Contribution Value：A Indicator for Measuring the Contribution of ncRNAs to Transcriptome

Xinyi Gu
Peking University People's Hospital
Bo Jin
Peking University People's Hospital
Zhidan Qi
Peking University People's Hospital
Xiaofeng Yin (xiaofengyin@bjmu.edu.cn)
Peking University People's Hospital https://orcid.org/0000-0001-9932-642X

Research article

Keywords: ncRNA, Contribution value, Gene ontology, KEGG,whole transcriptome sequencing

Posted Date: June 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-32442/v1

Abstract

Background The expression difference multiple log2 FC and P value are often the main basis for screening ncRNA after high-throughput sequencing. However, the above two indicators can't well reflect the regulatory effect of nonprotein coding RNA (ncRNA) on mRNA. Therefore, we propose a new indicator, Contribution value (C value), to characterize the contribution of ncRNAs to transcriptome transformation.

Results In this study, we analyzed multiple data sets from mice and humans. We take all differential expression mRNAs as a parent set and the GO and KEGG enrichment analysis were performed on it. C value can be simply regarded as the sum of the product of the richfactor of each ncRNA target genes participating in each term/pathway and the P value of that term/pathway obtained from the parent set. We found that C value was superior to log2 FC and P value in all operation results.

Conclusions We show that the C value, which takes into accounts the the KEGG pathways and GO terms involved in the development of the disease, provides a measure of another dimension compared with the log2FC and P value. These hidden interactions between ncRNAs and their target genes may provide more comprehensive analysis.

Background

RNA sequencing is an extremely sensitive method for analyzingdifferential expression RNA[1]. Compared with traditional capillaryelectrophoresis sequencing, the massive parallel sequencingprovides more data at a lower cost. In contrast to genemicroarrays, they are not limited to known RNAs, but can also beused to analyze the entire transcriptome of tissues and organs,thus providing biological information about the possible functionof annotation genes or new genes [2]. Since the beginning of theHuman Genome Project, the application of high-throughput sequencinghas developed rapidly. Sequencing techniques have been used tostudy the characteristics of dynamic genomic loci ranging fromsimple model organisms to bigger species such ashumans[3, 4].

One of the most important applications of RNA sequencing is tocompare the differences in the expression of the non-codingRNAs (ncRNAs). NcRNAs refers to a kind of RNAs that can betranscribed from genome but not translated into proteins and canperform their own biological functions at the RNA level, includingrRNA, tRNA, snRNA, lncrna, microRNA and others. They play importantroles in normal development, physiology and disease [5]. Comparedwith other ncRNAs, microRNAs (miRNAs) and long noncoding RNAs(lncRNAs) are involved in regulating gene expression in manybiological processes. MiRNAs are non-coding single-stranded microRNAs, ranging from about 20 to 30 nucleotides in length. MiRNAs areestimated to regulate the translation of more than 60% ofprotein-coding genes [6]. Some miRNAs degrade mRNAs by fullybinding to a target, or inhibit the expression of proteintranslation regulatory genes by partially binding mRNAs [7]. OthermiRNAs may act as master regulators of a process. And lncRNAs, as adifferent type of non-coding RNA, regulate at both thetranscriptional and posttranscriptional levels [8-10], the ceRNAhypothesis suggests that lncRNA regulates miRNAs-induced

genesilencing by competitively binding micro RNA response elements [11,12]. By direct or indirect means, single miRNA or lncRNA canregulate hundreds of mRNAs.

High throughput sequencing is a common method for ncRNAresearch. People often select genes with high expressiondifferences for follow-up function research [13, 14]. In thistraditional way, the expression difference multiple log2 FC and Pvalues are often the main basis for screening ncRNA, onlyconsidering the change of expression amount. Based on the powerfultargeting function of ncRNA to mRNA, we think that a new index canbe proposed to characterize the contribution of genes totranscriptome transformation, this method can assist the selectionof target gene after high-throughput sequencing.

We collected the existing sequencing results, including skeletalmuscle denervation, Alzheimer's disease, prostate cancer, gastriccancer, and adipocyte differentiation. C57BL / 6 mice were used asthe model of skeletal muscle denervation, APP / PS1 mice as themodel of Alzheimer's disease, prostate cancer, gastric cancer, andadipocyte differentiation samples were all from human[15-19]. Foreach sequencing result, we take all DE mRNAs as a parent set andthe GO and KEGG enrichment analysis were performed on it. Then, thepredicted target mRNAs of each DE ncRNA (miRNA and lncRNA) and DEmRNAs were cross-labeled as a subset, which was also used for theGO and KEGG enrichment analysis. Finally, C value can be simplyregarded as the sum of the products of the richfactor of each ncRNAtarget genes participating in each term/pathway and the P value ofthat pathway obtained from the parent set. The richfactor is equalto the proportion of the number of each subset participating ineach term/pathway to the total number of genes in thatterm/pathway.

Our proposed mathematical model for calculating C value for eachncRNA takes into accounts the P value for each enriched GO term andKEGG pathway and the percentage of ncRNAs target genes in thatpathway. We expect to use C value to characterize the contributionof genes to transcriptome transformation, and to assist in theselection of target genes after high throughput sequencing.

Results

The total C value of each DE ncRNA is equal to the sum of BPvalue, CC value, MF value and KEGG value

We calculated the Cvalue of eachDE miRNA in skeletal muscledenervation, Alzheimer's disease, prostate cancer andgastric cancerdata setsrespectively.In addition, we calculate the C value of eachlncRNA in skeletal muscle denervation and adipocyte differentiationdata sets.Using the mathematical modelwe designed, the C values ofeach DE miRNA based on biological process (BP), cellular component(CC), molecular function (MF) and KEGG analysis can be obtained,and we call these C values as BP value, CC value, MF value and KEGGvalue respectively. The total C value of each DE miRNA is equal tothe sum of BP value, CC value, MF value and KEGG value. The DEmiRNAs were sorted with the total C value to obtain the 10 DEmiRNAs with maximum C value, named as top10 C value miRNAs (Table1-4). The top10 DE miRNAs with maximum absolute Log2 FC (top10 FCmiRNAs), and the top10 DE miRNAs with minimum P

value (top10 Pvalue miRNAs), were obtained by sorting the DE miRNAs according tothe absolute Log2 fold FC and P value respectively(TableS1-8).Similarly, DE lncRNAs are processed in the same way to obtaintop5 C value lncRNAs, top5 FC lncRNAs, top5P value lncRNAs foradipocyte differentiation and top10 C value lncRNAs, top10 FClncRNAs, top10 P value lncRNAsfor skeletal muscle denervation.(Table S9-14).

Table 1
The top10miRNAs according to C valueinskeletal muscledenervation

miRNAs	KEGG value	BP value	CC value	MF value	C value
mmu-miR-1943-5p	33.22977743	816.2096329	57.79710246	86.48301468	993.7195275
mmu-miR-322-5p	30.84058475	752.9168146	68.55345126	79.59753997	931.9083906
mmu-miR-497a-5p	30.73416222	748.7865715	69.70753712	79.46588326	928.6941541
mmu-miR-674-5p	27.16055299	715.4413807	58.01127153	72.71039755	873.3236028
mmu-miR-377-3p	27.49012205	693.5040261	53.27285114	72.83266223	847.0996615
mmu-miR-378d	23.25957428	680.9892711	61.28062951	72.28974851	837.8192234
mmu-miR-486a-3p	26.82482408	657.0154824	50.6865571	69.48345827	804.0103219
mmu-miR-34a-5p	26.69878945	659.2445359	53.42728799	63.0868827	802.457496
mmu-miR-34c-5p	26.69878945	659.2445359	53.42728799	63.0868827	802.457496
mmu-miR-485-5p	24.79976611	631.6503987	56.88385135	69.472899	782.8069152
BP, biological process;CC,cellularcomponent;MF, molecular function					

Table 2
The top10miRNAs according to C value inAlzheimer's disease

miRNAs	KEGG value	BP value	CC value	MF value	C value
mmu-miR-340-5p	43.52080776	1010.03906	99.40722008	95.12478643	1248.091874
mmu-miR-128-3p	32.34055348	702.2784936	72.03995682	72.09752374	878.7565277
mmu-miR-1912-3p	31.48177854	665.3035883	71.02378772	65.40238219	833.2115368
mmu-miR-3065-5p	28.57251324	635.0080827	59.73887501	60.19657486	783.5160458
mmu-miR-30e-5p	25.07905102	603.9771948	61.33651101	55.31973524	745.712492
mmu-miR-30b-5p	24.31969616	578.0747339	60.54628078	54.11557602	717.0562868
mmu-miR-369-3p	21.98379268	578.5994009	50.71409202	53.23059546	704.5278811
mmu-miR-30f	23.94953791	503.5940067	55.68167458	48.96495449	632.1901737
mmu-miR-16-5p	24.36382475	493.9211057	47.41829077	46.52040221	612.2236234
mmu-miR-3470a	18.44946604	405.6942042	42.64795188	40.23637553	507.0279976
BP, biological process; CC,cellularcomponent; MF, molecular function					

Table 3
The top10miRNAs according to C value inprostate cancer

miRNAs	KEGG value	BP value	CC value	MF value	C value
hsa-miR-374a-5p	4.398464287	118.2692853	5.213255954	10.04399186	137.9249974
hsa-miR-513a-5p	5.657237644	112.0295308	6.910295123	12.59297097	137.1900346
hsa-miR-95-5p	3.466944549	116.9227758	5.477859854	9.568881404	135.4364617
hsa-miR-374b-5p	3.8075956	113.5989789	5.300186874	11.67343102	134.3801924
hsa-miR-498	4.728130359	107.2751254	5.824931497	10.83476997	128.6629573
hsa-miR-20a-5p	4.115587199	109.1116054	5.632825474	8.080791708	126.9408098
hsa-miR-30e-5p	3.611723927	102.8695362	5.273755273	8.075628363	119.8306438
hsa-miR-96-5p	3.053742658	94.34293044	5.100190161	6.580687364	109.0775506
hsa-miR-148a-5p	3.391802004	90.10482282	4.519060515	6.899577071	104.9152624
hsa-miR-429	3.2432555	85.75354374	5.037021005	7.643002043	101.6768223
BP, biological process; CC,cellularcomponent; MF, molecular function					

Table 4
The top10miRNAs according to C value ingastric cancer

miRNAs	KEGG value	BP value	CC value	MF value	C value
hsa-miR-153-5p	18.23914759	362.6949521	64.02355601	71.76595514	516.7236109
hsa-miR-3662	15.39458508	317.1733239	49.40052706	52.85782946	434.8262655
hsa-miR-548f-3p	14.21781796	286.8668432	49.40872076	47.79869518	398.2920771
hsa-miR-5680	13.27934194	242.4315078	42.30706957	49.89122698	347.9091463
hsa-miR-944	14.78584485	239.0950535	40.7592345	49.77751375	344.4176466
hsa-miR-7-2-3p	13.34378972	249.0496147	38.59579289	38.19653997	339.1857373
hsa-miR-4677-5p	8.419400634	187.5755651	34.26195456	30.58636571	260.843286
hsa-miR-20a-5p	7.557754765	178.5451401	36.14268786	28.25566276	250.5012455
hsa-miR-4728-5p	10.00609667	161.1988844	32.79039421	31.1539595	235.1493347
hsa-miR-6507-5p	10.37776368	162.008939	28.4527209	26.25126198	227.0906856
BP, biological process; CC,cellularcomponent; MF, molecular function					

C value is superior to log2 FC and P value in miRNAsoperation results

In each data set, the most significant enriched BP term, CCterm, MF term, KEGG pathway and the most involved BP term, CC term,MF term, KEGG pathway were obtained by DEmRNAs enrichmentanalysis.We took the intersections of DE mRNAs with the predictedtarget genes of top10 C value miRNAs, top10 FC miRNAs and top10 Pvalue miRNAs respectively, and then calculated the proportion ofthese intersections in the above pathways/terms. It was found thatthe proportion of top10 C Value miRNAs target mRNAs wassignificantly larger than that of top10 FC miRNAs, top10 P valuemiRNAs in the above terms/pathways (Fig.1). Furthermore, weenriched KEGG pathways based on DE mRNAs and pathways with p valueless than 0.01 to generate an annotation network (Fig.2). Accordingto the distance from the centre, the annotation network was dividedinto three regions: core region, subcore region, and non-coreregion (Fig. 2). In the annotation network, the predicted targetgenes of top10 C value miRNAs, top10 FC miRNAs and top10 p valuemiRNAs were labeled in red (Fig. S1). It was found that the totalnumber of top10 C value miRNAs' target genes and their proportionin each region were larger than those of top10 FC miRNAs, and top10P value miRNAs (Fig.3).

Based on extensive literature, we identified 14 skeletal musclegrowth regulatory miRNAs, 6 Alzheimer's disease associated miRNAs,7prostate cancer associated miRNAs, 6gastric cancer associatedmiRNAsand found that when DE miRNAs were sorted by C value, thesequence number accumulation value of thesemiRNAs was significantlysmaller than that of the other two indexes, which means that thesemiRNAs sequences increased integrally (Fig. 4). When sorting by Cvalue versus sorting by absolute Log2 FC/ P value, most of thedisease critical miRNAs ranked up(Fig. 4).

C value is superior to log2 FC and P value in lncRNAsoperation results

We got a conclusion similar to miRNA in the result of lncRNAoperation based on skeletal muscle denervation and adipocytedifferentiation data sets. In skeletal muscle denervationdata set,we calculated the proportion of the predicted target genes of top10C value lncRNAs, top10 FC lncRNAs, and top10 P value lncRNAs in the7 most enriched terms/pathways respectively, and found that theproportion of the genes regulated by top10 C value lncRNAs waslarger than that of top10 FC lncRNAs and top10 P value lncRNAs(Fig. 5A). Then, the predicted target genes of top10 C valuelncRNAs, top10 FC lncRNAs and top10 P value lncRNAs were labeled inred in KEGG annotation network(Fig. 5B-D). It was found that thetotal number of top10 C value lncRNAs' target genes and theirproportion in each region (total: 87, core region: 17.9%, subcoreregion: 17%, non-core region: 19.7%) were larger than those oftop10 FC lncRNAs (total: 65, core region: 11.5%, subcore region:12.7%, non-core region: 16.9%), and top10 P value lncRNAs (total:61, core area: 10.3%, subcore area: 11.8%, non-core area: 16.9%)(Fig. 5E-F).

Since there are relatively few DElncRNAs and DE mRNAsinadipocyte differentiation data set, we take top5C value lncRNAs,top5 FC lncRNAs, top5 P value lncRNAs and draw the KEGG networkwithout partition (Fig. S2). The proportion of the genes regulatedby top5 C value lncRNAs was larger than that of top5 FC lncRNAs andtop5 P value lncRNAs in GO terms and KEGG annotation network (Fig.6A-B). And when DE lncRNAs were sorted by C value, theadipocytedifferentiation associatedlncRNAs sequences increased integrallythan that of the other two indexes (Fig. 6C-D).

Discussion

MiRNAs play an important role in transcriptome regulation andthey can directly or indirectly influence the biological processesbefore and after gene transcription[20-22]. LncRNA (> 200nucleotides) has recently been studied as endogenousncRNA, whichhas multiple mRNA characteristics, including polyadenylation [23].Recent studies have shown that lncRNA, as a different type ofnoncoding RNA, can be classified according to genome location ortranscription direction, and regulate gene expression[24].If we canquantify this regulatory capacity of ncRNAs, the results mustin turn indicate how much miRNAs are involved in a particularbiological process. We designed a new mathematical model tocalculate Contribution value, and tried to use this index tocharacterize the contribution of each DE ncRNA to genome-widechanges in total transcriptome sequencing results. C value can besimply regarded as the sum of the product of the richfactor of eachncRNA target genes participating in each pathway and the P value ofthat pathway obtained from the parent set. The richfactor is equalto the proportion of the number of each subset participating ineach pathway to the total number of genes in that pathway. And theparent set is a collection of all DE mRNAs.

To test the superiority of C value as a measure of ncRNAscontribution to genome-wide changes, we compared it with absoluteLog2 FC and P value. Log 2 FC reflects the expression change ofncRNAs and P value reflects how significant the change is. The twoindexes of each DE RNA were obtained after the traditional wholetranscriptome sequencing, and many follow-up studies have partiallyreferenced Log2 FC

and P values in selecting the target gene[13,14]. As miRNAs, we compared C value with Log2 FC, P value, andfound that the proportion of top10 C value miRNAs target mRNAs wassignificantly larger than that of top10 FC miRNAs, top10 P valuemiRNAs in the GO terms and KEGG pathway. Further, we analyzed theannotation network of KEGG pathways with P value less than 0.01 andfound that the total number of top10 C value miRNAs' target genesand their proportion in each region were larger than those of top10FC miRNAs, and top10 P value miRNAs. At the same time, we searcheda large number of literatures and got 14 skeletal muscle growthregulatory miRNAs[25-38], 6 Alzheimer's disease associatedmiRNAs[39-42], 7 prostate cancer associated miRNAs[17, 43, 44],6gastric cancer associated miRNAs[45, 46].Their overall rankingwent up after sorted by C value. For lncRNAs, there were similarresults. Ranking according to C value, top lncRNAs showed strongerregulation on GO terms and KEGG pathways. Moreover, the proportionof top C value lncRNAs' target genes in the core region of KEGGannotation network is also larger than that of top lncRNAs selectedby other two indicators. And adipocyte differentiationassociatedlncRNAs[19] ranked up after sorted by C value.

Conclusions

Based on the above evidence and the rationality of themathematical model, Contribution value can be used as an indicatorto measure the contribution of ncRNA to the transcriptome. In theanalysis of whole transcriptome sequencing, C value, as asupplement to log2 FC and P value, makes the selection of targetgene more reasonable.

Methods

Prediction of ncRNAs' target mRNAs

MiRNA: MiRNAs target genes prediction software,miRanda(http://www.microrna.org/)[47], uses a weighted dynamicprogramming algorithm to calculate the optimal sequencecomplementarity between a mature microRNA and a given mRNA. The keyextension of Smith-Waterman algorithm is that the alignment scoreis the weighted sum of the matching and mismatching scores of basepairs (including G: U jitter) and gap penalties. Weights areposition-dependent, reflecting the relative importance of the 5'and 3' regions in a finely adjustable manner. The weight of eachposition can be optimized to reflect experimental facts andphysical principles.

LncRNA: The target genes of lncRNAs are predicted by expressioncorrelation analysis or co-expression analysis of lncRNA and mRNAamong samples. The Weighted Gene Correlation NetworkAnalysis(http://www.r-project.org/)[48] wasused to calculate Pearson correlation coefficients.The absolutevalue of the Pearson correlation coefficient≥0.90, p-value <0.01and FDR <0.01 was saved.

GO and KEGG pathway enrichment analysis

Gene Ontology Analysis: GO is a database established by GeneOntology consortium(http://www.geneontology.org),which includes three parts: molecular function, biological

processand cell composition. Fisher exact test and x2 test were used.Enrichment analysis of differentially expressed genes or ncRNAstarget genes was performed using GOseq R software package, and genelength bias was corrected. The corrected P value less than 0.05 wasconsidered to be significantly enriched by differentially expressedgenes.

Pathway Analysis: Based on the Kyoto Encyclopedia of Genes andGenomes (KEGG) database(http://www.genome.ad.jp/kegg/),we used Kobas software(http://kobas.cbi.pku.edu.cn)[49]to test the statistical enrichment of differentially expressedgenes or ncRNAs target genes, and Fisher precision test and x2 testwere used.

Using hypergeometric analysis, the pathway of KEGG/GO was foundto be significantly enriched in the candidate gene compared withthe whole genome background. The formula for this analysis:

N is the total number of genes in the GO annotation database; Mis the total number of genes in the database belonging to a GOsubclass; n is the total number of genes in the database thatrequire GO enrichment analysis; i is the number of genes in nbelonging to M.So, we can calculate the probability of whether thegene set n is enriched in class M. Theoretically, the smaller the Pvalue is, the higher the significance of pathway enrichment is.

In this study, GO and KEGG enrichment analysis were performed onall DE mRNAs which was called as the parent set. Then, thepredicted target genes of DE ncRNAs was cross-labeled with DEmRNAs, and the subsets were used for GO analysis(BP, CC, MF) andKEGG enrichment analysis.

C value mathematical model and itscalculation

The C value of each DE ncRNA is calculated using the followingmathematical model:

RichFactor is equal to the proportion of the number of eachsubset participating in each pathway to the total number of genesin that pathway; P value is the p value of the pathway in theparent set; n represents the number of pathways enriched bysubset.

The establishing of KEGG annotation network

KEGG enrichment analysis of DE mRNAs was performed usingcytoscape3.72(San Diego, CA, USA) with medium network specificity,showing only pathway with the P value less than 0.01.

Data Analysis

The analysis platform is R 3.6.1 and the R package isclusterProfiler. The database is org.Mm.eg.db developed with the Rpackage.

Abbreviations

KEGG：Kyoto Encyclopedia of Genes and Genomes；

GO：Gene Ontology；

FC：fold change；

ncRNA：nonprotein coding RNA；

miRNA ：microRNA；

lncRNA：long non-coding RNA；

mRNA：messenger RNA；

ceRNA：competing endogenous RNA；

BP：biological process；

CC：cellular component；

MF：molecular function.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included inthese published articles [and their supplementary informationfiles][15-19].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Beijing Municipal Natural ScienceFoundation [grant number 7192215 to XY]; the National NaturalScience Foundation of China [grant number 31471144 to XY].

Authors' contributions

XY: Conceptualization. XG: Methodology, Formal analysis,Investigation, Writing-Original Draft.BJ: Methodology, Formalanalysis, Investigation, Writing-Original Draft. ZQ:Writing-Original Draft. All authors have read and approved themanuscript.

Acknowledgements

We wish to thank all the members of the laboratory for thehelpful discussion.Thank Mr. Zhang Guangze for providing technicalguidance for this experiment.

References

1. Sirbu A, Kerr G, Crane M, Ruskin HJ: **RNA-Seq vs dual-and single-channel microarray data: sensitivity analysis fordifferential expression and clustering**. *PLoS One*2012, **7**(12):e50986.
2. Hurd PJ, Nelson CJ: **Advantages of next-generationsequencing versus the microarray in epigenetic research**.*Brief Funct Genomic Proteomic* 2009,**8**(3):174-183.
3. Nadal-Ribelles M, Islam S, Wei W, Latorre P, Nguyen M, de NadalE, Posas F, Steinmetz LM: **Sensitive high-throughputsingle-cell RNA-seq reveals within-clonal transcript correlationsin yeast populations**. *Nat Microbiol* 2019,**4**(4):683-692.
4. Yang W, Lu Z, Zhi Z, Liu L, Deng L, Jiang X, Pang L:**High-throughput transcriptome-Seq and small RNA-Seq revealnovel functional genes and microRNAs for early embryonic arrest inhumans**. *Gene* 2019,**697**:19-25.
5. Mercer TR, Dinger ME, Mattick JS: **Long non-coding RNAs:insights into functions**. *Nat Rev Genet* 2009,**10**(3):155-159.
6. He L, Hannon GJ: **MicroRNAs: small RNAs with a big rolein gene regulation**. *Nat Rev Genet* 2004,**5**(7):522-531.
7. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism,and function**. *Cell* 2004,**116**(2):281-297.
8. Kornienko AE, Guenzl PM, Barlow DP, Pauler FM: **Generegulation by the act of long non-coding RNAtranscription**. *BMC Biol* 2013,**11**:59.
9. Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R,Bonilla F, de Herreros AG: **A natural antisense transcriptregulates Zeb2/Sip1 gene expression during Snail1-inducedepithelial-mesenchymal transition**. *Genes Dev*2008, **22**(6):756-769.
10. Aravin AA, Hannon GJ, Brennecke J: **The Piwi-piRNApathway provides an adaptive defense in the transposon armsrace**. *Science* 2007,**318**(5851):761-764.
11. Peng H, Ishida M, Li L, Saito A, Kamiya A, Hamilton JP, Fu R,Olaru AV, An F, Popescu I *et al*: **PseudogeneINTS6P1 regulates its cognate gene INTS6 through competitivebinding of miR-17-5p in hepatocellular carcinoma**.*Oncotarget* 2015, **6**(8):5666-5677.
12. Wang L, Guo ZY, Zhang R, Xin B, Chen R, Zhao J, Wang T, Wen WH,Jia LT, Yao LB *et al*: **Pseudogene OCT4-pg4functions as a natural micro RNA sponge to regulate OCT4 expressionby competing for miR-145 in hepatocellular carcinoma**.*Carcinogenesis* 2013,**34**(8):1773-1781.

13. Zhang X, Wang S, Wang H, Cao J, Huang X, Chen Z, Xu P, Sun G,Xu J, Lv J *et al*: **Circular RNA circNRIP1 acts as amicroRNA-149-5p sponge to promote gastric cancer progression viathe AKT1/mTOR pathway**. *Mol Cancer* 2019,**18**(1):20.
14. Bai C, Gao Y, Zhang X, Yang W, Guan W: **MicroRNA-34cacts as a bidirectional switch in the maturation ofinsulin-producing cells derived from mesenchymal stemcells**. *Oncotarget* 2017,**8**(63):106844-106857.
15. Weng J, Zhang P, Yin X, Jiang B: **The WholeTranscriptome Involved in Denervated Muscle Atrophy FollowingPeripheral Nerve Injury**. *Front Mol Neurosci* 2018,**11**:69.
16. Ma N, Tie C, Yu B, Zhang W, Wan J: **IdentifyinglncRNA-miRNA-mRNA networks to investigate Alzheimer's diseasepathogenesis and therapy strategy**. *Aging (AlbanyNY)* 2020, **12**(3):2897-2920.
17. Ye Y, Li SL, Wang SY: **Construction and analysis ofmRNA, miRNA, lncRNA, and TF regulatory networks reveal the keygenes associated with prostate cancer**. *PLoS One*2018, **13**(8):e0198055.
18. Qi M, Yu B, Yu H, Li F: **Integrated analysis of a ceRNAnetwork reveals potential prognostic lncRNAs in gastriccancer**. *Cancer Med* 2020,**9**(5):1798-1817.
19. Chen K, Xie S, Jin W: **Crucial lncRNAs associated withadipocyte differentiation from human adipose-derived stem cellsbased on co-expression and ceRNA network analyses**.*PeerJ* 2019, **7**:e7544.
20. Bartel DP: **MicroRNAs: target recognition and regulatoryfunctions**. *Cell* 2009,**136**(2):215-233.
21. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R,Rajewsky N: **Widespread changes in protein synthesis inducedby microRNAs**. *Nature* 2008,**455**(7209):58-63.
22. Guo H, Ingolia NT, Weissman JS, Bartel DP: **MammalianmicroRNAs predominantly act to decrease target mRNAlevels**. *Nature* 2010,**466**(7308):835-840.
23. Quinn JJ, Chang HY: **Unique features of long non-codingRNA biogenesis and function**. *Nat Rev Genet* 2016,**17**(1):47-62.
24. Ma L, Bajic VB, Zhang Z: **On the classification of longnon-coding RNAs**. *RNA Biol* 2013,**10**(6):925-933.
25. Cheng X, Du J, Shen L, Tan Z, Jiang D, Jiang A, Li Q, Tang G,Jiang Y, Wang J *et al*: **MiR-204-5p regulates C2C12myoblast differentiation by targeting MEF2C and ERRgamma**.*Biomed Pharmacother* 2018,**101**:528-535.
26. Feng Y, Cao JH, Li XY, Zhao SH: **Inhibition of miR-214expression represses proliferation and differentiation of C2C12myoblasts**. *Cell Biochem Funct* 2011,**29**(5):378-383.
27. Ge G, Yang D, Tan Y, Chen Y, Jiang D, Jiang A, Li Q, Liu Y,Zhong Z, Li X *et al*: **miR-10b-5p Regulates C2C12Myoblasts Proliferation and Differentiation**. *BiosciBiotechnol Biochem* 2019, **83**(2):291-299.
28. Gan M, Du J, Shen L, Yang D, Jiang A, Li Q, Jiang Y, Tang G, LiM, Wang J *et al*: **miR-152 regulates theproliferation and differentiation of C2C12 myoblasts by targetingE2F3**. *In Vitro Cell Dev Biol Anim* 2018,**54**(4):304-310.
29. Huang Z, Chen X, Yu B, He J, Chen D: **MicroRNA-27apromotes myoblast proliferation by targeting myostatin**.*Biochem Biophys Res Commun* 2012,**423**(2):265-269.

30. Liu C, Wang M, Chen M, Zhang K, Gu L, Li Q, Yu Z, Li N, Meng Q:**miR-18a induces myotubes atrophy by down-regulatingIgfI**. *Int J Biochem Cell Biol* 2017,**90**:145-154.
31. Mi L, Li Y, Zhang Q, Zhao C, Peng Y, Yang G, Zheng X:**MicroRNA-139-5p regulates C2C12 cell myogenesis throughblocking Wnt/beta-catenin signaling pathway**. *BiochemCell Biol* 2015, **93**(1):8-15.
32. Qiu H, Zhong J, Luo L, Tang Z, Liu N, Kang K, Li L, Gou D:**Regulatory Axis of miR-195/497 and HMGA1-Id3 Governs MuscleCell Proliferation and Differentiation**. *Int J BiolSci* 2017, **13**(2):157-166.
33. Silva WJ, Graca FA, Cruz A, Silvestre JG, Labeit S, MiyabaraEH, Yan CYI, Wang DZ, Moriscot AS: **miR-29c improvesskeletal muscle mass and function throughout myocyte proliferationand differentiation and by repressing atrophy-relatedgenes**. *Acta Physiol (Oxf)* 2019,**226**(4):e13278.
34. Tang Z, Qiu H, Luo L, Liu N, Zhong J, Kang K, Gou D:**miR-34b Modulates Skeletal Muscle Cell Proliferation andDifferentiation**. *J Cell Biochem* 2017,**118**(12):4285-4295.
35. Wang H, Zhang Q, Wang B, Wu W, Wei J, Li P, Huang R:**miR-22 regulates C2C12 myoblast proliferation anddifferentiation by targeting TGFBR1**. *Eur J CellBiol* 2018, **97**(4):257-268.
36. Wang M, Liu C, Su Y, Zhang K, Zhang Y, Chen M, Ge M, Gu L, LuT, Li N *et al*: **miRNA-34c inhibits myoblastsproliferation by targeting YY1**. *Cell Cycle* 2017,**16**(18):1661-1672.
37. Wei X, Li H, Zhang B, Li C, Dong D, Lan X, Huang Y, Bai Y, LinF, Zhao X *et al*: **miR-378a-3p promotesdifferentiation and inhibits proliferation of myoblasts bytargeting HDAC4 in skeletal muscle development**. *RNABiol* 2016, **13**(12):1300-1309.
38. Zhang Z, Chen Y, Li B, Yao Y, Jiang A, Wei W, Liu H, Wu W:**Identification of a novel miR-206-Notch3 pathway regulatingmouse myoblasts proliferation**. *Gene* 2019,**695**:57-64.
39. Swarbrick S, Wragg N, Ghosh S, Stolzing A: **SystematicReview of miRNA as Biomarkers in Alzheimer's Disease**.*Mol Neurobiol* 2019, **56**(9):6156-6167.
40. Song Y, Hu M, Zhang J, Teng ZQ, Chen C: **A novelmechanism of synaptic and cognitive impairments mediated viamicroRNA-30b in Alzheimer's disease**. *EBioMedicine*2019, **39**:409-421.
41. Lei X, Lei L, Zhang Z, Zhang Z, Cheng Y: **DownregulatedmiR-29c correlates with increased BACE1 expression in sporadicAlzheimer's disease**. *Int J Clin Exp Pathol* 2015,**8**(2):1565-1574.
42. Yao X, Xian X, Fang M, Fan S, Li W: **Loss of miR-369Promotes Tau Phosphorylation by Targeting the Fyn andSerine/Threonine-Protein Kinase 2 Signaling Pathways in Alzheimer'sDisease Mice**. *Front Aging Neurosci* 2019,**11**:365.
43. Aghdam AM, Amiri A, Salarinia R, Masoudifar A, Ghasemi F,Mirzaei H: **MicroRNAs as Diagnostic, Prognostic, andTherapeutic Biomarkers in Prostate Cancer**. *Crit RevEukaryot Gene Expr* 2019, **29**(2):127-139.
44. Li F, Gu C, Tian F, Jia Z, Meng Z, Ding Y, Yang J:**MiR-218 impedes IL-6-induced prostate cancer cellproliferation and invasion via suppression of LGR4expression**. *Oncol Rep* 2016,**35**(5):2859-2865.
45. Yu B, Lv X, Su L, Li J, Yu Y, Gu Q, Yan M, Zhu Z, Liu B:**MiR-148a Functions as a Tumor Suppressor by TargetingCCK-BR via Inactivating STAT3 and Akt in Human GastricCancer**. *PLoS One*

2016,**11**(8):e0158961.

46. da Silva Oliveira KC, Thomaz Araujo TM, Albuquerque CI, BarataGA, Gigek CO, Leal MF, Wisnieski F, Rodrigues Mello Junior FA,Khayat AS, de Assumpcao PP *et al*: **Role of miRNAsand their potential to be useful as diagnostic and prognosticbiomarkers in gastric cancer**. *World JGastroenterol* 2016, **22**(35):7951-7962.
47. Betel D, Wilson M, Gabow A, Marks DS, Sander C: **ThemicroRNA.org resource: targets and expression**. *NucleicAcids Res* 2008, **36**(Databaseissue):D149-153.
48. Langfelder P, Horvath S: **WGCNA: an R package forweighted correlation network analysis**. *BMCBioinformatics* 2008, **9**:559.
49. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, LiCY, Wei L: **KOBAS 2.0: a web server for annotation andidentification of enriched pathways and diseases**.*Nucleic Acids Res* 2011, **39**(Web Serverissue):W316-322.

Figures

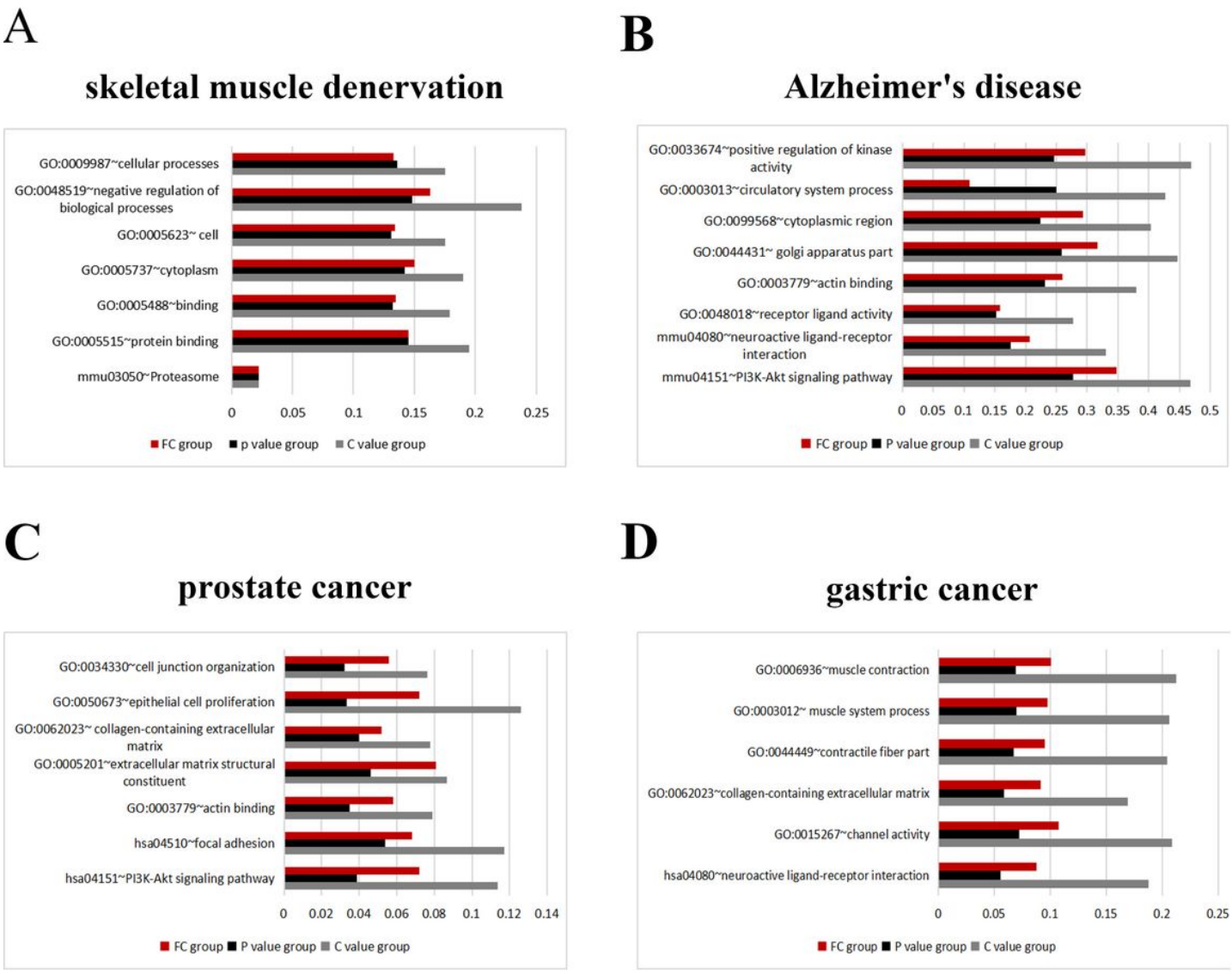

Figure 1

Proportion of three groups in each term/pathway(A) Skeletal muscle denervation. (B) Alzheimer's disease. (C) Prostate cancer. (D)Gastric cancer.(FC group: the collection of the top10 FC miRNAs' predictive target mRNAs; P value group: the collection of the top10 P value miRNAs' predictive target mRNAs; C value group: the collection of the top10 C value miRNAs' predictive target mRNAs).Picture drawn by Microsoft Excel.

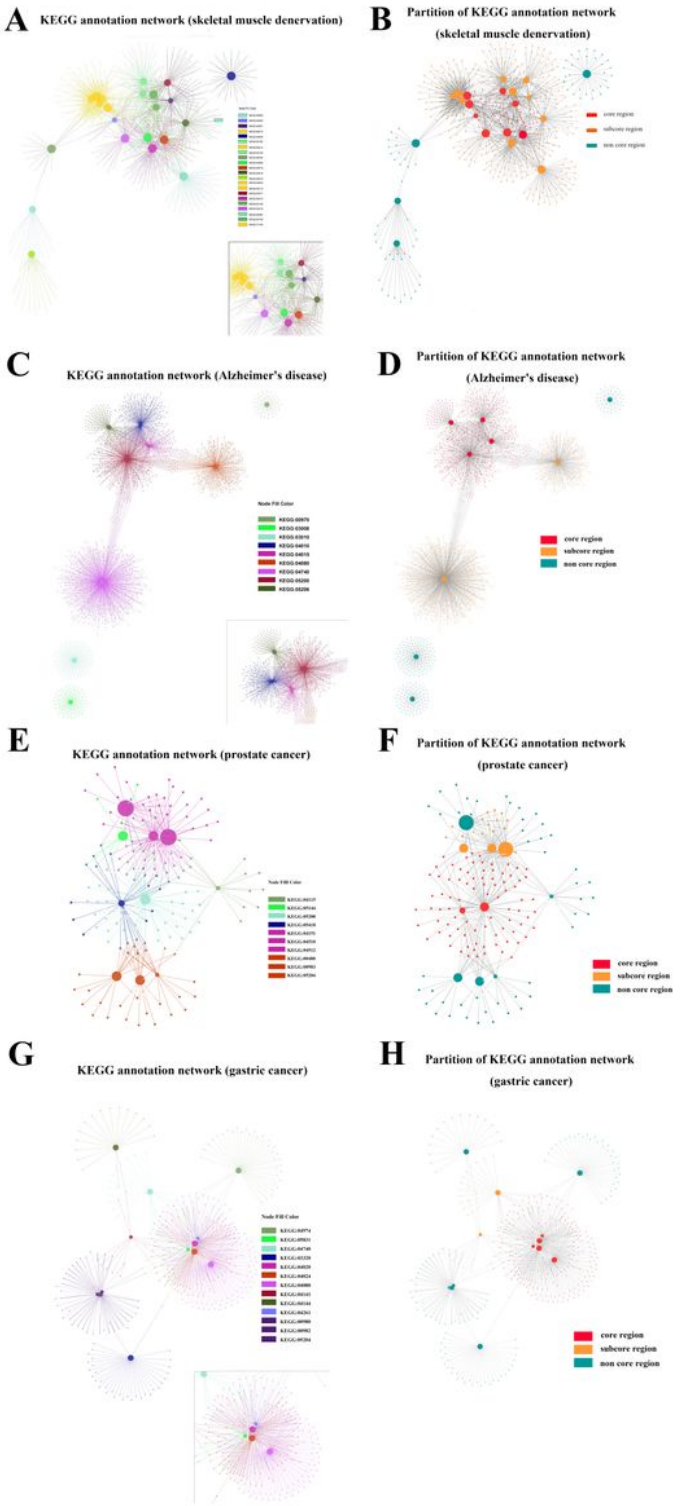

Figure 2

KEGG analysis results and regional division(A-B) Skeletal muscle denervation. (C-D) Alzheimer's disease. (E-F) Prostate cancer. (G-H)Gastric cancer. (Left: KEGG enrichment analysis of DE mRNAs; Right: according to the distance from the center region, the annotation network was divided into three regions).

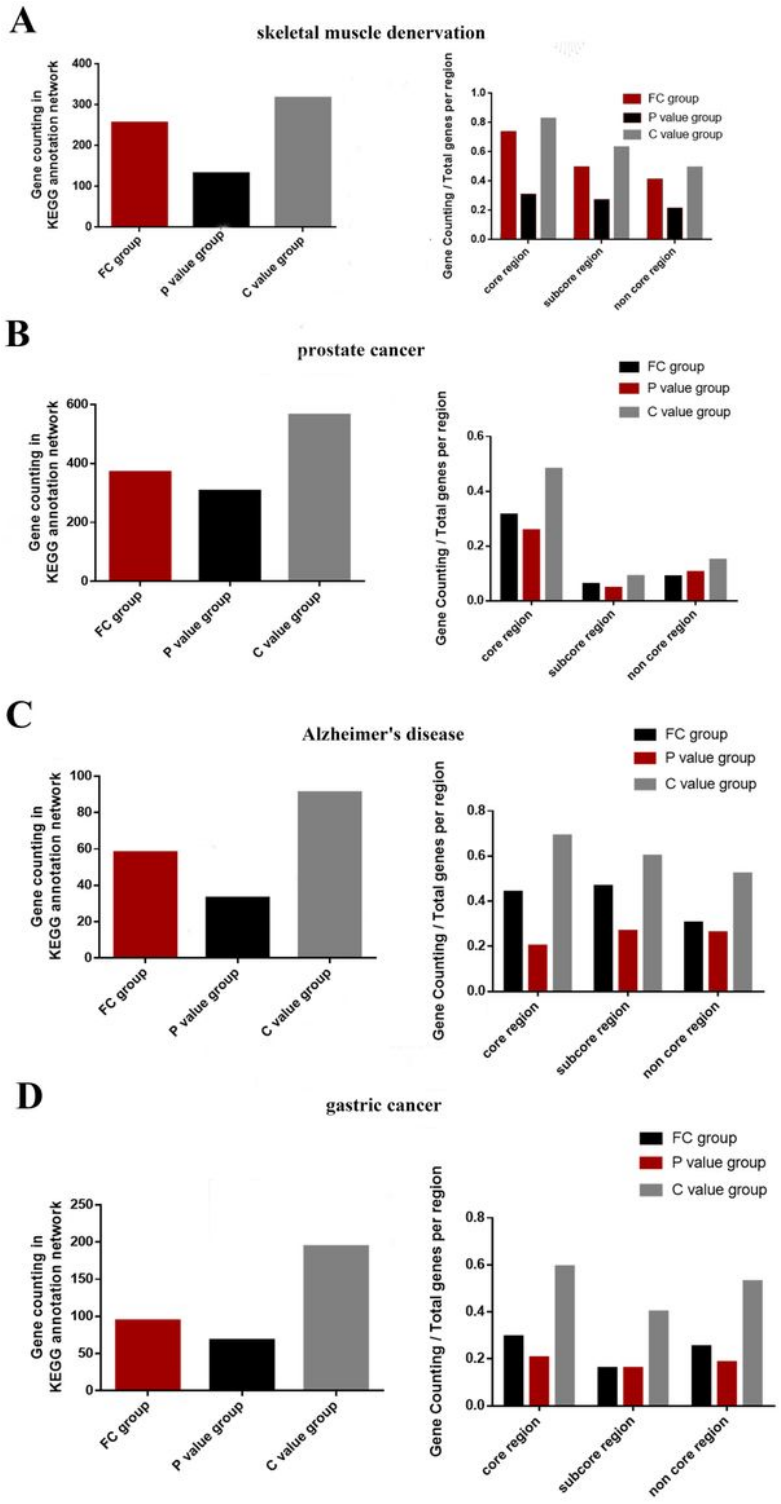

Figure 3

The total numbers of genes in annotation network and the proportion in each region.(A) Skeletal muscle denervation. (B) Alzheimer's disease. (C) Prostate cancer. (D)Gastric cancer.(FC group: the collection of the top10 FC miRNAs' predictive target mRNAs; P value group: the collection of the top10 P value miRNAs' predictive target mRNAs; C value group: the collection of the top10 C value miRNAs' predictive target mRNAs).

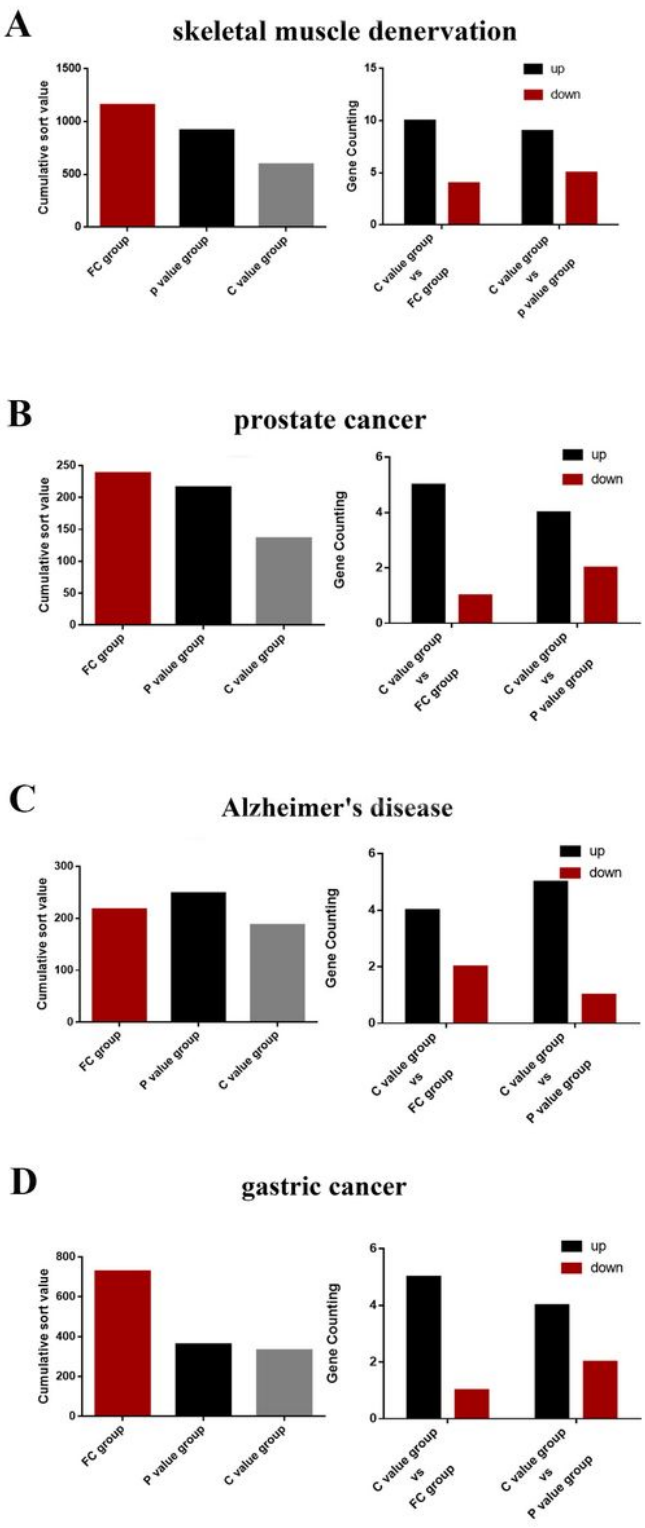

Figure 4

After sorting with C value, the ranking of disease critical miRNAs increased integrally.(A) Skeletal muscle denervation. (B)Alzheimer's disease.(C) Prostate cancer. (D)Gastric cancer. Left: Sort number accumulation value of disease critical miRNAs by the three indexes. Right: The number of mRNAs that rank up or down. (FC group: the collection of the top10 FC miRNAs' predictive target mRNAs; P value

group: the collection of the top10 P value miRNAs' predictive target mRNAs; C value group: the collection of the top10 C value miRNAs' predictive target mRNAs).

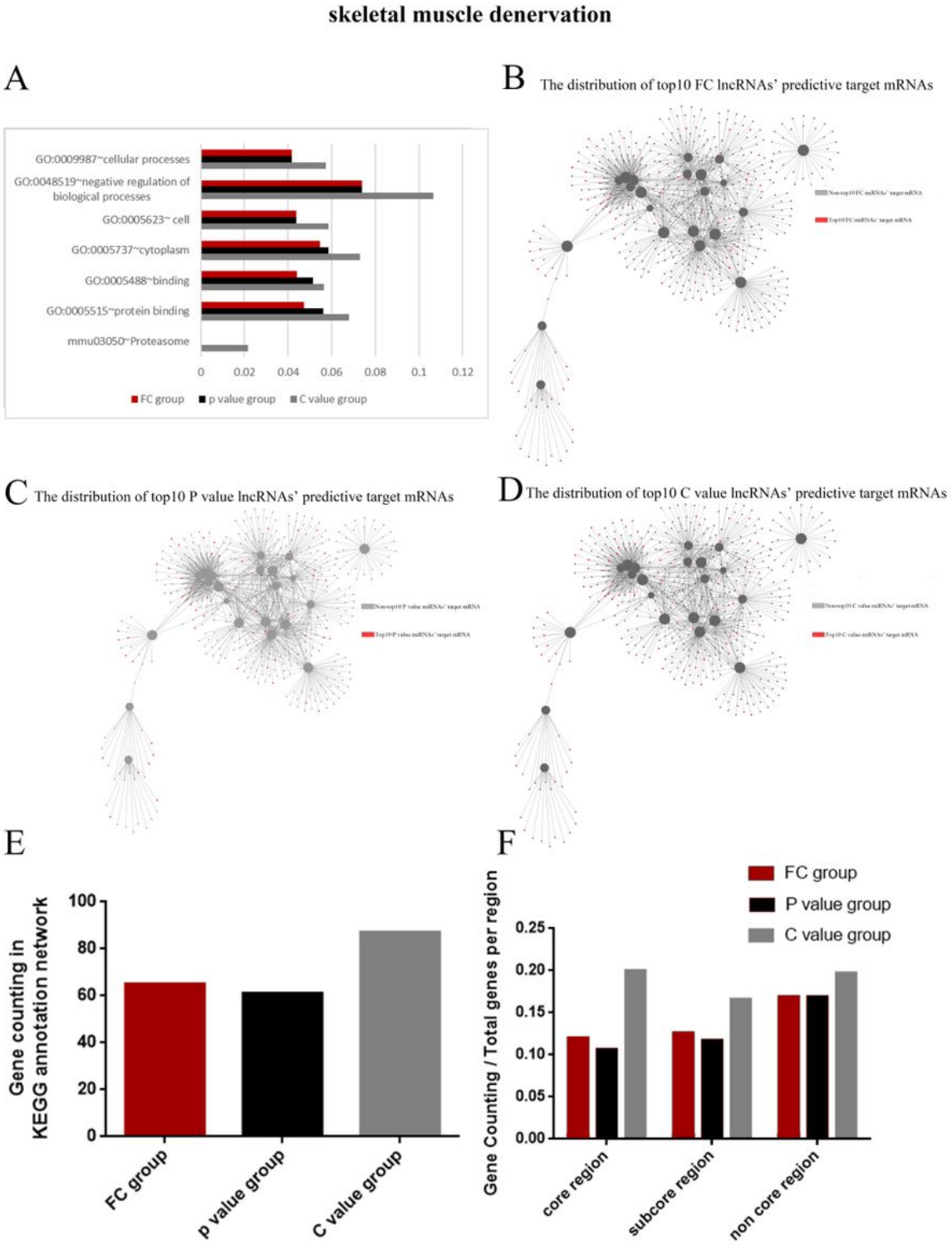

Figure 5

LncRNAs operation results forskeletal muscle denervation data set(A) The ratio of predicted target genes to the total genes in each term/pathway. (B) The distribution of top10 FC lncRNAs' predictive target mRNAs in the annotation network. (C) The distribution of top10 P value lncRNAs' predictive target mRNAs

in the annotation network. (D) The distribution of top10 C value lncRNAs' predictive target mRNAs in the annotation network. (E) The total numbers of mRNAs in the three groups in annotation network. (F) Proportion of three groups in each region. (FC group: the collection of the top10 FC lncRNAs' predictive target mRNAs; P value group: the collection of the top10 P value lncRNAs' predictive target mRNAs; C value group: the collection of the top10 C value lncRNAs' predictive target mRNAs)

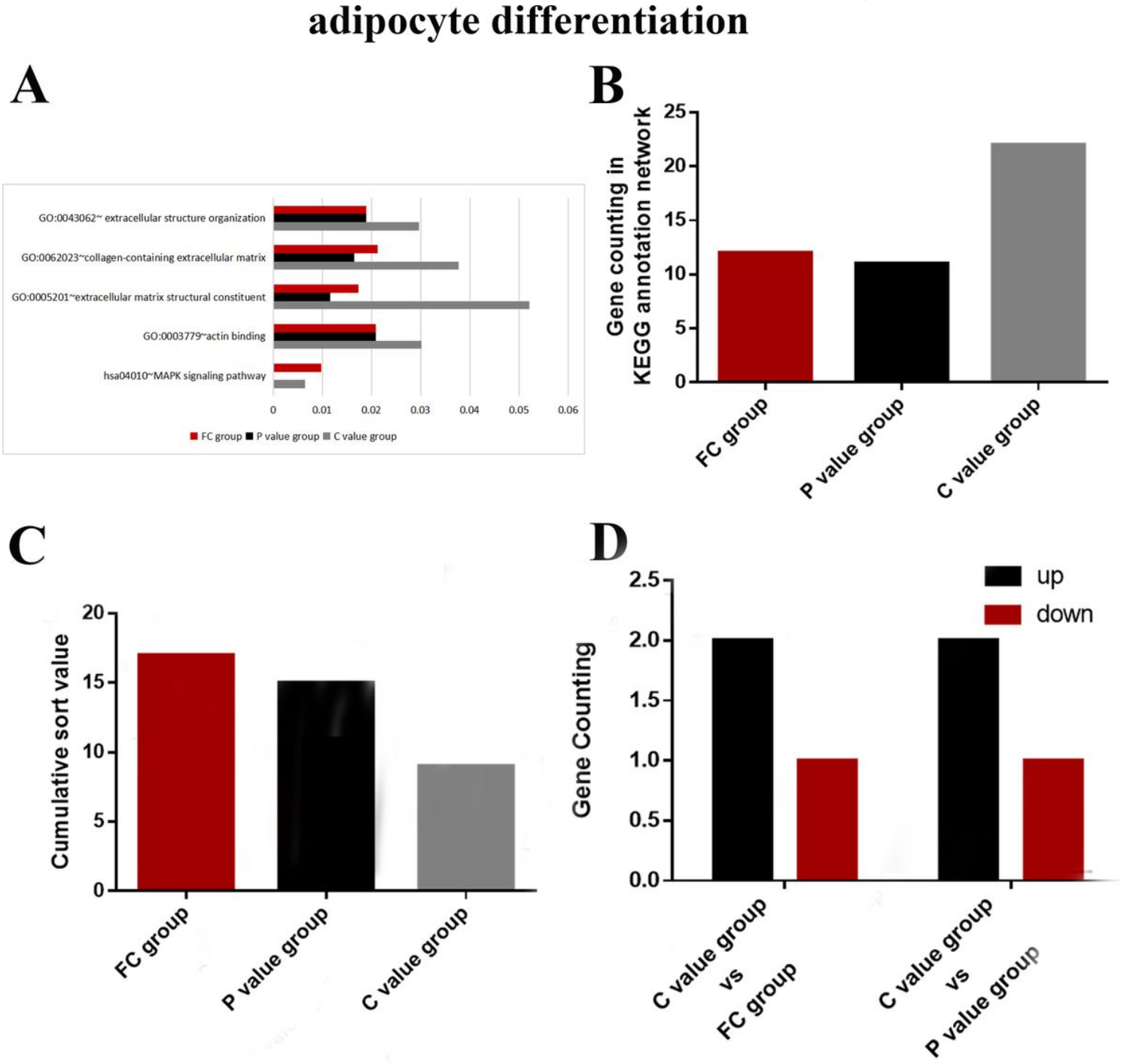

Figure 6

LncRNAs operation results for adipocyte differentiation data set (A) The ratio of predicted target mRNAs to the total genes in each term/pathway. (B) The total numbers of mRNAs in the three groups in annotation network. (C) Sort number accumulation value of adipocyte differentiation associatedlncRNAs by the three indexes. (D) The number of adipocyte differentiation associatedlncRNAs that rank up or down. (FC group: the collection of the top5 FC lncRNAs' predictive target mRNAs; P value group: the collection of the top5 P value lncRNAs' predictive target mRNAs; C value group: the collection of the top5 C value lncRNAs' predictive target mRNAs)

Supplementary Files

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

- Additionalfire1.docx
- SupplementaryFigureS2.tif
- SupplementaryFigureS1.tif
- Additionalfire2.docx