

Transcriptomic Profiling of Synovial Sarcoma Reveals Novel Transcriptional and Post-Transcriptional Abnormalities

Zhengwang Sun

Fudan University Shanghai Cancer Center

Qiyue Xu

East China Normal University

Quan Huang

Shanghai Changzheng Hospital

Mengchen Yin

Shanghai University of Traditional Chinese Medicine

Hongqiang Zhang

Fudan University Shanghai Cancer Center Minhang Branch Hospital

Geng Chen (✉ gchen@bio.ecnu.edu.cn)

East China Normal University <https://orcid.org/0000-0002-3452-9434>

Yong Chen

Fudan University Shanghai Cancer Center

Wangjun Yan

Fudan University Shanghai Cancer Center

Primary research

Keywords: synovial sarcoma, gene expression, alternative splicing, gene fusion, circular RNA

Posted Date: February 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-205642/v1>

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

[Read Full License](#)

Abstract

Background

Synovial sarcoma (SS) is a rare type of cancer but relatively common in soft tissue tumors. The transcriptome landscape of SS is crucial for understanding the SS biology; however, the transcriptomic profile of SS is still poorly understood.

Methods

Here we collected 10 paired SS and adjacent normal tissues, and then performed deeply total RNA sequencing to systematically dissect the transcriptomic changes of SS in terms of differential expression, alternative splicing, gene fusion, and circular RNAs.

Results

By comparing SS with adjacent normal tissues, we identified 2309 upregulated and downregulated genes in SS. Those upregulated genes could lead to the upregulation of cell cycle, ribosome, and DNA replication pathways, while the downregulated genes may result in the downregulation of a set of metabolic biological processes and signaling pathways. Moreover, 2511 genes (including 21 splicing factors) were differentially alternative spliced, indicating that the deregulation of alternative splicing could be one important factor contributes to the tumorigenesis. Additionally, we not only detected the common fusion events of SS18-SSX1/SSX2 in SS, but also uncovered 11 novel gene fusions. Interestingly, 49 circular RNAs (circRNAs) were differentially expressed and their parental genes could function in muscle contraction and muscle system processes. Interaction network reconstruction and exploration revealed that these differentially expressed circRNAs may function as the miRNA sponge to regulate the expression of an abundance of differentially expressed and alternatively spliced genes.

Conclusions

Collectively, our results provide novel insights into the transcriptome and underlying mechanism of SS, which could benefit the diagnosis and treatment of SS.

Background

Synovial sarcoma (SS) is a rare and aggressive soft tissue cancer, which tends to occur near large joints, particularly in the extremities of the arms or legs, in young adults[1]. At present, surgery is still the main treatment strategy for SS. Cytogenetically, a significant portion of SS cases involve the nonrandom translocations between SS18 and SSX[2]. Although a range of studies investigated the genetic profile of SS from different cascades, the exact genetic cause of SS is not clearly defined to date[3, 4]. RNA sequencing (RNA-Seq) can provide a comprehensive view and valuable insight into the transcriptome from various aspects, such as expression level, alternative splicing (AS), gene fusions, and circular RNAs. These analyses are essential to systematically reveal and better understand the abnormally

transcriptomic changes of SS; however, a comprehensive exploration of the SS transcriptome from these aspects is still currently lacking.

AS is a crucial mechanism of post-transcriptional modification responsible for increasing both transcriptome and proteome diversity of a cell in eukaryotes[5, 6]. Moreover, AS play important roles in a variety of physiological processes (e.g. developmental programming), and the misregulation of AS can result in splicing defects which may have pathogenic function to cause severe diseases, especially cancers[7, 8]. However, the abnormalities of AS profile in SS on a genome-wide scale is still rarely understood. Furthermore, besides the common gene fusions resulted from the translocation between chromosome X and 18 in SS, other gene fusions could contribute to tumorigenesis through generating abnormal proteins or resulting in loss of function for the two fusion partners[9]. In addition, an increasing number of studies revealed that circular RNAs (circRNAs) have diverse critical functions including work as miRNA sponges and modulators of transcription[10, 11]. CircRNAs could be vital for different aspects of malignant phenotypes, including cell cycle, apoptosis, and invasion[12, 13]. A multitude of circRNAs have been demonstrated to be closely associated with human cancers and some of them are potentially important biomarkers[12, 14, 15]. But little is known about the expression profile and functions of circRNAs in SS since few studies investigated the circRNA profile in SS. Thus, systematic analysis of the abnormal changes of both transcriptional and post-transcriptional layers is necessary to comprehensively understand the SS development and progression.

Here we performed Ribo-Zero RNA-seq on ten pairs of Chinese SS and adjacent normal tissues to explore the transcriptome abnormalities of SS from various aspects. We first carried out differential expression calling and detected a number of genes with significant expression level changes. Then the AS deregulation of a multitude of genes and a set of tumor-specific gene fusion events were identified. Moreover, we dissected the expression profile of circRNAs between SS and adjacent normal tissues. Finally, we constructed an interaction network among circRNAs, miRNAs, and their target genes, which enabled us to gain insights into the potential functional mechanism of circRNAs in SS.

Methods

Ethics

Clinical samples were collected from Fudan University Shanghai Cancer Center. This study was approved by the Ethics Committee of Fudan University Shanghai Cancer Center. All patients in this study provided written informed consent for sample collection and data analyses.

RNA extraction

Total RNA was extracted from the 10 mg synovial sarcoma tissues after grinding by Homogenizer (Scientz) using TRIzol® Reagent (Invitrogen) and RNeasy MinElute spin column (Qiagen) according to the manufacturer's instructions. Then the integrity of the total RNA was determined by 2100 Bioanalyser

(Agilent) and quantified using the NanoDrop (Thermo Scientific). About 1 ug high-quality or media-quality RNA sample ($OD_{260/280}=1.9\sim 2.0$, $RIN\geq 4$) was used to construct the sequencing library.

Library preparation and Illumina Sequencing

RNA purification, reverse transcription, library construction, and sequencing were performed at WuXi NextCODE in Shanghai according to the manufacturer's instructions (Illumina). The rRNA-depleted sequencing libraries from total RNA were prepared using Illumina TruSeq® Stranded Total RNA Gold preparation Kit. About 1 ug total RNA was used as input material, and then the Ribo-Zero Gold kit was used to remove both cytoplasmic and mitochondrial rRNA. After purification of the remaining RNA without rRNA, the RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments are copied into first-strand cDNA using reverse transcriptase and random primers, followed by second-strand cDNA synthesis. These cDNA fragments then were subjected to end-repair, phosphorylation, and 'A' base addition according to Illumina's library construction protocol. The products were purified and enriched with PCR, and the AMPure XP Beads (Beckmen) were used to clean up the amplified target fragments to create the final cDNA library. After library construction, Qubit 2.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific) was used to quantify the concentration of the resulting sequencing libraries, while the size distribution was analyzed using Agilent BioAnalyzer 2100 (Agilent).

Sequencing was performed using the Illumina system following Illumina-provided protocols for 2x150 paired-end sequencing in WuXi NextCODE at Shanghai, China.

Short-read mapping and gene expression quantification

The RNA-seq reads of each sample for 10 pairs of SS and adjacent normal tissues were separately aligned to the human reference genome GRCh38 using HISAT2[16] (version 2.1.0). Then we quantified the expression of genes in each sample by employing StringTie[17] (version 1.3.6). The human gene annotation file in the GTF format of version 95 from the Ensembl database (<http://www.ensembl.org>) was used. The mapped read count and expression value in TPM (transcript per million) for each gene were extracted from the output file of StringTie.

Differential gene expression calling

For differential expression analysis, the read count mapped to each gene were used as input. The gene expression changes between SS and adjacent normal samples were examined using DESeq2[18] (version 1.24.0). We defined the differentially expressed genes (DEGs) with the threshold of $|\text{fold change}| > 2$ and adjusted $P\text{-value} < 0.01$.

Detection of alternative splicing events

We investigated the alternative splicing (AS) profile of genes between SS and adjacent normal tissues by employing rMATS[19] (version 4.0.2). The input was the bam files outputted by HISAT2 after read

mapping. Five common AS types of exon skipping (ES), alternative 3' acceptor sites (A3AS), alternative 5' donor sites (A5DS), intron retention (IE), and mutually exclusive exons (ME) were examined. The differential alternative splicing events were identified with the cutoff of FDR < 0.05.

Identification of gene fusions

In order to explore the genetic alterations, we employed TopHat-Fusion[20] (version 2.1.0) with default parameters to identify the gene fusion events in all tumor and normal samples. Only the fusions with at least 3 supporting reads and 2 supporting pairs were considered. Finally, 14 and 11 gene fusion pairs were detected in SS and adjacent normal tissues, respectively. We only kept the 14 gene fusions that are unique to SS and discarded the fusions that are present in adjacent normal tissues.

Circular RNA detection and differential expression analysis

We checked the expression profiles of circRNAs in SS and adjacent normal samples using CIRI[21] (version 0.1.0). Then differential expression analysis was conducted by employing DESeq2 (version 1.24.0) based on the circRNAs identified by CIRI. Only the circRNAs with expression changes of |fold change| >2 and adjusted *P*-value < 0.01 were considered as significant. The official IDs of circRNAs were obtained by coordinate mapping using the circBase[22] database.

Construction of interaction network among circRNAs, miRNAs and target genes

To gain insights into the function of circRNAs, we built an interaction network among the circRNAs, miRNAs, and the target genes of miRNAs. The PPI interactions were downloaded from the STRING[23] database (version 11.0). The regulatory relationship between miRNAs and target genes, as well as the known miRNA-circRNAs interactions, were obtained from the starBase[24] database (version 3.0). Each circRNA-miRNA pair was supported by > 5 CLIP-seq experiments, while each miRNA-target gene pair was detected in >2 CLIP-seq experiments and >2 degradome-seq experiments in the StarBase2 database. Then we incorporated these interactions to construct the interaction network among circRNAs, miRNAs, and the target genes of miRNAs using Cytoscape[25] (version 3.7.2). Only the parental genes of differentially expressed circRNAs, DEGs, DASGs, and fusion genes were considered in the interaction network.

Gene functional enrichment analysis

We conducted gene ontology (GO) and KEGG pathway enrichment analyses using GSEA[26] (version 4.0.1) for the upregulated and downregulated DEGs between SS and adjacent normal tissues. The functional enrichment analysis of biological processes and pathways for the differentially alternative spliced genes, fusion genes, and the parental genes of circular RNAs were carried out with clusterProfiler[27] (version 3.12.0). The enriched items with adjusted *P*-value < 0.05 was regarded as significant.

Results

An abundance of genes are differentially expressed between SS and adjacent normal tissues

To gain insights into the transcriptomic changes of SS patients, we collected the SS and adjacent normal tissues from 10 patients and then deeply sequenced them with total RNA sequencing (including both poly (A+) and poly (A-) RNAs). The RNA-seq reads of each sample were first mapped to the human reference genome GRCh38 using HISAT2[16], then differential expression calling was conducted by employing DEseq2[18]. A total of 4,286 differentially expressed genes (DEGs) were detected using the threshold of $|\text{fold change}| > 2$ and adjusted P -value < 0.01 , of which 2,309 (including 432 lncRNA genes) and 1,977 (including 333 lncRNA genes) genes were separately upregulated and downregulated in SS compared to normal tissues (Figure 1A, Supplementary Table S1). Interestingly, we found that 340, 185, 124, and 7 of those DEGs are oncogenes, tumor suppressor genes (TSGs), transcription factors (TFs), and splicing factors, respectively (Figure 1B, Supplementary Figure S1). Specifically, 52 TFs (such as AES and BCL6) were down-regulated and 72 TFs (e.g. ARID3A and BRCA2) were up-regulated, suggesting that these TFs could influence the expression of their downstream target genes including related oncogenes and TSGs. Since oncogenes and TSGs are closely correlated with cancer, they may significantly contribute to the development of SS. Moreover, we further validated the expression profiles of the seven differentially expressed splicing factors (ELAVL2, HNRNPA1, HNRNPH2, MBNL1, PCBP1, QKI, and TIA1) using qPCR. As expected, the experimental results are consistent with that of our RNA-seq data analysis. Splicing factors are critical for the AS process[28], differential expression of splicing factors could result in the AS deregulation of corresponding genes in SS.

Gene ontology (GO) and KEGG pathway enrichment analyses showed that these upregulated and downregulated DEGs were mainly involved in a range of fundamental and tumor-related biological processes and pathways (Figure 1C, D, and E $\text{FDR} < 0.05$). For example, the up-regulated DEGs were primarily enriched in the cell-cycle-related biological processes (e.g. chromosome organization, chromatin organization and DNA conformation change) and pathways of systemic lupus erythematosus, cell cycle, DNA replication and P53 signaling (Figure 1C). Several previous studies also identified the cell-cycle-related genes in sarcoma as a major category of up-regulated genes[29, 30], which further supported our finding. The down-regulated DEGs were mainly involved in the metabolic-related biological processes (such as energy derivation by oxidation of organic compounds, muscle system process, and glucan metabolic process) and the pathways of oxidative phosphorylation, insulin signaling pathway and vascular smooth muscle contraction (Figure 1D). Thus, the result suggests that a multitude of genes prominently changed their expression levels in SS, which could be one of the main factors accounting for tumorigenesis through up-regulating and down-regulating a range of crucial pathways.

Deregulation of alternative splicing could contribute to the tumorigenesis of SS

Considering that the misregulation of AS can lead to the production of aberrant proteins that contribute to tumorigenesis[8], we further compared the AS profile between SS and adjacent normal tissues by employing rMATS[19]. Five classical splicing categories of exon skipping (ES), alternative 5' donor sites

(A5DS), alternative 3' acceptor sites (A3AS), mutually exclusive exons (ME), and intron retention (IR) were analyzed. Remarkably, we identified 2511 (including 41 lncRNA genes) significantly differential AS genes (DASGs), of which 2018, 223, 242, 486, and 159 belong to the splicing mode changes of ES, A5DS, A3AS, ME and IR, respectively (Figure 2A, FDR < 0.05, Supplementary Table S2). As expected, ES is the most common differential splicing mode (80.37%, 2018 out of 2511 DASGs), whereas IR is the least (6.33%, 159 out of 2511 DASGs). Notably, the majority of those genes that underwent the five classical splicing mode changes are largely different, only a small portion of them simultaneously exhibited three or more distinct splicing types (Figure 2A).

Gene functional enrichment analysis indicated that those 2511 DASGs were mainly involved in the RNA splicing and cancer-related biological processes and KEGG pathways (Figure 2B and C, adjusted *P*-value < 0.05), which is highly correlated with the AS process. For instance, the top enriched biological processes of those DASGs are mRNA processing, microtubule cytoskeleton organization, and RNA splicing, while the enriched pathways are endocytosis, RNA transport, proteoglycans in cancer and spliceosome. Interestingly, we observed that 21 splicing factor genes showed significantly differential AS between SS and adjacent normal tissues, such as HNRNPA1, PTBP2, QKI, RBFOX2, and TRA2A. It is well known that the splicing factors are crucial for AS regulation[28], the deregulation of those splicing factors could drastically disrupt the splicing process of many other genes and contribute to the tumorigenesis of SS[31]. Furthermore, we found that 346, 204, and 122 oncogenes, TSGs, and TFs were also differentially spliced (Figure 2D). The abnormal splicing of these TFs could influence the expression of many downstream target genes, which could be closely associated with the development and progression of SS. Only 368 genes shared between DASGs and DEGs, and, leaving the most of them are distinct (Figure 2E). These common 368 genes were enriched in the biological process of actomyosin structure organization and pathway of regulation of actin cytoskeleton (Figure 2E, adjusted *P*-value < 0.05). Thus, the genes showed differential expression are quite distinct from those exhibited differential splicing, suggesting that expression level is complementary with AS in revealing the transcriptomic changes. Our results indicate that the abnormal AS changes of genes could be another important factor responsible for the tumorigenesis of SS.

Novel gene fusion events are identified in SS

Considering that the chromosomal translocation between X and 18 chromosomes is frequent in SS[32], we further explored the gene fusion events in SS patients. A total of 14 and 11 unique gene fusion pairs were identified in SS and adjacent normal tissues using TopHat-Fusion[20], respectively. No fusion events shared between SS and adjacent normal tissues. The 14 tumor-specific gene fusion pairs were from seven SS patients (Supplementary Table S3), most of which (11 out of 14) were resulted from the rearrangements within the same chromosome, while 3 of them were generated from the breakage and re-joining of two disparate chromosomes (Figure 3A). In total, 27 genes were involved in these tumor-specific gene fusions, only SS18 was fused with SSX1 and SSX2, other genes were only fused with one partner (Figure 3B).

Moreover, all the 14 tumor-specific gene fusion pairs were detected only once and the maximum number of gene fusion pairs detected in individual patients was four (Figure 3C), suggesting that gene fusion events in those SS patients are quite distinct. (Figure 3C). Intriguingly, these tumor-specific gene fusion events contain one TF of SSX2 and seven oncogenes of SS18, SSX1, SSX2, BCOR, CNOT1, HIST2H2AC, and TOP1 (Figure 3D). Oncogene SS18 was fused with the TF and oncogene of SSX2 as well as the oncogene SSX1, which is consistent with the known finding that such fusion events are common for SS[32]. Besides, other oncogenes of BCOR, CNOT1, HIST2H2AC, and TOP1 were formed the fusion events of BCOR-CCNB3, CNOT1-SETD6, HIST2H2AC-HIST2H2AB, and TOP1-PLCG1-AS1, respectively. Previous studies have shown that BCOR-CCNB3 fusion tends to occur in the undifferentiated small round-cell sarcomas like Ewing sarcoma and has the potential to drive sarcoma progression[33-35]. Other gene fusions could be novel for SS, and the involved genes are with critical functions. For example, CNOT1 encodes the CCR4-NOT transcription complex subunit 1, which mainly participates in deadenylating mRNAs[36]. HIST2H2AC and HIST2H2AB can generate the replication-dependent histones that are basic nuclear proteins responsible for the nucleosome structure of the chromosomal fiber. TOP1 encodes the enzyme of DNA topoisomerase for controlling and altering the topologic states of DNA during transcription[37]. Since TF could regulate the expression of many downstream target genes and oncogenes are closely associated with cancer, the fusion events of these TFs and oncogenes could contribute to the tumorigenesis/progression of SS. Interestingly, lncRNA genes of LINC00970, LOC105375787 and PLCG1-AS1 are also involved in the gene fusion events, but their functions are still unknown. Gene functional enrichment analysis showed that those fusion genes were significantly enriched in the KEGG pathway of transcriptional misregulation in cancer (Figure 3E, adjusted P -value < 0.05). Therefore, we not only identified the common gene fusions between chromosome 18 and X, but also detected some novel tumor-specific gene fusions that could contribute to the formation/progression of SS.

Circular RNAs are functionally important in SS

Emerging evidence shows that circRNAs can involve in various aspects of tumor biology[15, 38], thus we further investigated the expression profile of circRNAs in SS and adjacent normal tissues. We detected 49 differentially expressed circular RNAs by employing CIRI[21] with the threshold of |fold change| >2 and adjusted P -value < 0.01 (Supplementary Table S4). As shown in Figure 4A, 21 of them were significantly up-regulated in SS, whereas the other 28 were down-regulated. Furthermore, we found that the great majority (46 out of 49, 93.88%) of those differentially expressed circRNAs were formed by the circulation of exons of their parental genes, only two circRNAs of 10:24380869|24384423 (parental gene: KIAA1217) and 17:35168061|35168685 (parental gene: UNC45B) were produced from the intronic region and another one (5:137757867|137759020) was generated by an intergenic region (Figure 4B).

Intriguingly, 7, 5, and 3 of the parental genes for those differentially expressed circRNAs are oncogenes, TSGs, and TFs (Figure 4C). These circRNAs have the potential to affect the expression of their parental oncogenes, TSGs, and TFs, because previous studies have shown that circRNAs can form posttranscriptional regulators to regulate the expression of their parental genes[39, 40]. Strikingly, gene

functional enrichment analysis showed that the parental genes of those differentially expressed circRNAs were mainly enriched in muscle system process (such as MAP2K4, HDAC4, TMEM38A, MYH1, MYH2, CAMK2G, TRDN, and SULF2) and muscle contraction (e.g. HDAC4, TMEM38A, MYH1, MYH2, TRDN and SULF2) (Figure 4D, adjusted P -value < 0.05), which is highly correlated with SS. Therefore, our result indicates that those differentially expressed circRNAs could play key roles in SS.

The genes involved in different types of transcriptomic changes are largely distinct

We further compared the four types of genes with significant changes in terms of expression level and AS, as well as the fusion genes and the parental genes of differentially expressed circRNAs. As shown in Figure 5, the genes in one type are largely distinct from that of other types, and no genes were common among these four categories. Only a fraction of them were involved in two or three types of changes (Figure 5). Intriguingly, 3 DEGs (BCOR, HIST2H2AB, and MEG8) and 2 DASGs (AKR1E2 and DCAF8) were overlapped with the fusion genes, suggesting that the fusion events may influence the expression and/or AS profile of these genes. BCOR is an oncogene, while MEG8 is an imprinted gene. Moreover, 18 DEGs (e.g. DNM3OS, ZNF730, DNAH14, and AFF2) shared with the parental genes of differentially expressed circRNAs, implying that expression changes of these genes could affect the expression of circRNAs as well. In addition, 17 DASGs (such as SUCO, VWA8, MTUS1, and USP53) were common to the parental genes of differentially expressed circRNAs. Since circRNAs are mainly formed by AS of pre-mRNAs through backsplicing[41], the AS changes of these DASGs has the potential to influence the expression of corresponding circRNAs. Accordingly, our result shows that all the four aspects of expression changes, AS, gene fusions, and circRNAs could be closely correlated with the tumorigenesis/progression of SS.

CircRNAs could regulate the expression of a multitude of genes by acting as miRNA sponges

An increasing number of studies suggested that the endogenous circRNAs can act as miRNA sponges to regulate corresponding gene expression[42, 43], we further constructed the interaction network among differentially expressed circRNAs, miRNAs and the miRNA target genes of DEGs, DASGs and fusion genes to elucidate the functional roles of those differentially expressed circRNAs. Based on the known miRNA-circRNA regulations, and the miRNA-targets relationships in starBase database[24] as well as the protein-protein interactions (PPIs) in String database[23], the resulting interaction network involved in 5 circRNAs (hsa_circ_0001699, hsa_circ_0000247, hsa_circ_0000246, hsa_circ_0000095, and hsa_circ_0000118), 44 miRNAs, 293 protein-coding genes, containing 57 miRNA-circRNA interactions, 789 miRNA-mRNA interactions and 350 PPIs (Figure 6). It is well known that circRNAs can regulate gene expression through influencing transcription, mRNA turnover as well as translation by sponging RNA-binding proteins (RBPs) and miRNAs[43]. Our resulting network showed that circRNAs hsa_circ_0001699, hsa_circ_0000247, hsa_circ_0000246, hsa_circ_0000095 and hsa_circ_0000118 could separately act as the sponges of 14, 13, 13, 12 and 5 miRNAs, where these miRNAs have the potential to regulate the expression of 119, 202, and 3 genes of DEGs, DASGs, and/or fusion genes. The expression of these miRNA target genes could be indirectly influenced by corresponding circRNAs through competing the interaction with miRNAs.

Consequently, our finding suggests that circRNAs could function as miRNA sponges to regulate the expression of an abundance of genes.

Discussion

In this study, we systematically explored the transcriptome of SS in terms of the abnormal changes of gene expression and AS, as well as gene fusions and circRNAs. A total of 4286 genes (including 765 lncRNA genes) were differentially expressed between SS and paired normal tissues, which were mainly involved in fundamental biological processes and cancer-related pathways. We further detected 2511 genes (including 41 lncRNA genes) that showed differential AS, the most common AS mode was ES (80.37% of these DASGs), followed by ME, A3AS, A5DS and IR. As expected, these DASGs were enriched in splicing related biological processes and pathways. However, those DEGs and DASGs were largely distinct, only a small portion of them are the same, suggesting that AS is complementary with expression level in revealing transcriptomic changes. Notably, a fraction of those DEGs and DAGs are oncogenes, tumor suppressors, and TFs, indicating that they could be closely associated with the tumorigenesis of SS. Moreover, we identified 14 tumor-specific gene fusion pairs in SS, which not only included the common gene fusions of SS18-SSX[32], but also contained novel fusion events involving both protein-coding and lncRNA genes. Additionally, we observed that 49 circRNAs were markedly changed expression levels in SS compared to adjacent normal tissues, and the functions of their parental genes were closely correlated with SS.

To the best of our knowledge, we are the first to study the aberrant changes of SS transcriptome from a comprehensive view covering both transcriptional and post-transcriptional levels. Specifically, the deregulation of AS and the functional role of circRNAs are rarely investigated in SS previously. An increasing number of studies have shown that imbalances in the AS process can affect the development of various human diseases, especially the oncogenesis, progression, and metastasis of a range of cancers[44]. We identified 124 differentially expressed TFs and 122 differentially spliced TFs, suggesting that these TFs could be responsible for the expression level changes of an abundance of genes[45, 46]. Moreover, we observed that 7 and 21 splicing factors were dramatically changed in expression level or AS profile. Since splicing factors are essential in regulating the AS of genes, these abnormally changed splicing factors could significantly contribute to the AS changes of many genes[47]. On the other hand, circRNAs have critical regulatory functions and play key roles in the initiation and progression of diverse diseases including cancers[48, 49]. We observed that those differentially expressed circRNAs are mainly generated from the genes correlated with muscle system process and contraction, indicating that they are functionally important in the development of SS. We also constructed the interaction network among circRNAs, miRNAs and downstream target genes to elucidate their potential regulatory mechanism. Those differentially expressed circRNAs could act as the sponge for dozens of miRNAs to indirectly regulate the expression of hundreds of DEGs and DASGs.

Conclusion

Collectively, we comprehensively dissected the transcriptomic changes of SS and identified a multitude of DEGs, DASGs, fusion genes, and circRNAs that could be closely associated with the tumorigenesis/progression of SS. Our results not only elucidated the abnormal variations of SS transcription and post-transcription, but also shed light on the underlying molecular mechanisms.

Abbreviations

SS: Synovial sarcoma

circRNAs: Circular RNAs

RNA-Seq: RNA sequencing

AS: Alternative splicing

ES: Exon skipping

A3AS: Alternative 3' acceptor sites

A5DS: Alternative 5' donor sites

IE: Intron retention

ME: Mutually exclusive exons

DEGs: Differentially expressed genes

DASGs: Differentially alternative spliced genes

TSGs: Tumor suppressor genes

TFs: Transcription factors

Declarations

Ethics approval and consent to participate

Clinical samples were collected from Fudan University Shanghai Cancer Center. This study was approved by the Ethics Committee of Fudan University Shanghai Cancer Center. All patients in this study provided written informed consent for sample collection and data analyses.

Consent for publication

Not applicable

Availability of data and materials

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE144190.

Competing interests

The authors declare no conflicts of interest.

Funding

This work was supported by the National Natural Science Foundation of China (31771460, 91629103).

Authors' contributions

GC, YC and WY designed and supervised the study. ZS collected and prepared the samples. QX, ZS, QH and MY performed the analysis. GC and ZS wrote the manuscript.

Acknowledgements

We thank the helpful discussion with Qizhi Li.

References

1. Ladanyi M, Antonescu CR, Leung DH, et al. Impact of SYT-SSX fusion type on the clinical behavior of synovial sarcoma: a multi-institutional retrospective study of 243 patients. *Cancer Res.* 2002; 62(1):135-140.
2. Przybyl J, Sciort R, Rutkowski P, et al. Recurrent and novel SS18-SSX fusion transcripts in synovial sarcoma: description of three new cases. *Tumour Biol.* 2012; 33(6):2245-2253.
3. Cancer Genome Atlas Research Network. Electronic address edsc, Cancer Genome Atlas Research N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell.* 2017; 171(4):950-965 e928.
4. He RQ, Wei QJ, Tang RX, et al. Prediction of clinical outcome and survival in soft-tissue sarcoma using a ten-lncRNA signature. *Oncotarget.* 2017; 8(46):80336-80347.
5. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008; 456(7221):470-476.
6. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet.* 2010; 11(5):345-355.
7. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet.* 2007; 8(10):749-761.
8. Zhang J, Manley JL. Misregulation of pre-mRNA alternative splicing in cancer. *Cancer Discov.* 2013; 3(11):1228-1237.

9. Edwards PA. Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol.* 2010; 220(2):244-254.
10. Chen LL. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol.* 2016; 17(4):205-211.
11. Li X, Yang L, Chen LL. The Biogenesis, Functions, and Challenges of Circular RNAs. *Mol Cell.* 2018; 71(3):428-442.
12. Greene J, Baird AM, Brady L, et al. Circular RNAs: Biogenesis, Function and Role in Human Diseases. *Front Mol Biosci.* 2017; 4:38.
13. Qu S, Yang X, Li X, et al. Circular RNA: A new star of noncoding RNAs. *Cancer Lett.* 2015; 365(2):141-148.
14. Abu N, Jamal R. Circular RNAs as Promising Biomarkers: A Mini-Review. *Front Physiol.* 2016; 7.
15. Dong YP, He D, Peng ZZ, et al. Circular RNAs in cancer: an emerging key player. *J Hematol Oncol.* 2017; 10.
16. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12(4):357-360.
17. Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology.* 2015; 33(3):290-+.
18. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15(12):550.
19. Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A.* 2014; 111(51):E5593-5601.
20. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biology.* 2011; 12(8).
21. Gao Y, Wang JF, Zhao FQ. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biology.* 2015; 16.
22. Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *Rna.* 2014; 20(11):1666-1670.
23. Szklarczyk D, Gable AL, Lyon D, 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-D613.
24. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research.* 2014; 42(D1):D92-D97.
25. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13(11):2498-2504.
26. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *P Natl Acad Sci USA.* 2005; 102(43):15545-15550.

27. Yu GC, Wang LG, Han YY, et al. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics*. 2012; 16(5):284-287.
28. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem*. 2015; 84:291-323.
29. Chibon F, Lagarde P, Salas S, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med*. 2010; 16(7):781-U781.
30. Yen CC, Yeh CN, Cheng CT, et al. Integrating Bioinformatics and Clinicopathological Research of Gastrointestinal Stromal Tumors: Identification of Aurora Kinase A as a Poor Risk Marker. *Ann Surg Oncol*. 2012; 19(11):3491-3499.
31. Dvinge H, Kim E, Abdel-Wahab O, et al. RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer*. 2016; 16(7):413-430.
32. Kawai A, Woodruff J, Healey JH, et al. SYT-SSX gene fusion as a determinant of morphology and prognosis in synovial sarcoma. *New Engl J Med*. 1998; 338(3):153-160.
33. Pierron G, Tirode F, Lucchesi C, et al. A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion. *Nat Genet*. 2012; 44(4):461-U151.
34. Li WS, Liao IC, Wen MC, et al. BCOR-CCNB3-positive soft tissue sarcoma with round-cell and spindle-cell histology: a series of four cases highlighting the pitfall of mimicking poorly differentiated synovial sarcoma. *Histopathology*. 2016; 69(5):792-801.
35. Kao YC, Owosho AA, Sung YS, et al. BCOR-CCNB3 Fusion Positive Sarcomas A Clinicopathologic and Molecular Analysis of 36 Cases With Comparison to Morphologic Spectrum and Clinical Behavior of Other Round Cell Sarcomas. *Am J Surg Pathol*. 2018; 42(5):604-615.
36. Pavanello L, Hall B, Airhihen B, et al. The central region of CNOT1 and CNOT9 stimulates deadenylation by the Ccr4-Not nuclease module. *Biochem J*. 2018; 475:3437-3450.
37. Baranello L, Wojtowicz D, Cui KR, et al. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell*. 2016; 165(2):357-371.
38. Zhang MC, Xin Y. Circular RNAs: a new frontier for cancer diagnosis and therapy. *J Hematol Oncol*. 2018; 11.
39. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013; 495(7441):333-338.
40. Zhang Y, Zhang XO, Chen T, et al. Circular Intronic Long Noncoding RNAs. *Molecular Cell*. 2013; 51(6):792-806.
41. Barrett SP, Salzman J. Circular RNAs: analysis, expression and potential functions. *Development*. 2016; 143(11):1838-1847.
42. Kulcheski FR, Christoff AP, Margis R. Circular RNAs are miRNA sponges and can be used as a new class of biomarker. *J Biotechnol*. 2016; 238:42-51.

43. Panda AC. Circular RNAs Act as miRNA Sponges. *Circular Rnas: Biogenesis and Functions*. 2018; 1087:67-79.
44. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016; 17(1):19-32.
45. Vaquerizas JM, Kummerfeld SK, Teichmann SA, et al. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 2009; 10(4):252-263.
46. Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. *Cell*. 2018; 172(4):650-665.
47. Anczukow O, Krainer AR. Splicing-factor alterations in cancers. *Rna*. 2016; 22(9):1285-1301.
48. Zhang ZR, Yang TT, Xiao JJ. Circular RNAs: Promising Biomarkers for Human Diseases. *Ebiomedicine*. 2018; 34:267-274.
49. Haddad G, Lorenzen JM. Biogenesis and Function of Circular RNAs in Health and in Disease. *Front Pharmacol*. 2019; 10.

Figures

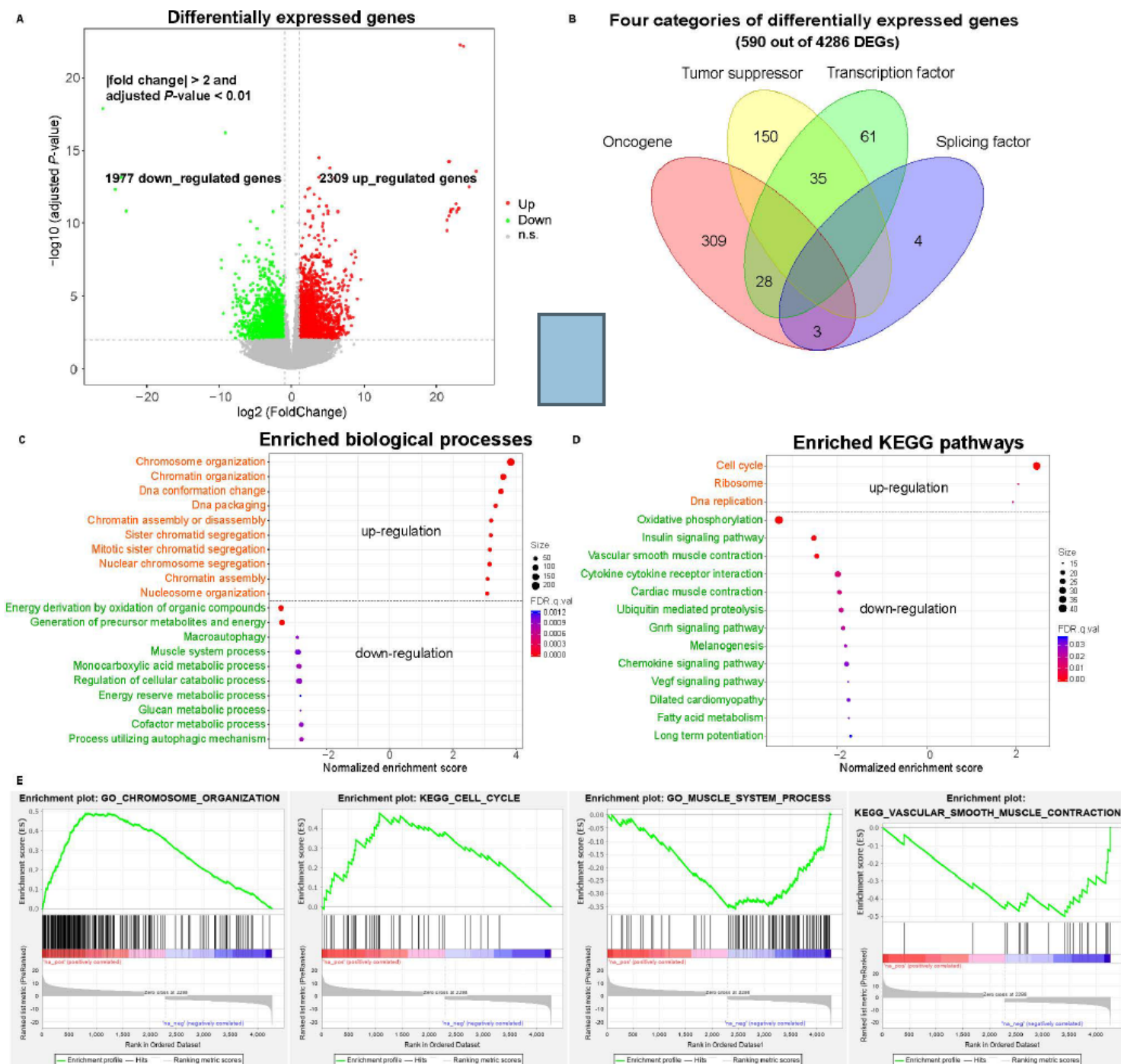


Figure 1

Differential expression profiles between SS and adjacent normal tissues. A. Volcano plot of differentially expressed genes (DEGs) between ten pairs of SS and adjacent normal samples. $|\text{fold change}| > 2$ and adjusted $P\text{-value} < 0.01$. B. Different categories of DEGs. C. Top enriched up-regulated and down-regulated biological processes of DEGs. D. Top enriched up-regulated and down-regulated KEGG pathways for DEGs. E. Examples of enriched biological processes and pathways. Adjusted $P\text{-value} < 0.05$.

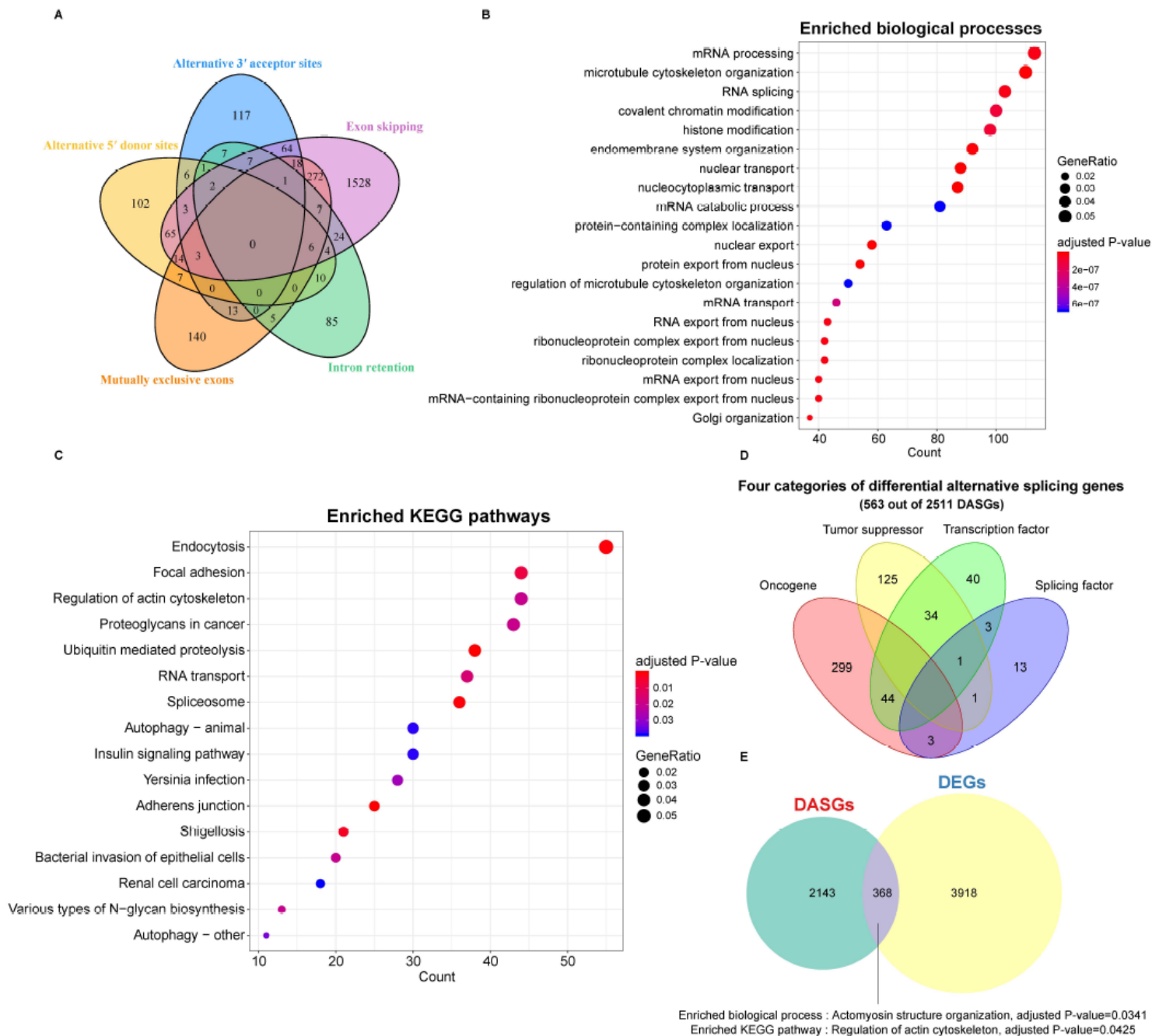


Figure 2

Alternative splicing profile of SS. A. Venn plot of the differentially alternative spliced genes (DASGs) for five common splicing modes. B. Top 20 enriched biological processes of DASGs. FDR < 0.05. C. Top 20 enriched pathways for DASGs. Adjusted P-value < 0.05. D. Distinct categories of DASGs. E. Comparison between DEGs and DASGs.

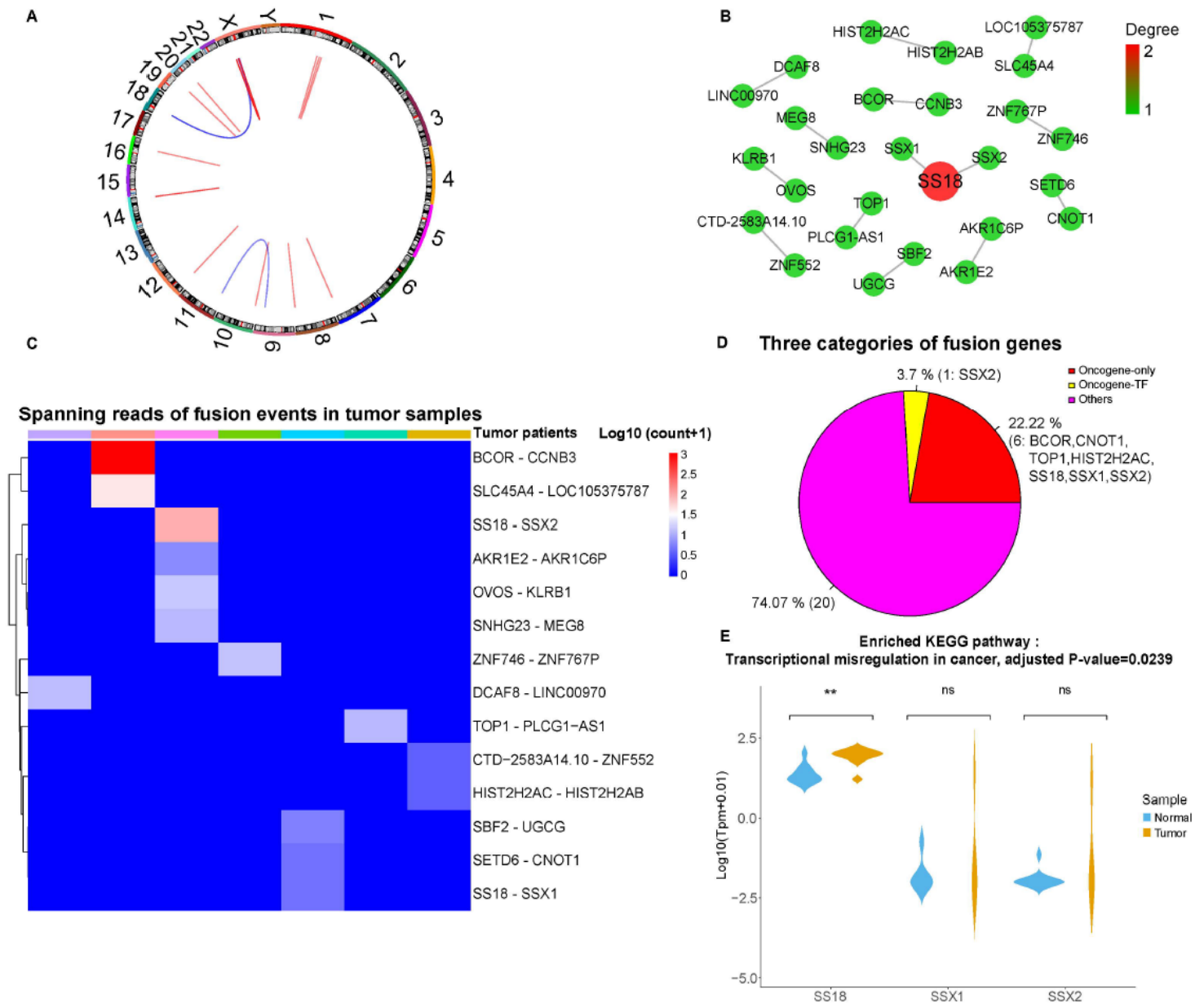


Figure 3

Gene fusion profile of SS. A. Circos plot for the 14 tumor-specific fusion pairs in SS. B. Network of tumor-specific fusion genes. The size and color of each circle correspond to the degree of fusion edges. C. Heatmap of the supporting junction reads for tumor-specific gene fusions. D. Different categories of the fusion genes. E. Expression profile of the fusion genes involved in significantly enriched biological processes. Adjusted P-value < 0.05.

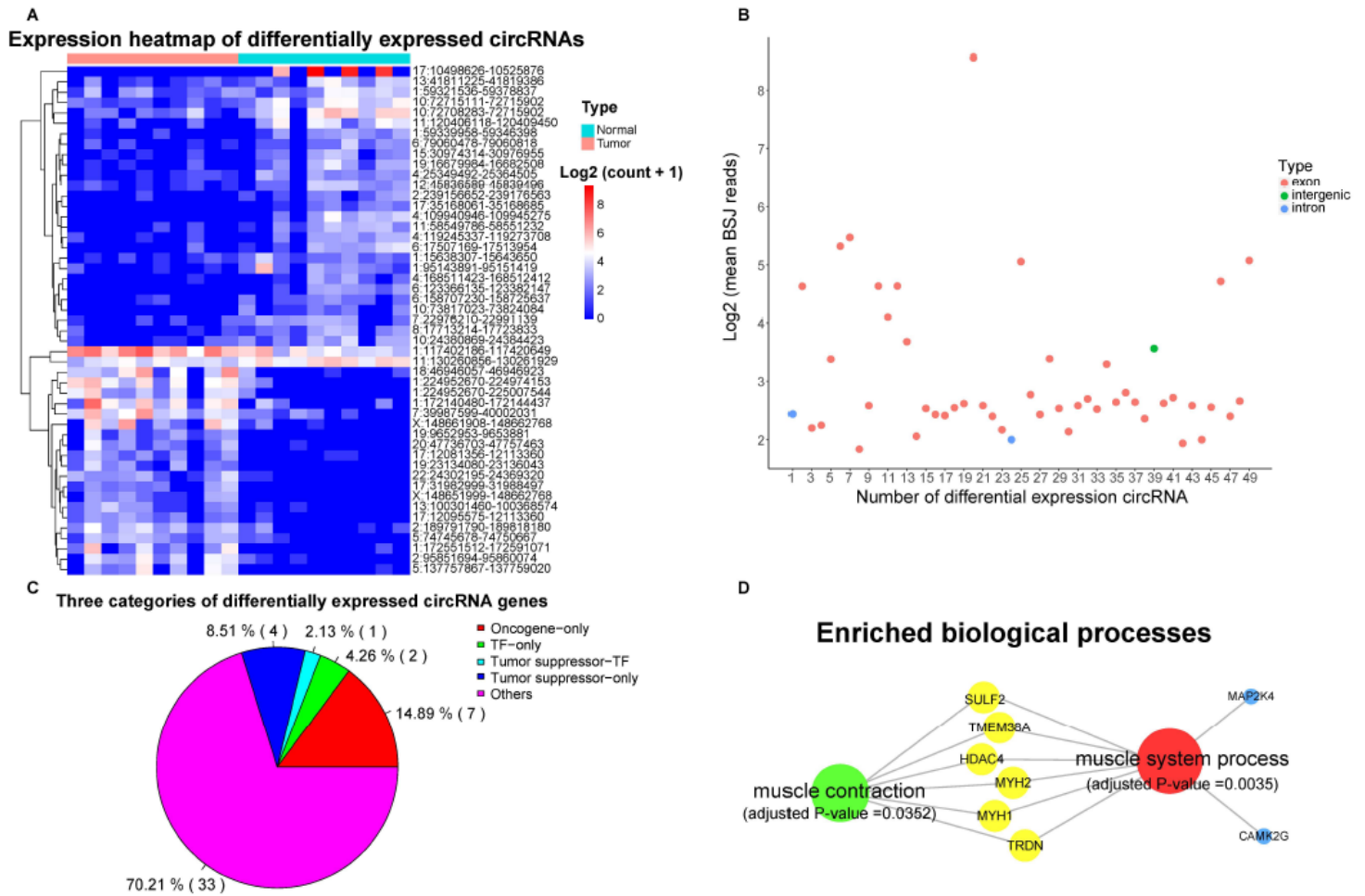


Figure 4

Expression profile and characteristics of circRNAs. A. Expression heatmap of differentially expressed circRNAs. Supporting reads of circRNAs were used in heatmap. B. Genomic region of the differentially expressed circRNAs. BSJ: back-spliced junction. C. Distinct types of the parental genes of differentially expressed circRNAs. D. Significantly enriched biological processes for the parental genes of differentially expressed circRNAs. Adjusted P-value < 0.05.

DNM3OS, ZNF730, DNAH14, AFF2, TMEM38A, TMEM56, CAP2, SYTL3, MCU, KATNAL2

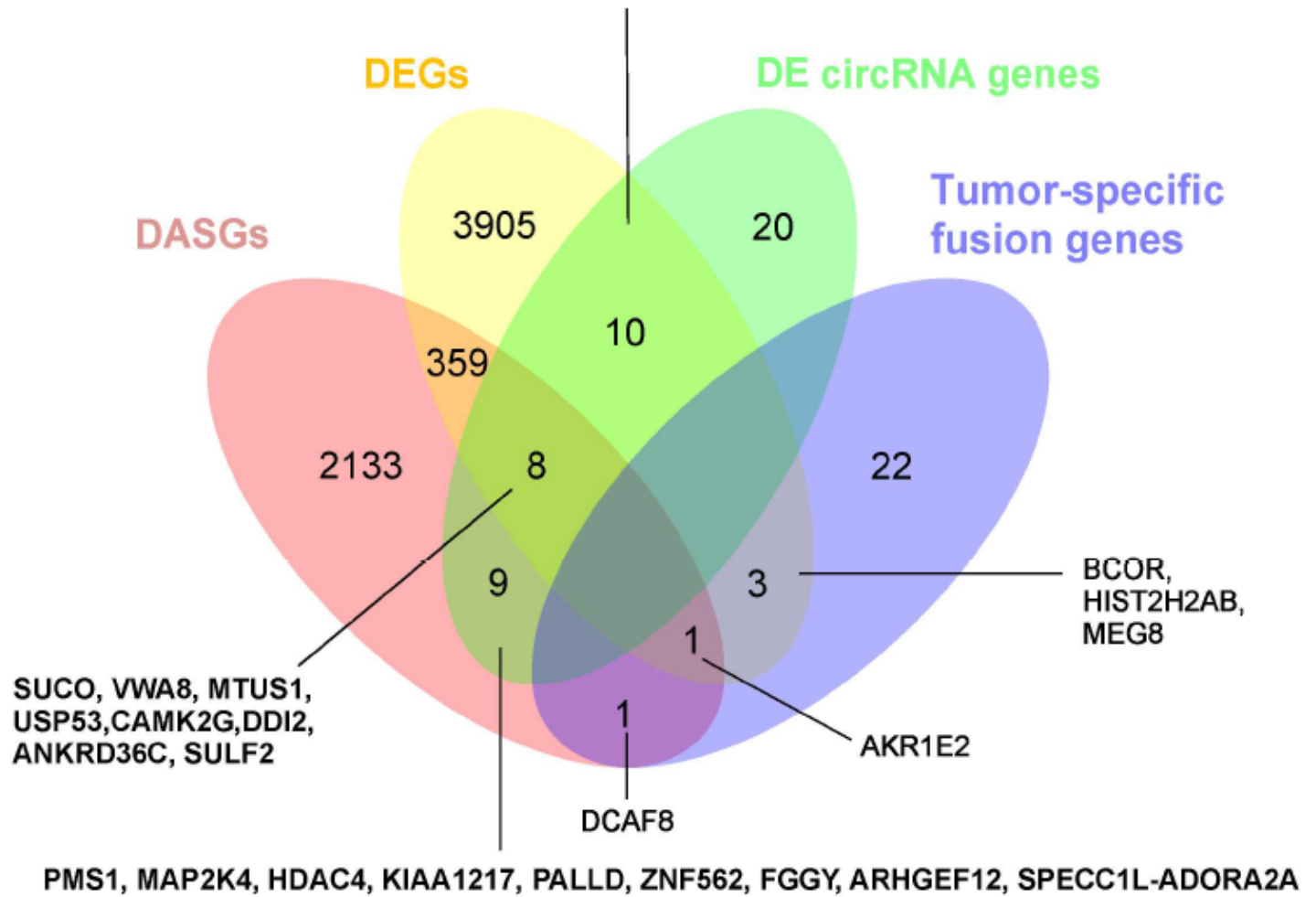


Figure 5

Comparison of DEGs, DASGs, fusion genes, and the parental genes of differentially expressed circRNAs.

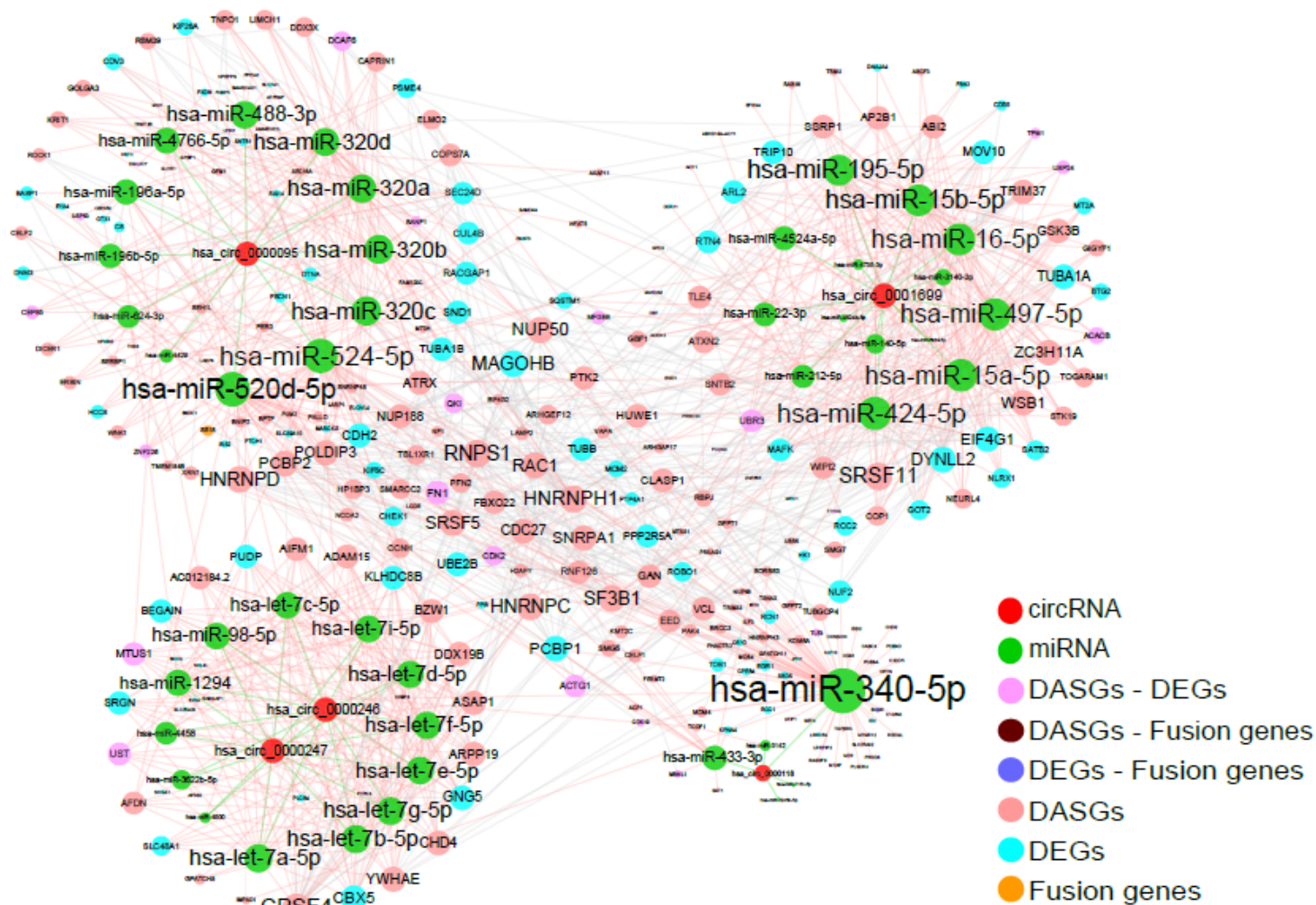


Figure 6

Interaction network among circRNAs, miRNAs, and related target genes. The size of each node is in proportion to the degree of edges. Only the target genes of DEGs, DASGs, fusion genes, and the parental genes of differentially expressed circRNAs were considered for corresponding miRNAs.

Supplementary Files

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

- [SupplementaryTableS1.csv](#)
- [SupplementaryTableS2.csv](#)
- [SupplementaryFigureS1.pdf](#)
- [SupplementaryFigureS2.pdf](#)