Research article

Genome-wide expression profiling of long non-coding RNAs and competing endogenous RNA networks in alopecia areata

Sisi Qi, Youyu Sheng, Ruiming Hu, Feng Xu, Ying Miao, Jun Zhao and Qinping Yang[*]

Department of Dermatology, Huashan Hospital, Fudan University, 12 Cent Urumqi Rd, Shanghai 200040, China

* Correspondence: Email: dxq93216@163.com; Tel/Fax: +86-13671962231.

Abstract: *Background:* Long non-coding RNAs (lncRNAs) regulate gene expression in concert with microRNAs (miRNAs) and mRNAs. This study was designed to explore the potential roles of lncRNAs and their related competing endogenous RNA (ceRNA) networks in alopecia areata (AA). *Methods:* This study comprised six participants (three AA patients and three healthy individuals) whose serum lncRNA profiles were evaluated by lncRNA sequencing. Following differential expression analysis, and function enrichment analysis, a lncRNA-miRNA-mRNA network was then constructed using various bioinformatics tools and validated using quantitative reverse-transcription PCR (qRT-PCR). *Results:* We identified 220 mRNAs and 166 lncRNAs that were differentially expressed in AA patients. The differentially expressed mRNAs were predominantly associated with cytokine-cytokine receptor interactions, MAPK signaling and Ras signaling pathways. The differentially expressed lncRNAs were primarily associated with cytokine-cytokine receptor interactions, chemokine signaling pathways, axon guidance, and legionellosis. In addition, qRT-PCR analyses verified the upregulation of AC005562.1, AF131217.1, and RP11-251G23.5 and downregulation of RP11-231E19.1 in AA patients. *Conclusion:* We constructed a complex ceRNA network for AA and discovered that various RP11 lncRNAs including RP11-251G23.5 and RP11-231E19 may play a crucial role in the pathogenesis of AA via the regulation of the cytokine-cytokine receptor interaction pathway, which could serve as a therapeutic target for alopecia areata in clinical interventions.

Keywords: sequencing; long non-coding RNAs; competing endogenous RNAs; differential expression analysis; alopecia areata

1. Introduction

Alopecia areata (AA) is an organ-specific autoimmune disease characterized by distinct round patches of hair loss. Previous studies have found that the incidence of AA varies geographically, affecting approximately 0.1 to 0.2% of the US population [1], 0.3% of the South Korean population [2] and 3.8% of the Singaporean population [3]. Clinical studies have shown that AA is triggered by a combination of genetic predisposition and environmental factors [4,5], and is affected by age and ethnicity. Recent studies have described the wide range of clinical presentations of AA. A gene expression study, by Petukhova *et al.*, on scalp sections from 96 patients demonstrated that ATG4B and SMARCA2 are involved in autophagy and chromatin remodeling in these patients [6]. Using meta-analysis, the rs2476601 polymorphism of the PTPN22 gene was also shown to be correlated to AA susceptibility [7]. However, the molecular mechanisms underlying AA development still require further investigation.

Long non-coding RNAs (lncRNAs) are widely produced in mammals and other eukaryotes [8], and are much longer than small non-coding RNAs. Almost all aspects of cell biology including transcription, posttranscriptional processing and chromosome remodeling involve lncRNAs [9]. The levels of lncRNA expression typically correlates with their function, as mature lncRNAs are independently functional end products. Non-coding RNAs (ncRNAs) therefore present an intrinsic diagnostic advantage over protein-coding RNAs and a recent study by Bao *et al*. suggested that differentially expressed lncRNAs may represent novel therapeutic targets for AA [10]. Additionally, Sheng et al. [11] developed an integrative computational method to systematically identify genome-wide dysregulated lncRNAs and their related ceRNA network. Their results indicated that lncRNAs can function as a part of a ceRNA network, and their differential expression has been implicated in the development and progression of AA. However, their results were not validated.

Thus, in this study, we collected blood samples from three AA patients and three healthy controls to generate blood lncRNA expression profiles for both conditions. We went on to develop a lncRNA-microRNA (miRNA)-mRNA network (ceRNA network) and validated our microarray data using quantitative reverse transcription PCR (qRT-PCR) and the GEO database. We identified more than 220 differentially expressed mRNAs and 166 differentially expressed lncRNAs in AA patients. Moreover, these differentially expressed lncRNAs were primarily associated with cytokine-cytokine receptor interactions, chemokine signaling pathways, axon guidance, and legionellosis. Taken together, our findings may help to reveal the underlying molecular mechanisms of AA pathogenesis and may provide a theoretical framework for developing novel clinical interventions.

2. Material and methods

2.1. Data processing

2.1.1. Sample source

The samples, obtained from the blood of healthy controls and AA patients, were divided into two groups with three samples in each group: a healthy control group (Normal, W40, W42 and W43) and an AA patient group (AA, A33, A34, and A35).

2.1.2. Total RNA extraction, library construction and sequencing

Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The integrity and purity of the RNA samples were monitored using by gel electrophoresis and at absorbance ratios off or A260/A280 described by the using Nanodrop. After using the removal of rRNA Ribo-Zero™ kit to remove the rRNA, the remaining RNA was randomly digested to produce cut into short fragments. Then first-strand of cDNA was synthesized using these fragments and random hexamers; the second-strand of cDNA was then synthesized by mixing the first-strand cDNA with buffer, dNTPs, RNase H and DNA polymerase I. This cDNA was purified and degraded by uracil-N-glycosylase. The RNA fragments were separated by agarose gel electrophoresis. These, and fragments were expanded using with PCR and these. The PCR products were sequenced using an Illumina HiSeq™ 2000 instrument (Illumina, Inc., San Diego, CA, USA).

The sequencing data were deposited in the NCBI SRA dataset (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA663747), with the accession number of PRJNA663747.

2.1.3. Quality control of sequencing data

Both low-quality reads and incorrectly sequenced bases were observed in our sequencing data. In order to filter out unreliable bases and reads we evaluated and cleaned the data as follows: First, the sequencing tape joints were removed; then, when the number of N (a base that cannot be recognized by the instrument) bases in any of the sequencing reads exceeded 10% of the number of read bases, or the number of low-quality ($Q \leq 5$) bases in any of the sequencing reads exceeded 50% of the total, the paired reads were removed. Finally, clean data from all six samples were then subjected to further analyses.

2.1.4. Mapping of clean reads to the reference genome

Top-Hat software (v2.1.0) was used to map clean reads to the human reference genome (GENCODE download, GRCh38) [12] using the default parameters.

2.1.5. mRNA and lncRNA annotation

Feature-Counts software (v1.6.0) was used to obtain the read-count information for each gene alignment based on the human gene annotation information (Release 25) provided by GENCODE [12,13]. Genes annotated as "protein_coding" were categorized as mRNAs. Genes annotated as "antisense", "sense_intronic", "lincRNA", "sense_overlapping", "processed_transcript", "3prime_overlapping_ncRNA", or "non_coding" were categorized as lncRNAs.

2.2. Transcriptional profiling and principal component analysis (PCA)

The correlation of gene expression levels between samples is an important indicator of the reliability of the experiment and whether the sample selection is reasonable. Therefore, we evaluated the correlation of this data as follows: the Pearson correlation coefficient was calculated for every pair of samples using the "cor" function in R3.4.1 (https://stat.ethz.ch/R-manual/R-devel/ Library/stats/html/cor.html). A Pearson correlation coefficient that approaches 1 indicates highly correlated expression patterns between samples.

After PCA, the first two principal components were selected, and the distribution of the samples

in a two-dimensional plane was drawn based on their principal component score. Prcomp (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html) was used to reduce the data, and the ggfortify package (Version: 0.4.5, https://mirrors.tuna.tsinghua.edu.cn/CRAN/bin/windows/contrib/3.4/ggfortify_0.4.5.zip) was used to draw the PCA diagrams.

2.3. Screening of differentially expressed mRNAs and lncRNAs

First, the TMM algorithm in the edgeR package (Version: 3.4, http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) was used to standardize the raw count from the sequencing data and then convert this value to a logCPM value [14,15]. Differential expression analysis was performed by comparing the mRNAs/lncRNAs expression in the AA group to their expression in the control group. All mRNAs/lncRNAs were analyzed to obtain corresponding p and logFC values. A p-value of < 0.05 and a |logFC| value of > 1 were used as the cutoff values to establish differential expression.

2.4. Functional enrichment and pathway analysis of differentially expressed mRNAs

Gene Ontology analysis of the differentially expressed mRNAs was performed using the R package cluster Profiler (version: 3.8.1, http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) [16,17], and included annotations of the Biological processes (BP), molecular function (MF), cell fraction (CC), and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis [18]. A p-value of < 0.05 was used as the cutoff for determining statistical significance.

2.5. lncRNA and mRNA co-expression analysis and predicting lncRNA function

The Pearson correlation coefficients of differentially expressed mRNAs and lncRNAs were calculated based on matched mRNA and lncRNA data. Meanwhile, correlation tests were also performed to screen the pairs (|r| > 0.95 and p < 0.05). Co-expression network maps were constructed using Cytoscape software (version 3.4.0, http://chianti.ucsd.edu/cytoscape-3.4.0/) [19] and the top 30 lncRNAs were selected based on the number of their target genes. KEGG pathway enrichment analysis was performed on the mRNAs corresponding to each lncRNA using the R package clusterProfiler. A p-value of < 0.05 was used as the cutoff for determining statistical significance.

2.6. miRNA and ceRNA network predictions

We developed the ceRNA network based on the evaluation of the positively correlated lncRNA-mRNA pairs (r > 0.95 and p < 0.05). We then used the miRWalk2.0 (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) tool to select six commonly studied miRNA-mRNA pairs from the transcripts identified in our lncRNA-mRNA correlation analysis [20]. In addition, if the predicted miRNA-mRNA relationship appeared in the predictive databases (miRWalk, Microt4, miRanda, miRDB, RNA22 and Targetscan), we concluded that the miRNA regulates the target gene, and these hits were used to confirm our miRNA-mRNA pairs. The online lnCeDB database was used to identify any lncRNA targets of these miRNAs [21], and the lncRNA-miRNA pairs with at least one binding site were selected for further evaluation.

A lncRNA-miRNA-mRNA relationship network (ceRNA) was then constructed based on the positive correlations between the differentially expressed mRNAs and lncRNAs ($r > 0.95$). lncRNAs and mRNAs in the ceRNA network that are regulated by the same miRNA were regarded as ceRNAs. Finally, the Cytoscape plug-in CytoNCA (Version 2.1.6, http://apps.cytoscape.org/apps/cytonca) was used to analyze the degree of node connection [22] using the "without weight" parameter. A higher connection suggests greater importance of nodes in the network.

2.7. Verification analysis in clinic samples and public databases

2.7.1. qRT-PCR evaluation of lncRNAs

A total of 5 mL of blood was obtained from three healthy controls and three AA patients. Peripheral blood mononuclear cells (PBMCs) were separated from the whole blood using human peripheral blood mononuclear cell separation fluid (Solarbio, Beijing, China). $CD^{4+}T$ cells were separated from 1 mL of PBMCs using the Magnetic Bead CD^{4+} T-Cell Kit following the manufacturer's instructions. Trizol reagent (Invitrogen) was used to extract total RNA. RNA quality and purity were then quantified using a Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA). Reverse transcription was performed using a PrimeScript™RT Master Mix for qPCR (TAKARA, Beijing, China) according to the manufacturer's instruction and cDNA was obtained after the reverse transcription reaction (37 °C for 60 min and 85 °C for 5 s). The expression levels of AC005562.1, AF131217.1, RP11-251G23.5, RP11-231E19.1, and RP11-295G20.2 mRNAs were detected by qRT-PCR analysis. The qRT-PCR primers used in this study were as follows: AC005562.1 (AC005562.1-F: 5-CATATGGCCTGCTGCCTCT-3; AC005562.1-R: 5-GCTGTTAGACATGCAGTGTTGC-3), AF131217.1 (AF131217.1-F: 5-GGACACAGAGAAAGCACCATC-3; AF131217.1-R: 5-GGACACAGAGAAAGCACCATC-3), RP11-251G23.5 (RP11-251G23.5-F: 5-GTGGTGGGAATGCCAGATA-3; RP11-251G23.5-R: 5-CAAGACCAGCCTGAGCAAC-3), RP11-231E19.1 (RP11-231E19.1-F: 5-AAGGGTGGAGTTGGGAGAAAC-3; RP11-231E19.1-R: 5-GCTGTCAGATGGAGTGGGCT-3), and RP11-295G20.2 (RP11-295G20.2-F: 5-GGCATGTTCTGCTCTGGCAC-3; RP11-295G20.2-R: 5-TGGGTTGACTGGATGGATCC-3). The qRT-PCR cycling conditions were as follows: 50 °C for 3 min and then pre-incubation at 95 °C for 3 min, followed by 40 cycles at 95 °C for 15 s and 60 °C for 30 s. Melt curves were generated from 60 °C to 95 °C at increments of 0.5 °C for 10 s.

2.7.2. Source of validation dataset and differential expression analysis

The validation dataset used in this study was retrieved from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database from NCBI [23]. After careful review, the gene expression profiles from the GSE68801 dataset (122 samples), obtained using the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array were used as our validators. These data are freely available online.

Genes that were differentially expressed in the AA and control groups were analyzed using GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r) with the following thresholds: $p < 0.05$ and $|logFC| > 0.585$ (|fold change| > 1.5). Briefly, based on the probe sequence comments from the chip platform file, the human reference genome (GRCh38) was downloaded from the GENCODE database (https://www.gencodegenes.org/releases/current.html). Next, all probe sequences were aligned to the reference genome to annotate the corresponding gene for each probe using seqmap

software and any probes annotated as "protein_coding" were considered to be mRNA probes and those annotated as "antisense", "sense_intronic", "lncRNA", "sense_overlapping" or "processed_transcript" were considered to be lncRNA probes. In addition, overlapping mRNAs and lncRNAs were visualized using Draw Venn Diagram software (http://bioinformatics.psb.ugent.be/webtools/Venn/).

2.8. Statistical analysis

The data are presented as the mean ± standard deviation (SD) for three replicates. All statistical analyses were performed using GraphPad Prism 5 (GraphPad Software, San Diego, CA) and a p-value of < 0.05 was determined to be statistically significant.

3. Results

3.1. Summary of sequencing data quality

A total of 89.36 Gb of raw data were obtained, and 86.35 Gb of filtered clean data remained. The quality of the sequencing output for each sample is shown in Table S1.

Table S1. The output quality of the sequencing data.

Sample	Raw Reads	Clean Reads	Raw Base(G)	Clean Base(G)	Effective Rate (%)	Error Rate (%)	Q20 (%)	Q30 (%)	GC Content (%)
A33	55,318,245	52,775,447	16.6	15.83	95.4	0.01	95.88	90.37	46.63
A34	43,631,112	41,961,832	13.09	12.59	96.17	0.01	95.9	90.44	45.66
A35	47,502,747	46,203,539	14.25	13.86	97.26	0.01	96.62	92.05	44.84
W40	49,906,830	48,066,446	14.97	14.42	96.31	0.01	96.41	91.57	46.41
W42	47,013,507	45,773,807	14.1	13.73	97.36	0.01	96.61	92.02	46.36
w43	54,505,394	53,051,128	16.35	15.92	97.33	0.01	96.3	91.35	46.25

3.2. Data alignment and transcript notes

Top-Hat software (v2.1.0) was used to map clean reads to the human reference genome (GENCODE download, GRCh38) and the comparison results are summarized in Table S2.

Table S2. Top-Hat align summary.

	A33	A34	A35	W40	W42	W43
Left	52775447	41961832	46203539	48066446	45773807	53051128
Left	48647769	38671490	41883361	43089099	41596215	47561326
Left	92.2%	92.2%	90.6%	89.6%	90.9%	89.7%
Right	52775447	41961832	46203539	48066446	45773807	53051128
Right	43891896	35122585	38248521	38922731	38046922	42852427
Right	83.2%	83.7%	82.8%	81.0%	83.1%	80.8%
overall	87.7%	87.9%	86.7%	85.3%	87.0%	85.2%
Aligned	42493745	33942403	36673635		36506591	41122249

Then, the read information for each mRNA/lncRNA alignment was obtained and lowly-expressed mRNAs/lncRNAs were filtered out. Finally, mRNAs/lncRNAs with expression values > 0 in all six samples were retained, producing an expression matrix of 14,418 mRNAs and 5,239 lncRNAs.

3.3. Inter-sample correlation analysis and PCA

The correlation coefficient value (P)between samples ranged from 0.845 to 1 (Figure 1). The average correlation of the samples within a group was 0.971, and the average correlation of samples between groups was 0.939. The intra-group correlation was higher than that of inter-group correlation indicating that the patient group was clearly separated from the control group.
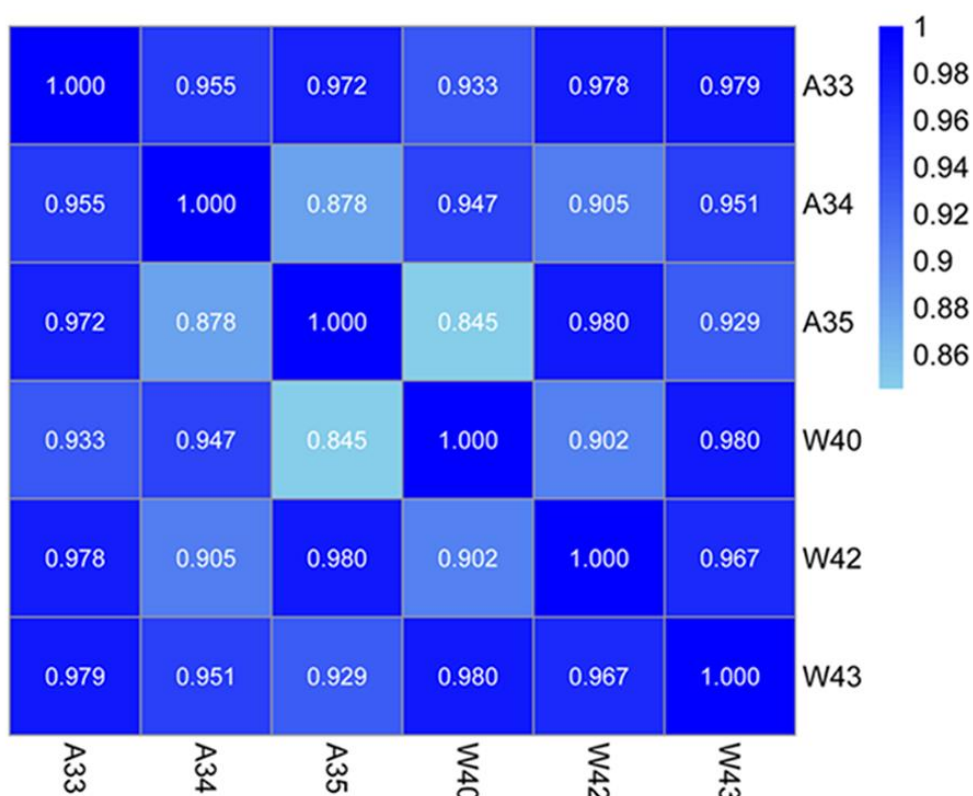

Figure 1. Heatmap of the correlation between two pairs based on transcript abundance and PCA value. A. Heatmap of the correlation between pairs based on transcript abundance. B. PCA.

3.4. Basic statistics of the differential expression analysis

The mRNA and lncRNA differential expression analyses were performed using the edgeR package, and a total of 220 differentially expressed mRNAs (including 108 upregulated and 112 downregulated mRNAs) and 166 differentially expressed lncRNAs (including 66 upregulated and 100 downregulated lncRNAs) were identified. Bidirectional hierarchical clustering and volcano plots of the differentially expressed mRNAs and lncRNAs are shown in Figure 2. Differentially expressed mRNAs and lncRNAs could be easily distinguished between the AA and control groups.
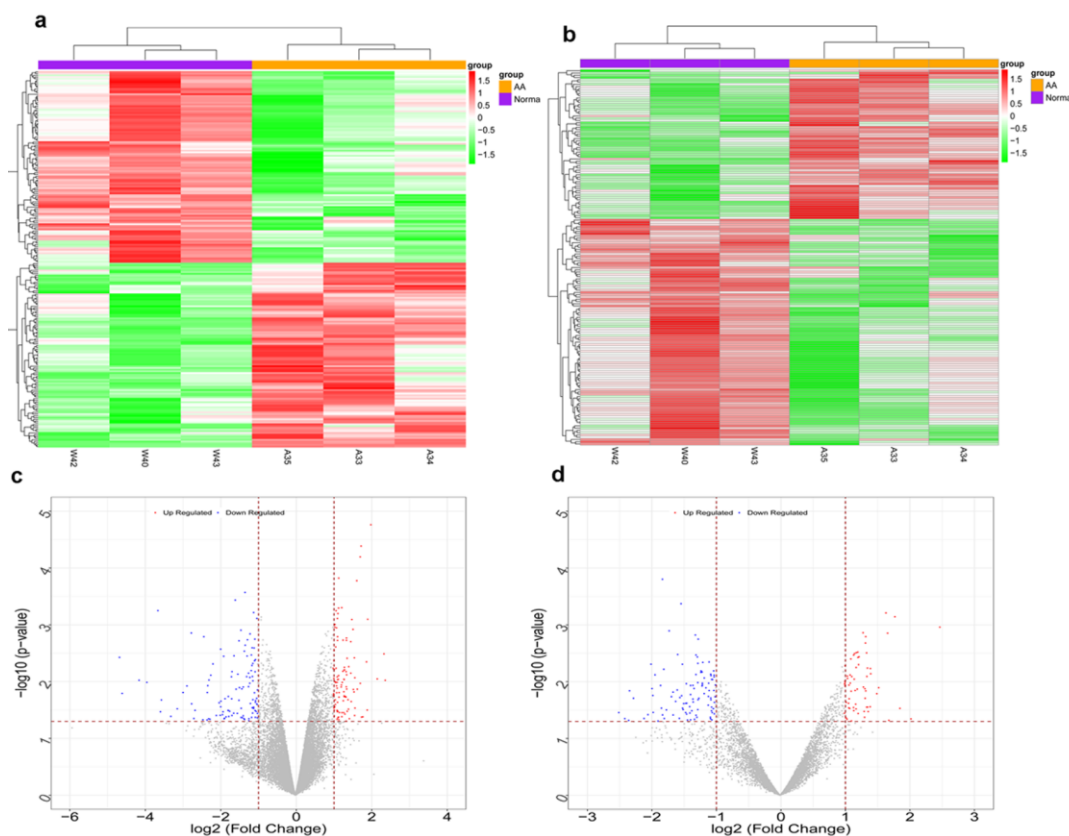
Figure 2. Heatmaps and volcano plots depicting differential mRNA and lncRNA expression. A. mRNA expression heatmap. B. lncRNA expression heatmap. The top orange bar indicates the patient sample and the purple bar indicates the control sample. C. Volcano plot of mRNA expression. D. Volcano plot of lncRNA expression. Red dots represent upregulation, blue dots represent downregulation and gray dots represent no significant difference.

3.5. Functional enrichment and pathway analysis of differentially expressed mRNAs

GO functional enrichment analysis and KEGG pathway enrichment analysis were performed on the differentially expressed mRNA transcripts. A total of 165 biological processes including leukocyte aggregation, peptidyl-tyrosine phosphorylation, and cell chemotaxis, 11 cell components, including side of membrane, receptor complex, and cytoplasmic region, and 27 molecular functions, including guanyl-nucleotide exchange factor activity, protein tyrosine kinase activity, and growth factor binding were shown to be enriched in this dataset. Moreover, 13 KEGG pathways including cytokine-cytokine receptor interaction, MAPK signaling pathway, and Ras signaling pathway were found to be enriched, and the top 10 enriched functions are shown in Figure 3.

3.6. Correlation between mRNA and lncRNA expression and predicting lncRNA function

Correlation analysis of the above differentially expressed mRNAs and lncRNAs was performed. In total, 3,356 significantly correlated pairs (1,415 negatively correlated pairs and 1,941 positively correlated pairs), including 217 mRNAs and 161 lncRNAs were screened. The top 9 pairs are described in Table 1A. Co-expression networks are shown in Figure 4A.
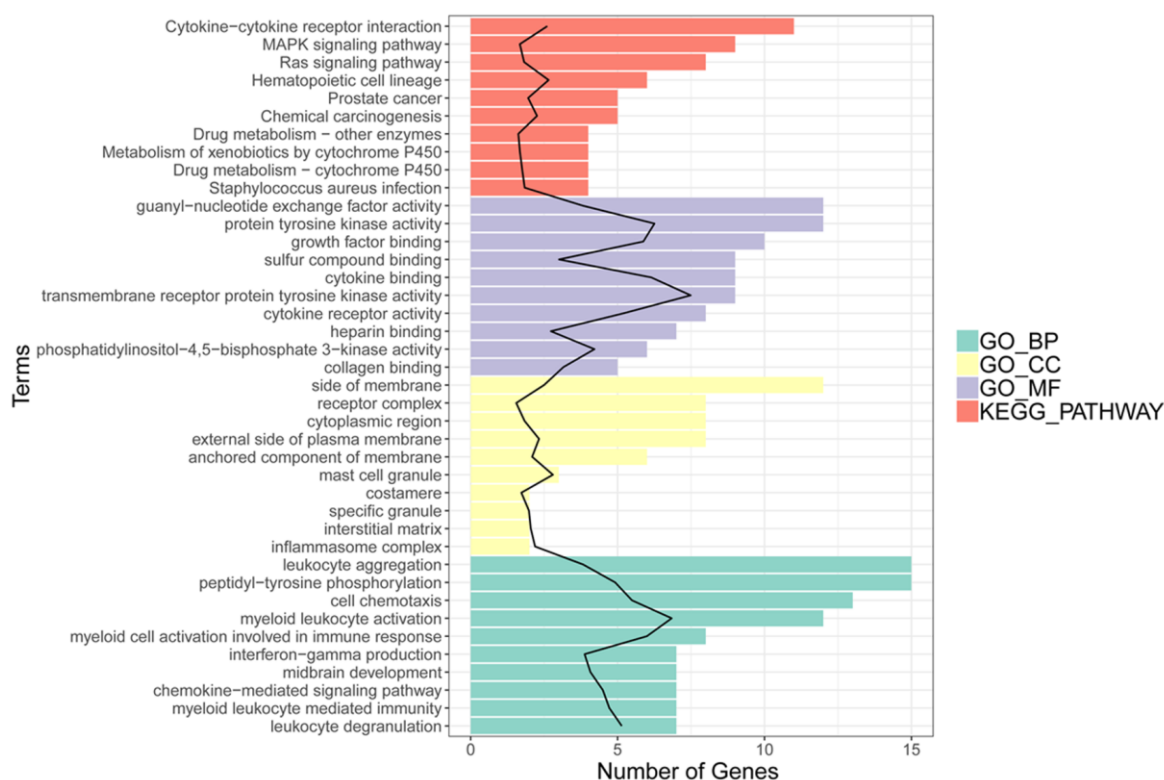
Figure 3. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analysis. The black line indicates the -log10 (*p*-value), the green bars indicate enriched biological process GO terms, the yellow bars indicate enriched cell fraction GO terms, the light purple bars represent enriched molecular function GO terms and the red bars indicate enriched KEGG pathway terms. In each case the length of the bar indicates the number of genes in each enriched term.

Based on known lncRNA-mRNA relationships, the mRNA transcripts were treated as the designated targets of the lncRNA transcripts. Moreover, KEGG pathway enrichment analysis was performed on the top 30 lncRNAs to predict their function. As shown in Figure 4B, a total of 12 lncRNAs were enriched in various pathways including cytokine-cytokine receptor interactions, chemokine signaling pathways, axon guidance, and legionellosis.

3.7. miRNA prediction and ceRNA network construction

First, 1,941 positively correlated lncRNA-mRNA pairs were screened, including 202 mRNAs and 152 lncRNAs. Second 773 miRNA-mRNA pairs were identified including 84 mRNAs and 486 miRNAs, using miRWalk and then, 5,265 miRNA-lncRNA pairs, including 1,612 miRNAs and 113 lncRNAs, were identified using lnCeDB. Next, positively correlated lncRNA-mRNA pairs regulated by the same miRNA were identified and the other data eliminated. This left us with 82 lncRNA-miRNA-mRNA regulatory relationships, including 48 miRNAs, 33 lncRNAs, and 28 mRNAs. The ceRNA network is shown in Figure 5. A total of 77 miRNA-lncRNA pairs are shown in the network, including 57 miRNA-mRNA pairs, and 55 lncRNA-mRNA pairs.
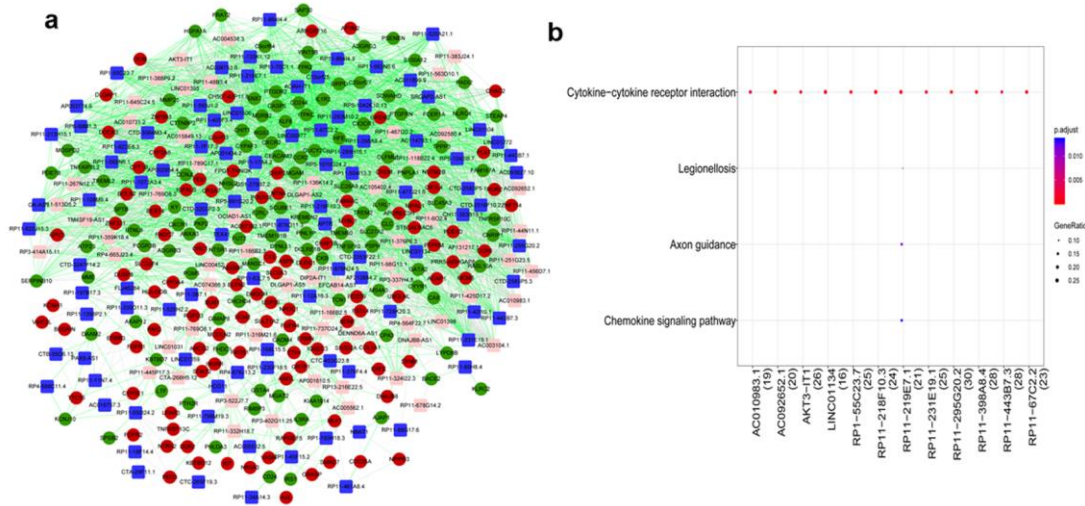
Figure 4. lncRNA-mRNA co-expression network and lncRNA pathway analysis. A. lncRNA-mRNA co-expression network. Red circles represent upregulated mRNAs, green circles indicate downregulated mRNAs, pink squares indicate upregulated lncRNAs, dark blue squares indicate downregulated lncRNAs, green dotted lines indicate positive correlations, and gray dotted lines indicate negative correlations. B. lncRNA pathway analysis. The color indicates decreasing *p*-value (blue to red), and the bubble size indicates the weight of each enriched gene.
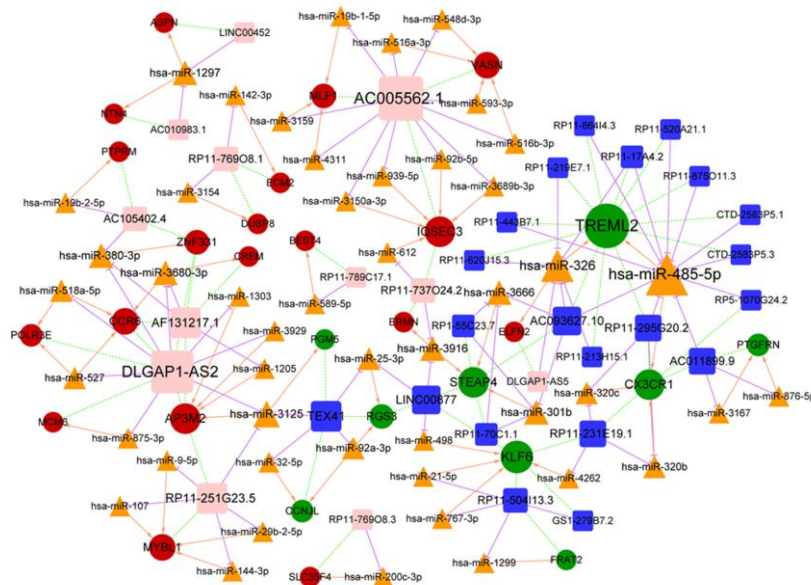

Figure 5. ceRNA network. Red circles represent upregulated mRNAs, green circles represent downregulated mRNAs, yellow triangles represent miRNAs, dark blue squares indicate downregulated lncRNAs, and pink squares indicate upregulated lncRNAs. The purple T-shaped arrow indicate competitive binding between the lncRNAs and miRNAs, the yellow arrow represents the miRNA-mRNA regulatory relationship and the green dotted line indicates a positive relationship between lncRNA and mRNA expression. Finally, the node and font size indicate the degree of connectivity.

Table 1. Correlation analyses for differentially expressed mRNAs and lncRNAs and a summary of the node ranking in the ceRNA network.

A. Correlation statistics for differentially expressed mRNA/ lncRNA pairs.

lncRNA	mRNA	r	p-value
RP11-67C2.2	WNT5B	0.99925121	8.41E-07
RP11-443B7.1	WNT5B	0.999239433	8.67E-07
RP11-231E19.1	FPR2	0.999206747	9.44E-07
RP11-443B7.3	AP3M2	-0.999129738	1.14E-06
RP11-443B7.3	PLAT	-0.998759399	2.31E-06
CTD-2516F10.2	PNPLA1	0.998687993	2.58E-06
RP11-231E19.1	S100A12	0.99849922	3.38E-06
RP11-398A8.4	S100A12	0.998475218	3.49E-06
RP1-197B17.3	PGM5	0.998334638	4.16E-06

B. Ranking of each node in the ceRNA network.

| node	degree	type
hsa-miR-485-5p	13	miRNA
hsa-miR-326	8	miRNA
hsa-miR-3125	5	miRNA
hsa-miR-1297	4	miRNA
hsa-miR-301b	4	miRNA
hsa-miR-3666	4	miRNA
hsa-miR-3680-3p	4	miRNA
hsa-miR-380-3p	4	miRNA
hsa-miR-3916	4	miRNA
AC005562.1	14	lnc_up
DLGAP1-AS2	13	lnc_up
AF131217.1	7	lnc_up
RP11-251G23.5	7	lnc_up
TEX41	7	lnc_down
AC093627.10	6	lnc_down
LINC00877	6	lnc_down
AC011899.9	5	lnc_down
RP11-231E19.1	5	lnc_down
RP11-295G20.2	5	lnc_down
TREML2	14	m_down
KLF6	8	m_down
AP3M2	7	m_up
CX3CR1	7	m_down
IQSEC3	7	m_up
STEAP4	7	m_down
CCR6	5	m_up
MYBL1	5	m_up
VASN	5	m_up
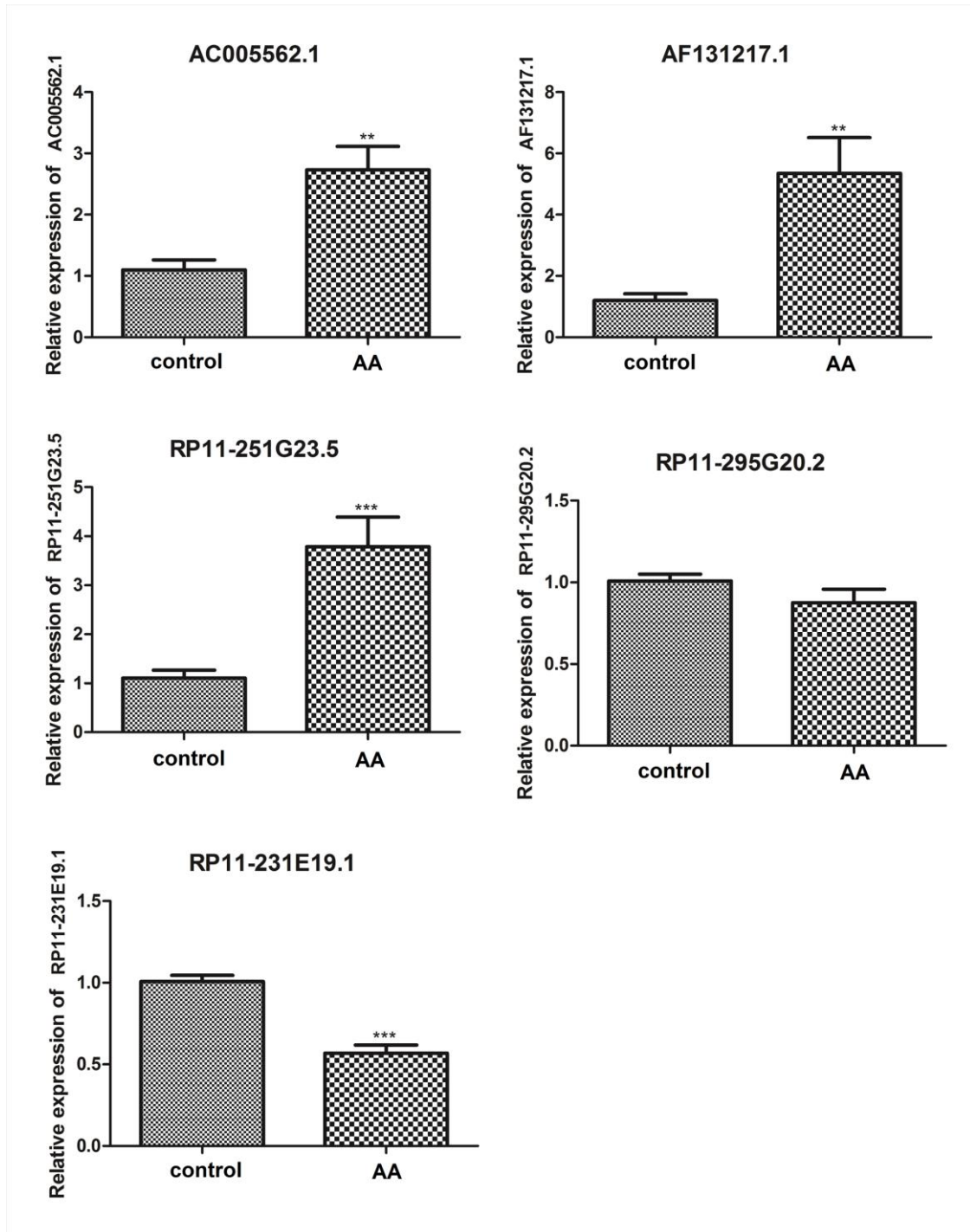
Figure 6. The relative expression of differentially expressed AC005562.1, AF131217.1, RP11-251G23.5, RP11-231E19.1 and RP11-295G20.2 lncRNAs in both groups. $**p < 0.01$, $***p < 0.001$ compared with control. AA: alopecia areata.

Connectivity analysis was performed on each node of the above network, and several mRNAs (TREML2, KLF6, and AP3M2) miRNAs (hsa-miR-485-5p, hsa-miR-326, and hsa-miR-3125) and lncRNAs (AC005562.1, AF131217.1, RP11-251G23.5, RP11-231E19.1, and RP11-295G20.2) with high degrees of connectivity were identified and are summarized in Table 1B.

3.8. Verification analysis

The qRT-PCR analysis confirmed the original bioinformatic predictions for lncRNAs AC005562.1, AF131217.1, RP11-251G23.5, which were significantly upregulated in the disease group ($p < 0.01$) and RP11-231E19.1, which was significantly downregulated in the disease group ($p < 0.01$). However, no significant differences in the expression levels of RP11-295G20.2 were found between the control and disease groups ($p > 0.05$) (Figure 6).

Furthermore, when we evaluated the validation data (GSE68801 dataset) a total of 11 lncRNAs and 70 mRNAs were found to overlap with our experimental dataset (Figure 7).
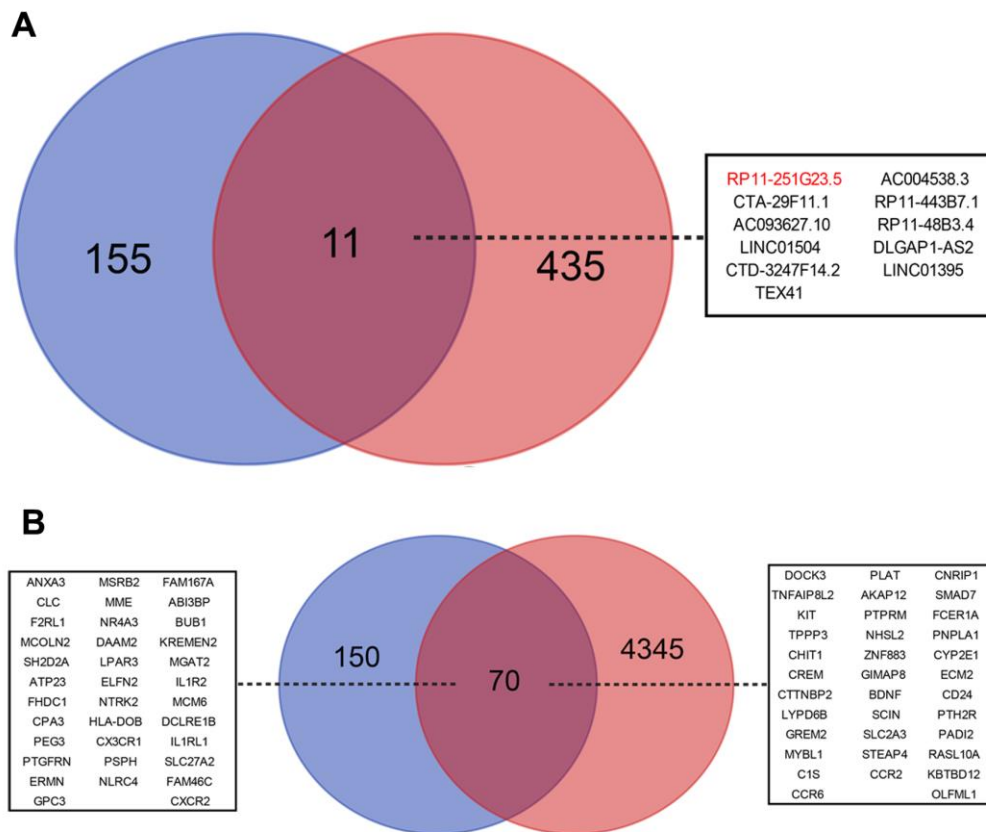

Figure 7. Venn diagram describing the overlapping lncRNAs (A) and mRNAs (B) between the experimental and verification datasets. Blue represents the lncRNAs and mRNAs in the experimental dataset and the light red represents the lncRNAs and mRNAs in the verification dataset.

4. Discussion

ncRNAs have recently been linked to the development and progression of various autoimmune diseases [24], including AA [11]. Sheng et al. [11] reported that NONHSAT062906 and NONHSAT011665 may be associated with the development of AA, but their results were not validated in cell culture or using tissue samples. Thus, studies are still needed to verify the role of these lncRNAs in AA pathogenesis and expand our understanding of AA [25]. Six samples from three AA patients and three healthy controls were obtained for this study and used to complete

transcriptional profiling and construct a lncRNA-miRNA-mRNA regulatory network (ceRNA network) for AA. We uncovered 220 differentially expressed mRNAs and 166 differentially expressed lncRNAs associated with AA and demonstrated that these differentially expressed mRNAs were mainly associated with cytokine−cytokine receptor interactions, MAPK signaling pathways, and Ras signaling pathways. While the differentially expressed lncRNAs were predominantly linked to cytokine-cytokine receptor interactions, chemokine signaling pathways, axon guidance, and legionellosis. qRT-PCR verified the upregulation of AC005562.1, AF131217.1, and RP11-251G23.5 and the downregulation of RP11-231E19.1 in the disease group.

Previous studies have identified an imbalance in the PBMC immune system in AA patients. Specifically, the levels of T helper 17 cells were negatively correlated with disease duration, and Treg levels were higher in severe AA patients[26]. Meanwhile, cytokines such as interleukin-2 and interleukin-8, and Janus kinase and their related pathways play significant roles in the development of AA [27,28]. In this study, differentially expressed mRNAs demonstrated significant enrichment for pathways involved with cytokine-cytokine receptor interactions however, the role of these interactions in AA pathogenesis remain unknown.

More importantly, we used known lncRNA-mRNA relationships and KEGG pathway enrichment analysis to identify 12 lncRNAs enriched in cytokine-cytokine receptor interactions, chemokine signaling pathways, axon guidance, and legionellosis which may help to advance our understanding of AA pathogenesis. Among these pathways, cytokine-cytokine receptor interactions were the most widely enriched pathway and included lncRNAs RP11-231E19.1, RP11-443B7.3, RP11-218F10.3, RP11-295G20.2, RP11-398A8.4 and RP11-67C2.2. Previous studies have shown a link between lncRNA RP11 and the development of various cancers including colorectal cancer, papillary thyroid cancer, and pancreatic cancer [29–31]. However, a link between lncRNA RP11 and the pathogenesis or progression of AA has not been reported till now. Here we describe the differential expression of several RP11 transcripts including the upregulation of RP11-251G23.5 and downregulation of RP11-231E19.1. All of which suggests that the RP11 lncRNAs are involved in the development of AA and may facilitate their effect through the cytokine-cytokine receptor interaction signaling pathway.

5. Conclusion

In summary, we constructed a complex ceRNA network for AA, and suggest that RP11 lncRNAs, including RP11-251G23.5 and RP11-231E19, might play a critical role in the pathogenesis of AA via their regulation of the cytokine-cytokine receptor interaction signaling pathway, which could serve as a novel therapeutic target for AA clinical interventions. In addition, this study identified several novel lncRNAs that may help to uncover the regulatory mechanisms of and biomarkers in AA. However, further clinical studies are still needed to confirm their function in AA.

Acknowledgements

Conflicts of interest

The authors have no conflicts of interest to declare.

Authors' contributions

Conception and design: QY; acquisition of data: YS; analysis and interpretation of data: RH and FX; statistical analysis: YM and JZ; funding: QY; drafting the manuscript: SQ; revision of the manuscript for important intellectual content: QY. All authors read and approved the final manuscript.

References

1. L. C. Strazzulla, E. H. C. Wang, L. Avila, K. Lo Sicco, N. Brinster, A. M. Christiano, et al., Alopecia areata: Disease characteristics, clinical evaluation, and new perspectives on pathogenesis, *J. Am. Acad. Derm.*, **78** (2018), 1–12.

2. J. H. Lee, H. J. Kim, K. D. Han, J. H. Han, C. H. Bang, Y. M. Park, et al., Incidence and prevalence of alopecia areata according to subtype: A nationwide, population-based study in South Korea (2006–2015), *Brit. J. Derm.,* (2019).

3. E. Tan, Y. K. Tay, C. L. Goh, Y. Chin Giam, The pattern and profile of alopecia areata in Singapore--a study of 219 Asians, *Int. J. Derm.*, **41** (2002), 748–753.

4. F. L. Xiao, S. Yang, J. B. Liu, P. P. He, J. Yang, Y. Cui, et al., The epidemiology of childhood alopecia areata in China: A study of 226 patients, *Pediat. Derm.*, **23** (2006), 13–18.

5. S. Yang, J. Yang, J. B. Liu, H. Y. Wang, Q. Yang, M. Gao, et al., The genetic epidemiology of alopecia areata in China, *Brit. J. Derm.*, **151** (2004), 16–23.

6. L. Petukhova, A. V. Patel, R. K. Rigo, L. Bian, M. Verbitsky, S. S. Cherchi, et al., Integrative analysis of rare copy number variants and gene expression data in Alopecia Areata implicates an etiological role for autophagy, *Exp. Dermatol.,* (2019).

7. Z. X. Lei, W. J. Chen, J. Q. Liang, Y. J. Wang, L. Jin, C. Xu, et al., The association between rs2476601 polymorphism in PTPN22 gene and risk of alopecia areata: A meta-analysis of case-control studies, *Medicine (Baltimore)*, **98** (2019), e15448.

8. T. R. Mercer, M. E. Dinger, J. S. Mattick, Long non-coding RNAs: Insights into functions, *Nat. Rev. Genet.*, **10** (2009), 155–159.

9. V. Simion, S. Haemmig, M. W. Feinberg, LncRNAs in vascular biology and disease, *Vascul. Pharmacol.*, **114** (2019), 145–156.

10. L. Bao, A. Yu, Y. Luo, T. Tian, Y. Dong, H. Zong, et al., Genomewide differential expression profiling of long non-coding RNAs in androgenetic alopecia in a Chinese male population, *J. Eur. Acad. Dermatol. Venereol.*, **31** (2017), 1360–1371.

11. Y. Sheng, J. Ma, J. Zhao, S. Qi, R. Hu, Q. Yang, Differential expression patterns of specific long noncoding RNAs and competing endogenous RNA network in alopecia areata, *J. Cell Biochem.*, **120** (2019), 10737–10747.

12. S. Ghosh, C. K. Chan, Analysis of RNA-Seq data using TopHat and Cufflinks, *Methods Mol. Biol.*, **1374** (2016), 339–361.

13. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, **30** (2014), 923–930.

14. O. Nikolayeva, M. D. Robinson, edgeR for differential RNA-seq and ChIP-seq analysis: An application to stem cell biology, *Methods Mol. Biol.*, **1150** (2014), 45–79.

15. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, **26** (2010), 139–140.

16. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological

themes among gene clusters, *OMICS*, **16** (2012), 284–287.

17. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, et al., Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium, *Nat.Genet.*, **25** (2000), 25–29.

18. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, *Nucl. Acids Res.*, **28** (2000), 27–30.

19. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al., Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13** (2003), 2498–2504.

20. H. Dweep, N. Gretz, miRWalk2.0: A comprehensive atlas of microRNA-target interactions, *Nat. Methods*, **12** (2015), 697.

21. S. Das, S. Ghosal, R. Sen, J. Chakrabarti, lnCeDB: Database of human long noncoding RNA acting as competing endogenous RNA, *PLoS One*, **9** (2014), e98965.

22. Y. Tang, M. Li, J. Wang, Y. Pan, F. X. Wu, CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks, *Biosystems*, **127** (2015), 67–72.

23. T. Barrett, T. O. Suzek, D. B. Troup, S. E. Wilhite, W.-C. Ngau, P. Ledoux, et al., NCBI GEO: Mining millions of expression profiles—database and tools, *Nucl. Acids Res.*, **33** (2005), D562–D566.

24. L. C. Tsoi, M. K. Iyer, P. E. Stuart, W. R. Swindell, J. E. Gudjonsson, T. Tejasvi, et al., Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin, *Genome Biol.*, **16** (2015), 24.

25. K. R. Sigdel, A. Cheng, Y. Wang, L. Duan, Y. Zhang, The emerging functions of Long Noncoding RNA in immune cells: Autoimmune diseases, *J. Immunol. Res.*, **2015** (2015), 848790.

26. Y. M. Han, Y. Y. Sheng, F. Xu, S. S. Qi, X. J. Liu, R. M. Hu, et al., Imbalance of T-helper 17 and regulatory T cells in patients with alopecia areata, *J. Dermatol.*, **42** (2015), 981–988.

27. M. Hordinsky, D. H. Kaplan, Low-dose interleukin 2 to reverse alopecia areata, *JAMA Dermatol.*, **150** (2014), 696–697.

28. H. Guo, Y. Cheng, J. Shapiro and K. McElwee, The role of lymphocytes in the development and treatment of alopecia areata, *Expert Rev. Clin. Immunol.*, **11** (2015), 1335–1351.

29. H. Chen, Z. Xu, X. Liu, Y. Gao, J. Wang, P. Qian, et al., Increased expression of Lncrna RP11-397A15.4 in gastric cancer and its clinical significance, *Ann. Clin. Lab. Sci.*, **48** (2018), 707–711.

30. R. Huang, W. Nie, K. Yao, J. Chou, Depletion of the lncRNA RP11-567G11.1 inhibits pancreatic cancer progression, *Biomed. Pharmacother.*, **112** (2019), 108685.

31. Y. Wu, X. Yang, Z. Chen, L. Tian, G. Jiang, F. Chen, et al., m(6)A-induced lncRNA RP11 triggers the dissemination of colorectal cancer cells via upregulation of Zeb1, *Mol. Cancer*, **18** (2019), 87.