Computational and Structural Biotechnology Journal 19 (2021) 2190–2201

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

RGL2 as an age-dependent factor regulates colon cancer progression

Qingyu Cheng [a], Yupeng Wu [d], Honghai Xia [c,*], Xiaoyuan Song [a,b,*]

[a] Department of Obstetrics and Gynecology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230001, China

[b] MOE Key Laboratory for Membraneless Organelles and Cellular Dynamics, Hefei National Laboratory for Physical Sciences at the Microscale, CAS Key Laboratory of Brain Function and Disease, School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230026, China

[c] Department of Emergency Medicine, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230001, China

[d] Department of Biochemistry and Molecular Biology, School of Laboratory Medicine, Bengbu Medical College, 2600 Donghai Avenue, Bengbu, Anhui, 233030, China

ARTICLE INFO

Article history:
Received 11 November 2020
Received in revised form 2 April 2021
Accepted 3 April 2021
Available online 8 April 2021

Keywords:
Aging
Colon cancer
Transcriptome
Transcription factor
Survival

ABSTRACT

Colon cancer is the fourth leading cause of cancer-related death, and exhibited clinical differences among patients of different ages, including malignancy, metastasis, and mortality rate. Few studies, however, focus on the communications between aging and colon cancer. Here we identified age-dependent differentially expressed genes (DEGs) in colon cancer using TCGA transcriptome data. Through analyzing multi-omics high throughput data, including ATAC-Seq, DNaseI-Seq and ChIP-Seq, we obtained six age-dependent transcription factors in colon cancer, and their age-dependent targets, significantly affecting patients' overall survivals. Transcription factor ETS1 potentially functioned in both aging process and colon cancer progression through regulating its targets, *RGL2* and *SLC2A3*. In addition, comparing with its relative lower expression levels in elderly patients, higher levels of *RGL2* were detected in young patients, and significantly associated with larger tumor size, higher metastasis, and invasions of colon cancer, consistent with the clinical traits that young patients' colon cancer exhibited late stages with more aggressiveness. Thus, these elements may serve as keys linking aging and colon cancer, and providing new insights and basis for mechanism researches, as well as diagnosis and therapies of colon cancer, especially in young patients.

1. Introduction

Colorectal cancer is exceedingly prevalent, as the third most commonly diagnosed cancer worldwide, whose lethal rate ranks fourth, after lung cancer, liver cancer, and stomach cancer [1,2]. The risk of developing colorectal cancer has been highly associated with demographic, behavioral, and environmental factors, including inflammatory bowel disease (IBD), colorectal cancer history, body mass index (BMI), dietary habits, and physical activities [1,3]. At genomic levels, colorectal cancer was quite commonly accompanied by genetic and epigenetic alternations, which promoted tumorigenesis [4]. One of the responsible factors was chromosomal instability, which contained the loss of tumor suppressor genes and amplifications of oncogenes [4-6], affecting major signaling pathways (WNT, MAPK/PI3K, TGF-β, P53) [2]. Telomere dysfunction and DNA damage response alterations were two main causes underlying chromosomal instability, affecting genes responsible for critical cellular functions, including *APC*, *KRAS*, *PI3K* and *P53* [2].

In addition, telomere dysfunction and DNA damage response alterations were also two leading factors responsible for aging process. Telomere dysfunction was highly correlated with limitations of cellular proliferative capacity, and could activate canonical DNA damage response pathway, leading to cellular apoptosis or senescence [7]. Telomere dysfunction could induce expression of *p21* and *p16* to initiate cellular senescence through p53 and RB tumor suppressor pathways [7-10]. Loss of *p53* function could restore many cellular defects caused by telomere dysfunction, including growth arrest and cellular apoptosis [11].

These researches indicated a hint of the links between senescence process and cancer progression. Considering the interplay of senescence and cancer, previous research showed that telomere dysfunction suppressed tumorigenesis through p53-dependent cellular senescence induction, accompanied by global induction of *p53*, *p21* and senescence-associated-β-galactosidase staining [12]. Since pathological progressions were always complicated, senescent cells were reported to promote epithelial cell growth

* Corresponding authors.
E-mail addresses: 13966744976@163.com (H. Xia), songxy5@ustc.edu.cn (X. Song).

https://doi.org/10.1016/j.csbj.2021.04.006

and tumorigenesis through senescence-associated secretory phenotypes in a proinflammatory context [13,14]. Cellular senescence was thought to be one of the hallmarks of aging [15]. Thus, some connections might exit between aging process and cancer progression.

Actually, age acts as an important risk factor related to the occurrence of colorectal cancer. According to the newest statistics provided by the American Cancer Society, the incidence rates for colorectal cancer declined by 3.3% annually among patients aged 65 years and older during the 2000 s, while increased by approximately 2% annually among adults aged younger than 50 [16]. Similar to the incidence rate patterns, colorectal cancer death rates decreased by 3% annually among patients aged 65 and older, while increased by 1.3% annually among adults aged younger than 50 [16]. Nevertheless, few researches focused on the molecular changes of this age-related differences, or whether some protecting mechanisms existed in old individuals. The underlying mechanisms accounting for this phenomenon remain unclear. Since colon cancer accounts for a large proportion of colorectal cancer incidence, we focused on colon cancer in this study.

In this study, we extended our previous study on the interplay between senescence and cancer [39]. Integrating TCGA datasets and other published multi-omics datasets, including ATAC-Seq, DNaseI-Seq, ChIP-Seq, and RNA-Seq, we identified several transcription factors and their target genes, linking aging process and colon cancer progression, through multi-parallel analyses.

2. Materials and methods

2.1. Data processing and identifying DEGs

For the collected RNA-Seq data of *ETS1* KD and *HOXA9* KD, clean reads were mapped to the human reference genome hg38 using hisat2 [17,18] with default parameters. Gene expression levels were quantified using stringtie [18,19]. Differential analyses were performed with edgeR [20,21] in R environment [22]. Genes with P-value <0.05 were considered significant. Heatmap visualizations of DEGs were done with Pheatmap package [23].

For TCGA RNA-Seq data, quantifications of patients with age of 50 or less and age of 80 or more were considered, and read counts were processed in edgeR [20,21] for differential analysis. Genes with P-value<0.05 were considered significant.

2.2. Pathway enrichment analysis

KEGG pathway enrichments were done with geneSCF [24]. For graphical demonstrations, we presented significantly enriched items related with senescence, immune, and cancer, for clear visualizations. Items enriched with P-value <0.05 were considered significant. Bar plots of enrichment results were done with ggplot2 package [25], and alluvial plots were done with ggalluvial package [26]. GSEA (Gene set enrichment analysis) was done with GSEA software [53,54].

2.3. ATAC-Seq and ChIP-Seq data processing

For ATAC-Seq data and DNaseI-Seq, Bowtie (v1.2.3) [27] or Bowtie2 (v2.3.4.3) [28] was used to align reads to human reference genome hg38 with default parameters. Peak calling was performed using MACS2 (v2.1.2) [29] with default parameters.

For ChIP-Seq data, the peak files and target files were directly downloaded from the Cistrome data browser [30,31]. The detailed peaks were modified from WASHU epigenome browser [32]. Targets were identified based on corresponding gene score greater than 0.2 in the downloaded files.

2.4. Identification of transcription factor binding site motifs

All ATAC-Seq and DNaseI-Seq peaks were extracted for motif analysis using the HOMER motif finding tool with default parameters [33]. Motifs with P-value<0.05 were considered significant.

2.5. Survival analysis

The clinical data for survival analysis were downloaded from UCSC Xena TCGA GDC hub. Survival analyses were performed with the survival package [34]. Cutoff thresholds were conducted with Q1 versus Q4 quartile for high and low expression groups of various genes, thus patients with the top 25% of gene expression belonged to high expression group, and patients with the minimum 25% of gene expression belonged to low expression group. Survival plots were proceeded with survminer package [35] in R environment [22].

2.6. Data collection

We downloaded expression profiles and clinical traits data of human colon cancer of TCGA and normal colon data of GTEx from UCSC Xena browser (https://xenabrowser.net/datapages/).

The ATAC-Seq and DNaseI-Seq datasets from GSE83968 [36], including two ATAC-Seq data and eleven DNaseI-Seq data, were downloaded from the EBI (European Bioinformatics Institute) database (https://www.ebi.ac.uk/).

For ChIP-Seq data of ETS1 and PKNOX1, the peak files and targets were directly downloaded from the Cistrome data browser (http://cistrome.org/db/) [30,31]. The *ETS1* KD dataset GSE153852 [37] and *HOXA9* KD dataset GSE100144 [38] were utilized to confirm their regulatory relationships to potential targets. All these datasets were downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/) unless otherwise specified.

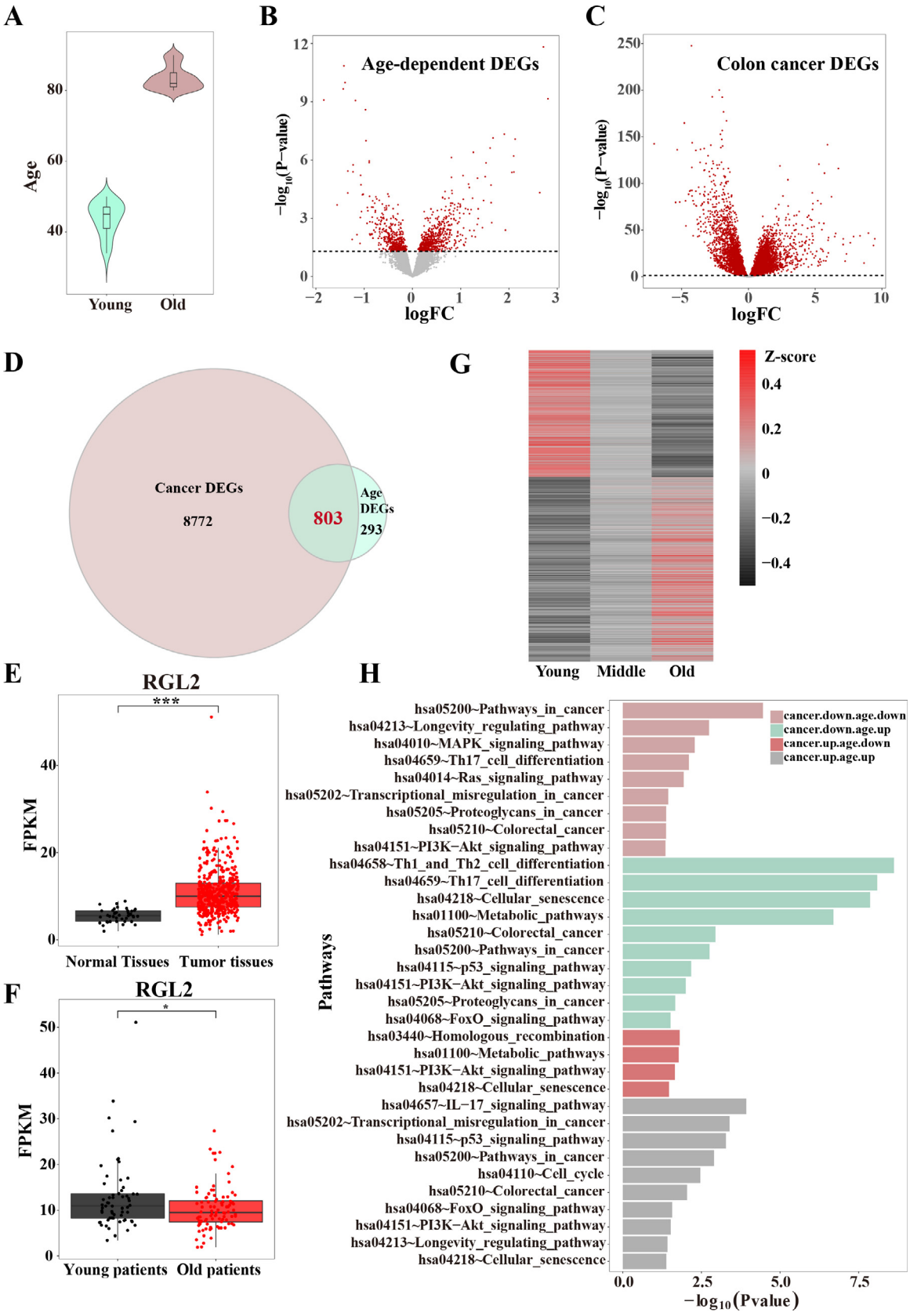
3. Results

3.1. Identifying age-dependent differentially expressed genes in colon cancer patients

Previously we found that age-dependent differentially expressed genes (DEGs) in senescent mouse embryonic fibroblasts

Fig. 1. Characterizations of age-dependent differentially expressed genes (DEGs) in colon cancer. (**A**) Age division of young and old patients of colon cancer. Age of 50 or less was defined as young, and age of 80 or more was defined as old. (**B**) Volcano plot of age-dependent DEGs in colon cancer. (**C**) Volcano plot of DEGs in colon tumor tissues, comparing to adjacent normal tissues. Significant DEGs (P-value < 0.05) were labeled red based on the thresholds, indicated by dotted lines. X-axis stands for $\log_2$ (fold change) and y-axis stands for $-\log_{10}$ (P-value). (**D**) Venn plot of the common genes between age-dependent DEGs (from B) and cancer-dependent DEGs (from C) in colon cancer. (**E**) *RGL2* was upregulated in colon tumor tissues, comparing with adjacent normal tissues. (**F**) *RGL2* was upregulated in young colon cancer patients, comparing with old patients. *** P-value < 0.001, * P-value < 0.05, by *Students' t-test*. (**G**) The gradual expression changes of the 803 common age-dependent DEGs from young to old colon cancer patients. Age between 51 and 79 was defined as 'middle'. Expression levels of DEGs were normalized to z-scores, indicated by different colors. (**H**) KEGG pathway enrichment results of the four groups of common DEGs, which were down-regulated in both old patients comparing to young patients and colon tumor tissues comparing to adjacent normal tissues, up-regulated in both old patients and colon tumor tissues, up-regulated in old patients while down-regulated in colon tumor tissues, and down-regulated in old patients while up-regulated in colon tumor tissues. Pathways related to senescence, immune and cancers, were selected. All pathways presented were significantly enriched, with P-value < 0.05. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

could also function in cancers, including pancreatic cancer, colorectal cancer and cholangiocarcinoma [39]. To investigate the role of senescence in the progression of colon cancer, we obtained colon cancer data from UCSC xena browser TCGA GDC hub. Focusing on age-dependent regulators in colon cancer, we defined age of 50 or less as 'young' and age of 80 or more as 'old' (Fig. 1**A**). Based on 61



young patients' and 92 old patients' RNA-Seq data, we identified 1096 DEGs as set 1 (639 up-regulated and 457 down-regulated in old patients, Fig. 1**B**). We then obtained the DEGs between colon tumor tissues and adjacent normal tissues as set 2 (Fig. 1**C**). Focusing on genes functioning in both aging and colon cancer progression, we obtained 803 common genes between DEGs of set 1 and set 2 (Fig. 1**D**), one of which was *RGL2*, which is upregulated in colon tumor tissues comparing with adjacent normal tissues (Fig. 1**E**), and in young patients comparing with old patients (Fig. 1**F**). Taking the average expression levels as indicators, these 803 common genes exhibited gradual downregulations or upregulations from 'young' to 'old' patients, where age between 51 and 79 were defined as 'middle' (Fig. 1**G**).

We divided these 803 common genes into four groups as follows, down-regulated in both old patients comparing to young patients and colon tumor tissues comparing to adjacent normal tissues (150 genes, **Figure S1A**), up-regulated in both old patients and colon tumor tissues (277 genes, **Figure S1B**), up-regulated in old patients while down-regulated in colon tumor tissues (201 genes, **Figure S1C**), and down-regulated in old patients while up-regulated in colon tumor tissues (175 genes, **Figure S1D**). KEGG pathway enrichments were then performed on these four groups respectively using geneSCF [24]. Significantly enriched pathways among different groups included senescence related pathways, as exemplified by cellular senescence, longevity regulating pathway, p53 signaling pathway, and several others (Fig. 1**H**), and cancer related pathways, as exemplified by pathways in cancer, colorectal cancer, and several others (Fig. 1**H**). In addition, significantly enriched pathways also included immune related pathways, as exemplified by Th17 cell differentiation, Th1 and Th2 cell differentiation, and IL17 signaling pathway (Fig. 1**H**). Furthermore, Gene Set Enrichment Analysis (GSEA) showed that the involved pathways mainly included immune-related pathways, and metabolism-related pathways (**Table S1**), as exemplified by interferon signaling (**Figure S2A**), cytokine signaling in immune system (**Figure S2B**), extracellular matrix organization (**Figure S2C**) and metabolism of lipids (**Figure S2D**).

3.2. Survival analysis reveals potential age-dependent candidates functioning in colon cancer

To unravel the relations between DEGs and patients' vital status, the above mentioned four groups of common genes were subjected to survival analysis. In the analysis, cutoff thresholds were determined by Q1 versus Q4 quartile to define high and low expression levels, which meant that patients with the top 25% of gene expression belonged to high expression group, and patients with the minimum 25% of gene expression belonged to low expression group. Totally 64 genes (13, 18, 16, and 17 genes in four groups respectively) were significantly (P-value < 0.05) associated with patients' survival probabilities (Fig. 2**A**, **Table S2**), which were then selected as potential candidates functioning in colon cancer. As exemplified by two coding genes (Fig. 2**B, C**) and one non-coding RNA (Fig. 2**D**), low expression of protein coding gene *TMEM165* significantly decreased patients' overall survival probabilities (Fig. 2**B**), while high expression of protein coding gene *ASPHD1* and non-coding RNA *KRT8P12* significantly decreased patients' overall survival probabilities (Fig. 2**C, D**).

3.3. Identifying transcription factors functioning in colon cancer progression

Aiming to find upstream transcription factors (TFs) of the 64 potential candidates functioning in colon cancer from survival analysis, we obtained ATAC-Seq or DNaseI-Seq data of 13 colon cancer cell lines. After aligning sequencing reads and evaluating open chromatin accessibilities, we subjected 13 sets of peaks separately to HOMER to find corresponding motifs [33], and identified 13 sets of regulatory TFs based on P-value <0.05 in the analyses (Fig. 3**A**). The 146 common factors among 13 sets of TFs were considered as candidate colon cancer TFs (Fig. 3**B**). We further overlapped these 146 candidate TFs with the 1096 age-dependent DEGs of colon cancer, and considered the common 6 factors as age-dependent colon cancer TFs (Fig. 3**C**), which were ATF3, HOXA9, ETS1, PKNOX1, STAT1, and HOXB4. To visualize these TF genes' differential expressions, three of them were represented as examples, *PKNOX1*, *HOXA9*, and *ETS1*. Both *PKNOX1* and *HOXA9* were significantly up-regulated in old patients, comparing with young patients, and significantly up-regulated in colon tumor tissues, comparing with adjacent normal tissues (Fig. 3**D, E**). *ETS1* was down-regulated in both old patients comparing with young patients, and colon cancer tissues, comparing with adjacent normal tissues (**Figure S3A, B**), although different expression levels of *ETS1* were not significantly correlated with patients' overall survivals (**Figure S3C**). These results revealed that several TFs (ATF3, HOXA9, ETS1, PKNOX1, STAT1, and HOXB4) may function in both aging and colon cancer.

3.4. Identifying potential targets of age-dependent colon cancer transcription factors

Using ChIP-Seq data from Cistrome DB [30,31], we found that two age-dependent candidate TFs, ETS1 and PKNOX1, directly bound to the genomic loci of a majority of the above mentioned 64 potential candidates functioning in colon cancer (Fig. 4**A-D**). Among the 64 potential candidate genes, ETS1 had binding sites on 70.3% (45 out of 64) in CUTLL1 cells, a T lymphocyte cell line (Fig. 4**A**) and on 62.5% (40 out of 64) in K562 cells, a human erythroleukemia cell line (Fig. 4**B**) respectively; while PKNOX1 had binding sites on 64.1% (41 out of 64) and 70.3% (45 out of 64) in two K562 replicates respectively (Fig. 4**C, D**). For detailed visualization, we found that ETS1 exhibited consistent binding at the promoters of *RGL2* (Fig. 4**E**) and *NCKAP5L* (Fig. 4**F**) in CUTLL1 (a T lymphocyte cell line), GM12878 (a lymphoblastoid cell line), K562 (a human erythroleukemia cell line), and PANC-1 (a pancreatic cancer cell line) cells. PKNOX1 exhibited consistent binding at the promoters of *FH* (Fig. 4**G**) and *VPS52* (Fig. 4**H**) in K562 (a human erythroleukemia cell line), MCF-7 (a breast cancer cell line), and GM12878 (a lymphoblastoid cell line) cells. Moreover, different expression levels of these target genes significantly affected patients' overall survival probabilities (**Figure S4**), suggesting their essential roles in colon cancer progression.

To validate these TFs' regulations on the potential targets, *ETS1* knocking down (KD) RNA-Seq data were collected from GSE153852 [37]. The data revealed that some targets were significantly affected by *ETS1* knockdown in both Rh30 and Rh41 cells, two rhabdomyosarcoma cell lines (Fig. 5**A, B**). In addition, both *RGL2* and *SLC2A3* genes were significantly down-regulated when *ETS1* was knocked down (Fig. 5**C, D**). *SLC2A3* was previously reported to be associated with colon cancer [40,41], supporting our analyzed data. Moreover, high expression levels of both *SLC2A3* and *RGL2* were significantly correlated with patients' low survivals (**Figure S4A, B**). In addition, KEGG pathways enrichments analysis of DEGs after *ETS1* disruption in both Rh30 and Rh41 cells revealed that many senescence, immune and cancer related pathways were significantly affected (Fig. 5**E**).

In addition to ETS1's regulatory roles, RNA-Seq data of another TF, HOXA9, were collected from GSE100144 [38]. Knocking down *HOXA9* could up-regulate *RGL2* (**Figure S5**). These results implied that *ETS1* potentially functions as an upstream TF in both senescence and colon cancer, and additionally that *SLC2A3* and *RGL2*

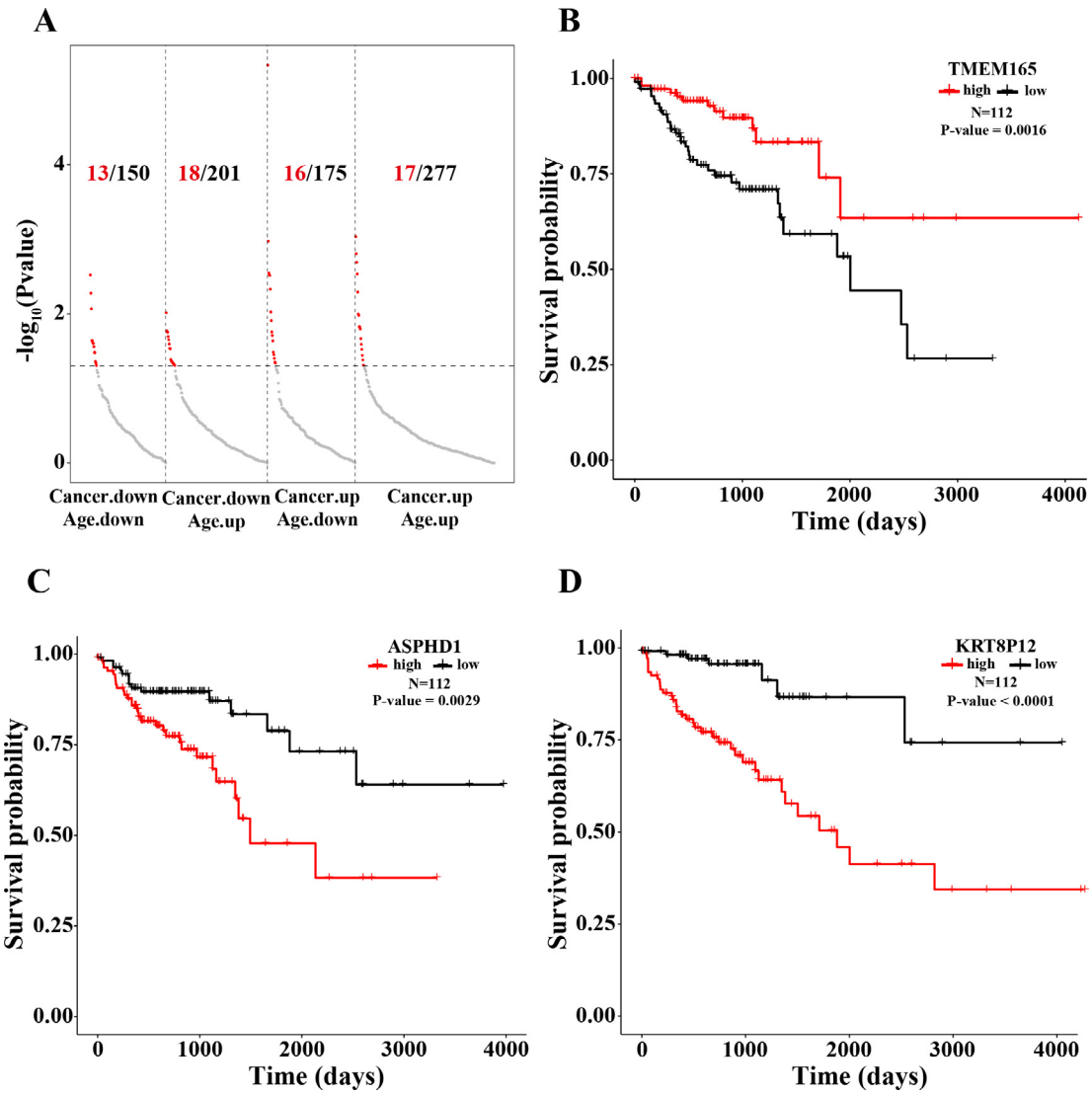


Fig. 2. Survival analysis results of age-dependent DEGs of colon cancer. (**A**) Survival analysis of the above mentioned four groups of common DEGs. X-axis stands for four groups of common DEGs and y-axis stands for -$\log_{10}$(P-value). DEGs with P-value<0.05 were labeled red. (**B-D**) Examples of survival analyses, including *TMEM165* (**B**), *ASPHD1* (**C**), and *KRT8P12* (**D**). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are direct targets of ETS1, playing their essential roles downstream in colon cancer progression.

3.5. RGL2 functions in both aging and colon cancer

To address the function of *RGL2*, a candidate we chose for further research, we also collected expression data of normal colon tissues from Genotype-Tissue Expression (GTEx). We defined high and low expression groups of *RGL2* with Q1 versus Q4 quartile in TCGA data (Fig. 6**A**) and in GTEx data (**Figure S6**). We then checked the expression levels of senescence, proliferation, and cancer markers. In colon cancer patients with low levels of *RGL2*, *PTEN* and *SMAD4*, two tumor suppressors, were significantly up-regulated (Fig. 6**B, C**), and *CDKN1A* and *CDKN2B*, two cell cycle regulators and senescence markers, were significantly up-regulated (Fig. 6**D, E**). In GTEx data, we observed similar differential expression patterns of these marker genes in normal colon tissues. Using all available samples for calculating correlations between *RGL2* and other marker genes, we found that the expression of the senescence marker *CDKN2B* was negatively correlated with that of *RGL2* (Fig. 6**F**); and *P53*, a tumor suppressor and senescence marker gene, was also negatively correlated with *RGL2* in expression level (Fig. 6**G**). Expressions of *ERCC1*, *EGFR* and *S100A2*, three cancer markers, however, were positively correlated with that of *RGL2* (Fig. 6**H-J**). The expression patterns of these markers were consistent between TCGA colon cancer samples and GTEx normal colon tissues. These results indicated that *RGL2* is involved in both aging process and colon cancer progression.

In addition, the expression levels of *RGL2* were associated with clinical traits. The TNM system is the most widely used staging system, and the higher the number after the letters, the more deteriorating the cancer becomes. Significant differential expressions of *RGL2* between M0 and M1 stages of colon cancer indicated that higher levels of *RGL2* were associated with cancer metastasis (Fig. 6**K**). Differential expressions of *RGL2* between N0, N1 and N2 stages showed that higher levels of *RGL2* were highly associated with more lymph nodes that contained cancer (Fig. 6**L**). This was consistent with the positive correlations between *RGL2* expression levels and positive lymph nodes detected by HE staining (Fig. 6**M**). Moreover, differential expressions of *RGL2* between T1, T2, T3, and T4 stages indicated that higher levels of *RGL2* were significantly associated with larger tumor size (Fig. 6**N**). In another staging sys-

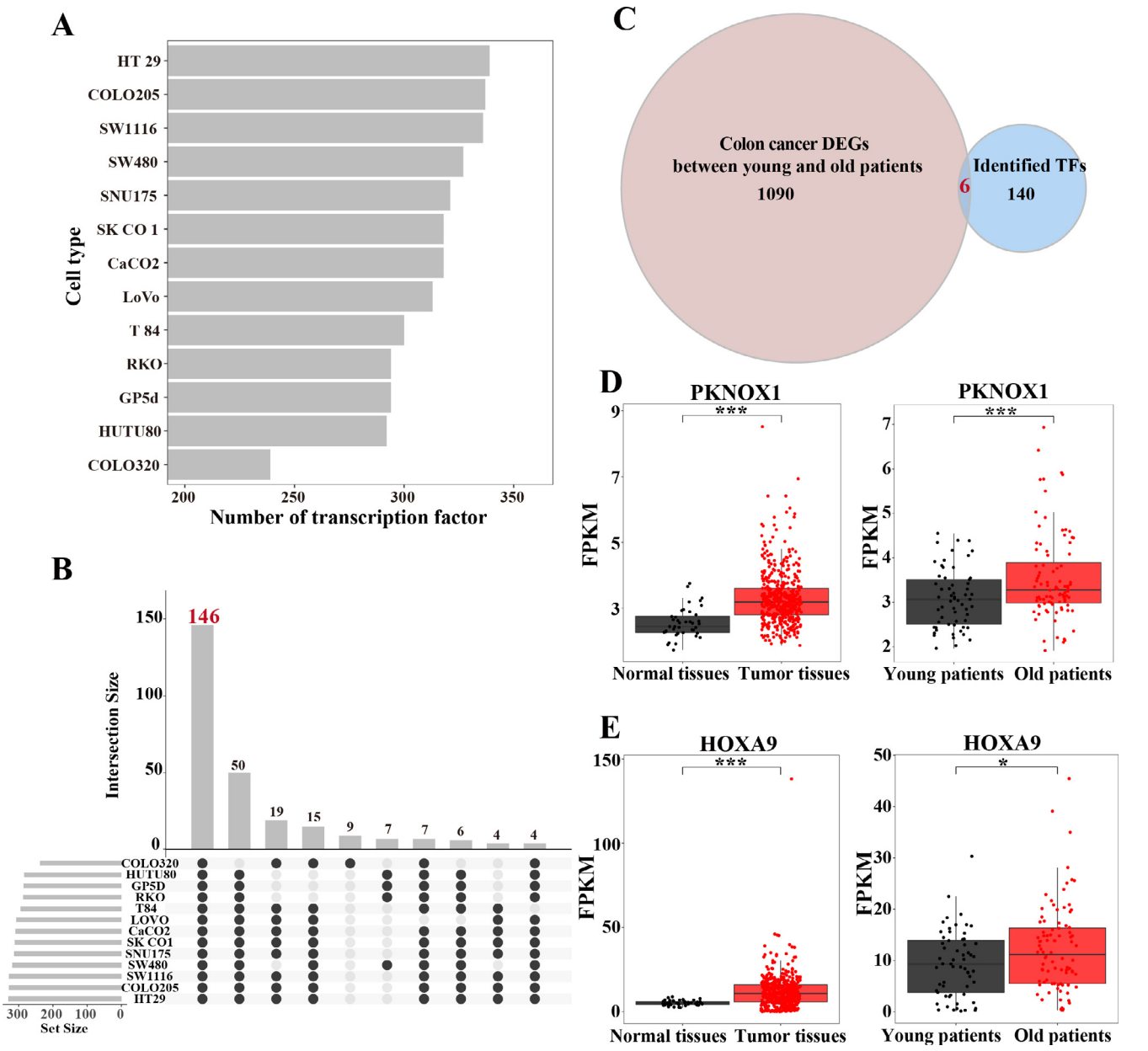


Fig. 3. Open chromatin accessibilities evaluation and transcription factors (TFs) identification in colon cancer. (**A**) The numbers of TFs identified as significant with HOMER [33] in 13 colon cancer cell lines. (**B**) Common TFs of 13 cell lines. Each of the 13 sets of TFs was divided into several sub-groups, indicated by black dots horizontally. Only top 10 sub-groups by frequency were presented. Y-axis stands for gene numbers, and each column stands for a single sub-group. We identified 146 common TFs among 13 sets, indicated in red number. (**C**) Venn plot of the common parts between age-dependent colon cancer DEGs and 146 identified common TFs in colon cell lines. (**D, E**) Differential expressions of *PKNOX1* (**D**) and *HOXA9* (**E**) in colon tumor tissues, comparing with adjacent normal tissues, and in old colon cancer patients, comparing with young patients. *** P-value < 0.001, * P-value < 0.05 by *Students' t-test*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tem, *RGL2* levels were also positively correlated with the tumor size and metastasis (Fig. 6**O**). In addition, cancers with venous invasion exhibited significantly higher expressions of *RGL2* (Fig. 6**P**).

Intriguingly, *RGL2* was not only upregulated in colon cancer tissues comparing to adjacent normal tissues (Fig. 1**E**), it expressed even higher in young colon cancer patients than in old patients (Fig. 1**F**), and its higher expression levels indicated more aggressive and deteriorating colon cancer (larger tumor size, higher metastasis, and invasions, Fig. 6**K-P**), consistent with the clinical data that colon cancer exhibits higher metastasis and malignancy in young patients. These results suggested that *RGL2* may be a predictor in colon cancer, and more importantly, it may serve as a new indicator for the elevated cancer risk in young colon cancer patients.

4. Discussion

To unravel the relationships between aging and cancers, we evaluated the potential roles of age-dependent DEGs in colon cancer progression. Enrichment analyses indicated that these DEGs were involved in senescence, immune, cancer, and metabolism related pathways. Combining multi-omics high throughput data, we elucidated a set of age-dependent colon cancer TFs and their

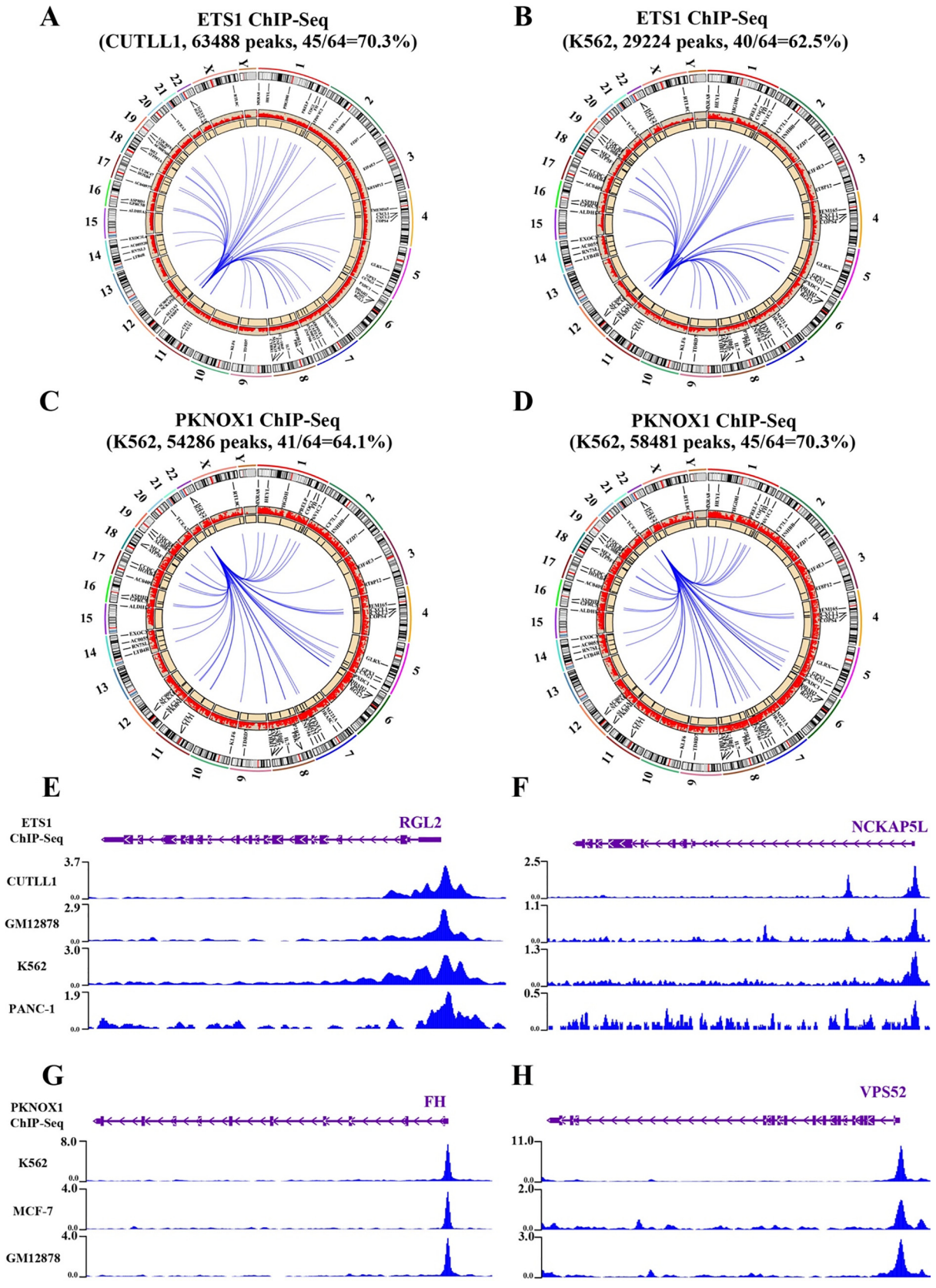
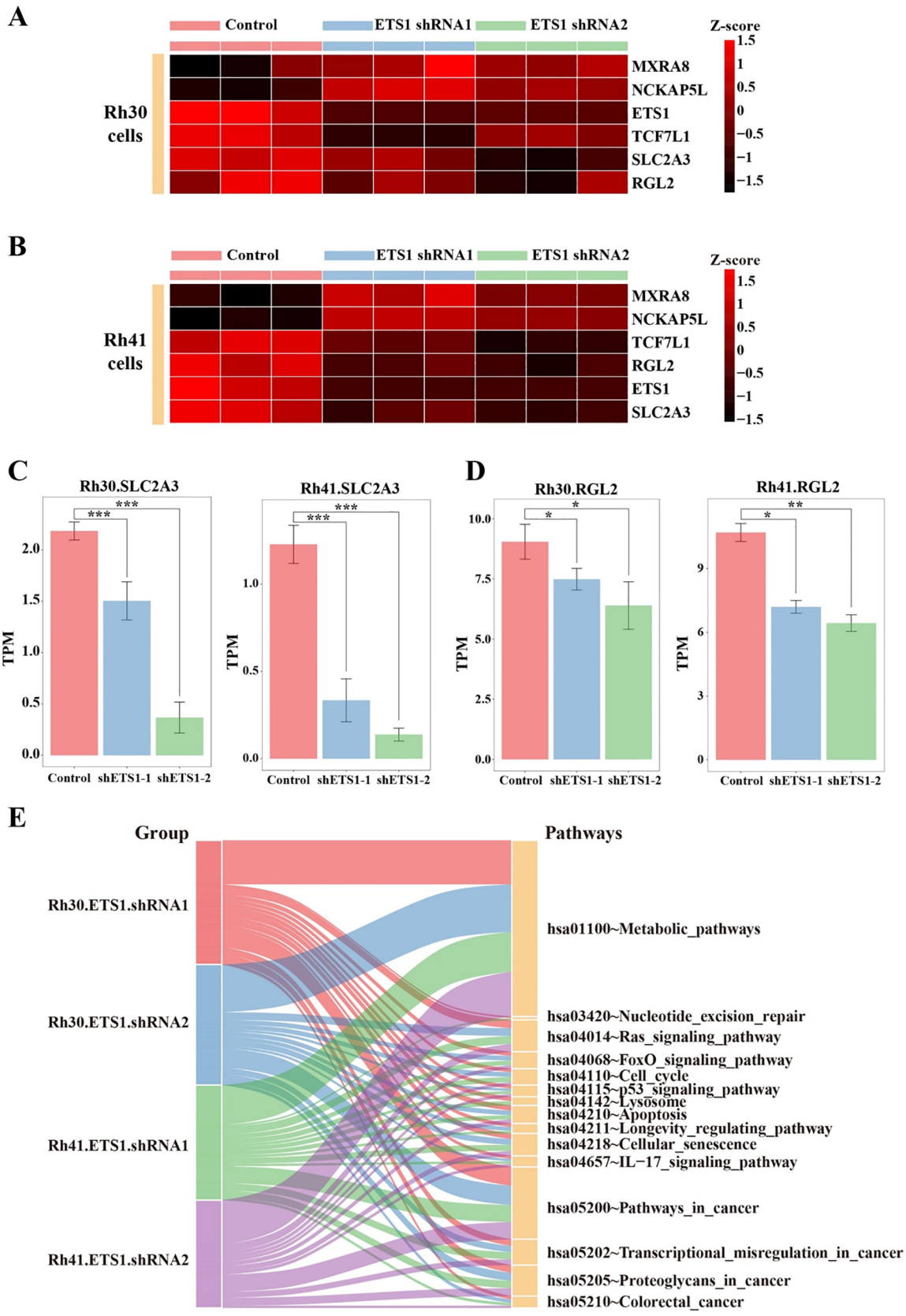


Fig. 4. Transcription factors bound at genomic regions of 64 candidates from survival analysis. (**A, B**) ChIP-Seq data representations of ETS1 binding in CUTLL1 cells (**A**) and K562 cells (**B**). (**C, D**) ChIP-Seq data representations of PKNOX1 binding in two replicates of K562 cells. Chromosome information is labeled in the outermost circle, and the 64 candidate genes are labeled in the immediate inner circle. ChIP-Seq peaks are labeled in the red circle, and the immediate inner black circle represents 64 candidates' genomic locations. The blue lines link the TF and its potential targets. (**E, F**) Detailed binding states of ETS1 on promoter regions of *RGL2* (**E**) and *NCKAP5L* (**F**) in four cell lines. (**G, H**) Detailed binding states of PKNOX1 on promoter regions of *FH* (**G**) and *VPS52* (**H**) in three cell lines. The ChIP-Seq data were collected from Cistrome DB (http://cistrome.org/db/). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)


A
Control
ETS1 shRNA1
ETS1 shRNA2
Z-score
1.5
1
0.5
0
−0.5
−1
−1.5
Rh30 cells
MXRA8
NCKAP5L
ETS1
TCF7L1
SLC2A3
RGL2
B
Control
ETS1 shRNA1
ETS1 shRNA2
Z-score
Rh41 cells
MXRA8
NCKAP5L
TCF7L1
RGL2
ETS1
SLC2A3
C
Rh30.SLC2A3
Rh41.SLC2A3
TPM
Control
shETS1-1
shETS1-2
D
Rh30.RGL2
Rh41.RGL2
E
Group
Pathways
Rh30.ETS1.shRNA1
Rh30.ETS1.shRNA2
Rh41.ETS1.shRNA1
Rh41.ETS1.shRNA2
hsa01100~Metabolic_pathways
hsa03420~Nucleotide_excision_repair
hsa04014~Ras_signaling_pathway
hsa04068~FoxO_signaling_pathway
hsa04110~Cell_cycle
hsa04115~p53_signaling_pathway
hsa04142~Lysosome
hsa04210~Apoptosis
hsa04211~Longevity_regulating_pathway
hsa04218~Cellular_senescence
hsa04657~IL−17_signaling_pathway
hsa05200~Pathways_in_cancer
hsa05202~Transcriptional_misregulation_in_cancer
hsa05205~Proteoglycans_in_cancer
hsa05210~Colorectal_cancer

candidate target genes potentially functioning in both aging and colon cancer. Among them, ETS1, as an upstream TF, potentially regulated aging process and colon cancer progression through binding its two direct targets, *RGL2* and *SLC2A3*. Age-related dysregulations of *RGL2* were significantly correlated with different clinical phenotypes of colon cancer. Higher levels of *RGL2* were detected in young colon cancer patients (Fig. 1**F**) and significantly associated with larger tumor size, higher metastasis, and invasions of colon cancer (Fig. 6**K-P**), consistent with the clinical data that colon cancer exhibited higher metastasis and malignancy in young patients. In summary, we elucidated essential roles of several TFs, together with their targets, linking aging and colon cancer, serving as potential diagnostic and prognostic indicators of colon cancer, especially in young patients. Specifically, *RGL2*, regulated by different transcription factors, may function in maintaining the dynamics between aging and colon cancer progression, as well as be a new indicator for the elevated cancer risk in young colon cancer patients.

SLC2A3, as a candidate we identified in this study potentially functioning in colon cancer, was reported to be activated by CAV1 elevation, to further increase glucose uptake and ATP production in colon cancer cells, which might contribute to malignant progressions of cancers [40]. Moreover, *SLC2A3* could be inhibited by miR-29c elevation, to further inhibit cell proliferation and glycolysis in prostate cancer [42], supporting that *SLC2A3* plays important roles in cancer progressions. In addition to *SLC2A3*, our study revealed that *RGL2* is potentially a new biomarker in colon cancer, especially in young patients.

After searching for age-dependent TFs, we found disruptions of two factors, *ETS1* and *HOXA9*, could affect the expression levels of *RGL2*. ETS1 is a TF previously reported to show dichotomous roles as an oncogene or a tumor suppressor gene in diverse cancers [48-52]. *ETS1* was involved in cellular senescence pathway, in response to microenvironment changes, including oncogenic stress, oxidative stress, ionizing radiation, and tumor suppressor loss. However, in TCGA colon cancer, different expression levels of *ETS1* were not significantly correlated with patients' overall survivals (**Figure S3C**). Thus, we focused on its target gene *RGL2*, of which dysregulation was significantly associated with the overall survivals of colon cancer patients (**Figure S4A**). Both *ETS1* and *RGL2* were down-regulated (**Figure S3A, 1F**), while *HOXA9* was up-regulated in old colon cancer patients (Fig. 3**E**), comparing with young patients. Knocking down *ETS1* decreased *RGL2* expression towards aging process (Fig. 5**A, B**), while knocking down *HOXA9* increased *RGL2* expression towards anti-aging process (**Figure S5**). Nevertheless, expressions of *ETS1* and *RGL2* were not significantly correlated in these young and old patients (Pearson correlation, P-value = 0.3527). We thus believe that the expression of *RGL2* was determined by different combinations of upstream transcription factors in different micro-environments, such as cancer cells in different stages.

Our results demonstrated that the age-related dysregulations of *RGL2* contributed to elevated colon cancer malignancy and metastasis in young patients. Firstly, *RGL2* was up-regulated in colon tumor tissues, comparing with adjacent normal tissues (Fig. 1**E**). Dysregulations of *RGL2* significantly affected tumor suppressors, as exemplified by *PTEN* and *SMAD4* (Fig. 6**B, C**). Secondly, high expression levels of *RGL2* were significantly correlated with patients' low survivals (**Figure S4A**). Thirdly, *RGL2* was up-regulated in young colon cancer patients, comparing with old patients (Fig. 1**F**). In addition to tumor suppressors, dysregulations of *RGL2* significantly affected senescent markers, as exemplified by two cell cycle regulators *CDKN1A* and *CDKN2B* (Fig. 6**D, E**). Furthermore, age-related dysregulations of *RGL2* were significantly correlated with different clinical phenotypes of colon cancer. Higher levels of *RGL2* were significantly associated with larger tumor size, higher metastasis, and invasions of colon cancer (Fig. 6**K-P**), indicating that *RGL2* was involved in both aging process and colon cancer progression. These results were consistent with the clinical data that colon cancer exhibited higher metastasis and malignancy in young patients. We thus believe that *RGL2*, regulated by different regulatory factors, may play essential roles in maintaining the dynamics between aging process and colon cancer progression. Since *RGL2* was upregulated in colon cancer tissues (Fig. 1**E**), and particularly expressed higher in young colon cancer patients (Fig. 1**F**), *RGL2* may be an indicator, accounting for the age-differentiated clinical outcomes, and a therapeutic target of colon cancer in young patients. Furthermore, these results shed light on the thinking that whether these factors might dictate age-differentiated medications and treatment approaches, which would be a great follow-up study in the near future.

Considering cell type's specificity, colon cancer cells were selected for evaluating open chromatin accessibilities. However, since we did not find ChIP-Seq data of the several TFs and *ETS1* disruption data in colon cancer cells, non-colon cells were selected for checking TFs' binding sites, including K562, CUTLL1, GM12878, PANC-1, and MCF-7 cells, and for *ETS1* disruption, including Rh30 and Rh41 cells (rhabdomyosarcoma cell lines). Restriction of these cell types to colon cancer cells would benefit further mechanism researches.

With increasing understanding on the pathogeneses and mechanisms of cancer developments, researchers found cancer stem cells were responsible for cancer recurrence, metastasis, and resistance to therapeutic drugs [43,44]. Thus, targeting cancer stem cells could be a promising strategy to eliminate cancers [45]. The TF, HOXA9, was reported to be mediated by long non-coding RNA *HOTTIP* to regulate downstream targets, and further alter cancer stem cell stemness, including tumorigenesis [46]. *HOXB4*, as a self-renewal regulator and another candidate in our analyses, was involved in the regulation of proliferation and tumorigenic potential of leukemia cells by *IGF2BP1*, through multiple pathways of stemness [47]. Targeting these stemness associated mechanisms would be a great follow-up study as well. We believe these results will shed light on future researches regarding aging process and cancer progression, as well as cancer therapies.

CRediT authorship contribution statement

Qingyu Cheng: Conceptualization, Data curation, Investigation, Visualization, Writing - original draft. **Yupeng Wu:** Writing - review & editing. **Honghai Xia:** Conceptualization, Investigation, Supervision, Writing - review & editing. **Xiaoyuan Song:** Conceptualization, Project administration, Supervision, Writing - review & editing.

Fig. 5. Characterizations of differentially expressed genes (DEGs) in colon cancer affected by *ETS1* disruption. (**A, B**) *ETS1* disruption significantly affected some of the DEGs in colon cancer in Rh30 cells (**A**) and Rh40 cells (**B**). Expression levels were normalized to z-scores, indicated by different colors. (**C, D**) Expression comparisons of *SLC2A3* (**C**) and *RGL2* (**D**) in *ETS1* disruption RNA-Seq data in Rh30 and Rh41 cells. *** P-value < 0.001, ** P-value < 0.01, * P-value < 0.05 by differential analysis described in the Methods and materials. (**E**) KEGG pathway enrichment analysis results of DEGs from *ETS1* disruption in Rh30 and Rh41 cells. In each cell line, two shRNAs were applied to knock down *ETS1*, indicated by shRNA1 and shRNA2. On the left side, different colors indicated different sets of DEGs, and on the right side were different pathways. Pathways related to senescence, immune, and cancers, were selected. All pathways presented were significantly enriched, with P-value < 0.05. The ETS1 disruption dataset was collected from GSE153852 [37].

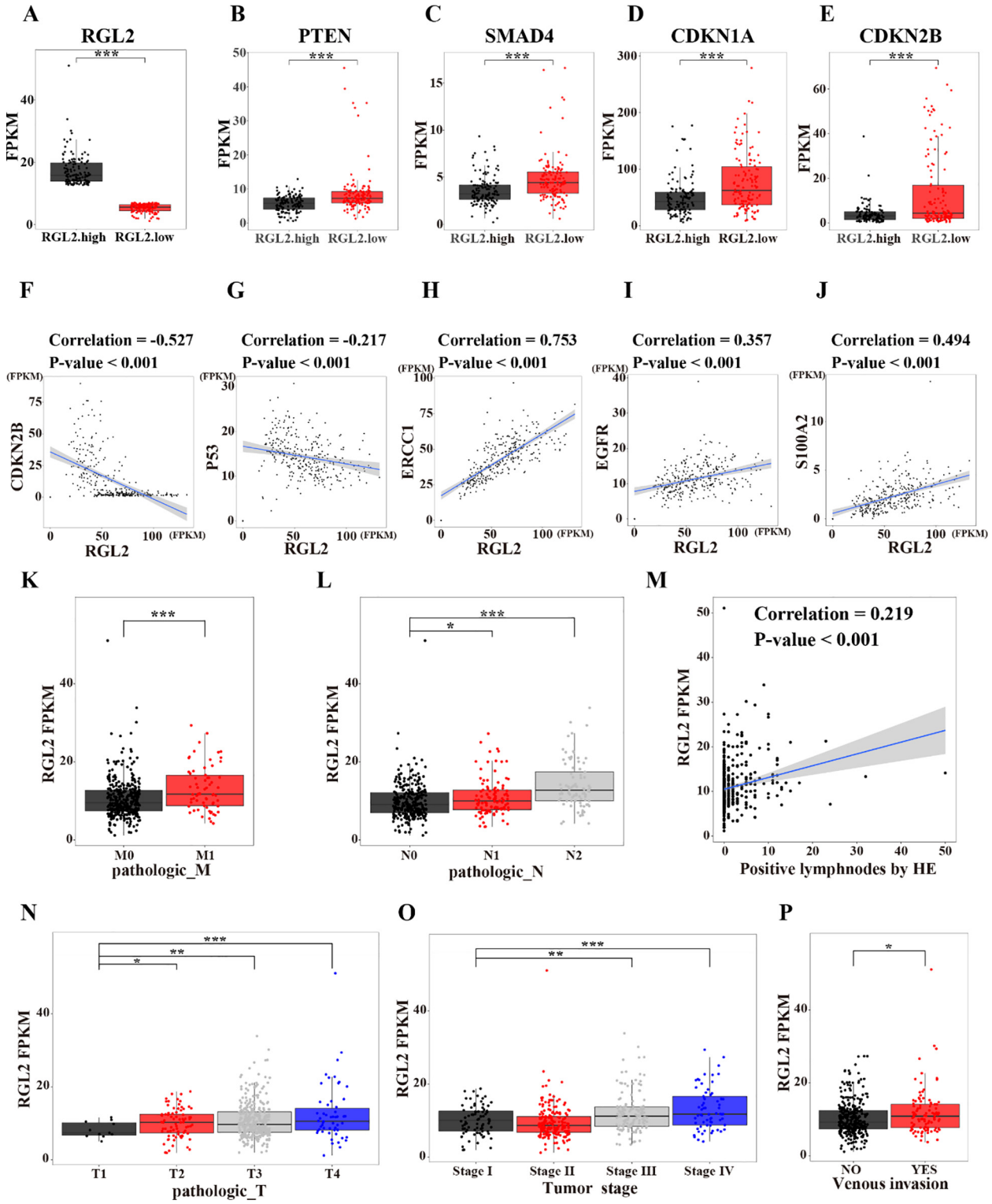


Fig. 6. Correlations of *RGL2* expression with different senescence and cancer markers' expression, and with clinical traits. (**A**) High and low expression groups of *RGL2* determined by Q1 versus Q4 quartile. (**B-E**) Expression comparisons of *PTEN* (**B**), *SMAD4* (**C**), *CDKN1A* (**D**) and *CDKN2B* (**E**) in the defined high and low expression groups of *RGL2*. (**F-J**) Corresponding expression correlations of *RGL2* and *CDKN2B* (**F**), *P53* (**G**), *ERCC1* (**H**), *EGFR* (**I**), and *S100A2* (**J**) in transcriptome data of normal colon tissues from GTEx. (**K, L**) Expression comparisons of *RGL2* in different M stages (**K**) and N stages (**L**) in TNM system. The TNM system is the most widely used staging system. The higher the number after the letters, the more deteriorating the cancer becomes. (**M**) Correlations of *RGL2* expressions and numbers of positive lymph nodes detected by HE staining. (**N**) Expression comparisons of *RGL2* in different T stages in TNM system. (**O**) Expression comparisons of *RGL2* in different stages of colon cancer in another staging system. (**P**) Expression comparisons of *RGL2* in tissues with or without venous invasion. *** P-value < 0.001, ** P-value < 0.01, * P-value < 0.05 by *Students' t-test*.

Acknowledgments

This work was funded by grants from the National Key Scientific Program of China (2016YFA0100502 to X.S.) and the Hefei National Laboratory for Physical Sciences at the Microscale (KF2020011 to Y. W.).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.006.

References

[1] Haggar F, Boushey R. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clinics Colon Rectal Surgery 2009;22(04):191–7.
[2] Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. Int J Mol Sci 2017;18(1):197.
[3] Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of colorectal cancer risk factors. Cancer Causes Control 2013;24 (6):1207–22.
[4] Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology 2008;135(4):1079–99.
[5] Knösel T, Petersen S, Schwabe H, Schlüns K, Stein U, Schlag P, et al. Incidence of chromosomal imbalances in advanced colorectal carcinomas and their metastases. Virchows Arch 2002;440(2):187–94.
[6] Watanabe T, Wu T-T, Catalano PJ, Ueki T, Satriano R, Haller DG, et al. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. N Engl J Med 2001;344(16):1196–206.
[7] Deng Y, Chan SS, Chang S. Telomere dysfunction and tumour suppression: the senescence connection. Nat Rev Cancer 2008;8(6):450–8.
[8] Wright WE, Shay JW. The two-stage mechanism controlling cellular senescence and immortalization. Exp Gerontol 1992;27(4):383–9.
[9] Kamijo T, Zindy F, Roussel MF, Quelle DE, Downing JR, Ashmun RA, et al. Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19 ARF. Cell 1997;91(5):649–59.
[10] Sage J, Miller AL, Pérez-Mancera PA, Wysocki JM, Jacks T. Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. Nature 2003;424(6945):223–8.
[11] Chin L, Artandi SE, Shen Q, Tam A, Lee S-L, Gottlieb GJ, et al. p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. Cell 1999;97(4):527–38.
[12] Cosme-Blanco W, Shen M-F, Lazar AJF, Pathak S, Lozano G, Multani AS, et al. Telomere dysfunction suppresses spontaneous tumorigenesis in vivo by initiating p53-dependent cellular senescence. EMBO Rep 2007;8(5):497–503.
[13] Krtolica A, Parrinello S, Lockett S, Desprez P-Y, Campisi J. Senescent fibroblasts promote epithelial cell growth and tumorigenesis: A link between cancer and aging. PNAS 2001;98(21):12072–7.
[14] Coppé J-P, Patil CK, Rodier F, Sun Yu, Muñoz DP, Goldstein J, et al. Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. PLoS Biol 2008;6(12):e301. https://doi.org/10.1371/journal.pbio.0060301.
[15] López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell 2013;153(6):1194–217.
[16] Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., ... & Jemal, A. (2020). Colorectal cancer statistics, 2020. CA: a cancer journal for clinicians.
[17] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12(4):357–60.
[18] Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protocols 2016;11(9):1650–67.
[19] Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33(3):290–5.
[20] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26(1):139–40.
[21] McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 2012;40(10):4288–97.
[22] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2019.
[23] Kolde R. Pheatmap: pretty heatmaps. R package version 1.0.12. 2019.
[24] Subhash S, Kanduri C. GeneSCF: a real-time based functional enrichment tool with support for multiple organisms. BMC Bioinf 2016;17(1):1–10.
[25] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
[26] Brunson J. Ggalluvial: layered grammar for alluvial plots. J Open Source Softw 2020;5(49):2017. https://doi.org/10.21105/joss10.21105/joss.02017.
[27] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25.
[28] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357–9.
[29] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.
[30] Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, ... & Liu XS. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucl Acids Res 47(D1), D729-D735.
[31] Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, ... & Liu T. (2016). Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic acids research, gkw983.
[32] Li D, Hsu S, Purushotham D, Sears RL, Wang T. WashU epigenome browser update 2019. Nucleic Acids Res 2019;47(W1):W158–65.
[33] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38(4):576–89.
[34] Therneau T. A package for survival analysis in S. version 2.38. 2015.
[35] Kassambara A, Kosinski M, Biecek P. survminer: drawing survival curves using'ggplot2'. R package version 0.4.6. 2019.
[36] Liu NQ, ter Huurne M, Nguyen LN, Peng T, Wang S-Y, Studd JB, et al. The non-coding variant rs1800734 enhances DCLK3 expression through long-range interaction and promotes colorectal cancer progression. Nat Commun 2017;8 (1). https://doi.org/10.1038/ncomms14418.
[37] Sobral LM, Hicks HM, Parrish JK, McCann TS, Hsieh J, Goodspeed A, et al. KDM3A/Ets1 epigenetic axis contributes to PAX3/FOXO1-driven and independent disease-promoting gene expression in fusion-positive Rhabdomyosarcoma. Mol Oncol 2020;14(10):2471–86.
[38] Zhou L, Wang Y, Zhou M, Zhang Y, Wang P, Li X, et al. HOXA9 inhibits HIF-1α-mediated glycolysis through interacting with CRIP2 to repress cutaneous squamous cell carcinoma development. Nat Commun 2018;9(1). https://doi.org/10.1038/s41467-018-03914-5.
[39] Cheng Q, Ouyang X, Zhang R, Zhu L, Song X. Senescence-associated genes and non-coding RNAs function in pancreatic cancer progression. RNA Biol 2020;17 (11):1693–706.
[40] Ha T-K, Chi S-G. CAV1/caveolin 1 enhances aerobic glycolysis in colon cancer cells via activation of SLC2A3/GLUT3 transcription. Autophagy 2012;8 (11):1684–5.
[41] Kim E, Jung S, Park WS, Lee J-H, Shin R, Heo SC, et al. Upregulation of SLC2A3 gene and prognosis in colorectal carcinoma: analysis of TCGA data. BMC cancer 2019;19(1). https://doi.org/10.1186/s12885-019-5475-x.
[42] Li J, Fu F, Wan X, Huang S, Wu D, Li Y. Up-regulated miR-29c inhibits cell proliferation and glycolysis by inhibiting SLC2A3 expression in prostate cancer. Gene 2018;665:26–34.
[43] Chang JC. (2016). Cancer stem cells: Role in tumor growth, recurrence, metastasis, and treatment resistance. Medicine, 95(Suppl 1).
[44] Chen Ke, Huang Y-H, Chen J-L. Understanding and targeting cancer stem cells: therapeutic implications and challenges. Acta Pharmacol Sin 2013;34 (6):732–40.
[45] Prasad S, Ramachandran S, Gupta N, Kaushik I, Srivastava SK. Cancer cells stemness: a doorstep to targeted therapy. Biochim Biophys Acta (BBA)-Mol Basis Dis 2020;1866(4):165424. https://doi.org/10.1016/j.bbadis.2019.02.019.
[46] Fu Z, Chen C, Zhou Q, Wang Y, Zhao Y, Zhao X, et al. LncRNA HOTTIP modulates cancer stem cell properties in human pancreatic cancer by regulating HOXA9. Cancer Lett 2017;410:68–81.
[47] Elcheva IA, Wood T, Chiarolanzio K, Chim B, Wong M, Singh V, et al. RNA-binding protein IGF2BP1 maintains leukemia stem cell properties by regulating HOXB4, MYB, and ALDH1A1. Leukemia 2020;34(5):1354–63.
[48] Rothhammer T, Hahne JC, Florin A, Poser I, Soncin F, Wernert N, et al. The Ets-1 transcription factor is involved in the development and invasion of malignant melanoma. Cell Mol Life Sci CMLS 2004;61(1):118–28.
[49] Nakayama T, Ito M, Ohtsuru A, Naito S, Sekine I. Expression of the ets-1 proto-oncogene in human colorectal carcinoma. Mod Pathol 2001;14(5):415–22.
[50] Span PN, Manders P, Heuvel JJ, Thomas CMG, Bosch RR, Beex LVAM, et al. Expression of the transcription factor Ets-1 is an independent prognostic marker for relapse-free survival in breast cancer. Oncogene 2002;21 (55):8506–9.

[51] A Fry E, Inoue K. Aberrant expression of ETS1 and ETS2 proteins in cancer. Cancer Rep Rev 2018;2(3). https://doi.org/10.15761/CRR10.15761/CRR.1000151.

[52] Kim G-C, Lee C-G, Verma R, Rudra D, Kim T, Kang K, et al. ETS1 suppresses tumorigenesis of human breast cancer via trans-activation of canonical tumor suppressor genes. Front Oncol 2020;10. https://doi.org/10.3389/fonc.2020.00642.

[53] Subramanian A, Tamayo P, Mootha V K, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 2005;102(43):15545–50.

[54] Mootha V K, Lindgren C M, Eriksson K F, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. genet. 2003;34(3):267–73.