characteristic genes were finally selected including RXRG, CDKN1A, CEBPB, MDH2, HIBCH, GADD45A, ESRRA, CYP4B1, SUCLG1, IDH1, APOE, DECR1, SDHC, CS, CEBPA, MRPL15, GBE1, IMMT, UQCRC1, COQ3, DLAT, and IDH3A, as shown in Figure S8C. The random forest analysis was used to measure the classification importance of characteristic genes through the reduction of Mean Decrease Gini, and finally 20 most significant genes were identified as characteristic genes, including CDKN1A, CEBPB, RXRG, SUCLG1, DECR1, ESRRA, GADD45A, IMMT, IDH3A, COX7B, NDUFS3, DLAT, MRPL15, CEBPA, IDH1, APOE, UQCRC1, DLD, PPARGC1A, and CYP4B1, as shown in Figure S8D. A total of 8 best feature genes were identified by taking the intersection of the feature genes selected by these three algorithms, namely ESRRA, GADD45A, IDH1, CDKN1A, CEBPB, RXRG, APOE, and CYP4B1 as shown in Fig. 5A.

Table 2 The details of the top 3 candidate small molecular drugs targeting ARDEGs

PubChem CID	Name	Target name	MOA	Chemical Formula
129211	tamsulosin	ADRA1A, ADRA1B, ADRA1D	Adrenergic receptor antagonist	$C_{20}H_{28}N_2O_5S$
3083616	fraxidin	CA12	Carbonic anhydrase inhibitor	$C_{11}H_{10}O_5$
2081	alaproclate	SLC6A4	Serotonin receptor antagonist	$C_{13}H1_8C_lNO_2$

We performed a verification in three independent datasets using ROC curve, which was commonly used to judge the accuracy of the best characteristic genes. When the AUC of genes was ≤ 0.7, this usually indicated lower diagnostic value. In GSE22284 datasets, the area under the ROC of genes including ESRRA, CDKN1A, GADD45A, RXRG and CEBPB were greater than 0.7(Fig. 5B). In the dataset GSE9103, ESRRA, GADD45A, CDKN1A, CEBPB, RXRG, and APOE had the area under the ROC greater than 0.7(Fig. 5C). In the dataset GSE164471, GADD45A, IDH1, CYP4B1, and APOE had the area under the ROC greater than 0.7(Fig. 5D). And among the three verification sets, the genes appeared more than once which had area under the ROC greater than 0.7 were GADD45A, CDKN1A, CEBPB, APOE and RXRG.

Identification and verification of the hub ARDEGs in aged muscles

In order to further find the hub ARDEGs, we constructed Friends analysis, in which the importance of each gene was calculated by using the network topology parameters. As shown in Fig. 6A, the cloud rain diagram showed the similarity between 8 genes and other genes, with the top gene representing the key gene with the highest similarity to other genes. And we found that ESRRA, GADD45A, CEBPB, and RXRG were highly interacted, so we chose them as the hub ARDEGs in muscle aging.

To further confirm the expression of ESRRA, RXRG, GADD45A, and CEBPB in vivo, we used aged mice as the muscle aging model. The results showed that compared to young mice group, ESRRA, RXRG were significantly decreased while GADD45A, and CEBPB were significantly increased, which were consistent with our findings in the training datasets (Fig. 6B).

To more precisely delineate the expression of hub ARDEGs in human muscle tissues, we utilized the HLMA database to identify the cell subpopulations expressing in muscle aging. Generally, we found the expression of GADD45A, and CEBPB were significantly increased in old population, and RXRG were significantly decreased according to the heatmap shown in Figure S9A. As shown in Figure S9B, there were 15 cell subpopulations in muscle samples, and we found GADD45A, and CEBPB abundantly expressed in the majority of the muscle samples. We further illustrated the different expression of the hub ARDEGs in different ages (Figure S9C). There seemed

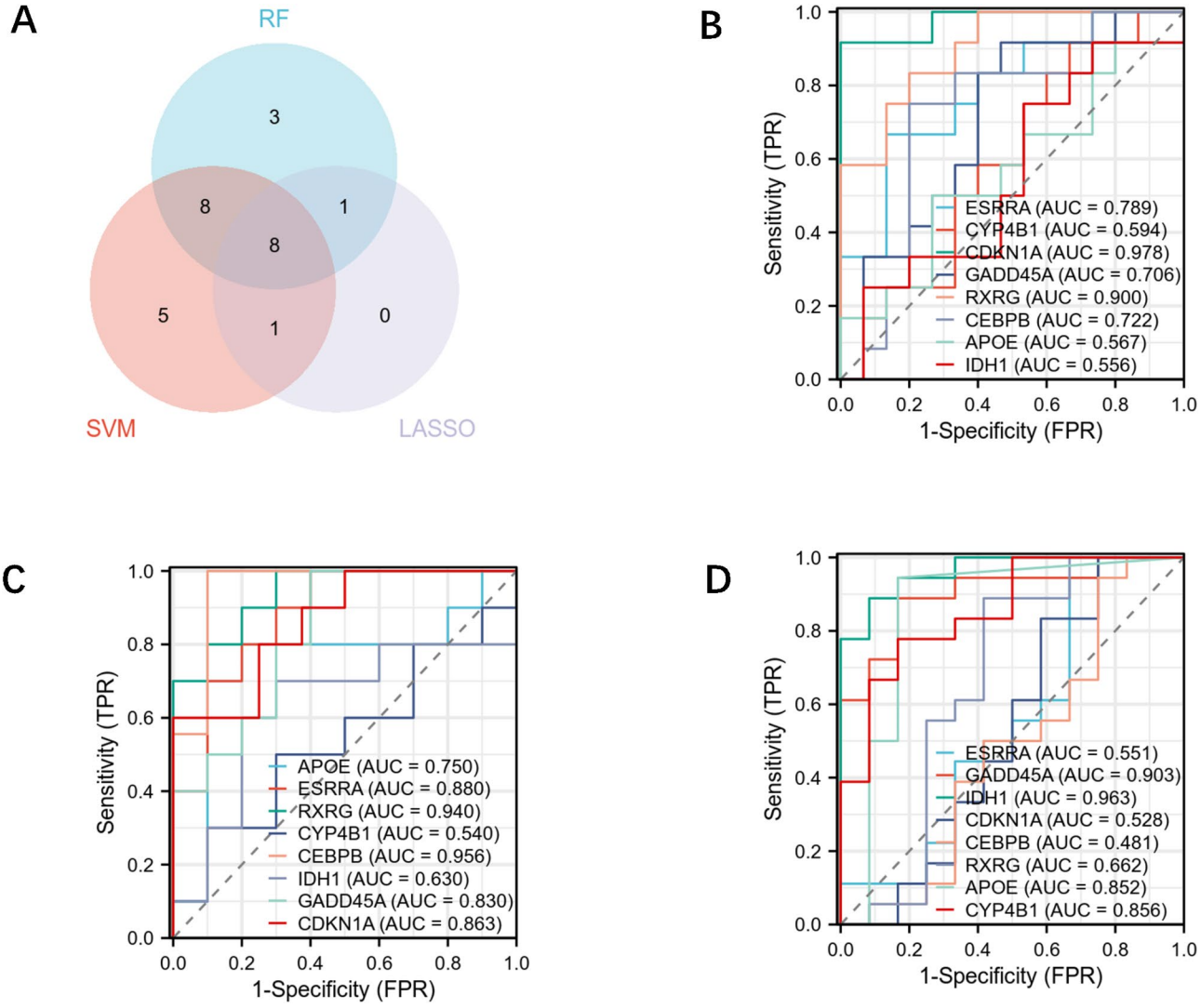

Fig. 5 Identification and verification of feature genes using machine learning analysis. (**A**) The intersection results for LASSO, SVM-RFE, and Random Forest. (**B**) ROC curves estimate the diagnostic values of 8 best feature genes in GSE28422. (**C**) ROC curves estimate the diagnostic values of best feature genes in GSE9103. (**D**) ROC curves estimate the diagnostic values of best feature genes in GSE164471

not very big expression changes of ESRRA, RXRG. Interestingly, we found that GADD45A, and CEBPB were together higher expressed in the fibroblast like cells, FAP, MuSC, pericyte and SMC in the old population which might indicate upregulation of GADD45A, and CEBPB in these cells highly involved in the pathogenesis of muscle aging.

GSEA of hub ARDEGs

Furthermore, we investigated the precise signaling pathways and the probable biological processes of the hub ARDEGs in muscle aging (Figure S10 & Tables S8–S11). Based on the median gene expression level, all samples in muscle aging group were split into low-expression groups (< median gene expression) and high-expression groups (≥ median gene expression). A GSEA between low-expression groups and high-expression groups was performed [29, 43]. The significant conditions were padj < 0.05 and FDR (q value) < 0.25.The top 5 HPO results revealed that ESRRA was mostly relevant for muscle abnormality related to mitochondrial dysfunction, proximal tubulopathy, hypertrophic cardiomyopathy, abnormal circulating proteinogenic amino acid concentration and emotional lability(Figure S10A). The main enriched items for RXRG were aplasia of the musculature, hyperphosphaturia, dyslexia, exaggerated cupid s bow, and mandibular prognathia(Figure S10B). The main enriched items for GADD45A were adrenocortical carcinoma, ependymoma, pituitary growth hormone cell adenoma, shortened qt interval, adrenocortical abnormality (Figure S10C). As for CEBPB, the main enriched items were abnormality of the nasal dorsum, hyperactivity, delayed

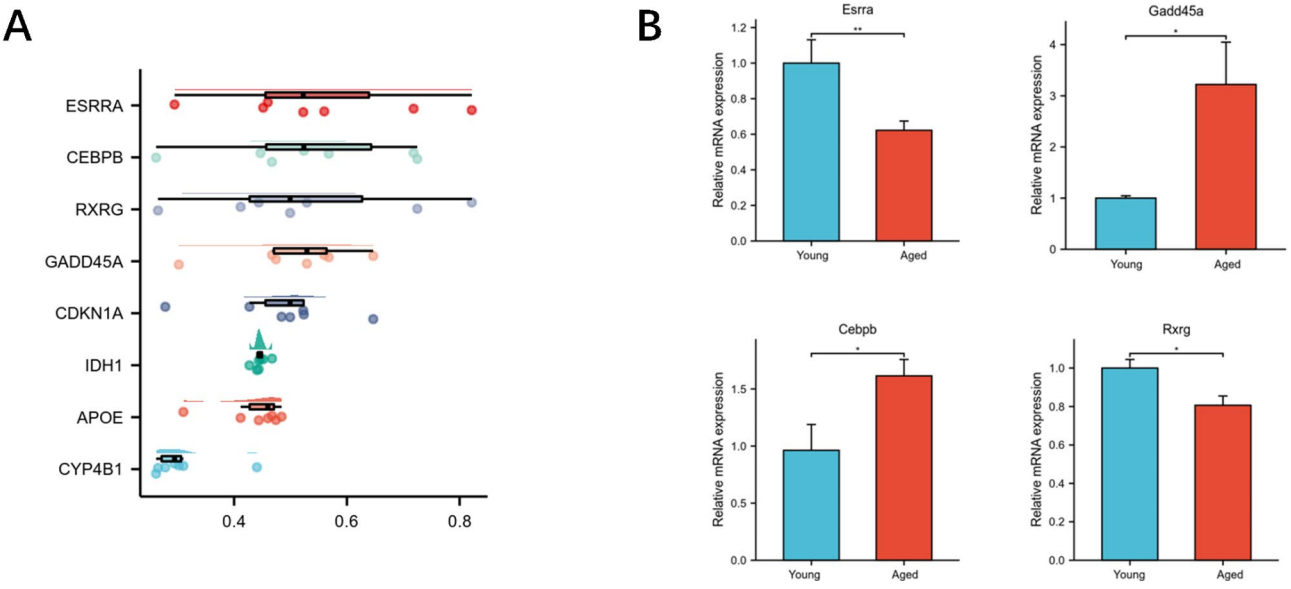

Fig. 6 Identification and verification of key adipogenesis-related genes in sarcopenia. (**A**) Friends analysis of the 8 best feature genes. (**B**) RT-PCR verification in aged mice

gross motor development, typical absence seizure, and oral cleft (Figure S10D).

In addition, the top 5 GO BP results revealed that ESRRA regulated aortic valve morphogenesis, developmental induction, morphogenesis of a branching structure, morphogenesis of an epithelium and prostate glandular acinus morphogenesis (Figure S10E). The main enriched terms for RXRG were postsynapse assembly, positive regulation of synapse assembly, positive regulation of synaptic transmission glutamatergic, sensory perception of bitter taste, and stabilization of membrane potential(Figure S10F). The main enriched terms for GADD45A expression were myelin assembly, mitotic cell cycle arrest, cellular response to ionizing radiation, dna strand elongation, regulation of superoxide metabolic process(Figure S10G). As for CEBPB, the main enriched terms were protein destabilization, mitotic g1 s transition checkpoint, photoperiodism, entrainment of circadian clock, and histone h4 acetylation (Figure S10H).

Meanwhile, KEGG gene sets found that ESRRA was primarily enriched in cysteine and methionine metabolism, parkinsons disease, huntingtons disease, oxidative phosphorylation, and amino sugar and nucleotide sugar metabolism (Figure S10I). The main enriched pathways for RXRG were oocyte meiosis, aldosterone regulated sodium reabsorption, ascorbate and aldarate metabolism, phosphatidylinositol signaling system, and endocytosis (Figure S10J). The main enriched pathways for GADD45A were dna replication, bladder cancer, cell cycle, lysosome, and thyroid cancer (Figure S10K). As for CEBPB, the main enriched pathways were nucleotide excision repair, cell cycle, base excision repair, circadian rhythm mammal, long term potentiation (Figure S10L).

Correlation between hub ARDEGs and immune cells

To analyze immunological patterns in aged and normal tissues, we used ssGSEA to compute the proportion of 24 immune cells in each sample. As shown in Fig. 7A, compared to the young sample, CD8 T cells, dendritic cells (DC), immature dendritic cells (iDC), mast cells, natural killer (NK) CD56bright cells, NK CD56dim cells, and Tcm infiltrated more significantly in muscle aging group. In the subsequent study, the correlation between immuno-infiltrated cells was investigated (Fig. 7B) and the results showed they were actively correlated. In order to explore the relationship between hub ARDEGs expression and immune cell abundance, we firstly explored their expression in immune cells using HLMA database. As shown in Fig. 7C, GADD45A has high expression in $CCL20^{+}$ TC and $CD8^{+}$ native TC, while CEBPB has high expression in $CD14^{+}$ Mono cells, $CD16^{+}$ Mono cells, and LAM MΦ cells. Then we performed a Pearson's correlation between hub ARDEGs and immune cells infiltration scores. As shown in Fig. 7D&E, ESRRA and RXRG was together highly negatively corelated with CD8 T cells and DC cells ($p<0.001$). GADD45A and CEBPB (Fig. 7F&G) were together highly negatively corelated with NK cells, and highly positively corelated with CD8 T cells ($p<0.001$).

Discussions

Previous research has demonstrated that muscle mass experiences a significant decline with advancing age. Notably, a reduction in muscle mass alone does not necessarily result in physical functional impairment in individuals. In contrast, alterations in muscle strength appear to exert a more substantial impact. This phenomenon has

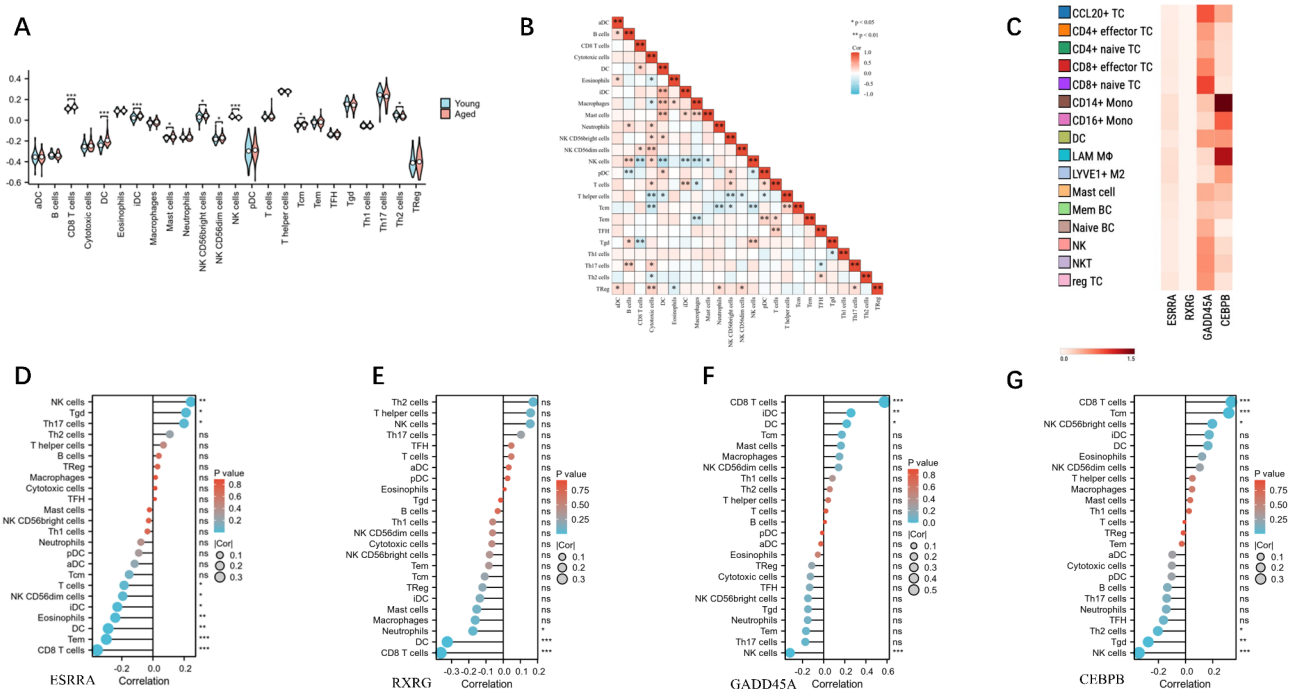

Fig. 7 Immune infiltration analysis in training datasets. (**A**) Immune cells infiltration scores in young and aged groups. (**B**) Correlation analysis between immuno-infiltrated cells. (**C**) Key adipogenesis-related genes' expression in immune cells using HLMA database. (**D-G**) Correlation analysis between key adipogenesis-related genes and immune cells infiltration scores. (**D**) ESRRA, (**E**) RXRG, (**F**) GADD45A, and (**G**) CEBPB

been attributed to an increase in intermuscular fat infiltration (IMAT), which had been reported to adversely affect muscle function [44, 45]. Under normal physiological conditions, adipose tissue is responsible for energy storage, providing substrates for metabolic processes, secreting adipokines and inflammatory factors, regulating the local muscle microenvironment, and facilitating signal transduction, among other functions. As individuals age, an imbalance occurs when adipogenesis surpasses the storage capacity of IMAT, resulting in local fat infiltration. This process, known as ectopic fat deposition within skeletal muscle, impairs muscle function and consequently elevates the risk of muscle aging [46]. Therefore, elucidating the pathological molecular mechanisms associated with adipogenesis is crucial for advancing our comprehension of pathogenesis within muscle aging and for establishing a theoretical foundation for the development of therapeutic interventions. In this study, we integrated datasets related to muscle aging and performed an enrichment analysis of differential gene pathways. Our findings indicate that adipogenesis plays a significant role in aged muscles (Fig. 2E). We employed WGCNA and integrated the findings with data obtained from DEGs and adipogenesis datasets. This comprehensive approach led to the identification of 29 ARDEGs. Interestingly subsequent analysis of gene interactions and functional pathways indicated that these 29 genes were intricately connected and predominantly involved in mitochondrial-related oxidative phosphorylation and energy regulation.

Mitochondria have traditionally been regarded as the powerhouses of the cell due to their continuous production of ATP, which is essential for maintaining normal cellular functions, regulating cell proliferation and calcium ion (Ca2+) homeostasis, and integrating apoptotic signals. Mitochondrial dysfunction can result in diminished ATP synthesis, bioenergetic impairment, and mutations in mitochondrial DNA, all of which have been previously reported to be associated with the loss of skeletal muscle mass and functional decline [47, 48]. In the context of muscle aging, fat infiltration has been closely associated with mitochondrial dysfunction. This relationship is underscored by the observation that diminished mitochondrial biogenesis may reduce the oxidative capacity of muscle cells, thereby promoting the differentiation of skeletal muscle stem cells into adipocytes [49]. Moreover, an augmentation in the quantity of adipocytes within skeletal muscle may further inhibit mitochondrial biogenesis, consequently diminishing the capacity for fatty acid oxidation metabolism in skeletal muscle and exacerbating the accumulation of IMAT [46]. Hence, the concurrent presence of fat infiltration and mitochondrial dysfunction may synergistically contribute to the onset and progression of muscle aging through interconnected mechanisms, thereby impacting muscle mass and function just as the results we found.

At present, there remains a paucity of research investigating the specific genes associated with muscle aging and fat infiltration. In this study, we employed three machine learning algorithms to identify characteristic genes of muscle aging from the perspective of adipogenesis. Our analysis revealed the genes ESRRA, GADD45A, CEBPB, and RXRG as significant. We subsequently verified the accuracy of these findings through ROC analysis and additional qPCR experiments conducted in aged mice. The findings consistently demonstrated that ESRRA and RXRG were downregulated, whereas GADD45A and CEBPB were upregulated in both aged mice and human. These molecular alterations contribute to the development of diagnostic and therapeutic strategies for muscle aging. ESRRA, a member of the nuclear receptor superfamily, is predominantly expressed in metabolically active tissues that utilize fatty acids as an energy source [50, 51]. Previous studies have demonstrated that fasting, energy supply deficiency, cold exposure, and exercise could induce the expression and activation of ESRRA [52]. ESRRA has been shown to regulate the expression of genes involved in various energy-related pathways, including oxidative phosphorylation, the citric acid cycle, beta-oxidation of fatty acids, glucolipid metabolism, as well as mitochondrial autophagy and self-renewal [50, 53]. LaBarge Samuel et al. developed muscle-specific ESRRA knockout (ESRRA(-/-)) mice and systematically analyzed the temporal progression of skeletal muscle regeneration following injury induced by intramuscular cardiotoxin injection. Their findings indicated that, in comparison to the control group, the M-ESRRA(-/-) mice exhibited compromised muscle regeneration. This impairment was attributed to the deficiency of ESRRA, which adversely affected the recovery of mitochondrial energy capacity and disrupted AMPK activity, ultimately resulting in delayed muscle fiber repair [54]. Herrera Uribe Juber et al. demonstrated that intervention measures combining diet management and exercise upregulate the expression of ESRRA in the muscles of overweight pet dogs [55]. Similarly, our study observed a significant downregulation of ESRRA, corroborating the findings of Jingbao Kan et al. [56]. Therefore, therapies that facilitate the upregulation of ESRRA expression levels may be considered as potential treatments for muscle aging.

The retinoid X receptors (RXRs) belong to the nuclear receptor superfamily of ligand-dependent transcription factors and are categorized into three distinct subtypes: RXRα, RXRβ, and RXRγ (RXRG) [57, 58]. RXRs occupy a pivotal role within the nuclear receptor superfamily, demonstrating the ability to form RXR-RXR homodimers, as well as heterodimers or heterooligomers with a variety of other nuclear receptors. These include, but are not limited to, retinoic acid receptors (RARs), peroxisome proliferator-activated receptors, cellular retinoic acid-binding protein II, cellular retinol-binding protein I, interleukin-2 receptor alpha, transglutaminase 2, cytochrome P450 family 26 subfamily A member 1, hepatocyte nuclear factors, and Homeobox proteins [58]. Through these interactions, RXRs play a crucial role in regulating diverse physiological processes such as lipid and glucose metabolism, as well as immune responses [59]. RXRG was predominantly expressed in muscle and brain tissues. Prior studies have demonstrated that RXRG is integral to the regulation of adipogenesis and lipid metabolism. In a study utilizing a chicken muscle adipocyte differentiation model, Zhang Meng et al. discovered that gga-miR-140-5p facilitates intramuscular adipocyte differentiation by targeting RXRG [60]. Zang et al. demonstrated that severe infection resulted in the rapid suppression of RXRG expression in type 2 innate lymphoid cells, leading to reduced cholesterol efflux and increased intracellular neutral lipid accumulation [61]. In our study, we observed a significant reduction in RXRG expression, suggesting that upregulation of RXRG may also represent a potential therapeutic target for muscle aging.

The GADD45A protein, a small acidic protein regulated by p53, is predominantly localized in the cell nucleus where it interacts with various nuclear proteins involved in cell cycle regulation [62, 63]. Prior research has demonstrated an upregulation of GADD45A expression in atrophied muscles of mice [64–67],, indicating a potential role for GADD45A in the regulation of muscle atrophy and aging. Ebert et al. employed an ATF4 muscle-specific mouse model to investigate the molecular mechanisms underlying skeletal muscle weakness and atrophy. The study demonstrated that mice with a deficiency in ATF4 within skeletal muscle fibers exhibited elevated levels of muscle protein synthesis. Conversely, the induction of ATF4 expression resulted in an upregulation of Gadd45a protein in skeletal muscle fibers, thereby accelerating the process of muscle aging [68]. Jeffrey T. Ehmsen et al. identified Gadd45a as one of the earliest and most consistently upregulated genes in denervated skeletal muscle [65]. Bullard Steven A., et al. demonstrated that the Gadd45a protein facilitates skeletal muscle atrophy through the formation of a complex with the protein kinase MEKK4 [64]. Marcotte George, et al. generated transgenic mice expressing Gadd45a and discovered that Gadd45a in skeletal muscle acts as a mediator of mitochondrial loss, atrophy, and muscle weakness in mice. Furthermore, they identified Gadd45a as a potential therapeutic target for addressing muscle weakness in humans [66]. CEBPB, also referred to as nuclear factor for IL-6 (NF-IL6), is a critical member of the CCAAT/enhancer binding proteins transcription factor family [69]. This protein possesses a highly conserved DNA-binding domain and a dimerization function

domain located at its C-terminus, which play pivotal roles in various essential biological processes, including cell proliferation and differentiation, tumorigenesis, apoptosis, and the inflammatory response [70, 71]. Previous studies have identified CEBPB as a pivotal regulatory factor in cancer-induced muscle wasting [69, 72, 73]. The overexpression of CEBPB has been shown to inhibit myocyte differentiation, resulting in decreased maturity of muscle fibers [74]. Furthermore, both GADD45A and CEBPB have been implicated in the process of fat infiltration. Research by You et al. demonstrated a positive correlation between GADD45A expression and IMF deposition in both animal and human skeletal muscles. The overexpression of GADD45A was found to promote IMF infiltration, thereby impairing muscle regeneration [75]. Inhibition of GADD45A expression has been shown to reduce fat infiltration and promote muscle regeneration. During the initial stages of adipocyte differentiation, CEBPB is identified as one of the earliest expressed transcription factors, which subsequently enhances adipogenesis by activating the expression of PPARγ and CEBPα genes. Schafere et al. developed Gadd45a/ING1 double-knockout mice and observed that these mice exhibited premature aging phenotypes characteristic of CEBP mutants, along with disorders in energy and fat metabolism [76]. Therefore, in conjunction with our findings, C/EBPB and GADD45A appear to play significant roles in age-related muscle atrophy and IMF infiltration. Consequently, targeting the downregulation of CEBPB and GADD45A may represent a novel therapeutic approach for the treatment of muscle aging.

Chronic low-grade inflammation associated with aging has been identified as a significant contributor to muscle aging. Elevated concentrations of pro-inflammatory cytokines, including interleukin-6, tumor necrosis factor-alpha, and acute-phase C-reactive protein, have been correlated with the deterioration of skeletal muscle mass and diminished physical function [77, 78]. Previous research has demonstrated that age-related adipose tissue accumulation can induce lipotoxic effects, generate local inflammatory mediators, and elicit inflammatory responses. Concurrently, the elevation of inflammatory factors can exacerbate adipose infiltration, thereby contributing to the progression of muscle atrophy and the deterioration of muscle strength [79–81]. To investigate the potential associations between inflammation infiltration and hub genes, we performed an immune cell infiltration analysis utilizing ssGSEA. Our findings demonstrated a pronounced infiltration of CD8 + T cells, DC, iDC, mast cells, and NK cells in the muscle aging group. And the hub genes identified in our study exhibit a strong association with immune cells such as CD8 + T cells and NK cells. Consequently, we hypothesize that aging induces dysregulations of the hub ARDEGS, which subsequently simultaneously activated adipogenesis and the immune cells including T cells, dendritic cells, and NK cells infiltration, produced elevated levels of inflammatory factors, thereby exacerbating muscle aging.

Currently, the pharmacological agents identified for the treatment of muscle aging predominantly encompass vitamin D, growth hormone, androgens/selective androgen receptor modulators, myostatin antibodies, angiotensin-converting enzyme inhibitors, and metformin, among others. However, the efficacy of these agents remains inconclusive [82]. The cMAP database leverages variations in gene expression following the administration of various perturbagens to human cells, thereby creating a bioinformatics resource that links perturbagens, gene expression profiles, and disease states. By uploading gene expression profile data, this website facilitates the comparison of drugs with significant relevance to diseases, elucidates the primary structures of most drug molecules, and summarizes their potential mechanisms of action [40]. In this study, the top three drugs we identified were tamsulosin, fraxidin, and alaproclate, which might exhibit substantial therapeutic potential. Tamsulosin is an alpha 1 receptor blocker primarily used to treat urinary disorders caused by benign prostatic hyperplasia. Studies have shown that tamsulosin may inhibit the growth of tumor cells by affecting the cell cycle [83] and have anti-inflammatory effect [84]. Fraxidin, a natural compound derived from the bark of the Fraxinus plant, possesses a range of pharmacological activities, including anti-inflammatory, antioxidant, antitumor, antimicrobial, and cytoprotective effects, among others [85]. Alaproclate is a pharmacological agent exhibiting antidepressant properties and functioning as a non-competitive NMDA receptor antagonist, previously investigated for potential therapeutic applications in conditions such as depression and Alzheimer's disease [86]. However, there is a paucity of experimental studies examining the efficacy of such drugs in the context of muscle aging. The potential of the three screened compounds to serve as small molecule therapeutics for muscle aging, along with the elucidation of their specific mechanisms of action, necessitates further verification through comprehensive in vivo and in vitro experimentation.

This study is subject to several limitations. Firstly, the use of transcriptome data from publicly available databases may have resulted in the exclusion of critical information regarding patient populations and clinical characteristics. And although we have used statistical approaches to remove the possible bath effects during the analysis, potential batch effects, such as different experimental times, batches, processing methods or sequencing platforms might still exist. Secondly, the verification of hub gene expression was exclusively dependent on public datasets and animal experiments, without corroboration

through human studies. Thirdly, the limited sample size may have affected the accuracy of the results, underscoring the need for a larger sample size and a prospective clinical study design for robust verification of the findings.

Conclusions

In conclusion, this study utilized an extensive analysis of bioinformatics and machine learning algorithms to identify ESRRA, GADD45A, CEBPB, and RXRG as hub ARDEGs implicated in muscle aging. Concurrently, compounds such as tamsulosin, fraxidin, and alaproclate, which targeted ARDEGs, were selected. Additionally, the investigation included an examination of immune cell infiltration in muscle aging and explored the relationship between hub ARDEGs and immune cells. Collectively, these findings established a foundational basis and offer guidance for subsequent research in the diagnosis, monitoring, and potential therapeutic interventions for muscle aging.

Abbreviations

cMAP	Connectivity map
GEO	Gene Expression Omnibus
DEGs	differentially expressed genes
HPO	Human Phenotype Ontology
GO	gene ontology
BP	Biological processes
MF	Molecular functions
CC	Cellular components
KEGG	Kyoto Encyclopedia of Genes and Genomes
GSEA	Gene Set Enrichment analysis
WGCNA	Weighted gene co expression network analysis
ARDEGs	Adipogenesis related differentially expressed genes
PPI	Protein-protein interaction
LASSO	Least absolute shrinkage and selection operator
SVM-RFE	Support Vector Machine-Recursive Feature Elimination
RF	Random forest
ssGSEA	Single sample gene set enrichment analysis
HLMA	Human Muscle Ageing Cell Atlas
MSigDB	Molecular Signatures Database
ROC	Receiver operating characteristic
RT-PCR	Reverse transcription-polymerase chain reaction
DC	Dendritic cells
iDC	Immature dendritic cells
NK	Natural killer
IMAT	Intermuscular fat infiltration
RXRs	Retinoid X receptors

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12891-025-08528-9.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not applicable.

Author contributions

JL and YMZ conceived the study and did the research design. YMZ and LQ performed data analysis, and article writing. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (No. 81900773).

Data availability

Publicly available datasets were analyzed in this study. The data could be downloaded from the GEO database of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/geo/). The codes used during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The experimental protocols of the present study were approved by the Animal Care and Use Committee of Nanjing Medical University (Nanjing, China), and carried out in compliance with the ARRIVE guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable.

Received: 19 November 2024 / Accepted: 12 March 2025
Published online: 22 March 2025

References

1. Boshnjaku A. Is age-related sarcopenia a real concern for my developing country? J Cachexia Sarcopenia Muscle[J]. 2022;13(6):2589–92. https://doi.org/10.1002/jcsm.13107.
2. Petermann-Rocha F, Balntzi V, Gray SR, et al. Global prevalence of sarcopenia and severe sarcopenia: a systematic review and meta-analysis. J Cachexia Sarcopenia Muscle[J]. 2022;13(1):86–99. https://doi.org/10.1002/jcsm.12783.
3. Kingsley J, Torimoto K, Hashimoto T, et al. Angiotensin II Inhibition: a potential treatment to slow the progression of sarcopenia. Clin Sci (Lond)[J]. 2021;135(21):2503–20. https://doi.org/10.1042/CS20210719.
4. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing[J]. 2019;48(4):601. https://doi.org/10.1093/ageing/afz046.
5. Cruz-Jentoft AJ, Sayer AA, Sarcopenia. Lancet[J]. 2019;393(10191):2636–46. https://doi.org/10.1016/S0140-6736(19)31138-9.
6. Li CW, Yu K, Shyh-Chang N, et al. Pathogenesis of sarcopenia and the relationship with fat mass: descriptive review. J Cachexia Sarcopenia Muscle[J]. 2022;13(2):781–94. https://doi.org/10.1002/jcsm.12901.
7. Dhillon RJ, Hasni S. Pathogenesis and management of sarcopenia. Clin Geriatr Med[J]. 2017;33(1):17–26. https://doi.org/10.1016/j.cger.2016.08.002.
8. Sarjeant K, Stephens JM, Adipogenesis. Cold Spring Harb Perspect Biol[J]. 2012;4(9):a008417. https://doi.org/10.1101/cshperspect.a008417.
9. Mota de Sa P, Richard AJ, Hang H, et al. Transcriptional Regul Adipogenesis Compr Physiol[J]. 2017;7(2):635–74. https://doi.org/10.1002/cphy.c160022.
10. Al Saedi A, Debruin DA, Hayes A et al. Lipid metabolism in sarcopenia, Bone[J] 2022,164, 116539.https://doi.org/10.1016/j.bone.2022.116539
11. Xu Z, You W, Chen W, et al. Single-cell RNA sequencing and lipidomics reveal cell and lipid dynamics of fat infiltration in skeletal muscle. J Cachexia Sarcopenia Muscle[J]. 2021;12(1):109–29. https://doi.org/10.1002/jcsm.12643.
12. Wang L, Valencak TG, Shan T. Fat infiltration in skeletal muscle: influential triggers and regulatory mechanism. iScience[J]. 2024;27(3). https://doi.org/10.1016/j.isci.2024.109221.
13. Cho YR, Kang M. Interpretable machine learning in bioinformatics. Methods[J]. 2020;179:1–2. https://doi.org/10.1016/j.ymeth.2020.05.024.
14. Ravi D, Wong C, Deligianni F, et al. Deep learning for health informatics. IEEE J Biomed Health Inform[J]. 2017;21(1):4–21. https://doi.org/10.1109/JBHI.2016.2636665.

15. Gueugneau M, Coudy-Gandilhon C, Chambon C, et al. Muscle proteomic and transcriptomic profiling of healthy aging and metabolic syndrome in men. Int J Mol Sci[J]. 2021;22(8). https://doi.org/10.3390/ijms22084205.
16. Welle S, Tawil R, Thornton CA. Sex-related differences in gene expression in human skeletal muscle. PLoS One[J]. 2008;3(1):e1385. https://doi.org/10.1371/journal.pone.0001385.
17. Giresi PG, Stevenson EJ, Theilhaber J, et al. Identification of a molecular signature of sarcopenia. Physiol Genomics[J]. 2005;21(2):253–63. https://doi.org/10.1152/physiolgenomics.00249.2004.
18. Melov S, Tarnopolsky MA, Beckman K, et al. Resistance exercise reverses aging in human skeletal muscle. PLoS One[J]. 2007;2(5):e465. https://doi.org/10.1371/journal.pone.0000465.
19. Raue U, Trappe TA, Estrem ST, et al. Transcriptome signature of resistance exercise adaptations: mixed muscle and fiber type specific profiles in young and old adults. J Appl Physiol (1985)[J]. 2012;112(10):1625–36. https://doi.org/10.1152/japplphysiol.00435.2011.
20. Lanza IR, Short DK, Short KR et al. Endurance exercise as a countermeasure for aging. Diabetes[J] 2008;57(11):2933–42.https://doi.org/10.2337/db08-0349
21. Tumasian RA 3rd, Kundu HA et al. G., Skeletal muscle transcriptome in healthy aging, Nat Commun[J]. 2021;12(1):2014.https://doi.org/10.1038/s41467-021-22168-2
22. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics[J]. 2007;8(1):118–27. https://doi.org/10.1093/biostatistics/kxj037.
23. Barshir R, Fishilevich S, Iny-Stein T, et al. GeneCaRNA: A comprehensive Gene-centric database of human Non-coding RNAs in the genecards suite. J Mol Biol[J]. 2021;433(11):166913. https://doi.org/10.1016/j.jmb.2021.166913.
24. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res[J]. 2015;43(7):e47. https://doi.org/10.1093/nar/gkv007.
25. Robinson PN, Mundlos S. The human phenotype ontology. Clin Genet[J]. 2010;77(6):525–34. https://doi.org/10.1111/j.1399-0004.2010.01436.x.
26. The Gene Ontology C. The gene ontology resource: 20 years and still going strong. Nucleic Acids Res[J]. 2019;47(D1):D330–8. https://doi.org/10.1093/nar/gky1055.
27. Ogata H, Goto S, Sato K, et al. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res[J]. 1999;27(1):29–34. https://doi.org/10.1093/nar/27.1.29.
28. Liberzon A, Birger C, Thorvaldsdottir H, et al. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst[J]. 2015;1(6):417–25. https://doi.org/10.1016/j.cels.2015.12.004.
29. Yu G, Wang LG, Han Y, et al. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS[J]. 2012;16(5):284–7. https://doi.org/10.1089/omi.2011.0118.
30. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis, BMC Bioinformatics[J] 2008,9, 559.https://doi.org/10.1186/1471-2105-9-559
31. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. Nucleic Acids Res[J]. 2018;46(W1):W60–4. https://doi.org/10.1093/nar/gky311.
32. Li Z, Sillanpaa MJ. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. Theor Appl Genet[J]. 2012;125(3):419–35. https://doi.org/10.1007/s00122-012-1892-9.
33. Sanz H, Valim C, Vegas E, et al. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics[J]. 2018;19(1):432. https://doi.org/10.1186/s12859-018-2451-4.
34. Ellis K, Kerr J, Godbole S, et al. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas[J]. 2014;35(11):2191–203. https://doi.org/10.1088/0967-3334/35/11/2191.
35. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R, BMC Bioinformatics[J] 2011,12, 35.https://doi.org/10.1186/1471-2105-12-35
36. Robin X, Turck N, Hainard A et al. pROC: an open-source package for R and S+to analyze and compare ROC curves, BMC Bioinformatics[J] 2011,12, 77.https://doi.org/10.1186/1471-2105-12-77
37. Yu G. Gene ontology semantic similarity analysis using gosemsim. Methods Mol Biol[J]. 2020;2117:207–15. https://doi.org/10.1007/978-1-0716-0301-7_11.
38. Lai Y, Ramirez-Pardo I, Isern J, et al. Multimodal cell atlas of the ageing human skeletal muscle. Nature[J]. 2024;629(8010):154–64. https://doi.org/10.1038/s41586-024-07348-6.
39. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinformatics[J] 2013,14, 7.https://doi.org/10.1186/1471-2105-14-7
40. Lamb J, Crawford ED, Peck D, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science[J]. 2006;313(5795):1929–35. https://doi.org/10.1126/science.1132939.
41. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta delta C(T)) method, Methods[J] 2001,25(4), 402–8.https://doi.org/10.1006/meth.2001.1262
42. Gu Z, Gu L, Eils R, et al. Circlize implements and enhances circular visualization in R. Bioinformatics[J]. 2014;30(19):2811–2. https://doi.org/10.1093/bioinformatics/btu393.
43. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A[J]. 2005;102(43):15545–50. https://doi.org/10.1073/pnas.0506580102.
44. Goodpaster BH, Bergman BC, Brennan AM, et al. Intermuscular adipose tissue in metabolic disease. Nat Rev Endocrinol[J]. 2023;19(5):285–98. https://doi.org/10.1038/s41574-022-00784-2.
45. Farsijani S, Santanasto AJ, Miljkovic I, et al. The relationship between intermuscular fat and physical performance is moderated by muscle area in older adults. J Gerontol Biol Sci Med Sci[J]. 2021;76(1):115–22. https://doi.org/10.1093/gerona/glaa161.
46. Tian Q, Lee PR, Yang Q, et al. The mediation roles of intermuscular fat and inflammation in muscle mitochondrial associations with cognition and mobility. J Cachexia Sarcopenia Muscle[J]. 2024;15(1):138–48. https://doi.org/10.1002/jcsm.13413.
47. Davin A, Ferrari RR, Pansarasa O. Mitochondria: between aging, frailty and sarcopenia. Aging (Albany NY)[J]. 2023;15(16):7863–5. https://doi.org/10.18632/aging.204998.
48. Alway SE, Mohamed JS, Myers MJ. Mitochondria initiate and regulate sarcopenia. Exerc Sport Sci Rev[J]. 2017;45(2):58–69. https://doi.org/10.1249/JES.0000000000000101.
49. Hulse JL, Habibi J, Igbekele AE, et al. Mineralocorticoid receptors mediate Diet-Induced lipid infiltration of skeletal muscle and insulin resistance. Endocrinology[J]. 2022;163(11). https://doi.org/10.1210/endocr/bqac145.
50. Tripathi M, Yen PM, Singh BK. Estrogen-Related receptor alpha: an Under-Appreciated potential target for the treatment of metabolic diseases. Int J Mol Sci[J]. 2020;21(5). https://doi.org/10.3390/ijms21051645.
51. Yoh K, Ikeda K, Horie K, et al. Roles of Estrogen, Estrogen receptors, and Estrogen-Related receptors in skeletal muscle: regulation of mitochondrial function. Int J Mol Sci[J]. 2023;24(3). https://doi.org/10.3390/ijms24031853.
52. Huang T, Lu Z, Wang Z, et al. Targeting adipocyte ESRRA promotes osteogenesis and vascular formation in adipocyte-rich bone marrow. Nat Commun[J]. 2024;15(1):3769. https://doi.org/10.1038/s41467-024-48255-8.
53. Kim S, Lee JY, Shin SG, et al. ESRRA (estrogen related receptor alpha) is a critical regulator of intestinal homeostasis through activation of autophagic flux via gut microbiota. Autophagy[J]. 2021;17(10):2856–75. https://doi.org/10.1080/15548627.2020.1847460.
54. LaBarge S, McDonald M, Smith-Powell L, et al. Estrogen-related receptor-alpha (ERRalpha) deficiency in skeletal muscle impairs regeneration in response to injury. FASEB J[J]. 2014;28(3):1082–97. https://doi.org/10.1096/fj.13-229211.
55. Herrera Uribe J, Vitger AD, Ritz C, et al. Physical training and weight loss in dogs lead to transcriptional changes in genes involved in the glucose-transport pathway in muscle and adipose tissues. Vet J[J]. 2016;20822–7. https://doi.org/10.1016/j.tvjl.2015.11.002.
56. Kan J, Hu Y, Ge Y, et al. Declined expressions of vast mitochondria-related genes represented by CYCS and transcription factor ESRRA in skeletal muscle aging. Bioengineered[J]. 2021;12(1):3485–502. https://doi.org/10.1080/21655979.2021.1948951.
57. de Lera AR, Bourguet W, Altucci L, et al. Design of selective nuclear receptor modulators: RAR and RXR as a case study. Nat Rev Drug Discov[J]. 2007;6(10):811–20. https://doi.org/10.1038/nrd2398.
58. Sharma S, Shen T, Chitranshi N et al. Retinoid X receptor: cellular and biochemical roles of nuclear receptor with a focus on neuropathological involvement. Mol Neurobiol[J] 2022,59(4), 2027–50.https://doi.org/10.1007/s12035-021-02709-y
59. Gilardi F, Desvergne B. RXRs: collegial partners. Subcell Biochem[J]. 2014;70:75–102. https://doi.org/10.1007/978-94-017-9050-5_5.
60. Zhang M, Li DH, Li F, et al. Integrated analysis of MiRNA and genes associated with meat quality reveals that Gga-MiR-140-5p affects intramuscular fat

deposition in chickens. Cell Physiol Biochem[J]. 2018;46(6):2421–33. https://doi.org/10.1159/000489649.
61. Zang Y, Liu S, Rao Z et al. Retinoid X receptor gamma dictates the activation threshold of group 2 innate lymphoid cells and limits type 2 inflammation in the small intestine. Immunity[J] 2023,56(11), 2542–e25547.https://doi.org/10.1016/j.immuni.2023.08.019
62. Salvador JM, Brown-Clay JD, Fornace AJ. Jr. Gadd45 in stress signaling, cell cycle control, and apoptosis. Adv Exp Med Biol[J]. 2013;793:1–19. https://doi.org/10.1007/978-1-4614-8289-5_1.
63. Jang HJ, Yang JH, Hong E et al. Chelidonine induces apoptosis via GADD45a-p53 regulation in human pancreatic cancer cells, integr cancer Ther[J] 2021,20, 15347354211006191.https://doi.org/10.1177/15347354211006191
64. Bullard SA, Seo S, Schilling B, et al. Gadd45a protein promotes skeletal muscle atrophy by forming a complex with the protein kinase MEKK4. J Biol Chem[J]. 2016;291(34):17496–509. https://doi.org/10.1074/jbc.M116.740308.
65. Ehmsen JT, Kawaguchi R, Kaval D, et al. GADD45A is a protective modifier of neurogenic skeletal muscle atrophy. JCI Insight[J]. 2021;6(13). https://doi.org/10.1172/jci.insight.149381.
66. Marcotte GR, Miller MJ, Kunz HE, et al. GADD45A is a mediator of mitochondrial loss, atrophy, and weakness in skeletal muscle. JCI Insight[J]. 2023;8(22). https://doi.org/10.1172/jci.insight.171772.
67. Ebert SM, Dyle MC, Kunkel SD, et al. Stress-induced skeletal muscle Gadd45a expression reprograms myonuclei and causes muscle atrophy. J Biol Chem[J]. 2012;287(33):27290–301. https://doi.org/10.1074/jbc.M112.374777.
68. Ebert SM, Bullard SA, Basisty N, et al. Activating transcription factor 4 (ATF4) promotes skeletal muscle atrophy by forming a heterodimer with the transcriptional regulator C/EBPbeta. J Biol Chem[J]. 2020;295(9):2787–803. https://doi.org/10.1074/jbc.RA119.012095.
69. Akira S, Isshiki H, Sugita T, et al. A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO J[J]. 1990;9(6):1897–906. https://doi.org/10.1002/j.1460-2075.1990.tb08316.x.
70. Nerlov C. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. Trends Cell Biol[J]. 2007;17(7):318–24. https://doi.org/10.1016/j.tcb.2007.07.004.
71. Zahid MDK, Rogowski M, Ponce C, et al. CCAAT/enhancer-binding protein beta (C/EBPbeta) knockdown reduces inflammation, ER stress, and apoptosis, and promotes autophagy in oxLDL-treated RAW264.7 macrophage cells. Mol Cell Biochem[J]. 2020;463(1–2):211–23. https://doi.org/10.1007/s11010-019-03642-4.
72. Sun R, Zhang S, Hu W, et al. Valproic acid attenuates skeletal muscle wasting by inhibiting C/EBPbeta-regulated atrogin1 expression in cancer cachexia. Am J Physiol Cell Physiol[J]. 2016;311(1):C101–15. https://doi.org/10.1152/ajpcell.00344.2015.
73. Zhang G, Jin B, Li YP. C/EBPbeta mediates tumour-induced ubiquitin ligase atrogin1/MAFbx upregulation and muscle wasting. EMBO J[J]. 2011;30(20):4323–35. https://doi.org/10.1038/emboj.2011.292.
74. AlSudais H, Rajgara R, Saleh A, et al. C/EBPbeta promotes the expression of atrophy-inducing factors by tumours and is a central regulator of cancer cachexia. J Cachexia Sarcopenia Muscle[J]. 2022;13(1):743–57. https://doi.org/10.1002/jcsm.12909.
75. You W, Liu S, Ji J, et al. Growth arrest and DNA damage-inducible alpha regulates muscle repair and fat infiltration through ATP synthase F1 subunit alpha. J Cachexia Sarcopenia Muscle[J]. 2023;14(1):326–41. https://doi.org/10.1002/jcsm.13134.
76. Schafer A, Karaulanov E, Stapf U, et al. Ing1 functions in DNA demethylation by directing Gadd45a to H3K4me3. Genes Dev[J]. 2013;27(3):261–73. https://doi.org/10.1101/gad.186916.112.
77. Tuttle CSL, Thang LAN, Maier AB. Markers of inflammation and their association with muscle strength and mass: A systematic review and meta-analysis. Ageing Res Rev[J]. 2020;64101185. https://doi.org/10.1016/j.arr.2020.101185.
78. Sharma B, Dabur R. Role of Pro-inflammatory cytokines in regulation of skeletal muscle metabolism: A systematic review. Curr Med Chem[J]. 2020;27(13):2161–88. https://doi.org/10.2174/0929867326666181129095309.
79. Savary S, Trompier D, Andreoletti P, et al. Fatty acids - induced lipotoxicity and inflammation. Curr Drug Metab[J]. 2012;13(10):1358–70. https://doi.org/10.2174/138920012803762729.
80. Kawai T, Autieri MV, Scalia R. Adipose tissue inflammation and metabolic dysfunction in obesity. Am J Physiol Cell Physiol[J]. 2021;320(3):C375–91. https://doi.org/10.1152/ajpcell.00379.2020.
81. Costamagna D, Costelli P, Sampaolesi M et al. Role of inflammation in muscle homeostasis and myogenesis, Mediators Inflamm[J] 2015,2015, 805172.DOI:https://doi.org/10.1155/2015/805172
82. Cho MR, Lee S, Song SK. A review of sarcopenia pathophysiology, diagnosis, treatment and future direction. J Korean Med Sci[J]. 2022;37(18):e146. https://doi.org/10.3346/jkms.2022.37.e146.
83. Cicero AFG, Allkanjari O, Busetto GM, et al. Nutraceutical treatment and prevention of benign prostatic hyperplasia and prostate cancer. Arch Ital Urol Androl[J]. 2019;91(3). https://doi.org/10.4081/aiua.2019.3.139.
84. Alabdali H, Algaem M. Study anti-inflammatory effect of Tamsulosin in rat by evaluation IL-4, IL-6 and TNF-α: airway model, Iraqi journal of pharmaceutical sciences (P-ISSN 1683–3597 E-ISSN 2521–3512)[J] 2023,32, 275–81.https://doi.org/10.31351/vol32iss1pp275-281
85. Zheng Z, Sun C, Zhong Y et al. Fraxini cortex: progresses in phytochemistry, Pharmacology and ethnomedicinal uses. J Ethnopharmacol[J] 2024,325, 117849. https://doi.org/10.1016/j.jep.2024.117849
86. Ogren SO, Holm AC, Hall H, et al. Alaproclate, a new selective 5-HT uptake inhibitor with therapeutic potential in depression and senile dementia. J Neural Transm[J]. 1984;59(4):265-8 8. https://doi.org/10.1007/BF01255596

Publisher's note