Studies on Transcriptomic Perturbation in Cancer Cells by Genome Wide Profiling

January 2020

Yusuke NAKAYAMA

Studies on Transcriptomic Perturbation in Cancer Cells by Genome Wide Profiling

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Biological Science
(Doctoral Program in Biological Sciences)

Yusuke NAKAYAMA

Table of Contents

Abstract

Comprehensive transcriptomic profiling has given us numerous new biological insights and become an essential approach in biological research. However, it is still challenging to find meaningful information from transcriptome data, because cells in response to a stimulation show extremely complex transcriptomic perturbation, especially in cancer cells. Therefore, various bioinformatics approaches are necessary in the research of transcriptomic perturbation. In this study, I present usefulness of multi-profiling approach to understand transcriptional perturbation and generate valuable biological hypothesis.

In the first study, I investigated quantitative perturbation of mRNA in cancer cells with centromere protein E (CENP-E) inhibition. CENP-E is one of the core molecules in chromosome alignment in mitosis and has high potential as therapeutic target for cancer with chromosomal instability (CIN). However, molecular mechanism of anti-proliferative effect by CENP-E inhibition is unclear. To determine which signaling pathways contribute to the postmitotic effect of CENP-E inhibition, I performed comprehensive gene expression analysis using microarray for cancer cells with siCENP-E in the presence of siBubR1. Quantitative analysis revealed that genes related to p53 signaling and DNA damage response (DDR) are upregulated by CENP-E inhibition. The study suggested that induction of apoptosis by CENP-E inhibition could be led by activation of p53 signal and DDR.

In the second study, I investigated structural perturbation of mRNA in cancer cells with alternative polyadenylation modulator. Alternative polyadenylation (APA) plays a critical role in regulating gene expression. However, the balance between genome-encoded APA processing and autoregulation by APA modulating RNA binding protein (RBP) factors is not well understood. I discovered two potent small-molecule modulators of APA (T4, T5) that promote distal to proximal (DtoP) APA usage in multiple transcripts. Monotonically responsive APA events, induced by short exposure to T4 or T5, were defined in the

transcriptome, allowing clear isolation of the genomic sequence features and RBP motifs associated with DtoP regulation. I found that longer vulnerable introns, enriched with distinctive A-rich motifs, were preferentially affected by DtoP APA, thus defining a core set of genes with genomically encoded DtoP regulation. Through APA response pattern and compound–siRNA epistasis analysis of APA-associated RBP factors, I further demonstrated that DtoP APA usage is partly modulated by altered autoregulation of polyadenylate binding nuclear protein-1 signaling.

Through the two studies, I demonstrated an effectiveness of genome-wide profiling approaches for elucidation of molecular mechanism of biological response in cells by generating practicable hypotheses. These approaches I presented here could be applicable for various biological scenarios, especially in oncology, which show complex transcriptomic perturbation.

Abbreviations

A3SS	alternative 3’ splice sites
A5SS	alternative 5′ splice sites
AFE	alternative first exons
ALE	alternative last exons
APA	alternative polyadenylation
BubR1	BUB1 mitotic checkpoint serine/threonine kinase B
CENP-E	centromere protein E
CF IIm	cleavage factors II
CF Im	cleavage factors I
CIN	chromosomal instability
CPSF	cleavage and polyadenylation specificity factor
CPT	camptothecin
CR-APA	coding region APA
CS	cleavage site
CstF	cleavage stimulation factor
DDR	DNA damage response
DEGs	differentially expressed genes
DtoP	distal to proximal
Eg5	kinesin-5
FPKM	fragments per kilobase of exon model per million reads mapped
Gadd45	growth arrest and DNA damage inducible
GEO	Gene Expression Omnibus
GO	gene ontology
HuR	human antigen R

MXE	mutually exclusive exons
NMR	nuclear magnetic resonance
NS	non-silencing
PABPN1	polyadenylate binding protein-1
PAS	polyadenylation signal
PDUI	percentage distal PAS usage index
PtoD	proximal to distal
RBP	RNA binding protein
RI	retained introns
RNA-seq	RNA sequencing
RPM	reads per million mapped reads
SE	skipped exons
siBubR1	BubR1 siRNA
siCENP-E	CENP-E siRNA
siEg5	Eg5 siRNA
siRNA	small interfering RNA
snRNP	small nuclear ribonucleoprotein
SSA	SF3B inhibitor spliceostatin A
TOPOi	topoisomerase inhibitors
TSS	transcription start site
TTS	transcription termination site
UTR-APA	untranslated region APA
WGCNA	weighted gene co-expression network analysis

General introduction

Transcriptomic complexity

Genomic information is transduced from gene to protein via transcript (Crick, 1958; Crick, 1970). In the sequential process, several quantitative and qualitative regulatory machineries contribute to production of enormous number of biological molecules, making it possible to control complex life systems sophistically. In these processes, transcriptional layer has a central role in the generation of molecular diversity through several machineries (Manning et al. 2017). Main events causing transcriptional variation are alternative transcription and alternative splicing. Alternative transcript, including promoter selection and polyadenylation, generates over 90,000 events, and alternative splicing produces over 50,000 events (Pal et al. 2012). As a result of the combination of their events, approximately 240,000 proteins are recorded in human gene database Ensembl (GRCh38.p12), and it is considered that much more kinds of protein could exist. Furthermore, abundance of each transcript is also perturbated by endogenous and exogenous stimulations. Therefore, capturing landscape of transcriptomic perturbation in cells could provide clues for understanding complex biosystems.

Transcripts in cancer cells

Transcriptomic perturbation in cancer cell is supposed to be much more complex than that in normal cells. It is generally recognized that accumulation of genetic alteration in cells is the main cause of neoplastic cell transformation (Lengauer et al., 1998). A number of studies reported that oncogenic genetic alterations affect most genetic products, transcripts and proteins and could lead to dysregulation of transcriptional control (Bradner et al., 2017; Sur et al., 2016; Wang et al., 2015; Lawrence et al., 2014; Garraway et al., 2013; Stratton et al., 2009). Transcriptional dysregulation is considered as one of the fundamental features of

cancer. It is assumed that cancer cells are induced to pluripotent stem cells by transcriptional dysregulation and acquire an ability of self-renewal, which shows suppression of differentiation and enhancement of cell proliferation (Vicente-Dueñas et al., 2013). Transcription activity is coordinated with cell proliferation and early differentiation process (Villicaña et al., 2014; Gonda et al.,2015). Therefore, cancer cells have higher activity of transcription and transcriptomic profiling of cancer cells could be quite complicated. Furthermore, somatic mutations involve in expanding variation of transcripts because of their effects to protein coding, splicing regulation, transcriptional regulation, transcript termination and generation of fusion gene. Consequently, transcriptomic perturbation caused by a stimulation could be more complex in cancer cells than that in normal cells. Therefore, decoding genome-wide transcriptomic profiling is essential for elucidation of oncologic molecular mechanisms.

Transcriptomic profiling

Over the past a couple of decades, technologies measuring genome-wide transcripts have made remarkable progress. Microarray has made us possible to readily survey quantitative transcript perturbation (Schena et al., 1995; Wodicka et al., 1997). RNA sequencing (RNA-seq) by next generation sequencing has not only quantitative but also qualitative measurement abilities (Wang et al., 2009; Marguerat et al., 2010). mRNA-seq can quantify comprehensive gene expression and detect alternative splicing events simultaneously (Li et al., 2008). 5'-seq and 3'-seq can detect genome-wide alternative transcriptional start sites and alternative polyadenylation sites, respectively (Beck et al., 2010; Jenal et al., 2012; Adiconis et al., 2018). These technologies have contributed greatly to the progress of biology. However, it is difficult to find meaningful information from these transcriptome data due to

their enormousness and noisiness feature (Klebanov et al., 2007; Wu., 2009). For this challenge, various informatics approaches have been proposed (Hira et al., 2015; Conesa et al. 2016) and the researcher should select and combine the right methods for the intended purpose.

Objective of my study

The objective of my study is to decode transcriptomic perturbation in cancer cells and create new hypotheses using genome-wide transcriptomic profiling. In this study, I demonstrated the effectiveness of genome-wide profiling approaches for understanding molecular mechanism of biological response in cancer cells. Moreover, I suggested practicable hypotheses for two subjects. In chapter 1, I quantitively profiled transcriptomic perturbation in cancer cells with CENP-E inhibition and created a hypothesis of anti-tumor effect of CENP-E inhibition for cancer therapy. In chapter 2, I generated novel APA modulators and elucidated their mechanism of action in cancer cells from multilayered qualitative transcriptomic profiling.

Chapter 1:

Expression data of HeLa cells treated with CENP-E siRNA or Eg5 siRNA in the presence of BubR1 siRNA

Introduction

CENP-E and Eg-5 are mitotic spindle motor proteins of the kinesin family, which play an important role for regulation of mitosis (Miki et al. 2005). Recently, inhibitors of CENP-E and Eg5 have been developed as cancer therapeutics (Wood et al. 2010; Chung et al. 2011; Rath et al. 2012). Against spindle assembly checkpoint (SAC) defective tumor, CENP-E inhibitor suppressed the proliferation and increased the apoptosis of cancer cells via mitotic aberrations, while Eg5 inhibitor showed no anti-proliferative effect (Ohashi et al. 2015). However, the molecular mechanism responsible for the difference of cell fate after mitotic aberrations is unclear. I investigated the postmitotic effects of different mitotic aberrations (Ohashi et al. 2015), misaligned chromosomes produced by CENP-E siRNA (siCENP-E), and monopolar spindles resulting from Eg5 siRNA (siEg5) on SAC-defective conditions with BubR1 siRNA (siBubR1) (Miki et al. 2005). To determine which signaling pathways contribute to the postmitotic effect of siCENP-E in the presence of siBubR1 (siCENP-E + siBubR1) compared with siEg5 + siBubR1, I performed comprehensive gene expression analysis using microarray comparisons.

Materials and Methods

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67905

Cell cultures

HeLa cells, purchased from American Type Culture Collection (ATCC; Manassas, VA, USA), were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS.

Transfection of siRNA oligonucleotides

siRNA oligonucleotides targeting CENP-E (M-003252-02), Eg5 (M-003317-01), and BubR1 (M-004101-02) were obtained from Dharmacon (SMART pools, Dharmacon, Lafayette, CO, USA). The siTrio negative control (B-Bridge International, Inc., Mountain View, CA, USA) was used as a non-silencing (NS) siRNA (siNS). Twenty-five nanomoles of pooled siRNA per gene were used for double knockdown (total siRNA concentration, 50 nM). siRNA transfection was performed as described previously [Ohashi et al., 2005]. Transfection of siRNA oligonucleotides was performed with Dharmafect (Dharmacon) in 6-well plates according to the manufacturer's specifications. Seventy-two hours after transfection, the cells were collected for RNA preparation for GeneChip microarray analysis.

GeneChip microarray analysis

To determine which signaling pathways contribute to the postmitotic effect of siCENP-E + siBubR1 compared with siEg5 + siBubR1, I performed comprehensive gene

expression analysis using microarray comparisons (Figure 1, Table 1).

Total RNAs were extracted using the RNeasy Miniprep kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. Preparation of cDNAs and cRNAs, hybridization, and microarray scanning were performed according to the manufacturer's protocols (Affymetrix Inc., CA, USA). Biotinylated cRNAs were hybridized to Affymetrix Human Genome U133 Plus 2.0 Array. Microarray Analysis Suite 5.0 (MAS5; Affymetrix) was used to quantify microarray signals, and the intensities were normalized to the median expression level by setting the median intensity to 1.0 using the GeneSpring software package (Agilent Technology, Santa Clara, CA, USA). Differentially expressed genes were identified with a fold change of > 2.0. Probe sets with absent call in all samples for comparison were filtered out. In addition, intensity filtering was applied, with the criterion set to 0.5. ClusterProfiler (Yu et al., 2012) and KEGG Mapper (Kanehisa, et al., 2019) were using for pathway analysis. The microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE67905.

Results

To determine which signaling pathways contribute to the postmitotic effect of siCENP-E in the presence of siBubR1 (siCENP-E + siBubR1) compared with siEg5 + siBubR1, I performed comprehensive gene expression analysis using microarray comparisons (Figure 1, Table 1). First, to identify genes affected by CENP-E and Eg5 knockdown, I carried out differentially expressed genes analysis for siCENP-E + siBubR1 and siEg5 + siBubR1 compared with siNS cells. The number of detected genes was 387 (up:224, down:163) and 120 (up:74, down:46), respectively (Table 2). Next, to determine which biological pathways were affected, I conducted pathways analysis using each differentially expressed gene set. Although pathway of p53 signaling showed the most significant enrichment for both siCENP-E (q-value=4.7×10^{-8}) and siEg5 (q-value=7.5×10^{-4}), significance and number of overlapped genes in siCENP-E were superior to siEg5 (Table 3, Table 4). ATF2/FRA/AP1 pathways and cytokine gene sets showed significant enrichment only for siCENP-E. To examine signaling cascade of p53 pathway in siCENP-E cells, I mapped upregulated genes onto p53 pathway using KEGG mapper. Upregulated genes of p21, CyclinD and Gadd45 (fold change: 3.5, 2.8 and 4.6, respectively) were mapped on a path that lead to cell cycle arrest, and Gadd45 was on a path that result to DNA repair and damage prevention (Figure 2). These results indicate that siCENP-E could induce apoptosis via p53 signaling pathway and DNA damage response.

Discussion

In this study, I acquired microarray gene expression profiles for siBubR1 + siCENP-E-treated, siBubR1 + siEg5-treated, and siBubR1 -treated cells. These expression data would be valuable for understanding the postmitotic effects of aneuploidy as well as polyploidy. Through quantitative analysis, I found that p53 signaling pathway was upregulated in CENP-E knockdown cells and the results lead to a hypothesis that activation of p53 signaling with induction of cell cycle arrest and DNA damage response could induce anti-proliferative effect in cancer cells with CIN feature.

Tables & Figures

Table1 Summary of expression data in this study.

Specifications	
Organism/cell line/tissue	*Homo sapiens* /HeLa/cervical cancer
Sex	Female
Sequencer or array type	Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix Inc., CA, USA)
Data format	Raw: CEL files, normalized data: SOFT, MINIML and TXT
Experimental factors	siBubR1 + siCENP-E-treated cells vs. siBubR1 + siEg5-treated cells vs. siBubR1-treated cells. siNS-treated cells were used as a negative control.
Experimental features	A comprehensive gene expression analysis by microarray comparisons was performed comparing siBubR1 + siCENP-E-treated, siBubR1 + siEg5-treated, and siBubR1-treated cells.
Consent	Not necessary, HeLa cells were obtained from ATCC.
Sample source location	Fujisawa, Kanagawa, Japan

Table2 Number of differentially expressed genes.

	Up-regulated	Down-regulated
siCENP-E+siBubR1	224	163
siEg5+siBubR1	74	46

Table3 Significant enriched pathway in siCENP-E +siBubR1 cells

Pathway	q-value
P53 DOWNSTREAM PATHWAY	4.72E-08
ATF2 PATHWAY	4.86E-07
FRA PATHWAY	1.88E-06
AP1 PATHWAY	1.96E-05
CYTOKINE CYTOKINE RECEPTOR	4.43E-04

The q-value was calculated using hypergeometric test through Benjamini-Hochberg procedure.

Table4 Significant enriched pathway in siEg5 +siBubR1 cells

Pathway	q-value
P53 DOWNSTREAM PATHWAY	7.48E-04

The q-value was calculated using hypergeometric test through Benjamini-Hochberg procedure.

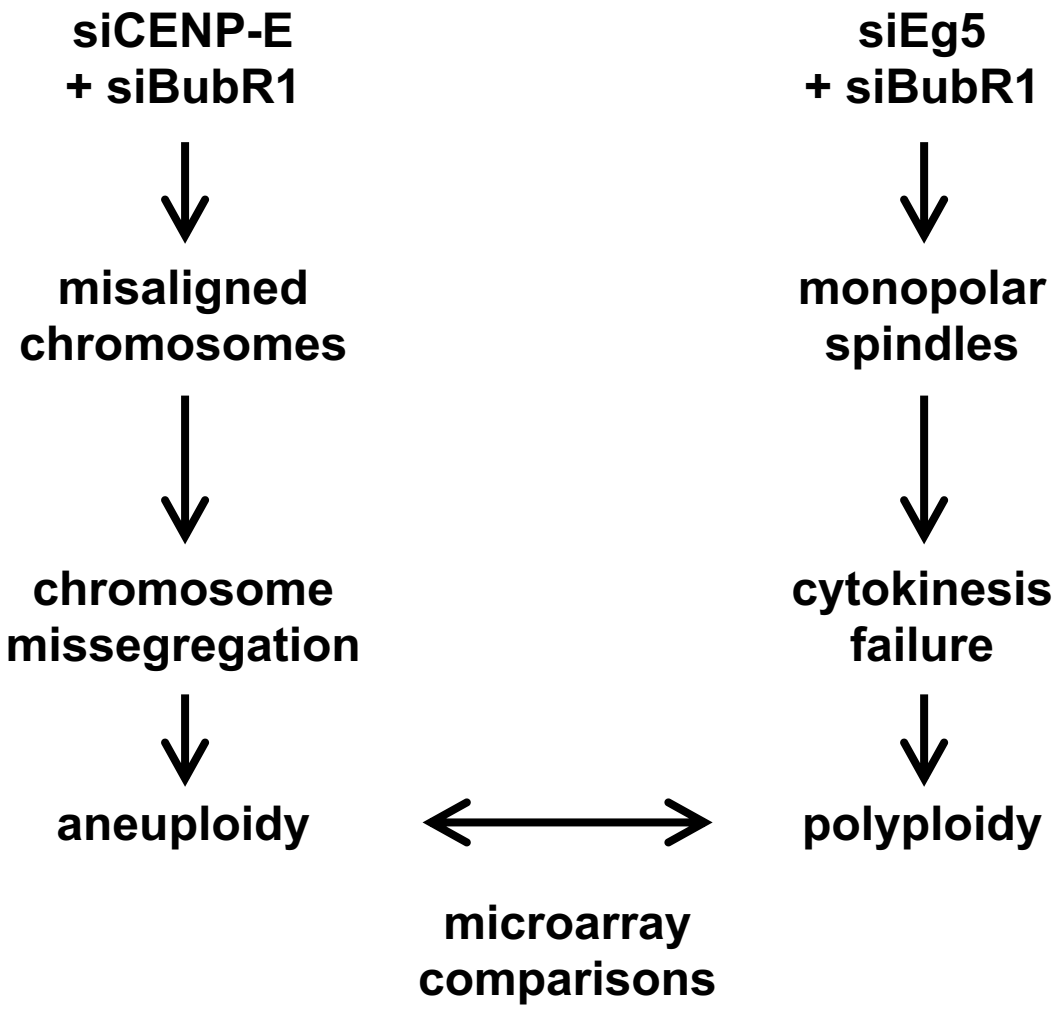


Figure 1. Schematic presentation of microarray comparisons in this study.

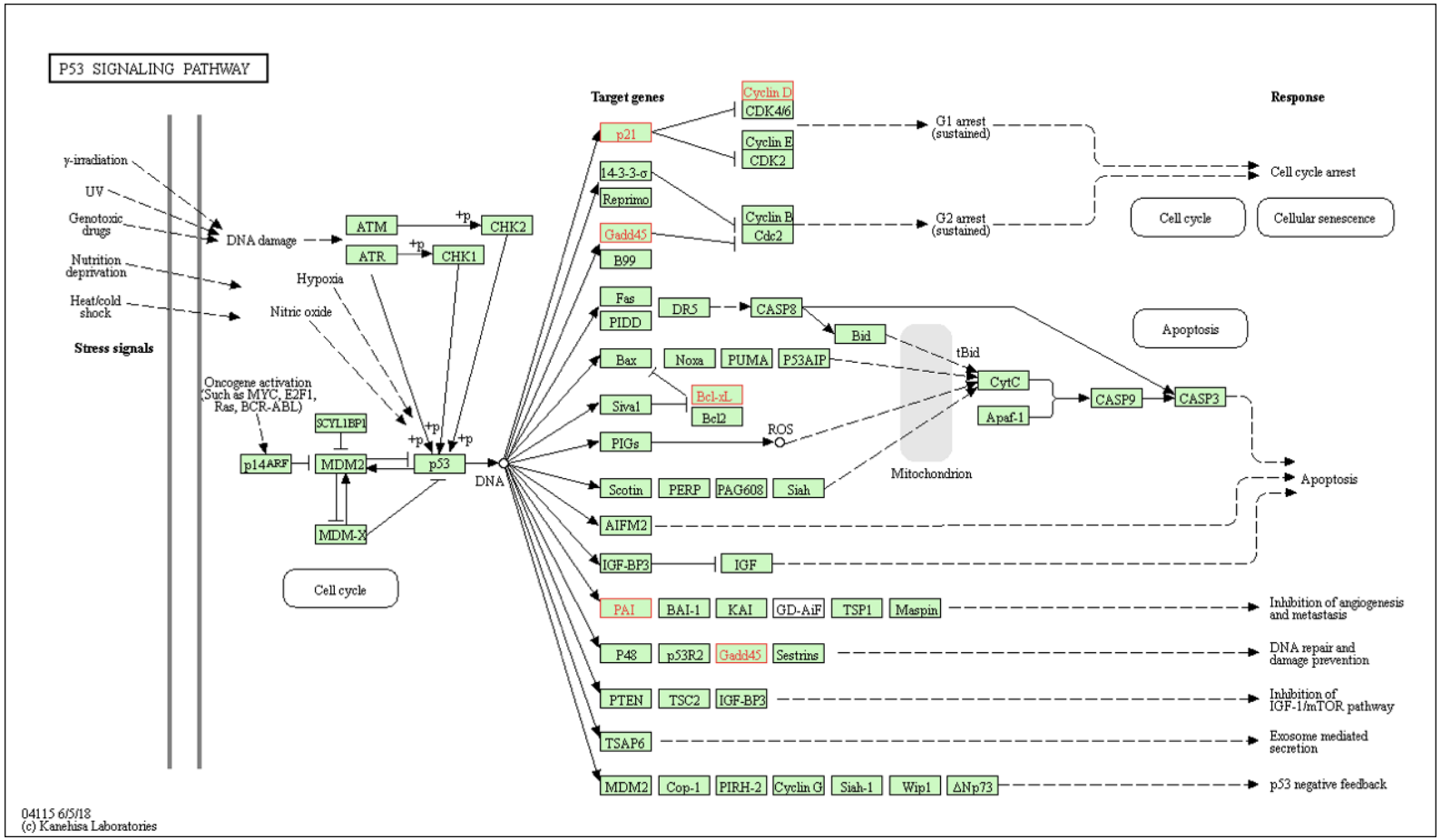


Figure 2. Pathway map of p53 signaling in siCENP-E cells.

Pathway of p53 signaling mapped with differentially expressed genes by KEGG mapper. Red colored boxes indicate genes up-regulated in siCENP-E cells.

Chapter 2:

Decoding Transcriptome Dynamics of Genome-Encoded Polyadenylation and Autoregulation with Small-Molecule Modulators of Alternative Polyadenylation

Summary

Alternative polyadenylation (APA) plays a critical role in regulating gene expression. However, the balance between genome-encoded APA processing and autoregulation by APA modulating RNA binding protein (RBP) factors is not well understood. I discovered two potent small-molecule modulators of APA (T4, T5) that promote distal to proximal (DtoP) APA usage in multiple transcripts. Monotonically responsive APA events, induced by short exposure to T4 or T5, were defined in the transcriptome, allowing clear isolation of the genomic sequence features and RBP motifs associated with DtoP regulation. I found that longer vulnerable introns, enriched with distinctive A-rich motifs, were preferentially affected by DtoP APA, thus defining a core set of genes with genomically encoded DtoP regulation. Through APA response pattern and compound–siRNA epistasis analysis of APA-associated RBP factors, I further demonstrated that DtoP APA usage is partly modulated by altered autoregulation of polyadenylate binding nuclear protein-1 signaling.

Introduction

More than 70% of human genes produce alternatively polyadenylated mRNAs, which is considered to be a critical molecular mechanism for generating diversity in the transcriptome and proteome (Di Giammartino et al., 2011; Elkon et al., 2013; Klerk and Hoen, 2015). Polyadenylation of mRNA is an essential step for mRNA maturation and involves endonucleolytic cleavage and the addition of a poly(A) tail, as well as protein complexes such as cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), and cleavage factors I (CF Im) and II (CF IIm) (Di Giammartino et al., 2011; Elkon et al., 2013; Tian and Manley, 2013). In the cleavage process, CPSF, CstF, and CF Im bind to the AAUAAA polyadenylation signal (PAS) hexamer, located upstream of the cleavage site (CS) and downstream of the UG-rich sequence. Poly(A) tails of mRNA are added by poly(A) polymerase and stabilized by poly(A) binding proteins including polyadenylate binding protein-1 (PABPN1) (Chan et al., 2011; Jenal et al., 2012).

Alternative polyadenylation (APA) is classified into either coding region APA (CR-APA) or untranslated region APA (UTR-APA). CR-APA produces mRNA isoforms with distinct C-terminal coding regions, resulting in distinct protein isoforms, for example, IgM proteins (Di Giammartino et al., 2011; Peterson, 2007). UTR-APA produces mRNA isoforms with distinct 3'-UTR lengths, each of which encodes the same protein. Longer 3'-UTRs can include target sites for miRNAs and RNA binding proteins, thereby potentially affecting mRNA stability, translational efficiency, or subcellular localization (Di Giammartino et al., 2011; Elkon et al., 2013; Shi, 2012).

Recent studies have shown that splicing and polyadenylation are coupled events occurring co-transcriptionally and approximately 20% of human genes have polyadenylation

sites within their introns, indicating dynamic interplay between polyadenylation and splicing (Elkon et al., 2012; Tian et al., 2007). U1 small nuclear ribonucleoprotein (snRNP) is protected against premature cleavage at cryptic polyadenylation sites within the introns (Berg et al., 2012; Kaida et al., 2010; Luo et al., 2013). SF3b155, a core component of U2 snRNP, generally suppresses intronic cleavage and polyadenylation through splicing regulation (Li et al., 2015). The rate of RNA polymerase II elongation affects splicing and tissue-specific APA in the cells showing a slow or rapid elongation rate (Fong et al., 2015; Fong et al., 2014; Liu et al., 2017).

Chemical biology approaches have been useful to clarify the molecular mechanisms, regulation, and functions of cellular processes involved in RNA processing pathways such as alternative splicing (Salton and Misteli, 2016). Topoisomerase inhibitors were reported to modulate CR-APA by dissociation from human antigen R (HuR) in pre-mRNA (Dutertre et al., 2014). However, no APA-specific inhibitors or modulators have been identified.

In this study, I identified small molecules that can modulate APA, using reporter-based high-throughput screening. These compounds predominantly modulated APA involved in alternative splicing, primarily leading to the use of proximal poly(A) CS, through the autoregulation of pathways related to RNA processing and splicing. The compounds used distinctive A-rich motifs, through autoregulated PABPN1 signaling. These molecules have potential value for understanding the molecular mechanisms of APA, and their use may lead to the discovery of different classes of therapeutics.

Materials and Methods

Cell culture

The human female osteosarcoma cell line U2OS was maintained in McCoy's 5a (Life Technologies, Carlsbad, CA, USA). The human female embryonic kidney cell line 293T and 293 were maintained in Dulbecco's modified Eagle's medium (Life Technologies). The human female adenocarcinoma cell line HeLa was maintained in Eagle's Minimum Essential Medium (Life Technologies). The media were supplemented with 10% FBS (Life Technologies), penicillin (10,000 U/mL; Life Technologies), and streptomycin (10,000 μg/mL; Life Technologies). All cells were cultured at 37° C in humidified incubators at 5% CO_2. Cells were seeded and subcultured in 100-mm-diameter dishes every 3–4 d. All cell lines were purchased from American Type Culture Collection (Manassas, VA, USA).

Compounds

SN38 (Cas No. 86639-52-3) was from Tocris Bioscience (Bristol, UK). Camptothecin (CPT) (CAS: 7689-03-4) and Febuxostat (Cas No. 86639-52-3) were from Sigma-Aldrich (St. Louis, MO, USA). T4 (Cas No. 785708-33-0) was from Enamine Chemicals (Monmouth Jct., NJ, USA). T5 was synthesized and purified as described in a previous patent application (Siu Tony, 2011). T4 is 6,7-dimethoxy-2- [(4-phenyl-3,6-dihydropyridin-1 (2H)-yl)methyl)] quinazolin-4(3H)-one and T5 is 3-cyclobutyl-6-(2-naphthylmethyl)-2,7-dihydro-4H-pyrazolo[3,4-b]pyridin-4-one. To confirm chemical identity for T5, nuclear magnetic resonance (NMR) spectra were recorded on Bruker AVANCE II+ (^{1}H NMR 600 MHz and ^{13}C NMR 151 MHz). Chemical shifts are given in parts per million (ppm) with tetramethylsilane as an internal standard. Abbreviations are used as

follows: s = singlet; d = doublet; t = triplet; m = multiplet; br s, broad singlet). Coupling constants (J values) are given in hertz (Hz). Low-resolution mass spectra (MS) was acquired using an Agilent LC/MS system (Agilent1200SL/Agilent6130MS) operating in an electrospray ionization mode (ESI+). The column used was an L-column 2 ODS (3.0 × 50 mm I.D., 3 μm, CERI) with a temperature of 40 °C and a flow rate of 1.5 mL/min. Mobile phases A and B were a mixture of 5 mmol/L $AcONH_4$ and MeCN (9/1, v/v) and a mixture of 5 mmol/L $AcONH_4$ and MeCN (1/9, v/v), respectively. The ratio of mobile phase B was increased linearly from 5 to 90% over 0.9 min, 90% over the next 1.1 min. The purity of compound tested in biological systems was assessed as being >95% using elemental analysis. Elemental analysis was carried out by Sumika Chemical Analysis Service and was within 0.3% of the theoretical values. T5 is a white solid. ^{1}H NMR (600 MHz, DMSO-d_6) δ 1.79–1.87 (m, 1H), 1.91–2.01 (m, 1H), 2.18–2.29 (m, 2H), 2.32–2.45 (m, 2H), 3.80–3.90 (m, 0.35H), 3.91–4.05 (m, 1.95H), 4.18 (br s, 0.7H), 5.55 (br s, 0.65H), 6.32 (br s, 0.35H), 7.37–7.57 (m, 3H), 7.76–7.96 (m, 4H), 11.03 (s, 0.35H), 11.62 (br s, 0.65H), 12.76 (br s, 0.35H), 13.17 (br s, 0.65H). Note: Two isomers were observed in ca. 65:35 ratio. ^{13}C NMR (151 MHz, DMSO-d_6) δ 17.9, 27.7 (2C), 28.3 (2C), 30.7, 33.9, 38.8, 44.1, 101.0, 102.4, 107.1, 108.2, 125.3, 125.6, 126.0, 126.1, 127.0, 127.1, 127.26, 127.35, 127.4, 127.66, 127.73, 127.9, 131.6, 131.8, 132.9, 133.1, 135.5, 137.5, 144.1, 147.2, 151.2, 151.8, 155.6, 159.8, 161.6, 177.1. Three carbon peaks were not observed, presumably due to overlapping of peaks. MS m/z 330.2 $(M + H)^+$. Anal. Calcd for $C_{21}H_{19}N_3O \cdot 0.2H_2O$: C, 75.74; H, 5.87; N, 12.62. Found: C, 75.71; H, 5.88; N, 12.51. Single crystal X-ray structure of T5 was obtained and is available from the Cambridge Crystallographic Data Centre (CCDC 1863277).

Construction of reporter plasmids

Reporter plasmids were generated as described previously (Jenal et al., 2012). The 3'-UTR (898 bp) encoding the KMT2A gene (NM_005933) was inserted into downstream *XhoI* and *NotI* sites for Renilla in psiCHECK2 (psiC) (Promega, Madison, WI, USA), using an IN Fusion cloning kit (Takara Bio, Kusatsu, Japan) and the resulting plasmid was named pSTAR. pSTAR-Dicer WT plasmid was generated by the insertion of DICER1 (NM_030621, 5952-6229 bp) and 3'-UTR of KMT2A at XhoI and NotI sites of psiC. pSTAR-Dicer Mt plasmid was mutated with PAS of pSTAR plasmid. pSTAR-p27 WT plasmid was generated by the insertion of CDKN1B (NM_004064, 2213-2413 bp) and the 3'-UTR of KMT2A at *XhoI* and *NotI* sites of psiC. pSTAR-p27 Mt was mutated with PAS of the pSTAR-p27 WT plasmid. All gene fragments were synthetic.

Compound screening

Approximately 45,000 compounds, each at a single concentration of 3 or 10 μM, were screened with a luciferase assay. Each pSTAR series plasmid (pSTAR, pSTAR-Dicer WT, or psiC) was transiently transfected into 293T cells using FuGene HD transfection reagent (Promega), in accordance with the manufacturer's instructions. After 24 h, the cells were collected and frozen until use. The frozen cells were recovered and seeded to individual wells of a 1536-well plate containing each compound. After 24 h treatment, luciferase activity was measured using the Renilla-Glo Luciferase Assay (Promega), in accordance with the manufacturer's instructions. The chemiluminescence of each well was measured by a LEADseeker Microplate Reader (GE Healthcare, Chicago, IL). Percent inhibition was calculated as a ratio of luciferase activity in the presence of the compound compared with that in the controls containing 1% DMSO.

3'-End quantitative PCR

Total RNA was extracted using an RNeasy Miniprep Kit (Qiagen, Valencia, CA, USA), in accordance with the manufacturer's instructions. RNA was fragmented at 70° C for 5 min using RNA fragmentation buffer (Life Technologies) and reverse transcription was performed using SuperScript III First-Strand Synthesis Super Mix (Life Technologies) and P7-t25-vn oligo-dT primer (CAA GCA GAA GAC GGC ATA CGA GAT TTT TTT TTT TTT TTT TTT TTT TTT VN). The expression level of each transcript was measured using Power SYBR Green PCR Master Mix (Applied Biosystems, Carlsbad, CA, USA) with P7 and proximal specific or distal specific primers. The expression level of each mRNA was normalized to that of beta-actin mRNA. All primer and probe sequences are listed in Table S3.

Transfection of cells with siRNAs

293T cells were seeded at a density of 3×10^3 cells/well into 96-well plates and incubated overnight in culture medium containing 10% FBS. Each siRNA (Silencer Select siRNA, Life Technologies) or a control Non-Silencing siRNA (Life Technologies) was mixed with DharmaFECT 1 transfection reagent, the cells were transfected for 72 h in accordance with the manufacturer's instructions, and cell lysates were prepared for real-time PCR.

Cell viability assay

Cells (3×10^3 per well) were seeded into 384- or 96-well plates and incubated overnight in 50 μL of culture medium. Next, 50 μL of each compound, diluted in culture medium, was added to individual wells. After 72 h of incubation, procedures for a CellTiter-Glo Luminescent Cell Viability Assay (Promega) were performed, in accordance with the

manufacturer's instructions. The chemiluminescence of each well was measured by an ARVO X3 Microplate Reader (Perkin-Elmer, Waltham, MA). GI_{50} values were calculated using GraphPad Prism 5 software (La Jolla, CA, USA) with a sigmoid dose–response curve.

mRNA sequencing

Total RNA was extracted using an RNeasy Miniprep Kit (Qiagen) and the quality was ascertained by the presence of two distinct peaks at 18S and 28S, with no additional peaks, using a Bioanalyzer (Agilent Technology, Santa Clara, CA, USA). Cells were treated with CPT, T4, or T5 for 6 h (for details, see Figure 5). Illumina mRNA-Seq reads (100 nucleotides, 200 million reads) were aligned to the hg19 (GRCh37) reference genome assembly, using the GSNAP aligner (Wu and Nacu, 2010). Aligned libraries had mate-pair information fixed, had potential PCR duplicates removed, and were sorted using SAMtools (Li et al., 2009). Cufflinks was used to quantify gene expression fragments per kilobase of exon model per million reads mapped (FPKM) values (Trapnell et al., 2012). Differential splicing detection was performed using MISO (Katz et al., 2010). Alternative splicing events, with Bayes factor < 20, $\Delta\mathrm{PSI} < 0.1$, 0 reads supporting the inclusion or exclusion isoform, or < 10 reads supporting either of the event isoforms, were removed.

Construction of 3'-seq libraries

3'-Seq libraries were generated as described previously (Beck et al., 2010; Jenal et al., 2012). Total RNA was extracted using an RNeasy Miniprep Kit (Qiagen) and poly(A) mRNA was purified using Oligotex -dT30 Super mRNA Purification Kit (Takara Bio.). Five hundred nanograms of Poly(A) mRNA was fragmented at 70° C for 5 min using RNA fragmentation buffer (Life Technologies) and purified using NucleoSpin RNA Clean-up XS

(Takara Bio.). Then, first-strand cDNA was synthesized using SuperScript Double-Stranded cDNA Synthesis Kit (Life Technologies) and P7-t25-vn oligo-dT primer. Subsequently, second-strand cDNA was synthesized and then end-repaired at 30° C for 20 min, using NEBNext End Repair Module (New England Biolabs, Ipswich, MA, USA). cDNA was purified using MinElute Reaction Cleanup Kit (Qiagen) and 100 pmol annealed P5-splinkerette ligated with cDNA for 15 min at 25° C using Quick Ligation Kit (New England Biolabs). Ligated cDNA was purified and size-selected (~220 bp), using NucleoMag NGS Clean-up and Size Select Kit (Takara Bio). Samples were amplified with P5 and P7 primers using Phusion Hot Start Flex 2X Master Mix (New England Biolabs). PCR steps were: 98° C for 30 s; 18 cycles of 98° C for 10 s, 65° C for 30 s, 72° C for 30 s; 72° C for 5 min. After purification, the library was sequenced on the Illumina HiSeq 2000. All primer sequences are listed in Supplemental Methods.

Analysis of 3'-seq data

Adaptor, low-quality (quality score < 30), and polyA sequences were trimmed from raw sequence reads using PRINSEQ (Schmieder and Edwards, 2011). Trimmed sequence reads were aligned to the hg19 (GRCh37) reference genome using Bowtie2, with one mismatch allowed in a seed alignment during multiseed alignment (Langmead and Salzberg, 2012). Only uniquely mapped reads were used for subsequent analyses.

Mapping of poly(A) cleavage sites

Peaks of 3'-seq data were identified from uniquely mapped reads using MACS software (Zhang et al., 2008) with default settings. Peak coordination between control and treated samples was arranged by MAnorm software (Shao et al., 2012). Peaks with fewer than

50 reads in all samples were discarded. Gene and region information was assigned to the peaks using annotatePeaks.pl from HOMER (Heinz et al., 2010). The intensity of each peak was quantified by reads per million mapped reads (RPM) metrics. PDUI metrics was used to qualify relative poly(A) site usage, as described previously (Xia et al., 2014). Briefly, PDUI was calculated using following formula: PDUI = [RPM of distal site] / ([RPM of distal site] + [RPM of proximal site]). To calculate PDUI, the distal and proximal sites were identified in each gene. The peak, located at the most distant region from the gene start position, was designated as the distal poly(A) site and the other peaks in the gene were designated as proximal poly(A) sites. Thus, several proximal poly(A) sites occasionally exist in one gene. If there was only one peak in a gene, that peak was excluded. PDUI was calculated for each peak of the proximal poly(A) site. The difference between PDUI values for control and treated groups was described as ΔPDUI. Pairs of proximal and distal sites that showed an absolute ΔPDUI value of over 0.4 and over 100 reads, in control or treatment samples, were each identified as reflecting significant alternative poly(A) usage (APA).

Motif analysis

Sequences at 50 bp upstream from the poly(A) CSs were extracted from the human genome. Program of findMotifs.pl from HOMER was used to discover de novo enriched motifs (Heinz et al., 2010). The RBPmap online web server was used to search putative target sites of RNA binding proteins for 114 known human/mouse motifs (Paz et al., 2014).

Clustering analysis: WGCNA

Alternative poly(A) site usage and gene expression response profiles were clustered using the WGCNA R package (Langfelder and Horvath, 2008). Soft threshold selection was

facilitated by calculating the scale-free network topology model and fitting R2 values for soft thresholds 1–30, using the pickSoftThreshold function. Initial threshold values were manually chosen by selecting values where the topology model fit was both relatively stable and high. Final thresholds were selected by adjusting the values to produce cluster sets that were comparable among datasets. WGCNA was run with networkType="signed", minModuleSize=30. Similar clusters were merged using the mergeCloseModules function. APA events were selected for PDUI profile clustering if they involved differential poly(A) site usages in any of the treated samples, at an absolute ΔPDUI threshold value of 0.4. Events with missing PDUI values were removed. A soft threshold of 20 was selected for both the T4 and the T5 treatment datasets. Genes selected for clustering were required to have FPKM values ≥ 1 in at least one library for the T4 and T5 treatments of 3'-seq datasets. A soft threshold of 26 was selected for both T4 and T5 treatments of RNA-seq datasets.

Biological process enrichment analysis and enrichment map generation

Gene ontology (GO) Biological Process (BP) term enrichment for a set of genes was performed by generating functional interaction networks using the BiNGO Cytoscape plugin (Maere et al., 2005). BP terms with ≤ 20 and ≥ 500 genes were filtered out. BP term enrichment was performed using genes in the resulting network and terms having *P*-values controlled at 1.0e−4, false discovery rate controlled at 0.1, and overlap coefficient controlled at 0.5. Enrichment maps were generated for BiNGO BP enrichment results using the EnrichmentMap Cytoscape plugin (Merico et al., 2010).

Heatmap

Hierarchical clustering and heatmap visualization were performed using Partek

Genomics Suite Version 6.6 (Partek Inc., St. Louis, MO, USA). Hierarchical clustering was performed using Pearson's dissimilarity as distance measure and Ward's aggregation procedure.

Transcript length analysis

Profiling of numbers of exons, transcript lengths, and differentially expressed genes (DEGs) was performed using the ggplot command in R. The transcript information (hg19) was downloaded from the University of California Santa Cruz (UCSC) Genome Browser Database. DEGs were required to have FPKM values ≥ 1 in at least one sample and were indicated by an absolute fold change ≥ 1.5 for each compound concentration, between RNA-Seq datasets. Statistical significance in comparisons between two different treatments or groups was determined using Welch's *t* test.

Quantification and statistical analysis

The number of biological replicates and statistical significance are noted in the figures, figure legends, and manuscript. Data are shown as mean ± standard deviation (SD). Statistical analysis and graph generation for 3'-end qPCR, cell growth assay, and reporter assay were calculated using GraphPad Prism 5 software. Hypergeometric test was used for the motif analysis and the functional enrichment analysis to assess the statistical significance using HOMER (Heinz et al., 2010) and the BiNGO Cytoscape plugin (Maere et al., 2005), respectively. Welch's *t* test was used to determine the statistical significance for the intron and transcript length analyses using R. Pearson's tests were performed for the pathway correlation between Cluster of T4 and that of T5 by using R. Statistical significance was defined as $p < 0.05$.

Data availability

Raw sequencing data are available at Gene Expression Omnibus (GEO) [GEO series accession number: GSE99527 (RNA-seq), GSE98948 (3'-seq)].

Results

Reporter system for discovery of modulators of alternative polyadenylation

To identify small molecules that regulate APA, I constructed a reporter system based on a previously described approach (Jenal et al., 2012) (Figure 3A). Insertion of 3'-UTR of the KMT2A gene downstream of Renilla luciferase (pSTAR) repressed mRNA transcription (Gomez-Benito et al., 2011), thereby decreasing luciferase expression (Figure 3B). I next inserted the proximal alternative PAS region of the Dicer gene upstream of the 3'-UTR of KMT2A (pSTAR-Dicer WT) and used the mutated PAS of Dicer as a control (pSTAR-Dicer Mt). I confirmed that pSTAR-Dicer WT induced Renilla expression by more than 5-fold owing to mRNA cleavage mediated by the Dicer PAS, while pSTAR-Dicer Mt slightly relieved the luciferase activity, which was consistent with the finding in a previous report (Jenal et al., 2012) (Figure 3B). PAS encoding p27 showed similar reporter activity. To address whether the pSTAR-Dicer system would be suitable for compound screening, I validated the reporter activity using a knockdown strategy against PABPN1, with extensive 3'-UTR shortening (Jenal et al., 2012). I confirmed that all three independent siRNA sequences efficiently silenced >80% of PABPN1 mRNA (Figure 4A). PABPN1 silencing suppressed 70% and 30% of the reporter activity, compared with the levels in the controls, in pSTAR-Dicer WT and pSTAR-p27 WT, respectively (Figure 4B). Because decreased reporter activity was also observed in pSTAR-Dicer Mt, it was possible that cleavage and polyadenylation were induced by alternative PAS CS, rather than a mutated PAS. Taken together, these findings suggest that pSTAR-Dicer WT would be appropriate for compound screening of APA modulators.

High-throughput compound screening with the reporter assay

High-throughput screening of approximately 45,000 compounds was performed using the pSTAR-Dicer WT reporter plasmid. Reporter activity using pSTAR-Dicer WT was evaluated by treating cells with 3 or 10 μM of each compound for 24 h, with psiC being used as a counterscreen. There was a high signal-to-background ratio (4.3) and a Z' factor of at least 0.6, indicating that the cell-based assay was highly robust and reliable. Differences of luciferase activity between pSTAR-Dicer WT and psiC were well correlated among the screened compounds (R^2 = 0.723) (Figure 3C), suggesting that many of the compounds affected luciferase expression as false positives. A total of five active compounds satisfied the primary screen hit criteria, with high replicate reproducibility for (i) inhibition >30% of pSTAR-Dicer WT and (ii) >30% difference between reporter activity of pSTAR-Dicer WT and psiC. I validated the relationship between reporter and cell growth inhibitory activities using 320 compounds, randomly chosen from the 45,000, and confirmed that luciferase activity was not strongly correlated with cell growth inhibitory activity (R^2 = 0.424) (Figure 3D). These results suggested that reporter activity did not depend on a secondary effect such as cytotoxicity. To confirm the activity of the compounds, the dose-dependence of luciferase activity and compound purity were tested. I thereby identified five molecules [SN38, camptothecin, febuxostat, T4 (6,7-dimethoxy-2-((4-phenyl-3, 6-dihydropyridin-1 (2H)-yl)methyl) quinazoline- 4(3H)-one), T5 (3-cyclobutyl-6- (2-naphthylmethyl)-2,7-dihydro-4H- pyrazolo[3,4-b]pyridin-4-one)] as hit compounds (Figure 3E and F). SN38 and camptothecin (CPT) are topoisomerase inhibitors (TOPOi), which were previously reported to be APA modulators (Dutertre et al., 2014). This indicated that the initial cell-based reporter screen was sufficiently sensitive and specific to detect APA modulators. I noted that T4 and T5 have chemically very distinct scaffolds.

Hit compounds T4 and T5 modulated endogenous APA via nonkinase-mediated pathways

To examine whether the identified hit compounds would regulate endogenous APA, APA modulation at the Dicer gene locus was validated using 3'-end qPCR assays (Jenal et al., 2012), which independently measure proximal and distal isoforms. TOPOi decreased poly(A) mRNAs in a dose-dependent manner at the proximal PASs, relative to at the distal PASs. This change, a proximal-to-distal (PtoD) shift, lengthened 3'-UTR (Figure 3G). T4 and T5 also induced a PtoD APA shift at the Dicer locus in 293T and U2OS cells (Figure 3G and H). However, febuxostat, a xanthine oxidase inhibitor, did not modulate Dicer APA (Figure 3G). T4 showed 4-fold greater APA modulatory activity in U2OS (IC_{50}: 2.1 μM) than in 293T cells (IC_{50}: 8.5 μM). In contrast, TOPOi had APA modulatory activity in 293T cells (IC_{50}: 1.5–3.8 μM) that was more than 10-fold greater than that in U2OS cells (IC_{50}: 19.4 μM). Taking these findings together, four of the five hit compounds modulated endogenous APA and two were T4 and T5, which have chemically distinct scaffolds. The activities of T4 and T5 compounds were subsequently assessed for the inhibition of 414 protein kinases. The results showed that T4 did not inhibit any of the kinases, whereas T5 inhibited CDK1, 2, 3, 5, 14, and 16 kinases by more than 80% (Figure 4C), identifying it as a multi-CDK family inhibitor. The 45,000-compound library used for the initial screen included several multi-CDK inhibitors such as AZD5597 (CDK1/2/9 inhibition), purvalanol A (CDK2/4 inhibition), and PHA-767491 (CDK7/9 inhibition). These compounds did not affect APA reporter activity in the screen. Furthermore, I confirmed that APA reporter activity was not affected by other multi-CDK inhibitors, flavopiridol, SNS-032, or AT7519 (Figure 4D). These results suggested that T4 and T5 modulated APA via non-CDK signaling.

To further explore the mode of action of T4 and T5, I examined APA site expression for additional genes (An et al., 2008; Jenal et al., 2012; Lackford et al., 2014; Yao et al., 2013).

Treatment with T5 significantly modulated APA at the RCC2, BDNF, p27, and MRPS18C gene loci. T4 had a similar pattern of APA modulation, while CPT had a completely different pattern (Figure 4E). In the expanded gene testing, febuxostat also did not modulate APA.

T4 and T5 predominantly induced CR-APA and UTR-APA

To comprehensively characterize alternative splicing and APA, I performed RNA-seq analysis of U2OS cells after treatment with T4, T5, or CPT. All RNA-seq analyses were performed after 6 h of treatment, avoiding cytotoxicity, as assessed by annexin V/PI staining to detect early-phase apoptosis (Figure 5A). The workflow for sequencing, alignment, and informatics quantification analysis is described in Figure 5B. Principal component analysis of FPKM values revealed that gene expression profiles induced by T4 were quite similar to those induced by T5 (Figure 5C). In contrast, profiles induced by CPT were very different from those induced by T4 or T5. I classified each splicing event, such as alternative 3' splice sites (A3SS), alternative 5′ splice sites (A5SS), alternative first exons (AFE), alternative last exons (ALE), mutually exclusive exons (MXE), retained introns (RI), and skipped exons (SE), using MISO analysis (Katz et al., 2010). The proportion of ALE (CR-APA) events was predominant among the splicing event types after exposure to T4, T5, or CPT (6%, 6%, and 4%, respectively) (Figure 6A). To further examine the effects of the compounds in PAS, 3'-seq was performed (Beck et al., 2010; Elkon et al., 2012) to map 3'-end cleavage and poly(A) CSs at single-nucleotide resolution. I applied 3'-seq to samples treated with the same concentration and treatment duration (6 h) of T4, T5, or CPT; approximately 48%–64% of reads were mapped to the transcripts, consistent with previous reports of studies in which this approach was used (Elkon et al., 2012; Jenal et al., 2012). The degree of difference of ALE or 3'-UTR usage among samples was quantified as a change in the percentage distal PAS usage

index (ΔPDUI). This can distinguish bias toward proximal CS (negative index) from that toward distal CS (positive index) usage. The total number of differential ΔPDUI for T4 and T5 (955 at 10 μM T4, 1002 at 5 μM T5) was greater than for CPT (489 at 5 μM), with the highest rate of increase caused by 2 μM T4 or 0.5–2 μM T5 (Figure 6B). All three compounds induced a DtoP shift. In particular, 85%–94% of DtoP APA change covered differential ΔPDUI at all of the concentrations of applied T5.

To determine whether sequence context reflected CS usage, I examined the enriched sequence motifs in the region 50 nt upstream of CSs. The canonical AATAAA PAS was successfully identified by HOMER motif enrichment analysis (Heinz et al., 2010) in samples under all treatment conditions, for both DtoP and PtoD APA utilization (Figure 6C). The AATAAA motif was significantly enriched in PtoD CSs compared with that in DtoP CSs, as determined by hypergeometric tests (P-values $< 1 \times 10^{-34}$ for all of the T4 test set; $P < 1 \times 10^{-31}$ for T5 and $P < 1 \times 10^{-10}$ for CPT) (Figure 6C and D).

To examine whether read-through transcripts were induced by the compounds, I performed RNA-seq analysis using STAR-fusion (Dobin et al., 2013). The term "read-through transcript" refers to a conjoined gene arising from the upstream transcript to the downstream transcript of partner genes with poly(A) sites (Akiva et al., 2006; Prakash et al., 2010). RNA-seq data showed that the number of conjoined genes did not change, compared with that in the control, indicating that the compounds did not induce conjoined genes (Figure 5D and E), in contrast to inhibitors of CLK family proteins (Funnell et al., 2017). Taken together, these results suggested that T4 and T5 modulate both CR-APA and UTR-APA of an extensive, but specific, set of genes and mainly promote the use of proximal PAS CSs.

Conserved APA utilization and transcriptional responses revealed by graded pharmacological response pattern clustering

To examine the relationships between APA changes and transcript responses, I performed clustering of weighted gene co-expression network analysis (WGCNA) on the PDUI profile. In the 3'-seq dataset from cells treated with T4, T5, or CPT, I identified 9, 5, and 5 different PDUI profile clusters, respectively (Figure 7A and B and Figure 8A-C). Data obtained after T4 treatment revealed three dominant clusters containing 52% of the clustered peak events (489, 392, and 220 events). After T5 treatment, there were two dominant clusters containing 89% of the clustered peak events (1292 and 360 events). Clusters 1, 2, and 3 induced by T4 showed monotonically increasing or decreasing responses, except for at 20 μM. The other clusters showed nonmonotonic responses in a small number of genes, implying that higher concentrations of the compound and nonmonotonic patterns captured secondary effects of deregulated transcripts, splicing, and polyadenylation. Treatment with T5 induced cluster patterns similar to those observed with T4. Clusters 1 and 3 showed monotonic responses and Cluster 2 showed semimonotonic responses, although, at above 2 μM T5, there were slightly increasing responses.

I next examined which region of the transcript was used for CR-APA and UTR-APA regulation, among the response clusters, by the compounds. Clusters with a decreasing PDUI in response to T4 (Clusters 1 and 2) predominantly included an intronic region (53% and 62%, respectively), whereas those with an increasing PDUI in response to T4 (Cluster 3) contained exons at a rate of 44% and 3'-UTR regions at a rate of 32% (Figure 7A). Clusters with a decreasing PDUI in response to T5 (Clusters 1 and 2) included an intronic region at rates of 63% and 71%, respectively. Clusters with an increasing PDUI in response to T5 (Cluster 3) contained exons at a rate of 25% and 3'-UTR regions at a rate of 33%, although

the number of genes affected was limited (Figure 7B). Because a decrease in PDUI value indicates a change in DtoP APA, this strongly suggested that clusters with a decreasing PDUI represented genes in which predominantly intronic polyA, not 3'-UTR, was used for APA regulation.

To determine which biological pathways were affected by changes in APA caused by the compounds, I performed functional enrichment analysis using each set of clusters. Only clusters with a decreasing PDUI upon treatment with T4 (Clusters 1 and 2) or T5 (Clusters 1 and 2) showed significantly enriched biological processes, as determined using the hypergeometric test (P-values $< 1 \times 10^{-4}$) (Figure 7C and D). Both compounds induced the regulation of similar pathways, namely, those involved in the cell cycle, metabolism, and transcription. In addition, T4 enriched splicing and T5 enriched the 3'-end RNA processing pathway. The pathway enriched with monotonically decreasing cluster (Cluster 1) was significantly correlated with that with Cluster 1 by T5, as determined using Pearson's test (P-values $< 1 \times 10^{-16}$) (Figure 7F). Conversely, CPT treatment did not enrich an RNA processing pathway (Figure 7E). These results suggested that T4 and T5 affected common pathways, including that for RNA processing.

Potential autoregulation of 3'-processing and transcription pathways by APA modulators

To examine how APA changes affected transcripts, I also performed WGCNA clustering using FPKM profiles of cells treated with T4 or T5 (Figure 9A). Seven clusters were identified from T4 treatment and six clusters from T5. T4 treatment resulted in three dominant clusters (Clusters 1–3) with a monotonic pattern and T5 also showed three dominant clusters (Clusters 1–3). Next, biological pathway analysis was performed, using functional enrichment analysis, on each set of dominant clusters. Only clusters with a

decreasing FPKM in response to T4 (Clusters 1 and 2) or T5 (Clusters 1 and 3) showed significantly enriched biological processes (Figure 9C and D). Like enrichment analysis using PDUI, the enriched pathways responding to T4 and T5 treatments were very similar. Strong similarities in clustered FPKM response patterns were also observed between T4 and T5 treatments, and PDUI response patterns were also relatively similar, although to a lesser extent (Figure 10A). Cell cycle/mitosis, chromosome organization, and DNA repair pathways were enriched with both T4 and T5 treatments (Figure 9C and D). Interestingly, splicing, transcription, and RNA processing pathways were also enriched. These results implied that the compounds targeted APA in genes related to the RNA processing pathway, thereby decreasing their expression. Hence, to investigate the relationship between APA change and gene expression, I compared PDUI and FPKM values (Figure 10B). Genes downregulated by 2 μM T4 or T5 accounted for 53% or 61%, respectively, of genes with decreasing PDUI (T4: 382 genes, T5: 656 genes). This suggested that most genes whose APA sites were altered had decreased expression.

To understand how APA changes affected the expression of transcripts for genes related to RNA processing, I performed hierarchical clustering analysis on PDUI values for 22 critical genes as trans-acting polyadenylation factors (Xia et al., 2014). With T4 treatment, genes showing dose-dependent decreases in PDUI were clustered together and included PCF11, RBBP6, PAPOLA, NUDT21, and CSTF3 (Figure 11A). With T5 treatment, genes showing decreased PDUI were also clustered together and included PCF11, CSTF3, RBBP6, and NUDT21 (Figure 11B). These results were consistent with the pathway analysis shown in Figure 7C and D. I chose PCF11, RBBP6, CSTF3, and WDR33 genes, whose PDUI values were significant (> 0.4), and examined their expression. The FPKM values for the four genes were decreased in a dose-dependent manner (Figure 11C), in accordance with the pathway

analysis results shown in Figure 9C and D. CR-APA was changed for both RBBP6 and PCF11 (Figure 12) and 3'-end qPCR analysis revealed that both T4 and T5 induced DtoP APA changes in PCF11 and RBBP6 genes (Figure 11D).

Motif elements in APA regulation and auto-regulation of PABPN1 signaling by APA modulators

I next examined whether the cleavage regions of the genes undergoing monotonic responses by APA modulators (T4, T5, or CPT) would show distinct patterns of recognition sites for RNA binding proteins. I searched the upstream regions of proximal or distal CS using the monotonic response Clusters 1–3 with PDUI for 114 human/mouse RNA binding motifs (Paz et al., 2014). Motifs of the PABP and KHDRBS families had A-rich sequences clustered together and were highly enriched in APA-responsive regions (Figure 13A). PABP family motifs were enriched in the proximal regions, whereas motifs of the KHDRBS family exhibited the opposite pattern. I then performed *de novo* motif enrichment analysis against the cleavage regions of the genes with monotonic responses induced by an APA modulator (T4, T5, or CPT) (Figure 13B and 14A). The canonical PAS, AATAAA, was strongly enriched in clusters 1–3, particularly at the distal sites, as determined by the hypergeometric test (P-values $< 1 \times 10^{-5}$ for all distal test sets). Interestingly, the AAAAAA motif was significantly enriched in the proximal sites of genes with monotonic decreases in PDUI, Cluster 1 for T4 and T5 ($P < 1 \times 10^{-8}$ and $P < 1 \times 10^{-22}$, respectively), whereas the A-rich motif for CPT treatment was not enriched. The location of the AATAAA PAS motif peaked at the expected positions upstream of the putative CSs (-35 nt) and the AAAAAA motif was located at a more downstream site (-15 nt) than the putative PAS site (Figure 14B).

I next examined CR- and UTR-APA regulation by knockdown of RNA binding

proteins and 3'-end processing factors, to determine which molecules were targeted by T4 or T5. I chose 20 genes whose binding motifs were enriched after treatment with T4 or T5 and whose silencing was reported to distinctly regulate APA (Li et al., 2015). I first confirmed that 17 out of 20 siRNAs showed > 80% knockdown efficiency in U2OS cells (Figure 14C). A total of 11 APA events were measured by 3'-end qPCR, using samples from cells transfected with each siRNA. Multiple APA events were significantly altered by knockdown of the corresponding genes (> 2SD above the mean control condition). Hierarchical clustering analysis, using APA expression values, revealed that depletion of PABPN1 clustered together with T4 or T5 treatment (Figure 13C). Subsequently, I comprehensively compared the APA events by T4 or T5 with the previously reported APA events by PABPN1 knockdown in the same U2OS cells (Jenal et al., 2012). Since knockdown of PABPN1 has been reported to predominantly (91.4%) induce DtoP APA shift, I analyzed DtoP APA events. DtoP APA events (Clusters 1 and 2) that were monotonically responsive upon exposure to T4 or T5 overlapped with DtoP APA events induced by PABPN1 knockdown in a statistically significant manner, as determined by Fisher's exact test (P=0.0045 for T4, P=7.4 $\times$ 10^{-5} for T5) (Figure 14D). PABPN1 was reported to control its protein levels through an autoregulatory mechanism, which is required for the retained-intron type of splicing, and an increase in intron 6-retained pre-mRNA was found to promote a decrease in PABPN1 protein levels (Bergeron et al., 2015) (Figure 13D). I calculated PSI (percent spliced in) value distributions for the retained-intron type of PABPN1 at each T4 or T5 concentration, using MISO with the RNA-seq dataset. Both T4 and T5, in dose-dependent manners, induced intron 6-retained pre-mRNAs with high probability (Bayes factor > 10^{12}) (Figure 13E). PABPN1 protein levels after 24 h of treatment were suppressed by T4 or T5 treatment, also in a dose-dependent manner, consistent with the PSI value for 24 h of treatment (Figure 13F

and 14E). PABPN1 knockdown induced a > 10-fold change in DtoP APA, compared with that in controls, in the loci of PCF11 and RCC2 genes (Figure 13G). T4 or T5 treatment after the knockdown of PABPN1 further induced an APA change in the locus of the PCF11 gene (26- and 16-fold for T4 and T5, respectively), consistent with complete depletion of PABPN1 protein level using both PABPN1 siRNA and compounds (Figure 14F). In contrast, T4 or T5 treatment of cells, after knockdown of PABPN1, maintained APA status in the RCC2 gene (1.8- and 1.4-fold increased DtoP change for T4 and T5, respectively). Taken together, these results suggested that T4 and T5 are involved in autoregulated PABPN1 signaling and affect the RNA processing pathway.

Poly(A) sites in long introns and transcripts were vulnerable to APA modulators

I next examined whether T4 or T5 targets U1 snRNP and RNA polymerase II elongation rate because much of the APA was a DtoP shift in the intron region (Figure 7A and B). I examined the APA events, regulated by U1 snRNP, after treatment with T4 or T5 in HeLa cells as used in previous studies (Berg et al., 2012; Kaida et al., 2010). I confirmed that treatment with T4 or T5 modulated APA at the PCF11 loci as a positive control. However, T4 or T5 treatment did not modulate APA sites other than the NR3C1 gene (Figure 15A). Next, I examined whether T4 or T5 regulates APA and splicing through the RNA pol II elongation rate. I selected the 16 genes whose APA and splicing sites were affected by elongation rate in 293 cells (Fong et al., 2015; Fong et al., 2014). I confirmed that treatment with T4 or T5 did not modulate six sites of APA other than Jun, but modulated APA at the PCF11 loci as a positive control in 293 cells (Figure 15B). Splicing analysis using RT-PCR revealed that T4 and T5 partly altered splicing in three to four of the nine genes that I tested (Figure 15C). These results suggest that T4 and T5 do not directly regulate the activity of U1

snRNP and the RNA pol II elongation rate.

I next examined which types of CSs were targeted by APA modulators. First, intron size was investigated. I confirmed that introns with CSs were larger than those without CSs, using the Welch's *t* test (*P*-values $< 1 \times 10^{-50}$), an observation consistent with a previous report (Tian et al., 2007) (Figure 16A). Interestingly, introns with CSs that were induced by T4, T5, or CPT were significantly larger than those that were not affected by the compounds (*P*-values $< 1 \times 10^{-5}$) (Figure 16A and 15D). Next, I compared transcript lengths of all human genes, using those that were upregulated, downregulated, or APA-modulated after treatment with the compounds (Figure 16B and 15E). With all compounds, T4, T5, and CPT, there were no significant differences in transcript length among the upregulated genes. However, among the downregulated and APA-modulated genes, transcript lengths were significantly larger than the mean values for all transcripts. The genes with decreased expression and altered CSs in cells treated with the compounds had higher numbers of exons than the mean values for all human genes (Figure 16B and 15E). These results suggested that T4, T5, and CPT mainly targeted vulnerable poly(A) sites in long introns and transcripts.

Discussion

Using reporter-based compound screening, I identified two small molecules that can modulate extensive, specific CS of APA. Multiple RNA binding proteins (RBPs) are involved in APA regulation; however, it has been observed that, on rare occasions, there are pockets for RBP binding to small molecules. This is supported by the fact that, to date, only evidence that a topoisomerase inhibitor can act as an APA modulator has been reported. Additionally, a limited number of hit compounds were identified from my screening. Luciferase screens generally detect many false positive compounds owing to the toxicity of compounds and their effects on luciferase itself (Heitman et al., 2008). Therefore, in my screening, both a reporter screen and a counterscreen were performed using psiC. The identified hit compounds including topoisomerase inhibitors satisfied the screen hit criteria, albeit to a lesser extent, and four out of five compounds modulated endogenous APA. These findings indicate that my reporter screen works well. All four hit compounds, SN38, CPT, T4, and T5, inhibited cell growth. However, other compounds in the library with stronger growth-inhibitory activity than the hit compounds did not have APA-modulating activity. Additionally, short exposure (6 h) at increasing concentrations modulated APA. These results clearly demonstrate that the APA modulation detected in my experiments did not depend on a secondary effect of cytotoxicity.

T4 and T5 have chemically very distinct structures; however, these compounds affect common pathways, such as the RNA processing pathway, as determined using functional enrichment analysis of both APA and gene expression. Additionally, motif enrichment analysis against the cleavage regions of the genes revealed that T4 and T5 enriched similar motif sequences. In regard to the chemical structure, T5 has a substructure

that can act as a hinge binding motif of kinases, but T4 does not. This structural feature is supported by the kinase panel data showing that T5 is a pan-CDK inhibitor, while T4 does not inhibit any kinases. However, my analysis revealed that several pan-CDK inhibitors did not modulate APA, as shown in Figure 4D. These results suggest that T5 modulates APA through a nonkinase- or unknown kinase-mediated pathway that is not covered by the kinase panel. It is conceivable that T4 and T5 bind to different molecules in the same signaling pathway regulating APA.

Cells treated with T4 or T5 were predominantly shifted to using proximal intronic poly(A) CSs. My analysis showed that the retained-intron type of splicing, like for PABPN1, was induced by T4 and T5, in a dose-dependent manner, albeit to a lesser extent than was ALE (CR-APA), as shown in Figure 6A. Presumably, induced intronic poly(A) mRNA retained the transcripts. This result raised the possibility that the compounds identified in my study target U1snRNP or transcription elongation rate. However, the expression of U1 snRNP-related genes, such as SNRNP70, SNRPA, and SNRPC, was not changed by the exposure of cells to the compounds and the APA events regulated by U1 snRNP and RNA pol II elongation rate did not change markedly. In contrast, the expression of several genes related to U2 snRNP (SF3A1, SF3B1, and SF3B3) was decreased. It is conceivable that downregulation of U2 snRNP by the compounds indirectly affected several splicing alterations and APA changes. Recent detailed RNA-seq analysis revealed that treatment with the SF3B inhibitor spliceostatin A (SSA) preferentially caused the retained-intron type of splicing (Yoshimoto et al., 2017), while CLK inhibitor-modulated splicing mainly induced skipped exons (Araki et al., 2015; Funnell et al., 2017). The ratio of splicing patterns was completely different with APA modulators (T4, T5, and CPT) than with splicing modulators (SSA and CLK inhibitor). These observations provided convincing evidence that T4 and T5

preferentially modulate APA and indirectly affect APA via U2 snRNP.

According to my findings, T4 and T5 appeared to target the poly(A) sites in long introns. Large intron size regulates the usage of terminal exon poly(A) sites (Tian et al., 2007). Large transcripts generally require more time to transcribe and splice out, and for polyadenylation of the RNA. Presumably, the long introns targeted by T4 or T5 were not only influenced by the time required for transcription but also regulated by some form of vulnerability of their sequences, such as A abundance. Motif enrichment analysis of monotonically increasing or decreasing dominant APA clusters indicated that T4 and T5 treatment targeted A-rich regions downstream of CSs in the proximal region, whereas inverted bell-shaped dose response clusters (Cluster 2) showed an association with enrichment of canonical AATAAA motif in DtoP CSs. Likewise, motif analysis using PtoD clusters (Cluster 3) revealed enrichment in AATAAA motif, as shown in Figure 14A. However, interestingly, the AAAAAA motif was located at a characteristic upstream site (−15 nt) of the putative PAS site, even in the distal sites of Cluster 3 (Figure 14B). These results imply that the A-rich sequence is targeted by T4 and T5 even at the distal site, but to a lesser extent. In addition to motif analysis, data obtained with a panel of siRNAs suggested that PABPN1 was in part involved in the signaling affected by T4 and T5. T4 or T5 exposure predominantly induced the usage of proximal CSs, consistent with the loss of PABPN1 (Jenal et al., 2012). Increased retention of introns of PABPN1, similar to that occurring during its autoregulation, was observed after treatment with T4 or T5. Autoregulation is accomplished via the binding of PABPN1 to an A-rich region upstream of the polyA site (Bergeron et al., 2015). These considerations were consistent with the results of my analyses, indicating that A-rich regions were preferred by T4 and T5.

In conclusion, the identified small-molecule compounds, T4 and T5, had high

potency and selectivity for APA modulation, inducing a preference for the use of proximal intronic poly(A) CSs with long intron regions through the PABPN1 pathway. These compounds should become valuable tools for increasing our understanding of the molecular mechanisms of APA and splicing because, unlike siRNAs, they can be applied to multiple cells and animals at various concentrations and with different exposure times. Further studies to identify the specific mode of binding of the compounds to their target proteins may elucidate their therapeutic potential.

Figures

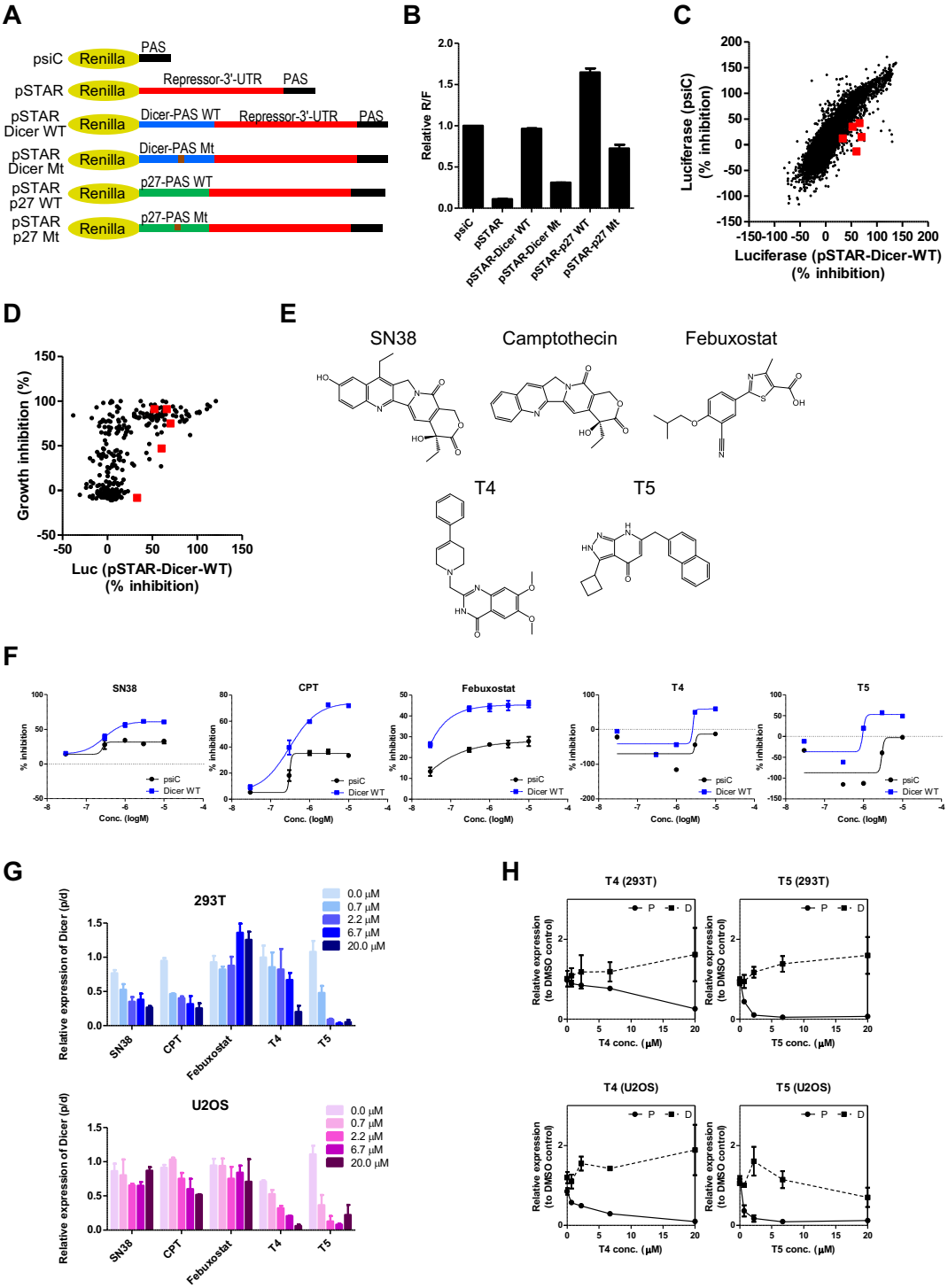
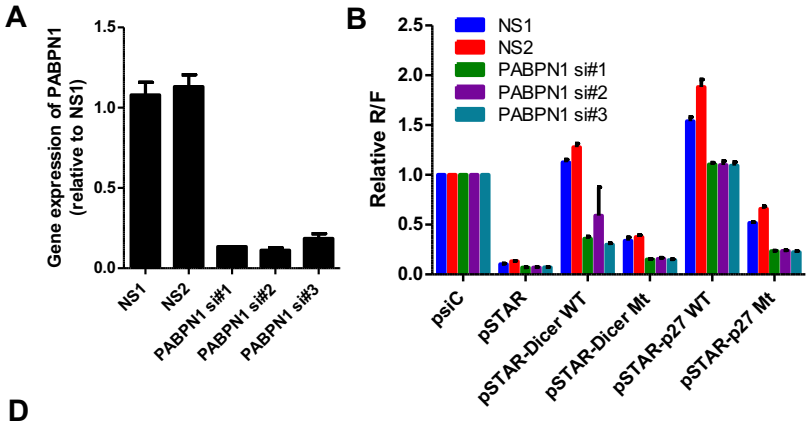

A
psiC
Renilla
PAS
pSTAR
Repressor-3'-UTR
pSTAR Dicer WT
Dicer-PAS WT
pSTAR Dicer Mt
Dicer-PAS Mt
pSTAR p27 WT
p27-PAS WT
pSTAR p27 Mt
p27-PAS Mt
B
Relative R/F
psiC
pSTAR
pSTAR-Dicer WT
pSTAR-Dicer Mt
pSTAR-p27 WT
pSTAR-p27 Mt
C
Luciferase (psiC) (% inhibition)
Luciferase (pSTAR-Dicer-WT) (% inhibition)
D
Growth inhibition (%)
Luc (pSTAR-Dicer-WT) (% inhibition)
E
SN38
Camptothecin
Febuxostat
T4
T5
F
SN38
CPT
Febuxostat
T4
T5
% inhibition
Conc. (logM)
psiC
Dicer WT
G
293T
U2OS
Relative expression of Dicer (p/d)
0.0 μM
0.7 μM
2.2 μM
6.7 μM
20.0 μM
SN38
CPT
Febuxostat
T4
T5
H
T4 (293T)
T5 (293T)
T4 (U2OS)
T5 (U2OS)
Relative expression (to DMSO control)
P
D
T4 conc. (μM)
T5 conc. (μM)

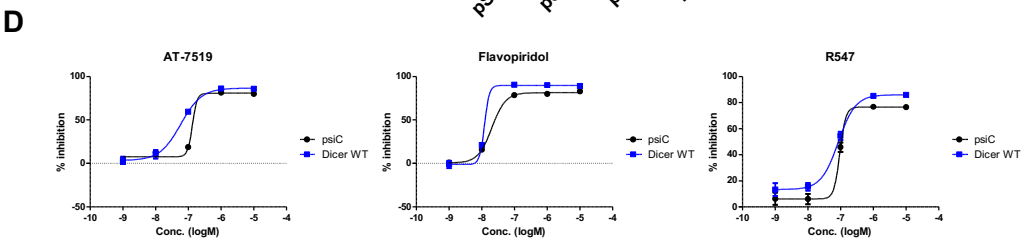
Figure 3. Modulators of endogenous APA identified by reporter-based compound screening

(A) Schematic diagram of pSTAR luciferase reporters containing the wild-type (WT) or mutant (Mt) proximal Dicer PAS upstream of the KMT2A-3'-UTR region and the WT or Mt proximal p27 PAS upstream of the KMT2A-3'-UTR region. (B) Each pSTAR plasmid was transfected into 293T cells for 48 h. Relative R/F (Renilla/firefly) ratios were determined. Data are mean ± SD from three independent analyses. (C) Scatterplots showing reporter inhibitory activities (% inhibition) of pSTAR-Dicer WT relative to pSTAR (x-axis) and psiC relative to pSTAR (y-axis) in a 1536-well format. The red box indicates the five hit compounds. (D) Scatterplot showing reporter activities of pSTAR-Dicer WT (x-axis) and cell growth inhibitory activities (% inhibition) (y-axis). The red box indicates the five hit compounds. The 293T cells were treated with 3 μM of each compound for 24 h (reporter assay) or 72 h (cell growth assay). (E) Structures of the hit compounds. (F) Luciferase activities after treatment with each hit compound. The x-axis shows the compound concentration (logM). The y-axis shows % inhibition of luciferase activity using psiC or pSTAR-Dicer WT constructs. (G, H) Results of 3'-end qPCR. (G) Expression of proximal (p) relative to distal (d) Dicer mRNA in 293T or U2OS cells treated with the indicated compound for 24 h. (H) Expression of proximal and distal Dicer mRNA in 293T or U2OS cells treated with T4 or T5 for 24 h. Data are mean ± SD from three independent experiments.



C

Enzyme	% inhibition T4	T5
CDK1/cyclin A2	7	86
CDK2/cyclin A1	14	97
CDK2/cyclin E1	7	99
CDK2/cyclin O	20	99
CDK3/cyclin E1	10	97
CDK5	12	82
CDK14/cyclin Y	3	80
CDK16/cyclin Y	5	90



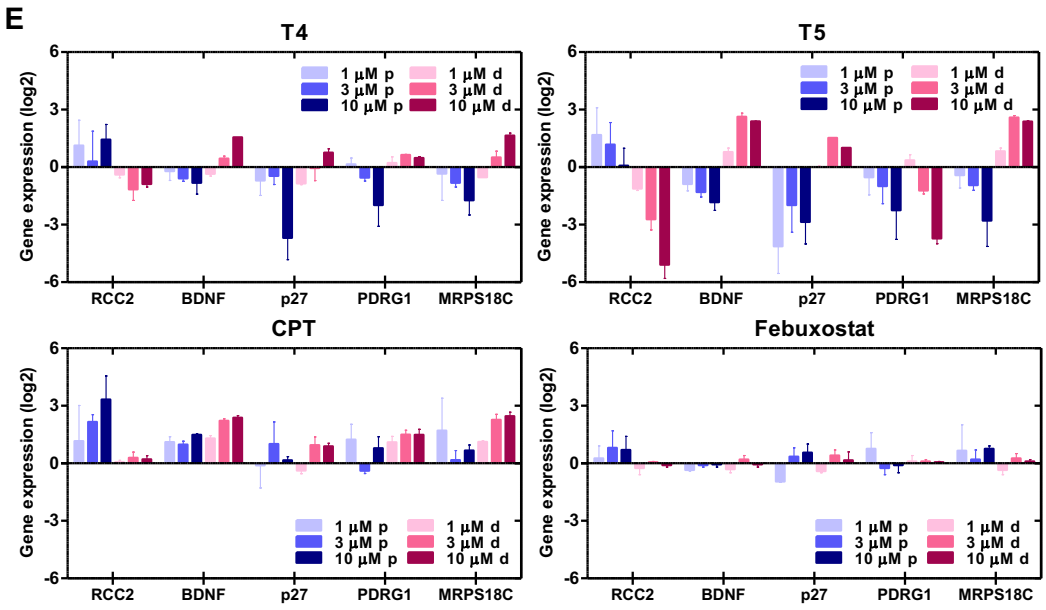


Figure 4. APA modulators identified by compound screening

(A, B) PABPN1 siRNA #1, #2, or #3 (three independent siRNA sequences) or control nonsilencing siRNA (NS) #1 or #2 (independent sequences) was transfected into 293T cells. (A) Relative PABPN1 mRNA levels (qRT-PCR). (B) Relative R/F ratios for pSTAR series in cells transfected with indicated PABPN1 or NS siRNA. Data are mean ± SD from three independent experiments. (C) Kinase panel for T4 or T5 (1 μM). (D) Luciferase activity in 293T cells treated with the indicated compounds. The x-axis shows compound concentration (logM). The y-axis shows % inhibition of luciferase activity in cells transfected with either psiC or pSTAR-Dicer WT. (E) 3'-End qRT-PCR analysis for 10 APA sites consisting of five genes in U2OS cells treated with the indicated compounds for 24 h. Gene expression of proximal and distal mRNAs (log2 scale) following treatment with the indicated compounds (1, 3, or 10 μM) for 24 h in U2OS cells. Data are mean ± SD from two independent experiments.

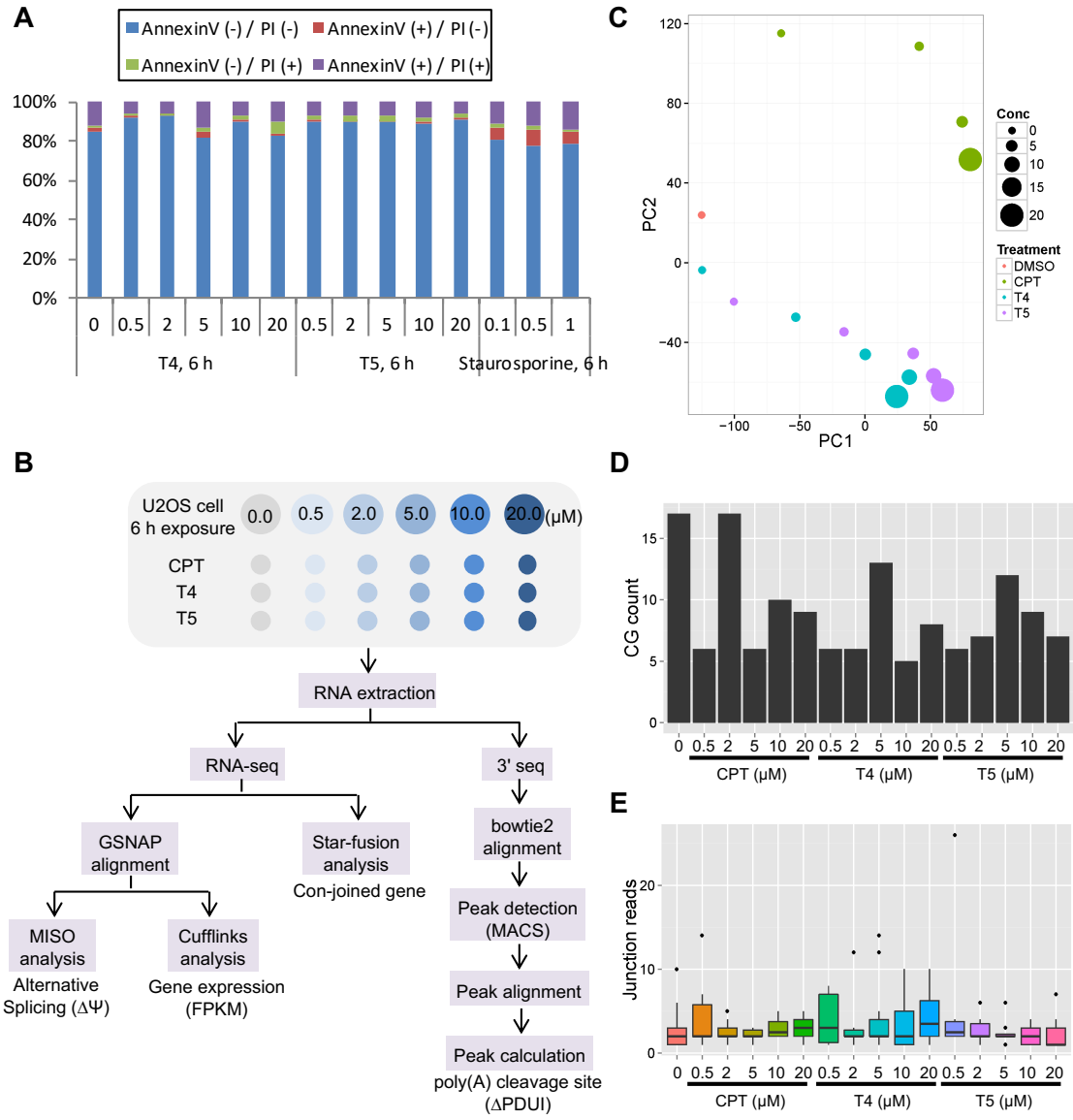
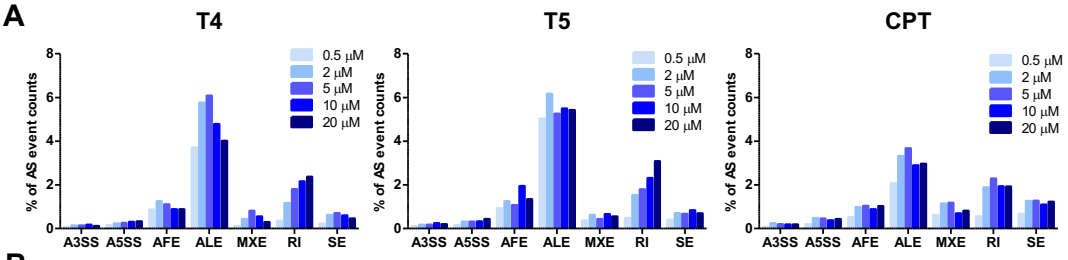

A
AnnexinV (-) / PI (-)
AnnexinV (+) / PI (-)
AnnexinV (-) / PI (+)
AnnexinV (+) / PI (+)
100%
80%
60%
40%
20%
0%
0 0.5 2 5 10 20 0.5 2 5 10 20 0.1 0.5 1
T4, 6 h
T5, 6 h
Staurosporine, 6 h
C
PC2
PC1
Conc
0
5
10
15
20
Treatment
DMSO
CPT
T4
T5
B
U2OS cell
6 h exposure
0.0 0.5 2.0 5.0 10.0 20.0 (μM)
CPT
T4
T5
RNA extraction
RNA-seq
3' seq
GSNAP alignment
Star-fusion analysis
Con-joined gene
bowtie2 alignment
MISO analysis
Alternative Splicing (ΔΨ)
Cufflinks analysis
Gene expression (FPKM)
Peak detection (MACS)
Peak alignment
Peak calculation
poly(A) cleavage site (ΔPDUI)
D
CG count
0 0.5 2 5 10 20 0.5 2 5 10 20 0.5 2 5 10 20
CPT (μM)
T4 (μM)
T5 (μM)
E
Junction reads
0 0.5 2 5 10 20 0.5 2 5 10 20 0.5 2 5 10 20
CPT (μM)
T4 (μM)
T5 (μM)

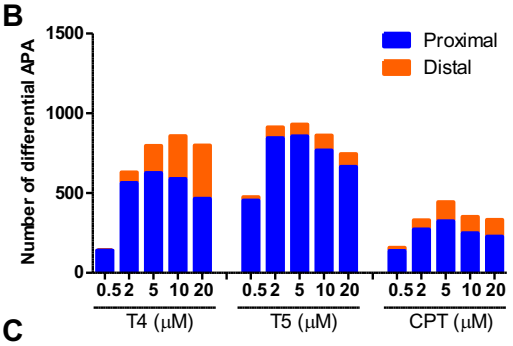
Figure 5. Comprehensive transcriptome analysis (RNA-seq and 3'-seq) by graded and short exposure to T4 or T5

(A) U2OS cells were treated with T4 and T5 for 6 h at the indicated concentrations. The cells were stained with annexin V-FITC and propidium iodide (PI) and analyzed by FACS. Percentages of each fraction, determined by FACS, are shown. (B) Workflow diagram for NGS analysis using the APA modulator. (C) Principal component analysis (PCA) plot of FPKM gene expression for treatment with APA modulators, T4, T5, or CPT. PCA was performed with genes whose FPKM values were > 0.5 in at least one sample. The first two principal components (PC1 and PC2) are plotted on the x-axis and y-axis, respectively. Each sample is represented by a single color-coded circle. The size of each circle indicates the concentration of each compound. (D) The number of conjoined genes (CGs) after treatment with CPT, T4, or T5. (E) Box plots of the distribution of CG junction reads per condition.

A



B



C

T4 Proximal

Rank	Motif	*p*-value
1	AATAAA	1.0E-34
2	TTTGAA	1.0E-14
3	CAGCCT	1.0E-10

T4 Distal

Rank	Motif	*p*-value
1	AATAAA	1.0E-66
2	CCAGTG	1.0E-11
3	CTGCAG	1.0E-09

T5 Proximal

Rank	Motif	*p*-value
1	AATAAA	1.0E-31
2	CTGCAG	1.0E-15
3	ACACTG	1.0E-15

T5 Distal

Rank	Motif	*p*-value
1	AATAAA	1.0E-80
2	GTGTCT	1.0E-14
3	CTGCCT	1.0E-10

CPT Proximal

Rank	Motif	*p*-value
1	AATAAA	1.0E-10
2	CAGCTG	1.0E-08
3	TTCAAA	1.0E-07

CPT Distal

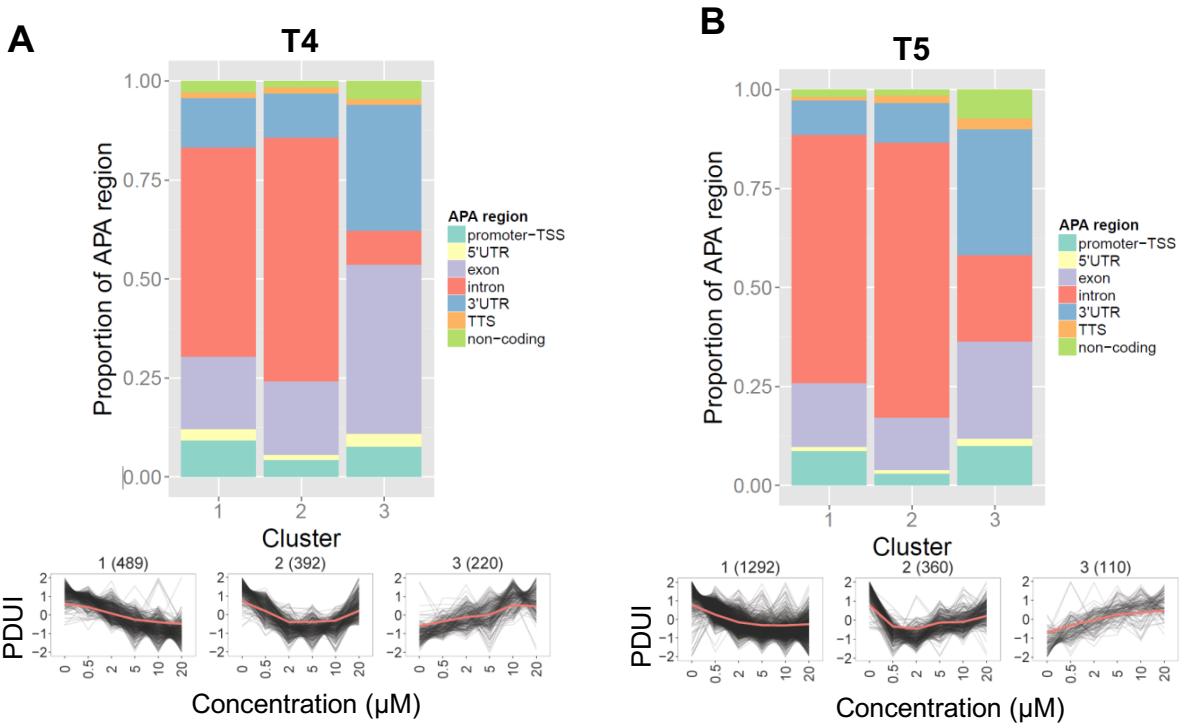
Rank	Motif	*p*-value
1	AATAAA	1.0E-38
2	CAGCTG	1.0E-09
3	GGGTCC	1.0E-08

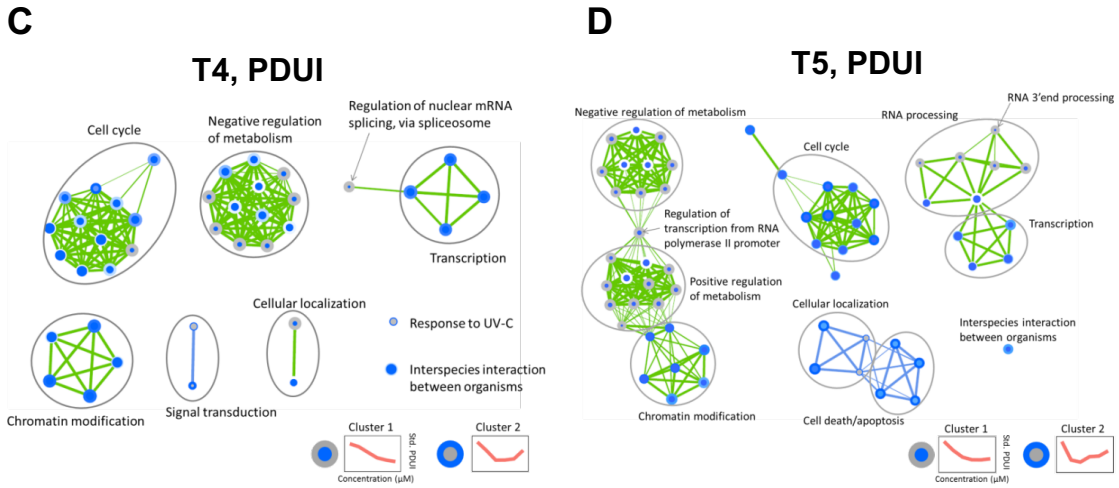
D

	T4		T5		CPT	
AATAAA motif	Proximal	Distal	Proximal	Distal	Proximal	Distal
Hit	291	306	477	335	126	146
Query	954	684	1148	756	474	366
% of hit	30.5	44.7	41.6	44.3	26.6	39.9

Figure 6. CR-APA and UTR-APA regulation by T4 and T5

(A) The proportion of each type of alternative exon regulated by T4, T5, and CPT using RNA-seq among the annotated splicing events. Alternative 3' splice sites (A3SS), alternative 5' splice sites (A5SS), alternative first exons (AFE), alternative last exons (ALE), mutually exclusive exons (MXE), retained introns (RI), and skipped exons (SE) are shown. (B) The number of differential APAs mediated by T4, T5, and CPT using 3'-seq. (C, D) Canonical polyA motif AATAAA with a significant P-value from upstream (−50 bp) of the proximal poly(A) or the distal poly(A) CSs induced by T4, T5, or CPT, using 3'-seq.


A
T4
Proportion of APA region
1.00
0.75
0.50
0.25
0.00
1
2
3
Cluster
APA region
promoter−TSS
5'UTR
exon
intron
3'UTR
TTS
non−coding
1 (489)
2 (392)
3 (220)
PDUI
Concentration (μM)
B
T5
Proportion of APA region
Cluster
APA region
promoter−TSS
5'UTR
exon
intron
3'UTR
TTS
non−coding
1 (1292)
2 (360)
3 (110)
PDUI
Concentration (μM)


C
T4, PDUI
Cell cycle
Negative regulation of metabolism
Regulation of nuclear mRNA splicing, via spliceosome
Transcription
Cellular localization
Response to UV-C
Interspecies interaction between organisms
Chromatin modification
Signal transduction
Cluster 1
Cluster 2
Std. PDUI
Concentration (μM)
D
T5, PDUI
Negative regulation of metabolism
RNA processing
RNA 3'end processing
Cell cycle
Regulation of transcription from RNA polymerase II promoter
Transcription
Positive regulation of metabolism
Cellular localization
Interspecies interaction between organisms
Chromatin modification
Cell death/apoptosis
Cluster 1
Cluster 2
Std. PDUI
Concentration (μM)

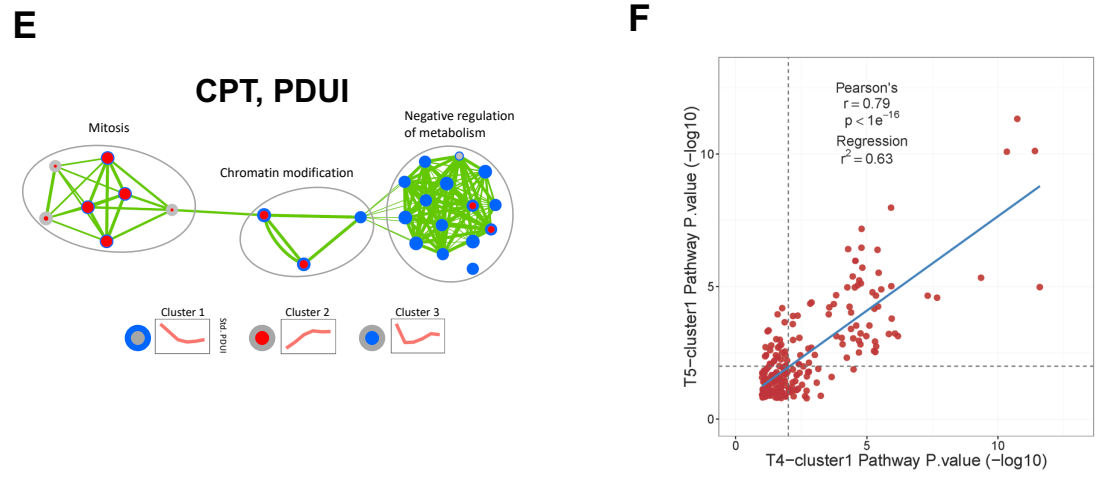

E
CPT, PDUI
Mitosis
Negative regulation of metabolism
Chromatin modification
Cluster 1
Cluster 2
Cluster 3
Std. PDUI
F
Pearson's
r = 0.79
p < 1e−16
Regression
r² = 0.63
T5−cluster1 Pathway P.value (−log10)
T4−cluster1 Pathway P.value (−log10)

Figure 7. T4- or T5-responsive APA event clusters, using PDUI values from 3'-seq

(A, B) PDUI cluster profiles and cluster region type proportions in cells treated with T4 (A) or T5 (B). Events associated with PDUI were clustered using WGCNA analysis and HOMER was used to determine each region in the genome, in each cluster. Proportions are shown as a stacked bar plot. Promoter-transcription start site (TSS), 5'-UTR, exon, intron, 3'-UTR, transcription termination site (TTS), and noncoding region are shown. The three dominant clusters are shown at the bottom of the figure. Black lines represent PDUI profiles. The red line represents the eigenvalue of a cluster. The vertical axis represents the PDUI value. The horizontal axis represents compound concentration (μM). (C–E) Biological process enrichment map for genes involved in APA changes in cells treated with T4 (C), T5 (D), or CPT (E). Each node represents a GO biological process gene set. Blue nodes represent biological processes enriched among the genes with decreased PDUI values. (F) Correlation of enrichment pathway between the monotonically decreasing cluster (Cluster 1) of T4 and T5 for PDUI profiles. Each dot represents the *P*-value of a biological process pathway enriched in the clusters.

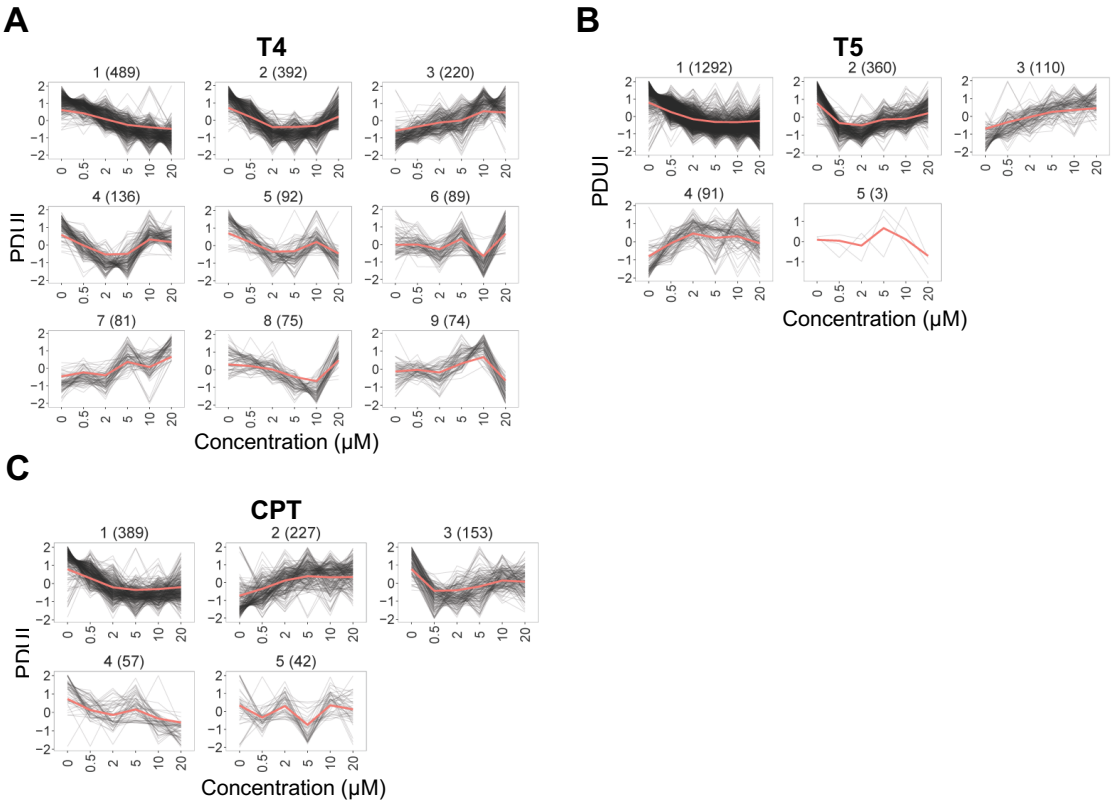

A
T4
1 (489)
2 (392)
3 (220)
4 (136)
5 (92)
6 (89)
7 (81)
8 (75)
9 (74)
PDUI
Concentration (μM)
B
T5
1 (1292)
2 (360)
3 (110)
4 (91)
5 (3)
PDUI
Concentration (μM)
C
CPT
1 (389)
2 (227)
3 (153)
4 (57)
5 (42)
PDUI
Concentration (μM)

Figure 8. T4- or T5-responsive APA event clusters, using PDUI values from 3'-seq

(A–C) PDUI cluster profiles after treatment with T4 (A), T5 (B), or CPT (C). Events associated with PDUI were clustered using WGCNA analysis. Black lines represent PDUI profiles. Each red line represents the eigen-value of a cluster. The vertical axis represents the PDUI value. The horizontal axis represents the compound concentration (μM).

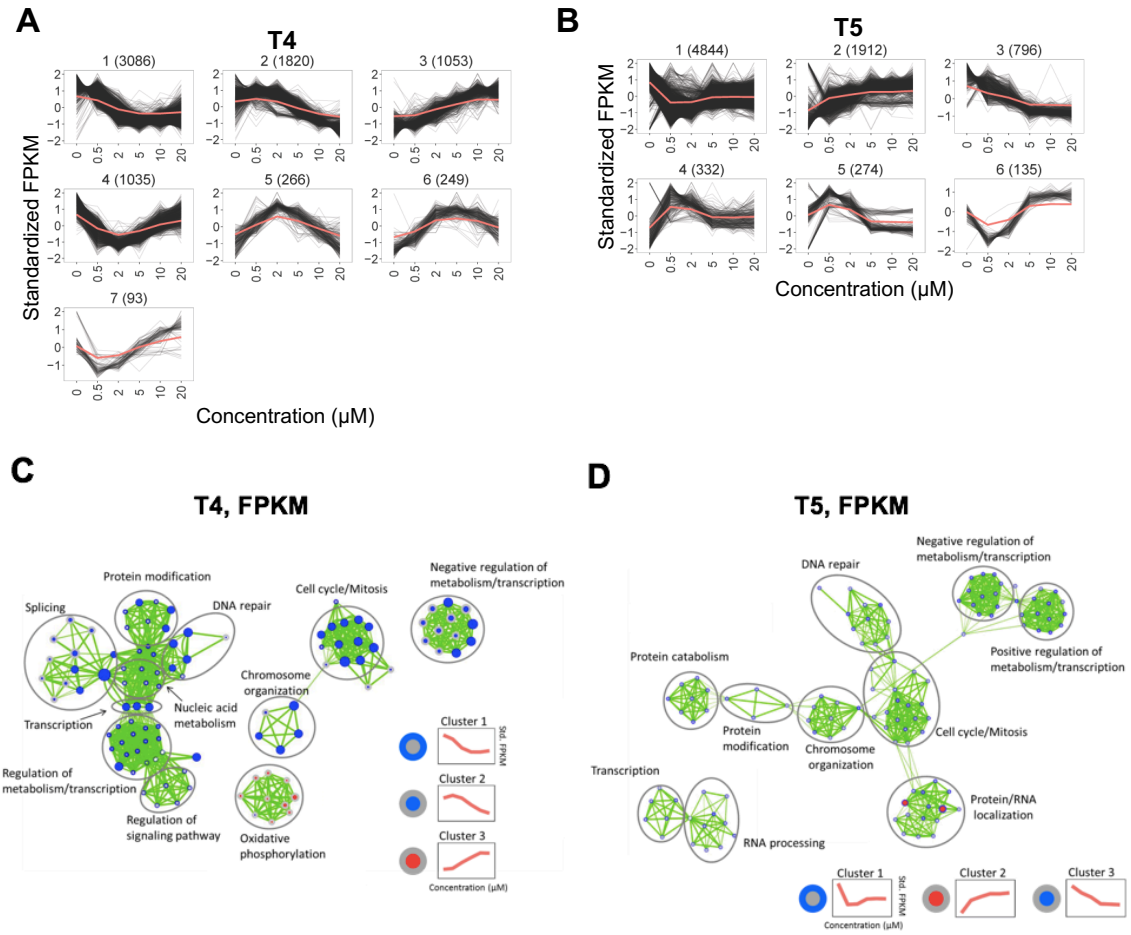

A
T4
1 (3086)
2 (1820)
3 (1053)
4 (1035)
5 (266)
6 (249)
7 (93)
Standardized FPKM
Concentration (μM)
B
T5
1 (4844)
2 (1912)
3 (796)
4 (332)
5 (274)
6 (135)
Standardized FPKM
Concentration (μM)
C
T4, FPKM
Protein modification
Splicing
DNA repair
Cell cycle/Mitosis
Negative regulation of metabolism/transcription
Chromosome organization
Nucleic acid metabolism
Transcription
Regulation of metabolism/transcription
Regulation of signaling pathway
Oxidative phosphorylation
Cluster 1
Cluster 2
Cluster 3
Std. FPKM
Concentration (μM)
D
T5, FPKM
Negative regulation of metabolism/transcription
DNA repair
Protein catabolism
Positive regulation of metabolism/transcription
Protein modification
Chromosome organization
Cell cycle/Mitosis
Transcription
Protein/RNA localization
RNA processing
Cluster 1
Cluster 2
Cluster 3
Std. FPKM
Concentration (μM)

Figure 9. T4- or T5-responsive gene expression clusters, using FPKM values from RNA-seq

(A, B) Normalized gene expression profiles in cells treated with T4 (A) or T5 (B). Gene expression of FPKM was clustered using WGCNA analysis. Black lines represent FPKM profiles. Each red line represents the eigen-value of a cluster. The vertical axis represents FPKM value. The horizontal axis represents compound concentration (μM). (C, D) Biological process enrichment map for differentially expressed genes in cells treated with T4 (C) or T5 (D). Red nodes represent biological processes enriched among upregulated genes, while blue represents downregulated genes. (C) Node cores are shown in blue when the gene set was enriched among genes in Cluster 2, with red for Cluster 3. The outer ring is shown in blue when the gene set was enriched among genes in Cluster 1. (D) Node cores are shown in blue when the gene set was enriched among genes in Cluster 3, with red for Cluster 2. The outer ring is shown in blue when the gene set was enriched among genes in Cluster 1. Edge thickness represents the level of overlap between two gene sets.

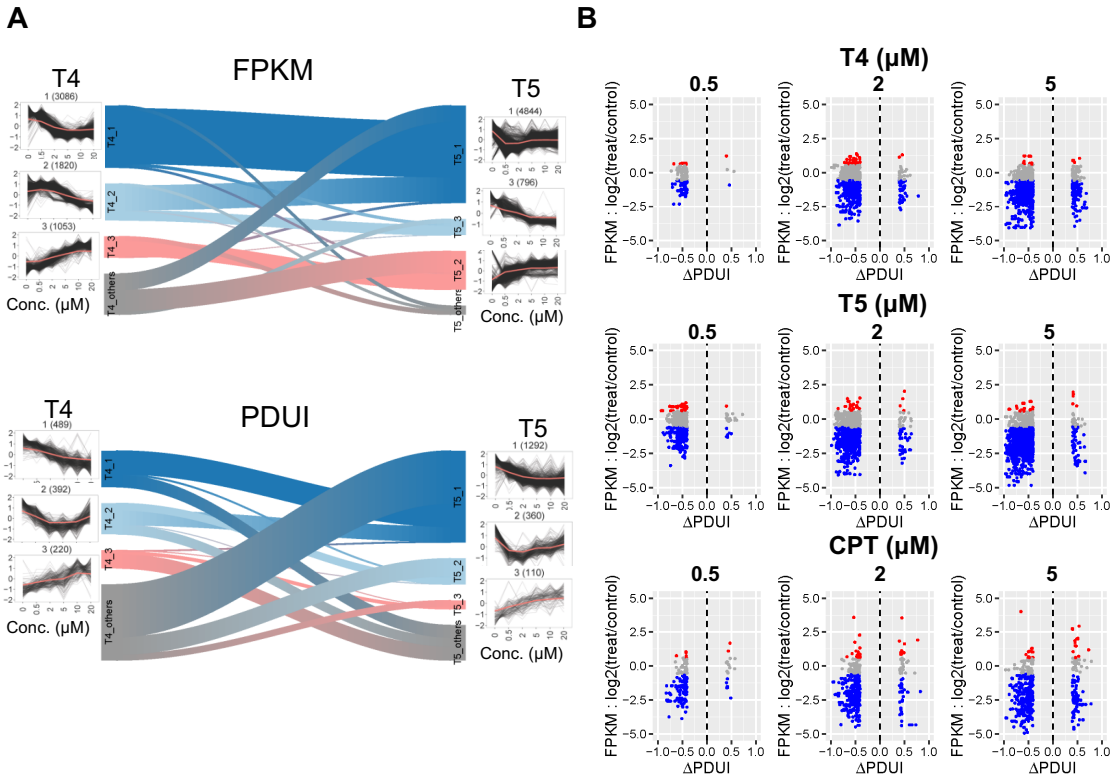

A
FPKM
T4
T5
1 (3086)
2 (1820)
3 (1053)
1 (4844)
3 (796)
Conc. (μM)
T4_1
T4_2
T4_3
T4_others
T5_1
T5_2
T5_3
T5_others
PDUI
1 (489)
2 (392)
3 (220)
1 (1292)
2 (360)
3 (110)
B
T4 (μM)
T5 (μM)
CPT (μM)
0.5
2
5
FPKM : log2(treat/control)
ΔPDUI

Figure 10. Correlation between gene expression and APA event induced by T4 or T5

(A) The schematic is presented as Sankey diagrams in the center, showing the relationships between T4 and T5 treatment in three FPKM (upper) or PDUI (lower) dominant cluster profiles. The blue (cluster 1), light blue (cluster 2), red (cluster 3), and gray (other clusters) lines show relationships between common members. FPKM or PDUI vs. T4 or T5 dose response plots are shown aligned to the cluster that they represent. The vertical axis represents standardized FPKM (upper) or PDUI (lower) scores. The horizontal axis represents T4 (left) or T5 (right) concentrations (μM). Red lines are cluster eigen-events. The number of events in each cluster is shown in parentheses beside the cluster number. (B) Correlation between PAS usage and gene expression. The x-axis indicates ΔPDUI. The y-axis indicates FPKM relative to the control (log2 scale).

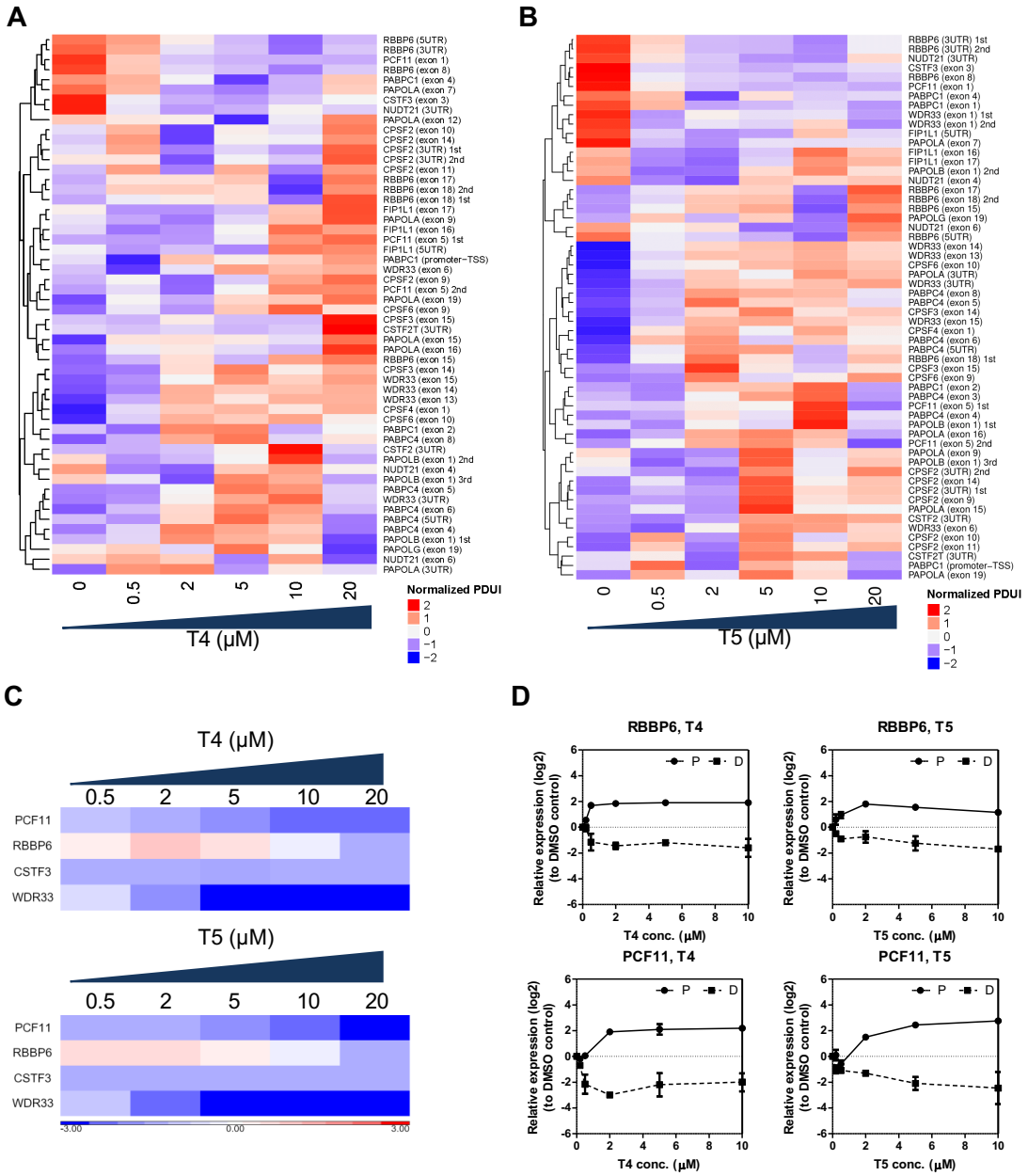

A
RBBP6 (5UTR)
RBBP6 (3UTR)
PCF11 (exon 1)
RBBP6 (exon 8)
PABPC1 (exon 4)
PAPOLA (exon 7)
CSTF3 (exon 3)
NUDT21 (3UTR)
PAPOLA (exon 12)
CPSF2 (exon 10)
CPSF2 (exon 14)
CPSF2 (3UTR) 1st
CPSF2 (3UTR) 2nd
CPSF2 (exon 11)
RBBP6 (exon 17)
RBBP6 (exon 18) 2nd
RBBP6 (exon 18) 1st
FIP1L1 (exon 17)
PAPOLA (exon 9)
FIP1L1 (exon 16)
PCF11 (exon 5) 1st
FIP1L1 (5UTR)
PABPC1 (promoter−TSS)
WDR33 (exon 6)
CPSF2 (exon 9)
PCF11 (exon 5) 2nd
PAPOLA (exon 19)
CPSF6 (exon 9)
CPSF3 (exon 15)
CSTF2T (3UTR)
PAPOLA (exon 15)
PAPOLA (exon 16)
RBBP6 (exon 15)
CPSF3 (exon 14)
WDR33 (exon 15)
WDR33 (exon 14)
WDR33 (exon 13)
CPSF4 (exon 1)
CPSF6 (exon 10)
PABPC1 (exon 2)
PABPC4 (exon 8)
CSTF2 (3UTR)
PAPOLB (exon 1) 2nd
NUDT21 (exon 4)
PAPOLB (exon 1) 3rd
PABPC4 (exon 5)
WDR33 (3UTR)
PABPC4 (exon 6)
PABPC4 (5UTR)
PABPC4 (exon 4)
PAPOLB (exon 1) 1st
PAPOLG (exon 19)
NUDT21 (exon 6)
PAPOLA (3UTR)
0
0.5
2
5
10
20
T4 (μM)
Normalized PDUI
2
1
0
−1
−2
B
RBBP6 (3UTR) 1st
RBBP6 (3UTR) 2nd
NUDT21 (3UTR)
CSTF3 (exon 3)
RBBP6 (exon 8)
PCF11 (exon 1)
PABPC1 (exon 4)
PABPC1 (exon 1)
WDR33 (exon 1) 1st
WDR33 (exon 1) 2nd
FIP1L1 (5UTR)
PAPOLA (exon 7)
FIP1L1 (exon 16)
FIP1L1 (exon 17)
PAPOLB (exon 1) 2nd
NUDT21 (exon 4)
RBBP6 (exon 17)
RBBP6 (exon 18) 2nd
RBBP6 (exon 15)
PAPOLG (exon 19)
NUDT21 (exon 6)
RBBP6 (5UTR)
WDR33 (exon 14)
WDR33 (exon 13)
CPSF6 (exon 10)
PAPOLA (3UTR)
WDR33 (3UTR)
PABPC4 (exon 8)
PABPC4 (exon 5)
CPSF3 (exon 14)
WDR33 (exon 15)
CPSF4 (exon 1)
PABPC4 (exon 6)
PABPC4 (5UTR)
RBBP6 (exon 18) 1st
CPSF3 (exon 15)
CPSF6 (exon 9)
PABPC1 (exon 2)
PABPC4 (exon 3)
PCF11 (exon 5) 1st
PABPC4 (exon 4)
PAPOLB (exon 1) 1st
PAPOLA (exon 16)
PCF11 (exon 5) 2nd
PAPOLA (exon 9)
PAPOLB (exon 1) 3rd
CPSF2 (3UTR) 2nd
CPSF2 (exon 14)
CPSF2 (3UTR) 1st
CPSF2 (exon 9)
PAPOLA (exon 15)
CSTF2 (3UTR)
WDR33 (exon 6)
CPSF2 (exon 10)
CPSF2 (exon 11)
CSTF2T (3UTR)
PABPC1 (promoter−TSS)
PAPOLA (exon 19)
0
0.5
2
5
10
20
T5 (μM)
Normalized PDUI
2
1
0
−1
−2
C
T4 (μM)
0.5
2
5
10
20
PCF11
RBBP6
CSTF3
WDR33
T5 (μM)
0.5
2
5
10
20
PCF11
RBBP6
CSTF3
WDR33
-3.00
0.00
3.00
D
RBBP6, T4
RBBP6, T5
PCF11, T4
PCF11, T5
P
D
Relative expression (log2)
(to DMSO control)
T4 conc. (μM)
T5 conc. (μM)

Figure 11. Autoregulation of 3'-processing and transcription pathway-related proximal poly(A) mRNA by APA modulators

(A, B) Heatmap showing normalized PDUI values of known polyadenylation factors. Each rectangle represents the PDUI for cells treated with T4 (A) or T5 (B) at the indicated concentration. (C) Heat map indicating the normalized FPKM expression of PCF11, RBBP6, PAPOLA, NUDT21, and CSTF3 genes in cells treated with T4 or T5. (D) 3'-End qPCR analysis for four APA sites, consisting of two genes, in U2OS cells treated with the indicated compound. Gene expression of proximal and distal mRNA (log2 scale) in U2OS cells treated with the indicated compound and concentration for 6 h. Data are mean ± SD from three independent experiments.

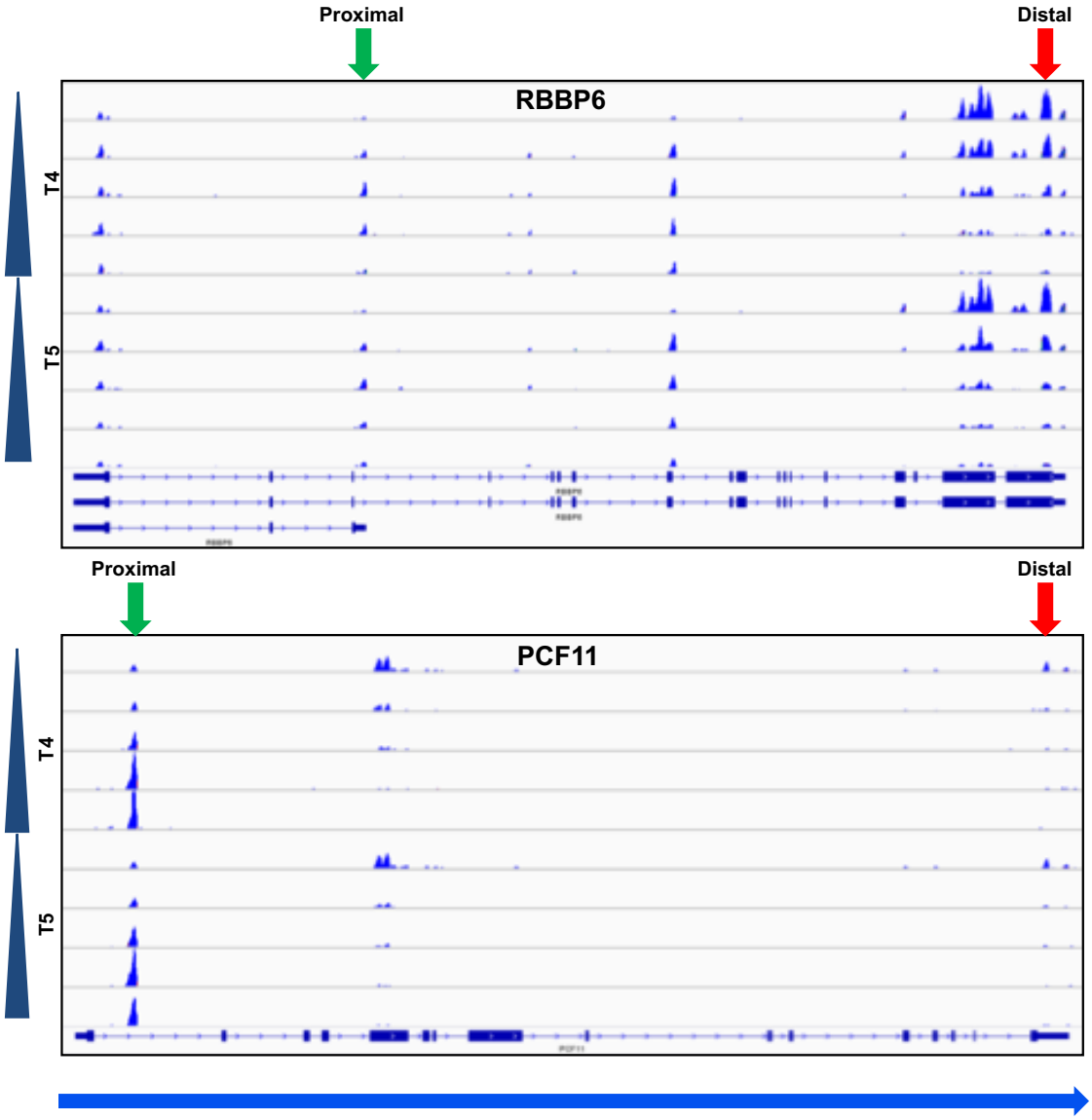

Proximal
Distal
RBBP6
T4
T5
Proximal
Distal
PCF11
T4
T5

Figure 12. T4 or T5 induces RBBP6 and PCF11 proximal PAS usage.

RBBP6 and PCF11 transcripts indicating APA changes after treatment with T4 or T5 for 6 h. Transcripts from 3'-Seq data with blue-colored reads were encoded on the forward DNA strand.

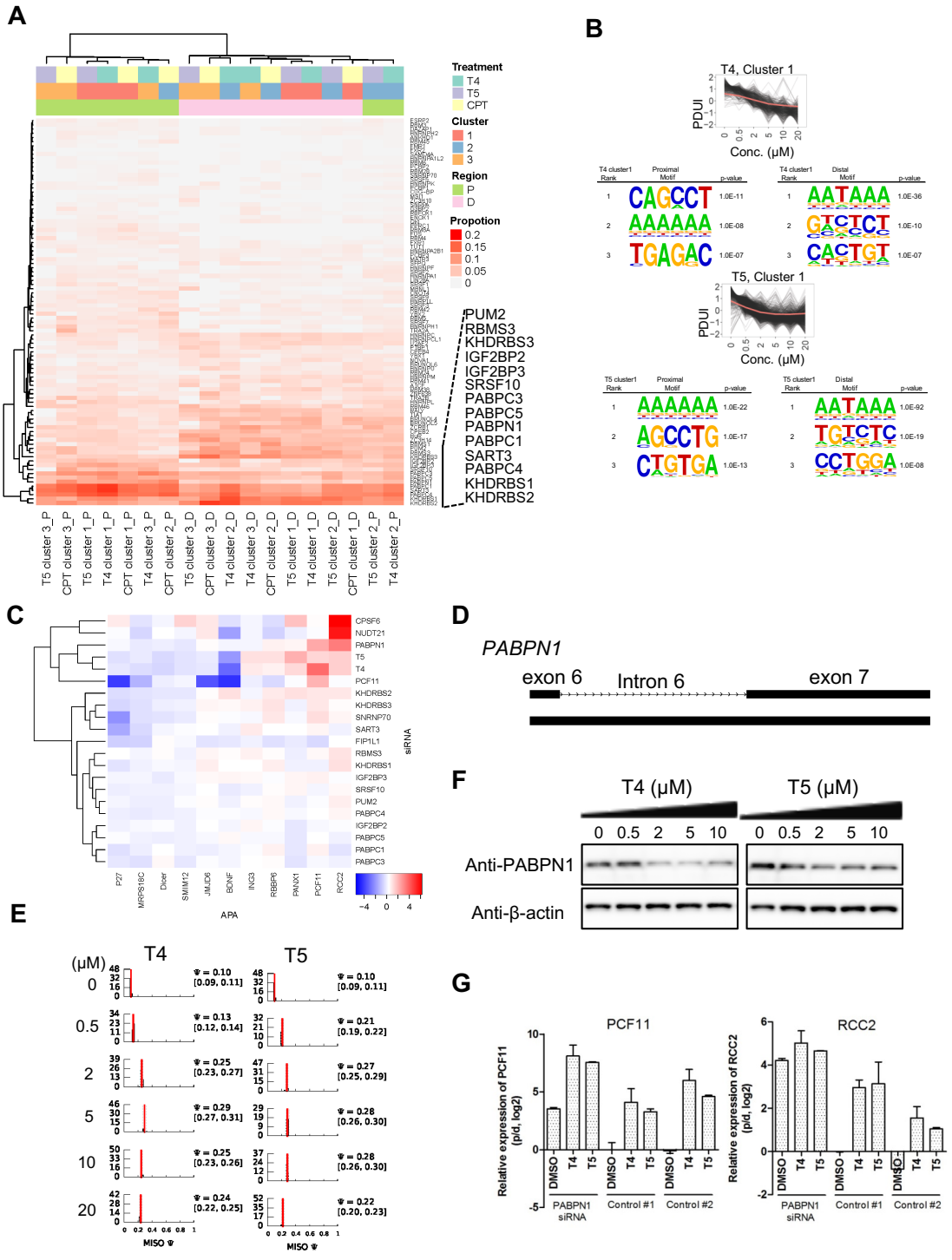

A
Treatment
T4
T5
CPT
Cluster
1
2
3
Region
P
D
Propotion
0.2
0.15
0.1
0.05
0
PUM2
RBMS3
KHDRBS3
IGF2BP2
IGF2BP3
SRSF10
PABPC3
PABPC5
PABPN1
PABPC1
SART3
PABPC4
KHDRBS1
KHDRBS2
T5 cluster 3_P
CPT cluster 3_P
T5 cluster 1_P
T4 cluster 1_P
CPT cluster 1_P
T4 cluster 3_P
CPT cluster 2_P
T5 cluster 3_D
CPT cluster 3_D
T4 cluster 2_D
T4 cluster 3_D
CPT cluster 2_D
T5 cluster 1_D
T4 cluster 1_D
T5 cluster 2_D
CPT cluster 1_D
T5 cluster 2_P
T4 cluster 2_P
B
T4, Cluster 1
PDUI
Conc. (μM)
T4 cluster1 Rank
Proximal Motif
p-value
1.0E-11
1.0E-08
1.0E-07
T4 cluster1 Rank
Distal Motif
p-value
1.0E-36
1.0E-10
1.0E-07
T5, Cluster 1
PDUI
Conc. (μM)
T5 cluster1 Rank
Proximal Motif
p-value
1.0E-22
1.0E-17
1.0E-13
T5 cluster1 Rank
Distal Motif
p-value
1.0E-92
1.0E-19
1.0E-08
C
CPSF6
NUDT21
PABPN1
T5
T4
PCF11
KHDRBS2
KHDRBS3
SNRNP70
SART3
FIP1L1
RBMS3
KHDRBS1
IGF2BP3
SRSF10
PUM2
PABPC4
IGF2BP2
PABPC5
PABPC1
PABPC3
siRNA
P27
MRPS18C
Dicer
SMIM12
JMJD6
BCNF
IMG3
RBBP6
PAIX1
PCF11
RCC2
APA
D
PABPN1
exon 6
Intron 6
exon 7
E
T4
T5
(μM)
0
0.5
2
5
10
20
MISO Ψ
F
T4 (μM)
T5 (μM)
0 0.5 2 5 10
0 0.5 2 5 10
Anti-PABPN1
Anti-β-actin
G
PCF11
RCC2
Relative expression of PCF11 (p/d, log2)
Relative expression of RCC2 (p/d, log2)
DMSO
T4
T5
PABPN1 siRNA
Control #1
Control #2

Figure 13. Motif elements in APA regulation and implication of target molecule/signaling for T4 and T5

(A) Heatmap indicating the proportion of genes with 114 known human/mouse motifs for RNA binding proteins. Each rectangle indicates the PDUI with three monotonically increasing or decreasing dominant clusters, from cells treated with T4, T5, or CPT. (B) The de novo motif analysis upstream (−50 bp) of CSs [the proximal poly(A) or the distal poly(A)] in Cluster 1 for T4 and T5. (C) Heat map showing normalized APA values for 11 genes, obtained by transfection with 19 siRNAs, for the indicated genes, or by treatment with 5 μM T4 or 2 μM T5. Color scale indicates the relative expression of each gene [proximal (p) relative to distal (d), log2 scale]. (D) Schematic representation of the retained intron form of PABPN1 mRNA. Each box represents an exon and each line represents an intron. (E) FPKM values are displayed along the vertical axis and MISO PSI posterior distribution along the horizontal axis, for cells treated with T4 (left) or T5 (right) at the indicated concentration. (F) U2OS cells were treated with T4 or T5 for 24 h. Immunoblotting for PABPN1 was performed. (G) Gene expression of proximal (p) relative to distal PCF11 or RCC2 mRNA. U2OS cells were transfected with siRNA for PABPN1 or control nontargeting siRNA. After 4 d, T4 or T5 was administered to cells for 6 h and mRNA levels were measured using 3'-end qPCR. Data are mean ± SD from two independent experiments.

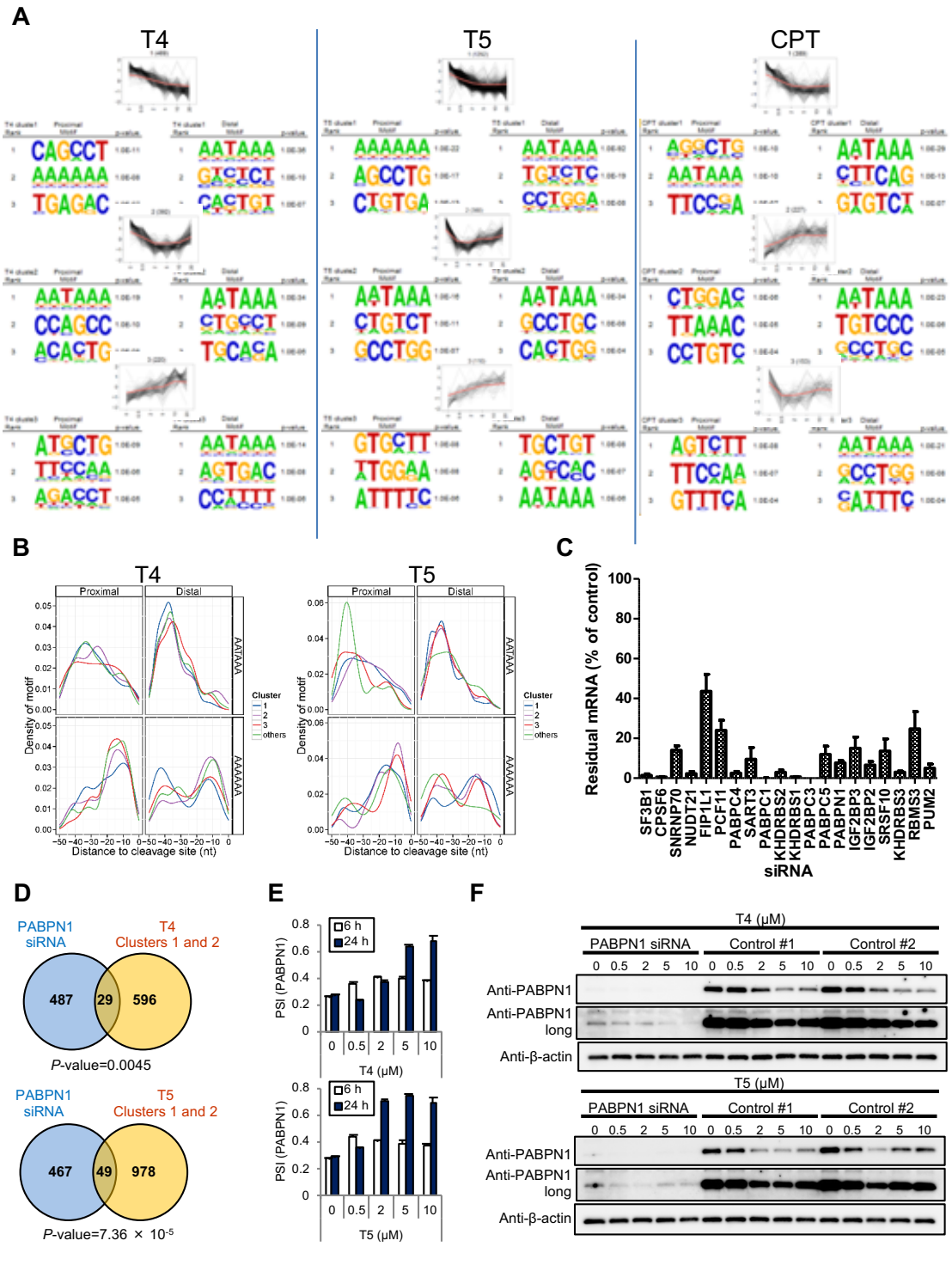

A
T4
T5
CPT
B
T4
T5
Proximal
Distal
AATAAA
AAAAAA
Density of motif
Distance to cleavage site (nt)
Cluster
1
2
3
others
C
Residual mRNA (% of control)
SF3B1
CPSF6
SNRNP70
NUDT21
FIP1L1
PCF11
PABPC4
SART3
PABPC1
KHDRBS2
KHDRBS1
PABPC3
PABPC5
PABPN1
IGF2BP3
IGF2BP2
SRSF10
KHDRBS3
RBMS3
PUM2
siRNA
D
PABPN1 siRNA
T4 Clusters 1 and 2
487
29
596
P-value=0.0045
PABPN1 siRNA
T5 Clusters 1 and 2
467
49
978
P-value=7.36 × 10-5
E
PSI (PABPN1)
6 h
24 h
T4 (μM)
T5 (μM)
F
T4 (μM)
T5 (μM)
PABPN1 siRNA
Control #1
Control #2
Anti-PABPN1
Anti-PABPN1 long
Anti-β-actin

Figure 14. Motif analysis in APA regulation and involvement of PABPN1 autoregulation in APA.

(A) The de novo motif analysis upstream (−50 bp) of CSs [the proximal poly(A) or the distal poly(A)] in the top 3 clusters and other clusters for T4, T5, and CPT. (B) Kernel density curve of A-rich motif (AATAAA or AAAAAA) distribution for three dominant clusters (blue, purple, and red lines) and other clusters (green line). The x-axis indicates the position of the mapped CSs. The y-axis indicates the density of AATAAA or AAAAAA motifs. (C) U2OS cells were transfected with the indicated siRNA or control nonsilencing siRNA. Cells were harvested after 48 h of siRNA transfection. The expression level of each gene was measured by qRT-PCR. Data are mean ± SD from two independent analyses. (D) A Venn diagram showing the overlap between DtoP APA changes induced by PABPN1 knockdown (blue circles) and that induced by T4 (upper) or T5 (lower) (orange circles) in U2OS cells. The number of genes in each area is indicated. Statistical analyses were performed using Fisher's exact test. (E) Percent spliced-in (PSI) values using RT-PCR for PABPN1 alternative exons after treatment with T4 (upper) or T5 (lower) for 6 or 24 h. Data are mean ± SD from two independent analyses. (F) U2OS cells were transfected with siRNA for PABPN1 or control nontargeting siRNA. After 4 d, T4 or T5 was administered to cells for 24 h. Immunoblotting for PABPN1 was performed.

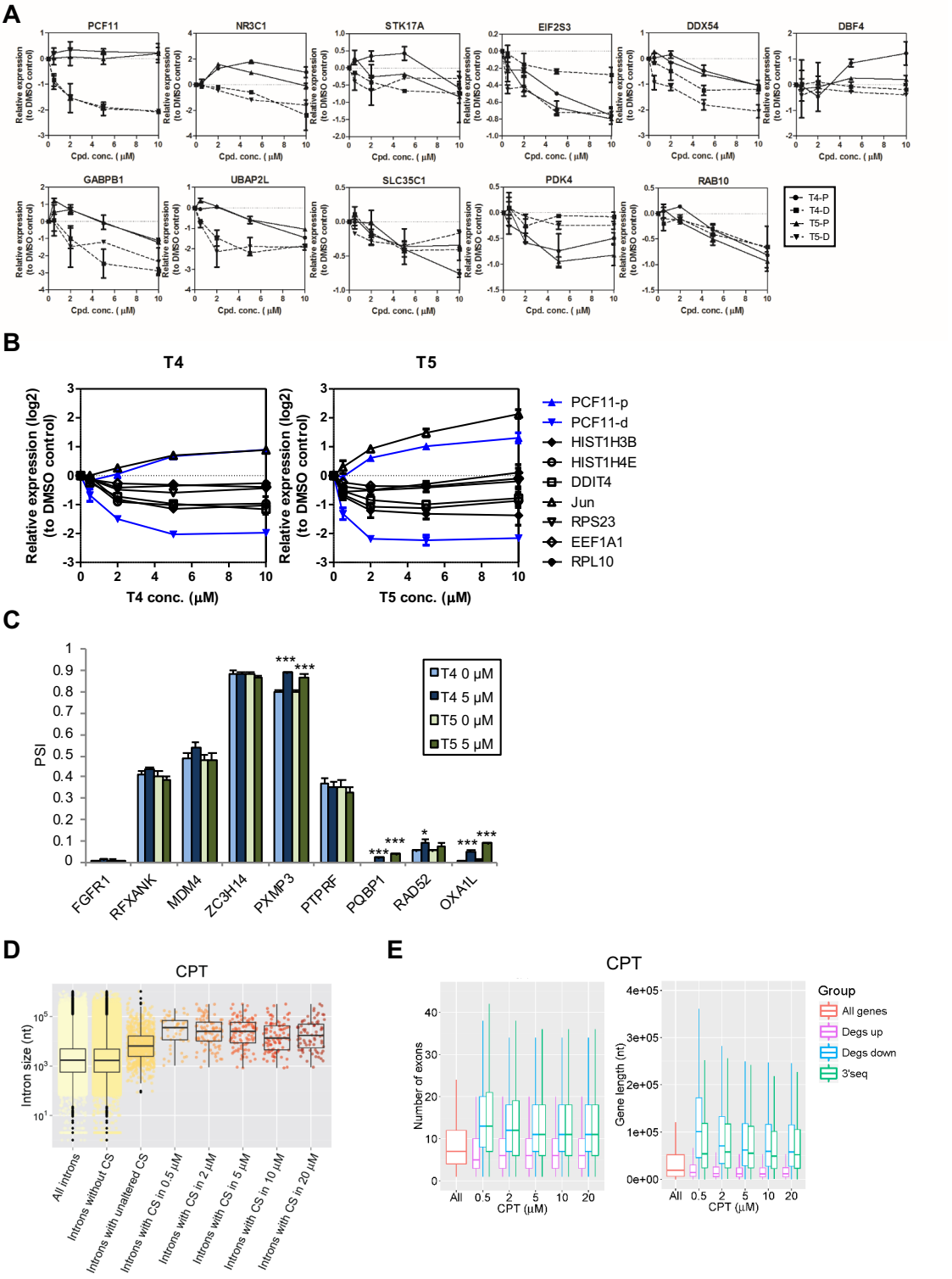

A
PCF11
NR3C1
STK17A
EIF2S3
DDX54
DBF4
GABPB1
UBAP2L
SLC35C1
PDK4
RAB10
Relative expression (to DMSO control)
Cpd. conc. (μM)
T4-P
T4-D
T5-P
T5-D
B
T4
T5
Relative expression (log2) (to DMSO control)
T4 conc. (μM)
T5 conc. (μM)
PCF11-p
PCF11-d
HIST1H3B
HIST1H4E
DDIT4
Jun
RPS23
EEF1A1
RPL10
C
PSI
T4 0 μM
T4 5 μM
T5 0 μM
T5 5 μM
FGFR1
RFXANK
MDM4
ZC3H14
PXMP3
PTPRF
PQBP1
RAD52
OXA1L
D
CPT
Intron size (nt)
All introns
Introns without CS
Introns with unaltered CS
Introns with CS in 0.5 μM
Introns with CS in 2 μM
Introns with CS in 5 μM
Introns with CS in 10 μM
Introns with CS in 20 μM
E
CPT
Number of exons
Gene length (nt)
CPT (μM)
Group
All genes
Degs up
Degs down
3'seq

Figure 15. Indirect involvement of U1 snRNP and elongation rate for APA regulated by T4 or T5.

(A, B) Results of 3'-end qPCR. (A) Expression of the proximal and distal indicated mRNA in HeLa cells treated with T4 or T5 for 6 h. Data are mean ± SD from two independent experiments. (B) Expression of the proximal and distal indicated mRNA in 293 cells treated with T4 or T5 for 6 h. Blue lines indicate PCF11 mRNA as a positive control. Data are mean ± SD from two independent experiments. (C) Percent spliced-in (PSI) values using RT-PCR for the indicated alternative exons after treatment with 5 μM T4 or T5 for 6 h. Data are mean ± SD from two independent analyses. Statistical analyses were performed using an unpaired Student's t-test (*$P < 0.05$; ***$P < 0.001$). (D) Box plots of intron size comparing introns with CSs induced by CPT to those without CSs. The individual intron sizes are plotted with dots. (E) Box plots of the number of exons or transcription length, comparing genes whose ΔPDUI value is > 0.4 in cells treated with CPT (3'-seq: green), those whose expression is differentially upregulated (pink: fold change >1.5) or downregulated (blue: fold change < 1.5), or all genes (orange).

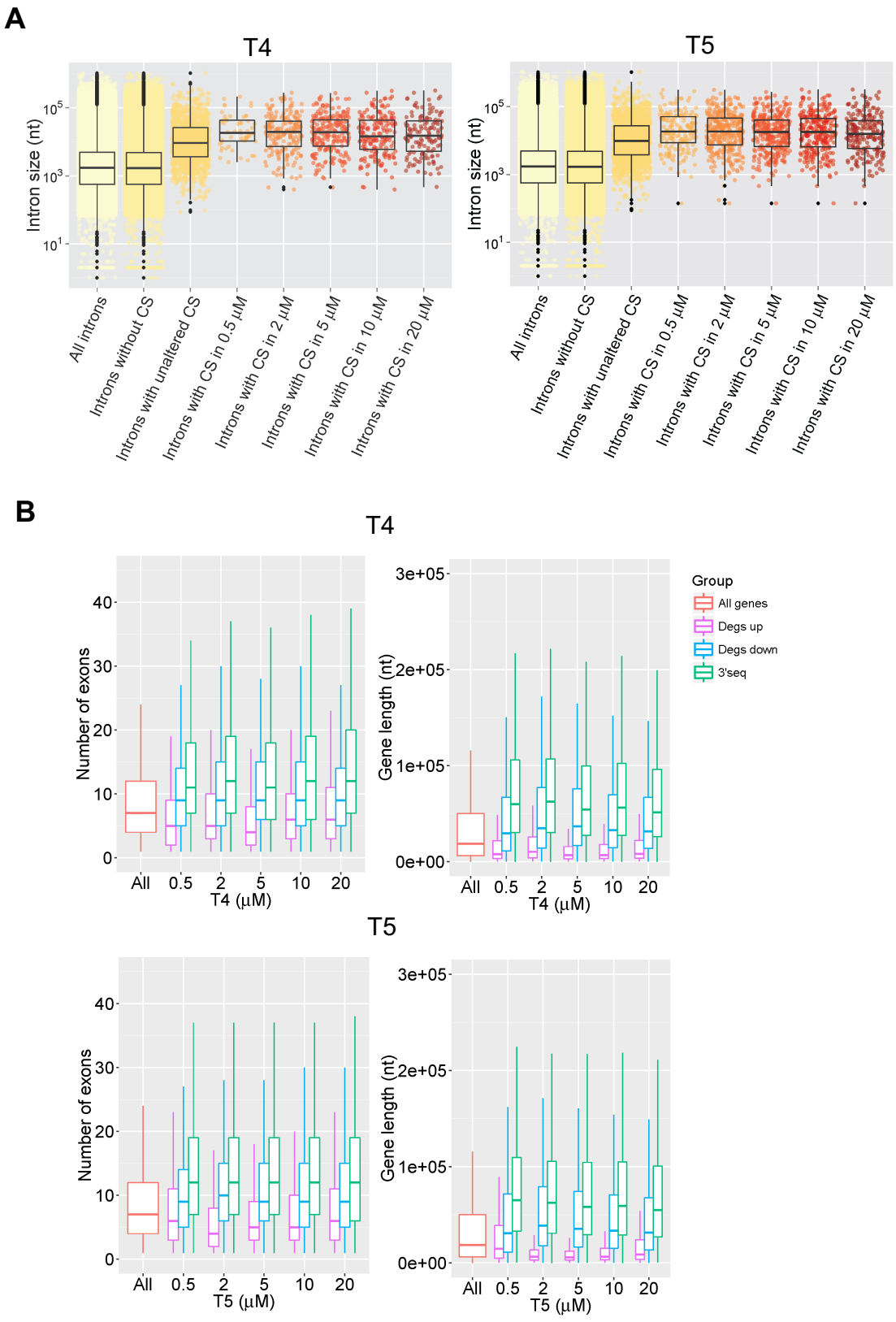

A
T4
T5
Intron size (nt)
10^5
10^3
10^1
All introns
Introns without CS
Introns with unaltered CS
Introns with CS in 0.5 μM
Introns with CS in 2 μM
Introns with CS in 5 μM
Introns with CS in 10 μM
Introns with CS in 20 μM
B
T4
Number of exons
Gene length (nt)
3e+05
2e+05
1e+05
0e+00
0
10
20
30
40
All
0.5
2
5
10
20
T4 (μM)
Group
All genes
Degs up
Degs down
3'seq
T5
T5 (μM)

Figure 16. Long introns and transcripts were targeted by the APA modulators

(A) Box plots of intron size comparing introns with CSs induced by T4 or T5 and introns without CSs. The individual intron sizes are plotted as dots. (B) Box plots of the number of exons or transcription length comparing genes whose ΔPDUI value was > 0.4 in cells treated with the indicated compound (3'-seq: green), showing genes with expression differentially upregulated (pink: fold change >1.5), genes with expression downregulated (blue: fold change < 1.5), or all genes (orange).

General Discussion

Transcripts play a central role in regulation of biological systems in eukaryotic cells. Orchestrating systems of transcript processing expand millions of protein variation from tens of thousands of genes and generate complexity of transcriptomic perturbation in cells. Therefore, effective data mining approaches are necessary to extract valuable information from genome-wide profiling data. In this study, I demonstrated the effectiveness of genome-wide profiling approaches for elucidation of molecular mechanism of biological response in cancer cells and generated practicable hypotheses that could lead to new insights in biology. In chapter 1, I profiled transcriptomic perturbation in cancer cells with CENP-E inhibition and generated a feasible hypothesis of anti-tumor effect of CENP-E inhibition for cancer therapy. In chapter 2, I identified novel APA modulators with high selectivity, and presented a landscape of transcriptomic perturbations caused by the modulators and proposed a hypothesis of molecular mechanism for cellular effect of these compounds using multilayered genome-wide profiling.

CENP-E is one of the attractive therapeutic targets for cancer because discovering its inhibitor was possible and it was known that CENP-E inhibition induced anti-tumor effect for cancer cells with chromosomal instability (CIN) (Wood et al., 2010; Ohashi et al. 2015). CIN feature is known as a driver of tumor aggressiveness. Accumulation of CIN cancer cells in tumor contributes to adaptability, drug resistance and metastasis (Thompson et al., 2017). In the past 5 years, immune check point blocker (ICB) has become an innovative therapeutic option for various cancers (Ribas et al., 2018). However, it has been known that proportion of responder of ICB is only from 10 to 20 percent (Haslam et al., 2019). It was proposed that various factors could be related to resistance for ICB (Jenkins et al., 2018; Fares et al., 2019). Among them, CIN feature is one of the compelling ICB-resistant factors. Therefore, development of therapeutic drug for CIN cancer could give a novel therapeutic option to ICB-

resistant patients. Based on the p53 hypothesis of apoptosis by CENP-E inhibition, which I proposed in chapter 1, Ohashi et al. proved that postmitotic apoptosis in CIN cancer cells induced by CENP-E inhibition was caused by activation of p53 signaling accompanied by DNA damage response and unfolded protein response activation, and demonstrated anti-tumor effect of CENP-E inhibitor in xenograft model with CIN phenotype (Ohashi et al. 2015). These evidences suggest that inhibition of CENP-E function could be one of the promising therapeutic concepts for CIN cancer. Clarification of molecular mechanism of drug effect could increase success rate of drug development (Morgan et al., 2018). Therefore, my study made an important clue for developing CENP-E inhibitor for cancer therapy and would contribute to providing a new therapeutic option for unmet medical needs.

APA is one of the essential machineries of transcript processing and a known cause of genetic disorders including cancer (Mayr et al., 2009). However, no therapeutic option based on APA has been developed so far. One reason was that there was no chemical tool targeting APA function, though various chemical tools modulating other transcriptomic machineries, transcription and splicing, had been developed (Ito et al., 2018; Funnell et al., 2017; Mazloomian et al.,2019). T4 and T5 that I have identified in this study were the first selective APA modulators. Although it was reported that knockdown of PABPN1 by siRNA could control APA (Jenal et al., 2012), situation of siRNA usage is limited due to effectiveness of transfection that depends on cell types. On the other hand, chemical tool can use most cell types. Additionally, chemical tool can control timing of stimulation and intensity of the efficacy by dose modification. Therefore, my novel APA modulators have a great usability advantage for various biological conditions and would contribute to progress of APA biology. Furthermore, this study could be an important steppingstone to developing therapeutic drug based on APA regulation. My profiling for transcriptomic perturbations revealed the

mechanism of action of T4 and T5, which induced selective 3' UTR shortening in A-rich motif and long intron via loss of function of PABPN1. Dysregulation of APA was observed in various tumor types (Xia et al., 2014), and several models of carcinogenesis via APA aberration were suggested (Masamha et al., 2018). Therefore, APA modulator could have anti-tumor effect. To confirm that T4 and T5 have the potential for cancer therapy, I investigated the association between PABPN1 and cancer with public database. In the Human Protein Atlas database (https://www.proteinatlas.org/), high expression of PABPN1 showed worse survival rate with statistical significance in renal and liver cancers (Figure 17, *P*-value: 2.4e-7 in renal cancer, 3.5e-4 in liver cancer). Searching knockout data of PABPN1 in the DepMap database (https://depmap.org/portal/), which included genome-wide screening of cancer cell growth inhibition by CRISPR, result data showed knockout of PABPN1 had a growth inhibition effect for many types of cancer cell, including renal and liver cancers (Figure 18). These data suggested that inhibition of PABPN1 function could have a therapeutic effect for renal and liver cancers. Furthermore, expression profiling and APA profiling showed that T4 and T5 had negative effect for cell cycle in my study (Figure 9C and D). Although further investigation of T4 and T5 function is needed, T4 and T5 have the therapeutic potential for cancer via PABPN1 inhibition.

Through this study, I demonstrated the effectiveness of genome-wide profiling approaches for elucidation of molecular mechanism of biological response in cells by generating practicable hypotheses. These approaches could be applicable for various biological situations, especially in oncology, which show complex transcriptomic perturbation, and would contribute to understanding molecular mechanisms in a broad area of biology.

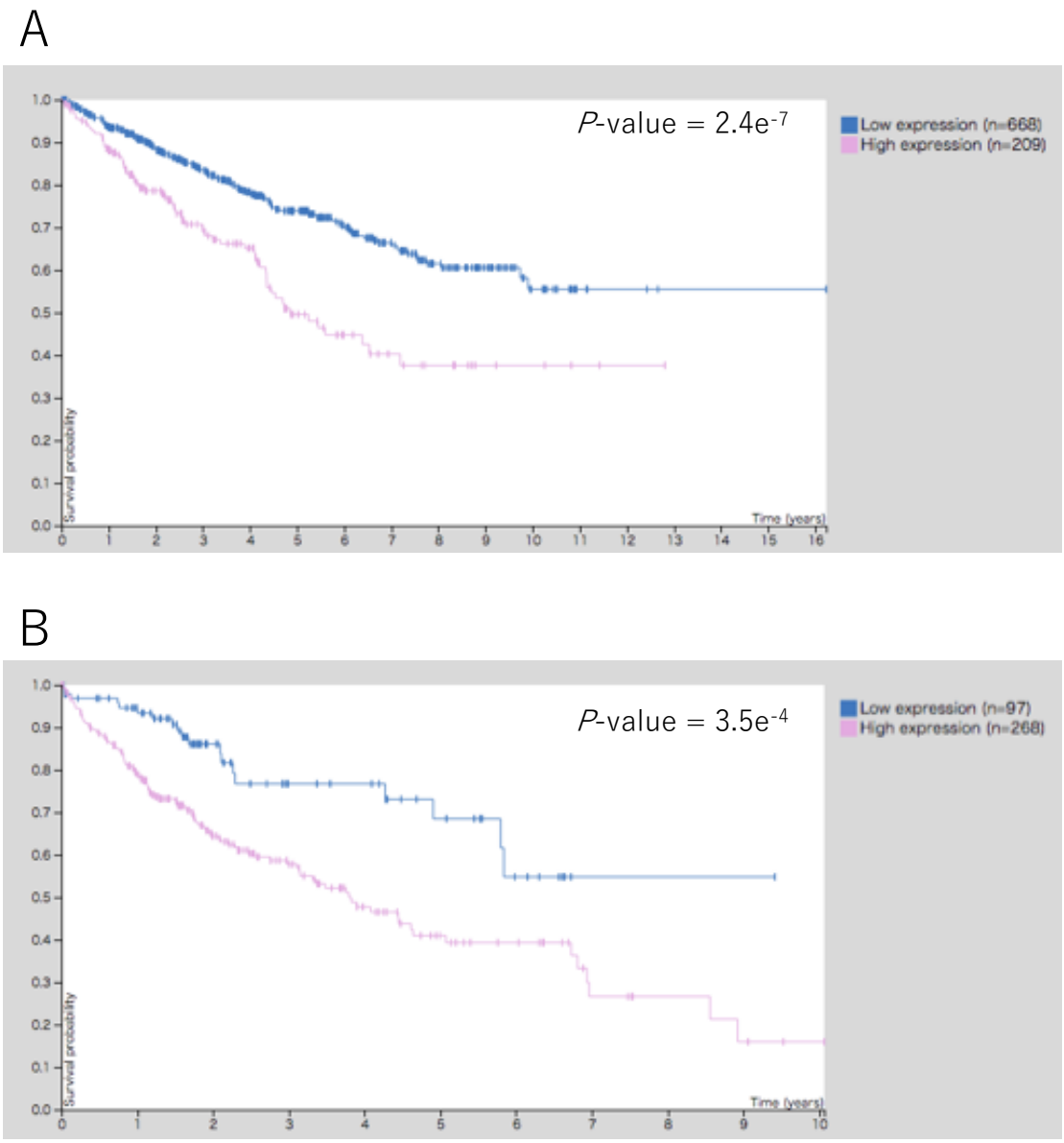


Figure 17. Expression of PABPN1 and survival curve of cancer patients.

Kaplan-Meier survival curve for renal cancer (A) and liver cancer (B) patients with PABPN1 expression. X-axis shows time for survival (years) and y-axis shows the probability of survival, where 1.0 corresponds to 100 percent. Patients were divided based on level of expression into one of the two groups “low” (blue) and "high" (pink). the survival outcomes of the two groups were compared by log-rank tests.

Figure 18. Dependency score of PABPN1 for cancer cell line.

Dependency score is based on data from a cell depletion assay. The score indicates effect of growth inhibition. Low score means high growth inhibition. A score of -1 (red line) is comparable to the median of pan-essential genes.

Acknowledgements

I am deeply grateful to Professors Kazuto Nakada, Tetsuo Hashimoto, Tomoki Chiba, and Associate Professor Ryuhei Harada, University of Tsukuba, for guiding my work and valuable discussions through my doctoral program.

I am very thankful to Dr. Akihiro Ohashi in National Cancer Center Japan and Dr Shinsuke Araki in Takeda Pharmaceutical Company Limited for cooperation and valuable discussions through my studies.

I appreciate Dr. Hiroyoshi Toyoshiba in FRONTEO Healthcare Inc., Professors Samuel Aparicio in British Columbia Cancer Agency and Atsushi Nakanishi in Takeda Pharmaceutical Company Limited for valuable suggestions, contributions and helpful supports.

I also thank Dr. Masaki Hosoya in FUJIFILM Corporation for valuable suggestions and comments.

I also thank Drs. Naomasa Suita and Hiroki Okada in Ono Pharmaceutical Company Limited for their understanding and support on my doctoral program.

Finally, I would like to appreciate my family for supporting my life in University of Tsukuba.

References

Adiconis, X., Haber, A. L., Simmons, S. K., Levy Moonshine, A., Ji Z., Busby, M. A., et al. (2018). Comprehensive comparative analysis of 5'-end RNA-sequencing methods. Nature methods, 15, 505–511.

Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., and Sorek, R. (2006). Transcription-mediated gene fusion in the human genome. Genome research 16, 30-36.

An, J.J., Gharami, K., Liao, G.Y., Woo, N.H., Lau, A.G., Vanevski, F., Torre, E.R., Jones, K.R., Feng, Y., Lu, B., et al. (2008). Distinct Role of Long 3' UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. Cell 134, 175-187.

Araki, S., Dairiki, R., Nakayama, Y., Murai, A., Miyashita, R., Iwatani, M., Nomura, T., and Nakanishi, O. (2015). Inhibitors of CLK protein kinases suppress cell growth and induce apoptosis by modulating pre-mRNA splicing. PloS one 10, e0116929.

Beck, A.H., Weng, Z., Witten, D.M., Zhu, S., Foley, J.W., Lacroute, P., Smith, C.L., Tibshirani, R., van de Rijn, M., Sidow, A., et al. (2010). 3'-end sequencing for expression quantification (3SEQ) from archival tumor samples. PloS one 5, e8768.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell 150, 53-64.

Bergeron, D., Pal, G., Beaulieu, Y.B., Chabot, B., and Bachand, F. (2015). Regulated Intron Retention and Nuclear Pre-mRNA Decay Contribute to PABPN1 Autoregulation. Molecular and cellular biology 35, 2503-2517.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell 168, 629–643.

Chan, S., Choi, E.A., and Shi, Y. (2011). Pre-mRNA 3'-end processing complex assembly and function. Wiley interdisciplinary reviews RNA 2, 321-335.

Chen, W., Jia, Q., Song, Y., Fu, H., Wei, G., and Ni, T. (2017). Alternative Polyadenylation: Methods, Findings, and Impacts. Genomics, proteomics & bioinformatics, 15(5), 287–300.

Chung, V., Heath, E., Schelman, W.R., Johnson, B.M., Kirby, L.C., , et al. (2011). First-time-in-human study of GSK923295, a novel antimitotic inhibitor of centromere-associated protein E (CENP-E), in patients with refractory cancer. Cancer Chemotherapy and Pharmacology, 69(3), 733-741.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., et al. (2016). A survey of best practices for RNA-seq data analysis. Genome biology 17, 13.

Crick, F.H.C. (1958). On protein synthesis. Symposia of Society Experimental Biology 12, 138–163.

Crick, F. (1970). Central dogma of molecular biology. Nature 227, 561–563.

Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and Consequences of Alternative Polyadenylation. Molecular cell 43, 853-866.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15-21.

Dutertre, M., Chakrama, F.Z., Combe, E., Desmet, F.-O., Mortada, H., Polay Espinoza, M., Gratadou, L., and Auboeuf, D. (2014). A recently evolved class of alternative 3'-terminal exons involved in cell cycle regulation by topoisomerase inhibitors. Nature communications 5, 3395.

Elkon, R., Drost, J., van Haaften, G., Jenal, M., Schrier, M., Vrielink, J., and Agami, R. (2012). E2F mediates enhanced alternative polyadenylation in proliferation. Genome biology 13, R59.

Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nature reviews Genetics 14, 496-506.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Molecular cell 60, 256-267.

Fares C.M., Van Allen E.M., Drake C.G., Allison J.P., Hu-Lieskovan S. (2019). Mechanisms of Resistance to Immune Checkpoint Blockade: Why Does Checkpoint Inhibitor Immunotherapy Not Work for All Patients? American Society of Clinical Oncology Educational Book 39, 147-164

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes & development 28, 2663-2676.

Funnell, T., Tasaki, S., Oloumi, A., Araki, S., Kong, E., Yap, D., Nakayama, Y., Hughes, C.S., Cheng, S.G., Tozaki, H., et al. (2017). CLK-dependent exon recognition and conjoined gene formation revealed with a novel small molecule inhibitor. Nature communications 8, 7.

Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. Cell 153, 17–37.

Gomez-Benito, M., Loayza-Puch, F., Oude Vrielink, J.A., Odero, M.D., and Agami, R. (2011). 3'UTR-mediated gene silencing of the Mixed Lineage Leukemia (MLL) gene. PloS one 6, e25449.

Gonda, T. and Ramsay, R. (2015). Directly targeting transcriptional dysregulation in cancer. Nature Reviews Cancer 15, 686-694

Haslam A., Prasad V. (2019). Estimation of the Percentage of US Patients With Cancer Who Are Eligible for and Respond to Checkpoint Inhibitor Immunotherapy Drugs. JAMA network open 2, e192535.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell 38, 576-589.

Heitman, L.H., van Veldhoven, J.P., Zweemer, A.M., Ye, K., Brussee, J., and AP, I.J. (2008). False positives in a reporter gene assay: identification and synthesis of substituted N-pyridin-2-ylbenzamides as competitive inhibitors of firefly luciferase. Journal of medicinal chemistry 51, 4724-4729.

Hira, Z.M., and Gillies, D.F. (2015). A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. Advances in Bioinformatics 2015, 198363.

Ito, M., Tanaka, T., Toita, A., Uchiyama, N., Kokubo, H., Morishita, N., Klein, M. G., Zou, H., Murakami, M., Kondo, M., et al. (2018). Discovery of 3-Benzyl-1-(trans-4-((5-cyanopyridin-2-yl)amino)cyclohexyl)-1-arylurea Derivatives as Novel and Selective Cyclin-Dependent Kinase 12 (CDK12) Inhibitors. Journal of Medicinal Chemistry 13, 7710-7728.

Jenkins R. W., Barbie D. A., Flaherty K. T. (2018). Mechanisms of resistance to immune checkpoint inhibitors. British journal of cancer 118, 9–16.

Jenal, M., Elkon, R., Loayza-Puch, F., Van Haaften, G., Kühn, U., Menzies, F.M., Vrielink, J.a.F.O., Bos, A.J., Drost, J., Rooijers, K., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149, 538-553.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468, 664-668.

Kanehisa, M. and Sato, Y. (2019), KEGG Mapper for inferring cellular functions from protein sequences. Protein Science, August 18 2019.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods 7, 1009-1015.

Klerk, E.D., and Hoen, P.a.C. (2015). Alternative mRNA transcription , processing , and translation : insights from RNA sequencing. Trends in Genetics, 1-12.

Klebanov, L., and Yakovlev, A. (2007). How high is the level of technical noise in microarray data? Biology Direct 2, 9.

Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.a., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M., et al. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. EMBO Journal 33, 878-889.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., et al. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495–501.

Lengauer, C., Kinzler, K. W. and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature 398, 643-649.

Li, H., Lovci, M. T., Kwon, Y. S., Rosenfeld, M. G., Fu, X. D., and Yeo, G. W. (2008). Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model. Proceedings of the National Academy of Sciences of the United States of America, 105, 20179–20184.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25, 2078-2079.

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L., et al. (2015). Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS genetics 11, e1005166.

Liu, X., Freitas, J., Zheng, D., Oliveira, M.S., Hoque, M., Martins, T., Henriques, T., Tian, B., and Moreira, A. (2017). Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster. RNA (New York, NY) 23, 1807-1816.

Luo, W., Ji, Z., Pan, Z., You, B., Hoque, M., Li, W., Gunderson, S.I., and Tian, B. (2013). The conserved intronic cleavage and polyadenylation site of CstF-77 gene imparts control of 3' end processing activity through feedback autoregulation and by U1 snRNP. PLoS genetics 9, e1003613.

Manning, K.S., and Cooper, T.A. (2017). The roles of RNA processing in translating genotype to phenotype. Nature reviews. Molecular cell biology 18, 102–114.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics (Oxford, England) 21, 3448-3449.

Marguerat, S., and Bähler, J. (2010). RNA-seq: from technology to biology. Cellular and molecular life sciences 67, 569–579.

Masamha, C. P., and Wagner E. J. (2018). The contribution of alternative polyadenylation to the cancer phenotype. Carcinogenesis 39, 2-10.

Mazloomian, A., Araki, S., Ohori, M., El-Naggar, A. M., Yap, D., Bashashati, A., and Aparicio, S. (2019). Pharmacological systems analysis defines EIF4A3 functions in cell-cycle and RNA stress granule formation. Communications biology 2, 165.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS one 5, e13984.

Miki, H., Okada, Y., and Hirokawa, N. (2005). Analysis of the kinesin superfamily: insights into structure and function. Trends Cell Biology 15, 467-476

Morgan P., Brown D., Lennard S., Anderton M., Barrett J., Eriksson U., Fidock M., Hamrén B., Johnson A., et al. (2018). Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nature Reviews Drug Discovery 17, 167-181.

Ohashi, A., Zdzienicka, M.Z., Chen, J., and Couch, F.J. (2005). Fanconi anemia complementation group D2 (FANCD2) functions independently of BRCA2- and RAD51-associated homologous recombination in response to DNA damage. Journal of Biological Chemistry 280, 14877-14883

Ohashi, A., Ohori M., Iwai, K., Nakayama, Y., Nambu, T., et al. (2015). Aneuploidy generates proteotoxic stress and DNA damage concurrently with p53-mediated post-mitotic apoptosis in SAC-impaired cells. Nature Communications 6, 7668

Pal, S., Gupta, R., and Davuluri, R.V. (2012). Alternative transcription and alternative splicing in cancer. Pharmacology & Therapeutics 136, 283-294.

Paz, I., Kosti, I., Ares, M., Jr., Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic acids research 42, W361-367.

Peterson, M.L. (2007). Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development. Immunologic research 37, 33-46.

Prakash, T., Sharma, V.K., Adati, N., Ozawa, R., Kumar, N., Nishida, Y., Fujikake, T., Takeda, T., and Taylor, T.D. (2010). Expression of conjoined genes: another mechanism for gene regulation in eukaryotes. PloS one 5, e13284.

Rath, O., and Kozielski, F. (2012). Kinesins and cancer. Nature reviews. Cancer, 12(8), 527–539.

Ribas A., Wolchok J.D. (2018). Cancer immunotherapy using checkpoint blockade. Science 359, 1350-1355.

Rosario, S. R., Long, M. D., Affronti, H. C., Rowsam, A. M., Eng, K. H., and Smiraglia, D. J. (2018). Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. Nature communications 9, 5330.

Salton, M., and Misteli, T. (2016). Small Molecule Modulators of Pre-mRNA Splicing in Cancer Therapy. Trends in molecular medicine 22, 28-37.

Schena, M., Shalon, D., Davis, RW. and Brown PO. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467-470.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics (Oxford, England) 27, 863-864.

Shao, Z., Zhang, Y., Yuan, G.C., Orkin, S.H., and Waxman, D.J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome biology 13, R16.

Shi, Y. (2012). Alternative polyadenylation: new insights from global analyses. RNA (New York, NY) 18, 2105-2117.

Siu Tony, D.C., Kumarasinghe Sathyajith E, inventors (2011). Merck Sharp & Dohme Corp, assignee. Pyrazolo [3,4-b] pyridin-4-one kinase inhibitors. patent WO/2011/049722.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature 458, 719–724.

Sur, I., and Taipale, J. (2016). The role of enhancers in cancer. Nature Reviews Cancer 16, 483–493.

Thompson L.L., Jeusset L.M., Lepage C.C., McManus K.J. (2017). Evolving Therapeutic Strategies to Exploit Chromosome Instability in Cancer. Cancers (Basel) 9, E151.

Tian, B., and Manley, J.L. (2013). Alternative cleavage and polyadenylation: The long and short of it. Trends in Biochemical Sciences 38, 312-320.

Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome research 17, 156-165.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7, 562-578.

Uhlen M., Zhang C., Lee S., Sjöstedt E., Fagerberg L., Bidkhori G., Benfeitas R., et al. (2017). A pathology atlas of the human cancer transcriptome. Science 18, eaan2507.

Vicente-Dueñas, C., Romero-Camarero, I., Cobaleda, C., and Sánchez-García, I. (2013). Function of oncogenes in cancer development: a changing paradigm. The EMBO journal 32, 1502–1513.

Villicaña, C., Cruz, G., and Zurita, M. (2014). The basal transcription machinery as a target for cancer therapy. Cancer cell international 14, 18.

Wang, J., Liu, Q., and Shyr, Y. (2015). Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. BMC genomics 16 Suppl 7, S5.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics 10, 57-63.

Wodicka, L, Dong, H., Mittmann, M., Ho, MH., and Lockhart, DJ. (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. Nature Biotechnology 15, 1359-1367.

Wood K.W., Lad L., Luo L., Qian X., Knight S.D., Nevins N., Brejc K., Sutton D., Gilmartin A.G., Chua P.R., Desai R., Schauer S.P., et al. (2010). Antitumor activity of an allosteric inhibitor of centromere-associated protein-E. Proceedings of the National Academy of Sciences of the United States of America 107, 5839-44.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics (Oxford, England) 26, 873-881.

Wu, Z. (2009)A review of statistical methods for preprocessing oligonucleotide microarrays. Statistical methods in medical research 18, 533-541.

Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nature communications 5, 5274.

Yao, C., Choi, E.-A., Weng, L., Xie, X., Wan, J., Xing, Y., Moresco, J.J., Tu, P.G., Yates, J.R., and Shi, Y. (2013). Overlapping and distinct functions of CstF64 and CstF64 τ in mammalian mRNA 3' processing. RNA (New York, NY) 19, 1781-1790.

Yoshimoto, R., Kaida, D., Furuno, M., Burroughs, A.M., Noma, S., Suzuki, H., Kawamura, Y., Hayashizaki, Y., Mayeda, A., and Yoshida, M. (2017). Global analysis of pre-mRNA subcellular localization following splicing inhibition by spliceostatin A. RNA (New York, NY) 23, 47-57.

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology 16, 284–287.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology 9, R137.