

Association Between Obesity and Incident Colorectal Cancer: An Analysis Based on Colorectal Cancer Database in the Cancer Genome Atlas

Su Yongxian (✉ 15818121868@163.com)

Peking University First Hospital

Chen Tonghua

Peking University First Hospital

Research Article

Keywords: Colorectal cancer, Obesity, mRNA, TCGA

Posted Date: February 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-200275/v1>

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

[Read Full License](#)

Abstract

Background

To investigate gene factors of colorectal cancer (CRC) in obesity and potential molecular markers.

Methods

Clinical data and mRNA expression data from The Cancer Genome Atlas (TCGA) was collected and divided into obese group and non-obese group according to BMI. The differential expressed genes (DEGs) were screened out by “Limma” package of R software based on ($|log_2(\text{fold change})| > 2$ and $p < 0.05$). The functions of DEGs were revealed with Gene Ontology and Kyoto Encyclopedia Genes and Genomes pathway enrichment analysis using the DAVID database. Then STRING database and Cytoscape were used to construct a protein-protein interaction (PPI) network and identify hub genes. Kaplan-Meier analysis was used to assess the potential prognostic genes for CRC patients.

Results

It has revealed 2055 DEGs in obese group with CRC, 7615 DEGs in non-obese group and 9046 DEGs in total group. MS4A12, TMIGD1, CA2, GBA3 and SLC51B were the top five downregulated genes in obese group. A PPI network consisted of 1042 nodes and 4073 edges, and top ten hub genes SST, PYY, GNG12, CCL13, MCHR2, CCL28, ADCY9, SSTR1, CXCL12 and ADRA2A were identified in obese group. PDCD11 may well predict overall survivals of CRC patients in non-obese group. The survival time of obese group was shorter than that of non-obese group, but there was no significant difference.

Conclusions

PDCD11 may be a potential molecular marker for non-obese patients with CRC.

1. Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related deaths worldwide based on the newly released global cancer statistics in 2018[1]. There will be an increased by 60% under the global burden of colorectal cancer by 2030, with > 2.2 million new cases and 1.1 million deaths [2]. It is widely believed that obesity is involved in the onset of cancer. Obesity-induced abnormal lipid metabolism, adipokines and hormones, chronic inflammation, gut microbiota dysbiosis, and disrupted bile acid homeostasis play significant roles in the complex metabolic regulation of CRC tumorigenesis [3]. Adipocytes in tumor micro-environment are an energy source for CRC growth. In adipose tissue, secreted adipokines and inflammation can elevate levels of insulin, IGFs, leptin, and inflammatory cytokines (e.g., IL-6, TNF- α , C-C motif ligand 2, and Plasminogen Activator Inhibitor 1), and

decrease levels of adiponectin in obese, which alone or together contribute to the formation and development of CRC. Moreover, obesity-induced gut microbiota dysbiosis increases harmful microbiota and metabolites and decreases beneficial ones. Additionally, CRC is promoted by bile acids, especially Deoxycholic acid and tauro- β -muricholic acid which are raised in obesity. The carcinogenesis of CRC is promoted by the bile acid-dependent inhibition of arylsulfatase A receptor, which is a target for anti-CRC [3]. However, there are few studies in gene terms concerning obesity with colorectal cancer, and the characteristic genes of colon cancer in obese people are not clear. Therefore, the main purpose of this paper is to identify the gene characteristics between obesity and colorectal cancer by bioinformatics based on TCGA database and to screen out potential molecular markers.

2. Materials And Methods

2.1. Data collection and grouping

The mRNA expression profiles and clinical data of colorectal patients were obtained from TCGA database (<https://portal.gdc.cancer.gov>). Approval by the Ethics Committee was not necessary because all data were collected from public available database (TCGA). Patients are divided into obese group and non-obese group according to BMI.

2.2. DEGs screening

Datasets were screened according to the following criteria: (1) samples with both mRNA sequencing data and clinical information about height and weight; (2) samples with prognosis information. Obesity is defined as $BMI \geq 30\text{kg}/\text{m}^2$ according to WHO. Finally, a total of 260 samples were enrolled in this study, including 244 with CRC and 16 without CRC. There were 78 with CRC and 3 without CRC in obese group, and 166 with CRC and 13 without CRC in non-obese group. The detailed clinical characteristics and differentially expressed mRNAs were listed in Table 1. The Limma and edgeR packages in R software 3.5 were used to identify DEGs between CRC and non-CRC samples in each group. The threshold was set at $(|\log_2(\text{fold change})| > 2 \text{ and } p < 0.05)$ and adjusted P-value < 0.05 . Volcano map was made later.

Table 1
Clinicopathological features of total group, obese group and non-obese group

Clinical index	Total(n = 228)	obese(n = 69)	non-obese(n = 159)	T/X ² value	P value
age(year, x ± s)	64.34 ± 12.55	63.74 ± 11.10	64.60 ± 13.16	0.510	0.611
sex(male/female, n)	120/108	35/34	85/74	0.144	0.704
Survival status (survive/death, n)	194/34	63/6	131/28	3.014	0.083
Survival time (month, x ± s)	22.85 ± 25.40	22.03 ± 26.24	23.22 ± 25.10	0.320	0.750
BMI(kg/m ²)	28.03 ± 5.90	35.30 ± 4.23	24.87 ± 3.03	0.000	0.000
Tumor Stage (unknown/I-II/III-IV, n)	10/116/102	2/35/32	8/81/70	0.559	0.756
BMI, body mass index.					

2.3. Functional enrichment analysis

To further analyze the potential biological processes (BP), cellular components (CC), molecular functions (MF), and pathways of the overlapping DEGs, the online software the Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>) was used to perform Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. P < 0.05 and counts > 2 were set as the threshold values.

2.4. PPI networks and the top 10 hub genes

A protein-protein interaction (PPI) network of common DEGs was constructed using the online software Search Tool for the Retrieval of Interacting Gene (STRING, <https://string-db.org/>). Confidence score ≥ 0.7 was considered as significant. Cytoscape software 3.7.2 was utilized to construct a protein interaction relationship network. Then, we select the top 10 hub genes from the PPI network through cytoHubba.

2.5. Statistic analysis

SPSS 25 was used to collate the data and compare clinical informations in each group by unpaired t test. Differences between groups in continuous variables were assessed with analysis of variance (ANOVA) and t test. Categorical variables were compared using X² test. The data were presented as mean ± standard deviation (SD) for continuous variable. The patients with CRC were divided into low-expression and high-expression groups, based on the mRNA expression of the ten hub genes in different groups using the X-tile software [4]. Kaplan-Meier (KM) survival curves were served to examine the prognostic

performance of the CRC-related hub genes using SPSS 25. P-values less than 0.05 were considered statistically significant.

3. Results

3.1. Clinicopathological features of total group, obese group and non-obese group

We divided total group into obese group($n = 69$) whose $BMI \geq 30 \text{ kg/m}^2$ and non-obese group ($n = 159$) whose $BMI < 30 \text{ kg/m}^2$. As shown in Table 1, the average age of obese group was smaller than that of non-obese group, and more importantly, the survival time of obese group was shorter, but there was no significant difference. Though survival rate of obese group was higher than that of non-obese group, no significant difference was found between groups. The proportion of intermediate- and advanced-stage tumors in the two groups was comparable and the ratio of men and women was close to equal.

3.2. Identification of DEGs

In total group, there were 9046 DEmRNAs with 3124 upregulated and 5922 downregulated. There were 7615 DEmRNAs in non-obese group with 1747 upregulated and 5868 downregulated, and 2055 DEmRNAs in obese group with 6 upregulated and 2049 downregulated (Fig. 1a, b, c). V-set and transmembrane domain containing 2A (VSTM2A), bone morphogenetic protein 3 (BMP3), scavenger receptor class A member 5 (SCARA5), carbonic anhydrases 1 (CA1), protein kinase inhibitor-B (PKIB) and solute carrier family 4 (SLC4A4) were significantly downregulated in non-obese group as well as in total group. Membrane spanning 4-domains A12 (MS4A12), transmembrane and immunoglobulin domain-containing 1 (TMIGD1), carbonic anhydrases 2 (CA2), glucocerebrosidase 3(GBA3), solute transporter-beta (SLC51B) are highly downregulated in obese group.

3.3. Functional enrichment analysis of mRNAs

In non-obese group, the DEGs were mainly enriched in leukocyte migration, cellular metal ion homeostasis, divalent inorganic cation homeostasis and calcium ion homeostasis in BP term (Fig. 2b) and involved in adherence junction, plasma membrane protein complex, neuronal cell body and actin cytoskeleton in CC term (Fig. 2c). In MF term, the DEGs were associated with cation transmembrane transporter activity, inorganic cation transmembrane transporter activity, antigen binding and metal ion cation transmembrane transporter activity (Fig. 2d). KEGG pathway analysis indicated that the DEGs were mainly related to cAMP signaling pathway, calcium signaling pathway, cGMP-PKG signaling pathway and adrenergic signaling in cardiomyocytes (Fig. 2a).

In obese group, the DEGs were markedly involved in BP term, which included leukocyte migration, lipid catabolic process, cellular lipid catabolic process and sodium ion transport (Fig. 3b). In addition, genes in CC term were enriched in apical part of cell, apical plasma membrane, cell projection membrane and basolateral plasma membrane (Fig. 3c). In MF term, genes were connective with cation transmembrane

transporter activity, inorganic cation transmembrane transporter activity, antigen binding and metal anion cation transmembrane transporter activity (Fig. 3d). In the KEGG pathway, genes mainly included neuroactive ligand-receptor interaction, cAMP signaling pathway, chemical carcinogenesis, drug metabolism-cytochrome P450 and mineral absorption (Fig. 3a). GO and KEGG pathway analysis of the DEGs in total group can be seen in Fig. 4 in detailed.

3.4. Construction of the PPI network and hub genes identification

We constructed a PPI network to further explore the interaction between the common DEGs by using STRING database and Cytoscape. As displayed in Supplementary Material 1, there were 1042 nodes and 4073 edges in the network in obese group, 1936 nodes and 9328 edges in non-obese group, 1940 nodes and 11029 edges in total group. Top 10 hub genes in each group were identified by cytoHubba (Fig. 5). All of top 10 hub genes in obese group were downregulated, while top 10 hub genes except SRSF5 in total group and top 10 hub genes in non-obese group were upregulated.

3.5. Identified prognostic mRNAs in patients

244 samples in total group from TCGA database, containing 167 in non-obese group and 77 in obese group, were used for KM survival analysis to screen hub genes related to prognosis of CRC patients. The results showed that in non-obese group, the survival rate with low expression of PDCD11 ($p < 0.05$) was higher than that of high expression (Fig. 6a). High expression of SRSF5, SRSF11 and HNRNPA1 had a higher survival rate than low expression of those in total group (Fig. 6b, c, d).

4. Discussion

The incidence and mortality of CRC still go up in many countries[5]. Identification of genes and pathways related to CRC would be helpful for the diagnosis and treatment of the disease [1]. Ga Eun Nam, et al.'s study [6] suggested that abdominal obesity may be a risk factor for CRC in this East Asian population. Also, a longitudinal analysis study indicated non-alcoholic fatty liver disease (NAFLD) with obesity posed a risk factor for colorectal cancer in apparently healthy Japanese individuals [7]. In addition, Abbinaya Elangovan, et al. conducted a large retrospective, cross-sectional database study which showed the association between obesity and colorectal cancer [8]. Obesity is regarded as one of the key environmental risk factors for the pathogenesis of CRC.

This is the first article to explore the relationship between obesity and colorectal cancer at the gene level of a large sample. We speculated that obesity may be an independent risk factor of CRC. In order to explore the characteristic genes of CRC in obese group and to identified potential molecular markers for early diagnosis, targeted therapy and recognition of prognosis of special CRC population, these characteristic genes were screened from TCGA database. We found that the most significant downregulated genes of non-obese group were mostly similar with those of total group, such as VSTM2A, BMP3, SCARA5, CA1, PKIB and SLC4A4. As for VSTM2A, consistent with our findings, Dong, Y., et al.

identified that it was one of the top downregulated secreted protein in CRC and maybe a novel prognostic biomarker for CRC patients, for VSTM2A might suppress colorectal tumorigenesis by directly binding to Wnt signaling co-receptor LDL receptor related protein 6 [9]. In addition, Secco, B., et al. [10] reported VSTM2A as a factor modulating adipogenic commitment, which can maintain and amplify the adipogenic capacity of adipose precursor cells. Cancer tissue in non-obese people has lower VSTM2A expression than normal tissue beside cancer, which seems to contradict energy theory.

In our study, MS4A12, TMIGD1, CA2, GBA3, SLC51B are only sharply downregulated in obese group and SST, PYY, GNG12, CCL13, MCHR2, CCL28, ADCY9, SSTR1, CXCL12, ADRA2A were identified as top ten hub genes in the network in obese group. Among above genes, MS4A12 is related to proliferation and motility of colon cancer cells, participating in cell membrane composition, cell differentiation, proliferation, and cell cycle regulation [11]. Current studies supported that MS4A12 is a colon-specific gene regulated by intestinal differentiation transcription factor CDX2 [12, 13] and in adenomatous polyp, MS4A12 transcription is downregulated while CDX2 gene and protein expression upregulated [13, 14]. Survival analysis of primary colorectal cancer patients showed that patients with low expression of MS4A12 had a worse survival rate than those with high expression of MS4A12, suggesting that MS4A12 was involved in the occurrence and development of primary colorectal cancer and could inhibit the progression of cancer, which may be a potential target for diagnosis and treatment [15]. What's more, MS4A12 can predict prognosis in early stage colon cancer patient and mainly relate to malignancy of non-metastasis tumor while the prognosis value of metastasis colon tumor is low [13]. However, few evidences confirm the relationship between this gene and obesity.

TMIGD1 is also a downregulated gene in obese group. It not only is a novel adhesion molecule, which is highly conserved in humans and other species, but also a novel tumor suppressor highly expressed in normal colorectal epithelial cells and downregulated in CRC [16]. Kyle Oliver, et al.'s study showed that TMIGD1 inhibits cell migration and metastasis in colon cancer as a tumor suppressor by arresting cell cycle at G2/M, and down-regulation of TMIGD1 correlates with poor survival [17]. However, few studies have been able to draw on any systematic research into the association between this gene and CRC with obesity.

SLC51B, combined with solute transporter-alpha (SLC51A) is the encode gene of the organic solute transporters alpha and beta (OST α -OST β), which is located at the basolateral membrane, involving in the metabolism of bile acids [18]. Some studies indicated a defect about intestinal bile acid and conjugated steroid absorption in SLC51A-deficient mice, causing a decrease in the levels of bile acid, serum triglyceride, cholesterol, and glucose [19, 20]. Wang DF, et al. confirmed that expression of SLC51B which plays a role in bile acid transport, was decreased in the ulcerative colitis. Diminished uptake of bile acids into ileocytes is most likely responsible for the decreased expression of SLC51B in the ileum. In our study, the down-regulation of this gene in the obese group may be associated with metabolic abnormalities in obesity. Notably, many studies have suggested that ulcerative colitis and CRC have the same metabolic pathway, which seems to support abnormal expression of this gene in CRC.

Carbonic anhydrases, a kind of critical enzymes in regulating pH homeostasis in cells, catalyze the interconversion between carbon dioxide and carbonic acid [21, 22]. Norihiro Nakada, et al.'s study showed that CA2, an acytoplasmic enzyme widely expressed in normal organs including colonic mucosa, was downregulated in ulcerative colitis-associated colorectal cancer. However, the mechanism of CA2 downregulation in CRC is not clear [23].

GBA3 is an enzyme belonging to the glycosid e hydrolase family 1 that is able to hydrolyse a wide variety of substrates with a b-D-glucose moiety linked to a hydrophobic group, mostly present in the kidney, liver, spleen, intestine and lymphocytes of mammals [24]. Very little is known about the role of this enzyme in cancer. In our study, GBA3 was downregulated in CRC tissue in obese group. Gene expression profiling of intestinal metaplastic lesions from patients, which are precursors of gastric cancers, identified GBA3 among the top 25 significantly upregulated genes [25]. GBA3 might play an important role in the detoxification and/or biotransformation of dietary xenobiotic plant β -glycosides in roots and tubers, which are the main food in African populations but not in Asian and European populations. And it can be observed an overlap in Eurasia between the highest frequency of GBA3 loss and a main dietary source of meat or fish, while in Africa the lowest frequency of GBA3 inactive alleles meets the general trend in diets higher in glycoside-rich foods and poorer in meat or fish [26].

In the enrichment analysis of the three groups, DEGs in MF term are mainly focused on ion channel activity, and on inflammation in BP term, but in the obesity group, DEGs also focused on metabolism. In terms of CC, cell membrane covered the most in the obese group, while the non-obese group and total group preferred cell connection structure. The KGGG enrichment analysis of the three groups mainly focused on the common energy metabolism signal pathway.

Though KM survival analysis, PDCD11 may be potential molecular markers for non-obese patients with CRC. PDCD11 is a NF- κ B-binding protein that colocalizes with U3 RNA in the nucleolus and is required for rRNA maturation and generation of 18S rRNA. PDCD11 is necessary for Fas ligand (FasL) expression, and PDCD11 overexpression is known to induce transcription of FasL, leading to the induction of apoptosis through Fas/FasL/caspase death pathway [27–30]. Current studies on PDCD11 mainly focus on the nervous system, such as transient ischemic attack and schizophrenia [31, 32]. Importantly, our study argued that PDCD11 upregulated in CRC tissue, belonged to low expression category in non-obese group with longer survival time (Supplement txt. 2; Fig. 6a). There might be other pathways affecting CRC tumorigenesis, which is not clear now.

It is found that SRSF5, SRSF11 and HNRNPA1 had proven to be associated with the prognosis of CRC in total group of our study. SRSF5, previously called SRp40, is a member of the SR protein family that has early been identified as a splicing regulator [33, 34]. Currently, research about SRSF5 is mainly focused on lung cancer, and it is hyperacetylated and upregulated in human lung cancer. Little is known about the relationship between SRSF5 and colorectal cancer. In our analysis, SRSF5 was downregulated in colorectal tissue and belonged to low expression category in total group (Supplement txt. 2; Fig. 6c). The survival time of low expression SRSF5 group was less than that of the high expression group in total

group ($P = 0.025$). Yuhan Chen, et al.'s study [35] indicated that upon glucose intake, the splicing factor SRSF5 is specifically induced and promotes the alternative splicing of cell cycle and apoptosis regulator 1 to produce special proteins, which promote tumor growth by enhancing glucose consumption and acetyl-CoA production. SRSF5 seems to be a tumor suppressor gene in colorectal cancer and responds to high glucose to promote cancer development.

HNRNPA1 is also a well-known splicing regulator with effects antagonistic to SR proteins [36]. Upregulated expression and aberrant cytoplasmic localization of HNRNPA1, as determined by immunohistochemical staining, were noted in CRC [37]. HNRNPA1 not only contributed to the promotion of cell growth through the regulation of energy metabolism to use glycolysis efficiently, but upregulated the promoter activity of TRA2B which was considered as an oncogene [38, 39]. Sun Y, et.al argued that the phosphorylation of Ser6 of hnRNPA1 was a predictor of poor prognosis for patients with CRC [40]. HNRNPA1 was upregulated in CRC tissue in total group of our study, too. However, it belonged to low expression category with shorter survival time (Supplement txt. 2; Fig. 6d). Interestingly, an animal study showed that Hepatic HNRNPA1 overexpression activated the CaM/Akt pathway and repressed the mTOR/SREBP-1C pathway to ameliorate hyperglycemia and steatosis in obese mice [41]. Therefore, we surmise that this gene might upregulated in obese people with CRC and lead to a poor prognosis of CRC.

At present, there are few researches about SRSF11 on CRC. Ji Hoon Lee, et.al suggested that SRSF11 is a novel the telomerase RNA component binding protein that localizes to nuclear speckles with subnuclear structures that are enriched in pre-messenger RNA splicing factors. It acts as a nuclear speckle-targeting factor that is essential for telomerase association with telomeres and provides a potential target for modulating telomerase activity in cancer [42]. We found that SRSF11 still belonged to low expression category in total group with shorter survival time (Supplement txt. 2; Fig. 6b). It seemed to confirm that SRSF11 is a tumor suppressor gene in CRC.

There are some limitations in this study. Firstly, patients were categorized by BMI. However, some studies suggested that measures of abdominal adiposity, such as waist circumference (WC) and waist-to-hip ratio (WHR), might be more strongly associated with colorectal neoplasia risk than BMI [6, 43]. Secondly, our study was performed only by pure bioinformatics analysis. Finally, our article only found the relevant colorectal cancer genes in the non-obese population, but not in obese population. Therefore, further experiments are needed to validate the results based on tumor samples and clinical data.

Conclusion

Briefly, our study reveals obesity plays an important role in pathogenesis of colorectal cancer in molecular term. Hub gene PDCD11 may serve as novel independent prognostic biomarkers that could be used to predict the clinical outcomes of non-obese patients with CRC. This signature may serve as a possible candidate biomarker and therapeutic target for these special colorectal cancer patients. It is necessary to further pre-clinical studies followed by clinical trials to validate our findings in the future.

Abbreviations

CRC = Colorectal cancer; CI = confidence intervals, DEmRNAs = the differentially expressed mRNAs, FC = fold change, GO = Gene Ontology, HR = hazard ratio, KEGG = Kyoto Encyclopedia of Gene and Genomes, PPI = protein-protein interaction, TCGA = The Cancer Genome Atlas, BMI = body weight index.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in the The Cancer Genome Atlas repository.

Competing interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contributions

SYX completed the experimental assumption and determined the method, participated in the collection and analysis of the data, and was a major contributor in writing the manuscript. SYX and CTH prepared figure 1-6 and Table 1. CTH participated in the collection and analysis of the data, and manuscript writing. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394-424.

2. Arnold, M., et al., *Global patterns and trends in colorectal cancer incidence and mortality*. Gut, 2017. **66**(4): p. 683-691.
3. Ye, P., et al., *Linking Obesity with Colorectal Cancer: Epidemiology and Mechanistic Insights*. Cancers (Basel), 2020. **12**(6).
4. Camp, R.L., M. Dolled-Filhart, and D.L. Rimm, *X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization*. Clin Cancer Res, 2004. **10**(21): p. 7252-9.
5. Siriwardena, A.K., et al., *Management of colorectal cancer presenting with synchronous liver metastases*. Nat Rev Clin Oncol, 2014. **11**(8): p. 446-59.
6. Nam, G.E., et al., *Association between Abdominal Obesity and Incident Colorectal Cancer: A Nationwide Cohort Study in Korea*. Cancers (Basel), 2020. **12**(6).
7. Hamaguchi, M., et al., *Non-alcoholic fatty liver disease with obesity as an independent predictor for incident gastric and colorectal cancer: a population-based longitudinal study*. BMJ Open Gastroenterol, 2019. **6**(1): p. e000295.
8. Elangovan, A., et al., *Colorectal Cancer, Age, and Obesity-Related Comorbidities: A Large Database Study*. Dig Dis Sci, 2020.
9. Dong, Y., et al., *VSTM2A suppresses colorectal cancer and antagonizes Wnt signaling receptor LRP6*. Theranostics, 2019. **9**(22): p. 6517-6531.
10. Secco, B., et al., *Amplification of Adipogenic Commitment by VSTM2A*. Cell Rep, 2017. **18**(1): p. 93-106.
11. Koslowski, M., et al., *MS4A12 is a colon-selective store-operated calcium channel promoting malignant cell processes*. Cancer Res, 2008. **68**(9): p. 3458-66.
12. Xu, H., et al., *Identification and Verification of Core Genes in Colorectal Cancer*. Biomed Res Int, 2020. **2020**: p. 8082697.
13. He, L., H.Y. Deng, and C.W. X, *Decreased expression of MS4A12 inhibits differentiation and predicts early stage survival in colon cancer*. Neoplasia, 2017. **64**(1): p. 65-73.
14. Drew, J.E., et al., *Predictive gene signatures: molecular markers distinguishing colon adenomatous polyp and carcinoma*. PLoS One, 2014. **9**(11): p. e113071.
15. Han, J., et al., *CLCA4 and MS4A12 as the significant gene biomarkers of primary colorectal cancer*. Biosci Rep, 2020. **40**(8).
16. Cattaneo, E., et al., *Preinvasive colorectal lesion transcriptomes correlate with endoscopic morphology (polypoid vs. nonpolypoid)*. EMBO Mol Med, 2011. **3**(6): p. 334-47.
17. De La Cena, K.O.C., et al., *Transmembrane and Immunoglobulin Domain Containing 1, a Putative Tumor Suppressor, Induces G2/M Cell Cycle Checkpoint Arrest in Colon Cancer Cells*. Am J Pathol, 2021. **191**(1): p. 157-167.
18. Schaffner, C.A., et al., *The organic solute transporters alpha and beta are induced by hypoxia in human hepatocytes*. Liver Int, 2015. **35**(4): p. 1152-61.

19. Ballatori, N., et al., *Ostalpha-Ostbeta is required for bile acid and conjugated steroid disposition in the intestine, kidney, and liver*. Am J Physiol Gastrointest Liver Physiol, 2008. **295**(1): p. G179-g186.
20. Rao, A., et al., *The organic solute transporter alpha-beta, Ostalpha-Ostbeta, is essential for intestinal bile acid transport and homeostasis*. Proc Natl Acad Sci U S A, 2008. **105**(10): p. 3891-6.
21. Supuran, C.T., *Structure and function of carbonic anhydrases*. Biochem J, 2016. **473**(14): p. 2023-32.
22. Mboge, M.Y., et al., *Carbonic Anhydrases: Role in pH Control and Cancer*. Metabolites, 2018. **8**(1).
23. Nakada, N., et al., *Expression of CA2 and CA9 carbonic anhydrases in ulcerative colitis and ulcerative colitis-associated colorectal cancer*. Pathol Int, 2020. **70**(8): p. 523-532.
24. Astudillo, L., et al., *Glucosylceramidases and malignancies in mammals*. Biochimie, 2016. **125**: p. 267-80.
25. Lee, H.J., et al., *Gene expression profiling of metaplastic lineages identifies CDH17 as a prognostic marker in early stage gastric cancer*. Gastroenterology, 2010. **139**(1): p. 213-25.e3.
26. Lopes-Marques, M., et al., *GBA3: a polymorphic pseudogene in humans that experienced repeated gene loss during mammalian evolution*. Sci Rep, 2020. **10**(1): p. 11565.
27. Sweet, T., et al., *Identification of a novel protein from glial cells based on its ability to interact with NF-kappaB subunits*. J Cell Biochem, 2003. **90**(5): p. 884-91.
28. Sweet, T., et al., *Evidence for involvement of NFBP in processing of ribosomal RNA*. J Cell Physiol, 2008. **214**(2): p. 381-8.
29. Lacana, E. and L. D'Adamio, *Regulation of Fas ligand expression and cell death by apoptosis-linked gene 4*. Nat Med, 1999. **5**(5): p. 542-7.
30. Ristić, T., et al., *Serum Fas/FasL levels in dependence on clinical presentations of coronary disease and their relationship with risk factors*. Vojnosanit Pregl, 2010. **67**(7): p. 537-42.
31. Yoshida, Y., et al., *Elevation of autoantibody level against PDCD11 in patients with transient ischemic attack*. Oncotarget, 2018. **9**(10): p. 8836-8848.
32. Hoya, S., et al., *Rare PDCD11 variations are not associated with risk of schizophrenia in Japan*. Psychiatry Clin Neurosci, 2017. **71**(11): p. 780-788.
33. Breig, O. and F. Baklouti, *Proteasome-mediated proteolysis of SRSF5 splicing factor intriguingly co-occurs with SRSF5 mRNA upregulation during late erythroid differentiation*. PLoS One, 2013. **8**(3): p. e59137.
34. Manley, J.L. and A.R. Krainer, *A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins)*. Genes Dev, 2010. **24**(11): p. 1073-4.
35. Chen, Y., et al., *Mutually exclusive acetylation and ubiquitylation of the splicing factor SRSF5 control tumor growth*. Nat Commun, 2018. **9**(1): p. 2464.
36. Zhu, J., A. Mayeda, and A.R. Krainer, *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins*. Mol Cell, 2001. **8**(6): p. 1351-61.

37. Hope, N.R. and G.I. Murray, *The expression profile of RNA-binding proteins in primary and metastatic colorectal cancer: relationship of heterogeneous nuclear ribonucleoproteins with prognosis*. Hum Pathol, 2011. **42**(3): p. 393-402.
38. Kuranaga, Y., et al., *SRSF3, a Splicer of the PKM Gene, Regulates Cell Growth and Maintenance of Cancer-Specific Energy Metabolism in Colon Cancer Cells*. Int J Mol Sci, 2018. **19**(10).
39. Nishikawa, T., et al., *HnRNPA1 interacts with G-quadruplex in the TRA2B promoter and stimulates its transcription in human colon cancer cells*. Sci Rep, 2019. **9**(1): p. 10276.
40. Sun, Y., et al., *Phosphorylation of Ser6 in hnRNPA1 by S6K2 regulates glucose metabolism and cell growth in colorectal cancer*. Oncol Lett, 2017. **14**(6): p. 7323-7331.
41. Wang, J., et al., *Long Noncoding RNA IncSHGL Recruits hnRNPA1 to Suppress Hepatic Gluconeogenesis and Lipogenesis*. Diabetes, 2018. **67**(4): p. 581-593.
42. Lee, J.H., et al., *Involvement of SRSF11 in cell cycle-specific recruitment of telomerase to telomeres at nuclear speckles*. Nucleic Acids Res, 2015. **43**(17): p. 8435-51.
43. Dash, C., et al., *Obesity is an initiator of colon adenomas but not a promoter of colorectal cancer in the Black Women's Health Study*. Cancer Causes Control, 2020. **31**(4): p. 291-302.

Figures

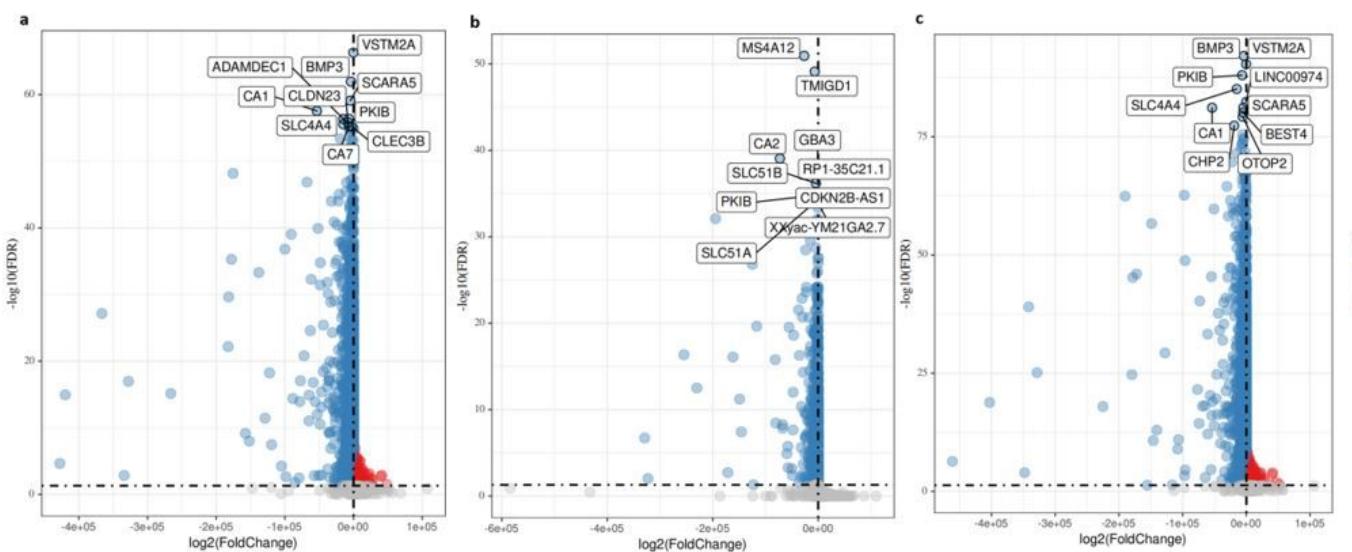


Figure 1

DEmRNAs in non-obese group, obese group and total group. (a) DEmRNAs in non-obese group. (b) DEmRNAs in obese group. (c) DEmRNAs in total group. DEmRNAs, the differentially expressed mRNAs.

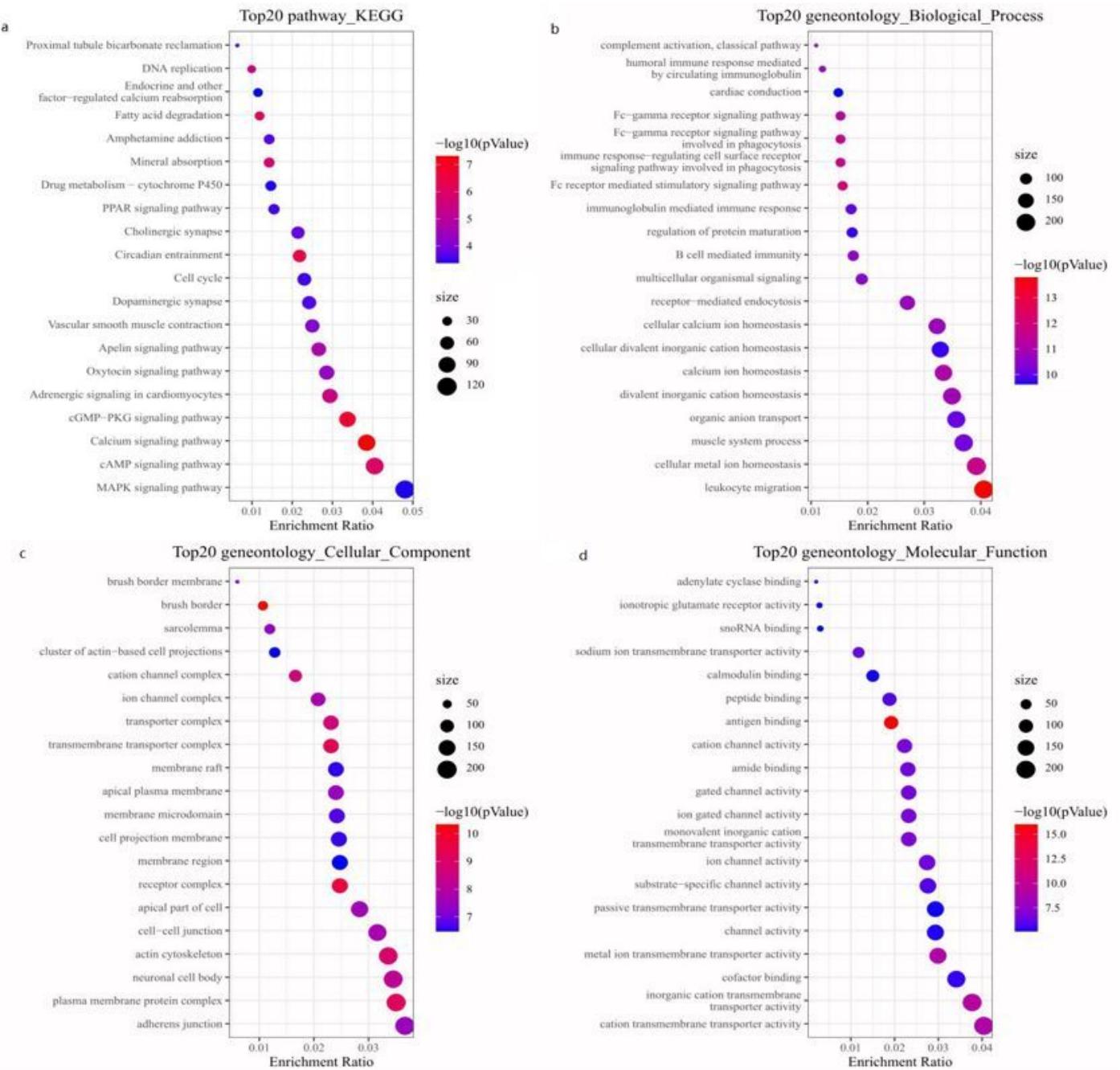


Figure 2

GO enrichment analysis (biological processes (BP, b), cellular component (CC, c) and molecular function (MF, d), and) and KEGG pathway analysis (a) in non-obese group.

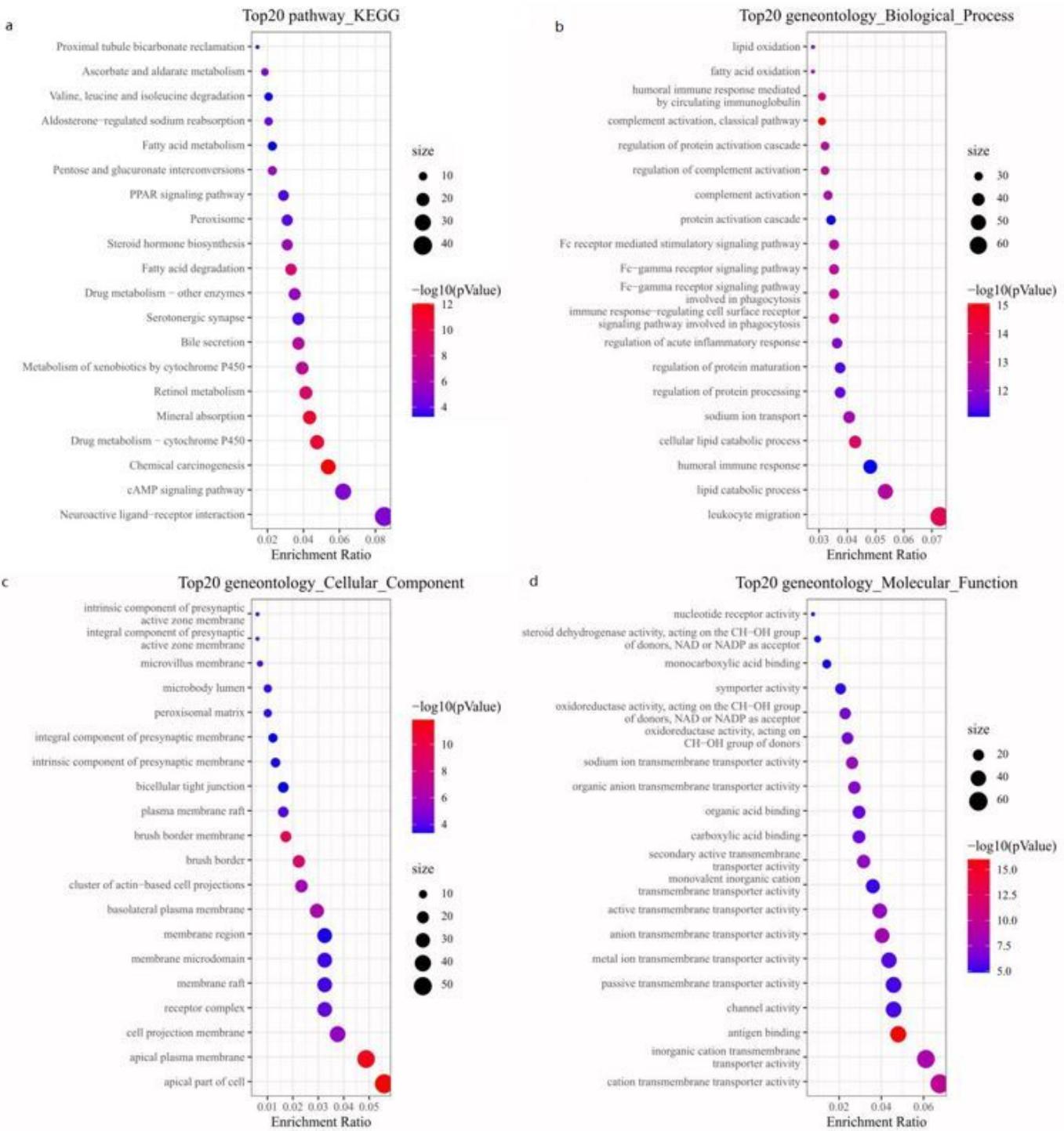


Figure 3

GO enrichment analysis (biological processes (BP, b), cellular component (CC, c) and molecular function (MF, d), and) and KEGG pathway analysis (a) in obese group.

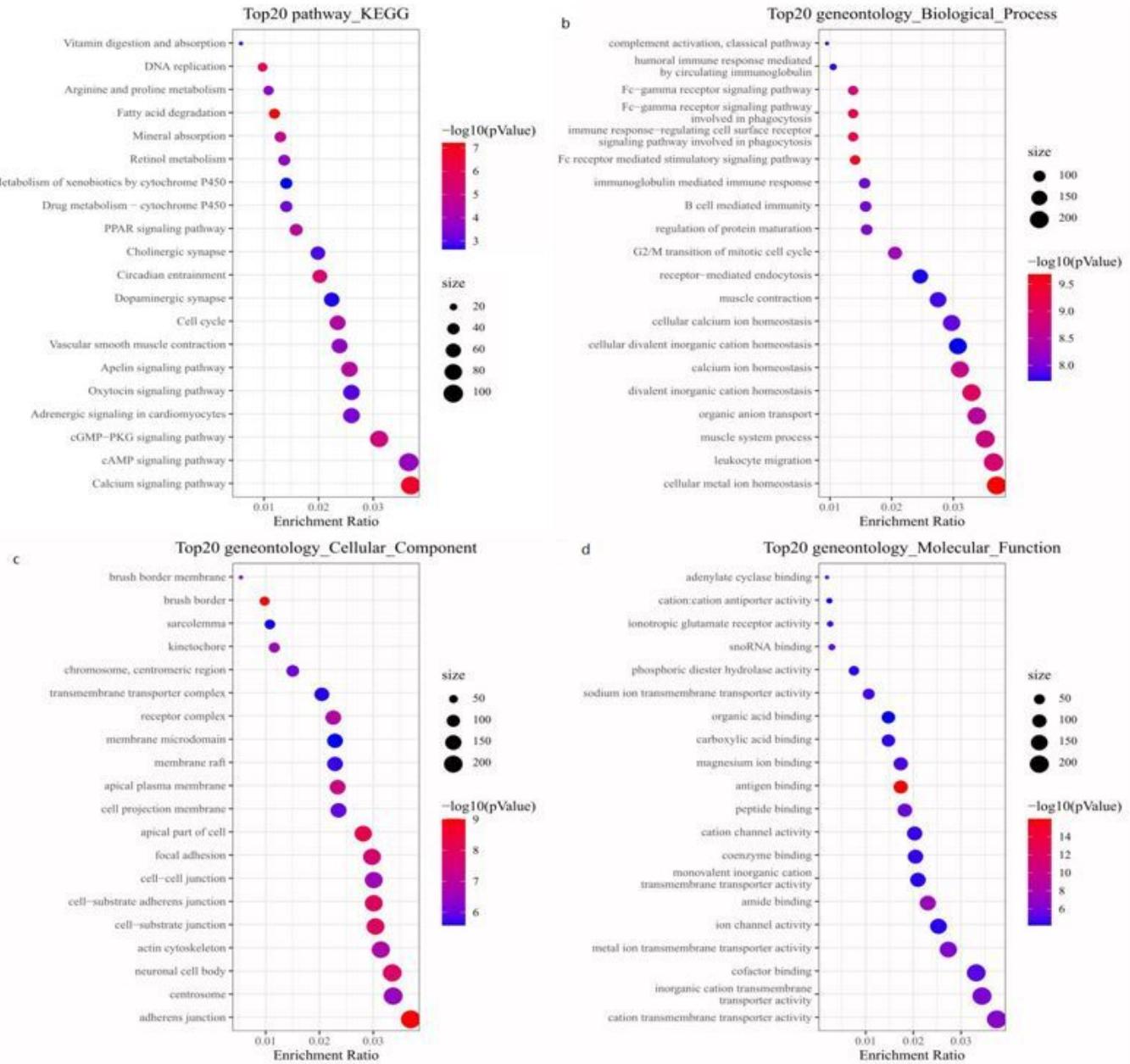


Figure 4

GO enrichment analysis (biological processes (BP, b), cellular component (CC, c) and molecular function (MF, d), and) and KEGG pathway analysis (a) in total group.

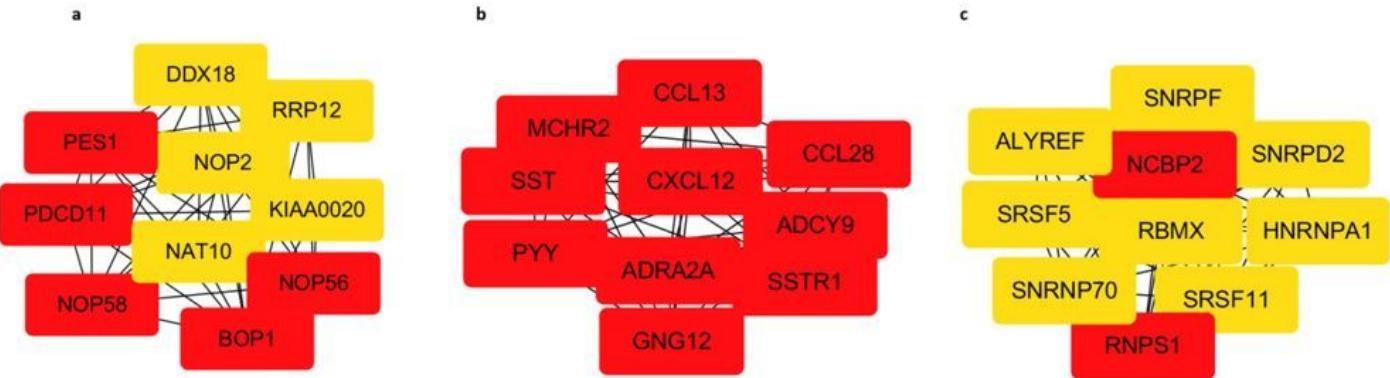


Figure 5

Top 10 hub genes were identified by cytoHubba. (a) PES1, PDCD11, NOP58, BOP1, NOP56, NAT10, NOP2, DDX18, RRP12, KIAA0020 were identified in the network in non-obese group. (b) SST, PYY, GNG12, CCL13, MCHR2, CCL28, ADCY9, SSTR1, CXCL12, ADRA2A were identified in the network in obese group. (c) NCBP2, RNPS1, SRSF11, SNRPF, ALYREF, RBMX, SNRNP70, SNRPD2, SRSF5, HNRNPA1 were identified in the network in total group. Hub genes of red blanks had a higher score than those of yellow blanks. PES1, pescadillo; PDCD11, programmed cell death 11; NOP58, nucleolar protein 58; BOP1, block of proliferation 1; NOP56, nucleolar protein 56 ; NAT10, n-acetyltransferase 10; NOP2, nucleolar protein 2; DDX18, DEAD-box helicase 18; RRP12, ribosomal RNA processing 12; KIAA0020, pumilio RNA Binding Family Member 3; SST, somatostatin; PYY, peptide YY; GNG12, guanine nucleotide-binding protein subunit gamma-12; CCL13/MCP-4, monocyte chemotactic protein-4; MCHR2, melanin-concentrating hormone receptor 2; CCL28, chemokine (C-C motif) ligand 28; ADCY9, adenylyl cyclase 9; SSTR1, somatostatin receptor type 1; SDF-1/CXCL12, stromal cell-derived factor 1; ADRA2A, adrenergic receptor alpha 2a. NCBP2, nuclear cap-binding protein 2; RNPS1, RNA-binding protein with serine-rich domain 1; SRSF11, serine/arginine-rich splicing factor 11; SNRPF, small nuclear ribonucleoprotein F; RBMX, X-linked RNA-binding motif protein; SNRNP70, small nuclear ribonucleoprotein 70; SNRPD2, small nuclear ribonucleoprotein Sm D2; SRSF5, serine/arginine-rich splicing factor 5; HNRNPA1, heterogeneous nuclear ribonucleoprotein A1.

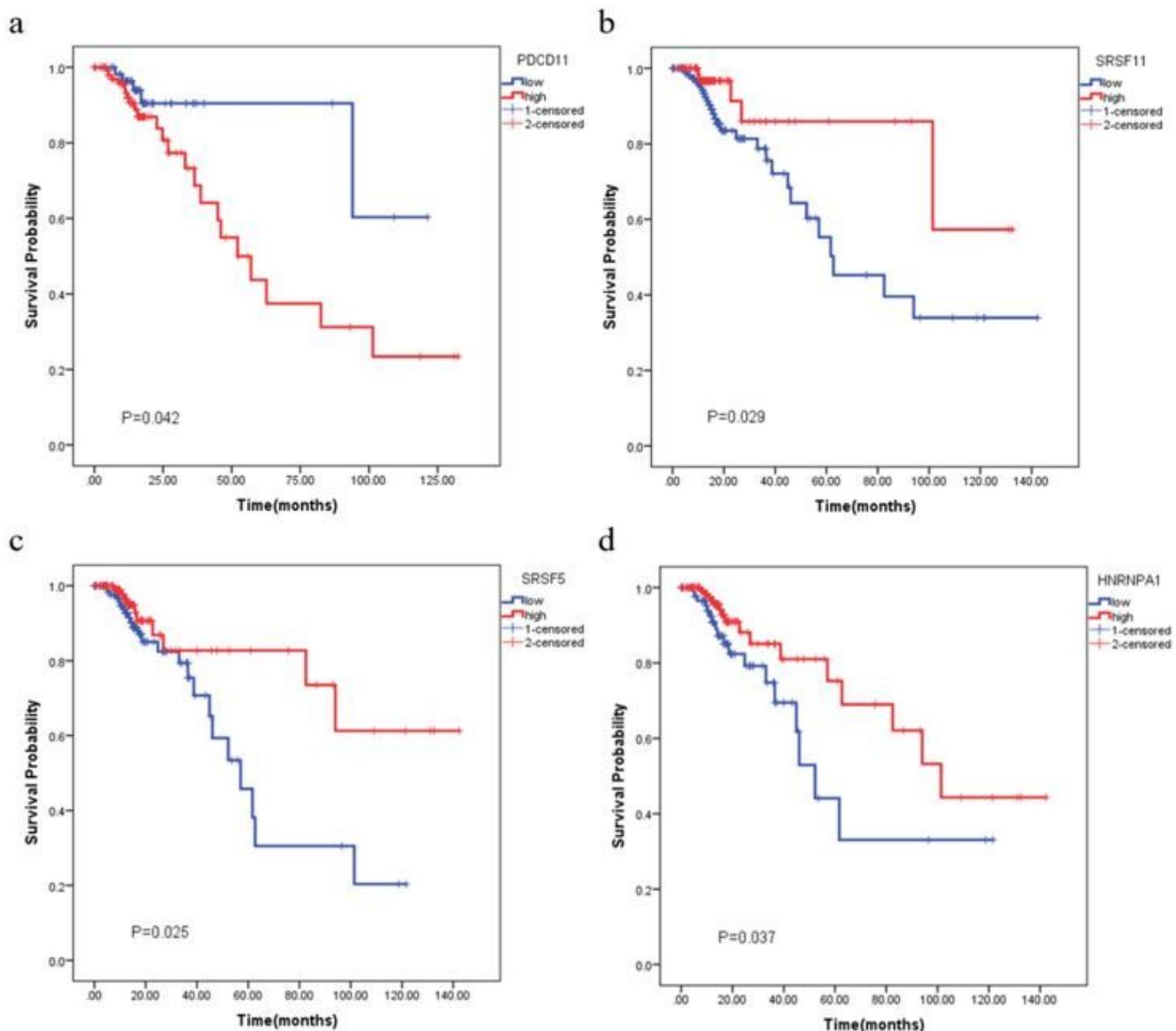


Figure 6

Kaplan-Meier survival analysis of PDCD11 (a) in non-obese group; SRSF11 (b), SRSF5 (c) and HNRNPA1(d) in total group. PDCD11, programmed cell death 11; HNRNPA1, heterogeneous nuclear ribonucleoprotein A1; SRSF5, serine-arginine-rich splicing factor 5; SRSF11, serine/arginine-rich splicing factor 11.

Supplementary Files

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

- [SupplementaryMaterial1.docx](#)
- [SupplementaryMaterial2.xlsx](#)