LncRNA- and circRNA-associated competing endogenous RNA (ceRNA) networks in leukocytes in traditional Chinese medicine (TCM)-defined Pi-qi-deficiency syndrome and Pi-wei damp-heat syndrome resulting from chronic atrophic gastritis

Leiming You
Beijing University of Chinese Medicine
Wei Wang
Beijing University of Chinese Medicine
Kunyu Li
Beijing University of Chinese Medicine
Xiaopu Sang
Beijing University of Chinese Medicine
Ting'an Li
Beijing University of Chinese Medicine
Shen Zhang
Beijing University of Chinese Medicine
Xinhui Gao
Beijing University of Chinese Medicine
Jiarui Wu
Beijing University of Chinese Medicine
Guangrui Huang
Beijing University of Chinese Medicine
Ting Wang
Beijing University of Chinese Medicine
Anlong Xu (xuanlong@bucm.edu.cn)
Beijing University of Chinese Medicine https://orcid.org/0000-0002-1419-3494

Research

Keywords: Chronic atrophic gastritis, Pi-qi-deficiency syndrome, Pi-wei damp-heat syndrome, Leukocyte, ceRNA biomarker, ceRNA network

Posted Date: October 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-934306/v1

Abstract

Background

To investigate lncRNA/circRNA-associated competing endogenous RNA (ceRNA)-gene regulation underlying leukocyte functions and characteristics, especially the potential ceRNA biomarkers, implicated in traditional Chinese medicine (TCM)-defined Pi-qi-deficiency syndrome (PQDS) and Pi-wei damp-heat syndrome (PDHS) resulting from chronic atrophic gastritis (CAG).

Methods

Based on RNA-sequencing approach, comparing with healthy control population, we identified the PDHS- or PQDS-specific lncRNAs/circRNAs in leukocytes, especially the *Zheng* (syndrome)-specific ceRNAs and their corresponding ceRNA regulatory networks, further decoding their potential functions and pathways.

Results

Despite being the TCM-defined *Zhengs* resulting from the same disease of CAG, the *Zheng*-specific lncRNAs/circRNAs in leukocytes were not same in PQDS and PDHS. There were the *Zheng*-specific lncRNA/circRNA-associated ceRNAs identified in the leukocytes, and their corresponding ceRNA regulatory networks were generated, including the ceRNA-gene binary relationship networks and the miRNA-centered ceRNA-miRNA-gene triple relationship networks. In the generated *Zheng*-specific ceRNA networks, the PQDS-specific ceRNA-governed genes in leukocytes, keeping more complex interactions with each other, were enriched in pathways related to MAPK signaling, receptor tyrosine kinases signaling as well as complement and coagulation cascades. Notably, the enriched pathways associated with adherens junction, focal adhesion, ECM-receptor interaction, ECM-organization and cell surface interactions, were implicated in cell-to-cell adhesion/junction and communication, probably contributing to the characteristics and functions of leukocytes. The PDHS-specific ceRNA-regulated genes, seemed to have no more interactions with each other, were enriched in the biological processes related to regulation of cell morphogenesis, neutrophil activation and degranulation, and lymphocyte mediated immunity such as B-cell receptor signaling, complement and coagulation cascades, and regulation of NK/T-cell mediated cytotoxicity. Importantly, the five exosome-encapsuled ceRNAs, containing *ZFAS1* (NR_036658.2), *AL353719.1* (ENST00000566847), *LOH12CR2* (NR_024061.1) and two new lncRNAs (TCONS_00038035 and TCONS_00027600), particularly higher expression in the leukocytes in PQDS rather than PDHS, could be the potential ceRNA biomarkers for differentiation of the TCM-defined PQDS and PDHS among CAG patients.

Conclusions

These results may provide new insights into the characteristic and functional changes of leukocytes in the two TCM *Zhengs*, especially the *Zheng*-specific ceRNA-mediated gene regulation underlying leukocyte characteristics and functions, with potential leukocyte biomarkers for future application in integrative medicine.

Background

Within Asia, traditional Chinese medicine (TCM) is an ancient medical practice system with more than 3000 years of rich history. TCM plays very important roles in people's healthcare in China, and is gaining global popularity [1]. "*Zheng*" in Mandarin Chinese, meaning TCM syndrome or pattern, is an essential concept of TCM theory containing rich integrated thoughts [2, 3]. In fact, it is a thousand-year-old key diagnostic concept in TCM, defined as a pattern of symptoms and physical signs in a patient at a specific stage during the course of a disease [2–4]. A TCM-defined *Zheng* of disease is usually identified by a comprehensive analysis of clinical phenotypes, based on the "four diagnostic methods" (inspection, listening and smelling, inquiring, and palpation) with a certain degree of subjectivity and ambiguity from TCM practitioners [1, 5]. It is the *Zheng* differentiation that determines the TCM diagnosis of a disease, guiding the subsequent TCM treatments such as acupuncture, herbal formulae and so on (known as "*Bian Zheng Lun Zhi*" in Chinese). Interestingly enough, chronic atrophic gastritis (CAG) patients were usually diagnosed with two different TCM syndromes, Pi-qi-deficiency syndrome (PQDS) or Pi-wei damp-heat syndrome (PDHS) [6, 7]. Hence, PQDS and PDHS seem to be two commonly occurring TCM *Zhengs* among CAG patients [8–10]. However, little is known about the biological basis of the TCM-defined PQDS and PDHS resulting from the same disease of CAG, especially the underlying molecular characteristics of the two different TCM *Zhengs*.

In recent years, advances in next-generation sequencing (NGS) technology enable the multiomics analysis of human diseases, notably the discovery of disease-related biomarkers by analyzing tissue or cell-specific transcripts including protein-coding messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs) such as microRNAs (miRNAs), long ncRNAs (lncRNAs) and circular RNAs (circRNAs) [11, 12]. In fact, a lncRNA or circRNA can not only act as a *cis*-acting ncRNA to directly control its neighboring genes in a genome [13, 14], but also function as a competing endogenous RNA (ceRNA) to indirectly regulate target genes [15, 16]. Such ceRNAs indeed act as miRNA sponges/decoys to competitively bind with shared miRNAs through their miRNA response elements (MREs), thus indirectly regulating the gene expression in cytoplasm [15–17]. Accordingly, to reveal the potential cytoplasmic ceRNAs involved in a focused phenotype, it has an important role to construct a global ceRNA regulatory network detailing all the possible ceRNA-gene regulation relationships in a given physiological or pathological condition. Recently, the ceRNA regulatory network-based analysis approach has revealed the potential cytoplasmic lncRNAs or circRNAs related to multiple human diseases including colorectal cancer and gastric cancer of digest system [16, 18, 19]. However, it is rarely reported about the potential lncRNA- or circRNA-associated ceRNA biomarkers for TCM-defined *Zhengs* and diseases, especially the ceRNA regulatory networks probably implicated in maintaining the clinical phenotypes of TCM *Zhengs*.

Hence, in this work, based on an NGS-based RNA sequencing (RNA-seq) approach, using the *control population* (healthy individuals, n = 5), we identified the differentially expressed lncRNAs and circRNAs in the peripheral blood leukocytes of individuals belonging to the two case populations, namely, *case population* 1 (CAG patients with PQDS, n = 5) and *case population* 2 (CAG patients with PDHS, n = 5). Based on the identified sets of differential lncRNAs and circRNAs, excluding the common differential lncRNAs and circRNAs observed in the two case populations, we obtained the *Zheng*-specific lncRNAs and circRNAs (namely, in this study, they were observed only in a certain case population). Especially, we wanted to explore their possible roles in contributing to the functions and characteristics of leukocytes, which might be implicated in the differentiation of the TCM-defined *Zhengs* (PQDS or PDHS) among CAG patients. Thus, we first revealed the potential *Zheng*-specific *cis*-acting lncRNAs or circRNAs in the leukocytes, further decoding their neighboring targets and the potential pathways involved in these targets. In addition, particularly, integrating with the miRNA and gene expression data described in our previous work [20], filtering the lncRNAs and circRNAs incapable of locating into cytoplasm, we ultimately generated several more confident *Zheng*-specific lncRNA- and circRNA-associated ceRNA regulatory networks in leukocytes, including the ceRNA-gene binary relationship networks and the miRNA-centered ceRNA-miRNA-gene triple relationship networks. Also, we investigated the functions and pathways of genes that belonged to the *Zheng*-specific ceRNA-gene pairs in the resultant ceRNA regulatory networks, revealing the potential roles of the ceRNA-regulated genes in contributing to the functions and characteristics of leukocytes in the TCM-defined PQDS and PDHS. Additionally, we specially analyzed the *Zheng*-specific ceRNAs probably encapsuled and carried in exosomes to reveal their potential roles when traveling all over the body. The obtained results may provide new insights into the ceRNA network-mediated gene regulation contributing to characteristic and functional changes of leukocytes in TCM-defined PQDS and PDHS, with potential leukocyte biomarkers for future application in integrative medicine.

Materials And Methods

Ethics approval

The research has been registered at ClinicalTrials.gov (NCT02915393). The protocols have been approved (JDF-IRB-2016031002) by the Institutional Review Board of Dongfang Hospital, Beijing University of Chinese Medicine. All the methods were performed in accordance with the relevant guidelines and regulations. Participants were informed of the purpose, general contents, and data use of the study, and they all signed the informed consent.

Participants

All the subjects (detailed in supplemental Table S1) were recruited at the hepatobiliary and gastroenterological outpatient's department of Dongfang Hospital. The CAG patients were diagnosed and recruited according to the CAG pathological diagnosis and grading standards proposed on the "*Consensus on Chronic Gastritis in China*" (2012, Shanghai) [21]. The diagnosis of TCM syndromes

(PQDS or PDHS) of CAG patients were based on the "*Guiding Principles for Clinical Research of New Chinese Medicines*" published in 2002 [22]. The experimental design and route, including the diagnosis, inclusion and exclusion criteria for all the subjects in this study, are detailed in the supplemental information (Supplemental methods).

Leukocytes

Following overnight fasting, blood samples (5 mL) were obtained from each individual by venipuncture into the additive-free blood collection tubes between 8 $_{AM}$ and 9 $_{AM}$, and the leukocytes were isolated using the lymphocyte separation reagent (Solarbio) according to the manufacture's instruction.

RNA Sequencing

RNA sequencing for the leukocyte samples in this study, were performed by OEbiotech company (Shanghai, China). The detailed descriptions for the NGS-based sequencing are provided in the supplemental methods.

Identification of the differentially expressed lncRNAs and circRNAs

The expression levels of lncRNAs and circRNAs, including the novel ncRNAs discovered in this study, were standardized and respectively indicated using FPKM (fragments per kilobase of exon model per million mapped reads) and RPM (spliced reads per million, 10^6 × number of back-splicing reads mapped to a circRNA/total number of mapped reads). The identification of differentially expressed ncRNAs (lncRNAs and circRNAs) between groups and the *P*-value calculations were performed using the popular R package of DESeq [23]. The differential lncRNAs and circRNAs among groups were filtered (*P*-value < 0.05 & |log2(fold change)| ≥ 1).

Identification of differentially expressed genes and miRNAs

The expression profiles of genes and miRNAs used in this study, including the identification of differential genes and miRNAs, were described in our previous publication [20].

Expression pattern clustering

Hierarchical clustering (HCL)-based expression analyses of differential lncRNAs and circRNAs were performed by the released package termed "pheatmap" in R, and the HCL analysis-based heatmaps were generated.

Predicting targets of cis-acting lncRNAs and circRNAs

The coexpression pairs of lncRNAs/circRNAs and genes in each case population were first identified by calculating the Pearson's correlation coefficients (|*r*| > 0.8 & *P*-value < 0.05) between the expression levels of differential lncRNAs and protein-coding transcripts. Then, an algorithm was adopted to predict target genes for *cis*-acting lncRNAs/circRNAs. Namely, a c*is*-acting lncRNA/circRNA usually controls its

neighboring genes [13, 14], therefore, for each obtained coexpression pair of gene and lncRNA/circRNA, if the gene is close (less than 100 kb) to the lncRNA/circRNA in a genome, it is considered as a candidate target of the lncRNA/circRNA.

Function and pathway enrichment analysis

The gene ontology (GO) function and pathway enrichment analyses of target genes of ncRNAs, were performed using the popular R/Bioconductor packages [24–26]. Particularly, the gene set enrichment analysis (GSEA), a knowledge-based approach for interpreting genome-wide expression profiles [27], were performed for the genes that encoded for the RNA binding proteins (RBPs) of the exosome-encapsuled lncRNAs.

Interaction network analysis

The interaction networks of the corresponding targets of lncRNAs/circRNAs, were created using the well-updated STRING database v11.0 [28]. The obtained interaction networks were further modified and integrated using the popular Cytoscape (package v3.7.1) [29].

Predicting subcellular location of lncRNA and circRNA

The tool of “lncLocator”, a stacked ensemble classifier-based computational predictor, was used to predict the subcellular localization for lncRNA, including cytoplasm, nucleus, cytosol, ribosome, and exosome [30]. Considering that the similar ncRNA sequence often share the same subcellular location, this published tool is also used for predicting subcellular location of circRNAs in this study. The default parameters were adopted to output the confident location prediction (score $\geq$ 0.6) for a specific lncRNA or circRNA.

Identification of the ceRNA-gene relationship pairs

As illustrated (Supplemental Fig. S1), the ceRNAs were identified and the corresponding ceRNA-target interaction networks were carefully generated. First, we selected the possible miRNA-target (lncRNA, circRNA or mRNA) interaction pairs. Briefly, we confirmed the strong negative correlation of expression levels between a miRNA and its target by calculating the Pearson’s correlation coefficients ($r < -0.6$ & *P*-value < 0.05). Also, using the popular miRanda software (v3.3a) [31], we further verified the possible binding of the miRNA and its targets.

Next, following the ceRNA hypothesis [15], we predicted the possible ceRNAs (lncRNA/circRNA)-mRNA relationship pairs among the above-obtained miRNA-target interaction pairs that shared the common miRNAs. In brief, we analyzed the strong positive correlation of expression levels ($r > 0.6$ & *P*-value < 0.05) between lncRNA/circRNA and mRNA, and further calculated the ceRNA score for each ceRNA-mRNA pair (ceRNA score $\geq$ 0.001 & *P*-value < 0.05). ceRNA score = [count of the miRNA response elements (MREs) shared for lncRNA/circRNA and mRNA] / [count of the MREs for lncRNA/circRNA].

Last, the obtained ceRNAs and the corresponding mRNAs and miRNAs were used to generate the final ceRNA-mRNA or ceRNA-miRNA-mRNA relationship networks. The generated networks were further filtered,

modified and presented by the popular Cytoscape tool [29].

Predicting RNA-binding proteins (RBPs) of lncRNAs and circRNAs

The lncRNAs were applied to analyze their possible binding proteins using the web interface of lncSEA, a new comprehensive human lncRNA sets resource and enrichment analysis platform which also integrates RNA binding proteins rooted in large-scale crosslinking-immunprecipitation sequencing (CLIP-Seq) data from StarBase, RNAInter (RNA interactome database) and EuRBPDB (a comprehensive and user-friendly database for eukaryotic RNA binding proteins) [32–35].

Results

Differential lncRNAs and circRNAs in leukocytes

Compared with the *control population* (healthy individuals, n = 5), 365 and 468 differential lncRNAs (Supplemental Tables S2 and 3), as well as 126 and 139 differential circRNAs (Supplemental Tables S4 and 5), were respectively identified in the leukocytes from two case populations, including the CAG patients with PDHS (n = 5) and CAG patients with PQDS (n = 5) (Fig. 1a). In particular, we performed HCL analyses of expression profiles of differential lncRNAs and circRNAs in leukocytes from the three populations (Fig. 1b, c). The generated HCL heatmaps showed that the expression profiles of differential lncRNAs and circRNAs in the leukocytes of individuals belonging to the same population clustered together well, distinguishable from those of other populations.

The Zheng-specific lncRNAs and circRNAs

The *Zheng* (syndrome)-specific lncRNAs and circRNAs in this study mean that they were identified to be differentially expressed in the leukocytes of individuals only with the TCM-defined PQDS or PDHS, excluding their common differential lncRNAs and circRNAs. Namely, the PQDC-specific lncRNAs and circRNAs were observed only in the leukocytes of CAG patients with PQDS rather than PDHS. As presented (Fig. 1a), 385 lncRNAs and 121 circRNAs were PQDS-specific in the leukocytes. Also, 282 lncRNAs and 108 circRNAs were found to be PDHS-specific in the leukocytes.

Targets of the PQDS-specific cis-acting lncRNAs and circRNAs

Total 323 and 124 neighboring genes were found to be corresponding to the differential lncRNAs and circRNAs, respectively. Several of them, including 12 (lncRNA target) and 2 (circRNA target) neighboring genes, were also identified to be differentially expressed in the leukocytes from PQDS population (Fig. 2a). To overview the possible interaction relationships of these obtained neighboring genes, a corresponding protein-protein interaction network was generated. The differential neighboring genes were specially indicated in red, including *NR4A2* (nuclear receptor subfamily 4 group A member 2), *FOS* (Fos proto-oncogene, *AP-1* transcription factor subunit), *HLA-DRB5* (major histocompatibility complex, class II,

DR beta 5), *OSM* (Oncostatin M), *EGF* (epidermal growth factor), *HIP1* (huntingtin interacting protein 1), *AATK* (apoptosis associated tyrosine kinase), *RARRES1* (retinoic acid receptor responder 1), *ARSJ* (arylsulfatase family member J), *LOC100289279*, *COL13A1* (collagen type XIII alpha 1 chain) and *RAB6C* (member RAS oncogene family) (Fig. 2b). Especially, another interaction network was carefully created for the observed 12 target genes. The network not only detailed the interactions between *cis*-acting lncRNAs/circRNAs and their corresponding target genes, including the interactions among target genes (edge width indicates interaction strength of data support), but also presented the expression pattern (shade of blue or red depends on degree of downregulation or upregulation of gene and ncRNA expression) (Fig. 2c). Also, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway-based enrichment analyses were performed to decode the potential pathways of 12 possible target genes, and a network was generated to present the relationships of the enriched pathways of genes. These enriched pathways are involved in helper T (Th)1, Th2 and Th17 cell differentiation, mitogen-activated protein kinase (MAPK) signaling, phosphatidylinositol 3-kinase (PI3K)-serine/threonine-protein kinase (AKT) signaling, janus kinase (JAK)-signal transducer and activator of transcription (STAT) signaling, colorectal cancer, as well as programmed cell death-ligand 1 (PD-L1) expression and PD-1 checkpoint (Fig. 2d). Furthermore, we particularly analyzed the linear correlation of expression levels between the 12 target genes and their neighboring *cis*-acting lncRNAs/circRNAs (Supplemental Fig. S2). A dot plot scattergrams were specially drawn to show the strong negative linear correlation pairs of LOC105378347-*COL13A1* (Fig. 2e).

Targets of the PDHS-specific cis-acting lncRNAs and circRNA

Total 269 and 115 neighboring genes were corresponding to the differential lncRNAs and circRNAs, respectively. Seven neighboring genes of 12 lncRNAs belonged to the differential genes identified in the leukocytes from the PDHS population, including *EGR3* (early growth response 3), *NR4A2* (nuclear receptor subfamily 4 group A member 2), *LOC102724843*, *RCAN2* (regulator of calcineurin 2), *GPM6A* (glycoprotein M6A), *KLHL14* (kelch like family member 14) and *VAMP7* (vesicle associated membrane protein 7) (Fig. 3a). An interaction network was created to overview the possible relationships among all the neighboring genes of the differential lncRNAs and circRNAs. For the obtained seven neighboring genes, they were considered as the candidate targets of the *cis*-acting lncRNAs, and were specially denoted in red to facilitate manual check in the resultant network (Fig. 3b). Another network was also carefully drawn to detail the possible regulation relationships between the *cis*-acting lncRNAs and their corresponding neighboring genes, and the expression patterns were also presented in the generated network (Fig. 3c). The reactome and KEGG pathway-based enrichment analyses of the seven candidate target genes revealed the potential pathways related to SUMOylation, vesicle biogenesis, nuclear receptor transcription, interleukin (IL)-12 signaling and C-type lectin receptor signaling. A network was created to show the relationships of the enriched pathways (Fig. 3d). We also checked the linear correlation of expression levels between the seven candidate targets and their neighboring *cis*-acting lncRNAs (Supplemental Fig. S3). Notably, two strong linear correlation pairs of *cis*-acting lncRNAs and targets, LOC107984113-*GPM6A* and LOC105372055-*KLHL14*, were found in the leukocytes from the PDHS population (Fig. 3e).

The Zheng-specific ceRNA-gene binary relationship networks

Several classes of ncRNA molecules, including lncRNA and circRNA, could counteract the gene repressive activity of miRNA by sequestering miRNA within a cell. Such lncRNAs and circRNAs were described as miRNA sponges/decoys, target mimics (in plants) or competing endogenous RNA (ceRNA) (in mammals) [15, 17]. As indicated (Supplemental Fig. S1), based on the expression profile data covering the lncRNAs, circRNAs, genes and miRNAs, several original networks were carefully generated to detail the ceRNA (lncRNA/circRNA)-gene relationship pairs in the leukocytes from the PQDS and PDHS populations. The observed relationship pairs in the original networks were further filtered using the above-mentioned *Zheng*-specific lncRNAs, circRNAs and genes to get the *Zheng*-specific ceRNA-gene relationship networks. Especially, considering that ceRNA could competitively bind with shared miRNAs through miRNA response elements (MREs), indirectly regulating gene expression in cytoplasm, we also eliminated the ceRNAs predicted to be incapable of locating into cytosol/cytoplasm. Therefore, the resultant ceRNA networks could hold more confident ceRNA-gene relationship pairs to detail the *Zheng*-specific ceRNA-gene regulation relationships in the leukocytes (Fig. 4a and c).

As shown (Fig. 4a), an integrated binary ceRNA network, was particularly generated to detail the PQDS- and PDHS-specific lncRNA-gene relationship pairs in the leukocytes (Supplemental Tables S6 and 7), including the common lncRNAs and genes implicated in the ceRNA-gene pairs. In the lncRNA-associated ceRNA network, overall, almost each lncRNA could compete with multiple targets and even additional common genes. For instance, each of these PQDS-specific lncRNAs could compete with multiple genes (more than ten), including ENST00000447519 (*AP001189.1*), XR_001744971.1 (*LOC105375433*), ENST00000433310 (AF165147.1), XR_926748.2 (*LOC105375035*), XR_001752213.1 (*LOC107984889*), XR_001739495.1 (*LOC105374768*), XR_945327.3 (*LOC105369968*), NR_120454.1 (*LINC02449*), NR_109859.1 (*LINC01730*), ENST00000635072 (*LINC01358*), ENST00000433669 (*LINC00954*) and XR_001746085.1 (*LOC105375754*). Also, each of these PDHS-specific lncRNAs containing NR_027412.1 (*LINC00910*), ENST00000451362 (*AL139130.1*), XR_944918.2 (*LOC102725258*), XR_001746915.1 (*LOC105376244*) and XR_001739484.1 (*LOC105377632*), were capable of competing with multiple targets (over four). Notably, six exosome-encapsuled lncRNAs were found and marked in the obtained lncRNA-gene network, including ENST00000416824 (*LINC00957*), NR_036658.2 (*ZFAS1*), NR_024061.1 (*LOH12CR2*), ENST00000566847 (*AL353719*.1) and two additional novel lncRNAs named TCONS_00038035 and TCONS_00027600 (Fig. 4a). Similarly, another integrated circRNA-associated ceRNA network was drawn to present the PQDS- and PDHS-specific circRNA-gene regulation relationship pairs in the leukocytes (Fig. 4c; Supplemental Tables S8 and 9). Obviously, each of these PQDS-specific circRNAs competed with multiple genes (over six), including circRNA_05342, circRNA_17558 (*hsa_circ_0003910*), circRNA_20312, circRNA_26206 (*hsa_circ_0002483*), circRNA_22470, circRNA_01462, circRNA_12753 and circRNA_11374 (*hsa_circ_0043114*). But most of the PDHS-specific circRNA competed only with one gene, and only the PDHS-specific circRNA_20312 (*hsa_circ_0001440*) could compete with more than two PDHS-specific genes (Fig. 4c). In addition, two corresponding heatmaps were generated to profile the expression of the observed lncRNA and circRNAs involved in the *Zheng*-specific ceRNA-gene relationship pairs in the leukocytes. Particularly, two special heatmap were

also created to show the possibility that the obtained Z*heng*-specific ceRNAs could locate in cytoplasm, cytosol, ribosome, nucleus and exosome (Fig. 4b, d). These *Zheng*-specific ceRNA-gene pairs identified in the different leukocytes suggested their potential roles in regulating the *Zheng*-specific gene expression profiles, which might contribute to the characteristics and functions of leukocytes in the TCM-defined PQDS and PDHS resulting from CAG.

The Zheng-specific ceRNA-miRNA-gene triple relationship networks

In each ceRNA (lncRNA/circRNA)-gene relationship pair, ceRNA competitively binds with the shared miRNAs by the MREs, thus enabling the indirect regulation of the gene expression. To decode the shared miRNAs, especially the key hub miRNAs implicated in multiple ceRNA-gene pairs in the leukocytes from the PQDS and PDHS populations, several ceRNA-miRNA-gene triple relationship networks were also carefully generated using the above-obtained *Zheng*-specific ceRNA-gene pairs and their corresponding shared miRNAs (Fig. 5; Supplemental Tables S6 to 9). In these miRNA-centered ceRNA regulatory networks (Fig. 5), all the shared miRNAs involved in the *Zheng*-specific ceRNA-gene pairs were visually presented, including the *Zheng*-specific miRNAs as well as the common miRNAs identified in the leukocytes from the PQDS and PDHS populations.

As show (Fig. 5a, b), obviously, most of the PQDS-specific miRNAs were involved in at least three PQDS-specific ceRNA-gene pairs. The PQDS-specific miRNAs corresponding to both lncRNA-gene pairs and lncRNA-gene pairs, were particularly labeled in bold font (the blue font denoted the known miRNAs among them) to facilitate manual check. Such key hub miRNAs implicated in multiple (at least four) PQDS-specific ceRNA-gene pairs, contained hsa-miR-6873-3p, hsa-miR-6509-3p, hsa-miR-136-5p, hsa-miR-1260b, hsa-miR-7110-3p, hsa-miR-4701-5p and several novel miRNAs discovered in this study (Fig. 5a, b). Besides, concern the PDHS-specific hub miRNAs involved in multiple ceRNA-gene pairs, each of the miRNAs of hsa-miR-6855-5p and has-miR-6855-5p could correspond to six PDHS-specific lncRNA-gene pairs, and the hsa-miR-133a-5p was involved in two PDHS-specific lncRNA-gene pairs (Fig. 5c). Also, three novel miRNAs labeled in bold were implicated be not only lncRNA-gene pairs but also circRNA-gene pairs (Fig. 5c, d). These observed *Zheng*-specific miRNAs in the ceRNA networks, especially the key hub miRNAs linking to multiple *Zheng*-specific ceRNA-gene relationship pairs, indicated their important roles in the *Zheng*-specific gene expression regulation, suggesting a class of potential biomarkers in leukocytes for differentiating TCM-defined PQDS and PDHS resulting from CAG.

Interaction networks and functions of Zheng-specific ceRNA-regulated genes

In order to detail the possible interactions between each gene belonging to the *Zheng*-specific ceRNA-gene pairs, two interaction complex networks were carefully generated, integrated with any type of edge and node attribute data such as gene expression pattern and the number of regulating ceRNAs (Fig. 6). The resultant networks not only show the possible interactions between each *Zheng*-specific node gene (edge thickness indicates interaction strength of data support) (Supplemental Tables S10 and 11), but also present the expression pattern of each node gene (node shade of blue or red relies on degree of down-regulation or upregulation of gene expression). Node size depends on the number of *Zheng*-specific

ceRNAs (lncRNAs/circRNAs) targeting to the corresponding node gene, and the number is also displayed near the corresponding gene node in format of “lncRNAs number/lncRNAs number”. Moreover, the pathway and GO function enrichment analyses were also performed to reveal the potential functions and pathways of these node genes governed by the *Zheng*-specific ceRNAs (Supplemental Figs. S4 and 5). Several enriched pathways and function terms were also notably highlighted and annotated in the resultant interaction networks. As presented (Fig. 6a), overall, most of the PQDS-specific node genes could be governed by more than three PQDS-specific ceRNAs. Although many node genes had no observed interactions with each other (no edge between each other), they were also found to be regulated by multiple PQDS-specific ceRNAs. The node genes keeping more complex relationships with each other, were enriched in the pathways involved in MAPK signaling, receptor tyrosine kinases signaling as well as complement and coagulation cascades. Also, the enriched pathways related to adherens junction, focal adhesion, ECM organization, ECM-receptor interaction and cell surface interactions at the vascular wall, were implicated in cell-to-cell adhesion/junction and communication, probably contributing to the characteristics and functions of leukocytes in TCM-defined PQDS resulting from CAG. Notably, the common node genes involved in multiple crosstalks between these enriched pathways, including *EGF* (epidermal growth factor), *COL4A2* (collagen, type IV, alpha 2 chain), *COL4A4* (collagen, type IV, alpha 4 chain), *PROCR* (protein C receptor) and *PROS1* (protein S), seemed to undergo a more complex regulation of the PQDS-specific lncRNAs and circRNAs in the leukocytes.

In addition, for the genes belonging to the PDHS-specific ceRNA-gene pairs, another network was carefully created to present the possible interactions between them (Fig. 6b). As shown, except the interaction of *ARHGEF4* (rho guanine nucleotide exchange factor 4) and *CHN1* (chimerin 1), no additional interaction relationships were observed among these genes governed by the PDHS-specific ceRNAs. *ARHGEF4* and *CHN1* were enriched in the biological processes such as regulation of cell morphogenesis, Rho GTPase cycle and regulation of small GTPase-mediated signal transduction. Especially, *CR2* (complement component receptor 2) and *ULBP3* (UL16-binding protein 3) were related to lymphocyte mediated immunity, including complement and coagulation cascades, B-cell receptor signaling and regulation of NK/T-cell mediated cytotoxicity. *CEACAM6* (carcinoembryonic antigen related cell adhesion molecule 6) were involved in neutrophil activation and degranulation, cell-matrix/cell adhesion and cell migration. *GNAI1* (G protein subunit alpha i1) was associated with regulation of protein localization to cell periphery. *VWDE* (von Willebrand factor D and EGF domain-containing protein) was associated regulation of canonical Wnt signaling.

Potential function analyses of the exosome-encapsuled ceRNAs

The five previously-mentioned PQDS-specific ceRNAs (lncRNAs), including NR_036658.2 (*ZFAS1*), ENST00000566847 (*AL353719.1*), NR_024061.1 (*LOH12CR2*) and two new lncRNAs named TCONS_00038035 and TCONS_00027600, were capable of being encapsuled and carried in exosomes (Figs. 4a, 7a), especially keeping higher levels in the leukocytes from the PQDS population (Fig. 7c). The leukocyte-derived exosomes could transfer these lncRNAs into other far away recipient cells, thus making

them function throughout the body. In order to further decode their potential roles, we not only drawn an interaction network to detail their corresponding ceRNA-gene pairs (Fig. 7a), but also particularly retrieved their experimentally-validated RNA binding proteins (RBPs) using the published "lncSEA" interface, a new comprehensive human lncRNA sets resource and enrichment analysis platform [32]. Also, another protein-protein interaction network was carefully generated to overview the possible interaction relationships between each of the obtained RBPs, and the three PQDS-specific genes-encoded RBPs were specially marked in the resultant network, such as NRXN1 (neurexin 1), LTK (leukocyte receptor tyrosine kinase) and ALK (anaplastic lymphoma receptor tyrosine kinase) (Fig. 7b). Interestingly, the gene encoding ALK was also observed to be governed by two PQDS-specific ceRNAs (lncRNAs), including NR_036658.2 (*ZFAS1*) and ENST00000566847 (*AL353719.1*) (Fig. 7a). Moreover, to further reveal the potential pathways related to these obtained RBPs, the gene set enrichment analysis (GSEA), a knowledge-based approach for interpreting genome-wide expression profiles [27], were particularly performed using the expression profiles of genes coding all the obtained RBPs in the leukocytes. The GSEA results revealed several potential pathways related to cancer and immunity, especially the pathways implicated in cell-to-cell adhesion/junction and communication which probably contributed to the characteristics and functions of leukocytes in TCM-defined PQDS resulting from CAG (Fig. 7c; Supplemental Table S12).

Discussion

In TCM, *Zheng* is a thousand-year-old key diagnostic concept, considered as the TCM theory-based interpretation of the symptom profiles of a disease, but not only a simple assemblage of disease symptoms [1–4]. In TCM practice, *Zheng* differentiation was used to classify patients, even the patients with the same disease, guiding the subsequent TCM treatments [1, 3, 36]. Accordingly, decoding the possible biological basis of TCM-defined *Zheng*, is a critical step in understanding TCM, especially the modernization of TCM. Interestingly, PQDS and PDHS seemed to two commonly occurring *Zhengs* among the CAG patients [8–10]. Leukocytes, as the important immune cells throughout the body, play very crucial roles in host defense and contribute to pathogenesis of various immune diseases. Thus, the changes in characteristics and functions of leukocytes may be implicated in the differentiation of the two TCM-defined *Zhengs* resulting from the same disease of CAG. Based on the high throughput sequencing identification of lncRNAs and circRNAs, integrating with the expression data of miRNAs and genes reported in our previous work [20], we especially wanted to reveal the *Zheng*-specific ceRNAs (lncRNAs/circRNAs) and generated their corresponding ceRNA regulatory networks, further decoding the *Zheng*-specific ceRNA-gene regulation underlying the leukocyte characteristics and functions in the two TCM-defined PQDS and PDHS resulting from CAG, in particular the potential ceRNA biomarker candidates in the corresponding leukocytes.

Compared with the healthy control, the *Zheng*-specific lncRNAs and circRNAs were identified in the leukocytes. An interesting result was observed that despite being the TCM-defined resultant *Zhengs* resulting from the same disease of CAG, the PQDS- or PDHS-specific lncRNAs and circRNA were not same (Fig. 1a). Because gene expression determines cell's characteristics, the *Zheng*-specific

lncRNA/circRNA-governed gene expression in leukocytes may cause the specific alterations in the characteristics and functions of leukocytes in the two TCM-defined *Zhengs*. Actually, a lncRNA or circRNA could not only act as a *cis*-acting ncRNA to directly control its neighboring genes in a genome [13, 14], but also function as a ceRNA to indirectly regulate target genes [15, 16]. Thereby, we first revealed the potential *Zheng*-specific *cis*-acting lncRNAs/circRNAs in leukocytes, decoding their neighboring targets and the potential pathways involved in these targets. Concerning the obtained 14 PQDS-specific *cis*-acting lncRNAs/circRNAs in leukocytes, they contained 12 lncRNAs and two circRNAs, capable of directly governing the 12 differential neighboring genes such as *NR4A2*, *FOS*, *HLA-DRB5*, *OSM*, *EGF*, *HIP1*, *AATK*, *RARRES1*, *ARSJ*, *LOC100289279*, *COL13A1* and *RAB6C* (Fig. 2a-c). These governed genes were enriched in the pathways associated with Th cell differentiation, MAPK signaling, PI3K-AKT and JAK-STAT signaling, colorectal cancer as well as PD-L1 expression and PD-1 checkpoint (Fig. 2c and d). Notably, a strong negative linear correlation pairs, namely, LOC105378347-*COL13A1* (r = - 0.95), were revealed by correlation analysis of expression level between a PQDS-specific *cis*-acting lncRNA/circRNA and its neighboring genes (Fig. 2e). In addition, regarding the 12 PDHS-specific *cis*-acting lncRNAs, they could direct regulate their neighboring genes, including *EGR3*, *NR4A2*, *LOC102724843*, *RCAN2*, *GPM6A*, *KLHL14* and *VAMP7* (Fig. 3a-c). The reactome and KEGG pathway-based enrichment analyses of these targets revealed the potential pathways related to SUMOylation, vesicle biogenesis, nuclear receptor transcription, IL-12 signaling and C-type lectin receptor signaling (Fig. 3d). Especially, two strong linear correlation pairs of the PDHS-specific *cis*-acting lncRNAs and the corresponding neighboring genes were found in the leukocytes, including LOC107984113-*GPM6A* (r = 0.879) and LOC105372055-*KLHL14* (r = 0.875) (Fig. 3e). Overall, these obtained results indicate the *Zheng*-specific *cis*-acting lncRNAs and circRNAs in different leukocytes, suggesting their potential roles in contributing to the functions and characteristics of leukocytes in the TCM-defined PQDS and PDHS resulting from CAG.

In addition, a lncRNA or circRNA can also function as a ceRNA to indirectly regulate target genes in cytoplasm. Such ceRNAs, indeed acting as miRNA sponges/decoys, counteract the gene repressive activity of miRNA by sequestering miRNA in cytoplasm [15–17]. We carefully revealed the *Zheng*-specific ceRNAs, especially filtering the ceRNA candidates that were incapable of locating into cytoplasm, and ultimately generated the corresponding *Zheng*-specific lncRNA- and circRNA-associated ceRNA regulatory networks in leukocytes (Figs. 4 and 5). Thus, the specially integrated ceRNA-gene binary relationship networks hold more confident ceRNA-gene relationship pairs to detail the PQDS- or PDHS-specific ceRNA-gene regulation relationships in the leukocytes, including the common lncRNAs, circRNAs and genes involved in the ceRNA-gene pairs (Fig. 4a and c). Such resultant networks could help to directly track the key ceRNAs (competing with multiple genes) and targets (competing with multiple ceRNAs), facilitating the discovery of the potential *Zheng*-specific ceRNA biomarkers in leukocytes implicated in the TCM-defined PQDS and PDHS. Furthermore, in the miRNA-centered ceRNA-miRNA-gene triple relationship networks (Fig. 5), all the shared miRNAs involved in the *Zheng*-specific ceRNA-gene pairs were visually presented, including the *Zheng*-specific miRNAs as well as the common miRNAs identified in leukocytes. The *Zheng*-specific miRNAs in such ceRNA networks, especially the key hub miRNAs linking to multiple *Zheng*-specific ceRNA-gene relationship pairs, indicated their important roles in the *Zheng*-specific gene

expression regulation. All together, these potential ceRNA and miRNA biomarkers, particularly the generated *Zheng*-specific ceRNA regulatory networks, suggested their potential roles in regulating the *Zheng*-specific gene expression profiles, probably contributing to the clinical characteristics and functions of leukocytes in TCM-defined PQDS and PDHS resulting from CAG.

Moreover, concerning the decoded potential functions and pathways implicated in the *Zheng*-specific ceRNA regulatory networks, as vividly presented (Fig. 6a), the ceRNA-governed genes keeping more complex relationships with each other, were enriched in the pathways involved in MAPK signaling, receptor tyrosine kinases signaling as well as complement and coagulation cascades. The pathways related to adherens junction, focal adhesion/ECM-receptor interaction, ECM-organization and cell surface interactions at the vascular wall, were implicated in cell-to-cell adhesion/junction and communication, contributing to the characteristics and functions of leukocytes in the TCM-defined PQDS. Notably, the common node genes involved in multiple crosstalks between these enriched pathways, including *EGF*, *COL4A2*, *COL4A4*, *PROCR* and *PROS1*, seemed to undergo a tight and complex regulation of more PQDS-specific ceRNAs in the leukocytes. However, for the PDHS-specific ceRNA-governed genes (Fig. 6b), except the interactions of *ARHGEF4* and *CHN1*, no additional interaction relationships were observed. They corresponding biological processes were involved in regulation of cell morphogenesis, Rho GTPase cycle and small GTPase-mediated signal transduction. Especially, *CR2* and *ULBP3* were related to lymphocyte mediated immunity, including complement and coagulation cascades, B-cell receptor signaling and regulation of NK/T-cell mediated cytotoxicity. The *CEACAM6* was involved in neutrophil activation and degranulation, cell-matrix/cell adhesion and cell migration. The *GNAI1* and VWDW were related to regulation of protein localization to cell periphery and canonical Wnt signaling. All above-mentioned results indicated that the ceRNA-governed genes in the resultant *Zheng*-specific ceRNA networks could implicate different biological processes and pathways, possibly leading to the alterations in functions and characteristics of leukocytes in TCM-defined PQDS and PDHS resulting from CAG.

In particular, five PQDS-specific ceRNAs (lncRNAs) were capable of being encapsuled and carried in exosomes (Figs. 4a, 7a), including NR_036658.2 (*ZFAS1*), ENST00000566847 (*AL353719.1*), NR_024061.1 (*LOH12CR2*) and additional two new lncRNAs named TCONS_00038035 and TCONS_00027600. Importantly, they kept higher transcription levels in the leukocytes of individuals from the PQDS rather than PDHS population (Fig. 7c), suggesting they seemed to be the potential exosome-carried ceRNA biomarkers for differentiation of the TCM-defined PQDS and PDHS among CAG patients. The leukocyte-derived exosomes could transfer these ceRNAs into other far away recipient cells, making them function all over the body. Thus, to decode their potential roles, we not only detailed their governed target genes, but also retrieved their experimentally-validated RBPs (Fig. 7a and b). Notably, of them, three PQDS-specific genes-encoded RBPs were also found such as ALK, LTK and NRXN1. Interestingly more, the gene encoding ALK could be governed by two PQDS-specific lncRNAs (ceRNAs), including NR_036658.2 and ENST00000566847 (Fig. 7a). The GSEA-based analysis of these retrieved RBPs also revealed several potential pathways related to cancer and immunity, especially the pathways implicated in cell-to-cell adhesion/junction and communication, probably contributing to the characteristics and functions of leukocytes in TCM-defined PQDS (Fig. 7c).

We revealed the *Zheng*-specific *cis*-acting lncRNAs/circRNAs in leukocytes, especially the *Zheng*-specific ceRNA networks-mediated gene regulation, contributing to the alterations in characteristics and functions of leukocytes in TCM-defined PQDS and PDHS. However, there are several limitations in the present study. Although PQDS and PDHS seemed to be the commonly occurring *Zhengs* among CAG patients, there were other three resultant *Zhengs* resulting from the same disease of CAG [7, 8]. Thus, the observed specific ceRNAs and expression patterns to differentiate the resultant PQDS and PDHS, were unlikely to be suitable for differentiating them from other resultant *Zhengs* of CAG. Besides, the study population was not large enough to draw the definitive conclusions, and there were significant differences in age between the health control group and CAG patients, which may lead to biased conclusions. Hence, further studies involving more resultant *Zhengs* of CAG and larger sample sizes, are needed to strengthen the conclusions of this study.

Conclusions

Despite being the two TCM-defined *Zhengs* resulting from the same disease of CAG, there seemed to be different alterations in characteristics and functions of leukocytes in the PQDS and PDHS. The *Zheng*-specific *cis*-acting lncRNAs/circRNAs, especially the *Zheng*-specific ceRNA regulatory networks, played potential roles in the regulation of the *Zheng*-specific gene expression profiles, probably contributing to the changes in characteristics and functions of leukocytes in the PQDS and PDHS. Importantly, five exosome-encapsuled ceRNAs, particularly higher expression in the leukocytes in the PQDS rather than PDHS, could be potential ceRNA biomarkers for differentiation of the TCM-defined PQDS and PDHS among CAG patients. These results may provide new insights into the characteristic and functional changes of leukocytes in the two TCM *Zhengs*, especially the *Zheng*-specific ceRNA-mediated gene regulation underlying leukocyte characteristics and functions, with potential leukocyte biomarkers for future application in integrative medicine.

Abbreviations

AATK: Apoptosis associated tyrosine kinase

AKT: Serine/threonine-protein kinase

ALK: Anaplastic lymphoma receptor tyrosine kinase

ARHGEF4: Rho guanine nucleotide exchange factor 4

ARSJ: Arylsulfatase family member J

CAG: Chronic atrophic gastritis

CEACAM6: Carcinoembryonic antigen related cell adhesion molecule 6

ceRNA: Competing endogenous RNA

CHN1: Chimerin 1

circRNA: Circular RNA

CLIP-Seq: Crosslinking-immunprecipitation sequencing

COL13A1: Collagen type XIII alpha 1 chain

COL4A2: Collagen, type IV, alpha 2 chain

COL4A4: Collagen, type IV, alpha 4 chain

CR2: Complement component receptor 2

ECM: Extracellular matrix

EGF: Epidermal growth factor

EGR3: Early growth response 3

FOS: Fos proto-oncogene AP-1 transcription factor subunit

FPKM: Fragments per kilobase of exon model per million mapped reads

GNAI1: G protein subunit alpha i1

GO: Gene ontology

GPM6A: Glycoprotein M6A

GSEA: Gene set enrichment analysis

HCL: Hierarchical clustering

HIP1: Huntingtin interacting protein 1

HLA-DRB5: Major histocompatibility complex, class II, DR beta 5

IL: Interleukin

JAK: Janus kinase

KLHL14: Kelch like family member 14

LTK: Leukocyte receptor tyrosine kinase

MAPK: Mitogen-activated protein kinase

miRNA: MicroRNAs

MRE: MicroRNA response elements

mRNA: Messenger RNA

ncRNA: non-coding RNA

NGS: Next-generation sequencing

NK: Natural killer cell

NR4A2: Nuclear receptor subfamily 4 group A member 2

NRXN1: Neurexin 1

OSM: Oncostatin M

PDHS: Pi-wei damp-heat syndrome

PD-L1: Programmed cell death-ligand 1 (PD-L1)

PI3K: Phosphatidylinositol 3-kinase

PQDS: Pi-qi-deficiency syndrome

PROCR: Protein C receptor

PROS1: Protein S

RAB6C: Member RAS oncogene family

RARRES1: Retinoic acid receptor responder 1

RBP: RNA binding protein

RCAN2: Regulator of calcineurin 2

RNA-seq: RNA sequencing

RPM: Spliced reads per million

STAT: Signal transducer and activator of transcription

TCM: Traditional Chinese medicine

Th: Helper T-cell

ULBP3: UL16-binding protein 3

VAMP7: Vesicle associated membrane protein 7

VWDE: Von Willebrand factor D and EGF domain-containing protein

ZFAS1: ZNFX1 antisense RNA 1

Declarations

Acknowledgements

We thank the members of our laboratory for discussion, especially the professors Junxiang Li and Jianxin Chen for his scientific advices to this study.

Authors' contribution

Anlong Xu conceived the study. Leiming You, Wei Wang, Xiaopu Sang and Anlong Xu designed the research; Wei Wang, Xiaopu Sang, Ting'an Li, Kunyu Li and Xinhui Gao performed the experiments; Leiming You, Wei Wang, Xiaopu Sang, Jiarui Wu, Guangrui Huang and Ting Wang analyzed the data; Leiming You and Wei Wang designed the figures and wrote the paper. Anlong Xu wrote and edited the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (nos. 81430099 and 91231206 to A. X.).

Data availability

All sequence data have been deposited in GenBank under BioProject accession number PRJNA591186. The RNA-seq and miRNA-seq reads are deposited in the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) under accession numbers (SRR10513202 to SRR10513204, SRR10513208, SRR10513209, SRR11548311 to SRR11548319 and SRR11548330).

Ethics approval and consent to participate

The study was registered at ClinicalTrials.gov (NCT02915393). The protocol was approved (JDF-IRB-2016031002) by the Institutional Review Board of Dongfang Hospital affiliated to Beijing University of Chinese Medicine. All the methods were performed in accordance with the relevant guidelines and regulations. Participants were informed of the purpose, general contents and data use of the study, and they all signed the informed consent.

Consent for publication

Participants were informed of the purpose, general contents, and data use of the study, and they all signed the informed consent.

Competing interests

The authors declare no conflict of interests regarding the publication of this paper.

References

1. Cheung F. TCM: Made in China. Nature. 2011;480(7378):82–3.
2. Cheng F, Wang X, Song W, Lu Y, Li X, Zhang H, et al. Biologic basis of TCM syndromes and the standardization of syndrome classification. Journal of Traditional Chinese Medical Sciences. 2014;1(2):92–7.
3. Wang Y, Xu A. Zheng. A systems biology approach to diagnosis and treatments. Science. 2014;346(6216):13–5.
4. Wang T, Dong J. What is “zheng” in traditional Chinese medicine? Journal of Traditional Chinese Medical Sciences. 2017;4(1):14–5.
5. Song Y-N, Zhang G-B, Zhang Y-Y, Su S-B. Clinical Applications of Omics Technologies on ZHENG Differentiation Research in Traditional Chinese Medicine. Evid-Based Compl Alt. 2013;2013:989618.
6. Tang XD, Lu B, Zhou LY, Zhan SY, Li ZH, Li BS, et al. Clinical practice guideline of Chinese medicine for chronic gastritis. Chin J Integr Med. 2012;18(1):56–71.
7. Zhang SS, Tang XD, Huang HP, Bian LQ. Expert consensus on diagnosis and treatment of Traditional Chinese Medicine for chronic gastritis (2017). China Journal of Traditional Chinese Medicine Pharmacy. 2017;32(07):3060–4. **(Article in Chinese)**.
8. Cao ZJ, Zuo MH. Study on Relationship between Traditional Chinese Medicine Syndrome Distribution, Gastroscopy and Pathology in Chronic Atrophic Gastritis. WORLD CHINESE MEDICINE. 2018;13(8):2040–3.
9. Hu L, Zheng XF, Yan XH. Expressions of HSP 70 and NF-kappaB in the peripheral blood lymphocyte of chronic gastritis patients of different syndrome patterns. Chinese journal of integrated traditional Western medicine. 2012;32(9):1188–91. **(Article in Chinese)**.
10. Liang JK, Hu L, Zheng XF. Study of Th1/Th2 balance in peripheral blood of chronic gastritis patients with Pi-Wei damp-heat syndrome. Chinese journal of integrated traditional Western medicine. 2012;32(3):322–4. 328 **(Article in Chinese)**.
11. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014;157(1):77–94.
12. George J, Patel T. Noncoding RNA as therapeutic targets for hepatocellular carcinoma. Semin Liver Dis. 2015;35(1):63–74.

13. Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell. 2010;143(1):46–58.
14. Ponjavic J, Oliver PL, Lunter G, Ponting CP. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet. 2009;5(8):e1000617.
15. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA. language? Cell. 2011;146(3):353–8.
16. Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell. 2011;147(2):344–57.
17. Gupta P. Competing endogenous RNA (ceRNA): a new class of RNA working as miRNA sponges. Curr Sci. 2014;106:823–30.
18. Xia T, Liao Q, Jiang X, Shao Y, Xiao B, Xi Y, et al. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. Scientific reports. 2014;4:6088.
19. Zhong Y, Du Y, Yang X, Mo Y, Fan C, Xiong F, et al. Circular RNAs function as ceRNAs to regulate and control human cancer progression. Mol Cancer. 2018;17(1):79.
20. You L, Zhang S, Li T, Sang X, Li K, Wang W, et al. Integrated analyses of miRNA and mRNA profiles in leukocytes and serums in traditional Chinese medicine (TCM)-defined Pi-qi-deficiency syndrome and Pi-wei damp-heat syndrome resulting from chronic atrophic gastritis. Chinese Medicine. 2021;16(1):4.
21. Fang JY, Liu WZ, Li ZK, Du YQ, Ji XL, Ge ZZ, et al. China Chronic Gastritis Consensus (2012, Shanghai). Chin J Front Med Sci. 2013;5(07):44–55. **(Article in Chinese)**.
22. Zheng XY. Guiding Principle for Clinical Research on New Drugs of Traditional Chinese Medicine. Beijing: China Medical Science Press; 2002.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.
24. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284–7.
25. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Molecular bioSystems. 2016;12(2):477–9.
26. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29(14):1830–1.
27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.
28. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2019;47(D1):D607–13.

29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498–504.
30. Cao Z, Pan X, Yang Y, Huang Y, Shen HB. The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. Bioinformatics. 2018;34(13):2185–94.
31. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome biology. 2010;11(8).
32. Chen J, Zhang J, Gao Y, Li Y, Feng C, Song C, et al. LncSEA: a platform for long non-coding RNA related sets and enrichment analysis. Nucleic acids research. 2021;49(D1):D969–80.
33. Lin Y, Liu T, Cui T, Wang Z, Zhang Y, Tan P, et al. RNAInter in 2020: RNA interactome repository with increased coverage and annotation. Nucleic acids research. 2020;48(D1):D189–97.
34. Liao J-Y, Yang B, Zhang Y-C, Wang X-J, Ye Y, Peng J-W, et al. EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). Nucleic acids research. 2020;48(D1):D307–13.
35. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic acids research. 2014;42(Database issue):D92–7.
36. Lu A, Jiang M, Zhang C, Chan K. An integrative approach of linking traditional Chinese medicine pattern classification and biomedicine diagnosis. J Ethnopharmacol. 2012;141(2):549–56.

Figures

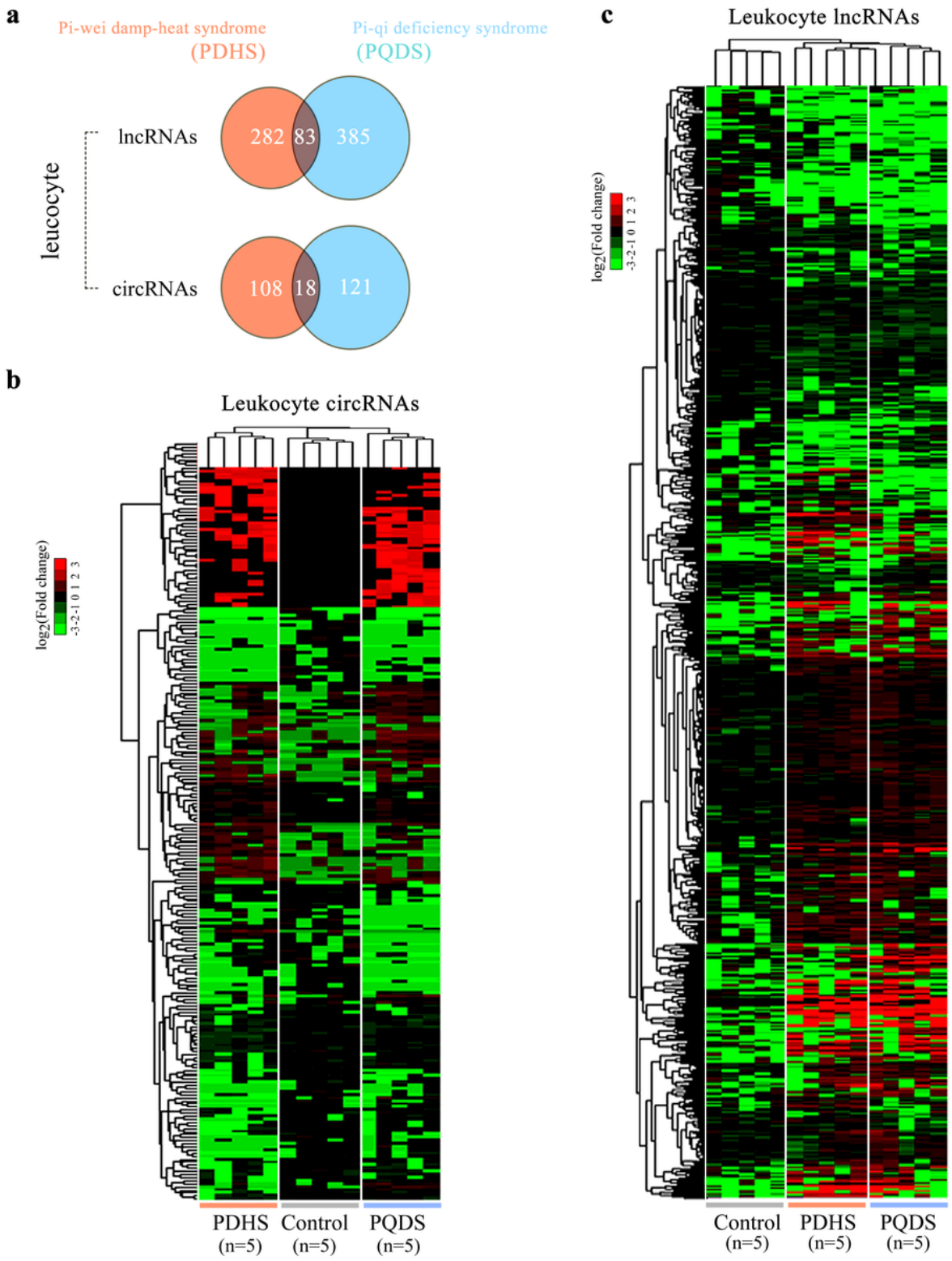

Figure 1

Expression pattern clustering analyses of differential lncRNAs and circRNAs identified in leukocytes from the PDHS and PQDS populations. (a) Venn diagrams for the differential lncRNAs and circRNAs in leukocytes. (b) Hierarchical clustering (HCL) analyses of expression profiles of the differential circRNAs in leukocytes. (c) HCL analyses of expression profiles of the differential lncRNAs in leukocytes.

Abbreviations: Control: healthy individuals; PQDS: chronic atrophic gastritis patients with Pi-qi-deficiency syndrome; PDHS: chronic atrophic gastritis patients with Pi-wei damp-heat syndrome.

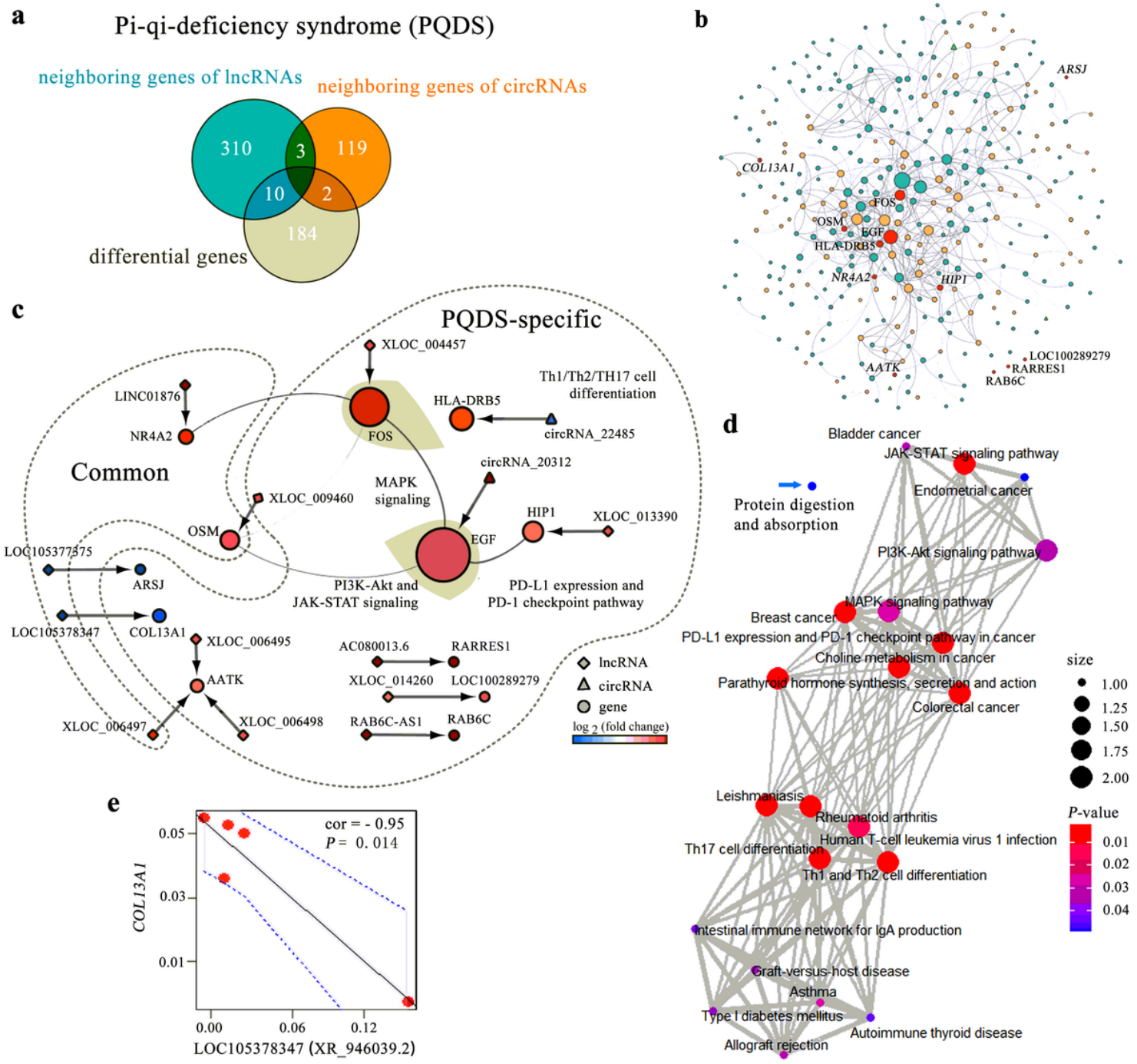

Figure 2

Function and pathway analyses of the targets of the PQDS-specific cis-acting lncRNAs/circRNAs in leukocytes. (a) Venn diagrams for the differentially expressed genes and the neighboring genes of differential lncRNAs/circRNAs in the leukocytes. (b) Overview of the interaction network of the neighboring genes of differential lncRNAs/circRNAs. Node size depends on the number of edges. Orange nodes represent the neighboring genes of differential circRNAs, greenish-blue nodes represent the neighboring genes of differential lncRNAs, and red nodes denote the 12 neighboring genes identified to

be differentially expressed in the leukocytes. (c) The interaction network created to detail the interactions among the obtained 12 targets of the 14 cis-acting lncRNAs/circRNAs. The expression levels of lncRNAs, circRNAs and genes were visually presented in the form of colorful nodes showing the color change (blue to red) in brightness and chromaticity. (d) The relationship network of the enriched pathways of the 12 target genes. A node labeled with a pathway term indicates an enriched pathway of targets, and the node size depends on the number of edges linking the node. The edge between two nodes denotes that there exist several common genes involved in the two nodes-labeled pathways, thus the edge thickness relies on the number of the common genes. The blue arrows denoted the pathways containing the gene COL13A1. (e) Dot plot scattergrams drawn to show the obtained strong negative linear correlation pairs of LOC105378347-COL13A1. Abbreviations: PQDS: chronic atrophic gastritis patients with Pi-qi-deficiency syndrome; COL13A1: collagen type XIII alpha 1 chain.

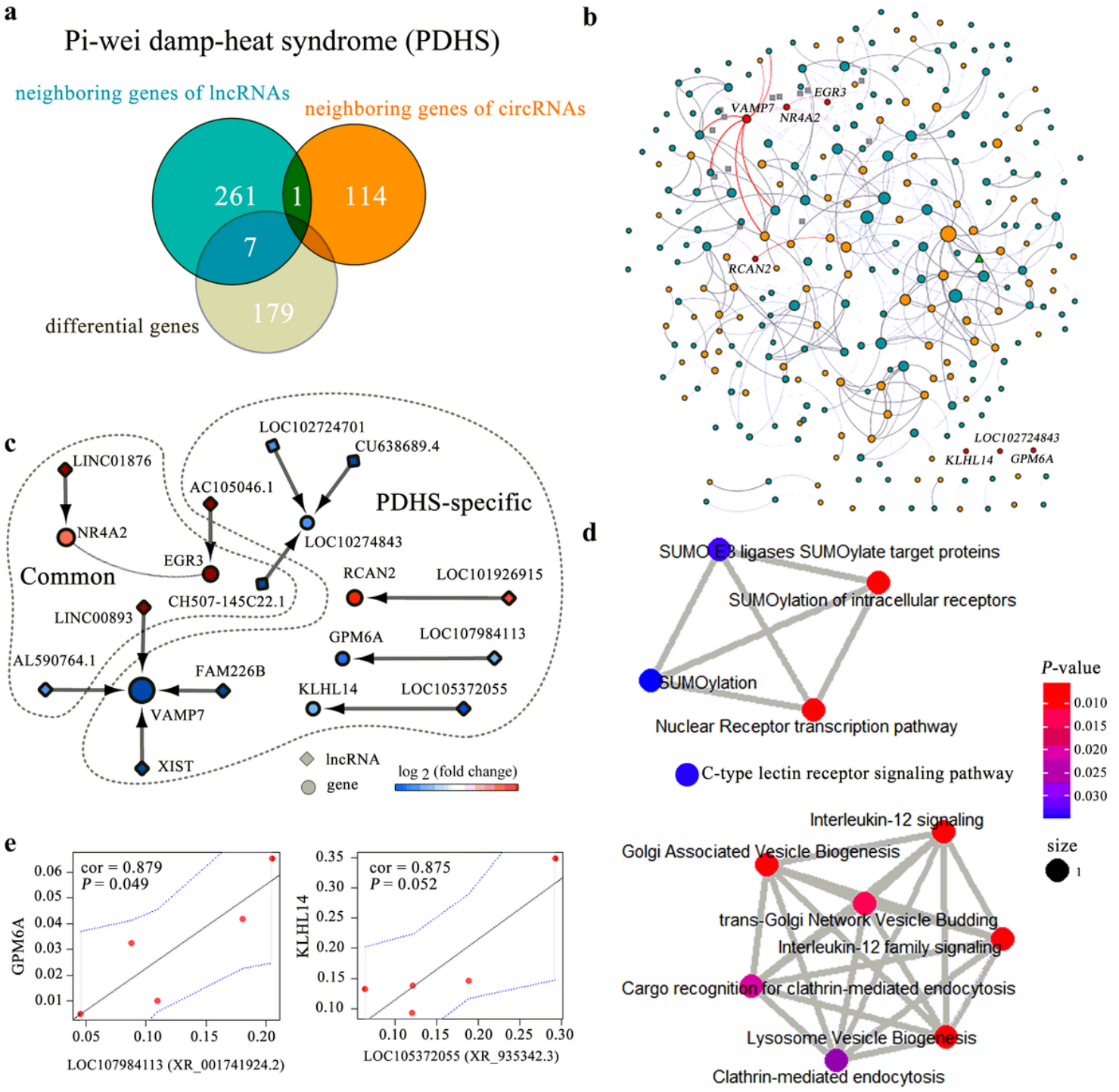

Figure 3

Function and pathway analyses of the targets of the PDHS-specific cis-acting lncRNAs/circRNAs in leukocytes. (a) Venn diagrams for the differentially expressed genes and the neighboring genes of differential lncRNAs/circRNAs in the leukocytes. (b) Overview of the interaction network of all the neighboring genes of differential lncRNAs/circRNAs. Node size depends on the number of edges. The orange nodes represent the neighboring genes of differential circRNAs, the greenish-blue nodes represent the neighboring genes of differential lncRNAs, and the red nodes denote the seven neighboring genes

identified to be differentially expressed in the leukocytes. (c) The interaction network generated to detail the interactions among the obtained seven targets of the 12 cis-acting lncRNAs. The expression levels of lncRNAs and genes were visually presented in the form of colorful nodes showing the color change (blue to red) in brightness and chromaticity. (d) The relationship network of the enriched pathways of the seven target genes. A node labeled with a pathway term indicates an enriched pathway of targets, and the node size depends on the number of edges linking the node. The edge between two nodes denotes that there exist several common genes involved in the two nodes-labeled pathways, thus the edge thickness relies on the number of the common genes. (e) Dot plot scattergrams drawn to show the obtained strong linear correlation pairs of LOC107984113-GPM6A and LOC105372055-KLHL14. Abbreviations: PDHS, chronic atrophic gastritis patients with Pi-wei damp-heat syndrome; GPM6A: glycoprotein M6A; KLHL14: kelch like family member 14.

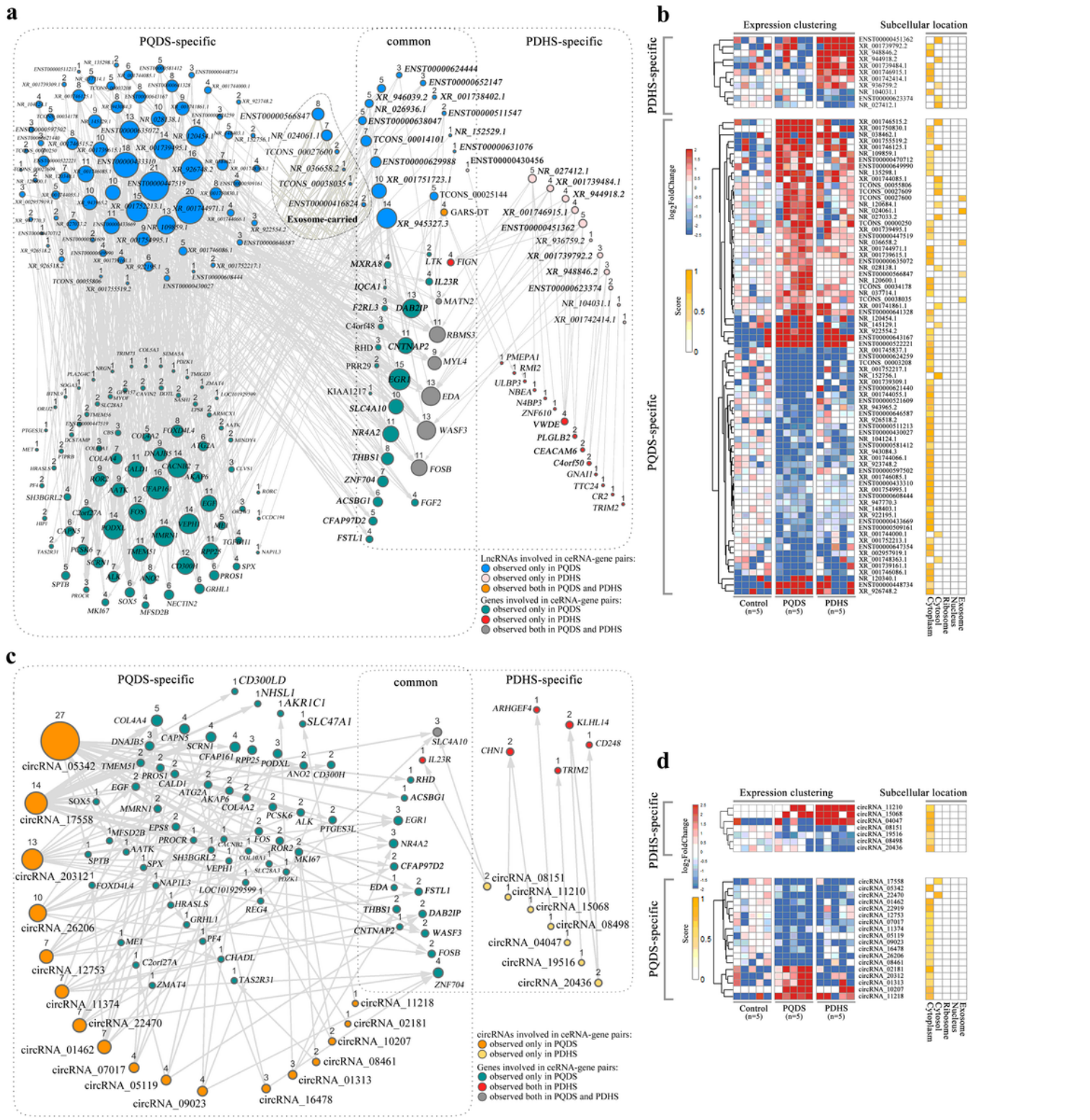

Figure 4

The Zheng-specific ceRNA-gene relationship networks in the leukocytes from the PQDS and PDHS populations. (a) The integrated ceRNA network detailing the PQDS-specific and PDHS-specific lncRNA-gene relationship pairs in leukocytes. The common lncRNAs involved in the lncRNA-gene regulation pairs were also marked. The “Exosome-contained”-labeled lncRNAs could be contained and carried in exosome, and five of them belong to the PQDS-specific lncRNAs. Node size depends on the number of edges, and

the number is shown near the corresponding node. The edge thickness relies on the corresponding ceRNA score value of each relationship pair. (b) Expression clustering and subcellular location analyses for the Zheng-specific lncRNAs. Prediction of subcellular location involves cytosol, cytoplasm, ribosome, nucleus and exosome. (c) The integrated ceRNA network detailing the PQDS-specific and PDHS-specific circRNA-gene relationship pairs in leukocytes. The common circRNAs implicated in circRNA-gene regulation pairs were also marked. (d) Expression clustering and subcellular location analyses for the Zheng-specific circRNAs. Abbreviations: PDHS: chronic atrophic gastritis patients with Pi-wei damp-heat syndrome; PQDS: chronic atrophic gastritis patients with Pi-qi-deficiency syndrome.

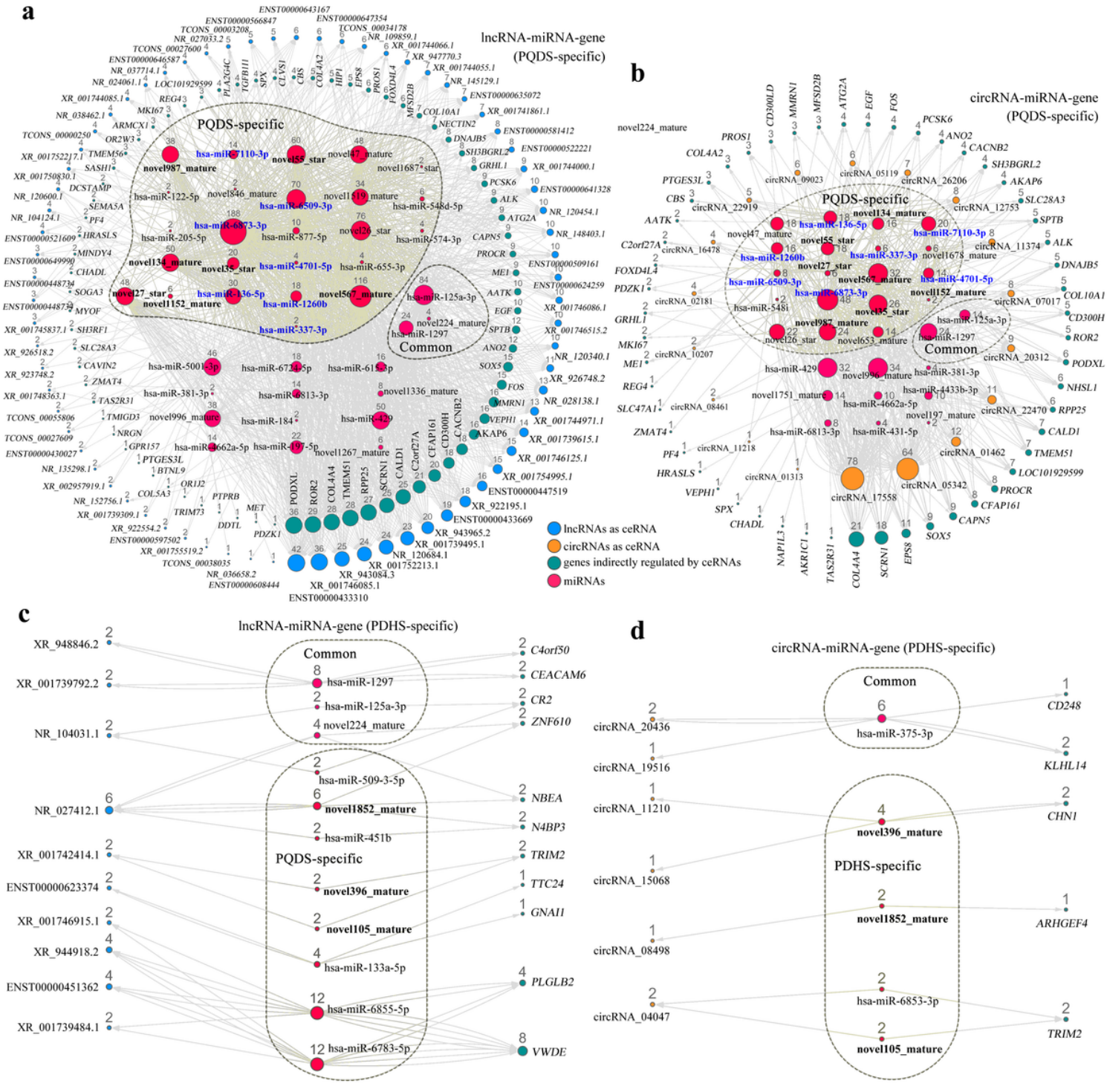

Figure 5

The Zheng-specific ceRNA-miRNA-gene triple relationship networks in the leukocytes from the PQDS and PDHS populations. (a) The PQDS-specific lncRNA-miRNA-gene ceRNA network. The PQDS-specific miRNAs implicated in both lncRNA-gene pairs and lncRNA-gene pairs, were particularly labeled in bold font (the blue font denoted the known miRNAs among them). (b) The PQDS-specific circRNA-miRNA-gene ceRNA network. (c) The PDHS-specific lncRNA-miRNA-gene ceRNA network. The miRNAs labeled in bold were involved in not only lncRNA-gene pairs but also circRNA-gene pairs. (d) The PDHS-specific circRNA-miRNA-gene ceRNA network. Abbreviations: PQDS: chronic atrophic gastritis patients with Pi-qi-deficiency syndrome; PDHS: chronic atrophic gastritis patients with Pi-wei damp-heat syndrome.

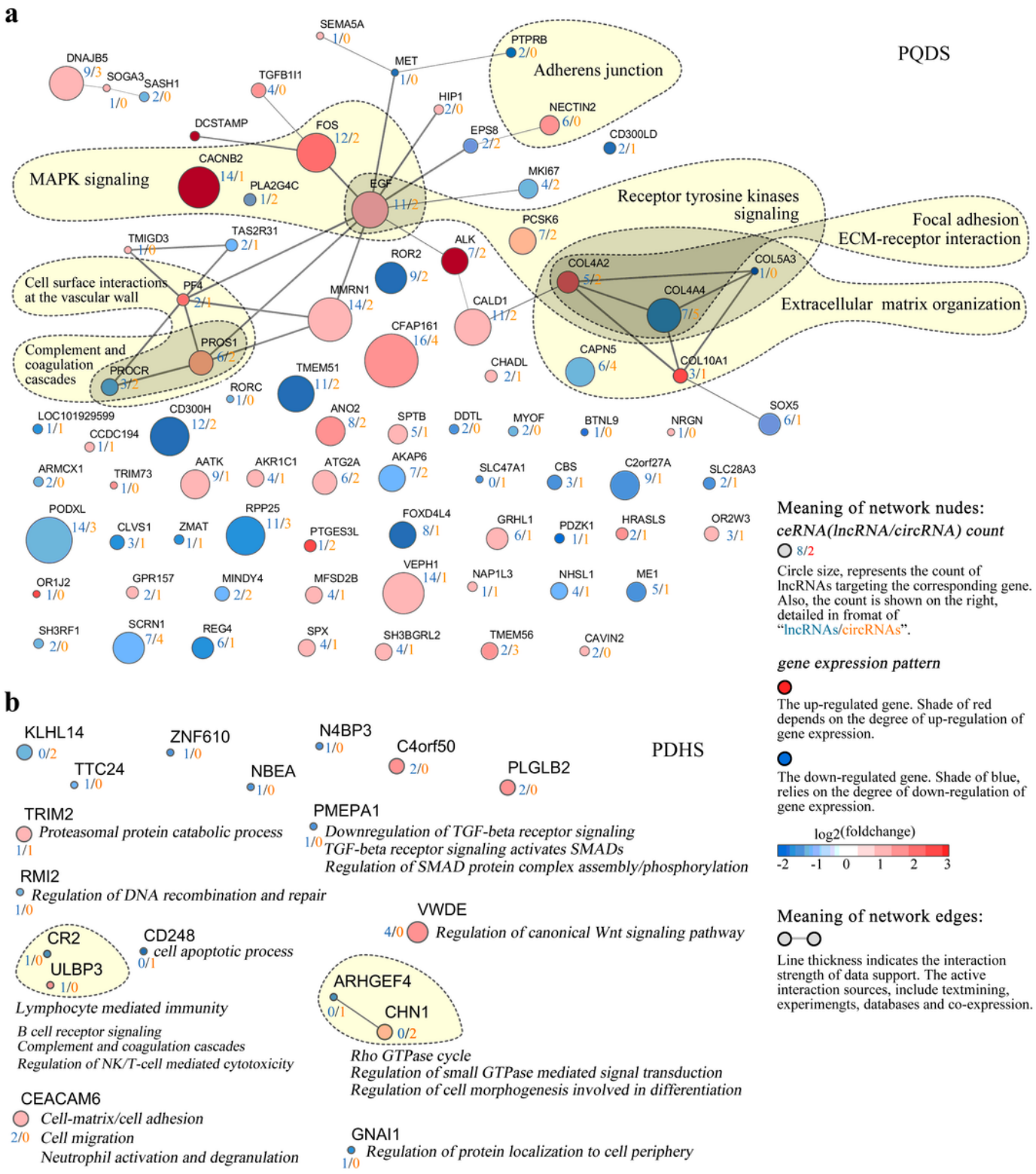

Figure 6

Interaction network analyses of the genes belonging to the Zheng-specific ceRNA-gene pairs identified in the leukocytes from the PQDS and PDHS populations. (a) The network generated to detail the interactions of genes involved in the PQDS-specific ceRNA-gene pairs. (b) The network drawn to show the interactions of genes implicated in the PQDS-specific ceRNA-gene pairs. The number near a node denotes the count of Zheng-specific ceRNAs governing the corresponding node gene in the format of "lncRNA

number/circRNA number". Node size depends on the number of Zheng-specific ceRNAs targeting to the corresponding node gene. The node genes-enriched pathways or GO (biological process) function terms were specially marked in the resultant networks. Abbreviations: PQDS, chronic atrophic gastritis patients with Pi-qi-deficiency syndrome; PDHS, chronic atrophic gastritis patients with Pi-wei damp-heat syndrome.

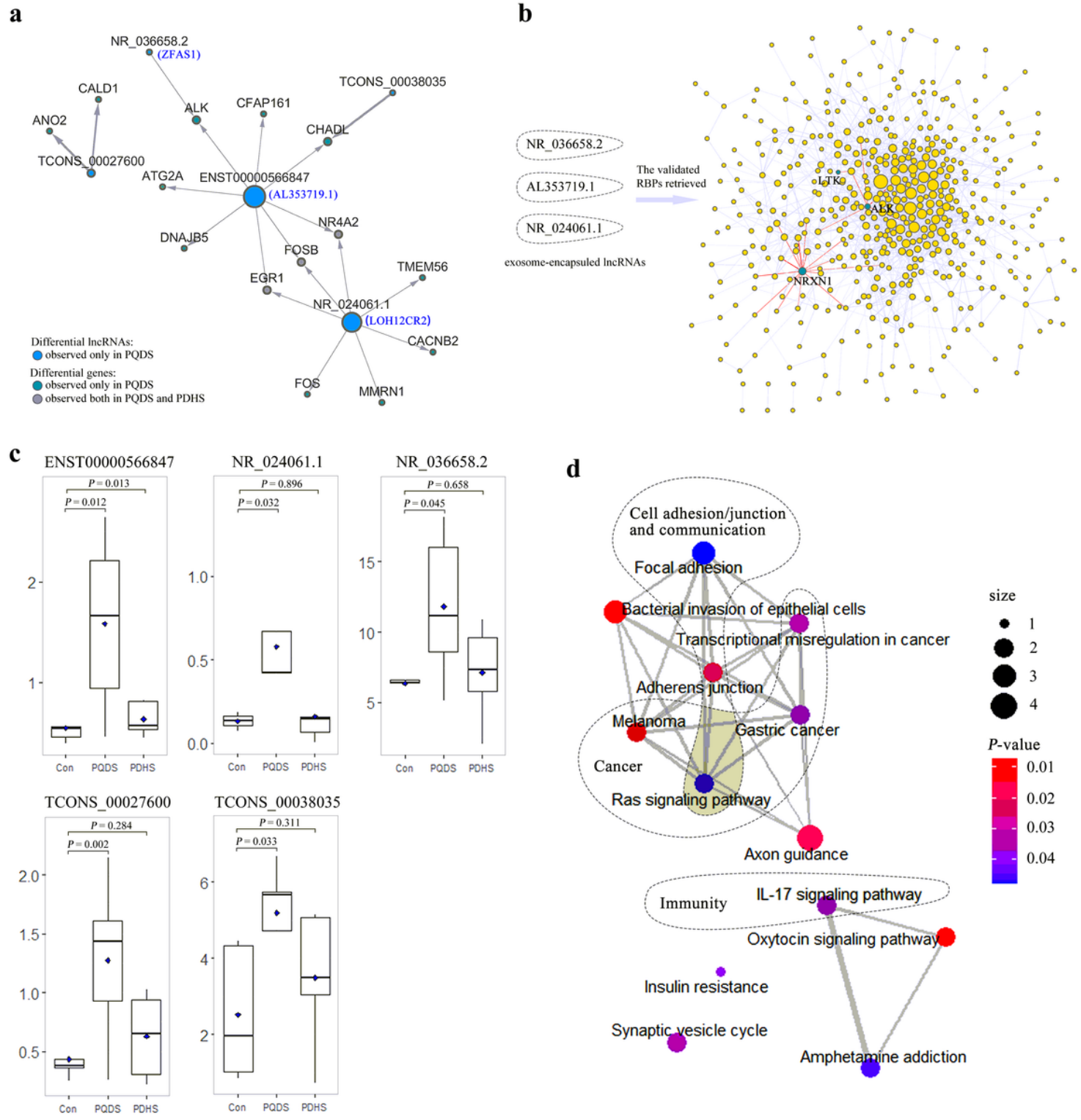

Figure 7

Function and pathway analyses for the exosome-lncRNAs) identified in the leukocytes from the PQDS population. (a) The network drawn to detail the ceRNA-gene relationship pairs implicated in the exosome-encapsuled lncRNAs. (b) Overview of the interaction network of the validated RNA binding proteins (RBPs) of the exosome-encapsuled lncRNAs. (c) Boxplot graphs of transcript levels of the five exosome-carried lncRNAs in the leukocytes from different populations. (d) The generated relationship network of the enriched pathways of the validated RBPs of the exosome-carried lncRNAs. The edge between two nodes indicates there exists several common genes implicated in the two nodes-labeled pathways, and therefore the edge thickness relies on the number of the existed common genes. Abbreviations: Con: healthy individuals; PQDS: chronic atrophic gastritis patients with Pi-qi-deficiency syndrome; PDHS: chronic atrophic gastritis patients with Pi-wei damp-heat syndrome.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementryinformation1.4.3.docx