OXFORD

Briefings in Bioinformatics, 2024, **25(5)**, bbae376
https://doi.org/10.1093/bib/bbae376
Problem Solving Protocol

A pan-cancer interrogation of intronic polyadenylation and its association with cancer characteristics

Liang Liu [1,2,*], Peiqing Sun [1], Wei Zhang [1,2,*]

[1]Department of Cancer Biology, Wake Forest University School of Medicine, Medical Center Blvd, Winston-Salem, NC 27157, United States
[2]Center for Cancer Genomics and Precision Oncology, Atrium Health Wake Forest Baptist Comprehensive Cancer Center, Medical Center Blvd, Winston-Salem, NC 27157, United States

*Corresponding authors. Liang Liu, Department of Cancer Biology, Wake Forest University School of Medicine, Medical Center Blvd, Winson-Salem, NC 27157, United States. Tel: 336-713-7514; E-mail: lliu@wakehealth.edu; Wei Zhang, Department of Cancer Biology, Wake Forest University School of Medicine, Medical Center Blvd, Winson-Salem, NC 27157, United States. Tel: 336-713-7508; E-mail: wezhang@wakehealth.edu

Abstract

3′UTR-APAs have been extensively studied, but intronic polyadenylations (IPAs) remain largely unexplored. We characterized the profiles of 22 260 IPAs in 9679 patient samples across 32 cancer types from the Cancer Genome Atlas cohort. By comparing tumor and paired normal tissues, we identified 180 ~ 4645 dysregulated IPAs in 132 ~ 2249 genes in each of 690 patient tumors from 22 cancer types that showed consistent patterns within individual cancer types. We selected 2741 genes that showed consistently patterns across cancer types, including 1834 pan-cancer tumor-enriched and 907 tumor-depleted IPA genes; the former were amply represented in the functional pathways such as deoxyribonucleic acid damage repair. Expression of IPA isoforms was associated with tumor mutation burden and patient characteristics (e.g. sex, race, cancer stages, and subtypes) in cancer-specific and feature-specific manners, and could be a more accurate prognostic marker than gene expression (summary of all isoforms). In summary, our study reveals the roles and the clinical relevance of tumor-associated IPAs.

Keywords: pan-cancer; intronic polyadenylation; cancer characteristics; biomarker

Background

Messenger ribonucleic acid (mRNA) cleavage and polyadenylation (CPA) constitute crucial final steps in mRNA processing, commencing with the cleavage of the 3′ end of the precursor mRNA, followed by the sequential addition of adenosine to form a poly(A) tail [1, 2]. Dynamic alternative cleavage and polyadenylation (APA) occur in numerous physiologic processes, such as cell development, differentiation, proliferation, and reprogramming [3–5], and in human diseases, such as viral infection [6] and cancers [7–11]. APA primarily manifests in the 3′-most exons of genes (3'UTR-APA), resulting in gain or loss of important cis-regulatory elements, such as microRNA (miRNA) binding sites [12]. Intronic polyadenylation (IPA) halts normal transcription, generating truncated mRNA isoforms and proteins or non-coding RNAs (ncRNAs) [1]. IPA is implicated in inactivation of deoxyribonucleic acid (DNA) damage repair (DDR) genes [13, 14] and tumor suppressors [15] in cancer, as well as in enhancing transcript diversity in immune cells [16]. Nevertheless, the clinical relevance of IPAs is poorly understood.

Various experimental approaches have been developed to evaluate the usage of polyadenylation sites (PAs) on a genome-wide scale [17–20]. These techniques, applied across normal and abnormal tissues, enable the study of regulatory mechanisms and functional consequences of APA status, thereby advancing our understanding of this post-transcriptional process under diverse biological conditions [3–8]. However, these approaches are relatively laborious and/or costly in contrast to conventional RNA sequencing (RNA-seq), which has been extensively employed to elucidate the transcriptomics landcape of large-scale sample cohorts, such as the Cancer Genome Atlas (TCGA) [21]. In this study, we employed a bioinformatics approach to quantify IPAs from conventional RNA-seq data of samples from TCGA and illustrated systematic patterns of IPA isoforms, resulting impaired gene transcription, and their clinical relevance.

Results

Broadly and recurrently aberrant intronic polyadenylations across tumor types

We gathered RNA-seq data from 9679 tumor tissues across 32 cancer types from TCGA [21] (Fig. 1a; Table S1a). Given that standard RNA-seq analysis methods are not specifically tailored to differentiate reads for different isoforms of the same gene, we applied APAlyzer v1.9.4 [22] to quantify the usage of ~32 000 IPAs within ~9000 genes by defining genomic regions according to IPA site(s) within each gene (PolyA_DB v3.2 database [23]; Supplemental methods) and counting RNA-seq reads mapped into each region (Fig. 1b). Across all tumor samples, we observed 22 260 IPAs (Fig. S1a). Generally, tumor has higher IPA burdens than normal

Received: April 3, 2024. **Revised:** June 26, 2024. **Accepted:** July 17, 2024

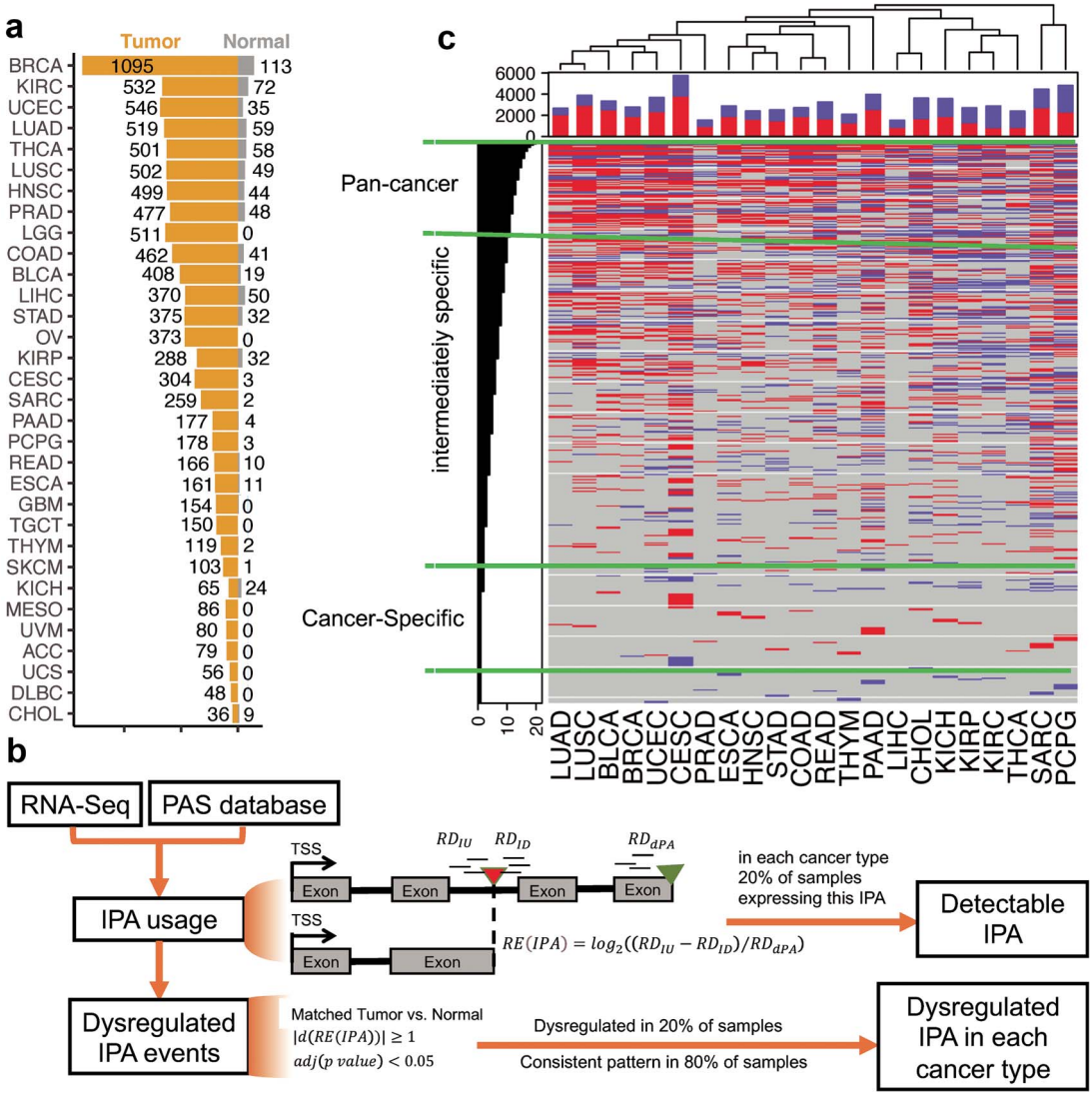

Figure 1. IPA landscape in the TCGA cohort. (a) Data source for the 33 cancer types in this study. Bar charts describe numbers of tumor (yellow) and normal samples (grey) with available RNA-seq for each cancer type (cancer abbreviations are shown in Table S1a). (b) Workflow and criteria for IPA quantification using RNA-seq data. RE: relative expression; dPA: distal PA; IU: IPA upstream; ID; IPA downstream; RD: read density. (c) IPAs (row) with increased (red) or decreased (blue) usage in each tumor type (column). The upper histogram shows the numbers of IPAs with increased (red) or decreased (blue) usages in each tumor type. The side histograms show the numbers of cancer types with dysregulated IPAs. A hierarchical clustering method was used. Pan-cancer IPA sites ($n = 2011$): consistently dysregulated in 11 cancer types, intermediately specific IPAs ($n = 7570$): Consistently dysregulated in 2–10 cancer types, and cancer specific IPAs ($n = 2394$): dysregulated in only 1 cancer type.

samples except for a few cancer types, such as pan-kidney cancers [24] (KIRC, KIRP, and KICH; Fig. S1b; Abbreviations of cancer types are shown in Table S1a.).

Next, we identified IPAs with differential usages between matched tumor and normal samples. For this analysis, we selected cancer types with at least two matched tumor and normal samples. This selection resulted in 22 cancer types, encompassing a total of 690 matched patient tumor and normal samples (Table S1b). We defined a significantly dysregulated IPA based on its relative expression difference ($|\ d(RE(IPA))\ | \geq 1$) and adjusted $p < 0.05$ from Fisher's exact test (Fig. 1b; Supplemental methods). The number of differential IPAs varies between 180 and 4645 in 132 to 2249 genes among the 690 tumor samples (Fig. S1c). Most of the IPAs are consistently dysregulated in tumors compared to matched normal samples within individual cancer types (Fig. S1d), although heterogeneity exists between samples that may reflect the specificity of each individual cancer patients.

We selected IPAs exhibiting consistent patterns in each cancer type (Supplemental methods). Most cancer types show markedly elevated IPA usage, such as LUAD and LUSC. However, five cancer types show opposite trends, including KIRC, THCA, CHOL, KIRP and PCPG (Fig. S1e; Table S1b). Dysregulated IPAs demonstrated consistency among cancer types with similar histologic features (Fig. 1c). For example, lung cancer (LUSC and LUAD), pan-kidney cancers (KIRC, KIRP and KICH), and gynecologic and breast cancers [25] (UCEC, CESC, and BRCA) show similar patterns in our hierarchical clustering analyses.

While no IPAs exhibit consistently higher or lower usage in tumors compared to matched normal tissue across all 22 cancer types, the majority (9581 out of 11 975) demonstrate consistent patterns in at least two cancer types (Fig. 1c). Specifically, there are 2011 pan-cancer dysregulated IPAs with consistently higher or lower usage in at least 11 cancer types, along with 7570 IPAs showing consistently higher or lower usage in 2 to 11 cancer types, and 2394 cancer-specific IPAs that are dysregulated in only

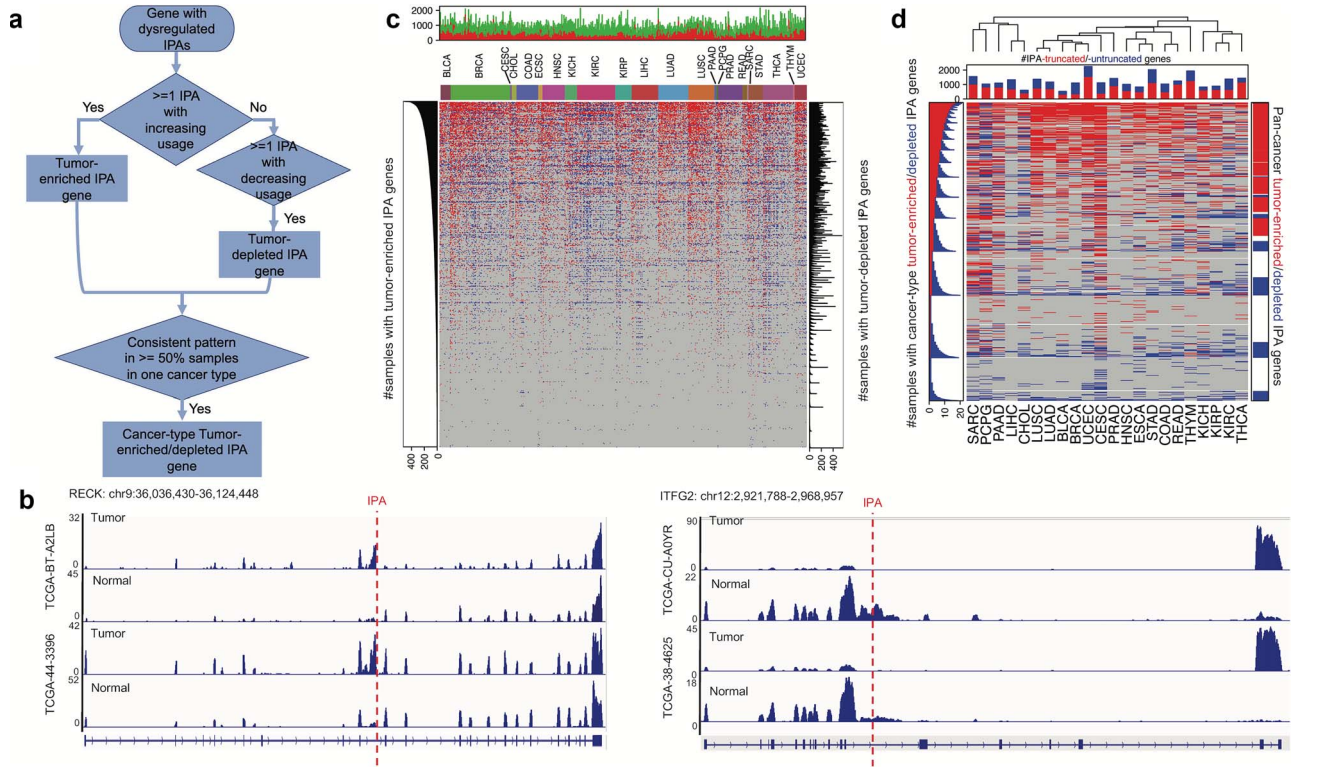

Figure 2. IPA-regulated genes across TCGA tumor types. (a) Strategy to determine if one gene was considered as a tumor-enriched or depleted IPA gene in each cancer type. (b) Example of tumor-enriched (RECK) and depleted (ITFG2) IPA genes in tumors compared to the matched normal samples. (c) The central heat map shows tumor-enriched (red) or depleted (blue) IPA genes in each tumor (columns). The upper histogram shows the number of genes per tumor. The side histograms show the numbers of tumors with tumor-enriched (left) or depleted IPA (right) genes. The central heat map shows cancer-type tumor-enriched or depleted IPA genes in each cancer type. The upper histogram shows the numbers of tumor-enriched (red) and depleted (blue) IPA genes in each cancer type. The left histograms show the numbers of cancer types with cancer-type tumor-enriched (red) and depleted (blue) IPA genes. The right histograms show whether the gene is selected as a pan-cancer tumor-enriched (red) and -depleted (blue) IPA gene. A hierarchical clustering method was used.

1 cancer type. The pan-cancer dysregulated IPAs represent 52.1% of dysregulated IPAs in LIHC (the highest), but only 27.8% in CESC (the lowest); whereas cancer-specific IPAs account for 9.2% in CESC and 1.6% in BLCA (the highest and lowest percentages, respectively) (Fig. S1f).

Pan-cancer intronic polyadenylation-derived early transcriptional termination

We classified a gene as a tumor-enriched or -depleted IPA gene based on the usage changes of IPAs within this gene (Fig. 2a and b; Supplemental methods). Across 690 tumors (compared to paired normal samples), these IPA-regulated genes vary from 132 to 2249 (Figs 2c and S2A). Many genes exhibited consistent patterns of IPA regulation (Fig. S2b).

To examine the consistency of IPA regulation, in each cancer type, we designated a gene as a cancer-type tumor-enriched IPA gene if more tumor samples (≥ 50% and at least 2) consistently exhibit this gene as a tumor-enriched IPA gene, or as a cancer-type tumor-depleted IPA gene if more samples (≥ 50% and at least 2) displayed this gene as a tumor-depleted IPA gene (Fig. 2a; Supplemental methods). For example, RECK is a cancer-type tumor-enriched IPA-truncated gene found in 21 cancer samples, while ITFG2 is a cancer-type tumor-depleted IPA gene found in 19 cancer samples (Fig. 2b). Most cancer types show markedly elevated number of tumor-enriched than depleted IPA genes in the tumors, such as LUSC, BLCA, and LUAD; whereas KIRC, THCA, CHOL, and KIRP have opposite trends (Fig. S2c; Table S2a). Additionally, unsupervised clustering analysis revealed the similarity of tumor-enriched and -depleted genes among cancer types with similar histologic features (Fig. 2d), such as lung cancer (LUSC and LUAD), pan-kidney cancers (KIRC, KIRP and KICH), pan-gyn organ sites (UCEC, CESC and BRCA), suggesting a similar regulatory role of IPA in cancers of common tissue origin.

Our analyses yielded 4491 cancer-type tumor-enriched or depleted IPA genes across 22 cancer types (Fig. 2d). Seven hundred seventy-one genes (17.6%) consistently express the same transcripts in at least 11 cancer types, and 2808 genes exhibit consistent patterns in at least 2 cancer types. To explore the pan-cancer consistency, we identified 1834 pan-cancer tumor-enriched IPA genes occurring in at least half of the cancer types, which was approximately twice the number of pan-cancer tumor-depleted IPA genes ($n = 907$; Fig. 2d; Table S2b).

Next, we utilized the tool coding-potential assessment tool (CPAT) [26] to predict coding probability of the IPA isoforms (Supplemental methods). Of the 1834 pan-cancer tumor-enriched IPA genes (Fig. S3a), 791 IPA isoforms (43.6%) have coding probability scores (CPS) greater than 0.364. This threshold, identified as optimal for discerning coding potential in human data [26], indicates the likelihood of these isoforms to generate protein-coding mRNAs and proteins (Fig. S3b). The remaining 1043 IPAs (56.4%; $CPS < 0.364$) likely generate micropeptides or represent ncRNAs (IPA-ncRNA; Fig. S3c), suggesting that IPA diversifies the transcriptomic landscape and potentially leads to the loss of protein-coding genes and their functional protein products in human cancers. The ability to code proteins is not associated with polyadenylation signaling

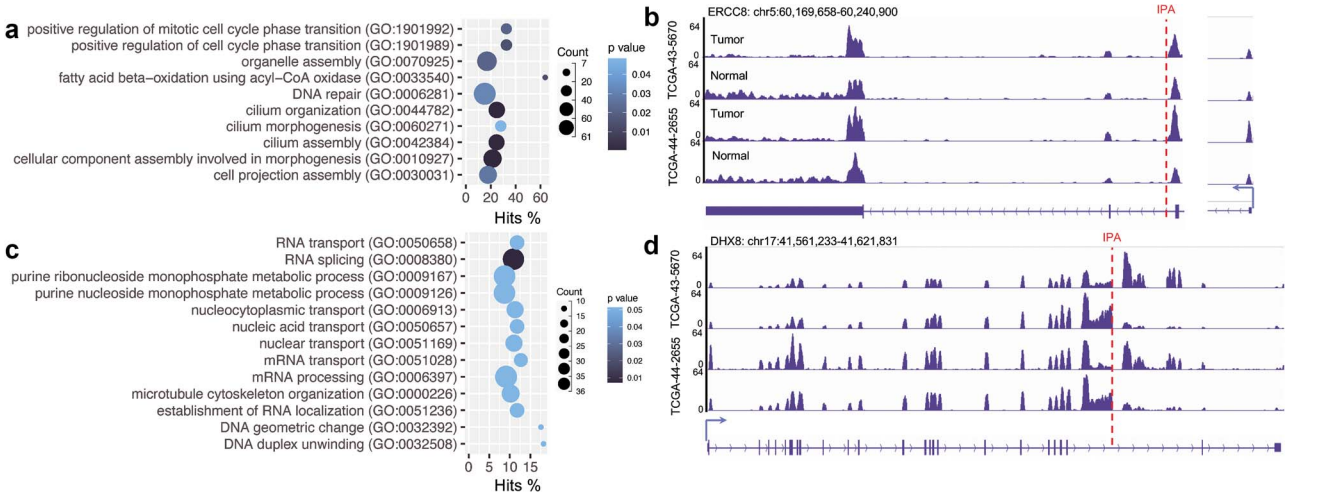

Figure 3. Functional enrichment of pan-cancer tumor-enriched and depleted IPA genes. (a and c) GO enrichment of pan-cancer (a) tumor-enriched and (c) depleted IPA genes. (b and d) Examples of tumor-enriched (B: ERCC8) and depleted (D: DHX8) IPA DNA damage repair genes.

sequences or IPA locations but is correlated with intron length (Fig. S3d).

Furthermore, 453 IPA isoforms are, in fact, 5′ IPA isoforms (IPA located at first intron; Fig. S3d). Of these, 212 may generate protein-coding genes with the length of open reading frames ranging between 126 and 5394 bp (Fig. S3e); others likely generate micropeptides or represent IPA-ncRNAs (Fig. S3f).

Pan-cancer tumor-enriched intronic polyadenylation genes are involved in cilium assembly and deoxyribonucleic acid repair

To better understand the biological consequences of IPA in cancer, we applied the enrichR tool [27] to characterize the pathways associated with the identified genes. Gene ontology (GO) analyses showed that the 1834 pan-cancer tumor-enriched IPA genes were enriched in the functional terms related to cilium assembly and DNA repair (Fig. 3a; Table S3a). Evidence is accumulating that ciliary defects lead to multiple diseases [28] and ciliary deregulation plays crucial roles in cancer formation and progression [29–31]; restoring the cilia can suppress proliferation in cancer cells [32, 33]. The DNA repair process operates through a number of mechanisms such as excision repair and homologous recombination repair to protect the human genome from damage and provide genome stability. DNA repair deficiency enables cancer cells to accumulate genomic alterations and contributes to their aggressive phenotype [34, 35]. Damage in one of the major DNA repair mechanisms, homologous recombination, by IPA has been reported previously through the depletion of CDK12 [13, 14]. ERCC8 (Fig. 3b) is the top gene that is truncated in tumors by aberrant IPAs in 17 cancer types, followed by FANCN. The former participates in excision repair and the latter plays roles in double-strand break repair.

On the other hand, the 904 pan-cancer tumor-depleted IPA genes that produce full-length mRNAs in tumors are enriched in mRNA processing tasks such as splicing and 3′-end processing (Fig. 3c; Table S3b). For example, the IPAs of DHX8, encoding an RNA-binding protein involved in splicing, are inhibited in 13 cancer types (Fig. 3d), and the IPAs of hnRNPL, encoding a protein regulating both mRNA splicing and IPA processing [36, 37], are inhibited in 11 cancer types. These indicate that IPA truncation of these genes is inhibited in tumors, consequently promoting the post-transcriptional regulation in human cancers. Moreover, genes associated with cancer hallmark signatures, such as epithelial–mesenchymal transition and E2F targets, are also overrepresented in these genes (Fig. S4), suggesting that IPA inhibition exerts an oncogenic effect.

Correlations between truncating mutations, deoxyribonucleic acid damage repair, tumor mutation burden, and intronic polyadenylation

We demonstrated that the aberrant IPAs in tumors led to early termination of transcription of numerous genes, resulting in the generation of truncated proteins lacking essential domains or IPA-ncRNAs. This phenocopied the effects of truncating mutations at DNA level. Across the 22 cancer types analyzed, 563 out of 690 samples carry truncating mutations in 5344 genes, whereas truncation by aberrant IPAs occurs in all samples and affects 6055 genes. Only 377 genes in 148 samples from 17 cancer types harbor both truncating mutations and aberrant IPAs (Table S4a), indicating that these two mechanisms independently regulate transcriptome diversity. Furthermore, in individual samples, IPAs affect more genes than truncating mutations (Fig. 4a; Table S4b).

Subsequently, we examined DDR genes in particular, as the loss of DNA repair is common in many cancer types due to somatic mutations [38]. A total of 132 out of 289 DDR genes (derived from GO:0006281 DNA repair) GO exhibit aberrant IPAs in all 690 samples, and 77 genes have truncating mutations in 94 samples; 11 genes have both in at least one sample (Table S4b). RAD51B is the top gene regulated only by aberrant IPAs in 425 samples (out of 690; 61.5%). TP53 is the top gene carrying truncating mutations (32 samples; 4.6%), existing in BRCA ($n = 2$), COAD ($n = 2$), ESCA ($n = 1$), HNSC ($n = 10$), KICH ($n = 1$), LUAD ($n = 6$), LUSC ($n = 3$), READ ($n = 2$), and STAD ($n = 2$), or both (4 samples from KICH, ESCA, HNSC, and BLCA; 0.58%; Fig. 4b; Table S4c). In contrast, MEN1 is the most frequently truncating mutation-only gene, which occurs in only three samples (from HNSC, LUAD, and STAD).

Given that genes targeted by truncating mutations often function as tumor suppressors [39], we investigated whether such genes are overrepresented among those regulated by IPAs. Among 301 tumor suppressor genes (TSG) [39], 62 do not contain

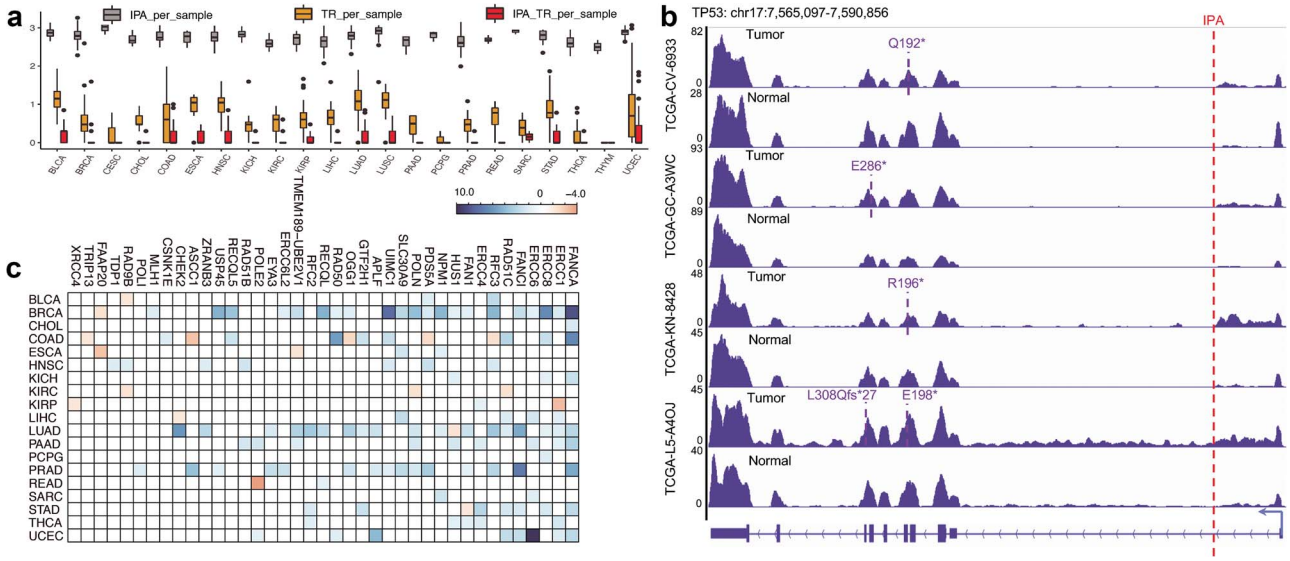

Figure 4. IPA and truncating mutations independently diverse cancer transcriptomics. (a) Numbers of genes truncated by IPA, truncating mutations, or both in each cancer type. (b) Samples having both TP53 truncating mutations and IPA isoforms in tumors compared to matched normal samples. Pan-cancer tumor-enriched IPA DNA repair genes associated with tumor mutation burden. Colors indicate the log10 (*P*-value) with positive (blue) or negative (orange) correlations.

IPAs or truncating mutations. Of the remaining 239 genes, 151 express IPA-truncated isoforms in at least one of the 690 samples, while 191 genes contain truncating mutations. FUBP1 is the top second gene regulated only by IPA in 419 samples (out of 690; 60.7%), following TP53, functions as both DDR and TSG. Conversely, APC is the top gene regulated by truncating mutations (21 samples; 3.0%). Only 27 genes have co-occurring IPAs and truncating mutations, including KMT2C (5 samples), TP53 (4 samples), CASP8 (3 samples), PTEN (3 samples), NSD1 (2 samples), and SETD2 (2 samples; Fig. S5a; Table S4d). KMT2C (MLL3) is a chromatin remodeling gene that mediate H3K4 monomethylation (H3K4me1). KMT2C mutations contribute to tumorigenesis and are associated with poor survival rates [40, 41]. In particular, the complete loss of KMT2C due to truncating mutations is correlated with significantly shorter progression-free survival in patients with breast cancer patients who undergo estrogen therapy [42] and in patients with metastatic prostate cancer [43]. The IPA-associated loss of KMT2C warrants further validation.

Furthermore, we investigated whether IPA isoform expression is associated with tumor mutation burden, as impaired DDR often leads to abundant mutations [44]. In this analysis encompassing all 8039 samples from the 22 cancer types (Supplemental methods), most pan-cancer tumor-enriched IPA genes ($n = 1641$) showed significant associations with tumor mutation burden (TMB) in at least 1 cancer type (Table S4e). BRCA exhibited the highest number of IPAs associated with TMB, while PCPG and CHOL had the fewest. Notably, most cancer types showed a higher number of positively correlated IPA isoforms, such as LUAD and PAAD (Fig. S5b). Specifically, 39 (out of 61) pan-cancer tumor-enriched IPA isoforms of DDR genes are associated with TMB, most of which show positive correlations across cancer types (Fig. 4c). For instance, FNACA, a member of the Fanconi anemia complementation group involved in DNA repair, exhibited a positive correlation between its IPA isoform expression and TMB in nine cancer type (Fig. S5c). Thus, impairment of DDR genes by IPA events may significantly contribute to mutation accumulation in human tumors.

Upstream regulatory factors for intronic polyadenylations and effects of intronic polyadenylation–non-coding ribonucleic acids on signaling pathways

Many factors may regulate the IPA process. CPA factors such as CPSF, CstF, CFIm, and CFIIm can directly bind or associate with mRNA core regulatory cis-elements (e.g. AAUAAA and other auxiliary sequences like U/GU-rich downstream elements [1]) to regulate this process. Additionally, alternative splicing and splicing factors [45–49] also participate in the global APA process. To explore potential regulatory factors for IPA in human cancer, we examined 42 CPA and 262 splicing factors across all 8039 samples from the 22 cancer types. We defined positive and negative regulators based on whether a regulator was positively or negatively correlated with more IPAs in a given cancer type (Supplemental methods). For example, in HNSC, we identified 190 putative master regulators, including 108 positive and 82 negative regulators.

Overall we identified 282 putative master regulators across cancer types (Table S5), among which 95 and 59 factors exhibited strong positive and negative correlation with multiple IPAs in 11 ~ 22 cancer types, respectively, while 30 and 11 factors showed correlations in only 1 cancer type. CPSF3 is the top CPA factor that broadly promotes IPA activities in almost all cancer lineages, ranging from 24% in PCPG to 94% in LUSC (Figs S6a and b). The splicing factors POLR2D, EIF4A3, and HSPA8 are the top master regulators positively correlated with more IPAs across all cancers (Fig. S6c).

Among the 282 master regulators, 18 are pan-cancer tumor-enriched and 29 are pan-cancer tumor-depleted IPA genes, respectively (Fig. 5). PCF11 is an tumor-enriched IPA gene whose transcription is terminated by IPA in seven cancer types (Fig. S6d), where its gene expression is inversively correlated with IPA isoform expression (Fig. S6e). hnRNPL is an tumor-depleted IPA gene whose full-length transcript is consistently expressed in 11 cancer types, and its gene expression is positively correlated with IPA isoform expression. These data are consistent with the previously reported roles of PCF11 [50] and hnRNPL [36] in the IPA process.

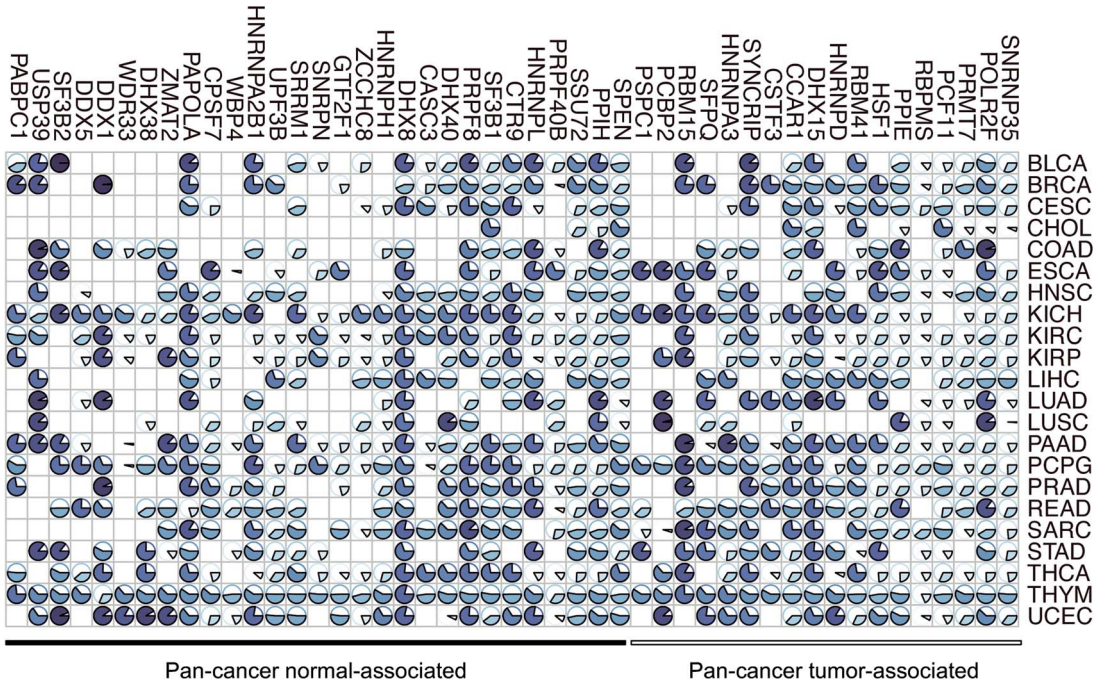

Figure 5. Putative regulators of IPA activities in human cancers. Pie charts show percentage of IPA isoforms that are positively (blue) or negatively (white) correlated with regulators (that were defined as pan-cancer tumor-enriched or depleted) in each cancer type.

ncRNAs play critical roles in cellular processes related to tumorigenesis by interacting with mRNAs [51–55], and may inform optimal drug design [56, 57]. We examined interactions between 1043 IPA-ncRNA isoforms determined previously (CPS < 0.364) and their associated genes. Specifically, we focused on the associations of these IPA-ncRNAs and 335 genes involved in 10 cancer signaling pathways (p53, PI3K, Myc, RTK/RAS, cell cycle, Wnt, TGF beta, Nrf2, Notch, and Hippo) [58] and built an IPA ncRNA-gene regulatory network across cancer types based on co-expression between individual IPA and their putative target genes (Supplemental methods). We found that 735 IPA-ncRNAs were computationally correlated with 327 target genes in the 22 cancer types. The most strongly associated pathway is RTK/RAS, especially in THYM and PAAD (Fig. S7a), and the top associated gene is RAC1 in the RTK/RAS pathway in PAAD, which is associated with expression of 98 IPA-ncRNAs (Fig. S7b; Table S6), including the IPA-ncRNA from TMEM8B (Fig. S7c).

Clinical relevance of intronic polyadenylation isoforms

We hypothesized that individual variations in IPA isoform expression might provide a meaningful predictor for patient clinical outcomes. Within the TCGA cohort, we categorized patients into two groups (≥ or < 90%; Supplemental methods) based on levels of IPA expression or gene expression (sum of all isoforms e.g. those from alternative polyadenylation and splicing) of each gene. Among the 1834 pan-cancer tumor-enriched IPA genes, many show significant associations with overall survival (OS) for either IPA isoform or total gene expression (the sum of all possible isoforms), but not both (Fig. 6a; Table S7a and b). The largest gene set ($n = 225$) for which gene expression and IPA isoform expression predicted patient survival was found in KIRC samples, whereas only one gene was found in the THYM. Intrudingly, we found 225 genes whose expression and IPA isoforms predicted opposite survival outcomes in at least 1 cancer type (Table S7b). For example, KDM4C has been reported as an oncogene and a negative biomarker for patients with multiple cancer types [59–61], and is associated with cancer metastasis [62] and immunosuppressive tumor microenvironment [63]. However, its gene expression is associated with better survival outcomes in patients with LUAD—unlike expression of its IPA isoform, which is associated with poor survival of those patients (Figs 6b and S8a). This negative correlation is also observed in patients with LUSC, KICH, and BRCA, where the gene expression does not show significant association (Table S7b). Another example is DNAJB4 (also known as HLJ1), which plays a complex dual role in tumorigenesis [64–67]. We observed negative associations between its gene expression level and patient survival in multiple cancer types (Fig. 6b; Table S7b); however, its IPA isoform expression is a biomarker of better survival (Figs 6b and S8b). These observations suggest consideration of gene isoforms in cancer biomarker studies.

We proceeded to examine whether IPAs correlated with patient immune cell profiles, a critical determinant of treatment outcomes. Utilizing CIBERSORT [68] to estimate immune cell fractions in patient tumors from the TCGA cohort (Supplemental methods), we identified 61 pan-cancer tumor-enriched IPA isoforms within DDR genes, of which 39 consistently showed associations with tumor immune cell profiles across various cancer types (Table S7c). Notably, the expression of the HUS1 IPA isoform exhibited a positive correlation with CD8 T cell populations in seven cancer types (Fig. 6c) and with natural killer cell populations in six cancer types (Fig. S8c). This suggests that truncation of HUS1 transcription, resulting in decreased levels of functional HUS1 proteins, fosters an anti-tumor microenvironment. This observation aligns with prior studies demonstrating that miRNA-mediated depletion of HUS1 inhibits de novo lung tumorigenesis [69]. Consistent with a pan-cancer analysis indicating that variant or low expression levels of FANCI are linked to poor patient survival [70], we

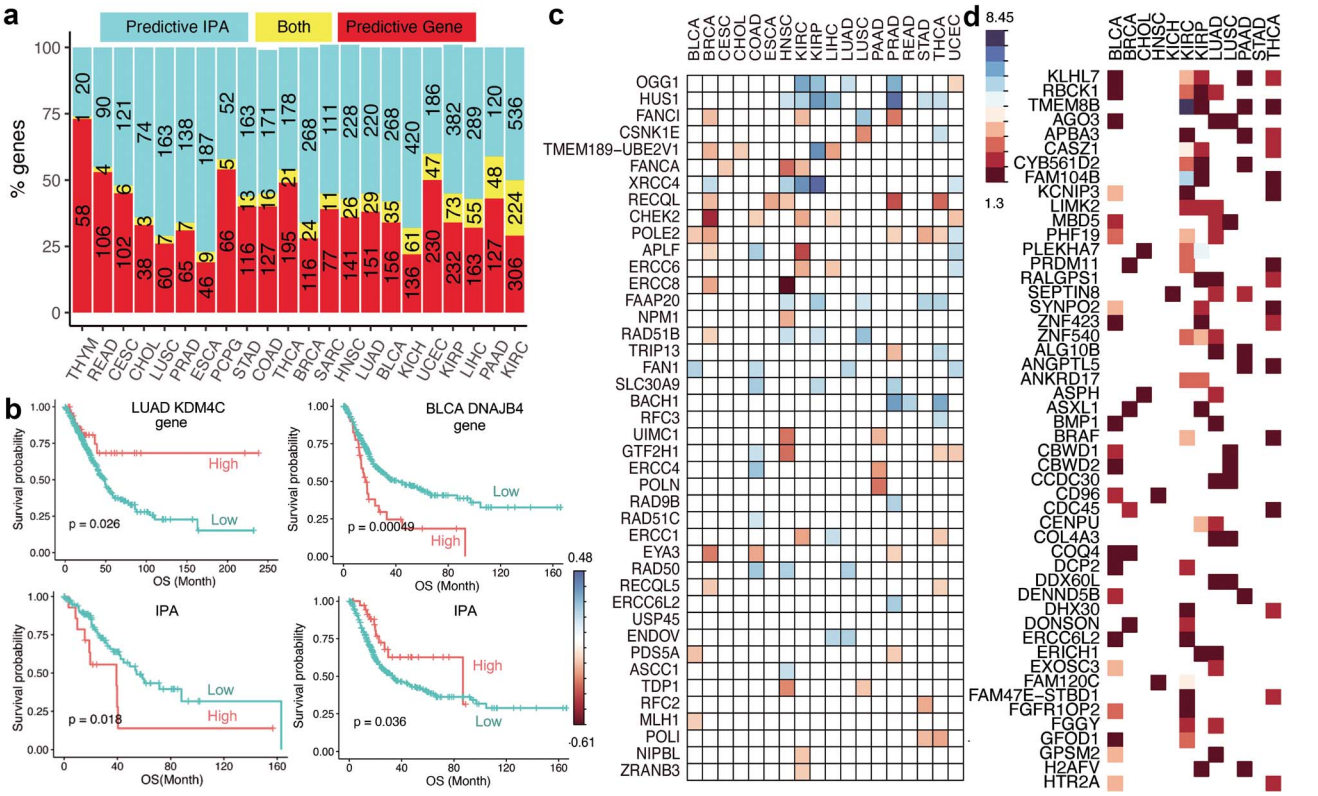

Figure 6. Associations between IPA isoforms and clinical characteristics. (a) Numbers of genes whose expression or IPA usage predicted patient survival. (b) IPA isoform and gene expression of KDM4C and DNAJB4 had opposite correlations with OS in the TCGA cohort. (c) Pan-cancer tumor-enriched IPA DNA repair genes and associations with CD8 cell fractions. Colors indicated the log10 (*P*-value) with positive (blue) or negative (orange) correlations. (d) the top 50 genes showing significant associations with cancer stages in the TCGA cohort. Colors indicated the log10 (adjusted *P*-value).

found that expression of the FANCI IPA isoform correlated negatively with CD8 T cells (Fig. 6c), activated CD4 T cells, and natural killer cells (Fig. S8c). XRCC4, a component of canonical-nonhomologous end-joining DNA repair, is associated with severe combined immunodeficiency and accelerated tumor development when deficient [71]. Our analysis also demonstrated that high expression levels of the XRCC4 IPA isoform were consistently associated with regulatory T cell populations across nine cancer types (Table S7c).

We then examined cancer-related demographic and clinical features, including stage (available for 16 cancer types), race (available for 22 cancer types), and sex (available for 15 cancer types; excluding cancers specific to men or women), as well as subtypes (available for 9 cancer types; Table S7d), and identified associations between IPAs and these factors. In total, we discovered 1210 clinically relevant IPA isoforms, encompassing 66.0% (1210/1834) of pan-cancer tumor-enriched IPA genes (Table S7e). These included 393 IPA isoforms associated with cancer stage, 516 with patient race, 226 with sex differences, and 1066 correlated with cancer subtypes. Most IPA isoforms exhibited cancer-specific patterns concerning individual clinical features. Of these associations, 306 (77.9% of 393) IPA-cancer stage associations were unique to a single cancer type, with no gene showing associations in more than 8 (50% of 16) cancer types. Notably, the top genes, KLHL7 and RBCK1, were associated with tumor stages in five cancer types (Fig. 6d). Similarly, 386 (74.8% of 516) IPAs were linked to patient races in only 1 cancer type, with no gene showing associations in more than 8 (50% of 15) cancer types. Noteworthy race-associated genes included BOD1L1 and PREP, identified in five cancer types (Fig. S8d). Regarding patient sex, 208 (92.0% of 226) IPA isoforms exhibited associations unique to a single cancer type, with only 1 gene, KDM5C, showing associations in 8 cancer types (Fig. S8e). Furthermore, 432 (40.5% of 1066) IPA isoforms were associated with patient subtypes in only 1 cancer type, while 84 genes showed associations in more than 5 (50% of 9) cancer types. Notably, ZFAND4 demonstrated significant associations in 8 cancer types (Fig. S8f).

Moreover, the majority of IPAs exhibited feature-specific patterns within individual cancer types (Fig. S8g). Among the five cancer types (COAD, ESCA, HNSC, READ, and STAD) with all four features available, no gene showed associations with all four features. In contrast, the substantial number of genes associated with just one feature. Among the 11 cancer types (BLCA, BRCA, CHOL, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, and THCA) with 3 features available, only a small number of genes were associated with all three features: 13 genes (of 418) in BRCA, 2 (of 197) in KIRC, and 1 (of 105) in LUAD. Similar patterns were observed in the 3 cancer types with only two features available, with 5(of 347) genes in CESC, 4 (of 400) in SARC, and 66 (of 395) in UCEC showing associations with two features.

Discussion

Interrogating large-scale RNA-seq data sets from the TCGA cohort, we provided a systemic view of the IPA landscape in human cancer. The TCGA cohort, comprising approximately 10 000 patient samples spanning 32 cancer types, was meticulously curated with stringent inclusion criteria to minimize potential biases, establishing it as a substantial resource for cancer research.

In this study, only paired tumor and normal samples were included to detect the dysregulated IPAs for each individual patient. We did not include additional normal samples from other data resources, such as GTEx, due to the natural differences between adjacent normal samples from the TCGA cohort and normal samples from the GTEx cohort. Moreover, comparing paired samples from individual patients provides insights into inter-tumor heterogeneity and consistency. However, we acknowledge that the limited sample size in some cancer types may constrain the observation of inter-cancer heterogeneity. This limitation will be mitigated by the growth of sample size in the future.

Our study revealed a robust pan-cancer pattern of aberrant IPAs and IPA-regulated genes that remained consistent across cancer patients and types, although certain IPAs exhibited inter- and intra-tumor variation. These pan-cancer tumor-enriched IPA genes are enriched in functional pathways such as DNA repair, implying that dysregulated IPAs are involved in accumulation of somatic mutations in human tumors. This hypothesis was supported by our observations indicating that many IPAs are positively correlated with patient TMB. Furthermore, we found that the expression of IPA isoforms correlates with various clinical characteristics of patients, including survival, immune profiles, stage, subtype, sex, and race, highlighting the importance of IPA isoforms independent of gene expression (the sum of all isoforms). We are confident that the comprehensive scope of the TCGA dataset, combined with our robust analytical methodology and consistent observations across diverse patient cohorts and cancer types, strengthens the validity and reliability of our findings.

Our analysis identified 22 260 detectable IPA sites, most of which showed a consistent expression pattern across 33 cancer types. Pan-cancer IPAs have higher usage levels than others (Wilcoxon test $p < 2.2 \times 10^{-16}$; (Fig. 1c), reminiscent of characteristics of protein-coding genes and enhancer RNAs (eRNAs)—the housekeeping genes and pan-cancer eRNAs are generally expressed at high levels compared with tissue-specific genes [72] and eRNAs [73], respectively.

Both truncating mutations and aberrant IPAs in genes lead to early termination of protein-coding gene transcription, resulting in the generation of ncRNAs or truncated mRNAs that may exhibit altered functionality compared to full-length isoforms. Our observations indicate that IPA regulation is exclusive to truncating mutations, particularly prominent in DDR and TSG. Thus, IPA, alongside somatic mutations, contributes to the diversification of the cancer transcriptomic landscape and compromises tumor suppressive mechanisms in human cancers.

Tumor cells often exhibit transcripts with systematically shorter 3' UTRs compared to normal cells [11] achieved by preferentially selecting proximal polyadenylation sites in the 3' UTR. This mechanism often results in the upregulation of numerous oncogenes, enabling them to evade regulatory miRNA-mediated repression [7, 10, 74]. In our study, we observed a plethora of tumor-enriched and depleted IPA genes in each tumor compared to paired normal tissues. Additionally, tumors tended to express IPA-truncated isoforms, with a higher prevalence of global IPA-related early transcriptional termination observed in more samples ($n = 460$ and 230, respectively). In contrast to 3' UTR-APA, which primarily impacts oncogenes, our findings indicate that pan-cancer tumor-enriched IPA genes are predominantly associated with DNA repair pathways; whereas, tumor-depleted IPA genes are linked to cancer hallmark signatures such as epithelial–mesenchymal transition (EMT) and E2F targets. Thus, APA processing, encompassing both 3' UTR-APA and IPA, likely plays a role in tumor development and progression, suggesting its potential as a therapeutic target.

We identified an notable number of IPA isoforms with potential clinical relevance that could serve as biomarkers. Traditionally, biomarker identification has relied on gene expression, which aggregates all possible isoforms, including those from alternative splicing and polyadenylations. However, this approach may lack accuracy due to the potentially distinct functions of individual isoforms. Through our analysis, we identified a subset of genes whose IPA isoform expression showed an inverse association with total gene expression in relation to patient survival, exemplified by KDM4C and DNAJB4. This suggests that isoform expression may offer a more precise prognostic marker than gene expression alone. This observation could potentially be elucidated by recent research highlighting the significance of long ncRNAs (lncRNAs), which are increasingly recognized for their diverse roles in gene regulation mechanisms. LncRNAs are increasingly implicated in various aspects of cancer biology, including initiation, progression, patient drug resistance, treatment outcomes, and as promising therapeutic targets [75]. Our analysis indicates that IPA truncation might lead to the generation of numerous IPA-ncRNAs, potentially involved in modulating transcriptional programs, cellular functions, and cancer progression.

DDR and the immune system exhibit a tight interconnection. Mutations in DDR genes and elevated TMB are correlated with cancer immunotherapeutic response and cancer prognosis [76–78]. IPA truncation of DDR also enables DNA damage deficiency; consequently; we observed a strong association between DDR IPA isoform expression and immune cell profiles in patient tumors. The IPA-induced truncation of transcription may produce neoantigens, akin to mRNA splicing or truncating mutations. Neoantigens can stimulate an immune system response against cancer cells [79, 80] and is a promising approach to tumor immunotherapy [81, 82].

Our study does have limitations. Firstly, while our study revealed the significance of IPA and highlighted that genes undergoing IPA regulation are enriched in pathways such as cilium assembly and DNA repair, both of which are strongly associated with cancer formation and progression, we concur that this finding does not offer a comprehensive portrayal of IPA's roles in cellular functions, cancer initiation, progression, and treatment, areas in which our understanding is still lacking depth. Secondly, while our study primarily focused on establishing associations between IPA and patient clinical characteristics such as cancer state and patient survivals, we recognize the significance of investigating causality to better understand the underlying mechanisms. Our findings will likely provide basis and directions for the causality and mechanistic studies in the future. Thirdly, although our study investigates the correlation between IPA events and splicing/CPA factors, we acknowledge that it does not explore the molecular mechanisms governing IPA regulation, including genome alterations, particularly those affecting core regulatory cis-elements (e.g. AAUAAA and other auxiliary sequences like U/GU-rich downstream elements [1]). This aspect merits further investigation, especially through experimental validation and in-depth analysis.

In summary, by using large-scale RNA-seq and clinical datasets from the TCGA cohort, we provide a comprehensive landscape of IPA events in human tumors and their clinical relevance. Our analysis can be a useful resource for further investigation regarding regulating mechanisms and functions of IPA.

Key Points

- We have identified numerous dysregulated IPA isoforms in tumor samples compared to normal samples in the TCGA cohort, with many showing consistent patterns across various cancer types.
- IPA regulation can impact multiple biological pathways, including DNA repair and cancer signaling pathways, which play pivotal roles in cancer development and treatment.
- The expression of IPA isoforms correlates with patient clinical characteristics in cancer-specific and feature-specific manners, indicating the potential of IPA isoforms as more precise prognostic markers.

Acknowledgements

The authors thank the Bioinformatics Shared Resource (BISR) at Atrium Atrium Health Wake Forest Baptist Comprehensive Cancer Center for providing the computational infrastructure. We would like to acknowledge the editorial assistance of Karen Klein, MA, ELS, MWC, from the Clarus Editorial Services.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by the National Institutes of Health (R01 CA272627, R21 CA253362, and R03 CA256100 to W.Z. and L.L.) and by the Comprehensive Cancer Center of Wake Forest Baptist Medical Center (P30 CA012197). L.L. is supported by Wake Forest start-up funds. P.S. is supported by the Anderson Oncology Research Professorship. W.Z. is supported by the Hines and Willis Family Professorship and a Fellowship from the National Foundation for Cancer Research.

Data availability

All data referenced in this manuscript is available in the supplemental tables. The IPA RE values for all samples can be accessed via the following link to our shared Google Drive.

Author contributions

L.L. and W.Z conceived and designed experiments. L.L. performed the calculations and analyzed the data. L.L. and W.Z co-wrote the manuscript with the support of P.S. All authors contributed to discussions.

References

1. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* 2017;**18**:18–30. https://doi.org/10.1038/nrm.2016.116.
2. Ren F, Zhang N, Zhang L. *et al.* Alternative polyadenylation: a new frontier in post transcriptional regulation. *Biomark Res* 2020;**8**:67. https://doi.org/10.1186/s40364-020-00249-6.
3. 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**:7028–33. https://doi.org/10.1073/pnas.0900028106.
4. Ji Z, Tian B. Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PloS One* 2009;**4**:e8419. https://doi.org/10.1371/journal.pone.0008419.
5. Hilgers V, Perry MW, Hendrix D. *et al.* Neural-specific elongation of 3′ UTRs during drosophila development. *Proc Natl Acad Sci U S A* 2011;**108**:15864–9. https://doi.org/10.1073/pnas.1112672108.
6. Majerciak V, Ni T, Yang W. *et al.* A viral genome landscape of RNA polyadenylation from KSHV latent to lytic infection. *PLoS Pathog* 2013;**9**:e1003749. https://doi.org/10.1371/journal.ppat.1003749.
7. Erson-Bensan AE, Can T. Alternative polyadenylation: another foe in cancer. *Mol Cancer Res* 2016;**14**:507–17. https://doi.org/10.1158/1541-7786.MCR-15-0489.
8. Yuan F, Hankey W, Wagner EJ. *et al.* Alternative polyadenylation of mRNA and its role in cancer. *Genes Dis* 2021;**8**:61–72. https://doi.org/10.1016/j.gendis.2019.10.011.
9. Masamha CP, Xia Z, Yang J. *et al.* CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 2014;**510**:412–6. https://doi.org/10.1038/nature13261.
10. 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. https://doi.org/10.1038/ncomms6274.
11. Mayr C, Bartel DP. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009;**138**:673–84. https://doi.org/10.1016/j.cell.2009.06.016.
12. insights from RNA sequencing. de Klerk, E. & t Hoen, P. A. Alternative mRNA transcription, processing, and translation. *Trends Genet* 2015;**31**:128–39. https://doi.org/10.1016/j.tig.2015.01.001.
13. Dubbury SJ, Boutz PL, Sharp PA. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* 2018;**564**:141–5. https://doi.org/10.1038/s41586-018-0758-y.
14. Krajewska M, Dries R, Grassetti AV. *et al.* CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat Commun* 2019;**10**:1757. https://doi.org/10.1038/s41467-019-09703-y.
15. Lee SH, Singh I, Tisdale S. *et al.* Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 2018;**561**:127–31. https://doi.org/10.1038/s41586-018-0465-8.
16. Singh I, Lee SH, Sperling AS. *et al.* Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat Commun* 2018;**9**:1716. https://doi.org/10.1038/s41467-018-04112-z.
17. Moll P, Ante M, Seitz A. *et al.* QuantSeq 3′ mRNA sequencing for RNA quantification. *Nat Methods* 2014;**11**:i–iii. https://doi.org/10.1038/nmeth.f.376.
18. 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**:133–9. https://doi.org/10.1038/nmeth.2288.
19. Ma F, Fuqua BK, Hasin Y. *et al.* A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics* 2019;**20**:9. https://doi.org/10.1186/s12864-018-5393-3.
20. Derti A, Garrett-Engele P, Macisaac KD. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res* 2012;**22**:1173–83. https://doi.org/10.1101/gr.132563.111.
21. Weinstein J, Collisson EA, Mills GB. *et al.* The Cancer Genome Atlas pan-cancer analysis project. *Nat Genet* 2013;**45**:1113–20. https://doi.org/10.1038/ng.2764.
22. Wang R, Tian B. APAlyzer: a bioinformatics package for analysis of alternative polyadenylation isoforms. *Bioinformatics* 2020;**36**:3907–9. https://doi.org/10.1093/bioinformatics/btaa266.

23. Wang R, Nambiar R, Zheng D. *et al.* PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res* 2018;**46**:D315–9. https://doi.org/10.1093/nar/gkx1000.
24. Ricketts CJ, de Cubas AA, Fan H. *et al.* The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Rep* 2018;**23**:3698. https://doi.org/10.1016/j.celrep.2018.06.032.
25. Berger AC, Korkut A, Kanchi RS. *et al.* A comprehensive Pan-cancer molecular study of gynecologic and breast cancers. *Cancer Cell* 2018;**33**:690–705.e9. https://doi.org/10.1016/j.ccell.2018.03.014.
26. Wang L, Park HJ, Dasari S. *et al.* CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res* 2013;**41**:e74. https://doi.org/10.1093/nar/gkt006.
27. Kuleshov MV, Jones MR, Rouillard AD. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;**44**:W90–7. https://doi.org/10.1093/nar/gkw377.
28. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol* 2017;**18**:533–47. https://doi.org/10.1038/nrm.2017.60.
29. Eguether T, Hahne M. Mixed signals from the cell's antennae: primary cilia in cancer. *EMBO Rep* 2018;**19**:e46589. https://doi.org/10.15252/embr.201846589.
30. Higgins M, Obaidi I, McMorrow T. Primary cilia and their role in cancer. *Oncol Lett* 2019;**17**:3041–7. https://doi.org/10.3892/ol.2019.9942.
31. Wang B, Liang Z, Liu P. Functional aspects of primary cilium in signaling, assembly and microenvironment in cancer. *J Cell Physiol* 2021;**236**:3207–19. https://doi.org/10.1002/jcp.30117.
32. Gradilone SA, Radtke BN, Bogert PS. *et al.* HDAC6 inhibition restores ciliary expression and decreases tumor growth. *Cancer Res* 2013;**73**:2259–70. https://doi.org/10.1158/0008-5472.CAN-12-2938.
33. Khan NA, Willemarck N, Talebi A. *et al.* Identification of drugs that restore primary cilium expression in cancer cells. *Oncotarget* 2016;**7**:9975–92. https://doi.org/10.18632/oncotarget.7198.
34. Torgovnick A, Schumacher B. DNA repair mechanisms in cancer development and therapy. *Front Genet* 2015;**6**:157. https://doi.org/10.3389/fgene.2015.00157.
35. Burrell RA, McGranahan N, Bartek J. *et al.* The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 2013;**501**:338–45. https://doi.org/10.1038/nature12625.
36. Hung LH, Heiner M, Hui J. *et al.* Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* 2008;**14**:284–96. https://doi.org/10.1261/rna.725208.
37. Movassat M, Crabb TL, Busch A. *et al.* Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biol* 2016;**13**:646–55. https://doi.org/10.1080/15476286.2016.1191727.
38. Knijnenburg TA, Wang L, Zimmermann MT. *et al.* Genomic and molecular landscape of DNA damage repair deficiency across the cancer genome atlas. *Cell Rep* 2018;**23**:239–254.e6. https://doi.org/10.1016/j.celrep.2018.03.076.
39. Davoli T, Xu AW, Mengwasser KE. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 2013;**155**:948–62. https://doi.org/10.1016/j.cell.2013.10.011.
40. Chang A, Liu L, Ashby JM. *et al.* Recruitment of KMT2C/MLL3 to DNA damage sites mediates DNA damage responses and regulates PARP inhibitor sensitivity in cancer. *Cancer Res* 2021;**81**:3358–73. https://doi.org/10.1158/0008-5472.CAN-21-0688.
41. Rampias T, Karagiannis D, Avgeris M. *et al. EMBO Rep* 2019;**20**:e46821.
42. Gala K, Li Q, Sinha A. *et al.* KMT2C mediates the estrogen dependence of breast cancer through regulation of ERα enhancer function. *Oncogene* 2018;**37**:4692–710. https://doi.org/10.1038/s41388-018-0273-5.
43. Limberger T, Schlederer M, Trachtová K. *et al.* KMT2C methyltransferase domain regulated INK4A expression suppresses prostate cancer metastasis. *Mol Cancer* 2022;**21**:89. https://doi.org/10.1186/s12943-022-01542-8.
44. Park S, Lee H, Lee B. *et al.* DNA damage response and repair pathway alteration and its association with tumor mutation burden and platinum-based chemotherapy in SCLC. *J Thorac Oncol* 2019;**14**:1640–50. https://doi.org/10.1016/j.jtho.2019.05.014.
45. Misra A, Green MR. From polyadenylation to splicing: dual role for mRNA 3′ end formation factors. *RNA Biol* 2016;**13**:259–64. https://doi.org/10.1080/15476286.2015.1112490.
46. Mathieu O, Bouche N. Interplay between chromatin and RNA processing. *Curr Opin Plant Biol* 2014;**18**:60–5. https://doi.org/10.1016/j.pbi.2014.02.006.
47. Berg MG, Singh LN, Younis I. *et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 2012;**150**:53–64. https://doi.org/10.1016/j.cell.2012.05.029.
48. Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(a) polymerase. *Mol Cell* 1998;**1**:255–64. https://doi.org/10.1016/s1097-2765(00)80026-x.
49. Kaida D, Berg MG, Younis I. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 2010;**468**:664–8. https://doi.org/10.1038/nature09479.
50. Wang R, Zheng D, Wei L. *et al.* Regulation of intronic polyadenylation by PCF11 impacts mRNA expression of long genes. *Cell Rep* 2019;**26**:2766–2778.e6. https://doi.org/10.1016/j.celrep.2019.02.049.
51. Grillone K, Riillo C, Scionti F. *et al.* Non-coding RNAs in cancer: platforms and strategies for investigating the genomic "dark matter". *J Exp Clin Cancer Res* 2020;**39**:117. https://doi.org/10.1186/s13046-020-01622-x.
52. Luo J, Liu L, Venkateswaran S. *et al.* RPI-bind: a structure-based method for accurate identification of RNA-protein binding sites. *Sci Rep* 2017;**7**:614. https://doi.org/10.1038/s41598-017-00795-4.
53. Suresh V, Liu L, Adjeroh D. *et al.* RPI-pred: predicting ncRNA–protein interaction using sequence and structural information. *Nucleic Acids Res* 2015;**43**:1370–9. https://doi.org/10.1093/nar/gkv020.
54. Yang M, Lu H, Liu J. *et al.* lncRNAfunc: a knowledgebase of lncRNA function in human cancer. *Nucleic Acids Res* 2022;**50**:D1295–306. https://doi.org/10.1093/nar/gkab1035.
55. Tang Q, Hann SS. HOTAIR: an oncogenic long non-coding RNA in human cancer. *Cell Physiol Biochem* 2018;**47**:893–913. https://doi.org/10.1159/000490131.
56. Matsui M, Corey DR. Non-coding RNAs as drug targets. *Nat Rev Drug Discov* 2017;**16**:167–79. https://doi.org/10.1038/nrd.2016.117.
57. Jariwala N, Sarkar D. Emerging role of lncRNA in cancer: a potential avenue in molecular medicine. *Ann Transl Med* 2016;**4**:286. https://doi.org/10.21037/atm.2016.06.27.
58. Sanchez-Vega F, Mina M, Armenia J. *et al.* Oncogenic signaling pathways in the cancer genome atlas. *Cell* 2018;**173**:321–337.e10. https://doi.org/10.1016/j.cell.2018.03.035.

59. Zack TI, Schumacher SE, Carter SL. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;**45**:1134–40. https://doi.org/10.1038/ng.2760.
60. Suikki HE, Kujala PM, Tammela TL. *et al.* Genetic alterations and changes in expression of histone demethylases in prostate cancer. *Prostate* 2010;**70**:889–98. https://doi.org/10.1002/pros.21123.
61. Kim TD, Fuchs JR, Schwartz E. *et al.* Pro-growth role of the JMJD2C histone demethylase in HCT-116 colon cancer cells and identification of curcuminoids as JMJD2 inhibitors. *Am J Transl Res* 2014;**6**:236–47.
62. Lin CY, Wang BJ, Fu YK. *et al.* Inhibition of KDM4C/c-Myc/LDHA signalling axis suppresses prostate cancer metastasis via interference of glycolytic metabolism. *Clin Transl Med* 2022;**12**:e764. https://doi.org/10.1002/ctm2.764.
63. Jie X, Chen Y, Zhao Y. *et al.* Targeting KDM4C enhances CD8. *J Immunother Cancer* 2022;**10**:e003716. https://doi.org/10.1136/jitc-2021-003716.
64. Kim HY, Hong S. Multi-faceted roles of DNAJB protein in cancer metastasis and clinical implications. *Int J Mol Sci* 2022;**23**:14970. https://doi.org/10.3390/ijms232314970.
65. Miao W, Li L, Wang Y. A targeted proteomic approach for heat shock proteins reveals DNAJB4 as a suppressor for melanoma metastasis. *Anal Chem* 2018;**90**:6835–42. https://doi.org/10.1021/acs.analchem.8b00986.
66. Wang CC, Lin SY, Lai YH. *et al.* Dimethyl sulfoxide promotes the multiple functions of the tumor suppressor HLJ1 through activator protein-1 activation in NSCLC cells. *PloS One* 2012;**7**:e33772. https://doi.org/10.1371/journal.pone.0033772.
67. Uretmen Kagiali ZC, Sanal E, Karayel Ö. *et al.* Systems-level analysis reveals multiple modulators of epithelial-mesenchymal transition and identifies DNAJB4 and CD81 as novel metastasis inducers in breast cancer. *Mol Cell Proteomics* 2019;**18**:1756–71. https://doi.org/10.1074/mcp.RA119.001446.
68. Newman AM, Liu CL, Green MR. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;**12**:453–7. https://doi.org/10.1038/nmeth.3337.
69. Hong H, Yao S, Zhang Y. *et al.* In vivo miRNA knockout screening identifies miR-190b as a novel tumor suppressor. *PLoS Genet* 2020;**16**:e1009168. https://doi.org/10.1371/journal.pgen.1009168.
70. Fierheller CT, Guitton-Sert L, Alenezi WM. *et al.* A functionally impaired missense variant identified in French Canadian families implicates FANCI as a candidate ovarian cancer-predisposing gene. *Genome Med* 2021;**13**:186. https://doi.org/10.1186/s13073-021-00998-5.
71. Sishc BJ, Davis AJ. The role of the core non-homologous end joining factors in carcinogenesis and cancer. *Cancers (Basel)* 2017;**9**:81. https://doi.org/10.3390/cancers9070081.
72. Zhu J, He F, Hu S. *et al.* On the nature of human housekeeping genes. *Trends Genet* 2008;**24**:481–4. https://doi.org/10.1016/j.tig.2008.08.004.
73. Zhang Z, Lee JH, Ruan H. *et al.* Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat Commun* 2019;**10**:4562. https://doi.org/10.1038/s41467-019-12543-5.
74. Gruber AJ, Schmidt R, Ghosh S. *et al.* Discovery of physiological and cancer-related regulators of 3′ UTR processing with KAPAC. *Genome Biol* 2018;**19**(1):44. https://doi.org/10.1186/s13059-018-1415-3.
75. Statello L, Guo CJ, Chen LL. *et al.* Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 2021;**22**:96–118. https://doi.org/10.1038/s41580-020-00315-9.
76. Chae YK, Anker JF, Oh MS. *et al.* Mutations in DNA repair genes are associated with increased neoantigen burden and a distinct immunophenotype in lung squamous cell carcinoma. *Sci Rep* 2019;**9**:3235. https://doi.org/10.1038/s41598-019-39594-4.
77. Liu L, Ruiz J, O'Neill SS. *et al.* Favorable outcome of patients with lung adenocarcinoma harboring POLE mutations and expressing high PD-L1. *Mol Cancer* 2018;**17**:81. https://doi.org/10.1186/s12943-018-0832-y.
78. Chan TA, Yarchoan M, Jaffee E. *et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* 2019;**30**:44–56. https://doi.org/10.1093/annonc/mdy495.
79. Rospo G, Lorenzato A, Amirouchene-Angelozzi N. *et al.* Evolving neoantigen profiles in colorectal cancers with DNA repair defects. *Genome Med* 2019;**11**:42. https://doi.org/10.1186/s13073-019-0654-6.
80. Ott PA, Hu Z, Keskin DB. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 2017;**547**:217–21. https://doi.org/10.1038/nature22991.
81. Peng M, Mo Y, Wang Y. *et al.* Neoantigen vaccine: an emerging tumor immunotherapy. *Mol Cancer* 2019;**18**:128. https://doi.org/10.1186/s12943-019-1055-6.
82. Zhang Z, Lu M, Qin Y. *et al.* Neoantigen: a new breakthrough in tumor immunotherapy. *Front Immunol* 2021;**12**:672356. https://doi.org/10.3389/fimmu.2021.672356.