

Systematic and integrated analysis of tRNA-derived small RNAs reveals novel potential pathogenesis targets of sarcoidosis

Min Zhao

Second Affiliated Hospital of Jilin University

Chang Tian

Second Affiliated Hospital of Jilin University

Xin Di

Second Affiliated Hospital of Jilin University

Xin Jin

Second Affiliated Hospital of Jilin University

Shan Cong

Second Affiliated Hospital of Jilin University

Yingshu Cao

Second Affiliated Hospital of Jilin University

Ke Wang (✉ wke@jlu.edu.cn)

Second Affiliated Hospital of Jilin University

Research Article

Keywords: Sarcoidosis, tRNA-derived small RNA (tsRNA), Pathogenesis, Bioinformatics analysis

Posted Date: January 12th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1230171/v1>

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

[Read Full License](#)

Abstract

The pathogenesis of sarcoidosis, which involves several systems, is unclear, and its pathological type is non-caseating epithelioid granulomas. tRNA-derived small RNA (tsRNA) is a novel class of short non-coding RNAs with potential regulatory functions. However, whether tsRNA contributes to sarcoidosis pathogenesis remains unclear. Deep sequencing technology was used to identify alterations in tsRNA expression profiles between patients with sarcoidosis and healthy controls. A total of 360 tsRNAs were identified for exact matches. Among them, the expression of three tRNAs (tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007) was markedly regulated in sarcoidosis and validated by quantitative real-time polymerase chain reaction. The expression of various tsRNAs was significantly correlated with age, the number of affected systems, and calcium levels in the blood. Additionally, target prediction and bioinformatics analyses revealed that these tsRNAs may play roles in chemokine, cAMP, cGMP-PKG, retrograde endorphin, and FoxO signalling pathways. The Cytoscape software was used for visual analysis to obtain 10 hub genes of each target tsRNA. Among the hub genes, *APP*, *PRKACB*, *ARRB2*, and *NR5A1* finding may participate in the occurrence and development of sarcoidosis through immune inflammation. This study provides novel insights to explore tsRNA as a novel and efficacious pathogenic target of sarcoidosis.

Introduction

Sarcoidosis is a multisystem inflammatory disease of unknown aetiology that is characterised by non-caseating epithelioid granulomatous lesions (aggregates of lymphocytes, macrophages, epithelioid cells, and giant cells) ^[1]. Typical clinical features include bilateral hilar lymph node lesions, pulmonary infiltration, and eye and skin lesions. Some patients may also have neurological and cardiac involvement. The diagnosis of sarcoidosis is based on the results of clinical and radiological examinations, as well as the pathological diagnosis of non-caseating granulomas on biopsy. Although the understanding of the immunopathological and genetic characteristics of sarcoidosis has increased, and there have been new developments in diagnostic techniques and biological agents for treatment ^[2], the pathogenesis of sarcoidosis remains unclear. It is generally accepted that sarcoidosis is caused by abnormal immune responses in genetically susceptible individuals exposed to specific environmental factors. As its cause is unclear, there is still a lack of truly effective treatments. In addition, due to the lack of reliable monitoring indicators for the disease process, sarcoidosis has become a clinically difficult disease to manage ^[3]. To address these problems, methods for rapid and non-invasive identification at any time point during the disease are required; serum biomarker assessment may be used for this purpose ^[4]. Therefore, further research on the serum of sarcoidosis patients may help to clarify the pathogenesis and discover new potential targets for the diagnosis and treatment of sarcoidosis.

A biomarker is measurable in the serum or tissue and may be of diagnostic, prognostic, or predictive value ^[5-7]. Given its non-invasive and disease-specific features, considerable efforts have been made to identify protein blood biomarkers of sarcoidosis in the past few decades. For example, angiotensin

converting enzyme (ACE), lysozyme, soluble IL-2 receptor (sIL-2R), and chitotriosidase (CTO) have all been proposed to be useful for diagnosis in the clinic ^[8]; however, there is limited evidence regarding currently available biomarkers in determining prognosis and predicting patterns of clinical behaviour, with the lack of sufficient sensitivity and specificity ^[9]. However, the number of protein biomarkers and their clinical use are limited. Thus, more research on cohorts of patients with explicit clinical diagnosis and treatment is required.

Recently, with the development of genomics and RNomics, increasing amounts of DNA, RNA, and non-coding RNA (ncRNA) molecules have been discovered in human serum. Cell-free nucleic acids (cfNAs) carrying sarcoidosis-related genetic and epigenetic modifications may be effective blood biomarkers ^[10]. Ideally, qualified biomarkers should be detected in a minimally invasive liquid biopsy, such as a blood sample, with prerequisites, including high concentrations and steady state in the biological sample. Therefore, a novel class of tRNA-derived small RNAs (tsRNAs) has attracted our attention. As the second most common small ncRNAs after miRNAs, an increasing amount of data suggests that these ncRNAs are thriving functional molecules. Similar to miRNAs, stable stem-loop structures, stable-formed nucleoprotein complexes, and exosome encapsulation enable tsRNAs to remain at high concentrations in serum ^[11–14]. Therefore, we hypothesised that tsRNA could serve as a biomarker that can be used for diagnosis and, in some cases, have prognostic implications.

tsRNAs are the result of specific cleavage of mature or pre-tRNA by specific nucleases, such as Dicer and angiogenin (ANG), in specific cells or tissues or under certain conditions, such as stress and hypoxia ^[15, 16]. tsRNAs generally consist of two parts: (1) tRNA-derived fragments (tRNA-derived fragments, tRFs, or tDRs) and (2) half tRNA (tRNA halves, tRHs, or tiRNAs). tsRNAs have been confirmed to regulate gene expression and contribute to development, differentiation, inflammation, and tumour oncogenesis ^[13]. In addition, accumulating evidence has shown that tsRNA can be secreted and exist as stable entities in peripheral blood because they are protected from RNase activity by being enclosed in micro vehicles, such as exosomes, or by forming stable macromolecular complexes ^[17, 18]. The heterogeneity and stability of tsRNAs in blood make them suitable candidates for sarcoidosis diagnosis.

We aimed to use transcriptomics high-throughput sequencing to detect the expression profile of sarcoidosis serum tRNA-derived fragments (tsRNA) and obtain differentially expressed tsRNA molecules, which may be a specific marker for the diagnosis of sarcoidosis. In combination with relevant clinical data, the clinical application value of tsRNA was analysed. Thereafter, the target mRNA of the differentially expressed tsRNA was predicted using bioinformatics technology, and the roles of tsRNA and target mRNA in the occurrence and development of sarcoidosis were preliminarily discussed.

Results

Expression profiles of tsRNAs

A total of 4,021 tsRNAs mapped to 334 unique tRNAs, including tRNAs in the cytosol or mitochondria, were identified. Generally, tRFs are classified into four subtypes according to their original sites in pre-tRNA or mature tRNA, including tRF-1, tRF-3, tRF-5 and i-tRF. tRF-1 is derived from the 3' end of the precursor tRNA, and tRF-5 and tRF-3 are derived from the 5' and 3' ends of mature tRNA. While i-tRF is produced by internal fragments of mature tRNA. tiRNA could produce tiRNA-3 and tiRNA-5 [13]. Based on this classification, we further analysed the proportion of tRF subtypes in patients with sarcoidosis and healthy controls. The pie charts show the distribution of the number for each subtype of tsRNA with a threshold average CPM ≥ 20 . A total of 231 tsRNAs for exact matches were identified in the serum (207 in sarcoidosis, 214 in healthy controls). In the sarcoidosis group, two tRF-2, 33 tRF-3a, 22 tRF-3b, 38 tRF-5a, 24 tRF-5b, 76 tRF-5c, and 12 tiRNA-5 were identified in the sarcoidosis group (Figure 1A). In the control group, four tRF-1, two tRF-2, 34 tRF-3a, 26 tRF-3b, 38 tRF-5a, 23 tRF-5b, 72 tRF-5c, and 15 tiRNA-5 were identified (Figure 1B). In addition, the numbers of tsRNAs derived from the variable anticodon tRNAs are shown in the stacked plots (Figure 1C-D).

Differentially expressed tsRNAs in patients with sarcoidosis vs. healthy controls were identified through RNA sequencing. All differentially expressed tsRNAs are shown in the cluster heatmap in Figure 2A. The heatmap revealed systematic variations between the sarcoidosis and control groups in terms of tsRNA expression. Additionally, we identified 13 significantly dysregulated tsRNAs: four were upregulated, while nine were downregulated ($FC > 1.5$, $P < 0.05$; Figure 2B and Table 4).

RT-qPCR validation of RNA sequencing results

Six tsRNAs (tRF-iMet-CAT-001, tRF-Gln-TTG-005, tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, tiRNA-Gly-GCC-002, and tRF-Ser-TGA-007) were randomly selected to validate the RNA sequencing results using RT-qPCR in sarcoidosis and control samples. The expression profiles of tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007 agreed with results obtained from RNA sequencing. There were significant differences in tsRNA expression levels between patients with sarcoidosis and healthy controls ($P < 0.05$; Figure 3). These findings indicated that the RNA sequencing data were reliable enough to warrant further analysis.

Correlation between differentially expressed tsRNAs and the clinical parameters of sarcoidosis

The relative expression levels of tiRNA-Glu-TTC-001 were negatively correlated with age (Figure 4A). The relative expression levels of tiRNA-Lys-CTT-003 were significantly correlated with age and the number of affected systems (Figure 4B-C). There was a significant positive correlation between the relative expression of tRF-Ser-TGA-007 and blood calcium levels (Figure 4D).

Prediction of target genes of differentially expressed tsRNAs

Although the mechanism of tsRNA action is yet to be elucidated, recent evidence suggests a miRNA-like mode of action. miRNAs can recognise their mRNA targets using their seed sequence (positions 2-7nt at their 5' ends) and suppress global mRNA translational activities. Based on this rule, we used two types of

algorithms, TargetScan and miRDB, to predict tsRNA targets. In total, miRDB targeted a total of 646 mRNAs, TargetScan targeted a total of 228 mRNAs, and the two target a total of 96 mRNAs (Figure 5B); tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007 could target 637, 111, and 39 transcripts, respectively (Figure 5A).

Biological function analysis revealing potential therapeutic effects of differentially expressed tsRNAs

The tsRNAs regulated mRNA translational activities, therefore we conducted bioinformatics analyses of the functions of target mRNAs to understand their biological functions. We used GO biological processes and KEGG pathways to explore the functions of target mRNAs. The target mRNA of tiRNA-Glu-TTC-001 was enriched in BP and involved in the regulation of transcription, DNA-templated, transcription, negative regulation of transcription from RNA polymerase II promoter, Ras protein signal transduction, heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules, transport, ERK1 and ERK2 cascade, positive regulation of transcription from RNA polymerase II promoter, small GTPase-mediated signal transduction, brain development, ER to Golgi vesicle-mediated transport, and cell proliferation (Figure 6A). The CC comprised intracellular, nucleus, cis-Golgi network, Golgi membrane, receptor complex, chromatin, early endosome, and actin cytoskeleton (Figure 6B), while MF included protein binding, metal ion binding, nucleic acid binding, transcription factor activity, sequence-specific DNA binding, DNA binding, RNA polymerase II core promoter proximal region sequence-specific DNA binding, chromatin binding, protein serine/threonine kinase activity, and transcription corepressor activity (Figure 6C).

The target mRNA of tiRNA-Lys-CTT-003 was enriched in BP, involved in transcription regulation, DNA-templated, regulation of cell cycle, DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest, transcription from RNA polymerase II promoter, megakaryocyte development, platelet alpha granule organisation (Figure 6D), in CC, comprising the nucleus (Figure 6E), and MF, in RNA polymerase II core promoter proximal region sequence-specific DNA binding, metal ion binding, transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding (Figure 6F).

The target mRNA of tRF-Ser-TGA-007 was enriched in BP involved in negative regulation of transcription from RNA polymerase II promoter, positive regulation of transcription from RNA polymerase II promoter, negative regulation of vascular endothelial growth factor receptor signalling pathway, positive regulation of IRE1-mediated unfolded protein response, activation of signalling protein activity involved in unfolded protein response, negative regulation of protein serine/threonine kinase activity, cellular response to vascular endothelial growth factor stimulus (Figure 6G). The CC comprised cerebellar mossy fibre, cell-cell adherens junction (Figure 6H), and MF in transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding, transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding, and protein phosphatase 2A binding (Figure 6I).

KEGG analysis revealed that the target mRNA of tiRNA-Glu-TTC-001 was particularly enriched in pathways in cancer, chemokine signalling, cAMP signalling, Rap1 signalling, endocytosis, Ras signalling, cGMP-PKG signalling, retrograde endocannabinoid signalling, FoxO signalling, adrenergic signalling in

cardiomyocytes, and inflammatory mediator regulation of TRP channels (Table 5), while no KEGG pathway was associated with the target mRNAs of tiRNA-Lys-CTT-003 and tRF-Ser-TGA-007 ($P < 0.05$).

Hub gene analysis

The top ten genes of each sub-plugin in the PPI network of the target mRNAs of tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007 were identified using the cytoHubba plugin, respectively (Table 6). The ten genes with the highest frequency in all sub-plugins were classified as hub genes. KEGG analysis of the hub genes showed their involvement in the insulin signalling pathway (hsa04910), chemokine signalling pathway (hsa04062), regulation of lipolysis in adipocytes (hsa04923), bacterial invasion of epithelial cells (hsa05100), neurotrophin signalling pathway (hsa04722), AMPK signalling pathway (hsa04152), platelet activation (hsa04611), FoxO signalling pathway (hsa04068), cAMP signalling pathway (hsa04024), proteoglycans in cancer (hsa05205), and pentose phosphate pathway (hsa00030).

We used AUC analysis to evaluate their potential for diagnosing sarcoidosis using the expression data from the GSE19314 database. The hub genes associated with tiRNA-Glu-TTC-001 were all expressed in the serum. The AUCs of *CLTC*, *APP*, *PRKACB*, *RAP1B*, *ARRB2*, *IRS1*, *ACTR2*, and *PIK3R3* were all greater than 0.5, of which *APP*, *PRKACB*, and *ARRB2* were 0.707, 0.728, and 0.714, respectively (Figure 7A). Except for *KIAA0101*, the hub genes associated with tiRNA-Lys-CTT-003 were all expressed in the serum of patients with sarcoidosis. The AUCs of nine genes were all greater than 0.5 (Figure 7B). The hub gene associated with tRF-Ser-TGA-007 was expressed in the serum. The AUCs of *NR5A1*, *SPI1*, *TKTL2*, *FMN1*, *MNDA*, and *NCOR2* were all greater than 0.5, of which *NR5A1* and *MNDA* were 0.712 and 0.726, respectively (Figure 7C). The AUCs >0.5 were considered statistically significant. These genes have higher diagnostic potential since they were all greater than 0.70.

RNAhybrid software was used to construct a base complementary pairing model between tiRNA-Glu-TTC-001 and *APP*, *PRKACB*, and *ARRB2* (Figure 8A-C), and between tRF-Ser-TGA-007 and *NR5A1* and *MNDA*, respectively. There was no base complementary pairing model between tRF-Ser-TGA-007 and *MNDA* (Figure 8D).

Discussion

This is the first study to investigate the expression profiles of tsRNA as potential biomarkers and molecules regulating mRNA transcription activities in sarcoidosis. In the serum of sarcoidosis patients, three tsRNAs were significantly reversed compared with the control group. Bioinformatics analysis revealed that the three differentially expressed tsRNAs could play a role in regulating the chemokine, cAMP, cGMP-PKG, and FoxO signalling pathways. These findings provide new insights to better understand the pathogenesis and diagnosis of sarcoidosis.

Sarcoidosis can affect almost every system in the body, but it is more common in the respiratory system, lungs, and lymph nodes in the chest. The unknown aetiology and difficulty of diagnosis prolongs the

period of time from symptom onset to diagnosis confirmation, thus statistics regarding incidence and prevalence of sarcoidosis are somewhat challenging. Although many efforts have been made to identify biomarkers of sarcoidosis, there are currently no definite biomarkers available at the clinical level.

The roles played by miRNAs have been investigated in sarcoidosis [27], while the involvement of tsRNA is novel. tsRNA has been found to be expressed in almost every branch of life, including bacteria and archaea, whose miRNA or siRNA pathway is missing. Due to the lack of conventional ribonucleases, tsRNA can still be found in these less-complex organisms, which indicates that an older regulatory system controls the expression of tsRNA. Numerous studies have suggested that tsRNAs have key regulatory roles and physiological and pathological functions [28]. Through binding to mRNA, a mechanism similar to miRNAs, tsRNAs regulate the stability of RNA and participate in important physiological processes such as cell proliferation and DNA damage [29].

In this study, we performed high-throughput sequencing to detect the expression profile of tsRNA in the serum of sarcoidosis and control groups. Furthermore, considering the reliability of sequencing data, we validated the six tsRNAs using qPCR. Three tsRNAs were not consistent with the sequencing results, and the remaining three tsRNAs (tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007) were confirmed for further analysis. Subsequently, we used bioinformatics methods to show the biological functions of the three differentially expressed tsRNAs and further reveal their targeted genes which can serve as potential biomarkers in sarcoidosis.

Further analysis of the correlation between differentially expressed tsRNAs and clinical parameters of sarcoidosis showed that the relative expression levels of tiRNA-Glu-TTC-001 and tiRNA-Lys-CTT-003 were negatively correlated with age. The age of onset of sarcoidosis is concentrated between 30-40 years old, but it may also occur in children and the elderly. As several patients have no symptoms, it is difficult to determine the true incidence and prevalence of sarcoidosis worldwide [30]. Therefore, the relationship between the expression levels of tiRNA-Glu-TTC-001 and tiRNA-Lys-CTT-003 and the age of sarcoidosis requires further experimental research. We observed a significant positive correlation between the relative expression of tRF-Ser-TGA-007 and blood calcium levels. It is known that the production of calcitriol, the activity of parathyroid hormone (PTH), and sensitivity to vitamin D are altered in patients with sarcoidosis [31]. Under the stimulation of interferon- γ (IFN- γ), tumour necrosis factor- α (TNF- α), interleukin-1 (IL-1), and interleukin-2 (IL-2), macrophages can spontaneously release 1, 25-dihydroxyvitamin D, further promoting the absorption of calcium in the gastrointestinal tract and bone, leading to hypercalcemia and hypercalciuria [32].

Furthermore, for an advanced understanding of the intrinsic mechanisms of tsRNA in sarcoidosis, we performed GO and KEGG pathway analyses. The results showed that the target genes of tiRNA-Glu-TTC-001 were abundant in chemokine, cAMP, cGMP-PKG, and FoxO signalling pathways. Bacterial toxins and chemicals that continue to increase cAMP exert immunosuppressive effects [33, 34]. At the same time, high concentrations of cAMP are often used to suppress immune responses. For exogenous pathogens, such as *Bacillus pertussis* and *Bacillus anthracis*, cAMP has been proven to inhibit the activation and

proliferation of T cells by increasing the concentration of intracellular cAMP, leading to the strong and continuous activation of PKA. By increasing cellular cAMP, these toxins also cause the destruction of chemokine receptor signals, thereby reducing T cell chemotaxis [33, 35, 36]. However, recent studies showed that cAMP also plays a positive role in the immune response, and cyclic adenosine monophosphate can promote the expression of CD86 [37, 38], thereby promoting T cell activation. Cyclic adenosine monophosphate can also enhance the isotype conversion of B cells to IgG and IgE. In dendritic cells, the increase in cAMP triggered by PGE 2 or pharmacology has been shown to promote the expression of IL-23, which facilitates the production of Th17 cells and IL-17, and contributes to cytotoxicity. cAMP further improves T lymphocyte and Th1 cell memory [39, 40]. Sarcoidosis primarily occurs in an abnormal immune system that mainly involves Th17/Treg imbalance related to Th cells. Therefore, cAMP may play a role in the occurrence and development of sarcoidosis. The FoxO signalling pathway shows that sarcoidosis may be related to autophagy and other pathways. Autophagy is an auto-degradation process that is an important adaptive mechanism in response to altered cell signals or cell stress (such as nutritional stress, infection, hypoxia, and toxicity). Current evidence shows that FoxO transcription factors play multiple roles in the regulation and dysregulation of autophagy. Ryan *et al.* used CD4Cre mouse experiments to show that Foxo1 transcriptional activity can change the responsiveness and competitive adaptability of T cells [41]. Further experimental studies have shown that the difference between Foxo1 activity in regulating CD4+ T cells and Treg cells may cause these changes.

We performed ROC curve analysis of the hub genes associated with tsRNA using other sarcoidosis-related serum in the GEO database, and found that expression levels of certain hub genes were significantly elevated in sarcoidosis and displayed a certain diagnostic value for sarcoidosis (AUC > 0.7). *APP* is a gene that encodes an amyloid precursor protein [42]. The protein is cleaved by secretase to form a series of polypeptide fragments, two of which are antimicrobial peptides that have been shown to have bactericidal and antifungal properties. Some other active peptide fragments can be combined with the acetyltransferase complex APBB1/TIP60 to promote transcriptional activation, while some of the remaining peptide fragments form the protein basis of amyloid plaques found in the brains of patients with AD. The protein encoded by *PRKACB* is the catalytic subunit of cAMP-dependent protein kinase (PKA) (Homo sapiens protein kinase, cAMP-dependent, catalytic beta gene, *PRKACB*), which affects cell proliferation and differentiation through the cAMP signalling pathway [43]. *ARRB2*, a member of the inhibitory protein family, participates in signalling pathways related to G-protein-coupled receptors (GPCRs) and desensitises the GPCRs, thus weakening cell responses to stimuli, such as hormones, neurotransmitters, or sensory signals. In recent years, studies have found that *ARRB2* can promote the migration of Th2 cells to the lungs, thereby inducing the onset of allergic asthma. *NR5A1* encodes steroidogenic factor 1 (SF1) [44]. SF1 is a nuclear receptor composed of 461 amino acids. SF1 has four functional domains: a DNA binding domain (DBD), an auxiliary hinge region, a ligand-binding domain (LBD), and two activation domains, AF1 and AF2. SF1 is expressed in the gonads and adrenal glands during embryonic development, and controls important processes therein. Defects in this gene are one of the reasons for XY reversal. Myeloid nuclear differentiation antigen (MND1), a human bone marrow antigen, can only be detected in the nucleus of cells of the granulocyte-monocyte lineage, and is a

member of the nuclear protein family IFI200 (interferon-inducible p200 family) [45]. This nucleoprotein family is characterised by interferon (IFN) induction with 200 repeating amino acids, because the N-terminal pyridine domain (PYD) and the C-terminal HIN domain are also annotated as PYHIN factors. Overexpression of the PYHIN factor can induce cell cycle arrest, apoptosis, senescence, inflammation, and decrease tumour cell invasiveness. Until recently, the relationship between the above-mentioned genes and sarcoidosis has not been reported, but *ARRB2* and *PKACB* are obviously related to the cAMP signalling pathway and have functions similar to those of the target genes predicted by KEGG.

After constructing a base complementary pairing model using RNAhybrid software, we preliminarily determined that in patients with sarcoidosis, *APP*, *PRKACB*, *ARRB2*, and *NR5A1* may participate in the occurrence and development of sarcoidosis through immune inflammation and other related pathways, and can be used as serum biomarkers of sarcoidosis.

This study had several limitations. First, our sample size was relatively small, and the current research may not be sufficient to detect significant differences between patients and healthy controls. Second, in the analysis of clinical characteristics of patients with sarcoidosis, the clinical patient information, such as lung function, serum ACE level, 24 h urine calcium content, and treatment information, is not comprehensive. Therefore, tsRNAs have not been fully analysed with clinical information. Finally, our follow-up correlation analyses regarding tsRNA, target mRNA, and hub genes are mainly based on bioinformatics, and no experimental verification has been performed. Therefore, we cannot determine whether there is a causal relationship between the differential expression of hub genes and disease progression. Therefore, it is necessary to conduct subsequent bioinformatics analysis and experimental verification with a larger sample size in the future to understand the molecular mechanisms of these new central genes in the pathogenesis of sarcoidosis.

We used high-throughput sequencing technology to display the tsRNA expression profile in the serum of patients with sarcoidosis and healthy controls, and predicted the target genes that may be involved in the occurrence and development of sarcoidosis. Further, we identified hub genes as a biomarker using bioinformatics and related gene databases. Our study provides the basis for future investigations on the diagnosis, treatment, and prognosis of sarcoidosis.

Conclusions

Our results demonstrated that serum tsRNA expression levels, especially that of tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007, are capable of differentiating sarcoidosis patients from healthy controls, revealing the pathogenesis, and indicating multiple clinical parameters. Bioinformatics analysis and omics studies are crucial for exploring novel and effective pathogenic targets of sarcoidosis.

Methods

Sample collection

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Research Ethics Board of the Second Affiliated Hospital of Jilin University (Changchun, China). Written informed consent was obtained from all individuals included in the study. Three adult patients who were diagnosed with sarcoidosis at the Department of Respiratory and Critical Care of the Second Affiliated Hospital of Jilin University were enrolled in the RNA sequencing experiment. The diagnosis of sarcoidosis was verified by histopathologic examination of mediastinal lymph nodes or lung biopsy samples. Overall, three female patients with sarcoidosis were included in this study, and were compared with three healthy female volunteers. To verify the RNA sequencing results, eight patients with sarcoidosis were included in the study and were compared with four healthy female volunteers. The patient characteristics are listed in Table 1. All study subjects were members of the Chinese Han nationality in northern China. The baseline characteristics of the validation group are presented in Table 2.

Samples were collected between February 2019 and September 2019. Whole blood samples were acquired and collected into acid-citrate-dextrose-containing yellow-top tubes before surgery. After incubation for 15 min at 25 °C and centrifugation for 10 min at 1,000 × *g*, the serum was transferred to new RNase-free tubes. The entire sample processing procedure was performed on ice. Finally, the serum samples were stored at –80 °C for further experiments [50].

RNA extraction

We used a fixed volume of serum (500 µL per sample) for tsRNA sequencing and quantitative reverse-transcriptase-polymerase chain reaction (qRT-PCR) detection. Total RNA was extracted using the miRNeasy Serum/Plasma Kit (Qiagen, Germany) according to the manufacturer's instructions. We then confirmed that all OD 260/280 absorbance ratios were between 1.8 and 2.0, using the NanoDrop ND-1000 (Thermo Fisher Scientific, Waltham, MA, USA). Finally, the extracted RNA was preserved at –80 °C for subsequent experiments.

tRF and tiRNA high-throughput sequencing

Before the sequencing experiment, we determined the quality of each RNA sample using agarose gel electrophoresis and the Nanodrop™ instrument. tsRNAs are heavily decorated by RNA modifications, which interfere with small RNAseq library construction. Therefore, we performed the following treatments with the rtStar™ tRF & tiRNA Pre-treatment Kit before library preparation for the total RNA samples [54]: 3-aminoacyl (charged) deacylation to 3-OH for 3' adapter ligation; 3-cP (2,3-cyclic phosphate) removal to 3-OH for 3' adapter ligation; 5-OH (hydroxyl group) phosphorylation to 5-P for 5' adapter ligation; and m1A and m3C demethylation for efficient reverse transcription. Pre-treated total RNA was used to prepare the sequencing library using the rtStar™ First-Strand complementary DNA (cDNA) Synthesis Kit [53]: (a) 3' adapter ligation, (b) 5' adapter ligation, (c) cDNA synthesis, (d) PCR amplification, and (e) size selection of 134–160 bp PCR-amplified fragments [55]. The prepared samples were denatured into single-stranded DNA molecules, which were then captured on Illumina flow cells, amplified *in situ*, and profiled for 50 cycles on an Illumina NextSeq 500 system (Illumina). Thereafter, we used the Solexa pipeline v1.8 (Off-

Line Base Caller software, v1.8) for graphical analysis and base calling. Sequencing quality was checked using FastQC, and trimmed reads (that passed the Illumina quality filter, trimmed 3' adapter bases by Cutadapt) were aligned to tRNA precursor sequences from GtRNAdb using NovoAlign software (v2.07.11). The remaining reads were aligned to the transcriptome sequences (mRNA/rRNA/snRNA/snoRNA/piRNA/miRNA). The expression profiling and differential expression of tRFs, tiRNAs, and known miRNAs were calculated based on normalised counts per million of total aligned reads (CPM). Hierarchical clustering and scatter-plot construction were performed using differentially expressed tsRNAs in R environments for statistical calculations and graphing [47-49].

Validation using reverse transcription-quantitative PCR (RT-qPCR)

RT-qPCR is considered the gold standard for the quantification of gene expression [19,20]. To verify the RNA sequencing results, six tsRNAs were randomly selected, and their expression levels were detected in eight patients with sarcoidosis and four healthy controls. Total RNA was extracted from peripheral blood using TRIzol® reagent. RNA samples were pre-treated using the rtStar™ tRF & tiRNA Pre-treatment Kit (cat. no. AS-FS-005; Arraystar Inc.) to remove RNA modifications, and then reverse transcribed to cDNA using the rtStar™ First-Strand cDNA Synthesis Kit (cat. no. AS-FS-003; Arraystar Inc.) according to the manufacturer's protocol. RT-qPCR was performed using the ViiA 7 Real-time PCR system, using the Arraystar SYBR® Green Real-time qPCR Master mix (cat. no. AS-MR-005-25; Applied Biosystems; Thermo Fisher Scientific, Inc.) with specific primers (Table 3). The thermocycling conditions were as follows: initial denaturation at 95 °C for 10 min, followed by 40 cycles at 95 °C for 10 s, and 60 °C for 60 s. The relative tsRNA expression levels were calculated using the relative standard curve method [21] and normalised to *U6*, as endogenous controls. All reactions were performed in triplicate [50].

Target prediction of treatment-related tsRNAs

tsRNAs contained some seed sequences, which might match the crosslink-centred regions of target mRNAs [22,23]. In addition, several studies strongly suggested that tsRNAs have a miRNA-like function, leading to mRNA silencing [24-26]. In this study, we used two common algorithms to predict tsRNA targets: TargetScan (<http://www.targetscan.org>) and miRDB (<http://www.mirdb.org>). Additionally, to reduce false-positive results, only genes predicted by all four software programs were considered targets of tsRNAs. The network illustration was visualised using Cytoscape software (version 3.5.1, Cytoscape Consortium, San Diego, CA, USA).

Bioinformatics analysis

Pathway and process enrichment analyses were performed to assign the biological annotation of the targets. Analysis was performed using several ontology sources. Gene Ontology (GO) is commonly used to determine the function(s) of genes during the analysis of high-throughput transcriptomic or genomic data. Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a collection of databases associated with genomes, diseases, biological pathways, drugs, and chemical materials [26]. Database for Annotation,

Visualisation, and Integrated Discovery (DAVID) is an online bioinformatics tool that is used to identify the functions of genes or proteins. We used DAVID to determine the functions of differentially expressed genes (DEGs) in the biological process (BP), cellular component (CC), molecular function (MF), and pathways ($P < 0.05$). PPI networks help clarify the molecular mechanisms underlying biological progress. PPI networks were generated using the Search Tool for the Retrieval of Interacting Genes (STRING; <https://string-db.org/>) using a threshold interaction score of 0.4. Subsequently, we used the MCODE app in Cytoscape to analyse PPI network modules (degree cut-off = 2, max; depth = 100; k-core = 2; and node score cut-off = 0.2) [51,52].

Identification of hub gene(s)

The hub genes in the candidate modules were identified using the Cytoscape plugin cytoHubba (top 10 nodes ranked by MCC, MNC, Radiality, Stress, Betweenness, BottleNeck, Closeness, ClusteringCoefficient, Degree, DMNC, EcCentricity, and EPC) [54]. The NCBI-GEO is a free public database of a microarray/genomic data repository. Gene expression profiles in the peripheral blood of sarcoidosis patients were obtained from the GSE19314 datasets. Microarray data of GSE19314 were obtained using GPL570 Platforms ([HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array), which included 37 and 20 blood samples from patients with sarcoidosis and healthy individuals, respectively. We used the expression data from the GSE19314 database to generate the receiver operating characteristic (ROC) curves of the hub genes using R/Bioconductor. The area under the ROC curve (AUC) for these genes was calculated to test their potential for diagnosing sarcoidosis. $AUC > 0.5$ were considered statistically significant [46]. The RNAhybrid software (<https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid>) was used to establish a basic complementary pairing model between the differentially expressed tsRNA and its associated central gene ($AUC > 0.7$) to clarify the site of action [47,50].

Statistical analysis

The results are presented as means \pm SEM. Statistical analysis was conducted using SPSS 24.0 software and GraphPad Prism 7. Differences between two groups were analysed using the Student's *t*-test. One-way analysis of variance (ANOVA) was employed for all three groups. Statistical significance was set at $P < 0.05$.

Declarations

Acknowledgements

We acknowledge GEO database for providing their platforms and contributors for uploading their meaningful datasets. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5). Informed consent was obtained from all patients for being included in the study.

Authors' contributions

MZ and KW conceived and designed the experiments; MZ, CT, XD, and SC performed the experiments; SC, YC, and XJ analyzed the data; MZ, CT, XD, XJ, SC, and YC contributed reagents/materials/analysis tools; MZ, XD and KW contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the Technology Research Funds of Jilin Province (20190303162SF), the Medical And Health Project Funds of Jilin Province (20200708083YY), the Medical And Health Project Funds of Jilin Province (2020SCZT019), Disciplinary Crossing and Integration and innovation cultivation project of Jilin University (JLUXKJC2020212), and the Medical And Health Project Funds of Jilin Province (20191102012YY).

Competing interests

The authors declare no competing interests.

References

1. Landi C, Bargagli E, Carleo A, et al. A functional proteomics approach to the comprehension of sarcoidosis. *J Proteomics*. 2015;128:375-387.
2. Ahmadzai H, Huang S, Steinfurt C, et al. Sarcoidosis: a state of the art review from the Thoracic Society of Australia and New Zealand. *Med J Aust*. 2018;208(11):499-504.
3. Landi C, Carleo A, Cillis G, Rottoli P. Sarcoidosis: proteomics and new perspectives for improving personalized medicine. *Expert Rev Proteomics*. 2018;15(10):829-835.
4. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. *Clin Chem*. 2015;61(1):112-123.
5. Borrebaeck CA. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nat Rev Cancer*. 2017;17(3):199-204.
6. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer*. 2005;5(11):845-856.
7. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer*. 2011;11(6):426-437.
8. Bennett D, Cameli P, Lanzarone N, et al. Chitotriosidase: a biomarker of activity and severity in patients with sarcoidosis. *Respir Res*. 2020;21(1):6.
9. Terrington DL, Kim JW, Ravenhill G, et al. Soluble interleukin-2 receptor in exhaled breath condensate in pulmonary sarcoidosis: a cross-sectional pilot study. *J Breath Res*. 2020;15(1):016016.

10. Beermann J, Piccoli MT, Viereck J, Thum T. Non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. *Physiol Rev.* 2016;96(4):1297-1325.
11. Anderson P, Ivanov P. tRNA fragments in human health and disease. *FEBS Lett.* 2014;588(23):4297-4304.
12. Dhahbi JM, Spindler SR, Atamna H, et al. 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics.* 2013;14:298.
13. Kumar P, Kuscu C, Dutta A. Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends Biochem Sci.* 2016;41(8):679-689.
14. Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* 2009;23(22):2639-2649.
15. Maute RL, Schneider C, Sumazin P, et al. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl Acad Sci U S A.* 2013;110(4):1404-1409.
16. Fu H, Feng J, Liu Q, et al. Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett.* 2009;583(2):437-442.
17. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. *Biomark Cancer.* 2014;6:37-47.
18. Keam SP, Hutvagner G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life (Basel).* 2015;5(4):1638-1651.
19. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001;25(4):402-408.
20. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res.* 1996;6(10):986-994.
21. Larionov A, Krause A, Miller W. A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics.* 2005;6:62.
22. Kim HK, Fuchs G, Wang S, et al. A transfer-RNA-derived small RNA regulates ribosome biogenesis. *Nature.* 2017;552(7683):57-62.
23. Karaiskos S, Naqvi AS, Swanson KE, Grigoriev A. Age-driven modulation of tRNA-derived fragments in *Drosophila* and their potential targets. *Biol Direct.* 2015;10:51.
24. Kumar P, Anaya J, Mudunuri SB, Dutta A. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* 2014;12:78.
25. Kuscu C, Kumar P, Kiran M, Su Z, Malik A, Dutta A. tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA.* 2018;24(8):1093-1105.

26. Li P, Tang T, Liu T, et al. Systematic Analysis of tRNA-Derived Small RNAs Reveals Novel Potential Therapeutic Targets of Traditional Chinese Medicine (Buyang-Huanwu-Decoction) on Intracerebral Hemorrhage. *Int J Biol Sci.* 2019;15(4):895-908.
27. Salamo O, Mortaz E, Mirsaedi M. Noncoding RNAs: New Players in Pulmonary Medicine and Sarcoidosis. *Am J Respir Cell Mol Biol.* 2018;58(2):147-156.
28. Karaiskos S, Grigoriev A. Dynamics of tRNA fragments and their targets in aging mammalian brain. *F1000Res.* 2016;5.
29. Gebetsberger J, Polacek N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol.* 2013;10(12):1798-1806.
30. Bargagli E, Prasse A. Sarcoidosis: a review for the internist. *Intern Emerg Med.* 2018;13(3):325-331.
31. Cameli P, Caffarelli C, Refini RM, et al. Hypercalciuria in Sarcoidosis: A Specific Biomarker With Clinical Utility. *Front Med (Lausanne).* 2020;7:568020.
32. Baughman RP, Teirstein AS, Judson MA, et al. Clinical characteristics of patients in a case control study of sarcoidosis. *Am J Respir Crit Care Med.* 2001;164(10 Pt 1):1885-1889.
33. Paccani SR, Dal Molin F, Benagiano M, et al. Suppression of T-lymphocyte activation and chemotaxis by the adenylate cyclase toxin of *Bordetella pertussis*. *Infect Immun.* 2008;76(7):2822-2832.
34. Vang T, Torgersen KM, Sundvold V, et al. Activation of the COOH-terminal Src kinase (Csk) by cAMP-dependent protein kinase inhibits signaling through the T cell receptor. *J Exp Med.* 2001;193(4):497-507.
35. Paccani SR, Tonello F, Ghittoni R, et al. Anthrax toxins suppress T lymphocyte activation by disrupting antigen receptor signaling. *J Exp Med.* 2005;201(3):325-331.
36. Rossi Paccani S, Tonello F, Patrussi L, et al. Anthrax toxins inhibit immune cell chemotaxis by perturbing chemokine receptor signalling. *Cell Microbiol.* 2007;9(4):924-929.
37. Kasprowicz DJ, Kohm AP, Berton MT, Chruscinski AJ, Sharpe A, Sanders VM. Stimulation of the B cell receptor, CD86 (B7-2), and the beta 2-adrenergic receptor intrinsically modulates the level of IgG1 and IgE produced per B cell. *J Immunol.* 2000;165(2):680-690.
38. Kohm AP, Mozaffarian A, Sanders VM. B cell receptor- and beta 2-adrenergic receptor-induced regulation of B7-2 (CD86) expression in B cells. *J Immunol.* 2002;168(12):6314-6322.
39. Schnurr M, Toy T, Shin A, Wagner M, Cebon J, Maraskovsky E. Extracellular nucleotide signaling by P2 receptors inhibits IL-12 and enhances IL-23 expression in human dendritic cells: a novel role for the cAMP pathway. *Blood.* 2005;105(4):1582-1589.
40. Shi Q, Yin Z, Zhao B, et al. PGE2 Elevates IL-23 Production in Human Dendritic Cells via a cAMP Dependent Pathway. *Mediators Inflamm.* 2015;2015:984690.
41. Cheng Z. The FoxO-Autophagy Axis in Health and Disease. *Trends Endocrinol Metab.* 2019;30(9):658-671.
42. Jakubauskiene E, Vilys L, Peciuliene I, Kanopka A. The role of hypoxia on Alzheimer's disease-related APP and Tau mRNA formation. *Gene.* 2021;766:145146.

43. Moen LV, Sener Z, Volchenkov R, et al. Ablation of the Cbeta2 subunit of PKA in immune cells leads to increased susceptibility to systemic inflammation in mice. *Eur J Immunol.* 2017;47(11):1880-1889.
44. Na X, Mao Y, Tang Y, et al. Identification and functional analysis of fourteen NR5A1 variants in patients with the 46 XY disorders of sex development. *Gene.* 2020;760:145004.
45. Bottardi S, Guieze R, Bourgoin V, et al. MNDA controls the expression of MCL-1 and BCL-2 in chronic lymphocytic leukemia cells. *Exp Hematol.* 2020;88:68-82 e65.
46. Zhao M, Di X, Jin X, et al. Identification of Biomarkers for Sarcoidosis and Tuberculosis of the Lung Using Systematic and Integrated Analysis. *Med Sci Monitor.* 2020;26.
47. Huang Y, Ge H, Zheng MJ, et al. Serum tRNA-derived fragments (tRFs) as potential candidates for diagnosis of nontriple negative breast cancer. *Journal of Cellular Physiology.* 2020;235(3):2809-2824.
48. Wang CJ, Zhao M, Wang J, Zhang D, Wang S, Zhao JZ. Expression analysis of transfer RNA-derived fragments in the blood of patients with moyamoya disease: A preliminary study. *Mol Med Rep.* 2019;19(5):3564-3574.
49. Sun C, Li W, Wu H, et al. tRNA-derived fragments as novel predictive biomarkers for trastuzumab-resistant breast cancer. *Cancer Res.* 2019;79(4).
50. Sun CX, Yang F, Zhang YH, et al. tRNA-Derived Fragments as Novel Predictive Biomarkers for Trastuzumab-Resistant Breast Cancer. *Cellular Physiology and Biochemistry.* 2018;49(2):419-431.
51. Crosara KTB, Moffa EB, Xiao YZ, Siqueira WL. Merging in-silico and in vitro salivary protein complex partners using the STRING database: A tutorial. *Journal of Proteomics.* 2018;171:87-94.
52. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research.* 2017;45(D1):D362-D368.
53. Li M, Kao E, Malone D, Gao X, Wang JYJ, David M. DNA damage-induced cell death relies on SLFN11-dependent cleavage of distinct type II tRNAs. *Nat Struct Mol Biol.* 2018;25(11):1047-1058.
54. Cosentino C, Toivonen S, Diaz Villamil E, et al. Pancreatic beta-cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes. *Nucleic Acids Res.* 2018;46(19):10302-10318.
55. Zhang T, Liu C, Huang S, Ma Y, Fang J, Chen Y. A Downmodulated MicroRNA Profiling in Patients with Gastric Cancer. *Gastroenterol Res Pract.* 2017;2017:1526981.

Tables

Table 1 The characteristics of the patients in the development and validation groups (mean \pm SD).

	Development group		Validation group	
	Sarcoidosis (n = 3)	Healthy controls (n = 3)	Sarcoidosis (n = 8)	Healthy controls (n = 4)
Age (years)	48.7±5.3	39.0±5.4	49.3±7.5	41.5±9.9
Male/female (n)	0/3	0/3	1/7	0/4
Radiological stage (I/II/III)	1/1/1	-	0/6/2	-
Smoking status (previous/never)	1/2	0/3	0/8	0/4

∓: No obvious abnormalities on chest CT scan.

Table 2. Individual key clinical characteristics for all 8 participants with sarcoidosis.

	∓	∓	∓	∓	∓	∓	∓	∓
Age (years)	55	41	51	55	37	60	43	52
Gender	Female	Female	Female	Female	Female	Female	Male	Female
Radiological stage	2	2	2	2	2	3	2	3
Extra-pulmonary manifestations	CI	PL	CU,OC	ELN	-	CA, PL	-	ELN
Serum calcium (/mmol L)	2.20	2.24	2.42	2.10	2.27	2.51	2.38	2.41
ESR (/mmh)	23	34	34	23	7	9	7	27
Pao2 (/mmHg)	83	81	74	72	90	59	83	87
CRP (mg/L)	2.65	3.96	2.49	3.73	0.17	14.01	1.7	4.28
Eosinophils count (×10 ⁹ /L)	0.00	0.09	0.00	0.14	0.10	0.02	0.04	0.20
T-spot	+	-	-	-	-	+	-	+

CU, cutaneous sarcoidosis; CA, cardiac involvement; ELN, extrathoracic lymph node involvement; PL, Pleural involvement; OC, ocular sarcoidosis; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein;

Table 3. Primers for reverse transcription-quantitative polymerase chain reaction.

Gene	Primers
U6	F:5'GCTTCGGCAGCACATATACTAAAAT3' R:5'CGCTTCACGAATTTGCGTGTCAT3'
tRF-iMet-CAT-001	F:5' TCCGACGATCAGCAGAGTGG3' R:5' TCTTCCGATCTCCCAGCACG3'
tRF-Gln-TTG-005	F:5' ACAGTCCGACGATCTAGGATG3' R:5' TGCTCTTCCGATCTATCACAC3'
tiRNA-Lys-CTT-003	F:5' ATCGCCCGGCTAGCTCAGT3' R:5' TGTGCTCTTCCGATCTGAGTCC3'
tiRNA-Glu-TTC-001	F:5' ACAGTCCGACGATCTCCCAC 3' R:5' CGTGTGCTCTTCCGATCTAAAC 3'
tiRNA-Gly-GCC-002	F:5' ACAGTCCGACGATCGCATG 3' R:5' GTGCTCTTCCGATCTCAGGC 3'
tRF-Ser-TGA-007	F:5' CTACAGTCCGACGATCGAAGC 3' R:5' TCTTCCGATCTAAATAAGAGCACC 3'

U6 was used as the internal control. F, forward; R, reverse.

Table 4. Differential tsRNAs in sarcoidosis.

tsRNA	Type	Regulation	Length	Fold_Change	p_value	q_value
tRF-Ser-GCT-109	tRF-5b	up	24	95.62724785	0.044775714	0.996329094
tRF-Thr-TGT-054	tRF-3b	up	19	94.37648341	0.04475833	0.996329094
tRF-iMet-CAT-001	tRF-5c	up	30	4.75526063	0.011536113	0.996329094
tRF-Gln-TTG-005	tRF-5a	up	16	3.600230008	0.035728136	0.996329094
tRF-Ser-TGA-037	tRF-3b	down	19	0.006946015	0.001563587	0.345742941
tRF-Pro-TGG-016	tRF-5b	down	24	0.01103809	0.019808701	0.996329094
tRF-Arg-ACG-017	tRF-1	down	21	0.044988695	0.036103512	0.996329094
tiRNA-Val-CAC-002	tiRNA-5	down	34	0.098833932	0.045891918	0.996329094
tRF-Ser-TGA-005	tRF-1	down	17	0.176230811	0.033045783	0.996329094
tiRNA-Glu-TTC-001	tiRNA-5	down	34	0.179020622	0.017044841	0.996329094
tiRNA-Lys-CTT-003	tiRNA-5	down	34	0.191484844	0.001920794	0.345742941
tiRNA-Gly-GCC-002	tiRNA-5	down	33	0.199464689	0.039743529	0.996329094
tRF-Ser-TGA-007	tRF-1	down	19	0.228750737	0.015413429	0.996329094

Table 5. KEGG pathway analysis of the target mRNA of tiRNA-Glu-TTC-001.

Pathway ID	Name	Count	%	P Value	Genes
hsa05200	Pathways in cancer	23	0.026752	0.001234	NTRK1, MAP2K1, SMAD4, FZD5, TPM3, PTGER2, LAMB4, GNAI3, PIK3R3, ADCY1, IGF1, CBL, TGFBR2, HSP90B1, MAPK10, FGF5, FGF14, CXCL12, PLCB4, PAX8, GNB5, PLCB2, PRKACB
hsa04062	Chemokine signaling pathway	14	0.016284	0.001914	MAP2K1, GNAI3, PIK3R3, ADCY1, ARRB2, VAV2, RAP1B, CXCL12, PLCB4, GRK6, CCR6, GNB5, PLCB2, PRKACB
hsa04024	cAMP signaling pathway	14	0.016284	0.003336	RYR2, MAP2K1, ADCYAP1R1, PTGER2, GNAI3, ATP1B4, PIK3R3, ADCY1, VAV2, MAPK10, RAP1B, PRKACB, CNGB1, GRIA4
hsa04015	Rap1 signaling pathway	14	0.016284	0.005515	VASP, MAP2K1, FLT1, GNAI3, PIK3R3, ADCY1, IGF1, MAPK13, FGF5, RAP1B, FGF14, PLCB4, PRKD1, PLCB2
hsa04144	Endocytosis	14	0.016284	0.016358	ARF3, IQSEC2, WIPF2, CLTC, AP2B1, VPS26B, ARRB2, CBL, HLA-F, RAB22A, TGFBR2, EEA1, GRK6, SMAP2
hsa04014	Ras signaling pathway	13	0.01512	0.023119	MAP2K1, FLT1, GAB1, PIK3R3, IGF1, MAPK10, FGF5, RAP1B, FGF14, ABL2, GNB5, RGL2, PRKACB
hsa04022	cGMP-PKG signaling pathway	12	0.013957	0.004367	VASP, MEF2A, ATF2, MAP2K1, PLCB4, IRS1, GNAI3, KCNMB3, ATP1B4, ADCY1, PLCB2, CNGB1
hsa04723	Retrograde endocannabinoid signaling	11	0.012794	4.61E-04	MAPK10, PLCB4, DAGLA, GABRA4, GNAI3, GNB5, ADCY1, PLCB2, PRKACB, MAPK13, GRIA4
hsa04068	FoxO signaling pathway	11	0.012794	0.003971	MAPK10, USP7, MAP2K1, SMAD4, HOMER1, IRS1, PIK3R3, IGF1, SIRT1, MAPK13, TGFBR2
hsa04261	Adrenergic signaling in cardiomyocytes	11	0.012794	0.004894	RYR2, ATF2, PLCB4, PPP1R1A, TPM3, GNAI3, ATP1B4, ADCY1, PLCB2, PRKACB, MAPK13

Table 6. The top 10 Hub genes of differential tsRNAs by cytoHubba

tsRNA	Top ten Hub genes
tiRNA-Glu-TTC-001	CLTC APP SIRT1 PRKACB CBL RAP1B ARRB2 IRS1 ACTR2 PIK3R3
tiRNA-Lys-CTT-003	KIAA0101 CDT1 GTSE1 KMT2D OPA3 FBXL19 SPC24 PCGF3 KDM2A KDM6B
tRF-Ser-TGA-007	NR5A1 SPI1 TKTL2 FNBP1L GPI FMN1 MND A NCOR2 SCARB1 NFIX

Figures

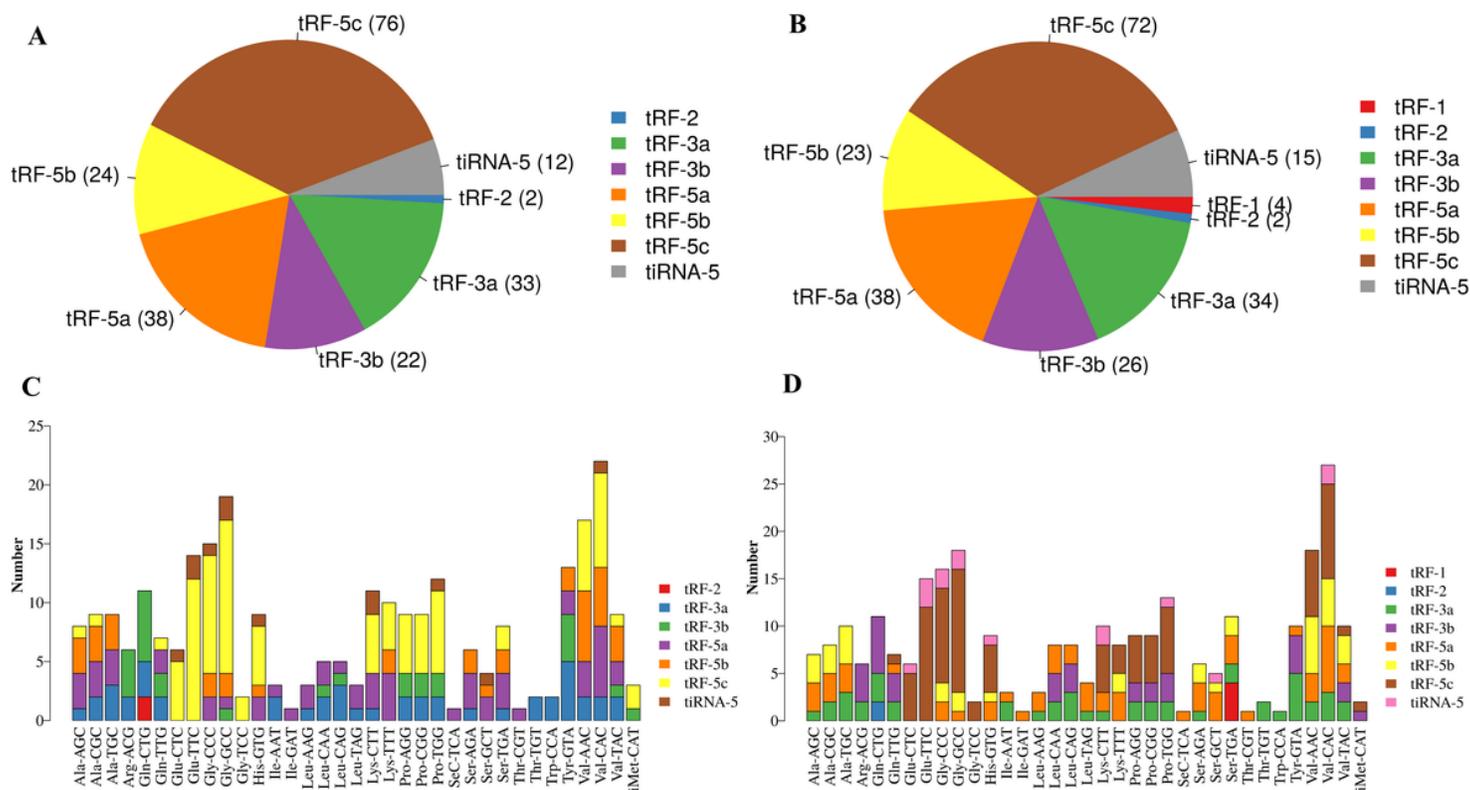


Figure 1

Subtype tsRNAs distribution between sarcoidosis and control samples by small RNA-Seq. A-B Pie chart of the distribution of subtype tsRNAs number in sarcoidosis group and control group. **C-D** The numbers of tsRNAs derived from the variable anticodon tRNAs are demonstrated in sarcoidosis group and control group.

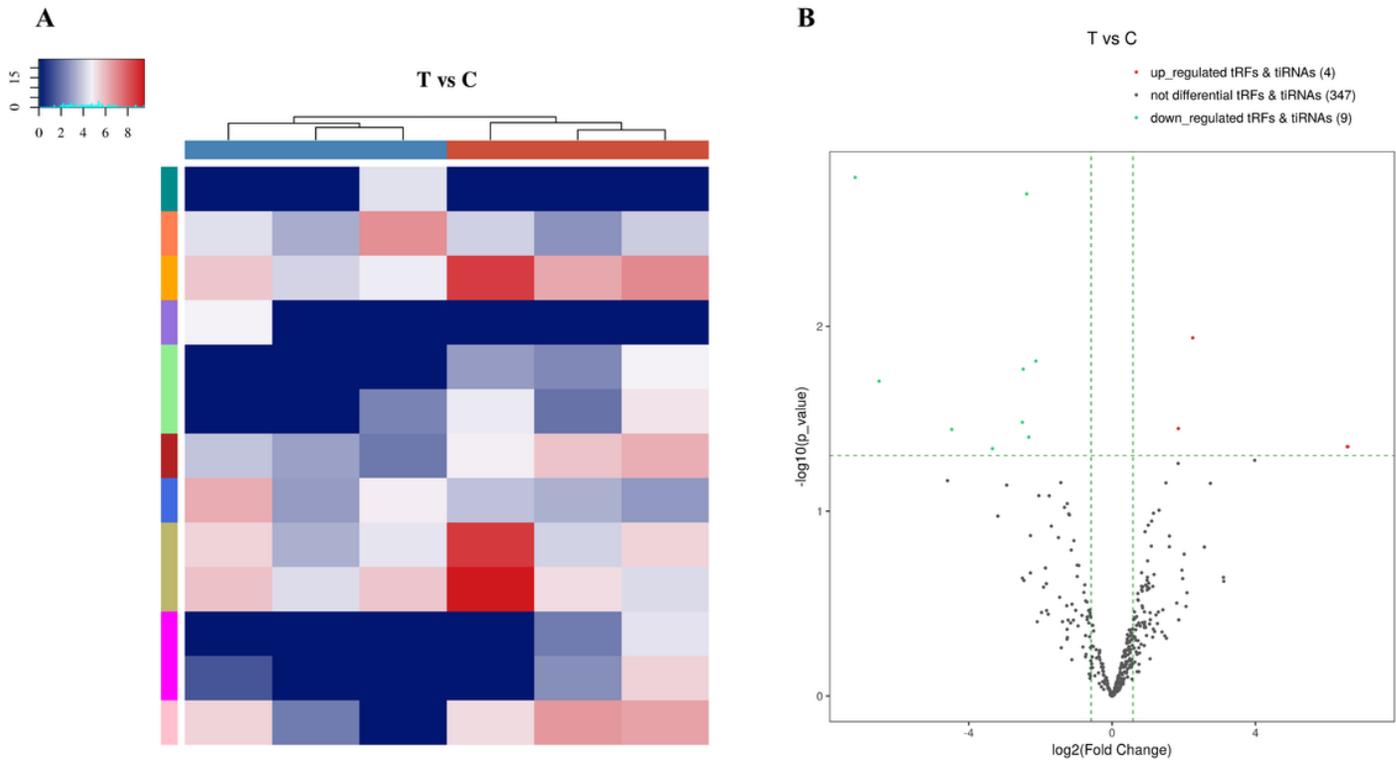


Figure 2

Differentially expressed tsRNAs between sarcoidosis and control samples. **A** Heat map and hierarchical clustering analysis of differentially expressed tsRNAs between sarcoidosis and control samples. **B** The volcano plots of differentially expressed tsRNAs. The green line shows the default 1.5-fold change. The red and green plots indicate the significantly upregulated and downregulated genes, respectively (fold change ≥ 1.5 , $P < 0.05$).

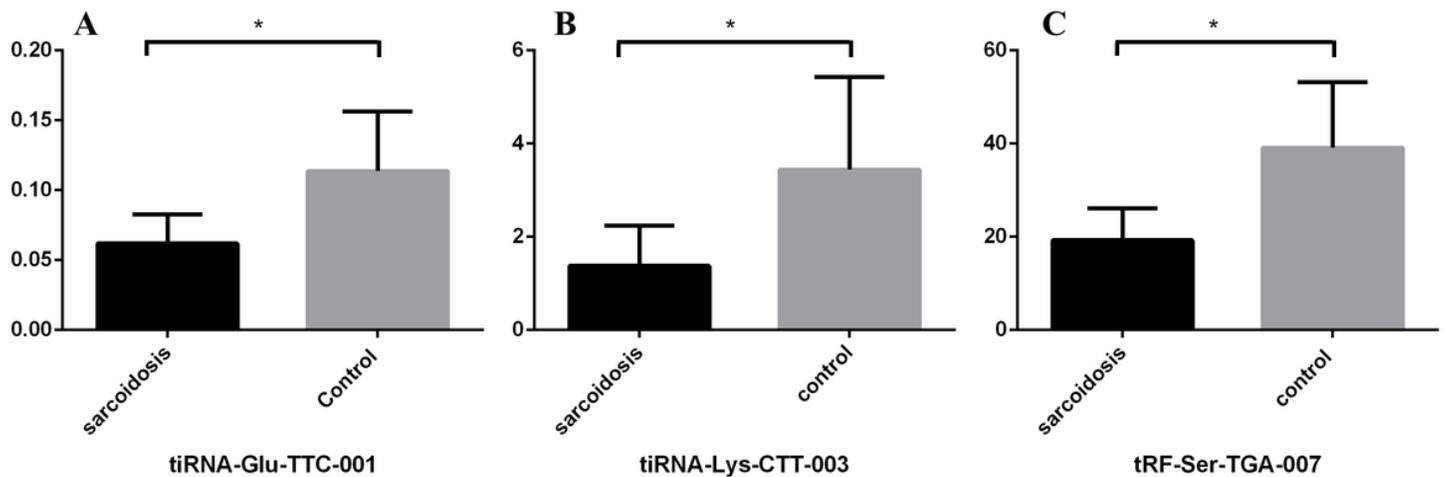


Figure 3

A-C The qPCR results showed that tiRNA-Glu-TTC-001, tiRNA-Lys-CTT-003, and tRF-Ser-TGA-007 were consistent with tsRNA-Seq data. Thus, the 3 tsRNAs were confirmed as sarcoidosis related tsRNAs and used for further analysis. Data were presented as mean \pm SEM, * $P < 0.05$, indicate significant difference compared with control group.

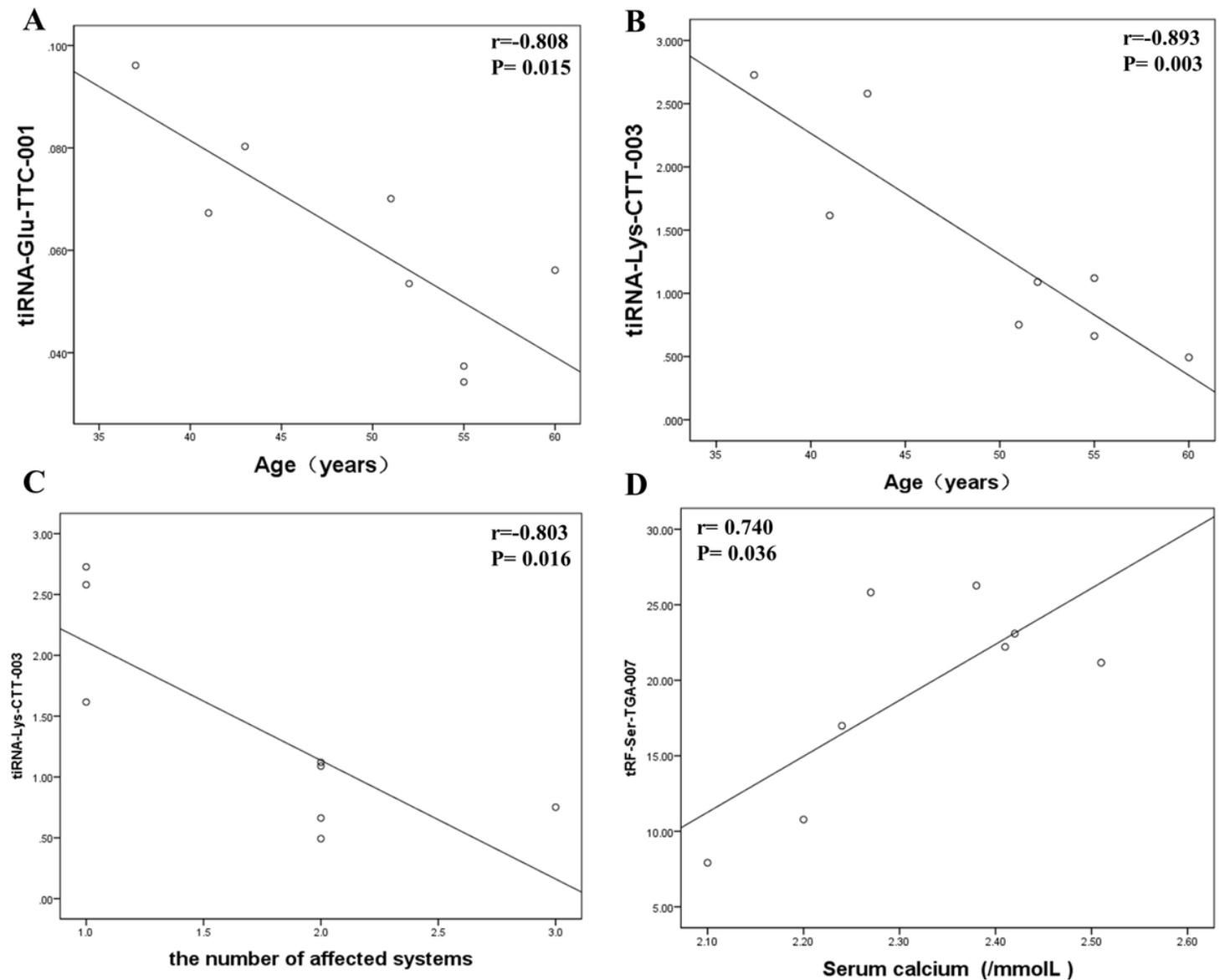


Figure 4

Correlation between differentially expressed tsRNAs and the clinical parameters of sarcoidosis. A

Correlation between tiRNA-Glu-TTC-001 and age. **B** Correlation between tiRNA-Lys-CTT-003 and age. **C** Correlation between tiRNA-Lys-CTT-003 and the number of affected systems. **D** Correlation between tRF-Ser-TGA-007 and blood calcium levels.

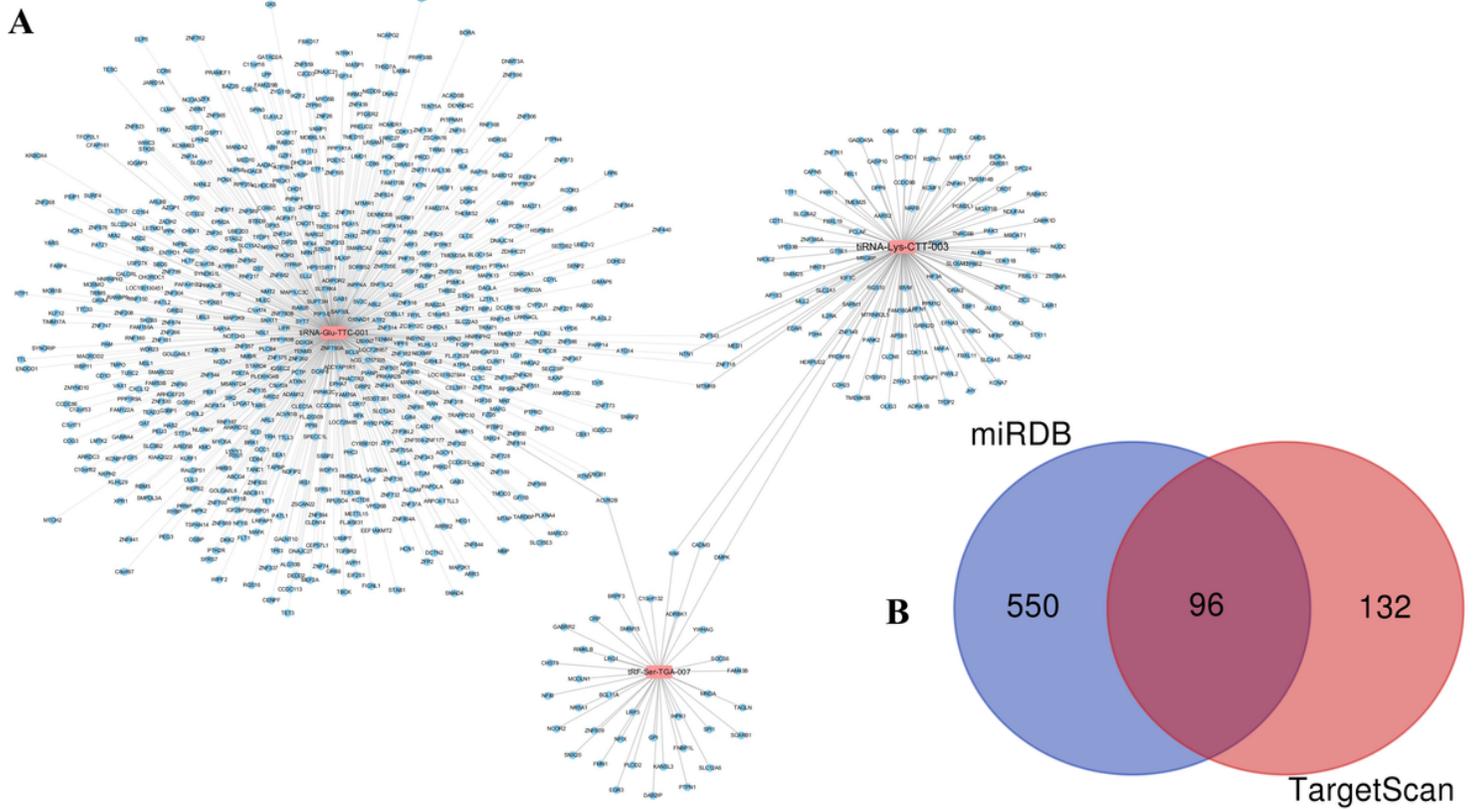


Figure 5

Targets genes of differentially expressed tsRNAs. **A** The targets of each tsRNA were shown respectively. **B** Venn plot to indicate that 96 mRNA targets of three tsRNAs were predicted by two prediction software simultaneously.

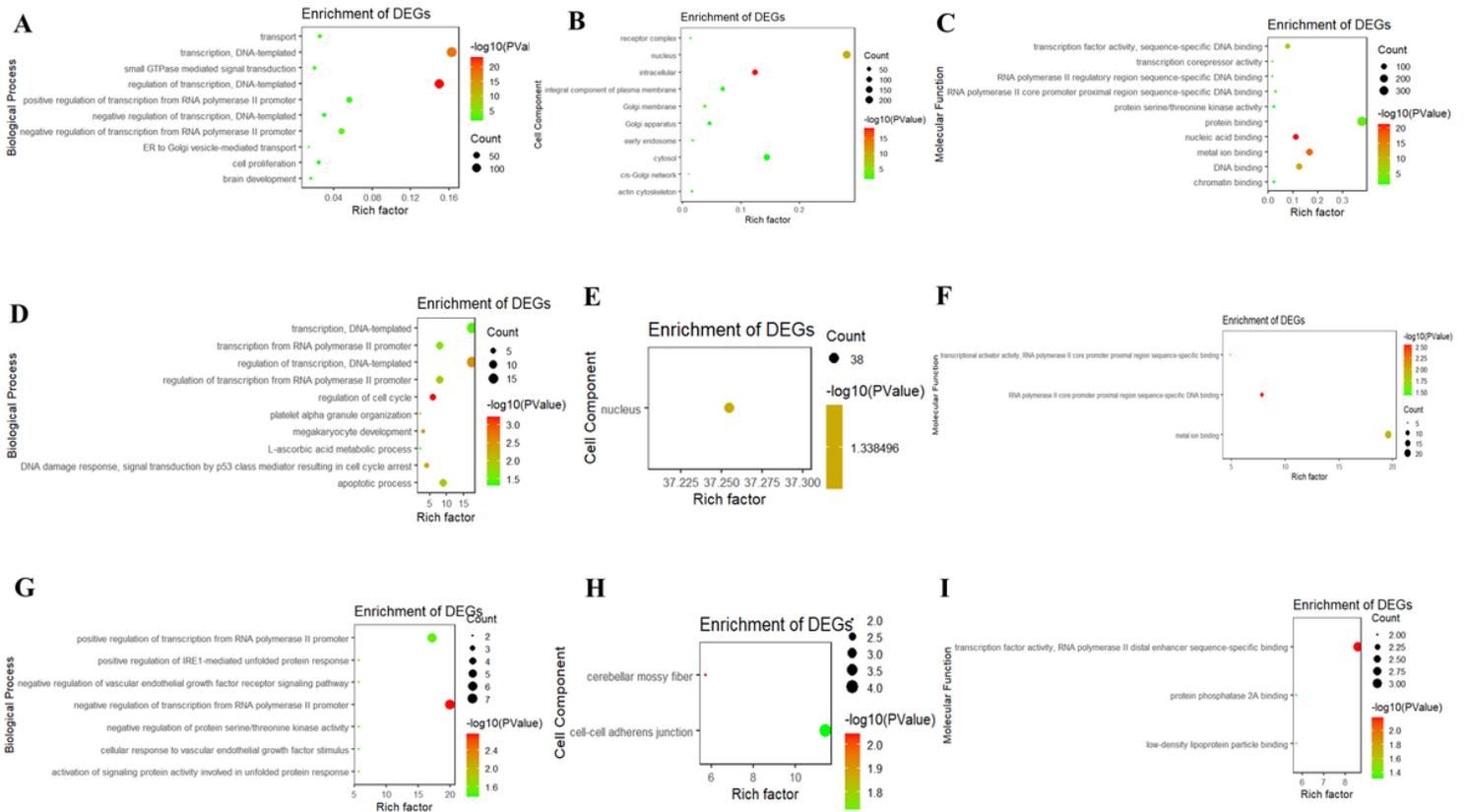


Figure 6

GO terms of the target mRNA of tsRNAs. A-C tiRNA-Glu-TTC-001. **D-F** tiRNA-Lys-CTT-003. **G-I** tRF-Ser-TGA-007. **A&D&G** biological process; **B&E&H** cellular component; **C&F&I** molecular function. GO, Gene ontology.

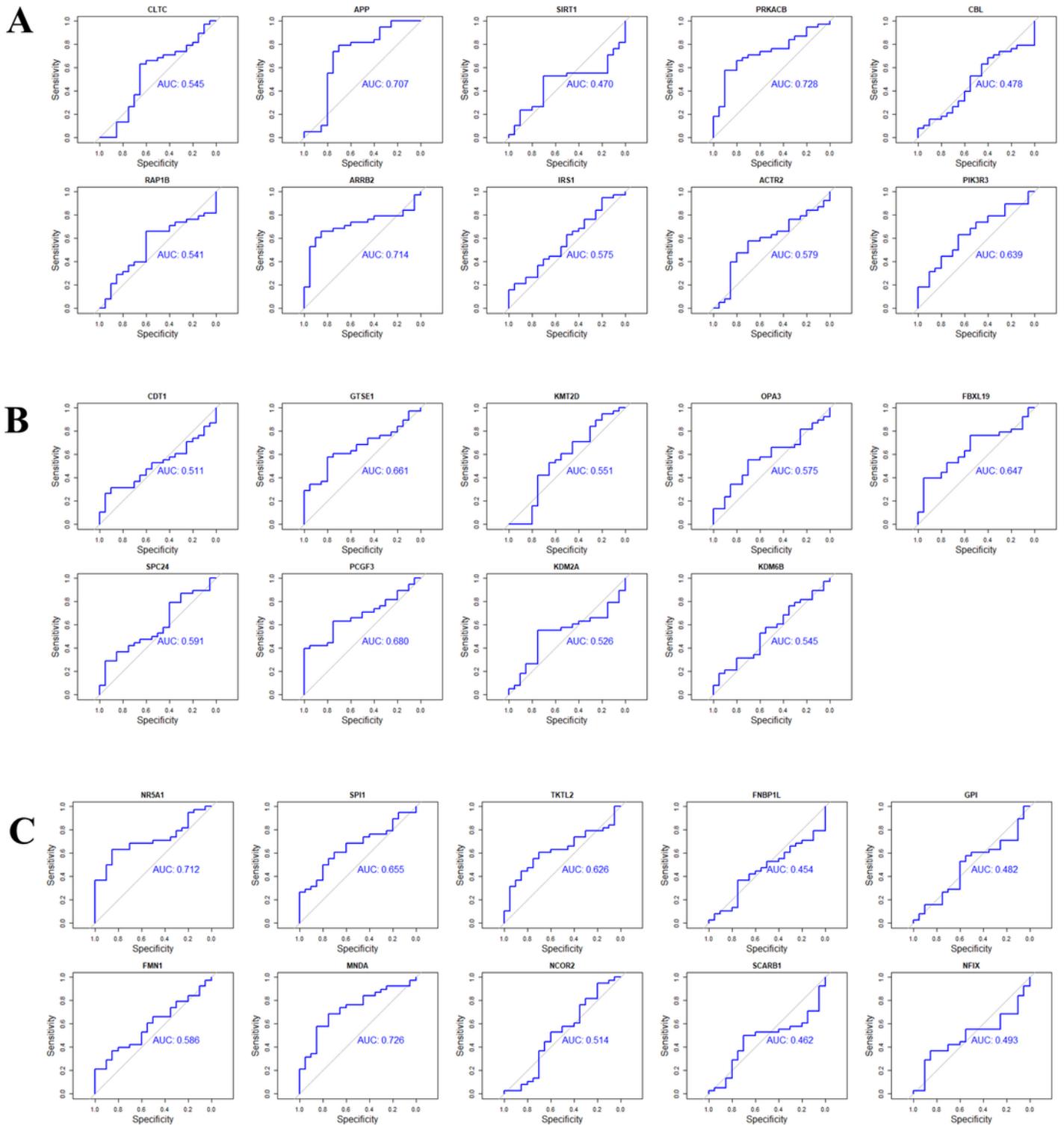


Figure 7

The diagnosis information of the hub genes. A The ROC curves were used to identify the diagnosis information of the hub genes associated with tiRNA-Glu-TTC-001 and, except SIRT1 and CBL, hub genes had a significantly diagnostic value (AUCs > 0.50). **B** The ROC curves were used to identify the diagnosis information of the hub genes associated with tiRNA-Lys-CTT-003 and, except KIAA0101, which was not expressed in the serum of patients with sarcoidosis, all hub genes had a significantly diagnostic value

(AUCs > 0.50). **C** The ROC curves were used to identify the diagnosis information of the hub genes associated with tRF-Ser-TGA-007 and NR5A1, SPI1, TKTL2, FMN1, MNDA, and NCOR2 had a significantly diagnostic value (AUCs > 0.50). ROC curve: receiver operating characteristic curve; AUC: the area under the ROC curve.

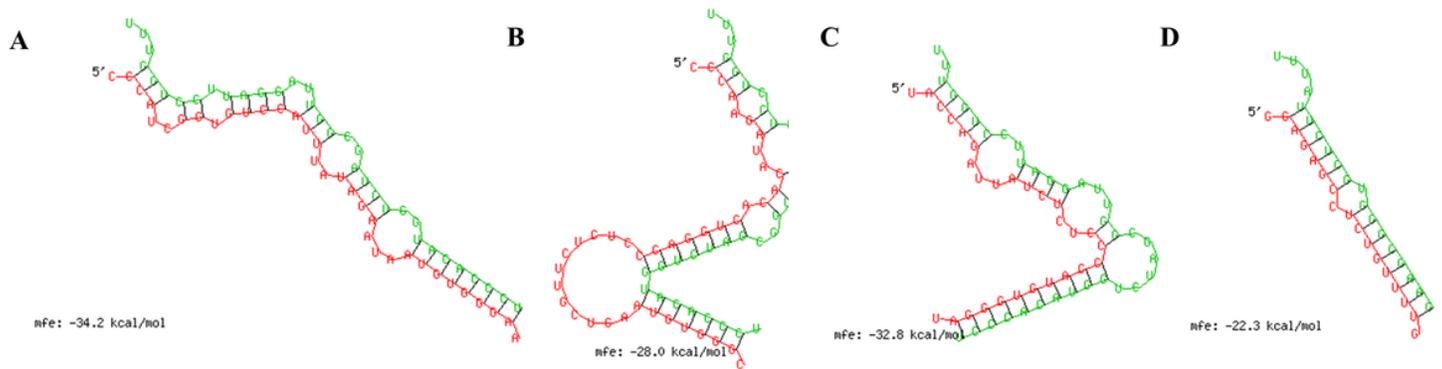


Figure 8

The binding region of tsRNA and seed sequence were exhibited. A-C The base complementary pairing model between tRNA-Glu-TTC-001 and APP, PRKACB, and ARRB2. **D** The base complementary pairing model between tRF-Ser-TGA-007 and NR5A1. Green: the binding region of tsRNA; Red: seed sequence of target gene; mfe: minimum free energy.