

RESEARCH

Open Access



Core promoter mutation contributes to abnormal gene expression in bladder cancer

Teng Huang, Jiaheng Li and San Ming Wang*

Abstract

Background: Bladder cancer is one of the most mortal cancers. Bladder cancer has distinct gene expression signature, highlighting altered gene expression plays important roles in bladder cancer etiology. However, the mechanism for how the regulatory disorder causes the altered expression in bladder cancer remains elusive. Core promoter controls transcriptional initiation. We hypothesized that mutation in core promoter abnormality could cause abnormal transcriptional initiation thereby the altered gene expression in bladder cancer.

Methods: In this study, we performed a genome-wide characterization of core promoter mutation in 77 Spanish bladder cancer cases.

Results: We identified 69 recurrent somatic mutations in 61 core promoters of 62 genes and 28 recurrent germline mutations in 20 core promoters of 21 genes, including *TERT*, the only gene known with core promoter mutation in bladder cancer, and many oncogenes and tumor suppressors. From the RNA-seq data from bladder cancer, we observed altered expression of the core promoter-mutated genes. We further validated the effects of core promoter mutation on gene expression by using luciferase reporter gene assay. We also identified potential drugs targeting the core promoter-mutated genes.

Conclusions: Data from our study highlights that core promoter mutation contributes to bladder cancer development through altering gene expression.

Keywords: Bladder cancer, Core promoter, Gene expression, Mutation

Background

Bladder cancer is the tenth most common cancer worldwide with an estimated 200,000 deaths per year [1]. Incidence rate of bladder cancer is the highest in Europe, especially in Southern European countries including Spain [1]. Urothelial cancer is the most common histologic type of bladder cancer accounting for 90% of all bladder cancers [2]. While environmental contaminants and smoking are known to be the risk factors for bladder cancer [3], knowledge about genetic factor contributing to bladder cancer is limited although altered expression

for the genes related to cell cycle, transcription and cytoskeleton was well observed in bladder cancer [4]; mutation altering *TERT* expression was identified in bladder cancer [5]; and differential gene expression was used to classify bladder cancer into sub-groups [6], the mechanisms of the abnormal gene expression in bladder cancer remains largely elusive.

Gene expression is under precise regulation to ensure spatial and temporal expression, in which transcriptional initiation is the gateway [7, 8]. In eukaryotes, transcriptional initiation is controlled by the basal transcriptional machinery composed of cis- and trans-elements in the core promoter region surrounding the transcriptional start site (TSS) [9]. The cis-elements consist of TFIIB recognition element (BRE), TATA box, Initiator element (Inr), downstream promoter element (DPE) etc. and their

*Correspondence: sanmingwang@um.edu.mo
Cancer Center and Institute of Translational Medicine, Faculty of Health Sciences, Ministry of Education Frontiers Science Center for Precision Oncology, University of Macau, Taipa, Macau



flanking sequences, and the trans-elements consist of RNA polymerase II, TFIIB and TFIID etc. and co-activators [8]. Mutation in cis sequences can interfere cis-trans interaction, modulate transcriptional initiation and gene expression level, and cause pathogenic consequences [5, 10, 11]. This is best exemplified by the core promoter mutation in *TERT*. *TERT* codes for telomerase reverse transcriptase involving in telomere structure. Mutation in *TERT* core promoter creates an ETS binding site and causes *TERT* over expression in multiple types of cancer including bladder cancer [5, 11, 12]. Regardless of the importance of core promoter in controlling gene expression, however, *TERT* remains as the only gene with established relationship between core promoter mutation and cancer. The prevalence of cis-mutation in core promoters remains largely unexplored in most cancer types including bladder cancer.

We hypothesized that core promoter mutation contributes to the abnormal gene expression in bladder cancer. Previously, we developed the Exome-based Variant Detection in Core-promoters (EVDC) method [13] for genome-wide core promoter mutation study, and used it in mapping the core promoter polymorphism in global human populations [14]. In this study, we applied this method to systematically analyze core promoter mutation in bladder cancer by using the exome data from bladder cancer patients. We identified both somatic and germline core promoter mutations in multiple genes and validated their effects on altering gene expression. Our study reveals that core promoter mutation can contribute to the etiology of bladder cancer.

Methods

Sources of sequence data

Exome data from Spanish bladder cancer ($n = 77$) and patient-matched blood [15] were from the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>, SRP029936 and SRP029935). Sequences in SRA format were converted into FASTQ format by using NCBI SRA Toolkit utility (version 2.9.1) [16]. Variants called from exome data of the Iberian population in Spain (IBS) ($n = 107$) sequenced by the 1000 Genome Project [17] were used as the normal population control in the study. Human genome reference sequences were used as the references for core promoter mapping analysis [18, 19] (hg19, <https://hgdownload.soe.ucsc.edu/downloads.html#human>).

Identification of core promoter mutations

Core promoter sequences were collected from the exome sequences by using the EVDC method [13]. Core promoter coordinates and sequences from hg19 were extracted by using BEDTools utility (version 2.27.1) [20].

BWA utility (version 0.7.17) was used to map exome sequences to hg19 [21]. The resulting SAM files were converted into BAM files and sorted by using SAMtools utility (version 1.9) [22, 23]. Duplicates were removed by using Picard tools (version 2.18.25), and the read group information was added [24]. The BAM files were further processed by using GATK (version 4.1.1.0) [24] with its recommended best practices pipeline. The called mutation files were compressed and indexed by using BCFtools utility (version 1.9) [22, 23], and annotated by using ANNOVAR [25]. Normal polymorphism in cancer samples were removed by filtering the normal population variation data including the Iberian population in Spain (IBS) sequenced by the 1000 Genome Project. Mutations with MAF value > 0.01 were also eliminated [26, 27]. Mutations absent in annotation data sets (dbSNP, 1000 Genome, ESP6500, ExAC, gnomAD, COSMIC, ClinVar) were classified as novel variants. The variants present in at least two cases were regarded as recurrent variants and used for further analysis. Somatic and germline mutations were distinguished by comparing the mutations from the tumor and the paired blood samples [15]. Examples of sequencing chromatograms were displayed by Tablet software [28].

Gene expression in human tissues

RNA-seq data of bladder cancer and adjacent normal tissues generated by TCGA [29] were from the cBioPortal database [30] for differential gene expression analysis (https://cbioportal-datahub.s3.amazonaws.com/blca_tcga_pan_can_atlas_2018.tar.gz). Differentially expressed genes were identified by using Student's t-test and fold changes. Gene identifiers were converted by using SynGO [31]. Volcano plots showing differential expressed genes were generated by using R ggplot2 package [32]. The expression for the luciferase reporter assayed genes in human tissues were searched in Human Protein Atlas [33].

Luciferase reporter assay

Human embryonic kidney 293 cells (HEK 293) were used to test the effects of core promoter mutation in gene expression using the dual-luciferase reporter system. Cells were grown in Dulbecco's modified Eagle's media/Nutrient Mixture culture medium with 10% fetal bovine serum, 100 IU/ml penicillin and 100 IU/ml streptomycin sulfate. The wild-type and mutated core promoter sequences were synthesized, cloned into pGL3 luciferase reporter vector, and validated by Sanger sequencing (BGI TECH SOLUTIONS, Beijing, China). Fifty microgram of pGL3 containing the targeted core promoter sequences and 5 μ g of control pRL *Renilla* luciferase reporter vector were mixed, and co-transfected

into HEK 293 cells by using Lipofectamine 3000 Transfection Reagent (Thermo Fisher SCIENTIFIC, MS, USA). Forty-eight hours after the transfection, cells were harvested to measure luciferase activity by using the Dual-Luciferase Reporter Assay System (Promega, WI, USA) following the instruction (PerkinElmer Victor X3 Microplate Reader, OH, USA). Three independent tests were performed for each core promoter. Luciferase activity was normalized by dividing *firefly* luciferase activity with *Renilla* luciferase activity:

$$E_l = E_f/E_r$$

E_f : *firefly* luciferase activity, E_r : *Renilla* luciferase activity, E_l : normalized luciferase activity.

Characterization of core promoter mutation-affected genes

For the core promoter mutated genes, their function categories and involved pathways were analyzed by using GO (Gene Ontology) knowledgebase [34] and GeneCards database [35]. Candidate drugs targeting the core promoter mutated genes were identified in DrugBank [36]. GO terms and drugs were identified by using Metascape [37]. Expression Quantitative Trait Loci in PanCanQTL database [38] was used to test the effects of the core promoter-mutated genes on gene expression in bladder tissue. A cancer driver gene panel was generated by integrating the 1064 cancer driver genes in OncoKB [39] database and the 299 genes from previous cancer driver gene study [40], and the core promoter mutated genes were searched in this gene panel to identify potential driver genes with core promoter mutation. KEGG (Kyoto Encyclopedia of Genes and Genomes) database [41] was used to identify the pathways affected by the mutated driver genes.

Statistics analysis

In the analysis of differential gene expression and dual-luciferase reporter assay, p -value < 0.05 by using Student's t -test, and fold changes ≥ 1.5 were considered as significantly different. Student's t -test were calculated by using T.TEST function in MS EXCEL. In the enrichment analysis, p -value < 0.05 by using the accumulative hypergeometric test, overlap ≥ 1 and enrichment factor > 1.5 were considered as significantly different. Statistics test in enrichment analysis was calculated by Metascape.

Results

Core promoter mutation in bladder cancer

We collected the core promoter sequences from a total of 77 tumor samples and matched blood samples from the exome data generated by the Spanish urothelial bladder

cancer study [15]. We called variants from the collected core promoter sequences (Fig. S1), removed polymorphic variants through filtering the variants from normal human population including the IBS population, and identified somatic and germline mutations by comparing the variants between cancer and blood samples. Figure 1 outlines the analytic process of the study.

We identified a total of 216 recurrent somatic mutations (present in ≥ 2 carriers), 3 mutations per cancer case on average, composed of 69 distinct mutations in 61 core promoters of 62 genes (Table 1A, Table S1A and Table S2A, B). Of the 69 somatic mutations, 45 (65.2%) were substitution, 14 (20.3%) were deletion and 10 (14.5%) were insertion (Table 1B); 63 (91.3%) were absent in the COSMIC database and 37 (53.6%) were novel and absent in all mutation databases; and 8 (11.6%) were located at simple repetitive sequences.

We also identified a total of 88 recurrent germline mutations, 1 mutation per cancer case on average, composed of 28 distinct mutations in 20 core promoters of 21 genes (Table 1A, Table S1B and Table S2C, D). Of the 28 germline mutations, 18 (64.3%) were substitution, 7 (25%) were deletion and 3 (10.7%) were insertion (Table 1B); 15 (53.6%) were novel; and 9 (32.1%) were located at simple repetitive sequences.

We observed that the core promoter mutations were enriched in multiple core promoter motifs (Table 1C and Table S3). For example, MTE box2 motif had 23 somatic mutations and 3 germline mutations. Reflecting the fact

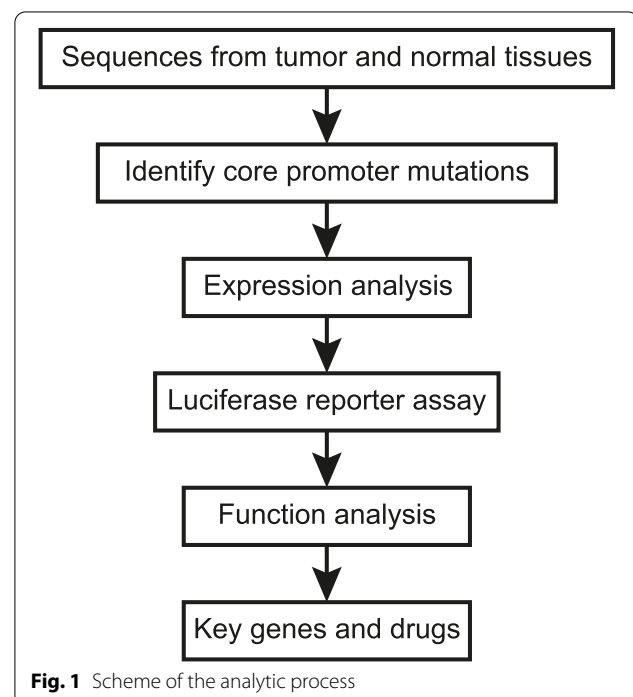


Fig. 1 Scheme of the analytic process

Table 1 Summary of core promoter mutations identified in bladder cancer

Items	Core promoter variants	
	Somatic	Germline
A. General features		
Total	216	88
Average number of mutation/case	3	1
Distinct	69	28
Co-promoter with variants	61	20
Gene affected	62	21
Absent in COSMIC database	63	28
Novel	37	15
Non-repetitive	61	19
Repetitive	8	9
B. Type		
Total	69	28
Substitution	45	18
Insertion	10	3
Deletion	14	7
C. Mutation located in core promoter motifs		
Total ^a	86	21
MTE_box2	23	3
DPE	10	4
Inr	9	2
Ets	9	–
DTIE	6	1
TCT	4	1
BREu	3	–
TATA box	2	1

^a Some mutations affected > 1 motif

that TATA box is not tolerable for base changes [13], only 2 somatic and 1 germline mutations were located at the TATA box. This also served as an internal control in validating the reliability of the mutations identified in the bladder cancer from this study.

Effects of core promoter variation on gene expression

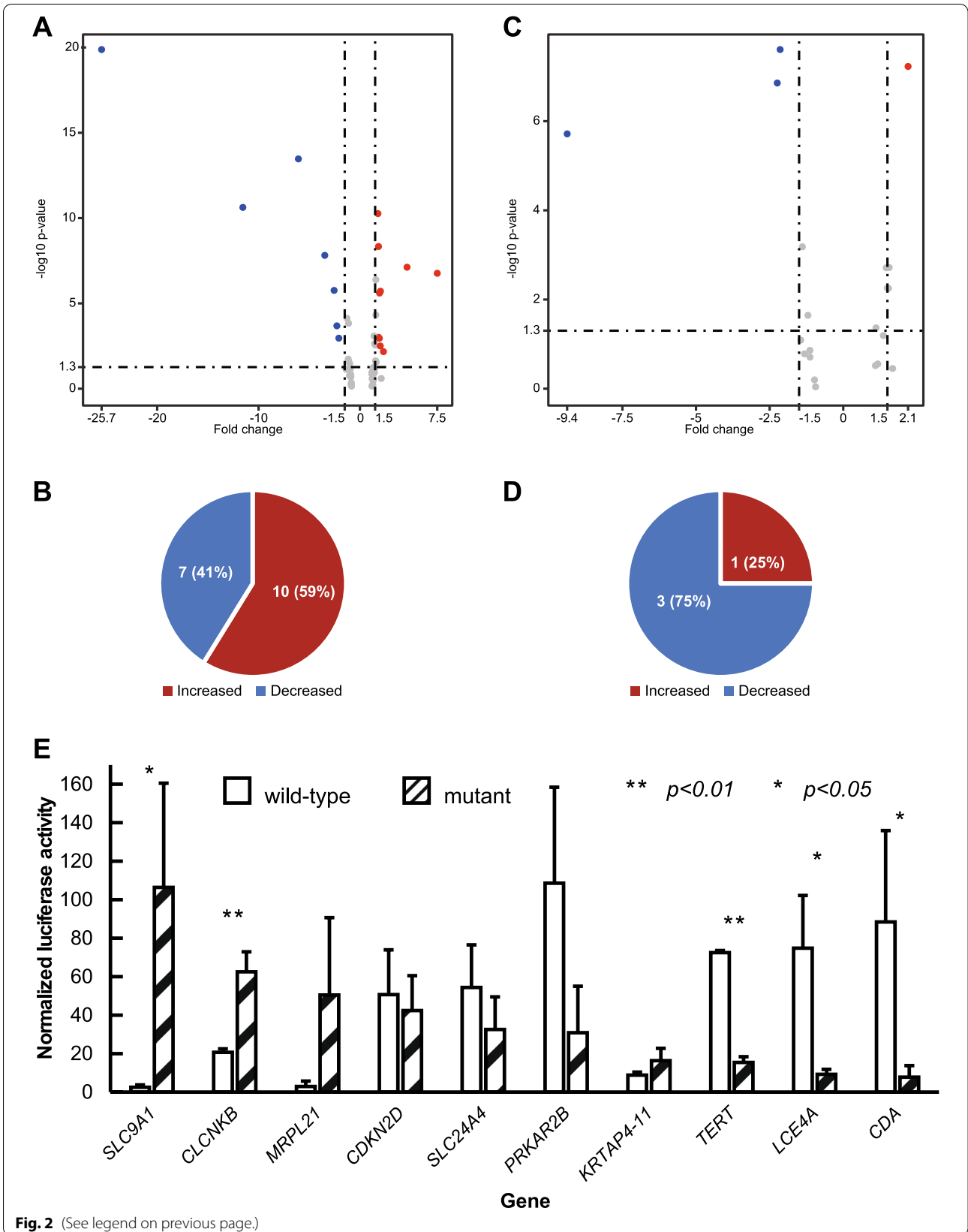
To address if core promoter mutation could lead to altered expression of the core promoter-mutated genes,

we compared the RNA-seq data between bladder cancer and adjacent normal samples. Of the core promoter somatically mutated 62 genes, 17 (27.4%) were significantly different including 10 increased and 7 decreased expressions. Of the 17 genes, *TERT* had the highest of 7.5-fold increased expression and *CFD* had the highest of 25.7-fold decreased expression. Of the core promoter germline-mutated 21 genes, 4 (19.0%) were significantly different including 1 increased and 3 decreased expressions (Fig. 2A-D and Table S4). We also searched the Human Protein Atlas database to collect the expression information for the core promoter mutation-affected genes in normal and bladder cancer (Table S5A). The result showed that *TERT* was not expressed in normal bladder but overexpressed in bladder cancer with core promoter C228T mutation *TERT* [5]; survival data of *CDA*, *SLC9A1* and *SLC24A4* also showed that their expression levels were associated with 5-year survival significantly.

While the data from the RNAseq data analysis provided evidence for the impact of the core promoter mutation on expression, the information was indirect as the genes in the original samples could not be sure to contain the core promoter mutations except *TERT*. Therefore, we used reporter gene assay to test the effects of core promoter mutation in gene expression. Based on the considerations 1) the functional importance of the genes carrying the mutation, 2) significance of the altered expression level by expression data analysis, and 3) core promoter sequence features for designing and constructing the mutants, we selected 10 core promoters for the test, including *TERT*, *CDA*, *SLC9A1*, *SLC24A4*, *PRKAR2B*, *CDKN2D*, *CLCNKB*, *LCE4A*, *KRTAP4–11* and *MRPL21*. The canonical core promoter mutation in *TERT* was selected as internal standard. *CDA* involves in metabolic process, *SLC9A1* is related with cancer growth, *SLC24A4* had decreased expression in bladder cancer. *PRKAR2B* is involved in mitotic cell cycle transition and response to cancer-related drug clozapine. *CDKN2D* is involved in cell cycle, metabolic process, and nutrient response. *CLCNKB* regulates trans-membrane transport and trans-differentiation. *LCE4A* and *KRTAP4–11* are related with cellular differentiation. *MRPL21* is related

(See figure on next page.)

Fig. 2 Core promoter mutated genes with altered gene expression in bladder cancer. The volcano plots showed the altered expression of core promoter mutated genes between cancer and adjacent normal samples based on RNA-seq data. X-axis represented fold changes of increased or decreased expression, and Y-axis represented distribution of the genes with altered expression at -log10 scale. The pie charts displayed the number of gene with altered expression. **A.** altered expression of somatic core promoter mutated genes; **B.** somatic core promoter mutated genes with altered expression; **C.** altered expression of germline core promoter mutated genes; **D.** germline core promoter mutated genes with altered expression. **E.** luciferase activities with mutated core promoters. Luciferase activities in 10 mutated core promoters were compared with the corresponding wild-type core promoters. Three independent tests were performed for each core promoter. *refers to these with significant differences



to mitochondrion metabolism. Each mutated core promoter was paired with the corresponding wildtype core promoter control for the test. We generated the mutated core promoters for the 10 selected genes, cloned into luciferase reporter constructs. Each type of mutant construct was transfected into 293 cells, the luciferase activities were compared with the corresponding wild-type core promoter controls. Of the 10 mutated core promoters tested, 5 had significantly altered luciferase activities (*SLC9A1*, *CLCNKB*, *TERT*, *LCE4A* and *CDA*, p -value < 0.05), of which *SLC9A1* and *CLCNKB* had increased luciferase activities, *TERT*, *LCE4A* and *CDA* had decreased luciferase activities (Fig. 2E and Table S5B).

Cancer driver genes and pathways affected by core promoter mutation

By Gene Ontology analysis, we observed that the core promoter-mutated genes were enriched in the functional pathways highly relevant to oncogenesis (Fig. 3 and Table 2). For example, somatic mutated genes were enriched in “Regulation of mitotic cell cycle phase transition”, “Cellular response to peptide hormone stimulus” and “Selective advantage”; germline mutated genes were enriched in “Evading apoptosis”, “Evading the immune system”, “Tissue invasion and metastasis” and “DNA repair”; and both somatic and germline-mutated genes were enriched in “Deregulated metabolism”, “Differentiation” and “Sustained angiogenesis” (Table S6).

We compared the core promoter-mutated genes with altered gene expression to the cancer driver gene list and observed that the somatic-mutated *TERT* and *PRRX1*, and germline-mutated *GAB2* were on the list (Table S7). *TERT* is the only known gene with somatic core promoter mutation in bladder cancer [5]. *TERT* participates in the formation of telomeric DNA repeats and affects the immortality of cell (Fig. 3C). The C228T in the core promoter of *TERT* was detected in 3 bladder cancer cases but absent in all paired blood samples, and no coding mutation in *TERT* was detected. The mutation generated a new binding motif of the Ets transcription factor, deleted a DTIE and created a new putative DCE_box1, and caused 4.8-fold decreased expression as shown by luciferase report gene assay (Fig. 2E and Table S5B). *PRRX1* is a transcription co-activator enhancing DNA-binding activity of SRF (serum response factor) required

for the induction of multiple genes by growth and differentiation factors. A CT-track simple repetitive sequence was inserted into the CT-repeat region in the core promoter, caused decreased *PRRX1* expression in bladder cancer. *GAB2* involves in immune-response and apoptosis (Fig. 3D). A germline A > C mutation at -60 altered the sequence from “CCCACC” to “CCCCC”, caused decreased expression in bladder cancer as shown by RNA-seq data (Table S7).

Potential drugs targeting core promoter mutated genes

The core promoter mutation-affected genes with altered expression provide potential drug targets for bladder cancer treatment [42]. From the DrugBank, we identified 6 drugs/compounds targeting 3 somatic-mutated genes with altered expression (1 increased and 2 decreased expression) (Table S8). For example, an approved drug Zidovudine targets *TERT* by inhibiting telomerase activity [43]; cyclic adenosine monophosphate (cAMP) targets *PRKAR2B*, which is a regulatory subunit of the cAMP-dependent protein kinases. We also identified 10 drugs/compounds targeting germline-mutated *PDE10A* (LINC00473) with decreased expression (Table S8), including Dipyridamole acting as a phosphodiesterase inhibitor to suppress *PDE10A* activity [44] and Triflusal, an antagonist to *PDE10A* [45].

Discussion

Core promoter variation is well recognized in affecting gene expression. However, the role of core promoter mutation in oncogenesis has not been well established. With its distinct gene expression in bladder cancer, our study identified both somatic and germline mutations in the core promoters of a group of cancer-related genes. Our study highlights that core promoter mutation can be an important etiological factor in bladder cancer oncogenesis through altering the expression of cancer genes.

In our study, the somatic and germline mutations were identified by 1) Comparing the cancer samples with their paired blood samples from the same study; 2) Filtering the called variant data by variants from normal human populations including these from the local population to eliminating normal polymorphism; 3) Using the exome data and RNA-seq data from urothelial cancer for mutation and expression analysis; 4) Comparing altered

(See figure on next page.)

Fig. 3 GO classification and KEGG pathways of core promoter-mutated genes. **A.** GO classification of somatic core promoter-mutated genes; **B.** GO classification of germline core promoter-mutated genes; **C.** KEGG pathway of *TERT* involved in cancer (<https://www.kegg.jp/pathway/map05200>). The C228T in the core promoter of *TERT* generated a new Ets binding motif, altered *TERT* expression, and promoted cellular immortality. **D.** KEGG pathway of *GAB2* involved in cancer (<https://www.kegg.jp/pathway/ko05220>). *GAB2* involves in MAPK and PI3K-Akt signal pathways in immune-response and apoptosis. A germline A > C mutation at -60 altered the sequence from “CCCACC” to “CCCCC”, caused decreased *GAB2* expression in bladder cancer (Table S7). Black bar: statistical significance of gene group; white bar: number of genes enriched in the group; full arrow: direct effects; dotted line arrow: indirect effects

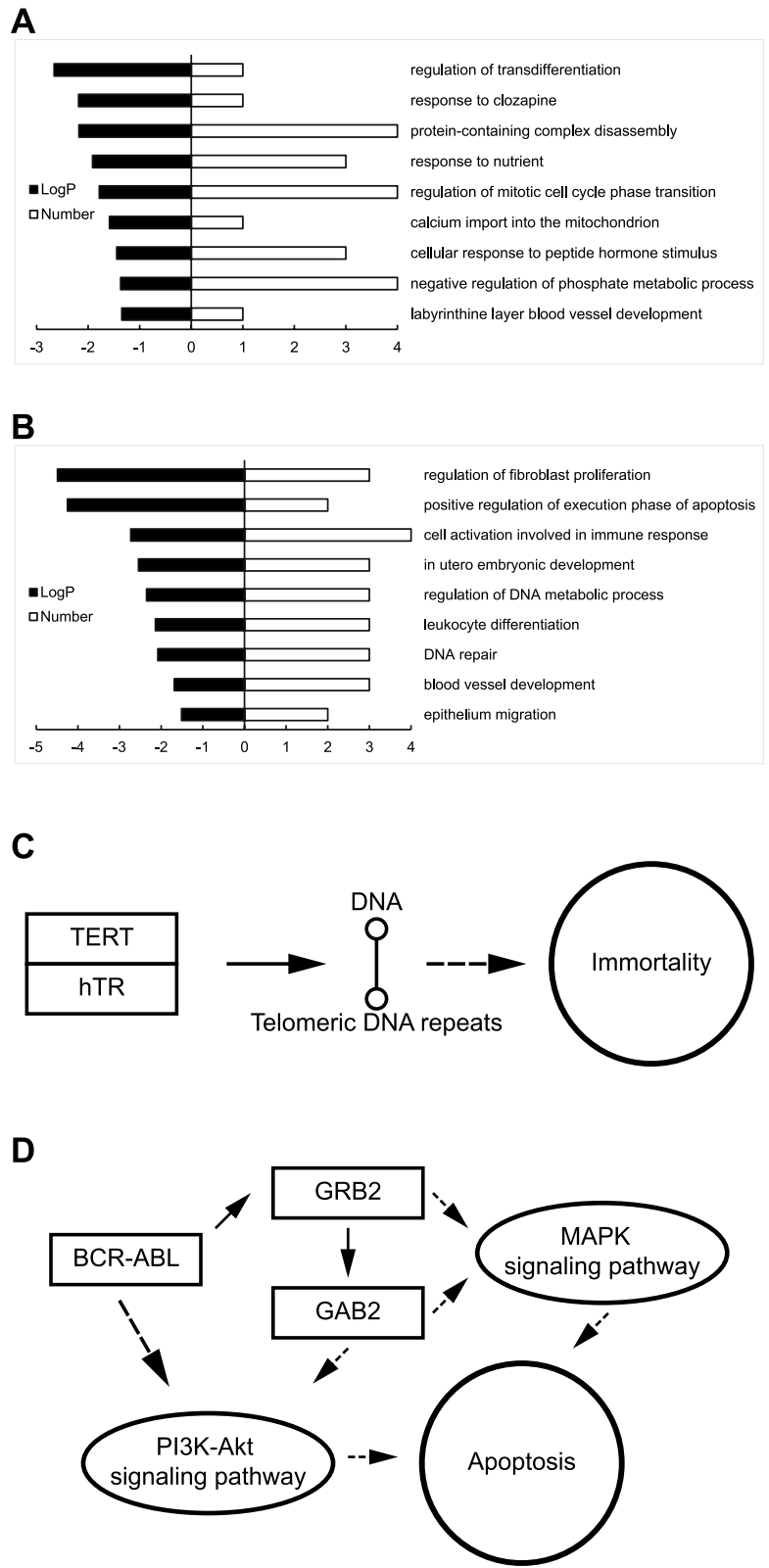


Fig. 3 (See legend on previous page.)

Table 2 Examples of functional important genes with core promoter mutation

Items	Mutation	Co-promoter position	#Carrier	Expression
A. Pathways with core promoter mutated genes				
Deregulated metabolism	Somatic Germline			
Differentiation	Somatic Germline			
Sustained angiogenesis	Somatic Germline			
Regulation of mitotic cell cycle phase transition	Somatic			
Cellular response to peptide hormone stimulus	Somatic			
Selective advantage	Somatic			
Evading apoptosis	Germline			
Evading the immune system	Germline			
Tissue invasion and metastasis	Germline			
DNA repair	Germline			
B. Examples of cancer related genes				
<i>TERT</i>	Somatic	-66	3	+7.5
<i>PRKAR2B</i>	Somatic	93	2	-6.2
<i>SMUG1</i>	Germline	90	2	+2.1
<i>GAB2</i>	Germline	-60	3	-2.3

expression between cancer and adjacent normal samples. These steps ensured high reliability of the mutations identified by our study, as exemplified by the identification of core promoter mutation in *TERT*, which is known to be present in bladder cancer [11]. It is interesting to note that the core promoter-mutated *TERT* causes its increased expression in multiple types of cancer [5, 11], including in our expression analysis (Table S4). However, in core promoter mutated *TERT*-luciferase reporter assay, the mutation caused decreased luciferase expression (Table S5B). This could be related to the differences of cell types, in vitro and in vivo conditions, etc., which may have had different regulation mechanisms of transcription initiation [46]. As a widely reported oncogene with core promoter mutation, the opposite effects of the mutated *TERT* core promoter on gene expression is worth of further study. In *TP53* core promoter, we found a germline mutation C>T at +101 and a poly T track deletion at +95, but no expression change was observed between cancer and control as shown by RNA-seq data analysis.

Our study identified multiple novel core promoter mutated genes. For example, somatic mutations were identified in the core promoter of *PRKAR2B*, and germline mutations were identified in the core promoter of *SMUG1* and *GAB2*. Gene ontological and pathway analysis showed that these core promoter mutated genes are oncogenic through affecting multiple functional pathways: *SMUG1* participates in DNA repair (KEGG: hsa03410); *GAB2* contributes to cellular differentiation, immunity and cancer (KEGG: ko05220); *PRKAR2B* regulates mitotic cell cycle transition and metabolism

(KEGG: hsa04910). Simple repetitive sequence is widely present in promoter, and plays important role in gene expression regulation [47]. The core promoter mutation in *GAB2* and *PRRX1* occurred at simple repetitive sequences, caused their altered expression in cancer. It is interesting to notice that both somatically mutated *PRKAR2B* and germline-mutated *GAB2* were present in a single bladder cancer case (BioSample accession number: SAMN02351138). Somatic mutation in *PRKAR2B* created putative motifs in the core promoter, caused *PRKAR2B* differentially expressed, affected regulation of mitotic cell cycle transition and phosphate metabolism [48]. *GAB2* is a cancer driver gene. The high frequent germline mutation in *GAB2* was also present in acute myeloid leukemia in the International Cancer Genome Consortium study and in acute lymphoblastic leukemia with *Ras*-independent leukemogenic effects [49]. Drug targeting the core promoter-mutated gene offers a potential pharmacological therapeutic agent for bladder cancer treatment and worthy to be studied further.

Conclusions

Our study identified both somatic and germline mutations in core promoters of multiple cancer driver genes in bladder cancer, highlighting that altered regulatory machinery including the core promoter can contribute to the alterative gene expression in cancer.

Abbreviations

BRE: TFIIIB recognition element; DPE: Downstream promoter element; EVDC: Exome-based Variant Detection in Core-promoters; FPKM: Number Fragments Per Kilobase of exon per Million reads; GO: Gene Ontology; HEK 293: Human embryonic kidney 293 cells; IBS: Iberian population in Spain; Inr: Initiator

element; KEGG: Kyoto Encyclopedia of Genes and Genomes; pTPM: Transcripts per million protein coding genes; SRA: Sequence Read Archive; SRF: Serum response factor; TSS: Transcriptional start site.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-022-09178-z>.

Additional file 1: Fig S1. Sequence chromatograms of three core promoter mutations. A. Mutation T > C/TC > CA (chr1:152,681,543-152,681,544) in core promoter of *LCE4A* occurred in 79 out of 80 reads in a sample. B. Mutation A > G/AG > GA (chr1:20,915,531-20,915,532) in core promoter of *CDA* occurred in 28 out of 68 reads in a sample. C. Mutation C > G (chr11:75,110,552-75,110,552) in core promoter of *RPS3* occurred in 17 out of 77 reads in a sample. Top line: reference sequences; other lines: sequence reads mapped to the reference sequences; base marked in red: the base different from the reference sequences; arrow: the mutated base identified by sequence alignment.

Additional file 2: Table S1. A. List of somatic non-repetitive core promoter mutations. B. List of germline non-repetitive core promoter mutations.

Additional file 3: Table S2. A. List of somatic core promoter mutations in simple repetitive sequences. B. Type of somatic core promoter mutations in simple repetitive sequences. C. List of germline core promoter mutations in simple repetitive sequences. D. Type of germline core promoter mutations in simple repetitive sequences.

Additional file 4: Table S3. Frequency of core promoter mutations in motifs.

Additional file 5: Table S4. Core promoter mutated genes with altered gene expression.

Additional file 6: Table S5. A. Expression for the luciferase reporter assay-tested genes in human tissues. B. Expression of core promoter mutated genes in luciferase reporter assay.

Additional file 7: Table S6. A. GO classification of somatic core promoter-mutated genes. B. GO classification of germline core promoter-mutated genes.

Additional file 8: Table S7. Driver gene with core promoter mutation.

Additional file 9: Table S8. Drugs targeting core promoter-mutated genes.

Acknowledgements

We are thankful for the Information and Communication Technology Office, University of Macau for providing the High-Performance Computing Cluster resource and facilities for the study.

Code availability

Public websites and code packages used in this study were described in the methods and no customized statistical code was programmed.

Authors' contributions

TH data collection, analysis, curation, interpretation, manuscript writing; JL: luciferase reporter assay, data analysis, interpretation, manuscript writing; SMW: conceptualization, experimental design, data interpretation, manuscript writing, revision, and funding. The author(s) read and approved the final manuscript.

Funding

This work was supported by grants from the Macau Science and Technology Development Fund (085/2017/A2, 0077/2019/AMJ), the University of Macau (SRG2017-00097-FHS, MYRG2019-00018-FHS), the Faculty of Health Sciences, University of Macau (FHSIG/SW/0007/2020P and a startup fund) to SMW.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 19 September 2021 Accepted: 6 January 2022

Published online: 15 January 2022

References

- 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(6):394–424. <https://doi.org/10.3322/caac.21492>.
- Chalasan V, Chin JL, Izawa JL. Histologic variants of urothelial bladder cancer and nonurothelial histology in bladder cancer. *Can Urol Assoc J*. 2009;3(6 Suppl 4):S193–8. <https://doi.org/10.5489/auaj.1195>.
- Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association between smoking and risk of bladder cancer among men and women. *JAMA*. 2011;306(7):737–45. <https://doi.org/10.1001/jama.2011.1142>.
- Aaboe M, Marcussen N, Jensen KM, Thykjaer T, Dyrskjot L, Orntoft TF. Gene expression profiling of noninvasive primary urothelial tumours using microarrays. *Br J Cancer*. 2005;93(10):1182–90. <https://doi.org/10.1038/sj.bjc.6602813>.
- Li C, Wu S, Wang H, Bi X, Yang Z, Du Y, et al. The C228T mutation of TERT promoter frequently occurs in bladder cancer stem cells and contributes to tumorigenesis of bladder cancer. *Oncotarget*. 2015;6(23):19542–51. <https://doi.org/10.18632/oncotarget.4295>.
- Hedegaard J, Lamy P, Nordentoft I, Algaba F, Hoyer S, Ulhøi BP, et al. Comprehensive transcriptional analysis of early-stage Urothelial carcinoma. *Cancer Cell*. 2016;30(1):27–42. <https://doi.org/10.1016/j.ccell.2016.05.004>.
- Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res*. 2013;23(1):169–80. <https://doi.org/10.1101/gr.139618.112>.
- Vo Ngoc L, Wang YL, Kassavetis GA, Kadonaga JT. The punctilious RNA polymerase II core promoter. *Genes Dev*. 2017;31(13):1289–301. <https://doi.org/10.1101/gad.303149.117>.
- Sato MP, Makino T, Kawata M. Natural selection in a population of *Drosophila melanogaster* explained by changes in gene expression caused by sequence variation in core promoter regions. *BMC Evol Biol*. 2016;16:35. <https://doi.org/10.1186/s12862-016-0606-3>.
- Sharma A, Jiang C, De S. Dissecting the sources of gene expression variation in a pan-cancer analysis identifies novel regulatory mutations. *Nucleic Acids Res*. 2018;46(9):4370–81. <https://doi.org/10.1093/nar/gky271>.
- Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. *Science*. 2013;339(6122):957–9. <https://doi.org/10.1126/science.1229259>.
- Fredriksson NJ, Ny L, Nilsson JA, Larsson E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat Genet*. 2014;46(12):1258–63. <https://doi.org/10.1038/ng.3141>.
- Kim YC, Cui J, Luo J, Xiao F, Downs B, Wang SM. Exome-based variant detection in Core promoters. *Sci Rep*. 2016;6:30716. <https://doi.org/10.1038/srep30716>.
- Gupta H, Chandratte K, Sinha S, Huang T, Wu X, Cui J, et al. Highly diversified core promoters in the human genome and their effects on gene

- expression and disease predisposition. *BMC Genomics*. 2020;21(1):842. <https://doi.org/10.1186/s12864-020-07222-5>.
15. Balbas-Martinez C, Sagraera A, Carrillo-de-Santa-Pau E, Earl J, Marquez M, Vazquez M, et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat Genet*. 2013;45(12):1464–9. <https://doi.org/10.1038/ng.2799>.
 16. Coordinators NR. Database resources of the National Center for biotechnology information. *Nucleic Acids Res*. 2016;44(D1):D7–19. <https://doi.org/10.1093/nar/gkv1290>.
 17. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74. <https://doi.org/10.1038/nature15393>.
 18. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921. <https://doi.org/10.1038/35057062>.
 19. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC table browser data retrieval tool. *Nucleic Acids Res*. 2004;32(Database issue):D493–6. <https://doi.org/10.1093/nar/gkh103>.
 20. Quinlan AR. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics*. 2014;47:11.12:11–34. <https://doi.org/10.1002/0471250953.bi1112547>.
 21. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
 22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome project data processing S. the sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
 23. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93. <https://doi.org/10.1093/bioinformatics/btr509>.
 24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
 25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. <https://doi.org/10.1093/nar/gkq603>.
 26. Fu Y, Jovelet C, Filleron T, Pedrero M, Motte N, Boursin Y, et al. Improving the performance of somatic mutation identification by recovering circulating tumor DNA mutations. *Cancer Res*. 2016;76(20):5954–61. <https://doi.org/10.1158/0008-5472.CAN-15-3457>.
 27. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in Cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn*. 2017;19(1):4–23. <https://doi.org/10.1016/j.jmoldx.2016.10.002>.
 28. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, et al. Using tablet for visual exploration of second-generation sequencing data. *Brief Bioinform*. 2013;14(2):193–202. <https://doi.org/10.1093/bib/bbs012>.
 29. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer genome atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113–20. <https://doi.org/10.1038/ng.2764>.
 30. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095>.
 31. Koopmans F, van Nierop P, Andres-Alonso M, Byrnes A, Cijssouw T, Coba MP, et al. SynGO: an evidence-based, expert-curated Knowledge Base for the synapse. *Neuron*. 2019;103(2):217–34 e214. <https://doi.org/10.1016/j.neuron.2019.05.002>.
 32. Villanueva RAM, Chen ZJ. ggplot2: elegant graphics for data analysis, 2nd edition. *Meas Interdiscip Res* 2019, 17(3):160–167. doi: <https://doi.org/10.1080/15366367.2019.1565254>.
 33. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419. <https://doi.org/10.1126/science.1260419>.
 34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–9. <https://doi.org/10.1038/75556>.
 35. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;54:1.30.1–1.30.33. <https://doi.org/10.1002/cpbi.5>.
 36. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074–82. <https://doi.org/10.1093/nar/gkx1037>.
 37. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. <https://doi.org/10.1038/s41467-019-09234-6>.
 38. Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, et al. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res*. 2018;46(D1):D971–6. <https://doi.org/10.1093/nar/gkx861>.
 39. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology Knowledge Base. *JCO Precis Oncol*. 2017;2017. <https://doi.org/10.1200/PO.17.00011>.
 40. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of Cancer driver genes and mutations. *Cell*. 2018;173(2):371–85 e318. <https://doi.org/10.1016/j.cell.2018.02.060>.
 41. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res*. 2021;49(D1):D545–51. <https://doi.org/10.1093/nar/gkaa970>.
 42. Liu B, Yan X, Hou Z, Zhang L, Zhang D. Impact of bupivacaine on malignant proliferation, apoptosis and autophagy of human colorectal cancer SW480 cells through regulating NF-kappaB signaling path. *Bioengineered*. 2021;12(1):2723–33. <https://doi.org/10.1080/21655979.2021.1937911>.
 43. Leeansyah E, Cameron PU, Solomon A, Tennakoon S, Velayudham P, Gouillou M, et al. Inhibition of telomerase activity by human immunodeficiency virus (HIV) nucleos(t)ide reverse transcriptase inhibitors: a potential factor contributing to HIV-associated accelerated aging. *J Infect Dis*. 2013;207(7):1157–65. <https://doi.org/10.1093/infdis/jit006>.
 44. Fujishige K, Kotera J, Michibata H, Yuasa K, Takebayashi S, Okumura K, et al. Cloning and characterization of a novel human phosphodiesterase that hydrolyzes both cAMP and cGMP (PDE10A). *J Biol Chem*. 1999;274(26):18438–45. <https://doi.org/10.1074/jbc.274.26.18438>.
 45. McNeely W, Goa KL. Triflusal. *Drugs*. 1998;55(6):823–33; discussion 834–825. <https://doi.org/10.2165/00003495-199855060-00011>.
 46. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 2018;19(10):621–37. <https://doi.org/10.1038/s41580-018-0028-8>.
 47. Huda A, Marino-Ramirez L, Landsman D, Jordan IK. Repetitive DNA elements, nucleosome binding and human gene expression. *Gene*. 2009;436(1–2):12–22. <https://doi.org/10.1016/j.gene.2009.01.013>.
 48. Sha J, Xue W, Dong B, Pan J, Wu X, Li D, et al. PRKAR2B plays an oncogenic role in the castration-resistant prostate cancer. *Oncotarget*. 2017;8(4):6114–29. <https://doi.org/10.18632/oncotarget.14044>.
 49. de Smith AJ, Lavoie G, Walsh KM, Aujla S, Evans E, Hansen HM, et al. Predisposing germline mutations in high hyperdiploid acute lymphoblastic leukemia in children. *Genes Chromosom Cancer*. 2019;58(10):723–30. <https://doi.org/10.1002/gcc.22765>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.