Identification and meta-analysis of copy number variation-driven circadian clock genes for colorectal cancer

WENLIANG YUAN[1-5], LI LIU[5], CAI WEI[2-4], XIAOBO LI[6,7], DAN SUN[2-4], CHAOXU DAI[2-4], SICONG LI[2-4], SIHUA PENG[2-4] and LINHUA JIANG[1]

[1]School of Optical-Electric and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093; [2]Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources (Shanghai Ocean University), Ministry of Education; [3]National Pathogen Collection Center for Aquatic Animals, Ministry of Agriculture; [4]International Research Center for Marine Biosciences at Shanghai Ocean University, Ministry of Science and Technology, Shanghai 201306; [5]College of Mathematics and Computer Science, Chizhou University, Chizhou, Anhui 247000; [6]Institute of Biomedical Informatics and [7]College of Engineering, Lishui University, Lishui, Zhejiang 323000, P.R. China

Received March 4, 2019; Accepted July 26, 2019

DOI: 10.3892/ol.2019.10830

Abstract. Both copy number variation (CNV) and circadian clock genes play a critical role in the etiology and pathogenesis of colorectal cancer (CRC); however, a comprehensive analysis of CNV-driven circadian clock genes is urgently required. The present study aimed to investigate the systematic associations between somatic cell CNVs and circadian clock gene expression in patients with CRC. Using somatic CNV, legacy clinical information and gene expression data from The Cancer Genome Atlas, 295 genes that were significantly differentially expressed and with significantly different CNV were obtained, and the expression of the genes, among which 15 were circadian clock genes, was significantly associated with CNV. Further analysis revealed that aryl hydrocarbon receptor nuclear translocator-like 2 (ARNTL2) expression and CNV in these circadian clock genes were significantly associated with survival time in patients with CRC, and the expression of ARNTL2 was also significantly associated with the pathological stage of CRC. Gene set enrichment analysis found that ARNTL2 is enriched for gene sets associated with CRC pathogenesis such as the p53 signaling pathway. These results suggest that ARNTL2 may be a promising prognostic biomarker for patients with CRC, and that circadian clock genes play an important role in CRC through CNV.

Correspondence to: Dr Sihua Peng, Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources (Shanghai Ocean University), Ministry of Education, 999 Hucheng Ring Road, Shanghai 201306, P.R. China
E-mail: shpeng@shou.edu.cn

Professor Linhua Jiang, School of Optical-Electric and Computer Engineering, University of Shanghai for Science and Technology, 516 Jungong Road, Shanghai 200093, P.R. China
E-mail: lhjiang@usst.edu.cn

Key words: colorectal cancer, copy number variation, circadian clock genes, ARNTL2, Gene Set Enrichment Analysis

Introduction

Colorectal cancer (CRC) has high morbidity and mortality rates worldwide, at 10.2 and 9.2%, respectively (1). Despite the therapeutic advances and earlier detection, the 5-year survival rate of patients with CRC remains unsatisfactory (2). One of the main reasons for this is that the occurrence of CRC is a complex multi-stage process, and involves further investigation into the proliferation, differentiation, apoptosis and survival mechanism of intestinal epithelial cells (3). Therefore, biomarkers for early detection and targeted therapy are urgently required.

Biological rhythms are produced by conserved transcription and translation feedback loops of circadian clock genes within the cells (4). A circadian disruption has been recognized as a potential independent risk factor for cancer development (5). Circadian clock genes appear to have multifaceted functions during cancer development and can act to both suppress tumors and promote carcinogenesis (6). Research by the International Agency for Research on Cancer has also demonstrated that this disruption increases the risk of CRC (7). Several previous studies have also demonstrated that large variations in expression levels, both up- and down-regulated, and the circadian clock genes are associated with tumor progression and mammalian tumorigenesis for several malignancies, such as breast cancer (8), liver cancer (9) and colorectal carcinoma (10). In addition, the association between single nucleotide polymorphisms (SNPs) in circadian clock genes and disease has also been analyzed (11,12). These studies indicate that mutations or deregulated expression of circadian clock genes are frequently detected in different tumors. Copy number variation (CNV) is a kind of structural variation at the submicroscopic level, which refers to the complex chromosomal structural variation forms derived from the deletion and/or duplication of DNA fragments longer than 1 kb (13).

Increasing research has shown that CNV is closely associated with the risk of tumor occurrence (14,15). The mechanism of action of CNV and circadian clock genes in many cancer types has been intensively investigated, such as liver cancer (16) and lung cancer (14); however, the study of the mechanism of action of CNV-driven circadian clock genes in cancer (including CRC) has not yet been reported.

The aryl hydrocarbon receptor nuclear translocator-like 2 (ARNTL2) gene is also known as brain and muscle ARNT-like 2 (BMAL2), which is mapped to human chromosome 12p11.22-11.23 and shares 52% amino acid identity with zebrafish Bmal2 and 49% identity with human BMAL1 (17). Schoenhard *et al* (18) hypothesized that the different ARNTL2 spatiotemporal distributions allow intrinsic circadian clocks to modulate the amplitudes of their oscillators while maintaining circadian periodicity. Research on ARNTL2 in various complex diseases (19,20), particularly cancer, has gradually become accepted. Studies have shown that ARNTL2 is a potential biomarker for tumor invasion in colorectal cancer (21), and it is significantly associated with lung cancer risk (22). In addition, a previous study has analyzed SNPs associated with ARNTL2 expression in patients with breast cancer (23).

The aim of the present study, using somatic CNV, legacy clinical information and gene expression data from the Cancer Genome Atlas (TCGA; https://tcga-data.nci.nih.gov/tcga/), was to investigate the systematic association between somatic cell CNV and circadian clock gene expression in patients with CRC, and to identify ARNTL2 as a contributing gene in CRC development that may serve as a promising therapeutic strategy.

Materials and methods

Data source and preprocessing. The CNV data were downloaded from TCGA data portal on October 23, 2018. The data contained 979 files and 460 cases for CNV analysis by setting specific parameters: Data Type was Masked Copy Number Segment. In addition, mRNA expression profile data and the corresponding legacy clinical information of patients with CRC from TCGA were also downloaded and contained 480 CRC tumor specimens and 41 tumor-adjacent tissue specimens. Firstly, 18 samples without adequate clinical information were removed, which left 462 patients with CRC with complete survival information. Subsequently, low-abundance mRNA expression data were removed; mRNAs with expression value >1 in 90% samples were retained. For the duplication data in one sample, the average values of the mRNA expression were adopted. The 2,083 differentially expressed mRNAs were analyzed using R/Bioconductor package edgeR (version 3.26) (24), with the criteria of $|\log_2$fold-change (FC)$|$ >1.5 and q-value <0.01. No patients were involved in clinical trials in this study.

Identification and functional analysis of CNV-driven circadian clock genes. Gene Ontology (GO) analysis was performed to explore the functional roles of the target genes using DAVID (http://www.david.abcc.ncifcrf.gov/) (25). Finally, the enriched GO terms with gene count >5 and P<0.05 were selected for further analyses. Cytoscape software (version 3.7.1) (26) (with ClueGO and CluePedia plugins) was used for the Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, showing only pathways with P<0.05.

Circadian clocks exist endogenously in almost every organism (6). The Circadian Gene Database (CGDB; version 1.0; http://cgdb.biocuckoo.org/index.php) (27) was used to identify the circadian clock genes. Circadian genes were selected that had been identified experimentally. A literature search using PubMed database was performed to identify the latest candidate circadian clock genes.

To verify the expression profile of ARNTL2 in CRC tissues and their non-tumoral counterparts, a meta-analysis was performed using the Oncomine database (version 4.5; www.oncomine.org) by setting specific parameters: 'ARNTL2', 'Cancer vs. Normal Analysis', 'Colorectal Cancer' and 'mRNA'.

The java software Gene Set Enrichment Analysis (GSEA; version 3.0) was employed to perform the statistical significance test between two phenotypes (http://software.broadinstitute.org/gsea/index.jsp), with gene expression data and phenotype data (high/low group of expression values of ARNTL2) to be prepared according to the GSEA guidelines (28). The parameters were set as follows: Using KEGG pathway as a reference, permutation type to be the phenotype, and at least 15 genes in a single pathway. The mean expression levels (905.75) of ARNTL2 in all cancer samples were obtained. In the GSEA analysis, the expression level higher than this value is considered to be high expression, and below this value is considered to be low expression.

Statistical analysis. The segment mean at (-0.2, 0.2) was generated by the error of the instrument measurement, so the copy number of such genes was confirmed as unchanged. A χ^2 test was used to compare the number of CNVs in cancer tissue and paracancer tissue, with a criterion of false discovery rate (FDR) <0.01. A Kolmogorov-Smirnov test was used to identify genes with CNV and expression consistency, with the criterion of P<0.005. A Kolmogorov-Smirnov test is based on cumulative distribution functions to test whether a distribution conforms to a theoretical distribution or whether there is a significant difference between two empirical distributions. To assess the result set of genes, a hypergeometric test was used to verify whether known CRC-related genes were enriched on the set. To identify the associations between clinicopathological parameters and the presence of copy number loss or gain in the regions containing selected genes, a Pearson's χ^2 test was performed. A Kaplan-Meier curve analysis was performed to analyze the association between the gene and survival time, and statistical significance was assessed using the R package 'survival' (29). P<0.05 (two-sided) was considered to indicate a statistically significant difference.

Results

Patient characteristics. The detailed clinical and pathological characteristics of the study population, including age, sex, pathological stage, pathological tumor (pathological T), pathological node (pathological N) and pathological metastasis (pathologic M), were summarized in Table I. All the 462 patients were pathologically diagnosed with colorectal cancer. The median age for all patients was 60 years (interquartile range, 31-90 years).

Table I. Clinicopathological features of the 462 patients with colorectal cancer.

Feature	Primary, n (%)	Metastatic, n (%)	NA, n (%)
Age, years			
<60	81 (24.0)	42 (36.5)	4 (40.0)
>60	256 (76.0)	73 (63.5)	6 (60.0)
Sex			
Male	177 (52.5)	53 (46.1)	6 (60.0)
Female	160 (47.5)	62 (53.9)	4 (40.0)
Pathological T			
T1-T2	77 (22.8)	10 (8.7)	3 (30.0)
T3-T4	260 (77.2)	105 (91.3)	7 (70.0)
Pathological n stage			
N0	231 (68.5)	35 (30.4)	5 (50.0)
N1-N2	106 (31.5)	80 (69.6)	5 (50.0)
Pathological stage			
I-II	228 (67.7)	23 (20.0)	2 (20.0)
III-IV	106 (31.5)	87 (75.7)	5 (50.0)
NA	3 (0.9)	5 (4.3)	3 (30.0)
Vital status			
Alive	288 (85.5)	73 (63.5)	7 (70.0)
Death	49 (14.5)	42 (36.5)	3 (30.0)

T, tumor; N, node.

Screening of potential CRC-related gene CNVs. To identify potential candidate genes within the regions exhibiting CNVs in the TCGA dataset, the frequency of copy number loss and gain in the regions was obtained. First, the instrument measurement error was filtered, and the area where the CNV number was significantly different located, and finally the genes in these areas were identified. Finally, the χ^2 test was conducted on CNV, and a total of 10,256 genes with significant differences in CNV expression were obtained. KEGG and GO enrichment analyses was then performed with a smaller set of genes (n=295). A detailed workflow chart of the methodology is illustrated in Fig. 1A. CNV occurred differently on each chromosome in patients with CRC. Large-scale losses of copy numbers occurred only on certain chromosomal regions, such as chromosomes 4, 11, 14, 15, 18, 21 and 22. However, on other chromosomes, such as chromosomes 7, 12 and 13, only gains occurred (Fig. 1B).

Screening of differentially expressed mRNAs. Based on the threshold criteria of $|\log_2 FC| > 1.5$ and q-value <0.01, 2,083 mRNAs were identified as aberrantly expressed mRNAs in the CRC tissues compared with that in the adjacent non-tumorous tissues. It was found that a number of mRNAs were upregulated or downregulated >100-fold (Fig. 2A). To further investigate the mechanism of CNV in the development and progression of CRC, the intersection of genes involved in significant abnormal CNV and differentially expressed genes was obtained. Subsequently the association analysis of the expression profiles and copy number profiles for the aforementioned small gene set was performed, with a result that 295 mRNAs had statistically significant differential expression and a difference in CNV. Finally, GO enrichment analysis and a KEGG pathway analysis of these mRNAs were performed, suggesting that the mRNAs were primarily enriched in only one KEGG pathway (P<0.05; Fig. 2B) and eight GO terms (Benjamin P<0.01; Fig. 2C). Recent evidence suggests that the circadian system can influence the Wnt/β-catenin signaling pathway (30), which is a critical pathway for the development and progression of CRC (31). Known CRC-related genes were mapped to the set of 295 mRNAs, and the 73 CRC-related genes were significantly enriched in this gene set (hypergeometric test, $P=4.725561 \times 10^{-9}$).

PubMed and CGDB databases were searched, and 15 of the 295 mRNAs were found to be circadian clock genes (Table II). Among the 15 circadian clock genes, NR3C2 and P2RX1 were downregulated, and the remaining 13 genes were upregulated in patients with CRC. Gain was the predominant type of alteration for BIRC7, GNGT1, NFE2L3, PDX1 and UBE2C, while loss of APCDD1 and P2RX1 was found in >30% of cases. No significant changes in the expression levels of other important genes, such as PER and ARNTL1, in the circadian clock signaling pathway, were found.

Subsequently, a meta-analysis on the expression of the 15 clock genes in CRC using public microarray datasets from the Oncomine database was performed. As presented in Fig. 3, the expression patterns of the clock gene ARNTL2 in 10 independent microarray datasets and TCGA datasets were consistent with previous analyses (32,33). Overexpression was found in all CRC tissues compared with that in the tumor-adjacent tissue (gene median rank, 86.0; $P=9.39 \times 10^{-7}$).

Function analysis of the clock gene ARNTL2 driven by CNV in CRC. The expression of the gene ARNTL2 was found in the 452 patients with CRC, among which a total of 48 CNVs occurred, with the presence of copy number gain in 44 patients and copy number loss in 4 patients. ARNTL2 was null in 10 samples, which were consequently removed from the study. The association of ARNTL2 mRNA expression levels with CNV type was identified. As shown in Fig. 4A, single gain and amplification of ARNTL2 were associated with increased mRNA expression, and deletion of ARNTL2 was associated with decreased mRNA expression. Therefore, ARNTL2 gene expression and CNV in CRC tissues show the same trend.

A Kaplan-Meier curve analysis was performed to investigate the overall survival time for ARNTL2 in patients with CRC. Compared with that of the patients with normal copy number, the survival rate of the patients with abnormal copy number (gain or loss) of ARNTL2 was significantly decreased (Fig. 4B), whereas the overall survival of patients with CRC with ARNTL2 CNV was significantly decreased. The expression levels of ARNTL2 were also associated with the overall patient survival; higher expression levels indicated greater survival time (Fig. 4C).

To further investigate whether ARNTL2 is involved in the development and progression of CRC, the tumor tissue samples were divided into several subgroups based on pathological TNM (T3+T4 vs. T1+T2, N2+N3 vs. N0+N1, M1 vs. M0) and pathological stages (I-II vs. III-IV) (34). A comparative

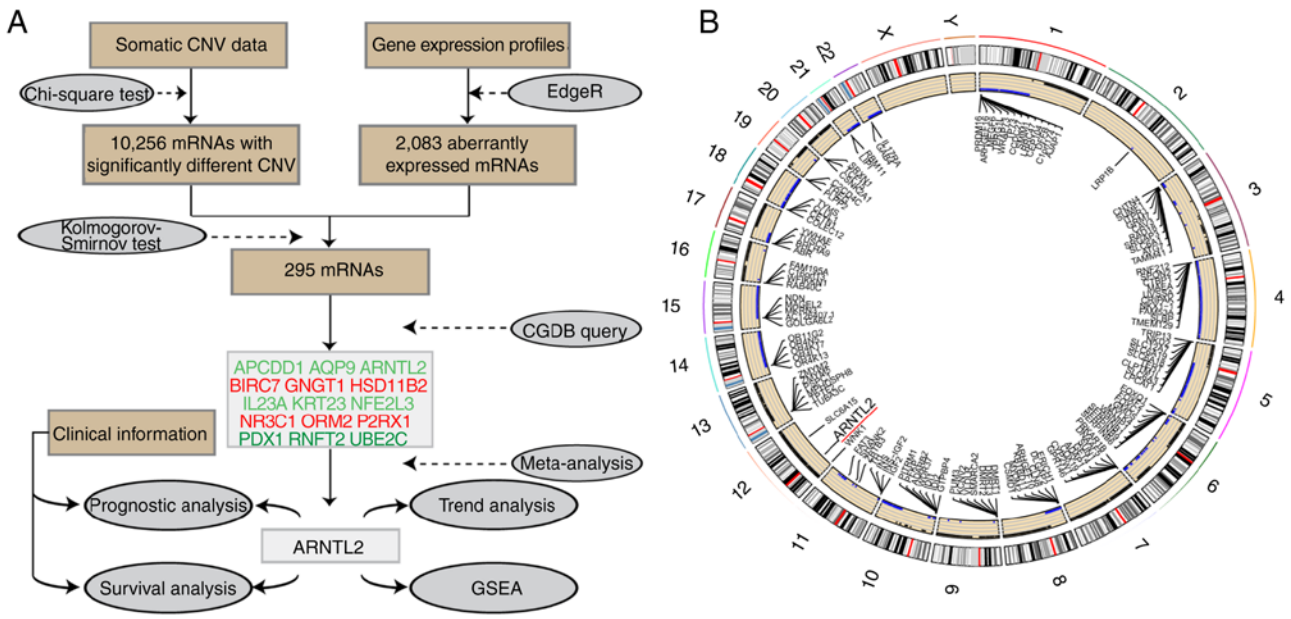

Figure 1. Analyzed workflow and CNV identification results. (A) Workflow for identification and analysis of CNV-driven circadian clock genes. (B) Circos plot of the human genome showing chromosome structure and CNV in CRC. The outer most layer is the chromosome model, and the inner layer illustrates the CNV. A gene with CNV is annotated in the inner layer, in which a blue dot near the center of the circle denotes the presence of copy number loss, and a black dot near the outer layer denotes the presence of copy number gain. CNV, copy number variation; CRC, colorectal cancer; CGBD, Circadian Genome Database; GSEA, Gene Set Enrichment Analysis; ARNTL2, aryl hydrocarbon receptor nuclear translocator-like 2.

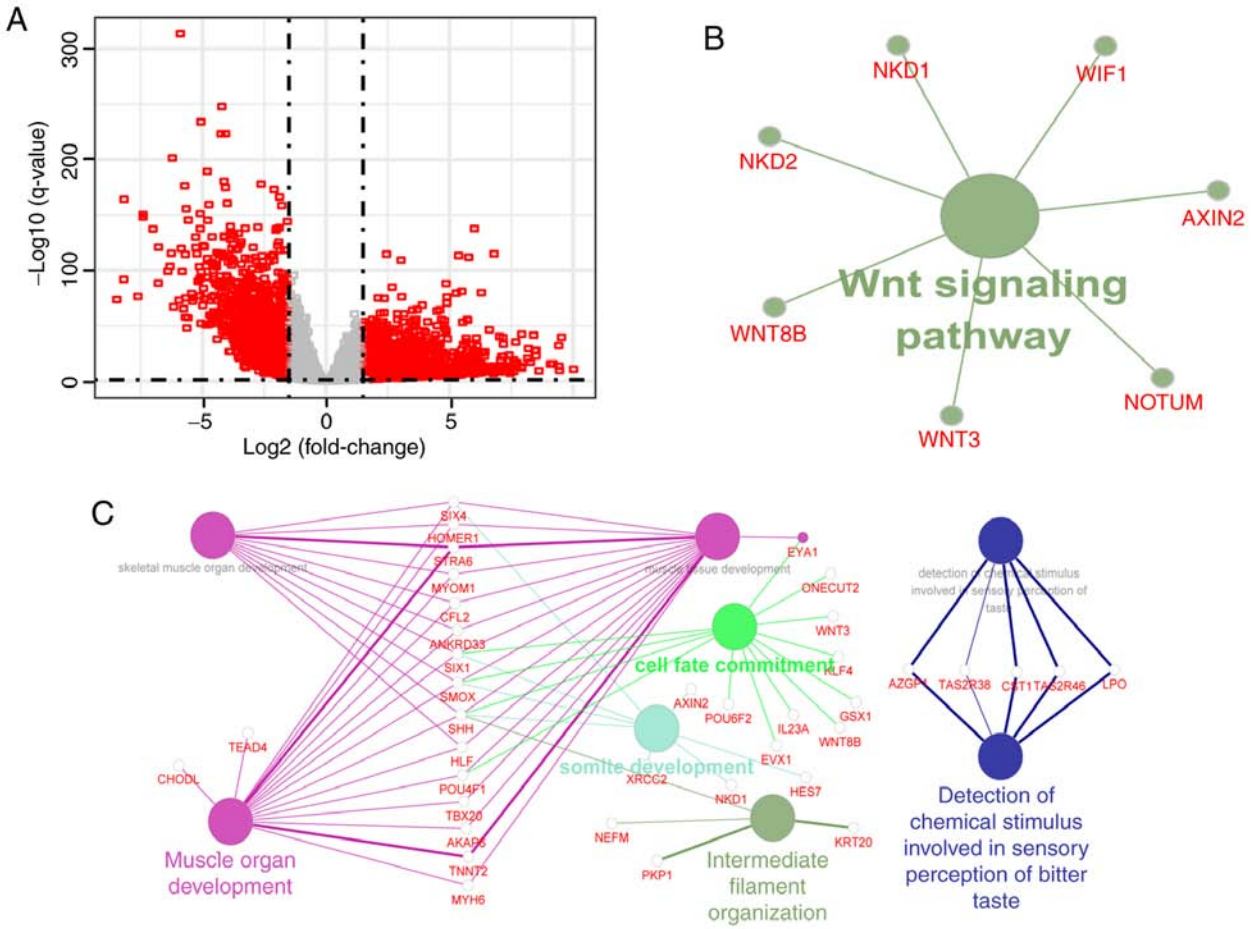

Figure 2. Functional analysis of copy number variation-driven genes. (A) Volcano plots show the expression profiles of mRNAs. The vertical line represents 2.0-fold up- and downregulation between the colorectal cancer tissues and the adjacent non-tumorous tissues, while the horizontal line represents the q-value. The red dots in the figure represent the differentially expressed mRNAs, with the left side indicating downregulation of mRNA expression and the right side indicating upregulation of mRNA expression. (B) The Kyoto Encyclopedia of Genes and Genomes pathway analysis of the 295 mRNAs. (C) The Gene Ontology enrichment analysis of the 295 mRNAs.

Table II. Information of the 15 circadian clock genes.

Gene	Location	Log_2FC	FDR	Loss	Gain	Normal	FDR	P-value[a]
APCDD1	Chr18: 10,454,628-10,489,948	2.483785	$2.96x10^{-12}$	154	9	289	$1.14x10^{-87}$	$1.93x10^{-5}$
AQP9	Chr15: 58,138,169-58,185,911	2.470799	$6.41x10^{-10}$	59	2	391	$6.65x10^{-12}$	$2.16x10^{-3}$
ARNTL2	Chr12: 27,332,854-27,425,289	2.484382	$1.58x10^{-41}$	4	44	404	$2.46x10^{-8}$	$1.28x10^{-7}$
BIRC7	Chr20: 63,235,883-63,240,507	2.466525	$1.09x10^{-11}$	0	272	180	$3.08x10^{-87}$	$1.37x10^{-6}$
GNGT1	Chr7: 93,591,573-93,911,265	3.164943	$4.55x10^{-10}$	1	109	342	$2.94x10^{-25}$	$5.17x10^{-6}$
HSD11B2	Chr16: 67,430,652-67,437,553	-2.35877	$5.09x10^{-65}$	2	27	423	$1.08x10^{-4}$	$1.44x10^{-3}$
IL23A	Chr12: 56,334,174-56,340,410	3.021143	$4.60x10^{-23}$	0	35	417	$6.42x10^{-5}$	$4.20x10^{-3}$
KRT23	Chr17: 40,922,696-40,937,634	7.179667	$2.02x10^{-34}$	9	38	405	$8.50x10^{-9}$	$7.79x10^{-7}$
NFE2L3	Chr7: 26,152,240-26,187,125	2.676112	$1.01x10^{-85}$	0	161	291	$1.53x10^{-41}$	$3.89x10^{-17}$
NR3C2	Chr4: 148,078,762-148,444,698	-2.63761	$4.00x10^{-84}$	29	2	421	$1.74x10^{-3}$	$4.86x10^{-5}$
ORM2	Chr9: 114,329,869-114,333,252	2.989211	$1.39x10^{-11}$	9	19	424	$1.42x10^{-3}$	$3.11x10^{-3}$
P2RX1	Chr17: 3,896,592-3,916,500	-2.29178	$2.75x10^{-55}$	138	3	311	$1.42x10^{-36}$	$5.26x10^{-7}$
PDX1	Chr13: 27,920,020-27,926,231	4.797965	$1.54x10^{-55}$	0	200	252	$1.37x10^{-58}$	$8.04x10^{-18}$
RNFT2	Chr12: 116,738,178-116,853,631	2.006432	$4.31x10^{-30}$	3	31	418	$1.08x10^{-4}$	$4.09x10^{-3}$
UBE2C	Chr20: 45,812,576-45,816,957	2.15803	$5.49x10^{-43}$	1	276	175	$5.72x10^{-89}$	$6.51x10^{-36}$

[a]Kolmogorov-Smirnov test. FC, fold-change; FDR, false discovery rate.

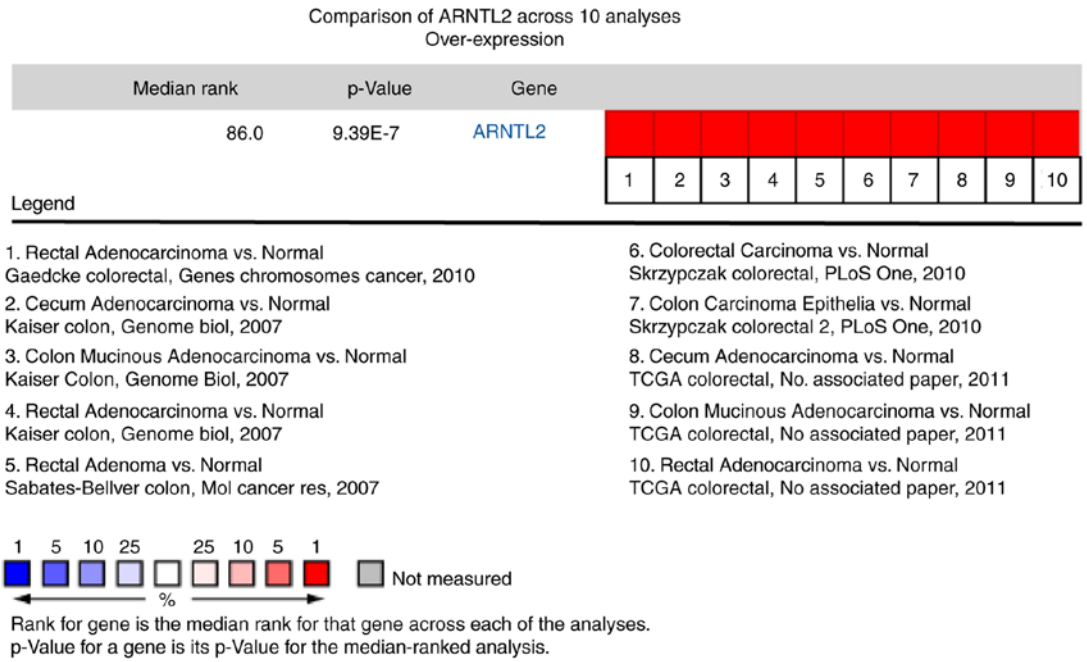

Figure 3. Oncomine analysis of ARNTL2 mRNA expression levels in the 10 independent microarray datasets and The Cancer Genome Atlas datasets. Data are shown as the median rank of ARNTL2 through each dataset analysis. P-value for ARNTL2 is obtained using the median ranked analysis. ARNTL2, aryl hydrocarbon receptor nuclear translocator-like 2.

analysis of ARNTL2 expression profiles was performed. As a result, ARNTL2 expression demonstrated a statistically significant association with pathological stages (P<0.001) and pathological N (P<0.001) (Fig. 4D).

To gain a clearer understanding of the expression of ARNTL2 in patients with cancer and adjacent tissues, a paired difference analysis of ARNTL2 from 41 patients with cancer and adjacent tissues was performed. The expression of ARNTL2 in cancer tissues was significantly higher compared with that in adjacent tissues ($P=1.058x10^{8}$; Fig. 4E). This is consistent with the results of our previous analysis of the difference.

To investigate the biological characteristics shared by the different ARNTL2 expression levels, a GSEA was performed. The most significant pathways for the upregulated gene sets in the significance order (nominal P<0.05) are shown in Fig. 5. The six pathways, including 'natural killer cell-mediated cytotoxicity', 'oocyte meiosis', the 'p53 signaling pathway', 'pancreatic cancer', 'prostate

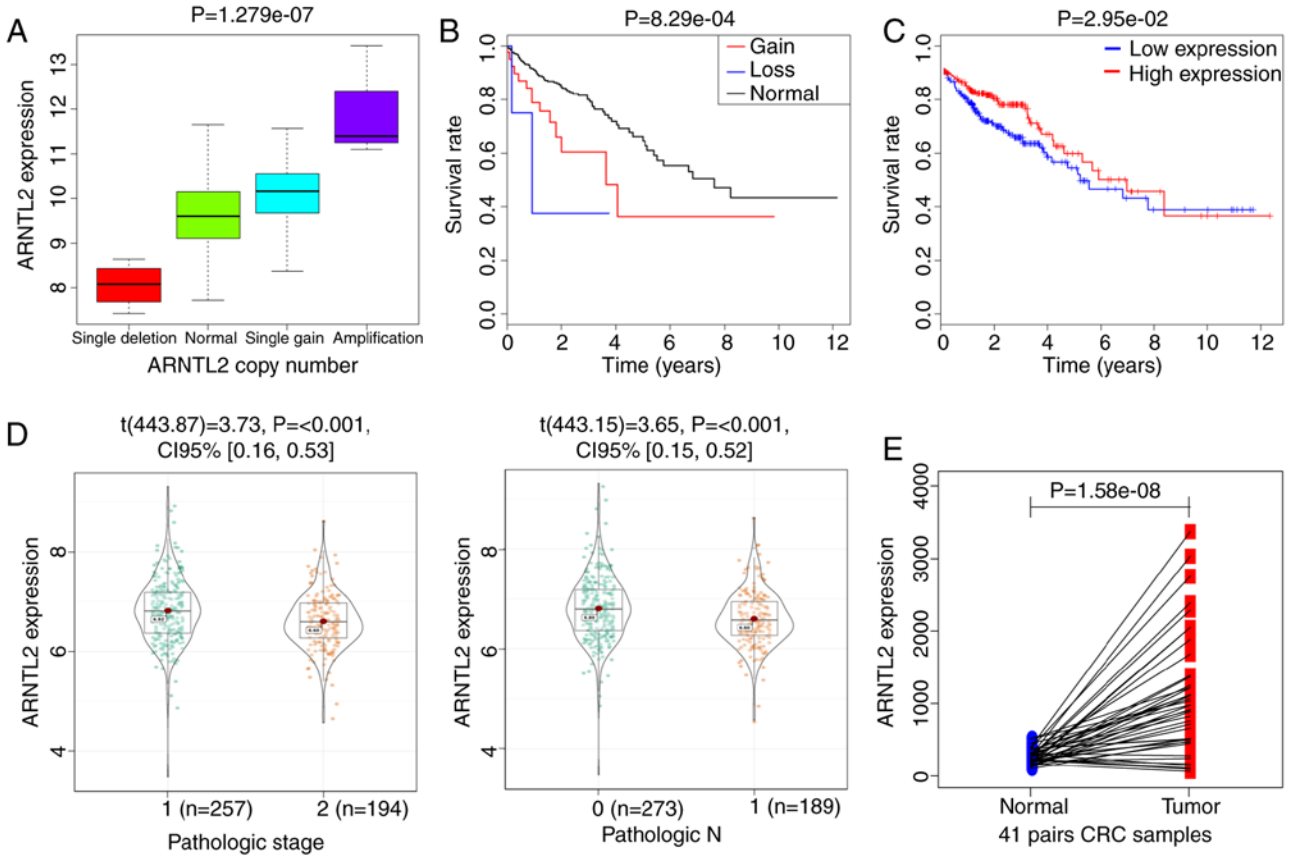

Figure 4. Functional analysis of copy number variation-driven clock gene ARNTL2 in CRC. (A) Box plot of ARNTL2 mRNA expression levels associated with the corresponding gene status. (B) Kaplan-Meier survival analysis of ARNTL2 gene status. (C) Kaplan-Meier survival analysis of ARNTL2 expression levels. (D) ARNTL2 expression associated with the development and progression of CRC. Y-axis represents the expression value following ARNTL2 correction. (E) Paired difference analysis of 41 pairs of paracancerous and cancer samples of ARNTL2. Y-axis indicates the uncorrected expression value of ARNTL2. CRC, colorectal cancer; ARNTL2, aryl hydrocarbon receptor nuclear translocator-like 2; CI, confidence interval; N, node.

cancer' and the 'toll like receptor signaling pathway' were significant in the ARNTL2 high expression phenotype. Among these pathways, some were directly linked to cancer pathogenesis, such as 'pancreatic cancer', the 'p53 signaling pathway' (35) and 'prostate cancer'. There were no significant pathways for downregulated gene sets with nominal P<0.05.

Discussion

CRC is the third most commonly occurring cancer worldwide and the fourth most frequent cause of death having an oncological origin (1); it is considered to be a complex disease resulting from a combination of environmental factors, genetic/epigenetic predisposing variants and specific molecular mechanisms. Chromosomal instability (CIN) has been defined as a major factor contributing to CRC carcinogenesis (36). CNV exists as a genetic polymorphism in the human genome that is a type of CIN (37). The form of CNV directly affecting the expression of a gene is mainly the deletion or amplification of a copy number of a gene, causing an increase or decrease in the amount of gene expression and increasing the occurrence of the disease (38). A previous study found that tumor necrosis factor receptor superfamily member 10C CNV is associated with metastatic colorectal cancer (39). In the present study, an integrated analysis of CNV data and gene expression profile for CRC with a large sample size (n=503, including 462 patient samples and 41 tumor-adjacent tissue samples) was performed. A total of 10,256 genes with significantly different CNV and 2,083 aberrantly expressed mRNAs were obtained, of which 295 genes showed a statistically significant association between the gene expression and CNV; therefore, these 295 genes were regarded as CRC-related CNV-driven genes. The present findings may provide a new theoretical basis for the pathogenesis of CRC and also contribute to the development of new therapeutic strategies.

In the present study, CNV-driven genes were only enriched in the Wnt signaling pathway. The Wnt pathway is involved in the regulation of important physiological processes such as normal embryo development, and cell proliferation and differentiation, and its abnormal activation plays an important role in the process of tumor development, metastasis and therapeutic resistance (40). A previous study showed that >90% of colorectal cancer cases have abnormal activation of the Wnt classical signaling pathway (41). Meanwhile, some studies suggested that the regulation of circadian clock genes, such as CRY1 (42) and Rev-erbα (43), was mediated by the classical Wnt/β-catenin signaling pathway. Further study of the Wnt signaling pathway will help to develop new strategies for CRC treatment. The present findings provide a new clue to study this signaling pathway.

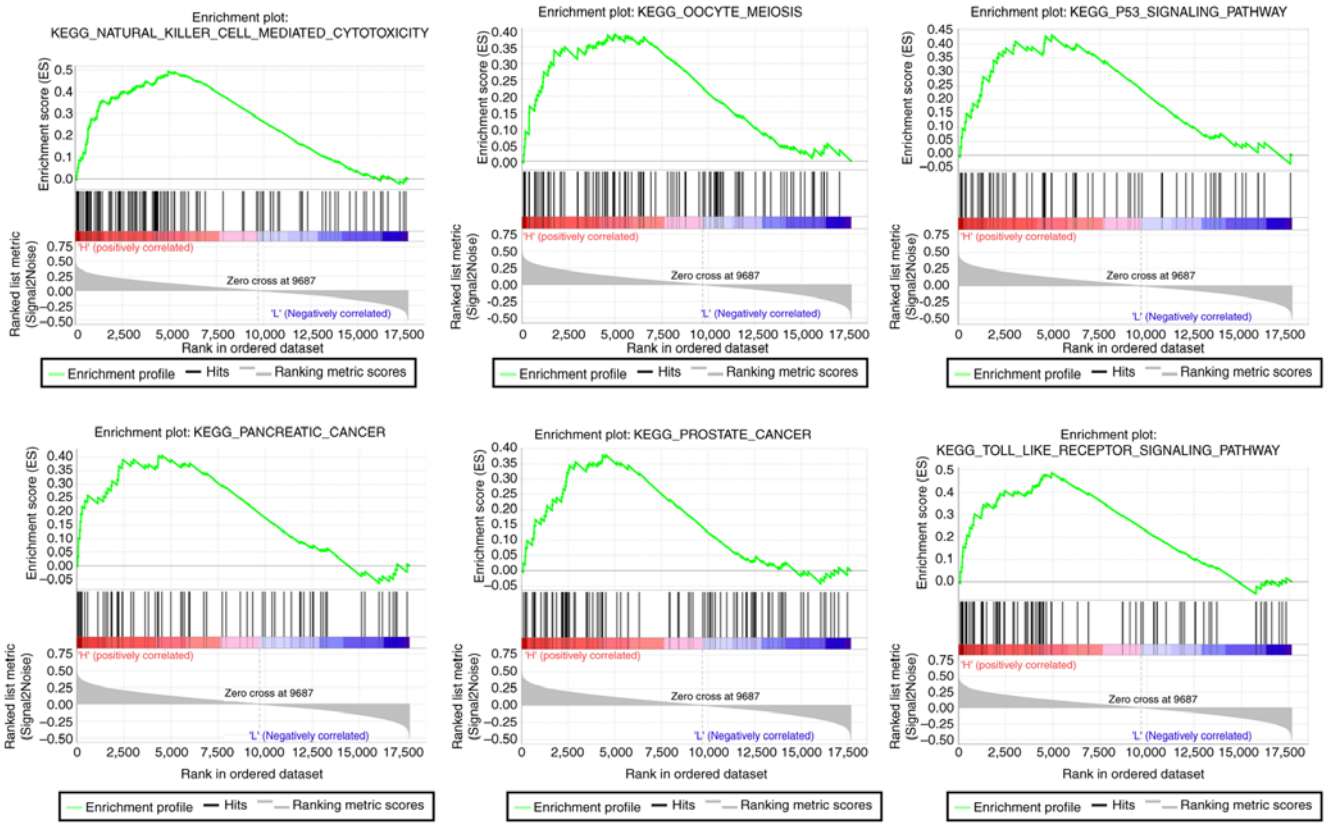

Figure 5. Gene Set Enrichment Analysis results showing the association of aryl hydrocarbon receptor nuclear translocator-like 2 expression levels and colorectal cancer-related gene sets.

Studies in circadian clock genes may expand the knowledge regarding the mechanism of occurrence and development of tumors, and may provide a new approach for tumor therapy (44). Indeed, multiple epidemiological studies have shown that impaired function of the circadian clock promotes development of cancer (45). For circadian clock genes, including Per1, Per2, and Per3, the expression levels of which are often found to be decreased in pancreatic cancer (46) and gastric cancer (47), as well as the disruption of autonomic rhythm. Additionally, in a previous study, CRC showed lower expression of NPAS2 compared with that in healthy tissues, and this was negatively associated with tumor size, stage and metastasis (48). A previous study has also shown that varying degrees of biorhythm destruction are found in 50% of metastatic cancer cases (49). In the present study, 15 CNV-driven circadian clock genes in CRC tissues were identified, indicating that these circadian clock genes may play a role in cancer. However, this requires further validation at the protein level. ARNTL2 has been described as a candidate biomarker in different cancer types, including kidney cancer (50), colorectal cancer (21) and hepatocellular carcinoma (51), and similar results for ARNTL2 were obtained in the present study. As research continues to deepen, numerous studies have found that genomic alterations involving circadian clock genes, such as point mutations or CNV, are frequently found in different human cancer types. The rs1801260 SNP, in the 3' untranslated region of the clock circadian regulator gene, was found to be associated with the development of CRC (52). The CNV form of the BMAL1 gene has also been found in multiple cancers, such as breast and colorectal cancer (35 gains and 7 losses in CNV numbers) (53). Previous studies have observed a close association between ARNTL2 expression and various types of cancer (21,22); however, no studies have characterized the association between CNV in ARNTL2 and cancer. In the present study, upregulation of ARNTL2 in patients with CRC was found, and ARNTL2 CNV has three forms: Single loss, single gain, and amplification. Further analysis found that the expression of ARNTL2 has the same trend as CNV. Our study showed that the expression level of ARNTL2 was abnormal due to the presence of CNV, which promoted the occurrence and development of CRC.

Genetic drifts in ARNTL2 polymorphisms have been described in the human population leading to variation in the circadian rhythm regulation (54). The expression of ARNTL2 was significantly associated with survival time in patients with CRC. A previous study found that high ARNTL2 expression predicted poor survival in patients with lung adenocarcinoma (55). However, in the present study, low ARNTL2 expression predicted poor survival in patients with CRC. This may be due to the heterogeneity between different types of cancer. Moreover, a previous report indicated that ARNTL2 high levels significantly influence mammary tumor metastasis (23). ARNTL2 levels were also significantly associated with pathological stage and N stage in patients with CRC in the present study. ARNTL2 CNV was also significantly associated with survival time in patients with CRC. These data suggest that ARNTL2 can be used as a prognostic factor

for CRC, which may bring more personalized treatment to patients with CRC. GSEA analysis showed that ARNTL2 is enriched for gene sets associated with CRC pathogenesis, such as the 'p53 signaling pathway'. These findings suggest that the CNV-driven clock gene ARNTL2 plays a crucial role in the development and progression of CRC. However, this study has some limitations as it was an *in silico* study. Further *in vivo* investigations would be beneficial to fully understand the roles of ARNTL1 in CRC initiation and development.

In summary, to the best of our knowledge, the present study demonstrates for the first time that circadian clock genes play an important role in CRC in the form of CNV, and that 15 CNV-driven clock genes are associated with the etiology and pathogenesis of CRC. Finally, it was concluded that CNV in the circadian clock gene ARNTL2 may be a useful genetic biomarker for the treatment of individualized CRC patients and may identify patients who may benefit from more aggressive systemic treatment strategies.

Acknowledgements

Not applicable.

Funding

This study was partly supported by the National Natural Science Foundation of China (grant no. 61775139 and grant no. 61373057), the Shanghai Natural Science Foundation (grant no. 15ZR1420800) and the Key Projects of Humanities and Social Sciences Research in Colleges and Universities in Anhui Province (grant no. SK2018A0630).

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author upon reasonable request.

Authors' contributions

WLY, SHP and LHJ conceived the study. WLY collected the data and performed the bioinformatics analyses. LL, XBL, CW, DS, CXD, and SCL performed quality control of the raw data and performed data analyses. WLY and SHP wrote the manuscript. SHP and LHJ supervised the study and agreed to be responsible for ensuring that all aspects of the study are accurate and have been appropriately investigated. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394-424, 2018.
2. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A and Jemal A: Colorectal cancer statistics, 2017. CA Cancer J Clin 67: 177-193, 2017.
3. Calvert PM and Frucht H: The genetics of colorectal cancer. Ann Intern Med 137: 603-612, 2002.
4. Liu F and Chang HC: Physiological links of circadian clock and biological clock of aging. Protein Cell 8: 477-488, 2017.
5. Kondratov R: Circadian clock and cancer therapy: An unexpected journey. Ann Med 46: 189-190, 2014.
6. Sahar S and Sassone-Corsi P: Metabolism and cancer: The circadian clock connection. Nat Rev Cancer 9: 886-896, 2009.
7. Kelleher FC, Rao A and Maguire A: Circadian molecular clocks and cancer. Cancer Lett 342: 9-18, 2014.
8. Stevens RG, Brainard GC, Blask DE, Lockley SW and Motta ME: Breast cancer and circadian disruption from electric lighting in the modern world. CA Cancer J Clin 64: 207-218, 2014.
9. Polo A, Singh S, Crispo A, Russo M, Giudice A, Montella M, Colonna G and Costantini S: Evaluating the associations between human circadian rhythms and dysregulated genes in liver cancer cells. Oncol Lett 14: 7353-7359, 2017.
10. Momma T, Okayama H, Saitou M, Sugeno H, Yoshimoto N, Takebayashi Y, Ohki S and Takenoshita S: Expression of circadian clock genes in human colorectal adenoma and carcinoma. Oncol Lett 14: 5319-5325, 2017.
11. Wendeu-Foyet MG, Koudou Y, Cénée S, Trétarre B, Rébillard X, Cancel-Tassin G, Cussenot O, Boland A, Bacq D, Deleuze JF, *et al*: Circadian genes and risk of prostate cancer: Findings from the EPICAP study. Int J Cancer 145: 1745-1753, 2019.
12. Bragantini D, Sivertsen B, Gehrman P, Lydersen S and Güzey IC: Variations in circadian genes and individual nocturnal symptoms of insomnia. The HUNT study. Chronobiol Int 36: 681-688, 2019.
13. Cooper GM, Nickerson DA and Eichler EE: Mutational and selective effects on copy-number variants in the human genome. Nat Genet 39: S22-S29, 2007.
14. Deng ZM, Liu L, Qiu WH, Zhang YQ, Zhong HY, Liao P and Wu YH: Analysis of genomic variation in lung adenocarcinoma patients revealed the critical role of PI3K complex. PeerJ 5: e3216, 2017.
15. Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, Cole K, Mossé YP, Wood A, Lynch JE, *et al*: Copy number variation at 1q21.1 associated with neuroblastoma. Nature 459: 987-991, 2009.
16. Cui M, Zheng M, Sun B, Wang Y, Ye L and Zhang X: A long noncoding RNA perturbs the circadian rhythm of hepatoma cells to facilitate hepatocarcinogenesis. Neoplasia 17: 79-88, 2015.
17. Ikeda M, Yu W, Hirai M, Ebisawa T, Honma S, Yoshimura K, Honma KI and Nomura M: cDNA cloning of a novel bHLH-PAS transcription factor superfamily gene, BMAL2: Its mRNA expression, subcellular distribution, and chromosomal localization. Biochem Biophys Res Commun 275: 493-502, 2000.
18. Schoenhard JA, Smith LH, Painter CA, Eren M, Johnson CH and Vaughan DE: Regulation of the PAI-1 promoter by circadian clock components: Differential activation by BMAL1 and BMAL2. J Mol Cell Cardiol 35: 473-481, 2003.
19. He CX, Avner P, Boitard C and Rogner UC: Downregulation of the circadian rhythm related gene Arntl2 suppresses diabetes protection in Idd6 NOD.C3H congenic mice. Clin Exp Pharmacol Physiol 37: 1154-1158, 2010.
20. Olkkonen J, Kouri VP, Kuusela E, Ainola M, Nordström D, Eklund KK and Mandelin J: DEC2 blocks the effect of the ARNTL2/NPAS2 dimer on the expression of PER3 and DBP. J Circadian Rhythms 15: 6, 2017.
21. Mazzoccoli G, Pazienza V, Panza A, Valvano MR, Benegiamo G, Vinciguerra M, Andriulli A and Piepoli A: ARNTL2 and SERPINE1: Potential biomarkers for tumor aggressiveness in colorectal cancer. J Cancer Res Clin Oncol 138: 501-511, 2012.
22. Mocellin S, Tropea S, Benna C and Rossi CR: Circadian pathway genetic variation and cancer risk: Evidence from genome-wide association studies. BMC Med 16: 20, 2018.
23. Ha NH, Long J, Cai Q, Shu XO and Hunter KW: The circadian rhythm gene Arntl2 is a metastasis susceptibility gene for estrogen receptor-negative breast cancer. PLoS Genet 12: e1006267, 2016.

24. Robinson MD, McCarthy DJ and Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140, 2010.
25. Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.
26. Smoot ME, Ono K, Ruscheinski J, Wang PL and Ideker T: Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics 27: 431-432, 2011.
27. Li S, Shui K, Zhang Y, Lv Y, Deng W, Ullah S, Zhang L and Xue Y: CGDB: A database of circadian genes in eukaryotes. Nucleic Acids Res 45: D397-D403, 2017.
28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102: 15545-15550, 2005.
29. Lin H and Zelterman D: Modeling survival data: Extending the Cox model. Taylor & Francis, pp85-86, 2002.
30. Karantanos T, Theodoropoulos G, Pektasides D and Gazouli M: Clock genes: Their role in colorectal cancer. World J Gastroenterol 20: 1986-1992, 2014.
31. Emons G, Spitzner M, Reineke S, Möller J, Auslander N, Kramer F, Hu Y, Beissbarth T, Wolff HA, Rave-Fränk M, *et al*: Chemoradiotherapy resistance in colorectal cancer cells is mediated by Wnt/β-catenin signaling. Mol Cancer Res 15: 1481-1490, 2017.
32. Skrzypczak M, Goryca K, Rubel T, Paziewska A, Mikula M, Jarosz D, Pachlewski J, Oledzki J and Ostrowski J: Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. PLoS One 5: e13091, 2010.
33. Kaiser S, Park YK, Franklin JL, Halberg RB, Yu M, Jessen WJ, Freudenberg J, Chen X, Haigis K, Jegga AG, *et al*: Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. Genome Biol 8: R131, 2007.
34. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, Gershenwald JE, Compton CC, Hess KR, Sullivan DC, *et al* (eds): AJCC cancer staging manual. Springer International Publishing, 2017.
35. Slattery ML, Mullany LE, Wolff RK, Sakoda LC, Samowitz WS and Herrick JS: The p53-signaling pathway and colorectal cancer: Interactions between downstream p53 target genes and miRNAs. Genomics 111: 762-771, 2019.
36. Laissue P: The forkhead-box family of transcription factors: Key molecular players in colorectal cancer pathogenesis. Mol Cancer 18: 5, 2019.
37. Zhang F, Gu W, Hurles ME and Lupski JR: Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10: 451-481, 2009.
38. Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC, Sudmant PH, Dorajoo R, Al-Shafai MN, Bottolo L, *et al*: Low copy number of the salivary amylase gene predisposes to obesity. Nat Genet 46: 492-497, 2014.
39. Tanenbaum DG, Hall WA, Colbert LE, Bastien AJ, Brat DJ, Kong J, Kim S, Dwivedi B, Kowalski J, Landry JC and Yu DS: TNFRSF10C copy number variation is associated with metastatic colorectal cancer. J Gastrointest Oncol 7: 306-314, 2016.
40. Clevers H and Nusse R: Wnt/β-catenin signaling and disease. Cell 149: 1192-1205, 2012.
41. Silva AL, Dawson SN, Arends MJ, Guttula K, Hall N, Cameron EA, Huang TH, Brenton JD, Tavaré S, Bienz M and Ibrahim AE: Boosting Wnt activity during colorectal cancer progression through selective hypermethylation of Wnt signaling antagonists. BMC Cancer 14: 891, 2014.
42. Sun S, Zhou L, Yu Y, Zhang T and Wang M: Knocking down clock control gene CRY1 decreases adipogenesis via canonical Wnt/β-catenin signaling pathway. Biochem Biophys Res Commun 506: 746-753, 2018.
43. He Y, Lin F, Chen Y, Tan Z, Bai D and Zhao Q: Overexpression of the circadian clock gene Rev-erbα affects murine bone mesenchymal stem cell proliferation and osteogenesis. Stem Cells Dev 24: 1194-1204, 2015.
44. Kettner NM, Katchy CA and Fu L: Circadian gene variants in cancer. Ann Med 46: 208-220, 2014.
45. Angelousi A, Kassi E, Nasiri-Ansari N, Weickert MO, Randeva H and Kaltsas G: Clock genes alterations and endocrine disorders. Eur J Clin Invest 48: e12927, 2018.
46. Relles D, Sendecki J, Chipitsyna G, Hyslop T, Yeo CJ and Arafat HA: Circadian gene expression and clinicopathologic correlates in pancreatic cancer. J Gastrointest Surg 17: 443-450, 2013.
47. Zhao H, Zeng ZL, Yang J, Jin Y, Qiu MZ, Hu XY, Han J, Liu KY, Liao JW, Xu RH and Zou QF: Prognostic relevance of Period1 (Per1) and Period2 (Per2) expression in human gastric cancer. Int J Clin Exp Pathol 7: 619-630, 2014.
48. Xue X, Liu F, Han Y, Li P, Yuan B, Wang X, Chen Y, Kuang Y, Zhi Q and Zhao H: Silencing NPAS2 promotes cell growth and invasion in DLD-1 cells and correlated with poor prognosis of colorectal cancer. Biochem Biophys Res Commun 450: 1058-1062, 2014.
49. Innominato PF, Roche VP, Palesh OG, Ulusakarya A, Spiegel D and Lévi FA: The circadian timing system in clinical oncology. Ann Med 46: 191-207, 2014.
50. Mazzoccoli G, Piepoli A, Carella M, Panza A, Pazienza V, Benegiamo G, Palumbo O and Ranieri E: Altered expression of the clock gene machinery in kidney cancer patients. Biomed Pharmacother 66: 175-179, 2012.
51. Kovanen L, Saarikoski ST, Haukka J, Pirkola S, Aromaa A, Lönnqvist J and Partonen T: Circadian clock gene polymorphisms in alcohol use disorders and alcohol consumption. Alcohol Alcohol 45: 303-311, 2010.
52. Karantanos T, Theodoropoulos G, Gazouli M, Vaiopoulou A, Karantanou C, Stravopodis DJ, Bramis K, Lymperi M and Pektasidis D: Association of the clock genes polymorphisms with colorectal cancer susceptibility. J Surg Oncol 108: 563-567, 2013.
53. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, *et al*: COSMIC: Somatic cancer genetics at high-resolution. Nucleic Acids Res 45: D777-D783, 2017.
54. Ciarleglio CM, Ryckman KK, Servick SV, Hida A, Robbins S, Wells N, Hicks J, Larson SA, Wiedermann JP, Carver K, *et al*: Genetic differences in human circadian clock genes among worldwide populations. J Biol Rhythms 23: 330-340, 2008.
55. Brady JJ, Chuang CH, Greenside PG, Rogers ZN, Murray CW, Caswell DR, Hartmann U, Connolly AJ, Sweet-Cordero EA, Kundaje A and Winslow MM: An Arntl2-driven secretome enables lung adenocarcinoma metastatic self-sufficiency. Cancer Cell 29: 697-710, 2016.