

Copy number variations primed lncRNAs deregulation contribute to poor prognosis in colorectal cancer

Huimin Liu^{1,2,*}, Xiaoyu Gu^{1,2,*}, Guihua Wang¹, Ying Huang¹, Shaoqing Ju¹, Jianfei Huang², Xudong Wang^{1,2}

¹Department of Laboratory Medicine, Affiliated Hospital of Nantong University, Nantong, Jiangsu, China

²Clinical Biobank, Affiliated Hospital of Nantong University, Nantong, Jiangsu, China

*Equal contribution

Correspondence to: Xudong Wang; email: wangxudong88@hotmail.com

Keywords: colorectal cancer, copy number variations, Long non-coding RNAs, prognosis

Received: May 5, 2019

Accepted: August 5, 2019

Published: August 22, 2019

Copyright: Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Copy number variations (CNVs) are crucial genetic change elements in malignancies, and lncRNAs deregulation induced by genomic and epigenomic aberrations plays key driving role in tumorigenesis, including colorectal cancer (CRC). However, effects of CNVs associated with lncRNAs in CRC is largely unknown. Here, we perform integrative analysis considering messenger RNA expression levels, DNA methylation and DNA copy numbers from 289 cases of CRC specimens. There are five prognostic subtypes of CRC determined by multi-omics integration, and differentially expressed lncRNAs (DE-lncRNAs) are acquired among five subtypes and normal cases. Finally, CNVs pattern matched with DE-lncRNAs reveals a signature including 10 lncRNAs (LOC101927604, LOC105377267, CASC15, LINC-PINT, CLDN10-AS1, C14orf132, LMF1, LINC00675, CCDC144NL-AS1, LOC284454), conspicuously contributing to poor prognosis in CRC, which can be validated in another independent dataset. Together, our research is interested in copy number changes relevant with lncRNAs, not only expanding the spectrum of CNVs, but also perfecting the regulation network of lncRNAs in CRC. The main purpose is to provide novel biomarkers for prognostic managements of CRC patients.

INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer-related death among all malignancies worldwide [1]. Despite surgical operation, radiotherapy, chemotherapy and personalized medicine have strongly prolonged survival of CRC patients, the 5-year relative survival remains less than 50% [2]. In view of genomic variations, epigenomic alterations and transcriptional deregulation, CRC is often signatored with heterogeneous behaviors and a couple of molecular subtypes have been proposed, which describe detailed pictures of tumor biology mechanisms.

Copy number variation (CNV) is defined as a pattern of genetic structural variation, and generally refers to the increase or decrease in copy number of genomic fragments between 1 kb and 3 Mb [3]. Copy number

amplifications or deletions in cancer genomes often induce oncogenes expression or deactivation of tumor suppressor genes and harbor significant influence on cellular functions, including adhesion, recognition, communication [4, 5]. Increase in copy number of IGHG3 located at 14 chromosome often contributes to over-expression of this gene and high prevalence and mortality in prostate cancer [6]. Not only referring to tumor biology, CNVs are also familiar with dosage effects in some intricately immunological diseases. Enhanced HIV/acquired immunodeficiency syndrome (AIDS) susceptibility is visible in individual who has CCL3L1 copy number aberrations [7]; CNV of human Fcgr3 gene is a determinant of glomerulonephritis [8]; Low copy number of component C4 is a danger factor and high copy number is a shielding factor in systemic lupus erythematosus (SLE) [9]. In the light of these expounded findings, bulks of investigations have shed light on

phenotypic changes and tumor progression caused by CNVs, concentrating on messenger RNA (mRNA), but few studies explicate the regulatory relationship between CNVs and non-coding RNAs, especially lncRNAs. With this concern in mind, here, we are particularly interested in taking a research on the relationship between structural variations in whole genomes and lncRNAs in CRC.

Long non-coding RNAs (lncRNAs) are known as size larger than 200nt without protein coding ability [10]. Since their discovery, increasing numbers of lncRNAs serve as significant modulators in tumorigenesis and progression in human CRC [11–14]. Furthermore, lncRNAs have also come forward as potential prognostic biomarkers in CRC, and survival analysis indicates that aberrant lncRNAs expression hold increasing risk of relapse [15–17]. However, there are few attentions calling for potential prognostic biomarkers by means of detecting DNA copy number amplifications or deletions of lncRNAs.

In this research, we could identify five molecular subtypes associated with prognostic outcomes of CRC patients based on messenger RNA expression levels, DNA methylation and DNA copy numbers. R package-DESeq2 was employed to distinguish differentially expressed mRNAs and lncRNAs across five molecular subtypes and normal tissues. Despite the well-known effect of CNVs on transcriptomic regulation, it is still unclear whether CNVs is systematically related to the expression levels of lncRNAs in CRC. By analyzing copy number profiles of the whole genome lncRNAs, we investigated these deregulated lncRNAs induced by copy number amplifications or deletions. Furthermore, prognostic-related lncRNAs were exploited for Kaplan-Meier (KM) survival analysis. Overall, we are aimed at identifying CNVs-related lncRNAs guiding prognosis in CRC.

RESULTS

Multi-omics integration analysis

The expression profiles of PCGs, CNVs and 450k methylation were combined with prognosis status based on univariate cox proportional hazards model. Eventually, 2118 genes, 5015 CNV regions and 7083 CpG sites were obtained with the significant threshold $p < 0.05$. Subsequently, 289 CRC patients sharing in three omics were classified into 5 molecular subtypes using iCluster. Five kinds of subtypes C1, C2, C3, C4, C5 consisted of 23, 44, 23, 60, 139 CRC patients respectively (Table 1). C3 group was harboring the worst survival probabilities (overall survival), and C2 subtype owned the optimal prognosis status ($p = 0.0056$) (Figure 1A). Progression free survival (PFS) among five subtypes was almost consistent

Table 1. Identification of five subtypes and the distribution of CRC samples across subtypes.

Cluster	SampleCount
C1	23
C2	44
C3	23
C4	60
C5	139

with overall survival, similarly, C3 group got hold of the worst survival probabilities (PFS) (Supplementary Figure 1). The distribution of TNM stage among five subtypes clearly showed that the proportion of advanced patients (stageIII+stageIV) in C3 group was much larger than other four subtypes accounting for about 60% (Supplementary Figure 2). Simultaneously, we described the spectrum of gene mutation status among five subtypes, and selected the top 10 genes with the highest mutation rate in each subtype. There are total 20 mutation related genes acquired and this phenomenon highlighted that the most common mutation related genes in five subtypes had higher coincidence. We made further observations revealing that these 20 genes were differently distributed in five subtypes and the mutation frequency of the same gene in different subtypes is also distinguished (Figure 1B). Obviously, the frequency of gene mutations in C5 subtype was significantly higher than other subtypes, and even in some C5 subtypes samples, these 20 gene mutations almost existed. In addition, TTN, APC, TP53 and KRAS mutations were especially more common than other genes, which was highly shining upon that these four genes mutations may imply pivotal role in carcinogenesis. These findings suggested that our strategy for subtypes classification based on the expression profiles of PCGs, 450k methylation and CNVs data could predict distinct prognostic situation of CRC and feature certain regulatory relationship among genomic, epigenomic and transcription level.

Differentially expressed lncRNAs and mRNAs across subtypes

DE-lncRNAs in five subtypes and carcinoma and normal tissues were 2668, 2890, 2838, 2634, 2265, 2718, respectively (Table 2). DE-PCGs in five subtypes and carcinoma and normal tissues were 3702, 3989, 3934, 3647, 3013, 3642, respectively (Table 2 and Figure 2G). It was evidently exhibiting that C5 subtype was holding less differential mRNAs and lncRNAs, on the contrary, C2 and C3 subtypes were holding more differential mRNAs and lncRNAs than other subtypes. A total of 4253 DE-lncRNAs and 5808 differential expression of

PCGs were obtained. The upregulated lncRNAs were remarkably more than down-regulated lncRNAs (Figure 2A–2F). Whereafter, we downloaded lncRNAs closely related to disease from the lncRNA Disease and lnc2Cancer database and 611 lncRNAs were obtained, which were compared to 4253 subtype-specific lncRNAs (Figure 2H). There were 120 lncRNAs contained in previous DE-lncRNAs and the significance of these lncRNAs was tested by hypergeometric test ($p=0.003929$) (Supplementary Table 1). In order to understand the clustering situation of these DE-lncRNAs, gene set enrichment analysis (GSEA) was employed based on fold change of each lncRNA in five molecular subtypes and carcinoma and normal tissues (Figure 3A–3F). The results indicated that these DE-lncRNAs were always clustered in the gene set with larger difference multiples. There was almost no intersection of lncRNAs between all five molecular subtypes and carcinoma and normal tissues. Most lncRNAs were merely enriched in one subtype, such as, subtype 1 having 237 lncRNAs; subtype 2 having 202 lncRNAs; subtype 3 having 226 lncRNAs; subtype 4 having 102 lncRNAs; subtype 5 having 53 lncRNAs and carcinoma and normal tissues having 61 lncRNAs, respectively. The remaining lncRNAs were focused on two kinds of subtypes or three kinds of subtypes. It was obviously seen carcinoma and normal tissues shared a great deal of lncRNAs with other subtypes (Figure 3G). This phenomenon highly suggested that these identified lncRNAs are distinctively enriched in specific molecular subtypes.

Co-expressing modules between lncRNAs and PCGs based on WGCNA

Co-expression between lncRNAs and PCGs was analyzed by hierarchical clustering. We defined the samples with distance more than 150000 as outlier samples for screening, and finally got 492 CRC cases for co-expressing analysis (Figure 4A). We accepted 3 as the final soft threshold in order to ensure the network belonging to non-scale property (Figure 4B, 4C), which can effectively strengthen strong correlation and weaken weak correlation or negative correlation. Next, the expression matrix was transformed into adjacency matrix, and then the adjacency matrix was transformed into topological overlap matrix (TOM) for reduced noise and false correlation. According to the dynamic tree cut criterion linked with TOM structure, the threshold for module partitions was constructed at 30. After determining the initial modules, the next was to establish new modules presenting as different colors, including calculating specific eigengenes of each module; clustering analysis of each module; grouping close modules into one new module with height = 0.25, deepsplit = 2, minModuleSize = 30. A total of 27 modules (Figure 4D) were eventually obtained and the module of grey was signified with genes which were unable to be classified into other modules. P value represents the significant aggregation degree and fold change represents the aggregation multiple. There was no significant difference between lncRNAs and PCGs among 27 modules (Table 3).

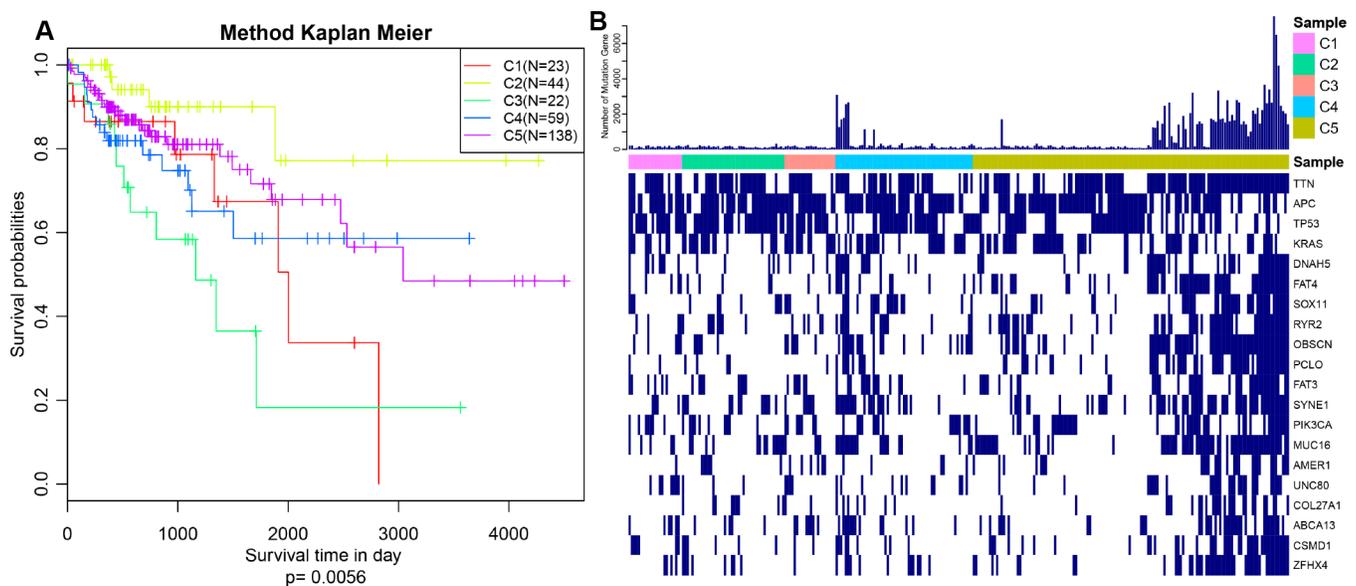


Figure 1. Prognosis description and profiling of mutation genes across five subtypes. (A) Kaplan–Meier plot analysis for five subtypes identified by iCluster (C1, C2, C3, C4, C5) is shown for overall survival (OS). ($P=0.0056$) (B) Exhibition of top 20 mutated genes among five molecular subtypes.

Table 2. Differentially expressed protein-coding genes (DE-PCGs) and lncRNAs (DE-lncRNAs) between tumors and adjacent tissues (subtype All) and five subtypes (C1, C2, C3, C4, C5).

Type	C1	C2	C3	C4	C5	All
PCG_Down	1843	1696	1996	1401	1296	1308
PCG_Up	1859	2293	1938	2246	1717	2334
PCG_All	3702	3989	3934	3647	3013	3642
Lnc_Down	1329	1217	1439	993	952	954
Lnc_Up	1339	1673	1399	1641	1313	1764
Lnc_All	2668	2890	2838	2634	2265	2718

For the interest of finding biologically associated modules, we need to correlate modules with external information, for instance clinical data of array information and functional enrichment analysis of gene information. Therefore, we executed correlation analysis by combining patients' gender, age, height, weight, BMI, T, N, M, stage with each module (Figure 5A). What could be clearly indicated in the correlation analysis that all modules except white, pink, black, darked were related to at least one phenotype. Besides, four modules: tan, blue, yellow,

magenta were finally selected with more than three kinds of phenotypic correlations for functional enrichment analysis and the significant enrichment pathways $p < 0.05$ were picked up. These four kinds of modules were clustered in distinguished pathways and the cross-talk between these four modules was missing (Figure 5B), suggesting each module may drive in different biological function. Tan module was associated with 18 pathways (Figure 6A, Supplementary Table 2) and neuroactive ligand-receptor interaction and calcium signaling pathway were in the highest flight in this module. MAPK signaling pathway also occupied superior gene ratio in this module and had been proved holding significant position in CRC-promoting process. Only two pathways were enriched in yellow module, of which microRNAs in cancer may possess outstanding role (Figure 6B, Supplementary Table 3). Besides, blue module enriched in 47 pathways (Figure 6C, Supplementary Table 4), and the remarkable pathways included neuroactive ligand-receptor interaction, PI3K-AKT signaling pathway and MAPK signaling pathway. Genes in magenta module mainly enriched in Wnt signaling pathway, TGF-beta signaling pathway and DNA replication which were implicated with classical undertakings in carcinogenesis, especially in CRC (Figure 6D, Supplementary Table 5). Aberrant

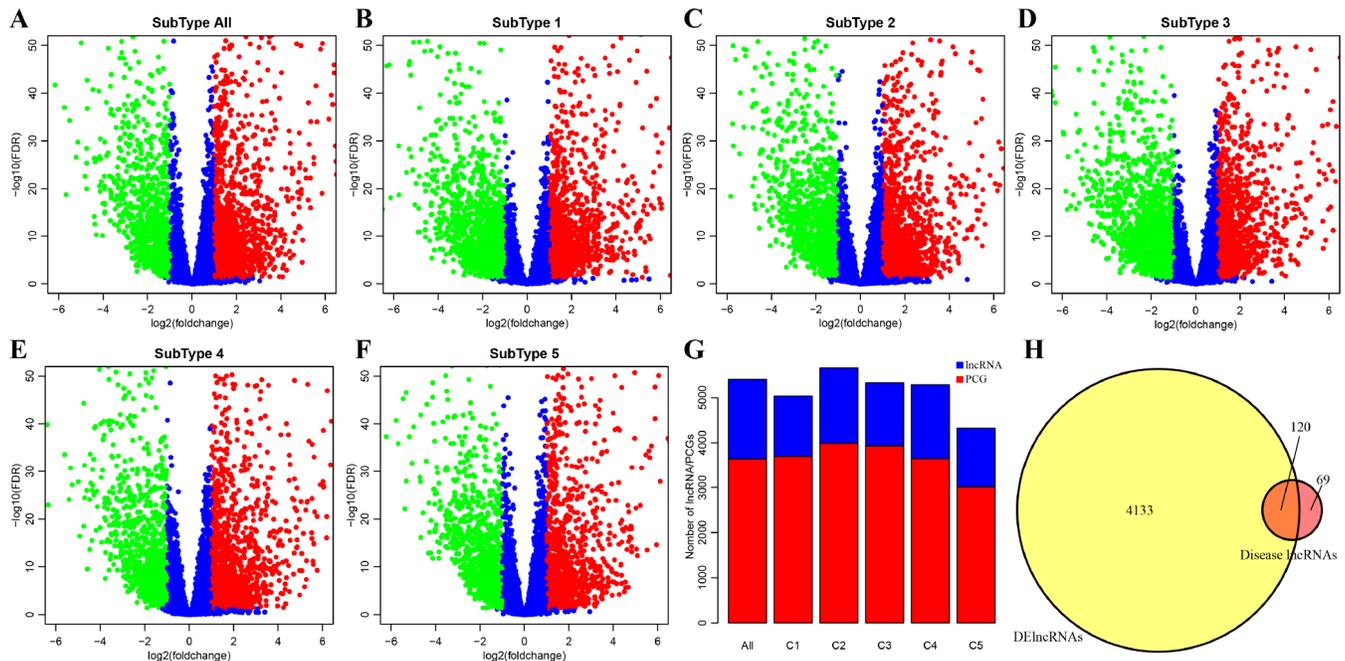


Figure 2. Identification of key differentially expressed protein-coding genes and lncRNAs among five subtypes. (A) Volcano plot shows the differential expression pattern of lncRNAs between tumors and adjacent normal tissues (named “subtype all”). The upregulated genes are shown as red and the downregulated genes are shown as green. (B–F) The up-(red) and down-(green) regulated lncRNAs across five molecular subtypes (subtype 1, 2, 3, 4, 5) are also shown as volcano plots, respectively. (G) Distribution of DE-lncRNAs and DE-PCGs among five subtypes (C1, C2, C3, C4, C5) and tumors and adjacent normal tissues (named “subtype all”) are shown. lncRNAs are presented as blue and PCGs are shown as red. (H) Venn diagram displays the intersection of DE-lncRNAs and Disease lncRNAs. Total 120 overlapped lncRNAs are indicated. Remaining 69 disease lncRNAs are excluded in our research.

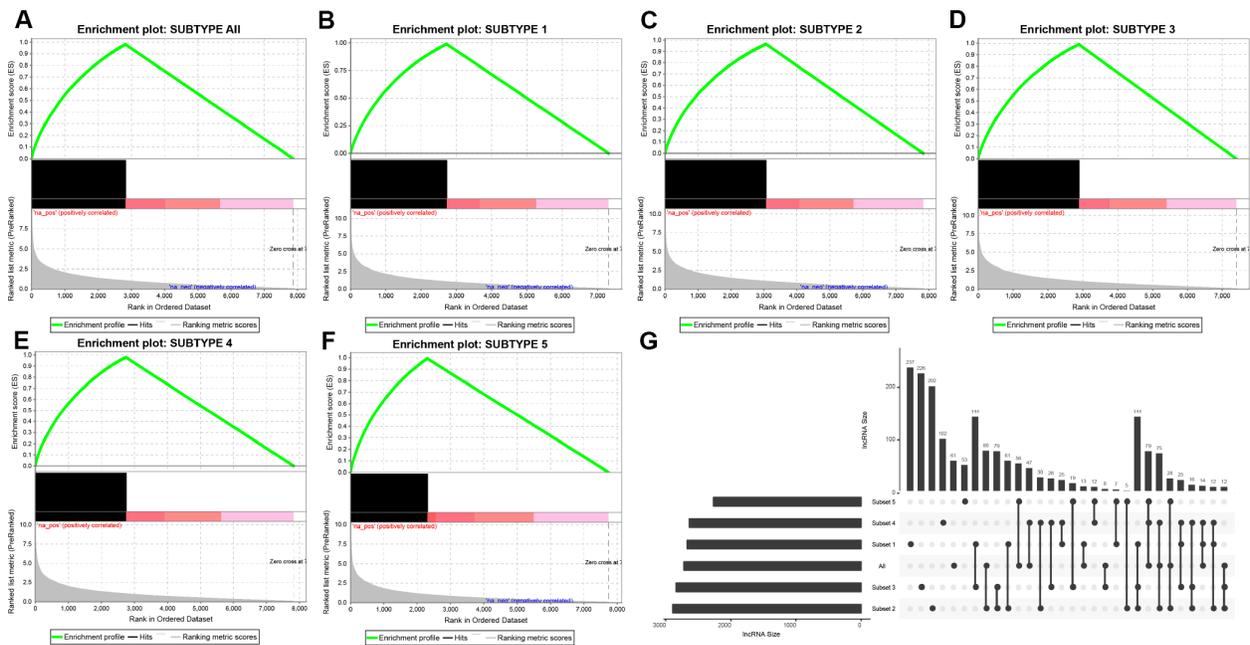


Figure 3. Distribution condition of DE-lncRNAs in subtypes. (A–F) GSEA analysis shows different enrichment states among these five subtypes (subtype 1, 2, 3, 4, 5) and tumors and adjacent normal tissues (named “subtype all”) based on difference multiple. The enriched lncRNAs are mainly focused on the left (presented as black bulks), which referring to larger difference multiple. (G) The overlapped lncRNAs exist in subtypes. The dot represents subtype and the line represents the overlapped lncRNAs across subtypes. lncRNA size points to the amount of DE-lncRNAs.

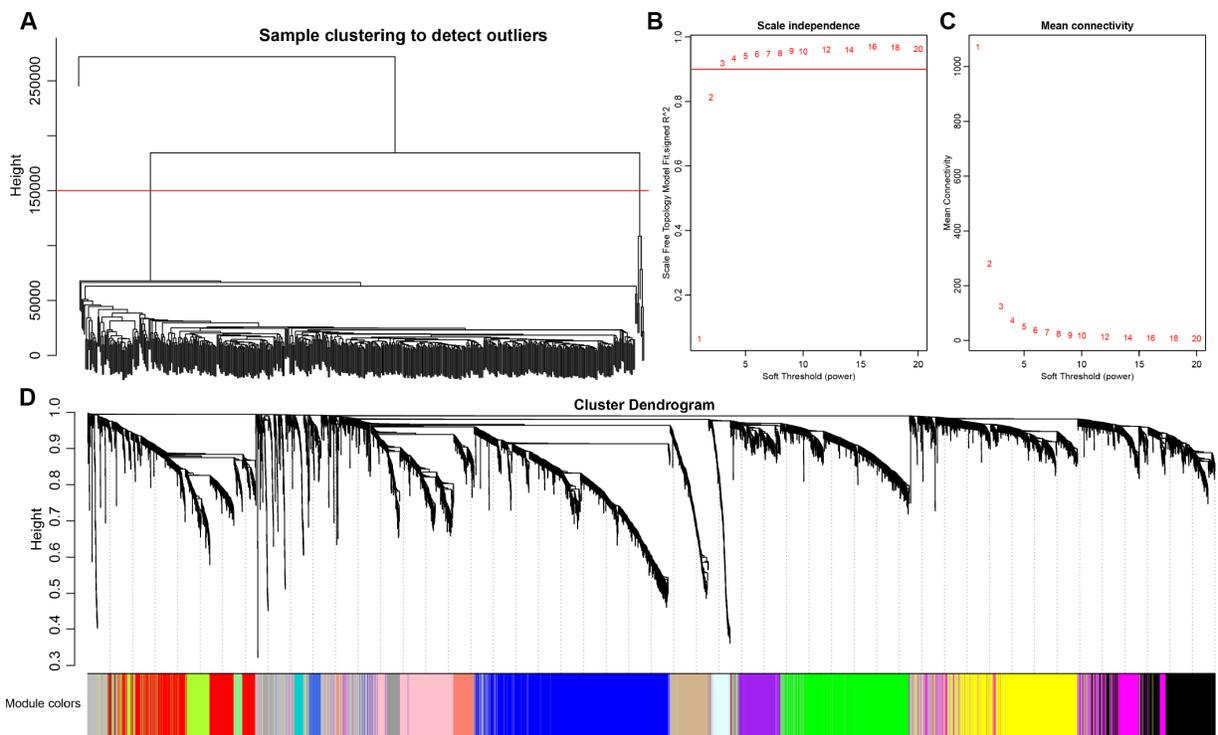
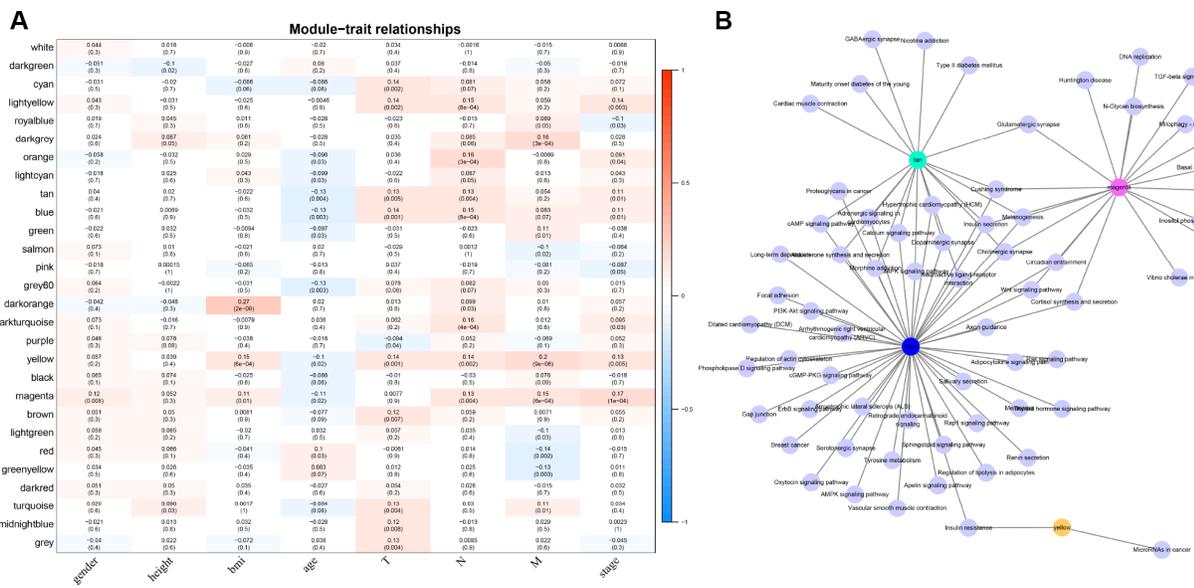


Figure 4. Weighted gene co-expression network analysis (WGCNA) to identify clinical modules. (A) Hierarchical clustering is applied to exclude some outliers samples. Objects with height greater than 150000 (upper the red line) are excluded. (B–C) Analysis of network topology for soft thresholding powers. The red line represents the square of correlation coefficient reaches to 0.9. The mean connectivity under different values of ‘power’ are shown. (D) Modules dendrogram of samples and modules are indicated by different colors.

Table 3. Coordination expression of PCGs and lncRNAs in 27 modules.

Module	All	Lnc	PCG	p.value	fc
green	697	299	398	0.378808	1.025934
cyan	139	58	81	0.584449	0.977854
tan	205	84	121	0.672971	0.948037
magenta	350	149	201	0.474915	1.012328
blue	1083	436	647	0.92705	0.920266
greenyellow	222	104	118	0.09268	1.203601
brown	908	378	530	0.671729	0.973974
purple	248	103	145	0.618262	0.970064
yellow	726	303	423	0.633562	0.978213
pink	390	163	227	0.596578	0.980602
royalblue	70	30	40	0.506279	1.024218
darkgrey	48	20	28	0.588628	0.975446
turquoise	1253	523	730	0.669036	0.978386
black	464	201	263	0.336854	1.04369
lightcyan	99	44	55	0.366243	1.092499
darkturquoise	51	19	32	0.807053	0.810839
grey60	77	34	43	0.410796	1.079796
salmon	142	64	78	0.275304	1.120512
white	36	16	20	0.458678	1.092499
darkorange	41	22	19	0.093987	1.581249
orange	41	16	25	0.716994	0.874
lightyellow	72	32	40	0.397353	1.092499
darkgreen	63	31	32	0.16111	1.322949
red	521	211	310	0.812304	0.929506
darkred	66	24	42	0.864677	0.780357
lightgreen	73	32	41	0.437124	1.065853
midnightblue	109	44	65	0.690877	0.924423



activation of WNT signaling pathway markedly increases the expression of β -catenin, thus promoting normal colon epithelial cells to infinitely proliferate and cancerate [25–27]. Moreover, TGF-beta/smad signaling pathway can advance invasion and metastasis of CRC by means of immune suppression, angiogenesis, augmenting interaction between tumor cells and extracellular matrix [28–30]. Overall, DE-lncRNAs and PCGs co-expression modules constructed by WGCNA were enriched in various pathways implicating important functions in regulating CRC progression.

Identification of CNVs associated lncRNAs in CRC

CNV is a frequent form of genomic structural changes, which is closely related to occurrence and deterioration of tumors. We profiled the spectrum of CNV-related lncRNAs on the whole genome in CRC patients. The frequency of copy number deletions were largely more than amplifications, suggesting copy number deletions may predominantly implicate in CRC, and majority

deletions were found on chromosome 8, in contrast, the amplifications were mainly concentrated on chromosome 20 (Figure 7A). The correlation between the expression profiles of lncRNAs and CNVs was demonstrating a positive relationship and the distribution was significantly higher than the random, $p < 1e-16$ (Figure 7B). Moreover, the frequently changed regions in CRC genome were identified on the basis of GISTIC algorithm. A genome-wide view of the CNVs was shown in Figure 7C and the overall frequent change regions were complex. The focal amplification events (indicated by dashes), surpassing the significance threshold (green line), mainly distributes among chromosomes 8, 11, 12p13, 16q12, 20p11, and the focal deletions events were concentrated on 1p33, 3q26, 4p16, 5q11, 5q22, 20p12. Similarly to previous condition, frequent copy number deletions of lncRNAs were largely more than the copy number amplifications in the whole genome. This finding emphasized the impact of copy number deletions of lncRNAs in CRC, and it may be significantly related with prognosis of CRC patients.

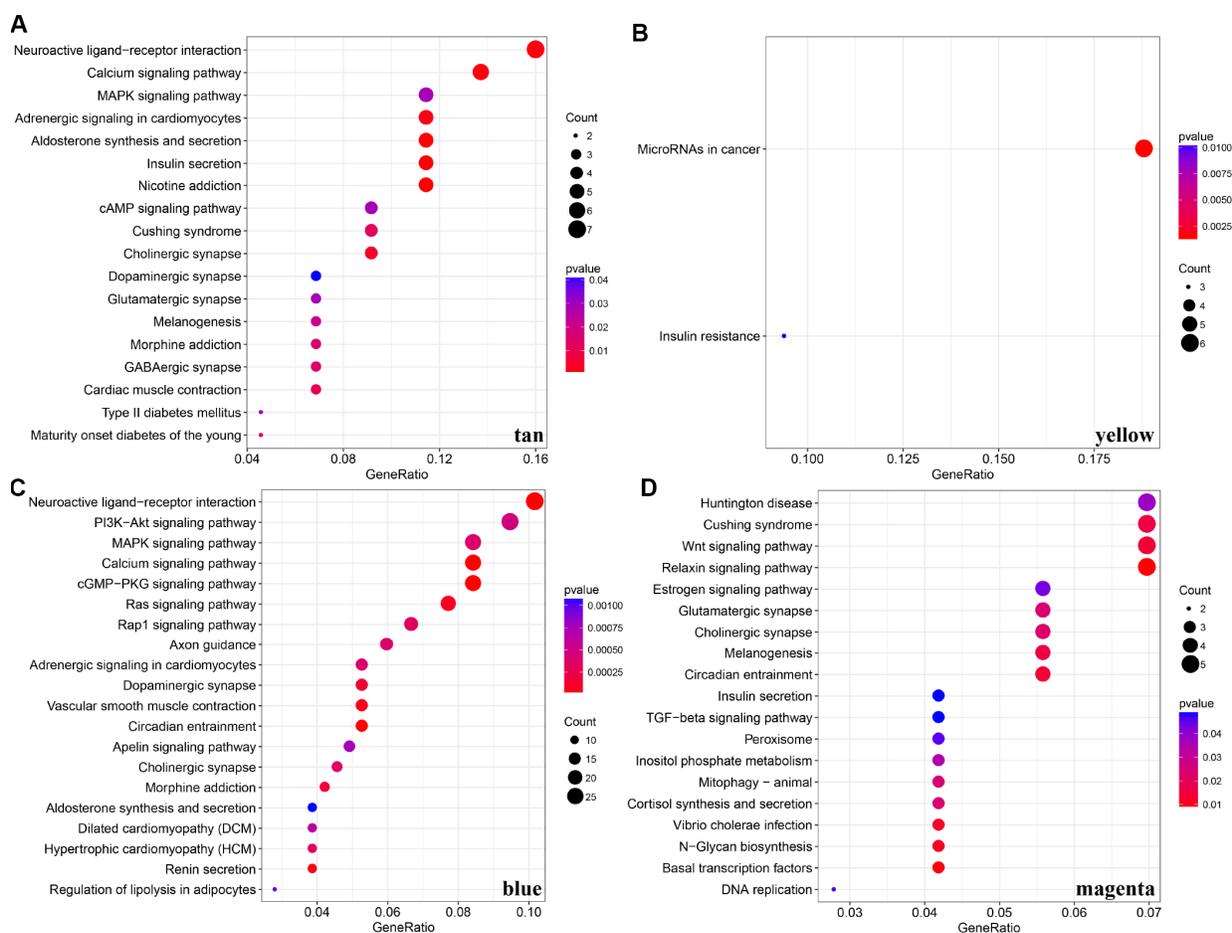


Figure 6. KEGG pathway enrichment analyses for genes existing in these four modules (tan, yellow, blue, magenta). (A–D) KEGG pathway enrichment analyses for significant pathways in modules, including tan, yellow, blue, magenta. KEGG: kyoto encyclopedia of genes and genomes

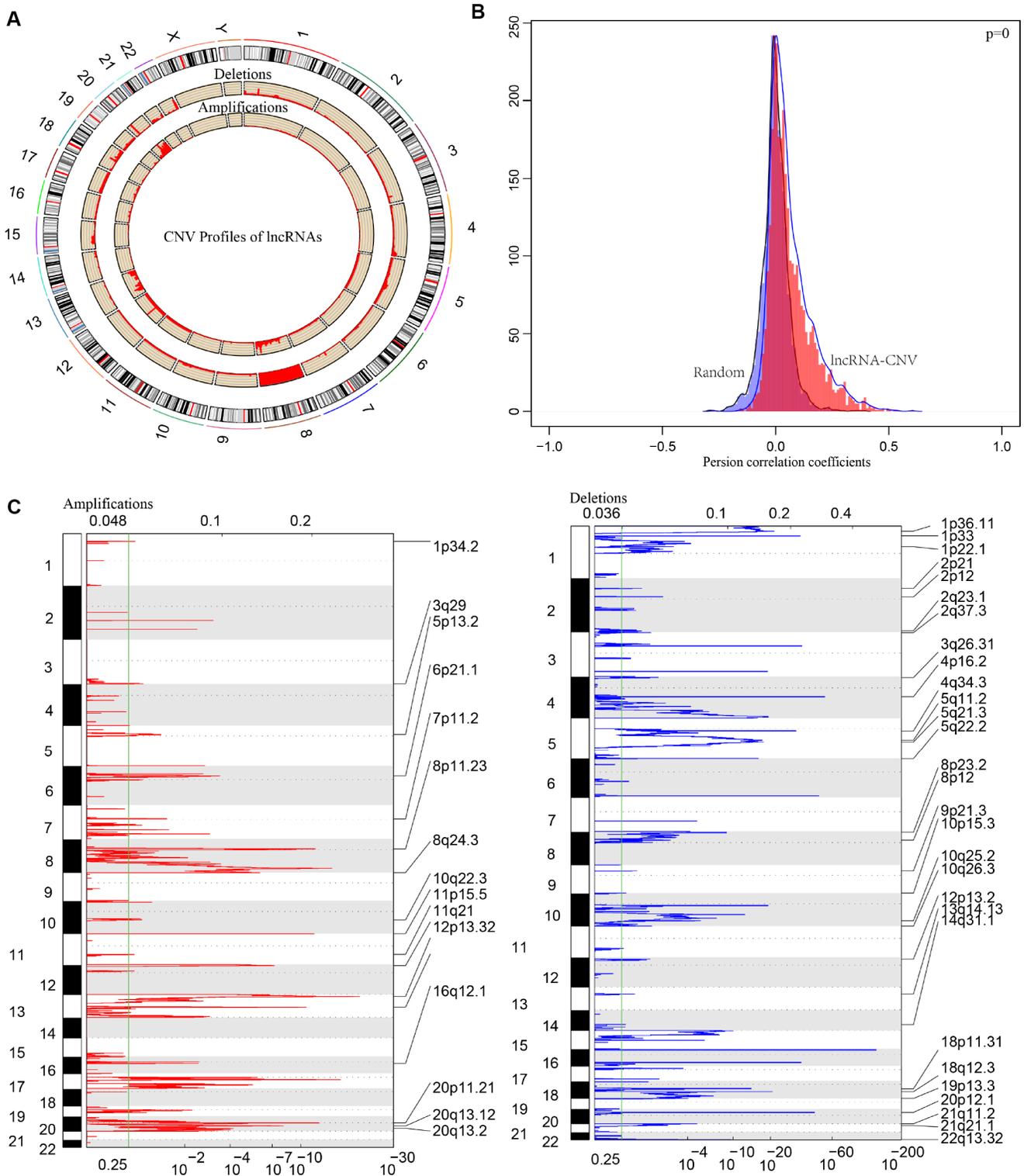


Figure 7. The pattern of CNVs profiles in the whole genomes. (A) The proportional frequencies of copy number deletions or amplifications of lncRNAs existing in the whole genome. CNV deletions of lncRNAs are mainly concentrated on 8 chromosome, and amplifications are focused on 8, 13, 20 chromosomes. The frequency of deletions are largely general than amplifications. (B) Distribution of correlation coefficient between copy number alternations of lncRNAs and the expression level of lncRNAs is shown. The correlation coefficient of CNV-lncRNA greater than 0 represents the regulation relationship is positive. (C) lncRNAs located in the focal peaks are CRC-related. False-discovery rates (q values) and scores from GISTIC 2.0 for alterations are plotted in x-axis, and the genome positions are shown as y-axis; dotted lines indicate the centromeres (distinguishing chromosome long arm from short arm). Amplifications (left, red) and deletions (right, blue) of lncRNAs are also shown. The green line represents 0.25 q value cut-off point that determines significance.

In order to deeply explore the effect of CNV on transcription of lncRNAs, we screened out 17 lncRNAs with more than 7% of the frequency of CNV in 452 CRC samples (Supplementary Table 6). In the light of these 17 lncRNAs, we analyzed the expression difference of each lncRNA between the samples with copy deletion or copy amplification and normal group. A total of 12 lncRNAs were obtained according to the criterion in which the expression in each group was greater than 0. There are four lncRNAs: CASC11, HM13-AS1, ABALON, NKILA highly expressed in amplification group than diploid group ($p < 0.05$), and FAM87A, LOC101927752, KBTBD11-OT1, LOC100287015, LOC101929066, these five lncRNAs were low-expressed in copy number deletion group than normal copy number group ($p < 0.05$) (Figure 8). All these findings indicated that deletions or amplifications of lncRNAs may closely conduct the expression level of lncRNAs. However, these differentially expressed lncRNAs are deserved to explore for the potential clinical significance based on large-scale samples supporting.

lncRNAs-based prognostic biomarkers in CRC patients

As we all know, lncRNAs are holding diving or inhibiting role in tumors and increasing numbers of lncRNAs have been conducted as prognosis associated biomarkers in CRC. In our research, we wondered whether the expression profiles of lncRNAs regulated by copy number deletions or amplifications can impact survival status of CRC patients. Based on the previous five molecular subtypes, we systematically analyzed the copy number alternations of these DE-lncRNAs among five subtypes, and then, 104 lncRNAs conforming to the following criterions were obtained (Supplementary Table 7): CNV change frequency among each sample is more than 0.1%; holding difference at least five subtypes; positively related with CNVs. We then wiped off the samples of whom the expression of each lncRNA was under zero, and subsequently correlated lncRNAs with disease-free survival. We finally acquired 10 lncRNAs with significant prognosis ($p < 0.05$) based on the quartiles method (Figure 9, Supplementary Table 8). LOC101927604, CASC15, CLDN10-AS1, C14orf132, LMF1 and LINC00675 were differentially expressed among all six subtypes (including normal group), and the expression levels of LOC105377267, LINC-PINT, CCDC144NL-AS1 and LOC284454 were differential in five subtypes (Table 4).

In addition, we verified the prognostic value of identified 10 lncRNAs in GEO (Gene Expression Omnibus) dataset (GSE39582) based on GPL570 platform (<https://www.ncbi.nlm.nih.gov/geo/>). There were about 8 lncRNAs successfully annotated to probes existing in GEO platform and only 5 lncRNAs, CASC15, LINC-

PINT, C14orf132, LMF1 and CCDC144NL-AS1 holding significant prognosis difference ($p < 0.05$) (Figure 10).

DISCUSSION

With next-generation sequencing and mass spectrometry technology sprouting quickly, biological complexity of tumors has been illuminated progressively. Previous knowledge have proved integrative clustering managements are capable to distinguish heterogeneity of tumors and to disclose practical prognostic signatures. The general integrative clustering methodologies include iCluster [19], intNMF (non-negative matrix factorization) [31], similarity network fusion (SNF) [32]. In this study, we firstly performed multi-dimensional data process according to iCluster which is based on a joint latent variable model in CRC. The most attractive signature of iCluster is that incorporates unobserved variables estimated from copy number data, mRNA expression data, methylation and others, simultaneously reducing the dimensionality of the datasets without changing sample size. A final cohort of 289 CRC patients were available for our algorithm matrix, and three omics data were covered by genomic and epigenomic, including mRNA sequencing, CNV and methylation downloading from TCGA, respectively. The integrative analysis unveiled five subtypes with distinctive molecular signatures and prognostic relevance. C3 subtype is exhibiting terrible prognosis and C2 subtype holds the most satisfactory outcome, which is extremely corresponding to distribution of TNM stage. A recent investigation has shown that copy number correlated genes (CNVcor) and methylation correlated genes (METcor) are co-regulated significantly and integration of CNVcor and METcor genes revealed three molecular subtypes of liver cancer [33]. Glioblastoma multiforme was subdivided into three molecular subtypes, namely subtype 1, subtype 2 and subtype 3 based on aggregation of DNA copy number, methylation and gene expression levels. Subtype 1 is characterized by hypermethylation involved in brain development and neuronal differentiation, and subtype 2 is presenting highly promoter methylation of homeobox and G-protein signaling genes, then, subtype 3 is featured with NF1 and PTEN alterations [34]. In our research, we also compared the mutation profiles of five molecular subtypes, and four mutation-related genes attracted our attention, TTN, APC, TP53 and KRAS, which are holding the highest mutation rate almost existing across five subtypes. Exome sequencing was performed to identify recurrent somatic mutations with prognostic significance and eventually APC, TP53 and KRAS were successfully diagnosed as mutation signatures in CRC [35].

The onset and progression of carcinogenesis is usually involved with thousands of genomic variants, including

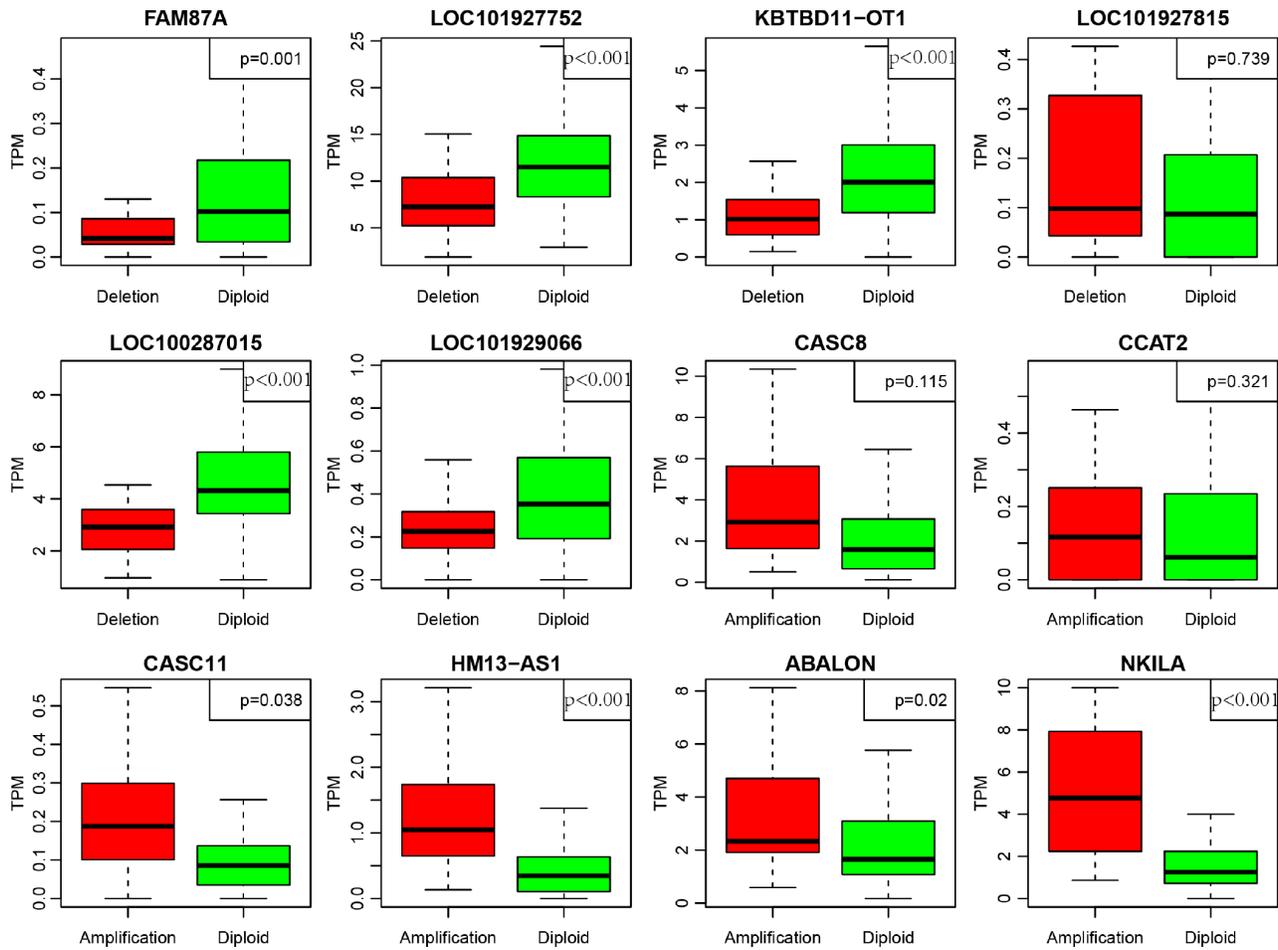


Figure 8. Deregulation of lncRNAs induced by copy number deletions or amplifications ($P<0.05$). Green (diploid) represents normal copy. Red (deletion or amplification) represents the variant copy.

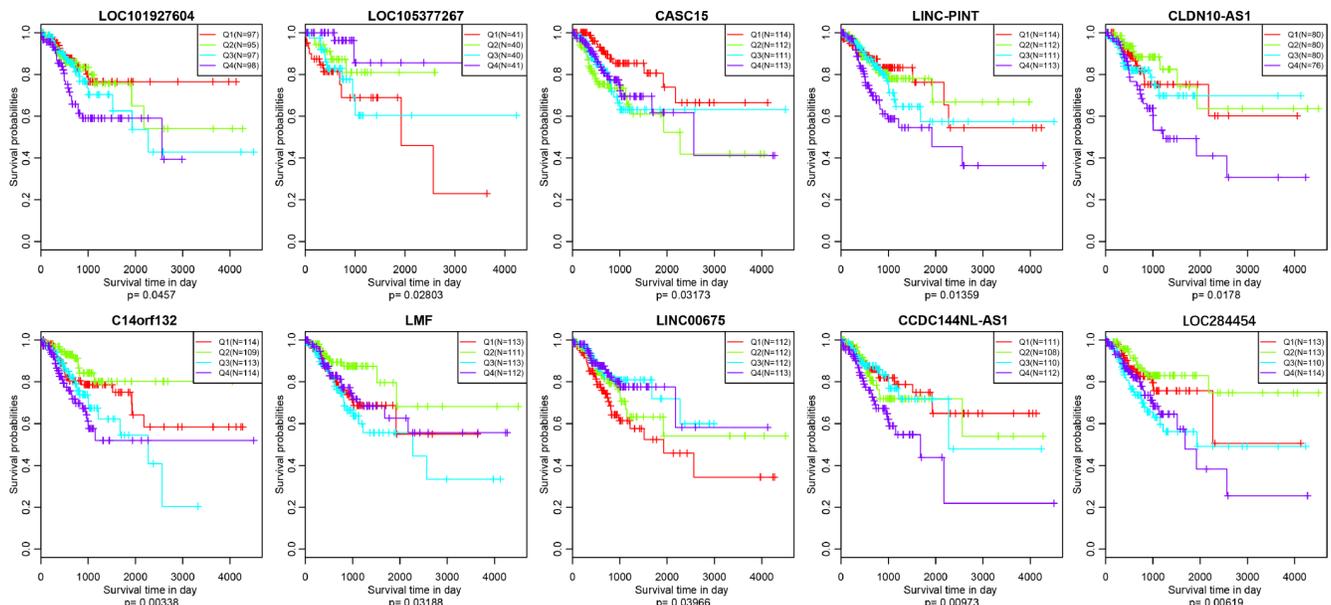


Figure 9. Kaplan–Meier plot analysis shows disease-free survival (DFS) based on quartile method in TCGA ($p<0.05$).

Table 4. 10 lncRNAs with significant prognosis among five subtypes (including “normal group”).

lncRNA	CNV frequency	PCC	Number of DE Subtype	p.value	GPL570.Probe
LOC101927604	0.004425	0.121608	6	0.0457	1557702_at
LOC105377267	0.004425	0.065193	5	0.028027	NA
CASC15	0.006637	0.101133	6	0.031734	229280_s_at
LINC-PINT	0.002212	0.033925	5	0.013595	228702_at
CLDN10-AS1	0.037611	0.18445	6	0.017802	1570291_at
C14orf132	0.004425	0.044214	6	0.003379	231859_at
LMF1	0.00885	0.027102	6	0.031878	46142_at
LINC00675	0.019912	0.340102	6	0.039658	215658_at
CCDC144NL-AS1	0.011062	0.082301	5	0.009727	229669_at
LOC284454	0.011062	0.156642	5	0.006193	NA

small-size mutations (SNPs) and large-scale genomic changes (CNVs), such as copy number deletions, duplications or amplifications. CNVs, hallmarks of cancer, often lead to aberrance in copy numbers, including amplification, gain, loss and deletion. CNVs are taking serious responsibility in regulating expression of PCGs and non-coding RNAs and the activations of multiple signaling pathways. It has been well-known that CNVs may conduct significant effects on various tumorigenesis, such as ovarian cancer [36], bladder cancer [37], hepatocellular carcinoma [38] and so on. Early stages

(stages I and II) CRC were obviously exhibiting the most frequent deletions involved in chromosomes 6, 8p, 14q and 1p and the most frequent amplifications were mainly located in chromosome 19, 5, 2, 9p and 20p according to detection of CNVs in plasma [39]. The expression level and copy number of UQCRB protein (ubiquinol cytochrome c reductase binding protein), implicated in mitochondrial complex III stability, were unanimously upregulated in CRC, supporting CNVs induced deregulation effect [40]. Our study established novel regulatory factors in CRC, and lncRNAs (CASC11,

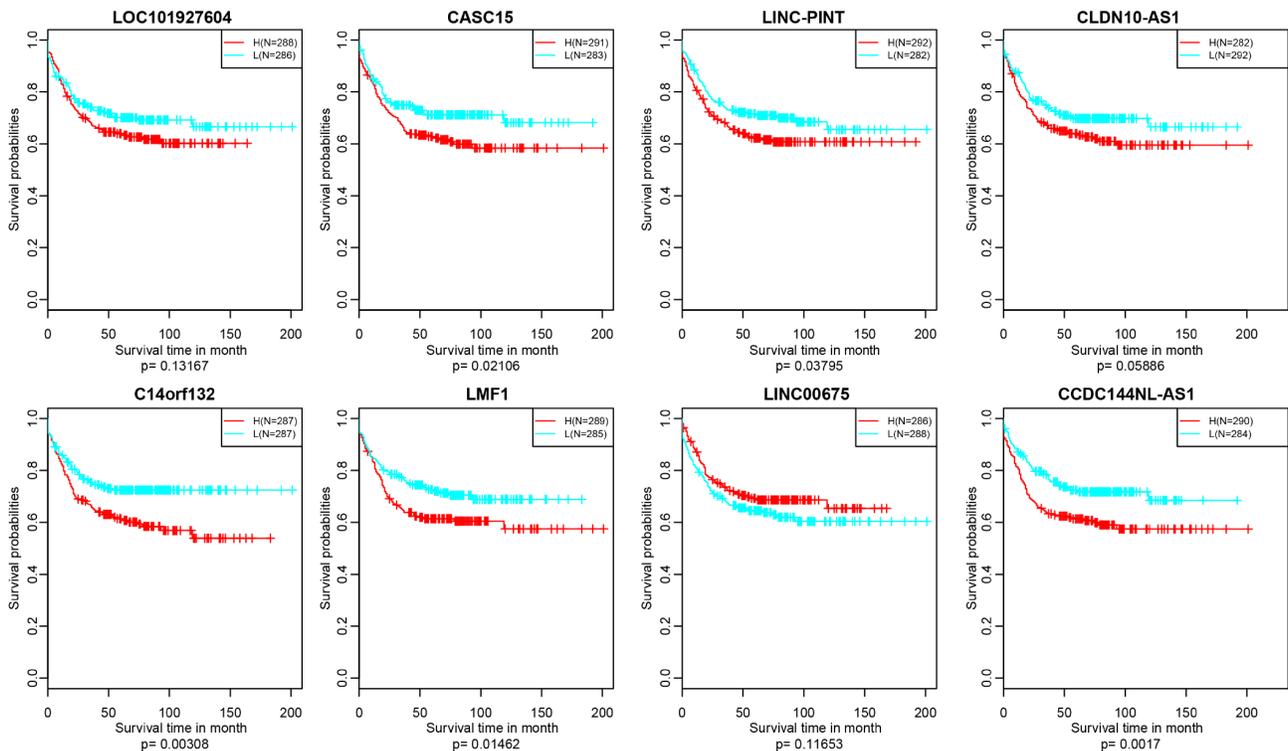


Figure 10. Kaplan–Meier plot analysis shows relapse-free survival (RFS) validated in GEO datasets ($p < 0.05$).

HM13-AS1, ABALON, NKILA, FAM87A, LOC101927752, KBTBD11-OT1, LOC100287015, LOC101929066) were acting as new candidates for CRC diagnosis. CASC11 is an oncogenic lncRNA, which involved with influencing tumor cell stemness, cancer cell proliferation and epithelial-mesenchymal transition (EMT) in CRC, small cell lung cancer, bladder cancer, prostate cancer, hepatocellular carcinoma (HCC) [41–45]. NKILA, an NF- κ B-interacting lncRNA, is determined for tumor-mediated T cell AICD (Activation-induced cell death) by inhibiting NF- κ B activity, and NKILA overexpression in tumor-specific cytotoxic T lymphocytes (CTLs) is correlated with increased apoptosis and shorter survival in lung cancer and breast cancer [46]. Thus, we proposed a hypothesis that copy number of NKILA will be significantly amplified when CTLs are activated in CRC, following the increased expression of NKILA. We introduced a new mechanism underlying deregulated lncRNAs in tumorigenesis and the inspection of CNVs are convenient and precise.

Here, we described a picture for prognosis prediction of CRC with respect to CNVs relevant lncRNAs. In our research, we identified 10 lncRNAs associated with crucial clinical outcome in CRC, of which 5 lncRNAs (CASC15, LINC-PINT, C14orf132, LMF1 and CCDC144NL-AS1) were validated in GEO datasets. Cancer susceptibility candidate 15 (CASC15) is familiarly observed to participate in regulation of tumor proliferation, metastasis, and worsen survival probability of human cancer including CRC, cervical cancer, breast cancer and so on [47–49]. Jing et al demonstrated that CASC15 advances proliferation and metastasis on the basis of activating Wnt/ β catenin signaling pathway [50], and some other examinations displayed CASC15 may act as an oncogene, performing a pushing role in the progression of HCC, and upregulated expression of CASC15 is tied with imperfect prognosis [51]. Notwithstanding, copy number changes associated with CASC15 in tumors are far from being understood. p53-regulated human lncRNA (LINC-PINT) has been noticed functioning as a tumor suppressor by impeding proliferation of cancer cells in gastric cancer and glioblastoma [52, 53], and down-regulated LINC-PINT in pancreatic tumor may deliver unsatisfactory outcome [54]. C14orf132 (chromosome 14 open reading frame 132) gene is a novel long non-coding RNA (lncRNA) with unknown functions implicated in tumors, and the only investigation elucidated in C14orf132 is concerned with extremely low birth weight [55]. Lipase maturation factor 1 (LMF1) is a profound regulator of plasma lipid metabolism and majority studies mainly focused on mutations of LMF1 determining severe hypertriglyceridemia [56]. Up to now, there is still lacking of evidence about the latent position of LMF1 in tumorigenesis. Our inspection is the forerunner for

exploring the prognostic value of LMF1 in CRC patients. CCDC144NL-AS1 is upregulated in ectopic endometrial (EC) tissues than eutopic endometrial (EU) tissues, and simultaneously exhibits elevated expression in advanced EC tissues (III+IV) [57]. Nevertheless, the distinct significance of CCDC144NL-AS1 in CRC is still inexperienced. Despite the fact that, these mentioned lncRNAs are prone to govern prognosis prediction among multiple tumors, prognostic signatures of lncRNAs guided by copy number changes remains elusive. The five prognostic lncRNAs are deserved to be furtherly verified on the foundation of large-scale clinical CRC samples.

In summary, systemic administration of RNA-seq, methylation and CNVs data presents original subtype classification methodology of CRC. Identification of deregulated lncRNAs induced by CNVs introduces an unprecedented regulatory pattern involved with lncRNAs, and prognosis prediction for CNV-related lncRNAs may take insight into precision diagnostics and therapeutics for CRC patients.

METHODS

Data preparation and processing

Involved original data were downloaded from official TCGA data portal (<https://tcga-data.nci.nih.gov>), including RNA sequencing data, 450k methylation data, CNV data, DNA mutation data and clinical information. Subsequently, series of managements were performed, for instance, RNA sequence data: primary data were downloaded in “counts” form, and the processed FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) data were transformed to TPM (TranscriptsPerKilobase of exonmodel per Million mapped reads). The gene expression profiles of 458 CRC patients and 41 normal cases were acquired. Then, we defined lincRNA, sense-intronic, sense-overlapping, antisense, processed-transcript, 3primer-overlapping as lncRNAs based on the genecode file. Meanwhile, gene type belonging to protein-coding was assigned to PCGs; Next, administrations of 337 CRC patients holding 450k methylation data were involved with removing of NA probes, cross-reactive CpG sites [18] and unstable CpG sites existing in the sex chromosomes and single nucleotide sites; Then, CNV data and single nucleotide mutation data were pictured, with the removed germline difference and processed by “mutect software”, respectively. Eventually, we also obtained corresponding clinical information of 458 CRC patients. The patients with following time less than 30 days were removed. Underlying these prepared data, we employed R package-iCluster to integrate PCGs, methylation and CNVs for subtypes clustering. Number of clusters was set to 5 and 50 iterations were set to identify stable

samples clusters with default parameters [19]. The R code can be downloaded at <http://www.mskcc.org/mskcc/html/85130.cfm>.

Differentially expressed lncRNAs and mRNAs across five subtypes

The R package-DESeq2 was claimed for determining these differentially expressed lncRNAs and mRNAs between five subtypes and normal tissues specimens, with the criteria of fold change >2 , FDR <0.05 . Genes with count <1 were rejected among the whole genome profiles. The basic data structure consisted of two tables, including countData and colData. Gene set enrichment analysis (GSEA) was furtherly applied to distinguish the distribution of DE-lncRNAs and the difference multiple was used for ranking.

Weighted gene co-expression network analysis (WGCNA) to identify phenotypes-related modules

WGCNA is often regarded as efficient methodology applying in picking up co-expressed gene modules and identifying potential therapeutic targets associated with clinical phenotypes [20]. In present research, based on the differential expression profiles of lncRNAs and PCGs, co-expression network was constructed using WGCNA. In order to make network conform to non-scale signature, the soft threshold was defined as 3. The non-scale signature means that $\log(k)$ is negatively correlated with $\log(P(k))$, and the correlation coefficient is greater than 0.8. The possible modules were identified with the average-linkage clustering and dynamic tree cut on the ground of height=0.25, deepSplit = 2, minModuleSize = 30. Considering the clinical information downloading from TCGA, the correlation between each module and corresponding clinical information was calculated. Immediately, annotations of potential pathways were targeted in order to screen out biologically interested modules based on the userListEnrichment function included in package WGCNA [21].

Copy number expression profiles of whole genomic lncRNAs

GISTIC (Genomic Identification of Significant Targets in Cancer) algorithm is responsible for identifying variant regions that are more prone to drive cancer pathogenesis [22]. It can visualize regions in the genome manifesting amplifications or deletions across thousands of samples. G-score is allocated to each alternational region to evaluate the amplitude of aberration and the frequency of occurrence [23]. Here, GISTIC2.0 software was applied to define the prepared CNVs profiles of all genes underlying 452 CRC samples, and lncRNAs-related CNVs expression profiles were extracted. We defined

copy number >1 or <-1 as copy number amplifications or deletions respectively. False Discovery Rate (FDR) q-values were assigned to each alternational region. “Peak regions”, also known as significantly aberrant regions, point to the greatest frequency and amplitude of aberrations. The “peak regions” aim at determining whether the signal is primarily due to broad events (longer than half a chromosome arm), focal events, or significant levels of both [23, 24].

CNV-related lncRNAs acting as prognostic biomarkers in CRC patients

CNVs profiles of DE-lncRNAs among five subtypes were described. Associated lncRNAs were prepared with the criterions as following: 1) the identified lncRNAs must be positively correlated with CNVs (correlation coefficient >0); 2) the identified lncRNAs should have significant difference across at least 5 subtypes (number of DEsubtype ≥ 5); 3) the identified lncRNAs hold copy number change ratio more than 0.1% in all involved cases (CNV frequency >0.001); 4) prognostic-related lncRNAs were marked with $p<0.05$ (samples were divided into four groups based on quartile method or two groups based on median level). The selected patients must manifest expression value of previous marked lncRNAs greater than 0.

Statistical analysis

All statistical strategies in our research were based on R 3.4.3 (<https://cran.r-project.org/bin/windows/base/old/3.4.3/>). All workings are dependent on default parameters except to special notes.

CONFLICTS OF INTEREST

All authors declare that they have no conflicts of interest.

FUNDING

This study was supported by the Jiangsu Health and Health Committee (H2018074), the Nantong Science and Technology Program (MS22015055), the National Natural Science Foundation of China (81702086), the Overseas Returnee Start Package (2015) from the Jiangsu Provincial Department of Human Resources and Social Security, the 226 Talent Training Project of Nantong City, and the Elite Doctor Program of the Affiliated Hospital of Nantong University.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN

- estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68:394–424.
<https://doi.org/10.3322/caac.21492> PMID:30207593
2. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet.* 2014; 383:1490–502.
[https://doi.org/10.1016/S0140-6736\(13\)61649-9](https://doi.org/10.1016/S0140-6736(13)61649-9) PMID:24225001
 3. Nakamura Y. DNA variations in human and medical genetics: 25 years of my experience. *J Hum Genet.* 2009; 54:1–8.
<https://doi.org/10.1038/jhg.2008.6> PMID:19158818
 4. Liang L, Fang JY, Xu J. Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy. *Oncogene.* 2016; 35:1475–82.
<https://doi.org/10.1038/onc.2015.209> PMID:26073079
 5. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, et al, and Wellcome Trust Case Control Consortium. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010; 464:704–12.
<https://doi.org/10.1038/nature08516> PMID:19812545
 6. Ledet EM, Hu X, Sartor O, Rayford W, Li M, Mandal D. Characterization of germline copy number variation in high-risk African American families with prostate cancer. *Prostate.* 2013; 73:614–23.
<https://doi.org/10.1002/pros.22602> PMID:23060098
 7. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science.* 2005; 307:1434–40.
<https://doi.org/10.1126/science.1101160> PMID:15637236
 8. Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Robertson-Lowe C, Marshall AJ, Poretto E, Hodges MD, Bhangal G, Patel SG, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature.* 2006; 439:851–55.
<https://doi.org/10.1038/nature04489> PMID:16482158
 9. Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K, Blanchong CA, McBride KL, Higgins GC, et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet.* 2007; 80:1037–54.
<https://doi.org/10.1086/518257> PMID:17503323
 10. Gugnoni M, Ciarrocchi A. Long Noncoding RNA and Epithelial Mesenchymal Transition in Cancer. *Int J Mol Sci.* 2019; 20:E1924.
<https://doi.org/10.3390/ijms20081924> PMID:31003545
 11. Han D, Wang M, Ma N, Xu Y, Jiang Y, Gao X. Long noncoding RNAs: novel players in colorectal cancer. *Cancer Lett.* 2015; 361:13–21.
<https://doi.org/10.1016/j.canlet.2015.03.002> PMID:25754818
 12. Xu MD, Qi P, Du X. Long non-coding RNAs in colorectal cancer: implications for pathogenesis and clinical application. *Mod Pathol.* 2014; 27:1310–20.
<https://doi.org/10.1038/modpathol.2014.33> PMID:24603586
 13. Li CH, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. *Int J Biochem Cell Biol.* 2013; 45:1895–910.
<https://doi.org/10.1016/j.biocel.2013.05.030> PMID:23748105
 14. Tang X, Qiao X, Chen C, Liu Y, Zhu J, Liu J. Regulation Mechanism of Long Noncoding RNAs in Colon Cancer Development and Progression. *Yonsei Med J.* 2019; 60:319–25.
<https://doi.org/10.3349/ymj.2019.60.4.319> PMID:30900417
 15. Ji Q, Zhang L, Liu X, Zhou L, Wang W, Han Z, Sui H, Tang Y, Wang Y, Liu N, Ren J, Hou F, Li Q. Long non-coding RNA MALAT1 promotes tumour growth and metastasis in colorectal cancer through binding to SFPQ and releasing oncogene PTBP2 from SFPQ/PTBP2 complex. *Br J Cancer.* 2014; 111:736–48.
<https://doi.org/10.1038/bjc.2014.383> PMID:25025966
 16. Takahashi Y, Sawada G, Kurashige J, Uchi R, Matsumura T, Ueo H, Takano Y, Eguchi H, Sudo T, Sugimachi K, Yamamoto H, Doki Y, Mori M, Mimori K. Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *Br J Cancer.* 2014; 110:164–71.
<https://doi.org/10.1038/bjc.2013.698> PMID:24196785
 17. Tsang WP, Ng EK, Ng SS, Jin H, Yu J, Sung JJ, Kwok TT. Oncofetal H19-derived miR-675 regulates tumor suppressor RB in human colorectal cancer. *Carcinogenesis.* 2010; 31:350–58.
<https://doi.org/10.1093/carcin/bgp181> PMID:19926638

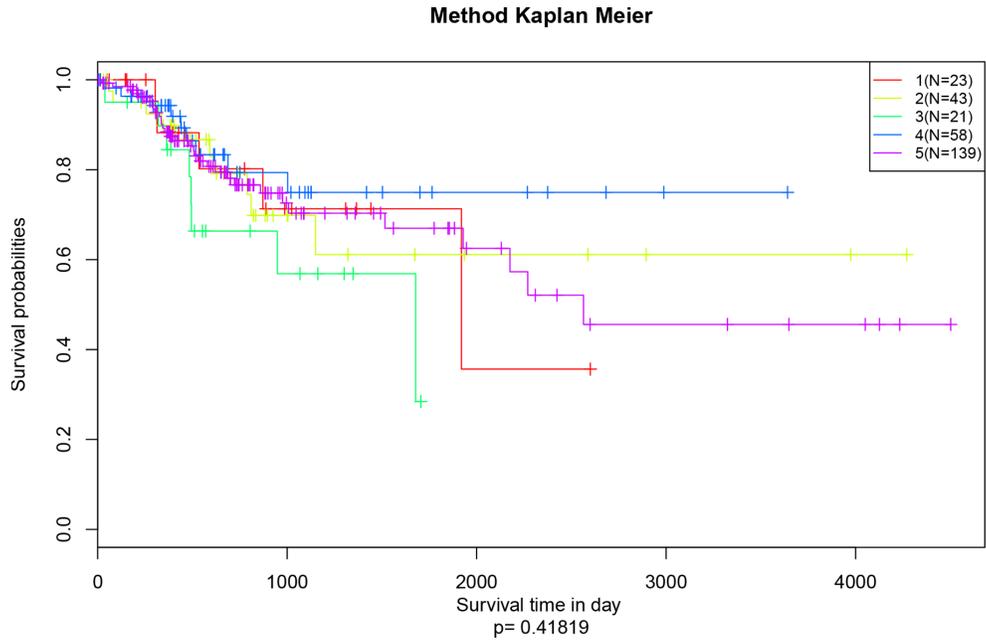
18. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013; 8:203–09. <https://doi.org/10.4161/epi.23470> PMID:[23314698](https://pubmed.ncbi.nlm.nih.gov/23314698/)
19. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*. 2009; 25:2906–12. <https://doi.org/10.1093/bioinformatics/btp543> PMID:[19759197](https://pubmed.ncbi.nlm.nih.gov/19759197/)
20. Zhai X, Xue Q, Liu Q, Guo Y, Chen Z. Colon cancer recurrence-associated genes revealed by WGCNA co-expression network analysis. *Mol Med Rep*. 2017; 16:6499–505. <https://doi.org/10.3892/mmr.2017.7412> PMID:[28901407](https://pubmed.ncbi.nlm.nih.gov/28901407/)
21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9:559. <https://doi.org/10.1186/1471-2105-9-559> PMID:[19114008](https://pubmed.ncbi.nlm.nih.gov/19114008/)
22. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011; 12:R41. <https://doi.org/10.1186/gb-2011-12-4-r41> PMID:[21527027](https://pubmed.ncbi.nlm.nih.gov/21527027/)
23. Beroukhi R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, Du J, Kau T, Thomas RK, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc Natl Acad Sci USA*. 2007; 104:20007–12. <https://doi.org/10.1073/pnas.0710052104> PMID:[18077431](https://pubmed.ncbi.nlm.nih.gov/18077431/)
24. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010; 463:899–905. <https://doi.org/10.1038/nature08822> PMID:[20164920](https://pubmed.ncbi.nlm.nih.gov/20164920/)
25. Ma Y, Yang Y, Wang F, Moyer MP, Wei Q, Zhang P, Yang Z, Liu W, Zhang H, Chen N, Wang H, Wang H, Qin H. Long non-coding RNA CCAL regulates colorectal cancer progression by activating Wnt/ β -catenin signalling pathway via suppression of activator protein 2 α . *Gut*. 2016; 65:1494–504. <https://doi.org/10.1136/gutjnl-2014-308392> PMID:[25994219](https://pubmed.ncbi.nlm.nih.gov/25994219/)
26. Rapetti-Mauss R, Bustos V, Thomas W, McBryan J, Harvey H, Lajczak N, Madden SF, Pellissier B, Borgese F, Soriani O, Harvey BJ. Bidirectional KCNQ1: β -catenin interaction drives colorectal cancer cell differentiation. *Proc Natl Acad Sci USA*. 2017; 114:4159–64. <https://doi.org/10.1073/pnas.1702913114> PMID:[28373572](https://pubmed.ncbi.nlm.nih.gov/28373572/)
27. Li J, He M, Xu W, Huang S. LINC01354 interacting with hnRNP-D contributes to the proliferation and metastasis in colorectal cancer through activating Wnt/ β -catenin signaling pathway. *J Exp Clin Cancer Res*. 2019; 38:161. <https://doi.org/10.1186/s13046-019-1150-y> PMID:[30987669](https://pubmed.ncbi.nlm.nih.gov/30987669/)
28. Kee JY, Han YH, Mun JG, Park SH, Jeon HD, Hong SH. Effect of Korean Red Ginseng extract on colorectal lung metastasis through inhibiting the epithelial-mesenchymal transition via transforming growth factor- β 1/Smad-signaling-mediated Snail/E-cadherin expression. *J Ginseng Res*. 2019; 43:68–76. <https://doi.org/10.1016/j.jgr.2017.08.007> PMID:[30662295](https://pubmed.ncbi.nlm.nih.gov/30662295/)
29. Dai G, Sun B, Gong T, Pan Z, Meng Q, Ju W. Ginsenoside Rb2 inhibits epithelial-mesenchymal transition of colorectal cancer cells by suppressing TGF- β /Smad signaling. *Phytomedicine*. 2019; 56:126–135. <https://doi.org/10.1016/j.phymed.2018.10.025> PMID:[30668333](https://pubmed.ncbi.nlm.nih.gov/30668333/)
30. Soleimani A, Khazaei M, Ferns GA, Ryzhikov M, Avan A, Hassanian SM. Role of TGF- β signaling regulatory microRNAs in the pathogenesis of colorectal cancer. *J Cell Physiol*. 2019; 234:14574–80. <https://doi.org/10.1002/jcp.28169> PMID:[30684274](https://pubmed.ncbi.nlm.nih.gov/30684274/)
31. Chalise P, Fridley BL. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One*. 2017; 12:e0176278. <https://doi.org/10.1371/journal.pone.0176278> PMID:[28459819](https://pubmed.ncbi.nlm.nih.gov/28459819/)
32. Ruan P, Wang Y, Shen R, Wang S. Using association signal annotations to boost similarity network fusion. *Bioinformatics*. 2019. [Epub ahead of print]. <https://doi.org/10.1093/bioinformatics/bt124> PMID:[30863842](https://pubmed.ncbi.nlm.nih.gov/30863842/)
33. Woo HG, Choi JH, Yoon S, Jee BA, Cho EJ, Lee JH, Yu SJ, Yoon JH, Yi NJ, Lee KW, Suh KS, Kim YJ. Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer. *Nat Commun*. 2017; 8:839.

- <https://doi.org/10.1038/s41467-017-00991-w>
PMID:29018224
34. Shen R, Mo Q, Schultz N, Seshan VE, Olshen AB, Huse J, Ladanyi M, Sander C. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One*. 2012; 7:e35236.
<https://doi.org/10.1371/journal.pone.0035236>
PMID:22539962
35. Yu J, Wu WK, Li X, He J, Li XX, Ng SS, Yu C, Gao Z, Yang J, Li M, Wang Q, Liang Q, Pan Y, et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut*. 2015; 64:636–45.
<https://doi.org/10.1136/gutjnl-2013-306620>
PMID:24951259
36. Despierre E, Moisse M, Yesilyurt B, Sehouli J, Braicu I, Mahner S, Castillo-Tong DC, Zeillinger R, Lambrechts S, Leunen K, Amant F, Moerman P, Lambrechts D, Vergote I. Somatic copy number alterations predict response to platinum therapy in epithelial ovarian cancer. *Gynecol Oncol*. 2014; 135:415–22.
<https://doi.org/10.1016/j.ygyno.2014.09.014>
PMID:25281495
37. Bonberg N, Pesch B, Behrens T, Johnen G, Taeger D, Gawrych K, Schwentner C, Wellhäußer H, Kluckert M, Leng G, Nasterlack M, Oberlinner C, Stenzl A, Brüning T. Chromosomal alterations in exfoliated urothelial cells from bladder cancer cases and healthy men: a prospective screening study. *BMC Cancer*. 2014; 14:854.
<https://doi.org/10.1186/1471-2407-14-854>
PMID:25412927
38. Xu H, Zhu X, Xu Z, Hu Y, Bo S, Xing T, Zhu K. Non-invasive Analysis of Genomic Copy Number Variation in Patients with Hepatocellular Carcinoma by Next Generation DNA Sequencing. *J Cancer*. 2015; 6:247–53.
<https://doi.org/10.7150/jca.10747>
PMID:25663942
39. Xu JF, Kang Q, Ma XY, Pan YM, Yang L, Jin P, Wang X, Li CG, Chen XC, Wu C, Jiao SZ, Sheng JQ. A Novel Method to Detect Early Colorectal Cancer Based on Chromosome Copy Number Variation in Plasma. *Cell Physiol Biochem*. 2018; 45:1444–1454.
<https://doi.org/10.1159/000487571> PMID:29466793
40. Kim HC, Chang J, Lee HS, Kwon HJ. Mitochondrial UQCRB as a new molecular prognostic biomarker of human colorectal cancer. *Exp Mol Med*. 2017; 49:e391.
<https://doi.org/10.1038/emm.2017.152>
PMID:29147009
41. Zhang Z, Zhou C, Chang Y, Zhang Z, Hu Y, Zhang F, Lu Y, Zheng L, Zhang W, Li X, Li X. Long non-coding RNA CASC11 interacts with hnRNP-K and activates the WNT/ β -catenin pathway to promote growth and metastasis in colorectal cancer. *Cancer Lett*. 2016; 376:62–73.
<https://doi.org/10.1016/j.canlet.2016.03.022>
PMID:27012187
42. Cheng N, Wu J, Yin M, Xu J, Wang Y, Chen X, Nie Z, Yin J. LncRNA CASC11 promotes cancer cell proliferation in hepatocellular carcinoma by inhibiting microRNA-188-5p. *Biosci Rep*. 2019; 39:BSR20190251.
<https://doi.org/10.1042/BSR20190251>
PMID:30910841
43. Luo H, Xu C, Le W, Ge B, Wang T. IncRNA CASC11 promotes cancer cell proliferation in bladder cancer through miRNA-150. *J Cell Biochem*. 2019; 120:13487–93.
<https://doi.org/10.1002/jcb.28622> PMID:30916832
44. Fu Y, Zhang P, Nan H, Lu Y, Zhao J, Yang M, Song Q. LncRNA CASC11 promotes TGF- β 1, increases cancer cell stemness and predicts postoperative survival in small cell lung cancer. *Gene*. 2019; 704:91–96.
<https://doi.org/10.1016/j.gene.2019.04.019>
PMID:30965130
45. Lin HY, Callan CY, Fang Z, Tung HY, Park JY. Interactions of *PVT1* and *CASC11* on Prostate Cancer Risk in African Americans. *Cancer Epidemiol Biomarkers Prev*. 2019; 28:1067–75.
<https://doi.org/10.1158/1055-9965.EPI-18-1092>
PMID:30914434
46. Huang D, Chen J, Yang L, Ouyang Q, Li J, Lao L, Zhao J, Liu J, Lu Y, Xing Y, Chen F, Su F, Yao H, et al. NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. *Nat Immunol*. 2018; 19:1112–25.
<https://doi.org/10.1038/s41590-018-0207-y>
PMID:30224822
47. Gao R, Fang C, Xu J, Tan H, Li P, Ma L. LncRNA CACS15 contributes to oxaliplatin resistance in colorectal cancer by positively regulating ABCC1 through sponging miR-145. *Arch Biochem Biophys*. 2019; 663:183–91.
<https://doi.org/10.1016/j.abb.2019.01.005>
PMID:30639170
48. Shan S, Li HF, Yang XY, Guo S, Guo Y, Chu L, Xu MJ, Xin DM. Higher lncRNA CASC15 expression predicts poor prognosis and associates with tumor growth in cervical cancer. *Eur Rev Med Pharmacol Sci*. 2019; 23:507–12.
https://doi.org/10.26355/eurrev_201901_16862
PMID:30720157
49. Yu L, Xu Q, Yu W, Duan J, Dai G. LncRNA cancer susceptibility candidate 15 accelerates the breast cancer cells progression via miR-153-3p/KLF5 positive feedback loop. *Biochem Biophys Res Commun*. 2018; 506:819–25.

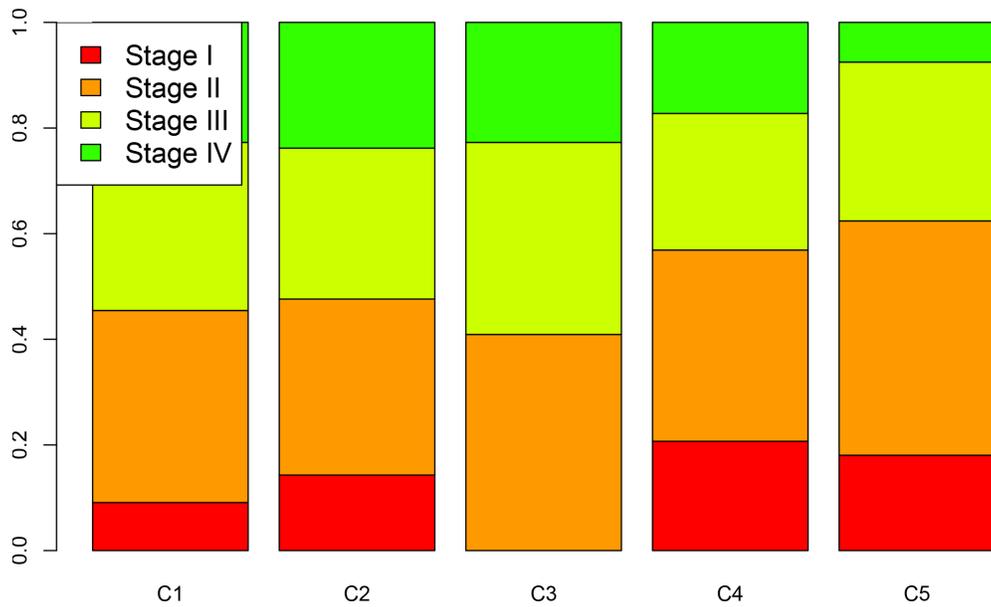
- <https://doi.org/10.1016/j.bbrc.2018.10.131>
PMID:[30389133](https://pubmed.ncbi.nlm.nih.gov/30389133/)
50. Jing N, Huang T, Guo H, Yang J, Li M, Chen Z, Zhang Y. LncRNA CASC15 promotes colon cancer cell proliferation and metastasis by regulating the miR-4310/LGR5/Wnt/ β -catenin signaling pathway. *Mol Med Rep.* 2018; 18:2269–76.
<https://doi.org/10.3892/mmr.2018.9191>
PMID:[29956772](https://pubmed.ncbi.nlm.nih.gov/29956772/)
51. He T, Zhang L, Kong Y, Huang Y, Zhang Y, Zhou D, Zhou X, Yan Y, Zhang L, Lu S, Zhou J, Wang W. Long non-coding RNA CASC15 is upregulated in hepatocellular carcinoma and facilitates hepatocarcinogenesis. *Int J Oncol.* 2017; 51:1722–30.
<https://doi.org/10.3892/ijo.2017.4175>
PMID:[29075788](https://pubmed.ncbi.nlm.nih.gov/29075788/)
52. Feng H, Zhang J, Shi Y, Wang L, Zhang C, Wu L. Long noncoding RNA LINC-PINT is inhibited in gastric cancer and predicts poor survival. *J Cell Biochem.* 2019; 120:9594–600.
<https://doi.org/10.1002/jcb.28236> PMID:[30569513](https://pubmed.ncbi.nlm.nih.gov/30569513/)
53. Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, Liu H, Xu J, Xiao F, Zhou H, Yang X, Huang N, Liu J, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun.* 2018; 9:4475.
<https://doi.org/10.1038/s41467-018-06862-2>
PMID:[30367041](https://pubmed.ncbi.nlm.nih.gov/30367041/)
54. Li L, Zhang GQ, Chen H, Zhao ZJ, Chen HZ, Liu H, Wang G, Jia YH, Pan SH, Kong R, Wang YW, Sun B. Plasma and tumor levels of Linc-pint are diagnostic and prognostic biomarkers for pancreatic cancer. *Oncotarget.* 2016; 7:71773–81.
<https://doi.org/10.18632/oncotarget.12365>
PMID:[27708234](https://pubmed.ncbi.nlm.nih.gov/27708234/)
55. Tiirats A, Viltrop T, Nõukas M, Reimann E, Salumets A, Kõks S. C14orf132 gene is possibly related to extremely low birth weight. *BMC Genet.* 2016; 17:132.
<https://doi.org/10.1186/s12863-016-0439-5>
PMID:[27660052](https://pubmed.ncbi.nlm.nih.gov/27660052/)
56. Chen WW, Yang Q, Li XY, Shi XL, Pu N, Lu GT, Tong ZH, Chen JM, Li WQ. Identification of a novel and heterozygous LMF1 nonsense mutation in an acute pancreatitis patient with severe hypertriglyceridemia, severe obesity and heavy smoking. *Lipids Health Dis.* 2019; 18:68.
<https://doi.org/10.1186/s12944-019-1012-9>
PMID:[30885219](https://pubmed.ncbi.nlm.nih.gov/30885219/)
57. Zhang C, Wu W, Zhu H, Yu X, Zhang Y, Ye X, Cheng H, Ma R, Cui H, Luo J, Guan J, Chang X. Knockdown of long non-coding RNA CCDC144NL-AS1 attenuates migration and invasion phenotypes in endometrial stromal cells from endometriosis. *Biol Reprod.* 2019; 100:939–949.
<https://doi.org/10.1093/biolre/iory252>
PMID:[30496345](https://pubmed.ncbi.nlm.nih.gov/30496345/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Prognosis description(Progression free survival) among five subtypes.



Supplementary Figure 2. The distribution of TNM stage among five subtypes.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1 and 2.

Supplementary Table 1. 120 lncRNAs contained in DE-lncRNAs.

Supplementary Table 2. Tan module was associated with 18 pathways.

Supplementary Table 3. Two pathways were enriched in yellow module.

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
hsa05206	MicroRNAs in cancer	11841	299/7466	0.001473	0.117837	0.117837	VEGFA/MIR25/ MIR200A/MIR135B/ MIR200B/MIR199A1	6
hsa04931	Insulin resistance	11749	107/7466	0.010498	0.419906	0.419906	MLXIPL/OGT/CPT1B	3

Please browse Full Text version to see the data of Supplementary Table 4.

Supplementary Table 4. Blue module enriched in 47 pathways.

Supplementary Table 5. Magenta module was associated with 19 pathways

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	geneID	Count
hsa04926	Relaxin signaling pathway	26420	130/7466	0.008169	0.38588	0.365784	RLN1/ADCY3/ PLCB4/RLN2/ GNG4	5
hsa03022	Basal transcription factors	26359	45/7466	0.009133	0.38588	0.365784	TAF4/GTF2F2/ GTF2IRD1	3
hsa00510	N-Glycan biosynthesis	26359	49/7466	0.011538	0.38588	0.365784	RPN2/ALG5/ MGAT5	3
hsa05110	Vibrio cholerae infection	26359	50/7466	0.012191	0.38588	0.365784	PLCG1/ADCY3/ ATP6V1C2	3
hsa04713	Circadian entrainment	26390	96/7466	0.013566	0.38588	0.365784	ADCY3/PLCB4/ CACNA1D/ GNG4	4
hsa04310	Wnt signaling pathway	26420	149/7466	0.014186	0.38588	0.365784	AXIN2/CSNK2A2/ TCF7/PLCB4/ NKD1	5
hsa04916	Melanogenesis	26390	101/7466	0.016087	0.38588	0.365784	ADCY3/TCF7/ PLCB4/ASIP	4
hsa04934	Cushing syndrome	26420	154/7466	0.016163	0.38588	0.365784	ADCY3/AXIN2/ TCF7/PLCB4/ CACNA1D	5
hsa04725	Cholinergic synapse	26390	112/7466	0.02262	0.390425	0.370091	ADCY3/PLCB4/ CACNA1D/GNG4	4
hsa04927	Cortisol	26359	64/7466	0.023557	0.390425	0.370091	ADCY3/PLCB4/	3

	synthesis and secretion						CACNA1D	
hsa04724	Glutamatergic synapse	26390	114/7466	0.023959	0.390425	0.370091	ADCY3/PLCB4/ CACNA1D/GNG4	4
hsa04137	Mitophagy - animal	26359	65/7466	0.024529	0.390425	0.370091	BCL2L1/ATG9B/ CSNK2A2	3
hsa00562	Inositol phosphate metabolism	26359	74/7466	0.034249	0.487047	0.461681	PLCG1/PLCB4/ INPP5D	3
hsa05016	Huntington disease	26420	193/7466	0.038044	0.487047	0.461681	TAF4/SDHA/ PLCB4/ATP5A1/ DNAH14	5
hsa04915	Estrogen signaling pathway	26390	137/7466	0.042819	0.487047	0.461681	ADCY3/PLCB4/ KRT23/KRT39	4
hsa04146	Peroxisome	26359	83/7466	0.045682	0.487047	0.461681	ECH1/PIPOX/ MPV17L	3
hsa03030	DNA replication	26330	36/7466	0.046803	0.487047	0.461681	RNASEH2B/RFC3	2
hsa04350	TGF-beta signaling pathway	26359	85/7466	0.04845	0.487047	0.461681	TFDP1/TGIF2/ NODAL	3
hsa04911	Insulin secretion	26359	85/7466	0.04845	0.487047	0.461681	ADCY3/PLCB4/ CACNA1D	3

Please browse Full Text version to see the data of Supplementary Tables 6–8:

Supplementary Table 6. 17 lncRNAs were identified with more than 7% of the frequency of CNV in 452 CRC samples.

Supplementary Table 7. 104 lncRNAs were identified for prognosis prediction.

Supplementary Table 8. 10 lncRNAs were identified with significant prognosis based on the quartiles method.