

ARTICLE

Comprehensive Characterization of Alternative Polyadenylation in Human Cancer

Yu Xiang*, Youqiong Ye*, Yanyan Lou, Yang Yang, Chunyan Cai, Zhao Zhang, Tingting Mills, Ning-Yuan Chen, Yoonjin Kim, Fatma Muge Ozguc, Lixia Diao, Harry Karmouty-Quintana, Yang Xia, Rodney E. Kellems, Zheng Chen, Michael R. Blackburn, Seung-Hee Yoo, Ann-Bin Shyu, Gordon B. Mills, Leng Han

Affiliations of authors: Department of Biochemistry and Molecular Biology (YXiang, YYe, ZZ, TM, NYC, YK, FMO, HKQ, YXia, REK, ZC, MRB, SHY, ABS, LH) and Department of Internal Medicine (CC), McGovern Medical School at The University of Texas Health Science Center at Houston, Houston, TX; Division of Hematology and Oncology, Mayo Clinic, Jacksonville, FL (YL); Department of Bioinformatics and Computational Biology (YYa, LD) and Department of Systems Biology (GBM), The University of Texas MD Anderson Cancer Center, Houston, TX.

*Authors contributed equally to this work.

Correspondence to: Leng Han, PhD, Department of Biochemistry and Molecular Biology, McGovern Medical School at The University of Texas Health Science Center at Houston, 6431 Fannin, MSB 6.166, Houston, TX 77030 (e-mail: leng.han@uth.tmc.edu).

Abstract

Background: Alternative polyadenylation (APA) is emerging as a major post-transcriptional mechanism for gene regulation, and dysregulation of APA contributes to several human diseases. However, the functional consequences of APA in human cancer are not fully understood. Particularly, there is no large-scale analysis in cancer cell lines.

Methods: We characterized the global APA profiles of 6398 patient samples across 17 cancer types from The Cancer Genome Atlas and 739 cancer cell lines from the Cancer Cell Line Encyclopedia. We built a linear regression model to explore the correlation between APA factors and APA events across different cancer types. We used Spearman correlation to assess the effects of APA events on drug sensitivity and the Wilcoxon rank-sum test or Cox proportional hazards model to identify clinically relevant APA events.

Results: We revealed a striking global 3'UTR shortening in cancer cell lines compared with tumor samples. Our analysis further suggested PABPN1 as the master regulator in regulating APA profile across different cancer types. Furthermore, we showed that APA events could affect drug sensitivity, especially of drugs targeting chromatin modifiers. Finally, we identified 1971 clinically relevant APA events, as well as alterations of APA in clinically actionable genes, suggesting that analysis of the complexity of APA profiles could have clinical utility.

Conclusions: Our study highlights important roles for APA in human cancer, including reshaping cellular pathways and regulating specific gene expression, exemplifying the complex interplay between APA and other biological processes and yielding new insights into the action mechanism of cancer drugs.

3' end processing and polyadenylation is an essential step of messenger RNA (mRNA) maturation. Recent transcriptome-wide studies have revealed at least 70% of human genes contain multiple polyadenylation sites, suggesting that alternative polyadenylation (APA) is highly prevalent (1,2). The 3' untranslated

region (3' UTR) of mRNA harbors key RNA regulatory elements (3). Therefore, APA provides a means to engage or evade 3' UTR mediated regulatory controls on mRNA metabolism (2,4). A large number of APA factors are required for appropriate polyadenylation processing, including cleavage and polyadenylation

Received: February 21, 2017; Revised: August 15, 2017; Accepted: September 20, 2017

© The Author 2017. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

specificity factor (CPSF), cleavage stimulation factor (CSTF), cleavage factors I and II (CFI and CFII), as well as the poly(A) polymerase and other associated factors (5).

APA is dynamically regulated in response to a variety of physiological conditions (6,7), and recent studies have highlighted the important role of APA in cancer (8–11). However, the regulation and functional consequences of APA alterations in tumorigenesis are not fully understood. Here, we systematically analyzed global patterns of APA events in samples from The Cancer Genome Atlas (TCGA) (12) and Cancer Cell Lines Encyclopedia (CCLE) (13).

Methods

Data Resource

We downloaded pair-ended RNA-seq BAM files of 5838 patients tumor samples across 17 TCGA cancer types and 560 paired nontumor tissues samples (if available), as well as 739 RNA-seq BAM files of cancer cell lines from the National Cancer Institute's Genomic Data Commons (GDC) (<https://gdc.cancer.gov>) (12–14), as previously described (15). We also downloaded the gene expression data from GDC. In addition, we obtained the RNA-seq files of 531 samples from three other large-scale independent studies, including lung adenocarcinoma, renal cell carcinoma, and glioblastomas (16–18). To avoid the potential batch effects introduced by gene expression quantification, we recalculated gene expression as reads per kilobase per million mapped reads (RPKM) across all TCGA and CCLE samples. We then normalized the expression data using quantile normalization (19). We further obtained the RNA-seq files of 8 cancer cell lines from ENCODE (ENCSR000EYR, ENCSR310FIS, ENCSR000CPT) (20) and GEO (GSE42420) (9). We obtained the tumor purity data (21) and clinically actionable genes (22) from previous studies.

Characterization of APA Events From RNA-Seq Data

To characterize the dynamic APA events using standard RNA-seq from TCGA data sets, we utilized the well-established algorithm DaPars (<https://github.com/ZhengXia/DaPars>) to identify the alternative proximal polyA site and calculate the Percentage of Distal polyA site Usage Index (PDUI) for each transcript. Detailed methods are in the Supplementary Materials (available online).

Analysis of APA Core Factors and APA Events

To explore the correlation between APA factors and APA events in tumor samples across different cancer types with sample sizes of 100 or greater, we built a linear regression model to assess the relationship between the expression of APA factors and PDUI for each transcript. Detailed methods are in the Supplementary Materials (available online).

Effects of APA Events on Drug Sensitivity

To assess the effects of APA events on drug sensitivity, we downloaded the drug sensitivity data from the CCLE (<http://www.broadinstitute.org/ccle/home>) data portal (13) and Cancer Therapeutics Response Portal (CTRP; <http://portals.broadinstitute.org/ctrp>) (23). We used Spearman correlation to calculate

the correlations between the PDUI of transcripts and the IC50 from CCLE or the area under the curve (AUC) from CTRP, as previously described (24). Please see the Supplementary Methods (available online).

Identification of Clinically Relevant APA Events

The clinically relevant APA events were identified as previously described (24). We obtained the clinical information associated with tumor samples, including tumor subtypes (defined as a distinct groups of patients based on molecular features or clinical features), disease stage (the extent of a cancer in the body, which is usually based on the size of the tumor), and patient overall survival time (range = 1–10 870 days) from TCGA maker papers or the GDC data portal (<https://gdc-portal.nci.nih.gov/>) (24,25). Please see the Supplementary Methods (available online).

Statistical Analysis

To compare PDUI or gene expression between two or more groups, we used the Wilcoxon rank-sum test or Kruskal-Wallis nonparametric analysis of variance (ANOVA) test. For correlation, we used Spearman correlation (R_s). For survival analysis, we used the Cox proportional hazards model, and survival curves were compared using the two-sided log-rank test. We used Benjamini and Hochberg adjustment for false discovery rate (FDR) (26) and considered an FDR of less than 0.05 as statistically significant in analyses where an FDR is presented. All statistical tests were two-sided, and P values of less than .05 were considered statistically significant.

Results

Global Shortening of APA Events in Cancer Cell Lines

To systematically investigate the global APA landscape in cancer patients and cancer cell lines, we utilized a well-established computational algorithm, DaPars (10), and identified a large number of APA events (Table 1). The number of samples and a detailed abbreviation name for each cancer type are listed in Table 1. Among these APA events, 2.4% in OV to 11.8% in KIRP are negatively correlated with their gene expression (Spearman correlation $R_s < -0.3$, $FDR < 0.05$) (Supplementary Table 1, available online), perhaps because the APA event is only one of the factors (eg, DNA methylation, transcription factor, etc.) regulating gene expression. Consistent with a previous study (10), we observed an overall shortening pattern in tumor samples vs paired normal samples (Figure 1; Supplementary Figure 1A, available online). We also observed an even more extensive global APA shortening in cancer cell lines than in tumor samples (Figure 1A; Supplementary Table 2, available online), ranging from 831 (14.4%) in Lung squamous cell carcinoma (LUSC) to 2027 (36.3%) in THCA (Figure 1B). We further examined the percentage of transcripts with preferred longer 3' UTR isoform ($PDUI > 0.5$) in cancer cell lines across different cancer types, which ranged from 59.6% in THCA to 63.8% in KIRC. These are statistically significantly lower than in tumor samples, which ranged from 66.7% in LUSC to 73.2% in OV (Wilcoxon test $P = 7.4 \times 10^{-7}$) (Figure 1C), while there is no statistically significant difference between tumor and normal comparison (Supplementary Figure 1, B and C, available online). We observed that APA difference partially contributes to mRNA

Table 1. Summary of alternative polyadenylation across different cancer types*

Data source	Sample type	Normal	Tumor	Read length	No. of detected APA events	No. of average APA events	
						Normal	Tumor
TCGA	Bladder urothelial carcinoma (BLCA)	19	258	48	13 119	6909	7276
TCGA	Brain lower grade glioma (LGG)	0	486	48	13 267	NA	8206
TCGA	Breast invasive carcinoma (BRCA)	105	837	50	14 537	8498	8244
TCGA	Skin cutaneous melanoma (SKCM)	0	288	48	12 929	NA	7756
TCGA	Glioblastoma multiforme (GBM)	0	154	76	12 891	NA	8599
TCGA	Head and neck squamous cell carcinoma (HNSC)	42	426	48	13 197	6934	7412
TCGA	Kidney chromophobe (KICH)	25	66	48	11 888	8664	7363
TCGA	Kidney renal clear cell carcinoma (KIRC)	67	448	50	13 589	8397	8174
TCGA	Kidney renal papillary cell carcinoma (KIRP)	30	198	48	12 684	7795	7302
TCGA	Liver hepatocellular carcinoma (LIHC)	50	200	48	11 650	5123	6196
TCGA	Lung adenocarcinoma (LUAD)	58	490	48	13 737	7216	7383
TCGA	Lung squamous cell carcinoma (LUSC)	17	220	50	13 567	8387	8158
TCGA	Ovarian serous cystadenocarcinoma (OV)	0	412	75	14 338	NA	8234
TCGA	Prostate adenocarcinoma (PRAD)	52	378	48	12 527	7956	7555
TCGA	Stomach adenocarcinoma (STAD)	33	285	75	13 910	6647	8031
TCGA	Thyroid carcinoma (THCA)	59	496	48	12 656	8237	7929
TCGA	Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC)	3	196	48	12 700	NA	7623
CCLL	Cancer cell lines	0	739	101	11 503	NA	8347
Other	Chinese_GBM	0	274	101	11 837	NA	7871
Other	Japanese_KIRC	0	97	100	12 275	NA	8933
Other	Korean_LUAD	77	83	101	11 967	NA	7895

*APA = alternative polyadenylation; TCGA = The Cancer Genome Atlas.

expression difference (APA shortening with upregulated mRNA expression or APA lengthening with downregulated mRNA expression), ranging from 36.0% in SKCM to 49.7% in LGG (Supplementary Figure 1D, available online).

To rule out potential bias, we showed that cancer cell lines displayed a more shortening APA profile than tumor samples (Supplementary Figure 2, A and B, available online) regardless of the read length or coverage. We also observed overall weak or no correlation between tumor purity (21) and percentage of transcripts with PDIUI greater than 0.5 across cancer types (Supplementary Figure 2C, available online), suggesting that tumor purity has little effect on our results.

To characterize the biological function of differentially regulated APA events, we performed functional enrichment analysis and observed that shortening APA events are highly enriched in gene ontology (GO) terms such as protein localization, transport, and cell cycle (Supplementary Figure 2D and Supplementary Table 3, available online), indicating that these cellular processes may be regulated by alterations of APA events and thus contribute to the distinctive features of cancer cell lines. Interestingly, these APA events are also highly enriched in GO terms related to the metabolic process, suggesting that APA events may be involved in different metabolic processes between cultured cancer cell lines and tumor samples (27).

Master Regulators for APA Events in Cancer

The 3' end-processing machinery is composed of multiple protein factors, including four biochemically distinct protein complexes (CPSF, CSTF, CFI, and CFII) and several single proteins, such as PABPN1, RBBP6, and SYMPK (5). Among these, 22 genes are considered core factors (5). Previous reports have indicated that alterations in expression and/or activities of APA factors can influence APA site selection in specific cell types, such as

CSTF2 (28) and NUDT21 (9). However, it is still unclear which APA factor(s) play the most important roles in APA regulation in human cancer. To address this issue, we used a linear regression model to evaluate relationships between 22 core APA factors with global APA events in tumor samples. Through positive correlation analysis, we found that PABPN1 may act as a dominant master regulator that broadly promotes usage of distal APA site in almost all cancer lineages, ranging from 14.5% in GBM to 39.8% in KIRC, which is much stronger than any other factors (Figure 2A). Based on limited cancer types or tissue types (sample size > 100), we did not observe that PABPN1 ranks as the top regulator in normal samples or in cancer cell lines. It is necessary to further investigate this with a larger number of samples. In the negative correlation analysis, the contribution of different APA regulators to global APA events varies among cancer types, suggesting that APA factors that promote the usage of proximal APA sites are more likely to be cancer specific or tissue specific (Supplementary Figure 3A, available online). Furthermore, we collected 63 more proteins identified in 3' end cleavage and polyadenylation processing complex (5) and performed similar analysis. We observed that PABPN1 still ranks as the top, further suggesting PABPN1 as the master regulator in APA regulation (Supplementary Figure 3B, available online). Indeed, PABPN1 knockdown led to global shortening of APA events in U2OS, a human bone osteosarcoma epithelial cell line (29), and C2C12, a mouse myoblast cell line (30) (Figure 2B). Our results highlighted the importance of PABPN1 as a general determinant of usage of distal APA sites through pan-cancer analysis, likely due to PABPN1 acting prior to any other APA factors, and it can function well alone to bind the proximal APA site (29) or affect APA selection through its role in RNA stability (Figure 2C) (30,31).

We further extended our study to genome-wide analysis. PABPN1 ranks as 48th (top 0.2%) among all genes and ranks

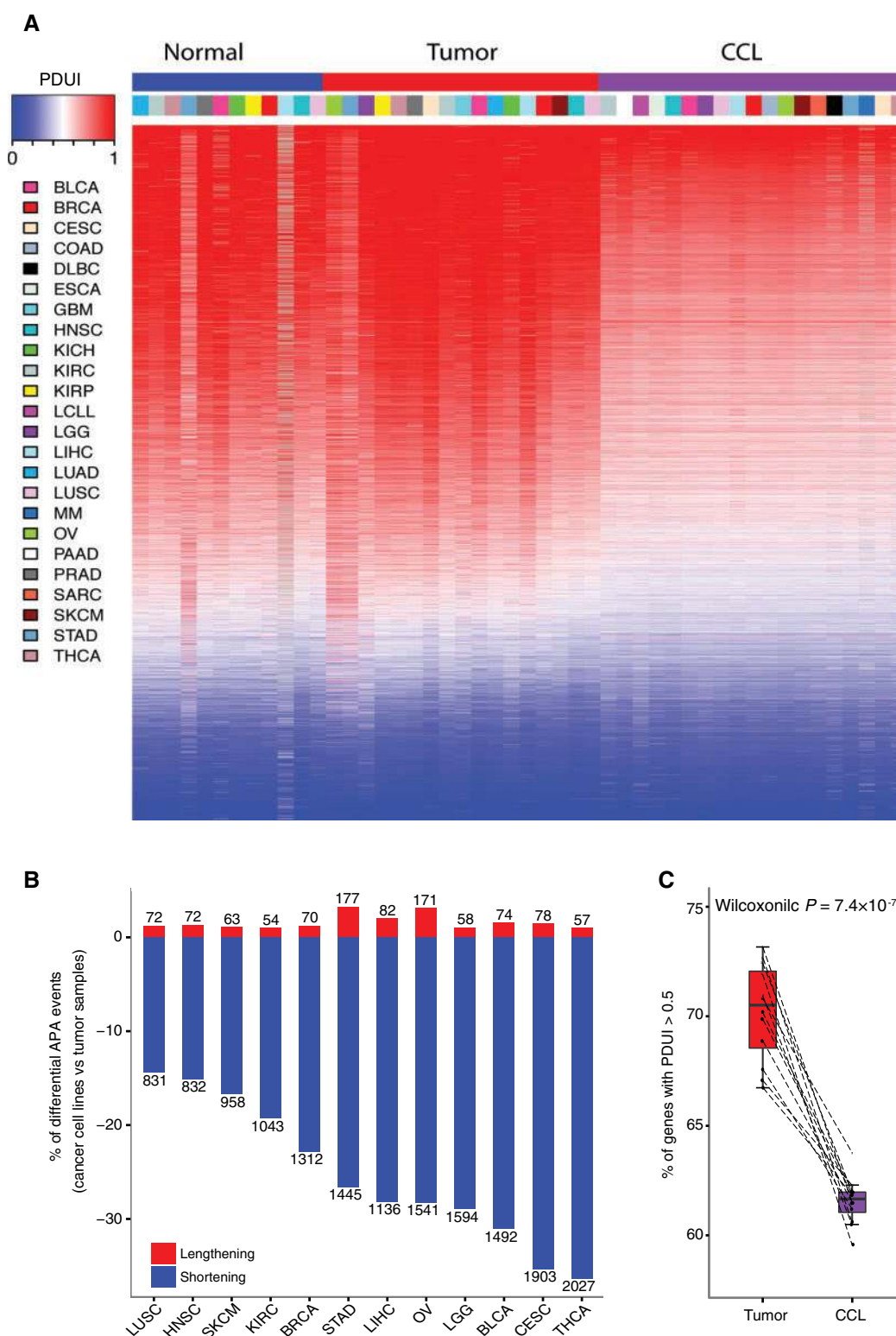


Figure 1. Global landscape of alternative polyadenylation (APA) events across patient samples and cancer cell lines. **A)** Global landscape of 7029 APA events across different cancer types in tumor samples (blue), normal samples (red), and cancer cell lines (purple). In the heatmap, red denotes transcripts with Percentage of Distal polyA site Usage Index (PDUI) greater than 0.5, while blue denotes transcripts with PDUI of less than 0.5. The columns are arranged by cancer type, and rows are arranged by mean PDUI of each APA event across all samples. **B)** Number of shortening (blue) and lengthening (red) APA events in cancer cell lines compared with tumor samples across 12 cancer types. **C)** Comparison of percentage of transcripts with PDUI greater than 0.5 divided by total number of APA events between tumor samples (red) and cancer cell lines (purple). The P value was calculated using a two-sided Wilcoxon test. The boxes show the median ± 1 quartile, with whiskers extending to the most extreme data point within 1.5 interquartile range from the box boundaries. APA = alternative polyadenylation; CCL = cancer cell lines; PDUI = Percentage of Distal polyA site Usage Index.

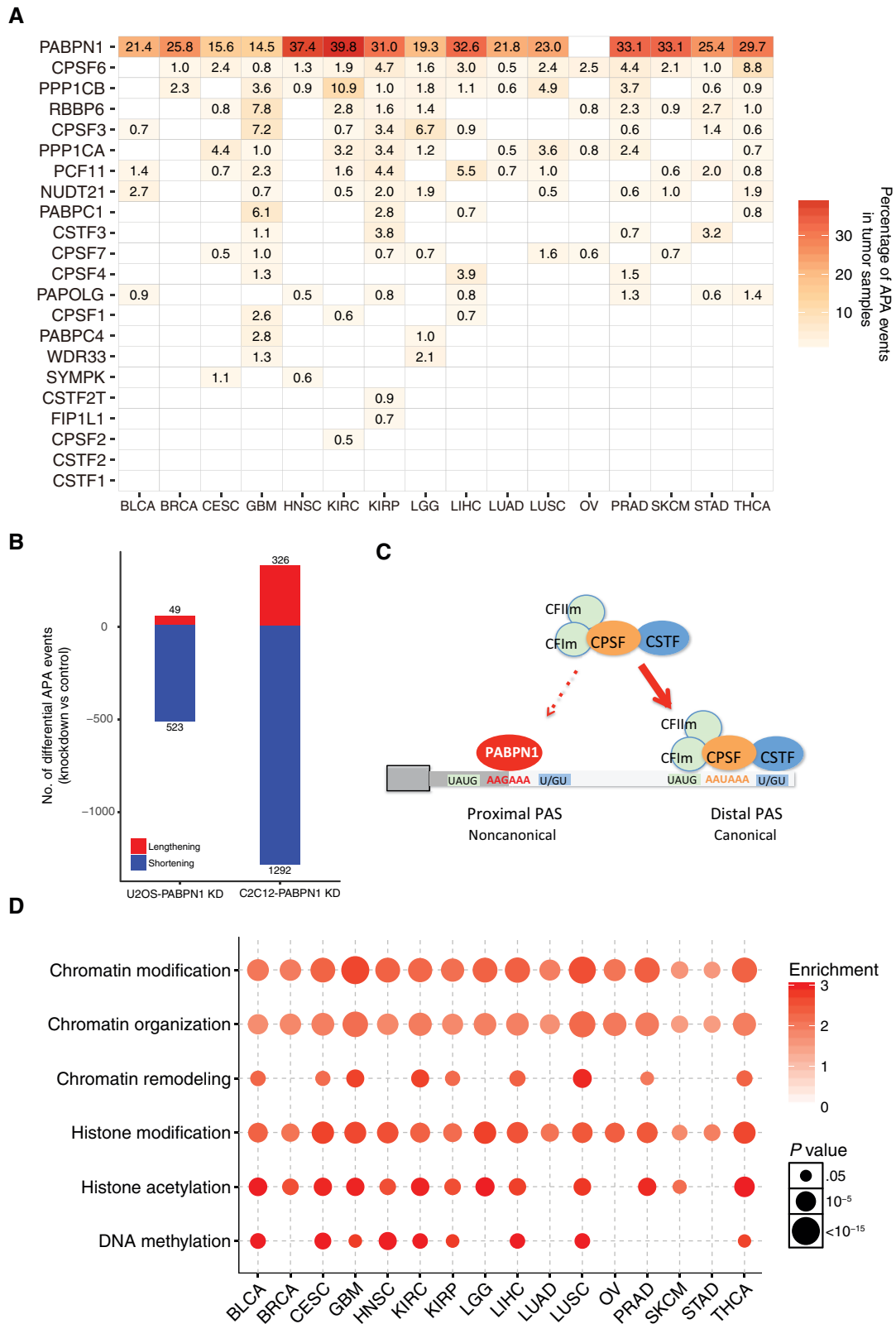


Figure 2. Key factors in regulation of alternative polyadenylation (APA) events. **A)** Percentage of APA events positively correlated with expression level of APA factors in tumor samples across cancer types. **B)** Global shortening of APA events upon PABPN1 knockdown in U2OS and C2C12 cell lines. **C)** A schematic model for the important role of PABPN1 in suppressing the usage of proximal APA site. **D)** Chromatin-related gene ontology (GO) terms enrichment analysis of top 10% of 20532 genes correlated to APA events in tumor samples. Gene lists were provided in Supplementary Table 4 (available online). The circle color represents the enrichment score, while the circle size represents the statistically significant P value. Enriched GO terms were listed in Supplementary Table 5 (available online). APA = alternative polyadenylation; CPSF = cleavage and polyadenylation specificity factor; CSTF = cleavage stimulation factor.

before any other APA factors (Supplementary Figure 3C, available online). Those genes that rank before PABPN1 are involved in more generalized biological processes such as RNA processing. We examined the top 10% genes of all 20532 genes (Supplementary Table 4, available online) correlated to APA events and observed that these genes are highly enriched for roles in chromatin-related categories (Figure 2D; Supplementary Table 5, available online), suggesting a coordinating regulation of chromatin structure modification and APA processes. We also observed several enriched GO terms including RNA splicing and processing (Supplementary Figure 3D and Supplementary Table 6, available online). Remarkably, we also observed enriched GO terms in mitochondrial electron transport and oxidative phosphorylation, suggesting a potential role of energy metabolism in influencing APA events or vice versa.

Therapeutic Liability of APA Events

The drug sensitivity profile provides a systematic approach to link drug response to genetic and/or cellular features, such as gene expression (13,23,32). However, no systematic relationship between drug responses and APA events has been reported. Therefore, we performed comprehensive analysis on correlation of sensitivity to 24 US Food and Drug Administration (FDA)-approved anticancer drugs from CCLE with APA events across 321 cancer cell lines. Despite drugs not being designed to target APA events, we identified 214 statistically significantly correlated APA-drug pairs ($|Rs| > 0.3$, $FDR < 0.05$), suggesting a critical role of APA events in response to anticancer drugs. Interestingly, we found that top drugs correlated with APA events are pan-histone deacetylase (HDAC) inhibitors, such as Panobinostat, and DNA topoisomerase I inhibitors, such as Topotecan and Irinotecan (Figure 3A). We further observed global shortening in HDAC inhibitor-treated samples (33), suggesting a direct link between APA and HDAC. We also observed slightly more shortening events than lengthening events in HDAC3-specific knockdown samples (34,35) (Supplementary Figure 4, available online). This may be due to the reason that other HDACs provide compensatory functions (36–39). These results suggest that chromatin-targeting drugs may act in part by modulating APA, consistent with our findings that chromatin modification and APA are closely linked.

To further confirm our findings, we explored the effect of APA events on drug sensitivity from the CTRP (13), which includes 481 small molecule compounds across 619 cancer cell lines. We identified 1516 APA events statistically significantly associated with drug sensitivity. We ranked the drugs based on the number of APA events correlated and found that the 40 top-ranked drugs are enriched in HDAC inhibitors (red, 5/40 vs 22/481, Fisher's exact test, $P = .047$) and topoisomerase inhibitors (blue, 3/40 vs 5/481, $P = .02$) (Figure 3B). The consistency with CCLE data highlighted the close link between chromatin modification and APA.

Clinical Relevance of APA Events

There is no comprehensive analysis to identify clinically relevant APA events in cancer. We identified 326 differentially regulated APA events in paired normal and tumor samples, 345 APA events associated with tumor subtypes, 66 APA events associated with tumor stages, and 1707 APA events associated with overall survival ($|Diff| \geq 0.2$, $FDR < 0.05$). We identified in total

1971 clinically relevant APA events (any of the above processes) (Supplementary Figure 5A and Supplementary Table 7, available online), and 37 of these events are in at least four cancer types (Figure 4A). By comparing PABPN1-dependent APA events obtained from previous reports (29–31), we observed that PABPN1-dependent APA events are enriched in clinically relevant APA events, suggesting the important role of PABPN1 in cancer (Supplementary Figure 5B, available online). Interestingly, tumor/normal differential APA events are also enriched in PABPN1-regulated APA events (Supplementary Figure 5C, available online). Among the differentially regulated APA events, most displayed a shortening pattern, which is consistent with our findings and a previous report (10). Most shortening APA events associated with worse survival, suggesting that use of a proximal polyadenylation site is associated with more aggressive status. For example, CSNK1D transcript exhibits statistically significant shortening, such as in LUSC ($Diff = -0.23$, $FDR = 1.74 \times 10^{-5}$) and KIRC ($Diff = -0.23$, $FDR = 1.15 \times 10^{-12}$) (Figure 4B). Shortening of CSNK1D transcript is associated with the worse survival in KIRC ($Diff = 0.27$, $FDR = 0.01$, hazard ratio [HR] = 0.60, 95% confidence interval [CI] = 0.48 to 0.75) (Figure 4B). Clinically, CSNK1D inhibition represents a promising therapeutic strategy in human breast cancer (40). Strikingly, we found that shortening of CSNK1D 3' UTR in tumor samples leads to loss of several predicted RNA regulatory elements, such as microRNA binding sites and AU-rich elements (AREs) (41,42) (Figure 4B), which could lead to upregulation of CSNK1D expression. Moreover, the CRTC1 plays a crucial role in tumor formation (43); its transcript undergoes statistically significant 3'UTR shortening, and its shortening is associated with poor survival in LGG ($Diff = 0.31$, $FDR = 0.004$, $HR = 0.75$, 95% CI = 0.65 to 0.87) (Supplementary Figure 5D, available online). These results suggest that these genes potentially utilize APA to escape RNA destabilized elements and thus to play oncogenic roles in cancer.

In contrast, several other genes displayed 3'UTR lengthening. For example, NDE1 showed lengthening in eight cancer types, as well as differential APA profile among tumor BRCA subtypes ($Diff = 0.51$, $FDR < 2.2 \times 10^{-16}$) and STAD stages ($Diff = 0.28$, $FDR = 0.01$) (Figure 4C). NDE1 can protect genome integrity during cell mitosis (44) and regulate cell cycle progression (45,46). Therefore, its lengthening may lead to enhanced negative regulation of NED1 expression through one more miRNA binding and several more AU-rich elements (Figure 4C), thus contributing to cancer genome instability and enhancing tumorigenesis. Other lengthening transcripts of interest are COL1A1 and COL1A2, which encode key components of type I collagen (Supplementary Figure 5, E and F, available online). Collagens are major components of the extracellular matrix and play an important role in tumor invasion and migration (47). Therefore, the lengthening of COL1A1 and COL1A2 may produce few proteins; thus it is involved in creating a tumorigenic micro-environment that facilitates tumor progression.

Alterations of APA Events in Clinically Actionable Genes

We further examined the APA profile of 135 clinically actionable genes, which includes therapeutic FDA-approved drug targets and/or their associated genes (22,48). Among these genes, 6.67% (GBM) to 31.9% (STAD) of them showed relatively large variance (interquartile range [IQR] ≥ 0.2), suggesting a novel layer of regulation of clinically actionable genes (Figure 5A). Several clinically actionable genes recurrently showed large variances

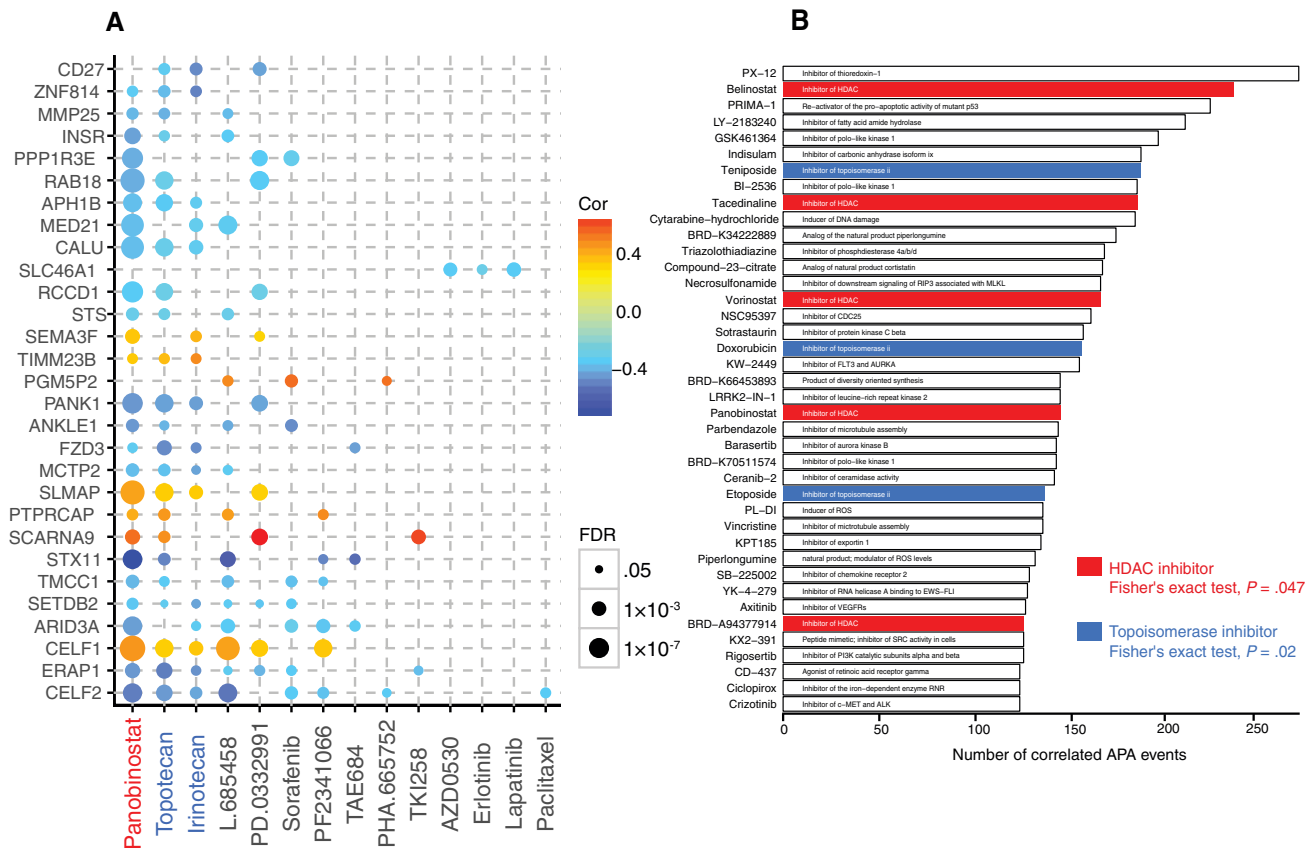


Figure 3. Therapeutic liability of alternative polyadenylation (APA) events in cancer cell lines. **A)** The statistically significant correlation of APA events with IC50 value of Cancer Cell Lines Encyclopedia drugs across cancer cell lines for three or more drugs. The circle color represents Spearman coefficient value, while the circle size represents the false discovery rate value. **B)** Top 40 drugs ranked based on the number of correlated APA events using Cancer Therapeutics Response Portal data: pan-histone deacetylase inhibitors, red. All statistical tests were two-sided. APA = alternative polyadenylation; HDAC = pan-histone deacetylase.

across cancer types (Figure 5B). For example, CTNNB1, the key component of the Wnt signaling pathway (49), showed variant APA (IQR ≥ 0.2) in 16 cancer types. PIK3R1, the target of PI3K inhibitors (50), showed variant APA in 14 cancer types (Figure 5B). PABPN1 regulated the largest number of clinically actionable genes, which further confirmed the important roles of PABPN1 in cancer (Figure 5C; Supplementary Table 8, available online). For example, the top recurrently actionable genes, CTNNB1, PI3KR1, and FGFR2, are highly correlated with PABPN1 mRNA expression level in THCA ($R_s = 0.66$, $P < 2.2 \times 10^{-16}$), LGG ($R_s = 0.47$, $P < 2.2 \times 10^{-16}$), and KIRP ($R_s = 0.66$, $P < 2.2 \times 10^{-16}$), respectively (Figure 5D). Knockdown of PABPN1 also led to shortening of CTNNB1 in U2OS and C2C12 cell lines. Among these APA events, CTNNB1 is strongly correlated with its APA status ($R_s = -0.41$, $P < 2 \times 10^{-16}$) and PIK3R1 is statistically significantly correlated with its APA status ($R_s = -0.18$, $P = 5 \times 10^{-5}$), while FGFR2 is not correlated with its APA status ($R_s = -0.11$, $P = .16$). The alterations of APA events in clinically actionable genes highlighted the necessity to consider the APA events in drug treatment.

Discussion

APA is an important and widespread event for gene regulation. With the large-scale RNA-seq data sets from both TCGA and CCLE, we provided a systemic view of the APA landscape in human cancer. Previous studies revealed global shortening of APA

in tumor samples and cancer cell lines compared with normal samples (8,10,11). Strikingly, we observed an extensive global shortening in cancer cell lines compared with tumor samples, and these transcripts that undergo APA shortening are enriched in several cellular pathways, including protein transport, metabolism, and cell cycle. Taken together, this suggests that APA may play important roles in reshaping cellular pathways and regulating specific gene expression patterns in cancer cell lines that have been grown in culture medium. Cancer cell lines derived from tumors are the most frequently used models to study cancer biology, especially for drug response prediction and new anticancer drug development (13,32,51,52). However, cancer cell lines also display unique features compared with tumor samples, such as gene expression (53,54) and mutation (53). Our analysis added APA as a new dimension for a unique feature of cancer cell lines, and further investigation is necessary.

It is still unclear which factors are key players, especially in the aberrant APA that occurs in cancer. We showed that PABPN1 functions as a master regulator that promotes distal poly(A) site usage in different cancers. Our finding is consistent with the notion that PABPN1 acts prior to any other APA-related factors, and it can work alone to prevent the binding of other APA factors to proximal poly(A) sites or affect APA selection through its role in RNA stability (29–31). A previous study identified CSTF2 as the major factor by comparing the expression level between tumor and normal samples (10), while our study mainly focused on the correlations between PDU1 and the expression level of APA factors within tumor samples. This inconsistency

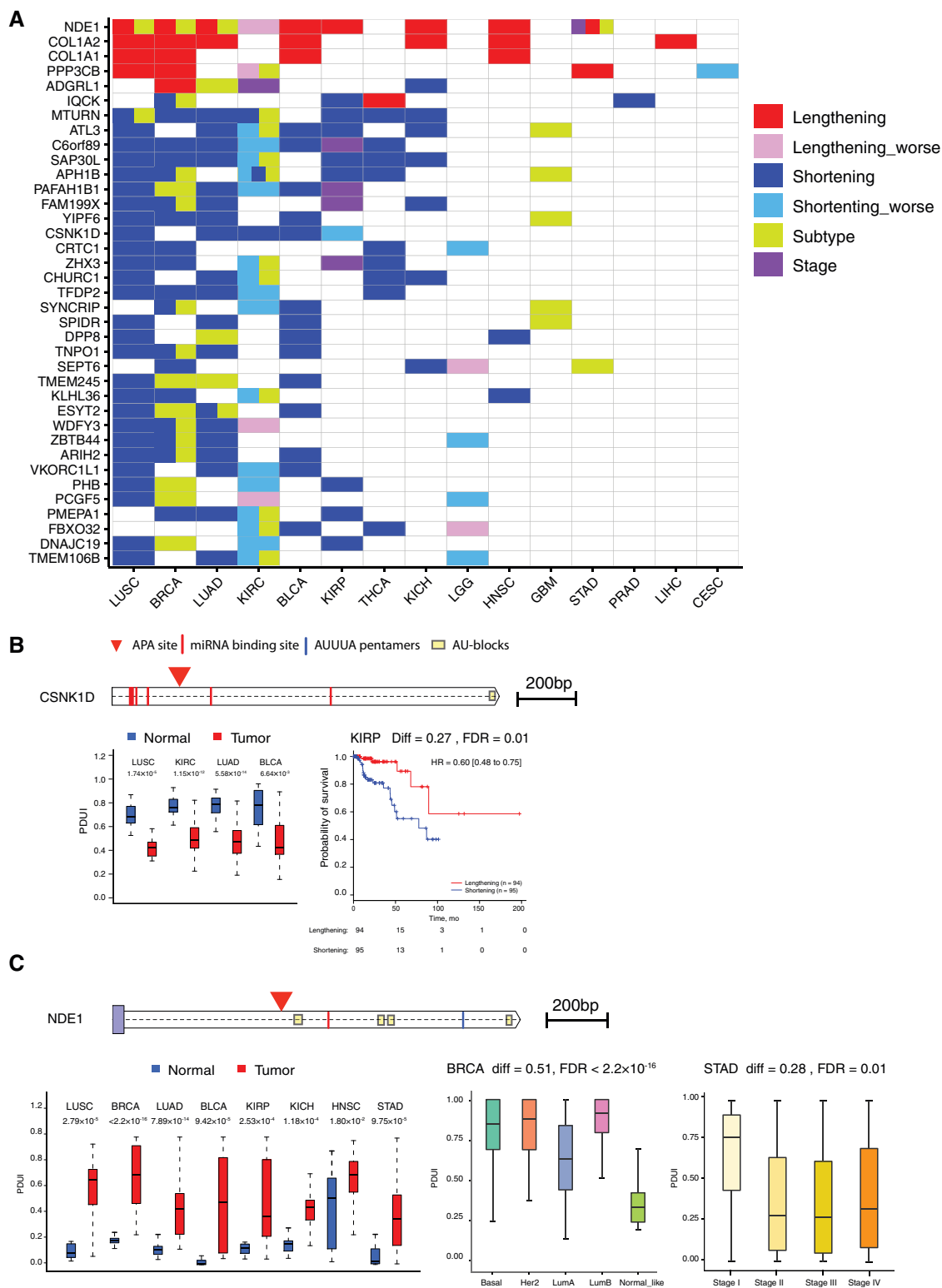


Figure 4. Clinical relevance of alternative polyadenylation (APA) events. **A)** Statistically significant differentially regulated and/or clinically relevant APA events identified in at least four cancer types. For each cancer type, the **red and blue box** indicates the statistically significant differential APA events between paired tumor and normal samples (red: Diff ≥ 0.2 , false discovery rate [FDR] < 0.05 ; blue: Diff ≤ -0.2 , FDR < 0.05), the **lemon yellow box** indicates the statistically significant differential APA events among tumor subtypes (Diff ≥ 0.2 , FDR < 0.05), the **purple box** indicates the statistically significant differential APA events among tumor stages (Diff ≥ 0.2 , FDR < 0.05), and the **pink and light blue** indicates the lengthening APA events associated with worse (pink) or better (light blue) overall survival (Diff ≥ 0.2 , FDR < 0.05). Complete clinically relevant APA events were listed in Supplementary Table 7 (available online). **B)** Schematic representation of gene last exon structure and predicted alternative polyadenylation site, miRNA binding site, and AU-rich elements (top). APA profile of CSNK1D in paired tumor and normal samples (left) and overall patient survival time for lengthening and shortening group (right). **C)** APA profile of NDE1 in paired tumor and normal samples (left), among BRCA tumor subtypes (middle), and STAD stages (right). The **boxes** show the median ± 1 quartile, with **whiskers** extending to the most extreme data point within 1.5 interquartile range from the box boundaries. APA = alternative polyadenylation; PDUI = Percentage of Distal polyA site Usage Index.

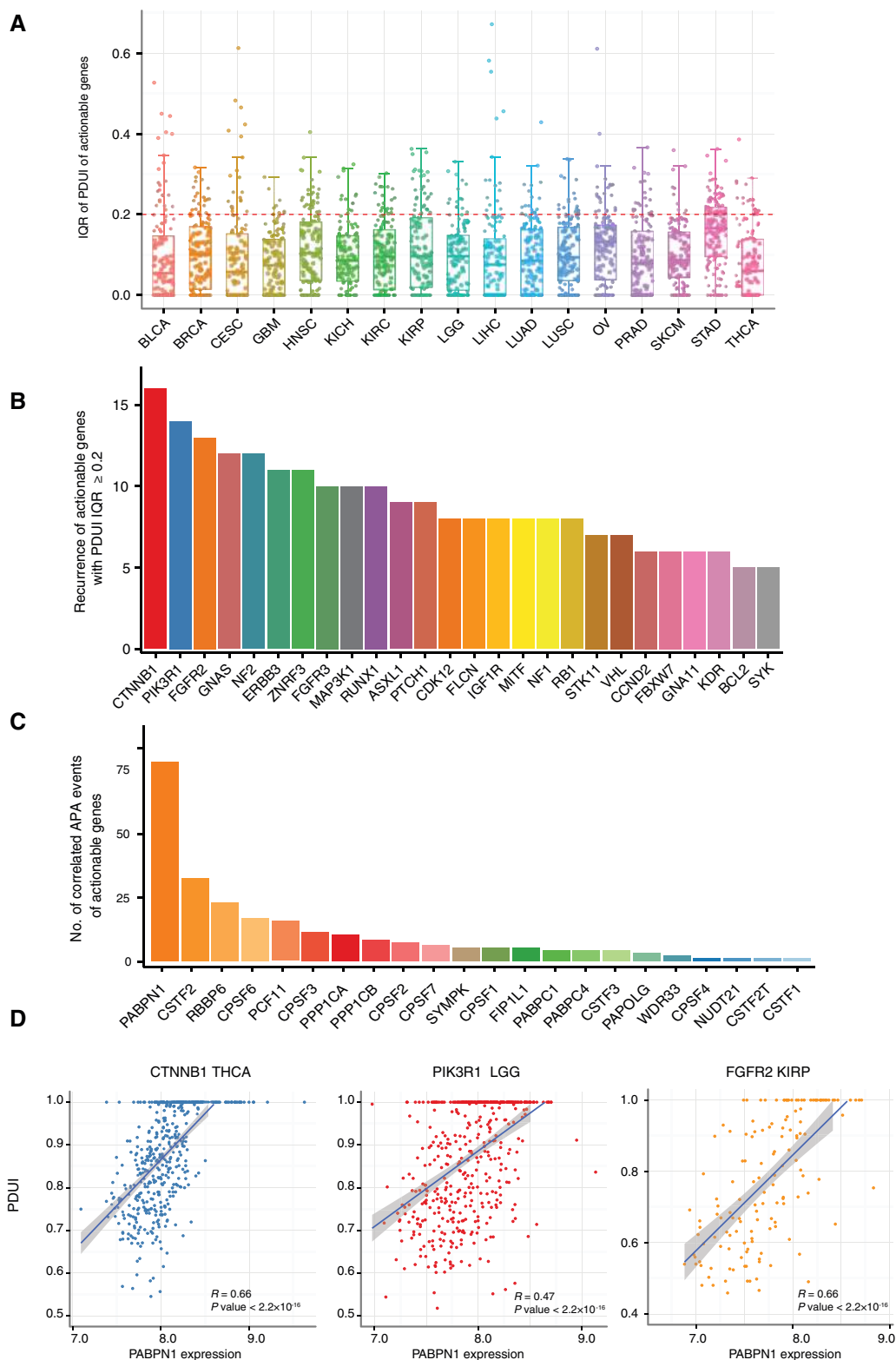


Figure 5. Alterations of alternative polyadenylation (APA) events in clinically actionable genes in tumor samples. **A)** Distribution of interquartile range (IQR) of Percentage of Distal polyA site Usage Index (PDUI) of clinically actionable genes across different cancer types. **B)** Recurrent clinically actionable genes with IQRs of PDUI of 0.2 or greater in tumor samples. **C)** Number of APA events of clinically actionable genes correlated with expression level of APA core factors. **D)** Selected correlations (two-sided Spearman correlation test) between APA events of clinically actionable genes and PABPN1 expression level (left: CTNNB1; middle: PIK3R1; right: FGFR2). Complete correlations were listed in Supplementary Table 8 (available online). APA = alternative polyadenylation; IQR = interquartile range; PDUI = Percentage of Distal polyA site Usage Index.

illustrates the complicated regulation of APA in cancer. As the master regulator, *PABPN1* showed overall diverse expression among different cancers (ANOVA $P = 2.5 \times 10^{-13}$) and LUSC ($P = .02$), and *PABPN1* expression is statistically significantly correlated with drug sensitivity of 31 drugs from the CTRP data set (Supplementary Figure 6, available online). Besides APA machinery-related components, we also identified other regulatory factors in APA regulation, such as factors involved in chromatin modification and remodeling. This is consistent with previous reports in human T cells (55) or yeast (56). More interestingly, genome-wide screening showed that mitochondrial electron transport and oxidative phosphorylation-related GO terms are highly associated with APA regulation. Considering that energy metabolism reprogramming is a hallmark of cancer (57) and ATP is a substrate of the polyadenylation process (58), our finding highlights the potential roles of energy metabolism as a new paradigm in the regulation of poly(A) site selection and vice versa.

We provided new evidence that differential APA profiles can affect the efficacy of drugs based on two independent pharmacogenomic data sets from CCLE and CTRP. We observed that the top drugs associated with APA events are HDAC inhibitors and/or DNA Topoisomerase inhibitors, highlighting the potential effects of those chromatin structure-targeted drugs on APA events, and that these drugs may yield new insights into the action mechanism of cancer drugs.

Finally, we identified an appreciable number of APA events with potential clinical relevance that could be potential biomarker candidates for further investigation. Despite the global shortening of APA profiles across tumor samples, we still observed clinically relevant lengthening APA events, such as *NDE1*, which functions as a guardian of genome integrity (45). We also identified a great variance of APA events among multiple clinically actionable genes, such as *CTNNB1*. Our results highlighted the importance and complexity of APA events in cancer treatment.

This study has some limitations. Despite the large computational effort, our analysis was the initial step to understand the regulation of APA events across different cancer types. We did not include all molecular measurements in analysis, such as miRNA expression. Moreover, further functional experiments are necessary to understand the molecular mechanisms of APA in cancer.

Funding

We gratefully acknowledge contributions from the TCGA Research Network. This work was supported by the Cancer Prevention and Research Institute of Texas (RR150085 to LH), the National Institutes of Health (GM046454 to ABS, UL1 TR000371 to CC, AG045828 to ZC, and GM114424 to SHY), an endowment from the Houston Endowment, Inc. (to ABS), The Welch Foundation (AU-1731 to ZC), the Biomedical Research Grant (RG-414673), the American Lung Association, and the UTHealth Pulmonary Center of Excellence Discovery Award (to HKQ).

Notes

The study sponsors had no role in design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Author contributions: LH conceived and designed the study; YXiang, YYe, ZZ, YYang, LD, CC, TM, NC, and YK performed data analysis; YXiang, YL, FMO, HKQ, YXia, MRB, REK, ZC, SHY, ABS, GBM, and LH interpreted results; YXiang, GBM, and LH wrote the manuscript with input from other authors; LH supervised the whole project.

References

- Hoque M, Ji Z, Zheng D, et al. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods*. 2013;10(2):133–139.
- Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*. 2017;18(1):18–30.
- Chen CA, Shyu AB. Emerging themes in regulation of global mRNA turnover in cis. *Trends Biochem Sci*. In press.
- Di Giandomartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Mol Cell*. 2011;43(6):853–866.
- Shi Y, Di Giandomartino DC, Taylor D, et al. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell*. 2009;33(3):365–376.
- Ji Z, Lee JY, Pan Z, et al. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A*. 2009;106(17):7028–7033.
- Shepard PJ, Choi EA, Lu J, et al. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*. 2011;17(4):761–772.
- Sandberg R, Neilson JR, Sarma A, et al. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*. 2008;320(5883):1643–1647.
- Masamha CP, Xia Z, Yang J, et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*. 2014;510(7505):412–416.
- Xia Z, Donehower LA, Cooper TA, et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun*. 2014;5:5274.
- Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. 2009;138(4):673–684.
- Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45(10):1113–1120.
- Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603–607.
- Grossman RL, Heath AP, Ferretti V, et al. Toward a shared vision for cancer genomic data. *N Engl J Med*. 2016;375(12):1109–1112.
- Li J, Han L, Roebuck P, et al. TANRIC: An interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res*. 2015;75(18):3728–3737.
- Seo JS, Ju YS, Lee WC, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res*. 2012;22(11):2109–2119.
- Sato Y, Yoshizato T, Shiraiishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet*. 2013;45(8):860–867.
- Bao ZS, Chen HM, Yang MY, et al. RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res*. 2014;24(11):1765–1773.
- Bolstad BM, Irizarry RA, Astrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–193.
- Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
- Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6:8971.
- Van Allen EM, Wagle N, Stojanov P, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med*. 2014;20(6):682–688.
- Rees MG, Seashore-Ludlow B, Cheah JH, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*. 2016;12(2):109–116.
- Han L, Diao L, Yu S, et al. The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell*. 2015;28(4):515–528.
- Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014;158(4):929–944.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B (Methodol)* 1995; 57(1):289–300.
- Vander Heiden MG, DeBerardinis RJ. Understanding the intersections between metabolism and cancer biology. *Cell*. 2017;168(4):657–669.
- Takagaki Y, Seipelt RL, Peterson ML, et al. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell*. 1996;87(5):941–952.

29. Jenal M, Elkon R, Loayza-Puch F, et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*. 2012;149(3):538–553.
30. Li W, You B, Hoque M, et al. Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet*. 2015;11(4):e1005166.
31. de Klerk E, Venema A, Anvar SY, et al. Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res*. 2012;40(18):9089–9101.
32. Basu A, Bodycombe NE, Cheah JH, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*. 2013;154(5):1151–1161.
33. Nagaraja S, Vitanza NA, Woo PJ, et al. Transcriptional dependencies in diffuse intrinsic pontine glioma. *Cancer Cell*. 2017;31(5):635–652, e6.
34. Carpio LR, Bradley EW, McGee-Lawrence ME, et al. Histone deacetylase 3 supports endochondral bone formation by controlling cytokine signaling and matrix remodeling. *Sci Signal*. 2016;9(440):ra79.
35. Stengel KR, Zhao Y, Klus NJ, et al. Histone deacetylase 3 is required for efficient T cell development. *Mol Cell Biol*. 2015;35(22):3854–3865.
36. Ma P, Pan H, Montgomery RL, et al. Compensatory functions of histone deacetylase 1 (HDAC1) and HDAC2 regulate transcription and apoptosis during mouse oocyte development. *Proc Natl Acad Sci U S A*. 2012;109(8):E481–E489.
37. Chang S, McKinsey TA, Zhang CL, et al. Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development. *Mol Cell Biol*. 2004;24(19):8467–8476.
38. Montgomery RL, Davis CA, Potthoff MJ, et al. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev*. 2007;21(14):1790–1802.
39. Jamaladdin S, Kelly RD, O'Regan L, et al. Histone deacetylase (HDAC) 1 and 2 are essential for accurate cell division and the pluripotency of embryonic stem cells. *Proc Natl Acad Sci U S A*. 2014;111(27):9840–9845.
40. Rosenberg LH, Lafitte M, Quereda V, et al. Therapeutic targeting of casein kinase 1delta in breast cancer. *Sci Transl Med*. 2015;7(318):318ra202.
41. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120(1):15–20.
42. Spasic M, Friedel CC, Schott J, et al. Genome-wide assessment of AU-rich elements by the AREScore algorithm. *PLoS Genet*. 2012;8(1):e1002433.
43. Canetti G, Coni S, Della Guardia M, et al. The coactivator CRTCl promotes cell proliferation and transformation via AP-1. *Proc Natl Acad Sci U S A*. 2009;106(5):1445–1450.
44. Kim S, Zaghloul NA, Bubenshchikova E, et al. Nde1-mediated inhibition of cilogenesis affects cell cycle re-entry. *Nat Cell Biol*. 2011;13(4):351–360.
45. Houlihan SL, Feng Y. The scaffold protein Nde1 safeguards the brain genome during S phase of early neural progenitor differentiation. *Elife*. 2014;3:e03297.
46. Bakircioglu M, Carvalho OP, Khurshid M, et al. The essential role of centrosomal NDE1 in human cerebral cortex neurogenesis. *Am J Hum Genet*. 2011;88(5):523–535.
47. Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: Drivers of tumour metastasis. *Nat Rev Cancer*. 2014;14(6):430–439.
48. Yuan Y, Liu L, Chen H, et al. Comprehensive characterization of molecular differences in cancer between male and female patients. *Cancer Cell*. 2016;29(5):711–722.
49. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer*. 2013;13(1):11–26.
50. Liu P, Cheng H, Roberts TM, et al. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat Rev Drug Discov*. 2009;8(8):627–644.
51. Iorio F, Knijnenburg Theo A, Vis Daniel J, et al. A landscape of pharmacogenomic interactions in cancer. *Cell*. In press.
52. Chen D, Frezza M, Schmitt S, et al. Bortezomib as the first proteasome inhibitor anticancer drug: Current status and future perspectives. *Curr Cancer Drug Targets*. 2011;11(3):239–253.
53. Domcke S, Sinha R, Levine DA, et al. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat Commun*. 2013;4:2126.
54. Neve RM, Chin K, Fridlyand J, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 2006;10(6):515–527.
55. Spies N, Nielsen CB, Padgett RA, et al. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell*. 2009;36(2):245–254.
56. Mavrich TN, Ioshikhes IP, Venters BJ, et al. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res*. 2008;18(7):1073–1083.
57. Hanahan D, Weinberg Robert A. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646–674.
58. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res*. 2010;38(9):2757–2774.