



Review

The Detection and Bioinformatic Analysis of Alternative 3' UTR Isoforms as Potential Cancer Biomarkers

Nitika Kandhari ¹, Calvin A. Kraupner-Taylor ¹, Paul F. Harrison ^{1,2}, David R. Powell ²
and Traude H. Beilharz ^{1,*}

- ¹ Development and Stem Cells Program, Department of Biochemistry and Molecular Biology, Monash Biomedicine Discovery Institute, Monash University, Melbourne, VIC 3800, Australia; nitika.kandhari@monash.edu (N.K.); Calvin.Kraupner-Taylor1@monash.edu (C.A.K.-T.); paul.harrison@monash.edu (P.F.H.)
- ² Monash Bioinformatics Platform, Monash University, Melbourne, VIC 3800, Australia; David.Powell@monash.edu
- * Correspondence: traude.beilharz@monash.edu; Tel.: +61-3-9902-9183

Abstract: Alternative transcript cleavage and polyadenylation is linked to cancer cell transformation, proliferation and outcome. This has led researchers to develop methods to detect and bioinformatically analyse alternative polyadenylation as potential cancer biomarkers. If incorporated into standard prognostic measures such as gene expression and clinical parameters, these could advance cancer prognostic testing and possibly guide therapy. In this review, we focus on the existing methodologies, both experimental and computational, that have been applied to support the use of alternative polyadenylation as cancer biomarkers.

Keywords: alternative polyadenylation; bioinformatics; 3' focused RNA-seq; scRNA-seq; cancer biomarkers



Citation: Kandhari, N.; Kraupner-Taylor, C.A.; Harrison, P.F.; Powell, D.R.; Beilharz, T.H. The Detection and Bioinformatic Analysis of Alternative 3' UTR Isoforms as Potential Cancer Biomarkers. *Int. J. Mol. Sci.* **2021**, *22*, 5322. <https://doi.org/10.3390/ijms22105322>

Academic Editor: Justin Wong

Received: 8 April 2021
Accepted: 6 May 2021
Published: 18 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Eukaryotic messenger RNA (mRNA) undergoes a highly regulated process of maturation before nuclear export and protein translation. This involves 5' end capping, RNA-splicing and 3' end cleavage and polyadenylation. Initially thought to be a static house-keeping function, mRNA 3' end formation has emerged as a major modulator of gene expression with implications in multiple disease settings [1,2].

Alternative polyadenylation (APA) is a regulatory mechanism that allows the production of coding and regulatory transcript isoforms from a single gene [3–6]. This occurs due to the presence of alternative adenylation sites in the genome and leads to significant transcriptome diversity. Nearly 70% of mammalian genes harbour multiple cleavage and polyadenylation sites i.e., poly(A) sites [7–10]. These sites can cause differential expression of mRNA transcripts by influencing their nuclear export, stability, subcellular localization, interaction with microRNAs, RNA binding proteins (RBPs), long non-coding RNAs (lncRNAs) and translation efficiency [11–15].

Two major types of APA events are described here; splicing-APA where protein sequence is changed, and tandem APA where only the extent of non-coding, regulatory information is altered (Figure 1). In the case of splicing-APA, the alternative poly(A) sites reside in introns of the coding sequences, generating protein isoforms with distinct Carboxy-termini. Such APAs are called coding region-APA (CR-APA) [16–18]. In the case of tandem APA, the poly(A) sites reside in the 3' UTRs resulting in transcript isoforms with invariant protein-coding sequence but 3' UTRs of different lengths. Such APAs are called UTR-APA [16–18]. In this review, we discuss the implications of APA and investigate the existing experimental and bioinformatic methods for detection, quantification and identification (Figure 2). Finally, the emerging role of APA signatures as cancer biomarkers will be explored.

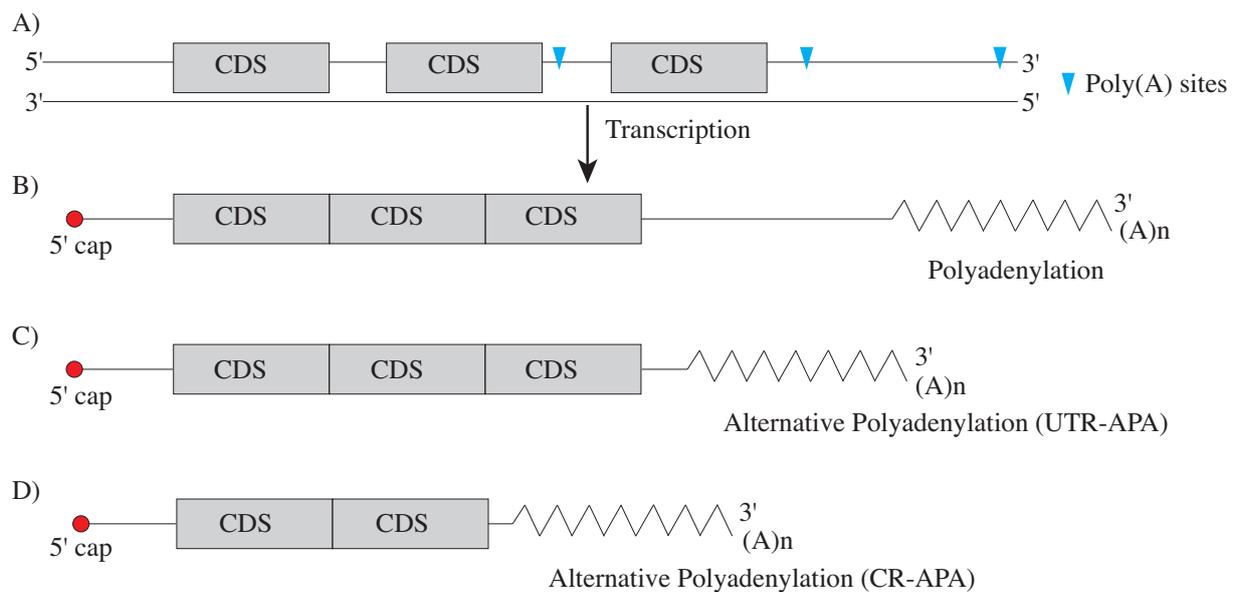


Figure 1. Alternative polyadenylation: (A) The schematic shows the 5' end, coding sequences (grey boxes), 3' UnTranslated Regions (3' UTRs) and polyadenylation sites (blue arrows) in DNA. (B) Polyadenylation is the enzymatic extension of ~200 Adenosine residues to the nascent mRNA, in this case the distal polyadenylation site was used. (B,C) In 3' UTR-APA, choice of the proximal cleavage and polyadenylation results in an mRNA with the same protein-coding potential but different 3' UTR length. (D) When a poly(A) signal is recognised in the intronic region, protein isoforms with distinct Carboxy-termini are generated in a process termed as CR-APA.

2. Implications of Alternative Polyadenylation

Since the discovery of APA in immunoglobulin M (IgM) and dihydrofolate reductase (DHFR) genes in 1980 [19,20], it has become clear that APA is the norm rather than the exception. At least 70% of human genes are subject to APA, and 3' UTR changes are often associated with physiological conditions including diseases such as cancer, immune dysfunction, congenital heart disease and dysplasia [21]. Where genes have the capacity to switch, short 3' UTRs generally associate with undifferentiated proliferative cells (e.g., stem cells) whereas the longer 3' UTR isoforms are favoured in differentiated tissues [22–24]. It has been suggested that the majority of APA genes switch to short mRNA isoforms in tumour cells [23–25]. Where there is an option to switch, mRNAs with longer 3' UTRs can cause reduced protein expression as a result of increased regulatory capacity. Whereas, increased stability and translation of short 3' UTR isoforms are some of the key functional consequences suggested for APA; for example, due to loss of microRNA-mediated repression [22,23]. APA-mediated evasion from microRNA repression can generate stable oncogenic mRNA isoforms with shorter 3' UTRs causing oncogenic activation [23]. It is important to note, however, that there are many exceptions to this trend. For example, the long-3' UTR isoform of the tumour suppressor PTEN is the more stable isoform and accounts for the bulk of its role in PI3K/AKT/mTOR signalling [26]. Albeit, the net consequence of 3' UTR shortening of PTEN still promotes tumour growth through reduced tumour suppressive activity.

Dynamic APA regulation has been reported in different healthy tissue types [27] in cellular proliferation, differentiation and development; in cancer cell transformation, and phenotypic response to extracellular stimuli [5,23,28–37]. For example, selection of a proximal poly(A) site resulting in 3' UTR shortening has been shown to associate with multiple cancers [25,38–41]. APA-mediated changes by CR-APA can diversify protein function. For example, a switch from proximal to distal APA in the IgM gene, results in a switch from a secreted to membrane-bound form of the antibody [42]. mRNAs with longer 3' UTRs can be subject to increased regulation and reduced protein expression. This is due to the inclusion of regulatory sequences such AU-rich and GU-rich sequences, RBP and

miRNA target sites all of which can negatively impact mRNA stability and/or translation efficiency [5]. As a result, shorter mRNA isoforms can escape regulation by loss of such sites leading to increased RNA stability and enhanced protein expression [23,33]. In addition to regulation in mRNA and its encoded protein, seminal work by Berkovits and Mayr (2015) shows that 3' UTRs can serve as a physical scaffold for ternary complex formation [13]. Alternative polyadenylation in long non-coding RNA has also been described and plays a role in tumorigenesis [43].

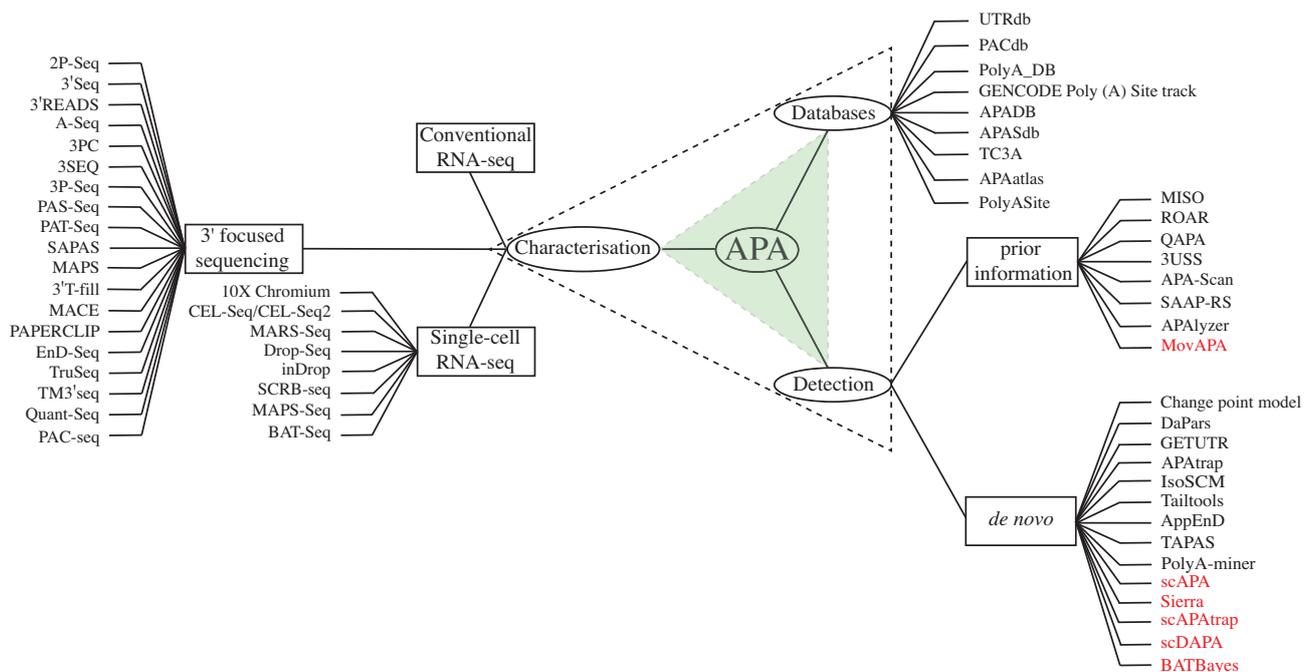


Figure 2. The triad of APA attributes: This review focuses on three attributes of genome-wide APA i.e., characterisation, detection and curation of APA databases. Currently, conventional RNA-seq, 3' focused seq and single-cell RNA-seq are the main methods for APA characterisation. APA databases hold information relating to APAs and 3' UTRs collated from a wide array of inputs. Detection requires bioinformatic methods for statistical ranking. These methods are classified based on prior knowledge from the databases or determined *de novo*. The bioinformatic methods for single-cell data analysis are shown in red.

3. Next-Generation Sequencing Based Techniques for Characterisation of APA

Global profiling of APA first became possible through accumulation of expressed sequence data in public databases and the development of high-content microarray. Bioinformatic analysis of expressed sequence tags (ESTs) and microarray studies helped detect many APA events in the late 90s [20,33,36,44–46]. Soon however, RNA sequencing (RNA-seq), became the major method for transcription profiling [47]. With RNA-seq it became possible to study the complete transcriptome by massively parallel short-read sequencing of cDNA libraries, allowing differential analysis of the gene expression between samples. Combined with biostatistics, this approach identified genes, and alternative isoforms of genes [47,48]. One of the drawbacks of bulk full-length RNA-seq, however, is an overall loss of read coverage of 5' and 3' ends of genes making it unreliable for detection of alternative transcriptional start-sites and APA [49]. Moreover, for many applications where only differential expression was required, sequencing the full-length transcriptome was unnecessary and costly. This motivated researchers to develop both 5' and 3' focused sequencing methods to sequence the specific transcriptomic regions of interest.

3.1. 3' focused RNA-seq Methods for APA Characterisation

Early studies for APA identification used Direct RNA-sequencing (DRS) [50] with the Helicos platform, now replaced by Oxford Nanopore and PacBio (Table 1). These provide a

quantitative view of APAs genome-wide, but are expensive and relatively low throughput. However, given that only the reads mapped to the 3' ends of mRNA are necessary for APA detection, a more pragmatic approach was to sequence only the mRNA 3' ends based on classic 3' RACE methods [51]. Most 3' focused methods enrich RNA carrying a poly(A) tail and include a variety of molecular biology methods to generate a library suitable for next generation sequencing [17]. The resulting sequencing data are bioinformatically analysed for identification of poly(A) sites and quantification of their differential usage. Current commercial and bespoke approaches to transcriptome-wide characterisation of APA are listed in Table 1.

Table 1. 3' focused RNA-sequencing approaches suitable for APA detection and characterization.

Name	Key Points	Typical Input	Sequence Target
PAIso-seq [52]	PacBio based method to capture poly(A) site, length, splicing, expression, PacBio is costly for the read coverage obtained, Low coverage	100 ng total RNA	Full length mRNA, Poly(A) tail included
Oxford Nanopore- Direct RNA sequencing [53]	The Nanopore instrument is capable of full-length direct RNA seq, tail lengths can also be extracted. Low coverage	500 ng poly(A)+ selected RNA	Full length mRNA, Poly(A) tail included
TAIL-seq [54]	rRNA depletion and 3' adaptor ligation, asymmetric paired end sequencing to determine tail length	~100 µg total RNA	Poly(A) tail length, Poly(A) site
mTAIL-seq [55]	3' oligo(dT) splinted ligation approach to TAIL-seq, reduced input RNA required. Paired-end sequencing.	1–5 µg total RNA	Poly(A) tail length, Poly(A) site
PAT-seq [56]	Single end read approach, 3' tagging by oligo templated RNA end extension	1 µg total RNA	Poly(A) tail length, Poly(A) site
PAL-seq [57]	Requires non-standard use of an Illumina instrument for tail length measurement by biotinylated dTTP incorporation. 3' end capture by splinted ligation	1–50 µg total RNA	Poly(A) tail length, Poly(A) site
Poly(A) seq [58]	Poly(A)+ RNA is captured with oligo(dT) conjugated magnetic beads, then 3' adaptors ligated 300 bp single end read. Samples sequenced on the Illumina NextSeq 500, 2 colour sequencing instrument	5.1 µg total RNA	Poly(A) tail length, Poly(A) site
TED-Seq [59]	3' adaptor ligation to Poly(A)+ RNA. Tail length is inferred from the size of the templated sequence after precise library size selection	100 ng poly(A)+ RNA	Poly(A) tail length, Poly(A) site
3P-seq [4]	Poly(A) tail removed by RNase H. Sequenced from the 3' end to determine site usage, adaptor addition by ligation to avoid internal priming	30 µg total RNA	Poly(A) site
2P-seq [60]	Poly(A) site detection by anchored oligo(dT) priming, sequencing from start of poly(A) tail in reverse	15 µg total RNA	Poly(A) site
3'-seq [30]	Poly(A) site detection by anchored oligo(dT) priming. Unique approach to fragmentation by rate limited nick translation of double stranded cDNA	2 µg DNase treated RNA	Poly(A) site
3'READS+ [37]	Poly(A) tail is trimmed by RNase H, 3' adaptor ligation	0.1–15 µg total RNA	Poly(A) site
3PC [61]	Anchored oligo(dT) primer to detect poly(A) site, 5' adaptor addition by circular ligation	100 µg total RNA	Poly(A) site
3'T-fill [62]	Anchored oligo(dT) primer to detect poly(A) site, sequenced from 3' end. 3'T-fill reaction - dA homopolymer region at 3' end filled with dTTPs on Illumina cBot cluster station before sequencing	0.5–10 µg total RNA	Poly(A) site
SAPAS [63]	Anchored oligo(dT) primer to detect poly(A) site, 5' adaptor addition by template switching	10 µg total RNA	Poly(A) site
PAS-seq [5]	Anchored oligo(dT) primer to detect poly(A) site, template switching 5' adaptor addition	0.5–1 µg poly(A)+ selected RNA	Poly(A) site
IVT-SAPAS [64]	<i>in vitro</i> transcription based amplification of cDNA for low input samples, poly(A) site detection by anchored oligo(dT) annealing	200 ng total RNA	Poly(A) site
PAPERCLIP [65]	RNA crosslinked, partially digested, and 3' ends immunoprecipitated via Poly(A) Binding protein, addresses internal priming issues, uses anchored oligo(dT) annealing for end detection	NA, starting material is tissue/cells	Poly(A) site
MACE [66]	GenXPro commercial kit, barcodes transcripts with UMIs to deal with PCR duplication	0.05 ng total RNA	Poly(A) site
Quant-Seq [67]	Lexogen commercial kit, oligo(dT) annealing to detect 3' ends, random forward priming of 2nd strand cDNA adds 5' adaptor	0.5–500 ng total RNA	Poly(A) site
MAPS [68]	3' end detection by anchored oligo(dT) priming, 5' adaptor addition by random forward priming of 2nd stand cDNA	1 µg total RNA	Poly(A) site
TM3' seq [69]	Fragmentation and 5' adaptor addition combined in a single step. 3' end detected via annealing of oligo(dT) primer	200 ng total RNA	Poly(A) site
PAC-seq [70]	Click-chemistry approach to fragmentation and 5' adaptor addition via reverse transcription termination by 3-azido-nucleotides. 3' end detected by oligo(dT) annealing	0.125–4 µg total RNA	Poly(A) site
EnD-Seq [71]	Targeted sequencing approach to 3' end detection, 3' adaptor ligation to total RNA, gene specific multiplex PCR of cDNA	1.5 µg total RNA	Poly(A) site, non-Poly(A) 3' ends

In general, 3' focused methods use oligo(dT) primers to target the poly(A) tail and thereby enrich sequencing of Poly(A)+ mRNAs. The steps that result in inclusion of sequencing adaptors, unique molecular identifiers (UMIs), size selection and library amplification are often varied between approaches. However, an RNA fragmentation step or other means to limit sequencing libraries to the region directly upstream of poly(A) sites is always included. Methods that use oligo(dT) primers bias away from ribosomal RNA and

other non-poly(A) RNA during reverse transcription. Albeit, rRNA decay intermediates carry poly(A)-tails and these can be abundantly detected. The use of oligo(dT) primers can cause significant mis-priming at internal A-rich regions leading to false poly(A) site identification. This can be addressed *in silico* by eliminating the putative poly(A) residues in A-rich regions [37,72]. Approaches that use 3' end ligation are less prone to mis-priming than those where cDNA synthesis is driven from annealed oligo(dT) primers. Both *in silico* and *in vitro* strategies have thus been developed to avoid the problem of internal priming [5,73]. PAPERCLIP, which uses immune-purification of the poly(A)-binding protein is an alternative method for detection of mRNA 3' ends [65,74]. While the methods discussed here focus on APA and 3' UTR isoforms, a subgroup of 3' focused sequencing methods additionally identify poly(A) tail length changes [54,56–58]. Finally, although direct RNA sequencing is currently the least affordable technology, it is the only method that can integrate APA with other mRNA processing events, such as alternative transcriptional start-site and splice sites.

3.2. Single-Cell Methods for mRNA 3' End Sequencing

High content research is experiencing a dramatic shift towards single-cell methods. Single-cell RNA-seq (scRNA-seq) allows transcriptome-wide analyses of gene expression in individual cells with high resolution [75] for discovery of novel cell types and their developmental trajectories [76–78]. The single cell methods include early cell-barcoding of samples which allows individual samples to be pooled and processed as a single sample. Early pooling (or early multiplexing) of samples significantly reduces the costs and increases sequencing-throughput [69]. Another interesting feature of single-cell RNA-seq methods is the use of UMIs, which allows detection of PCR duplicates while reporting the unique transcript counts and thus, removes PCR amplification bias [79,80]. Most scRNA-seq methods use 3' tag-based approach to generate reads enriched at 3' ends of mRNA similar to the approaches described above (Table 1). Several laboratories have already turned to scRNA-seq to study complex APA regulatory patterns in tissues and organs [10,81–83].

There are two major methods of scRNA-seq library generation that allow APA detection: Micro-well based methods and Microfluidic droplet-based methods. In microwell-based methods, cells are separated into microwells for barcode allocation and their transcriptome is reverse-transcribed; whereas in microfluidic droplet-based methods, individual cells are separated using nanolitre-sized droplets containing reagents for UMI and cDNA synthesis [84,85]. Each cell is lysed and mRNA 3' ends are annealed to primers containing UMI followed by RT reaction to generate the first cDNA strand. cDNAs are pooled for library amplification and sequencing. The information from individual cells is distinguished *in silico* based on the UMIs. The single cell approaches that allow detection of APA are listed in Table 2.

We have broadly classified the APA characterisation techniques into three categories: conventional RNA-seq, 3' focused RNA-seq and scRNA-seq methods (Figure 2). In the next sections, the bioinformatic tools available for 3' UTR detection and databases to store curated forms of this information are described.

Table 2. Single cell RNA-sequencing approaches suitable for APA detection and characterization.

Name	Overview	Scale
CEL-seq [86]	3' ends enriched by anchored oligo(dT) annealing including T7 promotor. cDNA amplified by <i>in vitro</i> transcription (IVT), amplified RNA fragmented and ligated to adaptor.	Manually isolated single cells
CEL-seq2 [87,88]	Application of CEL-seq to high throughput sequencing, UMI's added to reverse transcription oligo	Automated microfluidic sorting via Fluidigm C1 into wells
MARS-seq 2.0 [89]	3' end enrichment by anchored oligo(dT) annealing, included T7 promotor. cDNA amplified via IVT	384-well plate, FACS sorting
InDrop [80]	Application of CEL-seq to droplet-based sequencing for higher throughput	Droplet sequencing, inDrop system, 1CellBio
Drop-seq [90]	3' enrichment by oligo(dT) annealing RT, full length cDNA 5' labelled by template switching, oligo's with common barcode bound to beads, and separated into droplets. library prepared by Illumina Nextera XT DNA library prep kit	Droplet sequencing, custom instrument
10X Chromium [85]	3' enrichment by anchored oligo(dT) annealing, oligo's with common barcode bound to beads, and separated into droplets; library preparation with commercial kit GemCode Single-Cell 3' Gel Bead and library kit (now Chromium 10X)	Droplet sequencing, 10X genomics instrument
SCRB-seq [91]	3' enrichment by anchored oligo(dT) primer, template switching reaction for full length cDNA, library prepared by Illumina Nextera XT DNA library prep kit	384-well plate, FACS sorting
MAPS-seq [84]	3' ends enriched by biotinylated oligo(dT) annealing, RNA transcripts pulled down and samples pooled together using magnetic beads before RT. Full length cDNA 5' adaptor added via template switching, library prepared by Illumina Nextera XT DNA library prep kit	96-well plate, FACS sorting
BATSeq [92]	Method specifically developed to detect APA. 3' ends enriched by oligo(dT) annealing. 2nd strand cDNA IVT amplified	FACS sorting

4. Bioinformatic Methods for Detection of Poly(A) Sites

Bioinformaticians have sought to extract poly(A) site usage information from sequencing data, either using inference from read coverage in conventional RNA-seq or by quantitating read coverage data from the 3' focused methods (Figure 3). Some of these methods use known annotations from curated databases, whereas others identify peaks *de novo*. In this section, the existing bioinformatic tools for the detection of poly(A) sites from the sequencing data are discussed.

4.1. Databases for 3' UTR and APA Storage and Retrieval

The rapid accumulation of high-throughput data paved the way for investigation of RNA isoforms in a variety of physiological and pathological conditions [47,48]. RNA-seq emerged as a reliable tool to study transcriptome diversity due to its quantitative detection of alternative transcriptional start-site, splicing and APA events at nucleotide resolution. Public databases were created to store experimentally determined poly(A) sites and 3' UTR variants. In this section, we review the existing databases that catalogue the 3' UTRs in various organisms [27,93–99].

The primary data were collected from EMBL annotation records (UTRdb), transcript genome alignments in cDNA/ESTs (PACdb, PolyA_DB3, PolyA site track) inferred from RNA-seq (TC3A, APAAtlas) or curated from 3' focused RNA-seq (APADB, APASdb, PolyA-Site) (Table 3). Unfortunately, a number of useful resources have not been maintained (e.g. PACdb [95], APASdb [96] and TC3A [99]) and/or have been incorporated into updated resources. This leaves two main approaches for determination of global APA. (1) The bioinformatic extraction from consortium resources such as the Ensembl database, or more specifically GENCODE PolyA site track [100,101] which holds high-quality annotations for coding and non-coding regions and pseudogenes in the human genome. Or, (2) The use of specifically curated APA databases. The latter are collated from either direct 3' focused sequencing or by inference from RNA-seq. For example, APADB [97] reports poly(A) sites for coding and non-coding transcripts in human, mouse and chicken and reports

the loss of predicted miRNA binding sites from MACE-seq data. Whereas, PolyASite 2.0 [98] contains the most up to date curation from a multitude of 3' focused RNA-seq methods, re-analysed by protocol-specific data pre-processing steps for consistency in APA mining. Gene tracks can be downloaded for genome browser exploration. PolyA_DB3 [94] provides information about the genomic locations of poly(A) sites and the surrounding *cis* elements and a comparison of polyadenylation configuration between human and mouse orthologs. UTRdb [93] curates 5' and 3' UTR sequences and provides information about genome localisation and regulatory elements. It is integrated with UTRsite [93] which is a collection of experimentally validated functional regulatory motifs in 5' and 3' UTRs crosslinked with their protein partners. This integration allows users to retrieve data based on genomic coordinates and/or genes associated with encoded proteins using GO terms, PFAM domains, etc.

There is, however, still a relatively low availability of 3' focused RNA-seq data. Many cell, tissue and disease types are still missing, limiting the scope of these databases. To overcome this limitation, APAAtlas [27] provides a resource database of APA inferred from RNA-seq data in the Genotype-Tissue Expression (GTEx) project [102] using the DaPars [25] bioinformatic approach (see Section 4.2.2). A similar approach was recently used to mine RNA-seq from The Cancer Genome Atlas (TCGA) [103] where the inferred APA genes are provided in TC3A [99].

The annotation from these databases are useful for visualisation and interpretation of APA genome browsers such as the UCSC Genome Browser [104] or the Integrated Genome Browser [105]. Moreover, many tools for APA detection and quantification depend on database annotations to guide bioinformatic analysis as discussed in the section below.

Table 3. Bioinformatic databases for 3' UTR and APA storage and retrieval

Database	Primary Data Collection	Organism	Last Updated	URL
UTRdb [93]	5' and 3' UTR regions in EMBL/GenBank records	human, rodent, vertebrate, plant and fungi	2010	http://utrdb.ba.itb.cnr.it/
PACdb [95]	cDNA/ESTs	human, mouse, rat, dog, chicken, zebrafish, fugu, fruit fly, mosquito, nematode, Arabidopsis thaliana, rice and baker's yeast	inaccessible	http://harlequin.jax.org/pacdb/
PolyA_DB [72,94,106]	aligned cDNA/ESTs	human, mouse, rat, chicken and zebrafish	2018	http://polya-db.org/v3/
GENCODE Poly (A) site track [100,101]	cDNA/ESTs	human	2021	https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeGencodeV19
APADB [97]	MACE-Seq	human, mouse and chicken	2014	http://tools.genxpro.net/apadb/
APASdb [96]	SAPAS	human (22 normal and cancer tissues), mouse, zebrafish and some lancelet samples	inaccessible	http://mosas.sysu.edu.cn/utr
TC3A [99]	RNA-seq in TCGA	32 human cancer types	inaccessible	http://tc3a.org/
APAAtlas [27]	RNA-seq in GTEx project	>50 human normal tissue	2020	https://hanlab.uth.edu/apa/
PolyASite [98]	3'-Seq, 3'READS, DRS, QuantSeq_REV, SAPAS, PAPERCLIP, PolyA-seq, PAS-Seq, A-seq, 3P-Seq, DRS, 2P-Seq, PAT-seq	human, mouse and worm	2020	https://polyasite.unibas.ch/

4.2. Bioinformatic Methods for APA Detection and Quantification

The increasing interest in 3' UTR dynamics, and the growth of associated technologies required design of bioinformatic tools. Multiple approaches were designed to infer APA from conventional RNA-seq, as well as tools to extract it from 3' focused RNA-seq methods. Some APA detection methods rely on prior knowledge, while others involve the *de novo* detection of poly(A) sites.

4.2.1. APA Detection in RNA-seq Data Based on Prior APA Information

The section below provides a brief overview of the bioinformatic methods available for inference of APA from read-coverage in RNA-seq data, where known APA sites are used to guide analysis. The use of data-base derived APA information improves the accuracy of *in silico* APA detection.

Mixture of ISOforms (MISO) [107] was the first reported tool for detecting previously annotated 3' UTR isoforms, using a probabilistic framework to quantify alternative splicing (AS) and alternative polyadenylation. It identifies the differentially regulated AS/APA isoforms from the expression levels and delivers the probability of the origin of a read from a particular transcript isoform.

Ratio Of A Ratio (ROAR) [108] is an R-based program that identifies differential APA site usage in RNA-seq. The algorithm defines two distinct 3' UTRs in a gene, guided by APA databases, one which is shared by both the short and long 3' UTR isoform and the other which is present only in the long 3' UTR isoform. It scans the read-coverage for these two 3' UTR isoforms and computes the expression ratio (m/M) of reads falling in the two regions. To compare between conditions, the ratio of two isoform-expression ratios (m/M) is computed in different samples and is called the Ratio Of A Ratio. This ratio represents the tendency of expression of a short isoform or a long isoform in a given condition. A roar >1 indicates higher levels of short isoform (a roar <1 indicates higher levels of long isoform) in the first condition. This method derives APA annotations from APASdb and PolyA_DB2 [72].

Quantification of APA (QAPA) [9] uncovers APA from RNA-seq data by retrieval of 3' UTR annotations in GENCODE Poly (A) site track [101] and PolyASite 2.0 [98] and use these to construct an expanded reference library of annotated poly(A) sites and 3' UTR sequences. The sequences in this library are used to measure expression from RNA-seq data and estimate relative abundance of alternative 3' UTR isoforms. The method directly estimates the absolute alternative 3' UTR isoform expression from protein-coding genes. Then it computes the relative expression of each 3' UTR isoform among all isoforms to assess APA.

3' UTR Sequence Seeker (3USS) [109] is a web-server that analyses the transcript assembly file and automatically identifies transcripts with alternative 3' UTRs with respect to the reference genome of choice. The 3' UTRs are identified as the regions located immediately downstream of the stop-codon. These are then compared with previously annotated 3' UTRs in public databases, iGenomes (https://sapac.support.illumina.com/sequencing/sequencing_software/igenome.html) and GENCODE [100,101] to identify novel 3' UTRs and to detect length differences amongst existing and putative novel 3' UTRs. It provides the nucleotide sequence of the 3' UTR isoform along with their genomic coordinates and the UTR length differences.

APA-Scan [110] identifies genome-wide UTR-APA events by utilizing the predicted or experimentally verified poly(A) signals as reference for poly(A) sites estimating the 3' UTR read coverage from both aligned RNA-seq and 3' end-seq data to identify potential poly(A) sites. Then it pools all the aligned reads to identify peaks and cleavage sites in 3' UTRs which are considered as potential poly(A) sites. It performs a χ^2 -test on the experimentally determined or predicted cleavage site in the 3' UTR to compare APA between samples.

Significance Analysis of Alternative Polyadenylation using RNA-Seq (SAAP-RS) [111] uses RNA-seq samples from bulk, single cell and 3' focused (e.g., 3' READS+ [10]) approaches to identify APA events. The method calculates RNA-seq read counts upstream (UP) and downstream (DN) of every poly(A) site identified from PolyA_DB3 database and performs a statistical test to derive a p-value to compare the read distribution in UP and DN regions between two samples. The relative expression difference (RED) of the APA isoforms is used to identify genes with significantly altered 3' UTR lengths between cell types.

APAlyzer [112] is a Bioconductor package for identification of APAs in 3' UTR and intronic regions by calculating the RNA-seq read density (RD) after splitting the transcript 3' end regions based on the annotations derived from PolyA_DB3.

Due to their dependence on incomplete information of poly(A) sites, MISO, ROAR, 3USS and APAnalyzer may fail to detect uncharacterised APAs.

4.2.2. *de novo* APA Detection in RNA-seq Data

These are the bioinformatic methods that detect 3' UTR switching events in RNA-seq data without relying on prior knowledge. The methods use a variety of approaches, but a majority of tools scan the read-coverage to detect "change-points". A change point is a critical point that marks the shift or transition in the depth of read-coverage (Figure 3). The presence of more than one 3' UTR isoform creates a "step-down" inferred as the change points that define the APA boundaries.

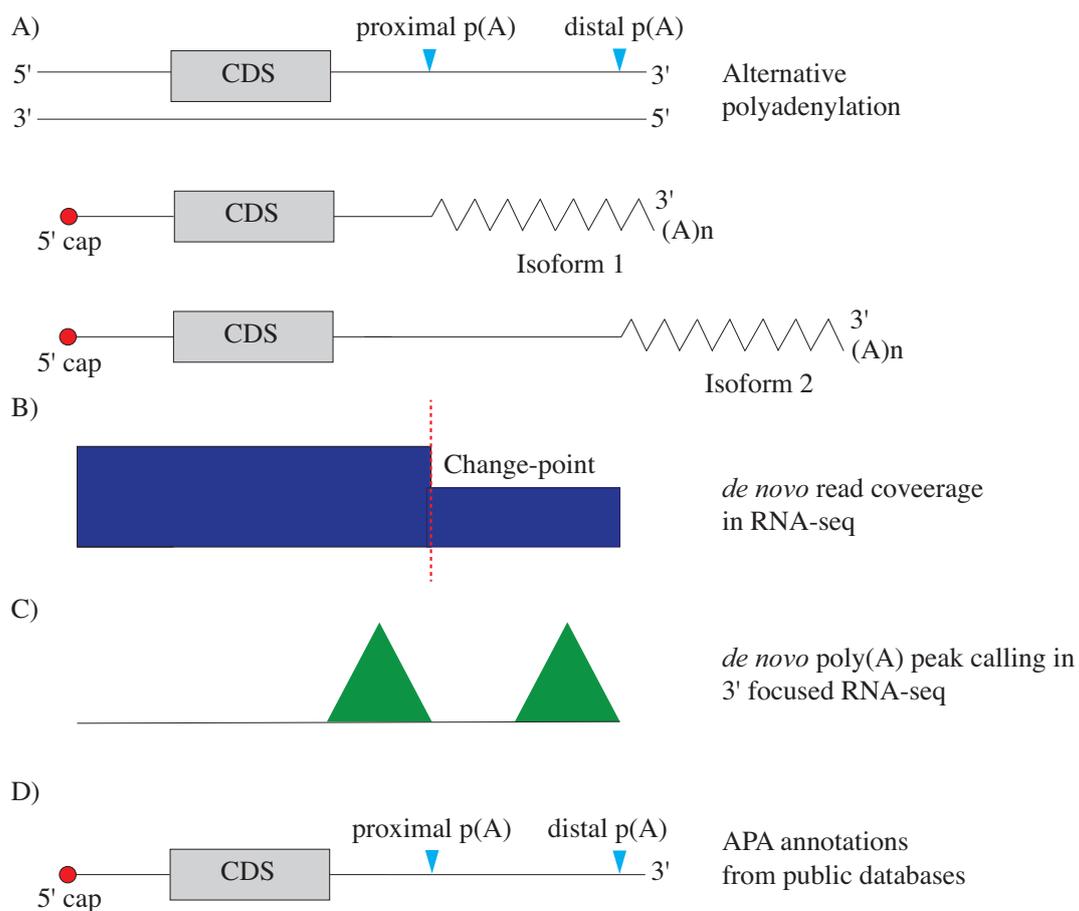


Figure 3. Detection of poly(A) sites: (A) Two polyadenylation sites, proximal and distal, result in expression of two isoforms. (B–D) Methods to determine the location of poly(A) sites: (B) *de novo* method to identify change-points in read-coverage of RNA-seq data. (C) *de novo* method to identify poly(A) peaks in 3' focused RNA-seq data. (D) combining read-coverage data with poly(A) site coordinates from APA databases.

Dynamic analysis of Alternative PolyAdenylation from RNA-Seq (**DaPars**) [25] performs *de novo* identification of APA in RNA-seq experiments. The method scans the read-coverage and identifies a distal poly(A) site present at the end-point of the longest 3' UTR among samples. It then seeks a model providing the best least-squares fit of the read-coverage along the gene up to the identified distal site. This model consists of the location of a proximal poly(A) site and the expression levels of the short and long isoforms in each of the two conditions. This best model provides both the location of a proximal site and the information required to calculate the "Percentage of Distal polyA site Usage Index" (PDUI) for each condition.

Tool for Alternative Polyadenylation site Analysis (**TAPAS**) [113] deals with more than two APA sites in genes as well as 3' UTRs with intronic regions. The tool is based on multiple change point inference model for finding change points in time series data,

but applies more stringent filtration techniques to discard false APA sites. The method is extended to identify APA sites that are differentially expressed across samples to infer genes that undergo 3' UTRs shortening/lengthening.

Global Estimation of The 3' UTR landscape based on RNA-seq (GETUTR) [114] is a Python-based method that uses RefSeq gene annotations to provides a landscape of 3' UTR and finds poly(A) sites by smoothing read-coverage to flatten the erroneous variations in the RNA-seq signal. The smoothing technique may generate many false poly(A) sites.

Isoform Structural Change Model (IsoSCM) [115] is a standalone transcript assembly tool that annotates mRNA 3' ends based on multiple change-point analysis to generate complete 3' UTR assemblies. It uses a statistical model to infer change points in a gene exhibiting a sharp increase or decrease in read-coverage and employs mathematical constraints to filter false APA sites. Although rare, introns occur in 3' UTRs and regulate gene expression [116,117]. Neither GETUTR nor IsoSCM consider intronic regions in their analysis and miss 3' UTRs that contain introns [113].

APAtap [118] uses an approach different from change-point or poly(A) peak calling (see Section 4.2.3). It extracts the known 3' UTR from genome annotations for each gene and extends it by a pre-defined length. A sliding window is used to scan the extended region by 1bp increments to identify changes in read coverage. The location of 3' UTR ends is determined by considering the mean read coverage in the current window, the previous window and the next window and a 3-step criterion is used to identify the precise 3' ends. The newly identified 3' UTRs are compared with the original genome annotation to procure novel 3' UTRs, the 3' end locations of which are then defined as the distal poly(A) sites. It then applies a least-squares model on read-coverage depth to identify the precise positions of poly(A) sites for each gene.

4.2.3. *de novo* APA Detection in 3' Focused Data

For every protocol listed in Table 1, bioinformatic methods were employed for data analysis. While some of them remain ad-hoc, others are available as stand-alone pipelines or packages which are discussed in this section.

The first reported **change-point model** [119] is based on a likelihood ratio test that detects any change in 3' UTR length. It assumes the existence of two 3' UTR isoforms in a gene, with a proximal and a distal poly(A) site. It then captures the percentage of read counts corresponding to each isoform, quantifies the expression ratio of the two isoforms across two conditions, treatment and control. The method also assumes a constant expression ratio of the two isoforms throughout the 3' UTR and tests for changes in the expression ratio. A change in this ratio marks the 3' UTR switching event and the site identifies as a poly(A) site. The Perl software can handle data from both RNA-seq and 3' focused protocols and has been tested for SAPAS [63].

Different from change point models, the bioinformatic methods developed for 3' focused RNA-seq identify poly(A) sites by peak-calling. Reads containing untemplated poly(A) sequences when compared to a reference genome are identified as 3' ends.

Tail Tools [56] is a suite of tools to process and analyse the reads rich in poly(A) tails. Tail Tools measures differential gene expression, differential poly(A) tail length and differential 3' end usage per gene. All the reads associated with each identified poly(A) peak are counted for each sample. The weitrax Bioconductor package [120] assigns a "shift score" and an associated precision weight to each gene with two or more APA sites relative to typical site usage. These scores and weights can then be used with limma [121] and topconfects [122] for differential testing. The topconfects package provides confidence bounds on the differential genes, thus provides a ranked gene list in the order of confident effect size i.e., how much shift is observed in the genes. Weitrax can handle data from both 3' focused RNA-seq methods and from single-cell RNA-seq experiments. Along with differential poly(A) site usage, it can also find differential tail length, and introduces some exploratory features like finding components of variation in data and identify genes with excess variation (or highly variable genes, HVGs).

These additional tools can also be applied to other 3' focused RNA-seq data such as Quant-seq and 10X Genomics single-cell RNA-seq data.

PolyA-miner [8] creates a matrix of poly(A) sites (as rows) and samples (as columns) from 3' focused sequencing data to apply non-negative matrix factorization which captures gene expression patterns. It first extracts all potential sample-wise poly(A) sites and pools them to construct a poly(A) library and then extensively filters out false poly(A) sites and maps the rest to their respective genes. The number of reads mapped gives the poly(A) peak count for each gene. The method accounts for all APA changes between proximal, intermediate and distal APA sites.

Application for mapping **EnD-Seq** data (**AppEnD**) [71] was reported along with EnD-Seq protocol but can also process data from PAS-Seq and A-Seq protocols and has the ability to automatically detect internally mis-primed A-tails, thus keeping only the true polyadenylated 3' ends. It outputs the transcript abundance ending at each nucleotide, resulting in a positional distribution of last templated nucleotides.

Most of these tools only identify UTR-APAs. They rely on gene annotations from reference genomes in ENSEMBL which provides annotations for 3' UTRs [123], but these are not differentiated by APA type. Independent of the reference genome annotations, **mountainClimber** [124] locates change points in the RNA-seq read coverage data to identify APA sites in coding and intronic regions and thus, differentiate between the two APA types.

4.2.4. APA Detection in 3' Tag-Based Single-Cell RNA-seq Data

The 3' focussed scRNA-Seq methods such as the popular 10X Chromium encouraged the development of bioinformatic tools to resolve complexity and study APA dynamics in single-cell data, which are discussed in this section.

Modeling and Visualization of dynamics of Alternative PolyAdenylation (MovAPA) [125] is an R package to measure APA. It extracts poly(A) site annotations from multiple sources like PolyASite2.0, PolyA_DB3, PlantAPAdb [126], APASdb, TAPAS, APATrap, DaPars and Cufflinks [127] to construct a library that stores expression levels, annotation, and sample information of poly(A) sites from different samples which is then used for the downstream analysis. While **movAPA** relies on prior poly(A) annotations, the following tools identify poly(A) peaks or compute differential APA usage *de novo*.

BATBayes [92] uses a statistical framework to compare variability in 3' UTR isoform usage in homogeneous cell populations from BAT-seq data. The analysis identifies poly(A) sites by UMI counting and only considers the two most abundant 3' UTR isoforms for each gene.

scAPA [128] is an R-script that combines various toolkits such as Samtools [129], Bedtools [130], Homer, UMI_tools [131], etc. for their analysis. It uses Homer to detect poly(A) site by peak-calling and uses *mclust* to separate overlapping peaks based on a Gaussian mixture model. It employs *featureCounts* [132] to quantify peak usage in each cell-type cluster and performs a χ^2 -test to detect dynamic APA events.

Sierra [81] applies the DEXSeq package [133], originally designed to detect differential exon usage in bulk RNA-Seq data, to APA usage in pseudo-bulk samples. As DEXSeq performs tests based on the negative binomial distribution, this method takes biological variation into account, which many other methods fail to do.

scAPATrap [82] employs peak-calling to detect potential poly(A) sites and integrates poly(A) read anchoring where reads with A/T stretches are used to determine the precise locations of the poly(A) sites, which other methods like **scAPA** and **Sierra** fail to do. It also splits the overlapping peaks into smaller peaks and then employs the **movAPA** package to compute APA.

scDAPA [134] computes the APA difference between samples or between cell-types within the same sample. It doesn't call poly(A) sites, instead, it employs a histogram-based approach to divide the reads in 3' ends into bins of the same width and computes a difference in the percentage of reads in each bin for a gene across two conditions. A Wilcoxon rank-sum test measures the significance of the differential APA usage in these bins.

5. The Repertoire of Cancer Biomarkers

The seminal study by Mayr and Sharp (2009) first showed the association of APA with cancer. Since that time, APA has been reported in multiple studies of cancer proliferation and transformation, as extensively reviewed by Gruber and Zavolan (2019) [39]. These APA genes have the potential to be used as prognostic markers in predicting cancer progression, risk stratification and even for developing personalised therapies [16,22,34,83,135–139].

Current prognostic tests rely on gene expression profiles [140,141]. But these may be improved by incorporating APA. Several APA genes have been proposed as novel prognostic biomarkers and some examples are shown in Table 4. These gene expression and APA signatures could be combined with drug-sensitivity data, and clinical covariates such as patient age, survival time, tumour stage, location and size to build a multivariate regression model [137,138]. For example, a recent study used linear regression model to connect APA events and drug sensitivity with clinical relevance, supporting their utility as biomarkers [137].

Table 4. APA genes as potential cancer biomarkers.

Cancer	Gene Markers	Signature APA	Physiological Effects	Molecular Role
Breast	PRELID1	Shortening of 3' UTR	increased protein expression	mitochondrial ROS signalling [139]
Breast	SNX3, YME1L1D, USP9X	Shortening of 3' UTR	increased protein levels in short isoform	EGF signalling [22]
adult T-cell lymphoma, large B-cell lymphoma, stomach adenocarcinoma	PD-L1 gene (CD274)	Shortening of 3' UTR	PD-1/PD-L1-mediated immune escape in cancer development; structural variants (SVs) disrupt 3' regulatory region of PDL1	T-cell modulator; PDCD1-mediated inhibitory pathway [136]
Colorectal cancer	IGF2BP1/IMP-1	Shortening of 3' UTR	increased protein levels; increased oncogenic transformation	Modulates pathogenesis [142]
TNBC, lung, esophageal, bladder, leukemia, ovarian	N4BP2L2, WDHD1, ZER1, ADGRL2, PRSS12, NPL, SIK3, SYNGR1, SCL2A3, UBE2G2	Shortening of 3' UTR	unfavourable prognosis	All are related to cancer development: cell cycle regulator and is involved in PI3K/Akt pathway; tumour antigen [138]
TNBC	PPIC, ZCCHC14, RTN1, PRCK8, CLIC2, CXCL8, SMAD6	Lengthening of 3' UTR	poor prognosis; response elements (MREs) in the lengthened 3' UTR leads to homologous gene repression and competing endogenous RNA (ceRNA) resulting in cancer progression; more miRNA binding sites	TGF- β pathway; autocrine NF- κ B/IL-8 (CXCL8) pathway responsible for cell migration; aberrant pathways and cancer progression [138]
TNBC (MB-231)	Caspase 6, DFFA (ICAD), DFFB (CAD), PARP1	Lengthening of 3' UTR	escape of apoptosis	Caspase pathway [63]
TNBC (MB-231)	cyclin D1, D2	Shortening of 3' UTR	promote cell cycling	Mitotic cell cycle; APC [63]

A 17-gene 3' UTR-based classifier was reported that divided patients into high and low risk groups, predicting risk in patients with triple-negative breast cancer (TNBC) significantly better than the classical clinicopathological risk [138]. The prognostic model in this study reported 10 APA genes that undergo 3' UTR shortening and were associated with poor prognosis. It also reported 7 APA genes that undergo 3' UTR lengthening and were associated with poor prognosis showing that APA-mediated gene regulation is more complicated than was first thought. In an important caveat, this study found the SMAD6 gene to be associated with poor prognosis in TNBC patients but that it favours survival in lung cancer patients, indicating that the APA events are tumour-dependent. The expression of APA genes detected by single-cell RNA-seq are now being shown to correlate with clinical outcomes of early-stage breast cancer in a single-cell data [83]. They report 53 cancer cell-specific APA genes with a distinct pattern of 3' UTR shortening and

an immune-specific APA signature with possible clinical utility in early stage breast cancer. However, of the many potential clinically relevant APA genes that have been reported, most have yet to be independently clinically validated.

In a disease setting like TNBC, which is highly aggressive and has a high recurrence rate, the lack of hormone receptors means the targeted therapies are not applicable. As a result, patients are treated with conventional radiotherapy or chemotherapy [143]. Better treatment methods are required. APA markers or the mechanism that cause APA could be used as targets for development of novel treatment therapies [144,145].

Based on current literature, APA appears to be associated with tumorigenicity in all cancer patients. The time is therefore ripe to take these smaller scale research findings into larger cohort studies to mine the full potential of APA as novel cancer biomarkers.

6. Conclusions

APA is an established mechanism for the generation of transcriptome diversity that impacts basic cellular functions, cancer proliferation and transformation and ultimately controls cellular fate. The development of bespoke RNA-seq technologies combined with bioinformatic methods and curated databases have paved the way for the potential of APA as cancer biomarkers to be tested at scale. These APA markers, if combined with standard prognostic measures such as gene expression and clinical covariates may contribute toward development of novel diagnostic tests and may facilitate personalised cancer therapies.

Author Contributions: Original draft preparation, N.K. and T.H.B.; technical contribution C.A.K.-T. generated Tables 1 and 2; writing—review and editing, N.K., C.A.K.-T., P.F.H., D.R.P., T.H.B.; supervision, D.R.P. and T.H.B.; funding acquisition, D.R.P. and T.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants from the National Health and Medical Research Council (NHMRC: APP1128250) and the Australian Research Council (ARC: DP170100569 and FT180100049). N.K. and C.A.K.-T. were supported by PhD stipends from the Monash University Biodiscovery Institute.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We thank members of the Beilharz laboratory and the Monash Bioinformatics platform for constructive feedback on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sweet, T.J.; Licatalosi, D.D. 3' end formation and regulation of eukaryotic mRNAs. *Methods Mol. Biol.* **2014**, doi:10.1007/978-1-62703-971-0_1.
2. Danckwardt, S.; Hentze, M.W.; Kulozik, A.E. 3' end mRNA processing: Molecular mechanisms and implications for health and disease. *EMBO J.* **2008**, *27*, doi:10.1038/sj.emboj.7601932.
3. Ozsolak, F.; Kapranov, P.; Foissac, S.; Kim, S.W.; Fishilevich, E.; Monaghan, A.P.; John, B.; Milos, P.M. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **2010**, *143*, doi:10.1016/j.cell.2010.11.020.
4. Jan, C.H.; Friedman, R.C.; Ruby, J.G.; Bartel, D.P. Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. *Nature* **2011**, *469*, 97–101, doi:10.1038/nature09616.
5. Shepard, P.J.; Choi, E.A.; Lu, J.; Flanagan, L.A.; Hertel, K.J.; Shi, Y. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **2011**, *17*, 761–772, doi:10.1261/rna.2581711.
6. Ren, F.; Zhang, N.; Zhang, L.; Miller, E.; Pu, J.J. Alternative Polyadenylation: A new frontier in post transcriptional regulation. *Biomark. Res.* **2020**, *8*, doi:10.1186/s40364-020-00249-6.
7. Liu, D.; Brockman, J.M.; Dass, B.; Hutchins, L.N.; Singh, P.; McCarrey, J.R.; MacDonald, C.C.; Graber, J.H. Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. *Nucleic Acids Res.* **2007**, *35*, 234–246. doi:10.1093/nar/gkl919.
8. Yalamanchili, H.K.; Alcott, C.E.; Ji, P.; Wagner, E.J.; Zoghbi, H.Y.; Liu, Z. PolyA-miner: Accurate assessment of differential alternative poly-adenylation from 3'Seq data using vector projections and non-negative matrix factorization. *Nucleic Acids Res.* **2020**, *48*, doi:10.1093/nar/gkaa398.
9. Ha, K.C.H.; Blencowe, B.J.; Morris, Q. QAPA: A new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.* **2018**, *19*, 45. doi:10.1186/s13059-018-1414-4.

10. Cheng, L.C.; Zheng, D.; Baljinnayam, E.; Sun, F.; Ogami, K.; Yeung, P.L.; Hoque, M.; Lu, C.W.; Manley, J.L.; Tian, B. Widespread transcript shortening through alternative polyadenylation in secretory cell differentiation. *Nat. Commun.* **2020**, *11*, doi:10.1038/s41467-020-16959-2.
11. Elkon, R.; Ugalde, A.P.; Agami, R. Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* **2013**, *14*, doi:10.1038/nrg3482.
12. Tian, B.; Manley, J.L. Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 18–30. doi:10.1038/nrm.2016.116.
13. Berkovits, B.D.; Mayr, C. Alternative 3'UTRs act as scaffolds to regulate membrane protein localization. *Nature* **2015**, *522*, 363–367. doi:10.1038/nature14321.
14. Mayr, C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol.* **2016**, *26*. doi: 10.1016/j.tcb.2015.10.012.
15. Millevoi, S.; Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.* **2010**, *38*, 2757–2774. doi:10.1093/nar/gkp1176.
16. Giammartino, D.C.D.; Nishida, K.; Manley, J.L. Mechanisms and Consequences of Alternative Polyadenylation. *Mol. Cell* **2011**, *43*, doi:10.1016/j.molcel.2011.08.017.
17. Chen, W.; Jia, Q.; Song, Y.; Fu, H.; Wei, G.; Ni, T. Alternative Polyadenylation: Methods, Findings, and Impacts. *Genom. Proteom. Bioinform.* **2017**, *15*, 287–300, doi:10.1016/j.gpb.2017.06.001.
18. Turner, R.E.; Pattison, A.D.; Beilharz, T.H. Alternative polyadenylation in the regulation and dysregulation of gene expression. *Semin. Cell Dev. Biol.* **2018**, *75*, doi:10.1016/j.semcdb.2017.08.056.
19. Rogers, J.; Early, P.; Carter, C.; Calame, K.; Bond, M.; Hood, L.; Wall, R. Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. *Cell* **1980**, *20*, doi:10.1016/0092-8674(80)90616-9.
20. Setzer, D.R.; McGrogan, M.; Nunberg, J.H.; Schimke, R.T. Size heterogeneity in the 3' end of dihydrofolate reductase messenger RNAs in mouse cells. *Cell* **1980**, *22*, doi:10.1016/0092-8674(80)90346-3.
21. Chatterjee, S.; Pal, J.K. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol. Cell* **2009**, *101*, doi:10.1042/bc20080104.
22. Akman, H.B.; Oyken, M.; Tuncer, T.; Can, T.; Erson-Bensan, A.E. 3'UTR shortening and EGF signaling: implications for breast cancer. *Hum. Mol. Genet.* **2015**, *24*, 6910–6920, doi:10.1093/hmg/ddv391.
23. Mayr, C.; Bartel, D.P. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **2009**, *138*, 673–684, doi:10.1016/j.cell.2009.06.016.
24. 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, doi:10.1371/journal.pone.0008419.
25. Xia, Z.; Donehower, L.A.; Cooper, T.A.; Neilson, J.R.; Wheeler, D.A.; Wagner, E.J.; Li, W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* **2014**, *5*, 5274, doi:10.1038/ncomms6274.
26. Thivierge, C.; Tseng, H.W.; Mayya, V.K.; Lussier, C.; Gravel, S.P.; Duchaine, T.F. Alternative polyadenylation confers Pten mRNAs stability and resistance to microRNAs. *Nucleic Acids Res.* **2018**, *46*, doi:10.1093/nar/gky666.
27. Hong, W.; Ruan, H.; Zhang, Z.; Ye, Y.; Liu, Y.; Li, S.; Jing, Y.; Zhang, H.; Diao, L.; Liang, H.; et al. APAAtlas: Decoding alternative polyadenylation across human tissues. *Nucleic Acids Res.* **2020**, *48*, doi:10.1093/nar/gkz876.
28. Zhang, H.; Lee, J.Y.; Tian, B. Biased alternative polyadenylation in human tissues. *Genome. Biol.* **2005**, *6*, R100, doi:10.1186/gb-2005-6-12-r100.
29. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470–476, doi:10.1038/nature07509.
30. Lianoglou, S.; Garg, V.; Yang, J.L.; Leslie, C.S.; Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* **2013**, *27*, 2380–2396, doi:10.1101/gad.229328.113.
31. Lee, S.H.; Singh, I.; Tisdale, S.; Abdel-Wahab, O.; Leslie, C.S.; Mayr, C. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* **2018**, *561*, 127–131, doi:10.1038/s41586-018-0465-8.
32. Singh, I.; Lee, S.H.; Sperling, A.S.; Samur, M.K.; Tai, Y.T.; Fulciniti, M.; Munshi, N.C.; Mayr, C.; Leslie, C.S. Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.* **2018**, *9*, 1716, doi:10.1038/s41467-018-04112-z.
33. Sandberg, R.; Neilson, J.R.; Sarma, A.; Sharp, P.A.; Burge, C.B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **2008**, *320*, 1643–1647, doi:10.1126/science.1155390.
34. Ji, Z.; Lee, J.Y.; Pan, Z.; Jiang, B.; Tian, B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7028–7033, doi:10.1073/pnas.0900028106.
35. Singh, P.; Alley, T.L.; Wright, S.M.; Kamdar, S.; Schott, W.; Wilpan, R.Y.; Mills, K.D.; Graber, J.H. Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res.* **2009**, *69*, 9422–9430. doi:10.1158/0008-5472.CAN-09-2236.
36. Flavell, S.W.; Kim, T.K.; Gray, J.M.; Harmin, D.A.; Hemberg, M.; Hong, E.J.; Markenscoff-Papadimitriou, E.; Bear, D.M.; Greenberg, M.E. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **2008**, *60*, 1022–1038, doi:10.1016/j.neuron.2008.11.029.
37. Zheng, D.; Liu, X.; Tian, B. 3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA. *RNA* **2016**, *22*, 1631–1639. doi:10.1261/rna.057075.116.

38. Xue, Z.; Warren, R.L.; Gibb, E.A.; MacMillan, D.; Wong, J.; Chiu, R.; Hammond, S.A.; Yang, C.; Nip, K.M.; Ennis, C.A.; Hahn, A.; Reynolds, S.; Birol, I. Recurrent tumor-specific regulation of alternative polyadenylation of cancer-related genes. *BMC Genom.* **2018**, *19*, doi:10.1186/s12864-018-4903-7.
39. Gruber, A.J.; Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* **2019**, *20*, doi:10.1038/s41576-019-0145-z.
40. Jenal, M.; Elkon, R.; Loayza-Puch, F.; Haaften, G.V.; Kühn, U.; Menzies, F.M.; Vrieling, J.A.; Bos, A.J.; Drost, J.; Rooijers, K.; et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **2012**, *149*, doi:10.1016/j.cell.2012.03.022.
41. Gruber, A.J.; Schmidt, R.; Ghosh, S.; Martin, G.; Gruber, A.R.; van Nimwegen, E.; Zavolan, M. Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. *Genome Biol.* **2018**, *19*, doi:10.1186/s13059-018-1415-3.
42. Takagaki, Y.; Seipelt, R.L.; Peterson, M.L.; Manley, J.L. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **1996**, *87*, 941–952. doi:10.1016/s0092-8674(00)82000-0.
43. Naveed, A.; Cooper, J.A.; Li, R.; Hubbard, A.; Chen, J.; Liu, T.; Wilton, S.D.; Fletcher, S.; Fox, A.H. NEAT1 polyA-modulating antisense oligonucleotides reveal opposing functions for both long non-coding RNA isoforms in neuroblastoma. *Cell Mol. Life Sci.* **2020**, doi:10.1007/s00018-020-03632-6.
44. Edwalds-Gilbert, G.; Veraldi, K.L.; Milcarek, C. Alternative poly(A) site selection in complex transcription units: Means to an end? *Nucleic Acids Res.* **1997**, *25*, doi:10.1093/nar/25.13.2547.
45. Gautheret, D.; Poirot, O.; Lopez, F.; Audic, S.; Claverie, J.M. Alternate polyadenylation in human mRNAs: A large-scale analysis by EST clustering. *Genome Res.* **1998**, *8*, doi:10.1101/gr.8.5.524.
46. Tian, B.; Hu, J.; Zhang, H.; Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **2005**, *33*, doi:10.1093/nar/gki158.
47. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, doi:10.1038/nrg2484.
48. Smyth, G.K.; Ritchie, M.E.; Law, C.W.; Alhamdoosh, M.; Su, S.; Dong, X.; Tian, L. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* **2018**, *5*, doi:10.12688/f1000research.9005.3.
49. Ma, F.; Fuqua, B.K.; Hasin, Y.; Yukhtman, C.; Vulpe, C.D.; Lusic, A.J.; Pellegrini, M. A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods 06 Biological Sciences 0604 Genetics. *BMC Genom.* **2019**, *20*, doi:10.1186/s12864-018-5393-3.
50. Ozsolak, F.; Milos, P.M. Transcriptome profiling using single-molecule direct RNA sequencing. *Methods Mol. Biol.* **2011**, *733*, doi:10.1007/978-1-61779-089-8_4.
51. Scotto-Lavino, E.; Du, G.; Frohman, M.A. 3'End cDNA amplification using classic RACE. *Nat. Protoc.* **2007**, *1*, doi:10.1038/nprot.2006.481.
52. Liu, Y.; Nie, H.; Liu, H.; Lu, F. Poly(A) inclusive RNA isoform sequencing (PAIso-seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. *Nat. Commun.* **2019**, *10*, doi:10.1038/s41467-019-13228-9.
53. Krause, M.; Niazi, A.M.; Labun, K.; Cleuren, Y.N.T.; Müller, F.S.; Valen, E. TailFindR: Alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA* **2019**, *25*, doi:10.1261/rna.071332.119.
54. Chang, H.; Lim, J.; Ha, M.; Kim, V.N. TAIL-seq: Genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* **2014**, *53*, 1044–1052, doi:10.1016/j.molcel.2014.02.007.
55. Lim, J.; Lee, M.; Son, A.; Chang, H.; Kim, V.N. MTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev.* **2016**, *30*, doi:10.1101/gad.284802.116.
56. Harrison, P.F.; Powell, D.R.; Clancy, J.L.; Preiss, T.; Boag, P.R.; Traven, A.; Seemann, T.; Beilharz, T.H. PAT-seq: A method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA* **2015**, *21*, 1502–1510, doi:10.1261/rna.048355.114.
57. Subtelny, A.O.; Eichhorn, S.W.; Chen, G.R.; Sive, H.; Bartel, D.P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **2014**, *508*, 66–71, doi:10.1038/nature13007.
58. Yu, F.; Zhang, Y.; Cheng, C.; Wang, W.; Zhou, Z.; Rang, W.; Yu, H.; Wei, Y.; Wu, Q.; Zhang, Y. Poly(A)-seq: A method for direct sequencing and analysis of the transcriptomic poly(A)-tails. *PLoS ONE* **2020**, *15*, e0234696, doi:10.1371/journal.pone.0234696.
59. Woo, Y.M.; Kwak, Y.; Namkoong, S.; Kristjánssdóttir, K.; Lee, S.H.; Lee, J.H.; Kwak, H. TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Rep.* **2018**, *24*, doi:10.1016/j.celrep.2018.08.084.
60. Spies, N.; Burge, C.B.; Bartel, D.P. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res.* **2013**, *23*, 2078–2090, doi:10.1101/gr.156919.113.
61. Mata, J. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol.* **2013**, *10*, 1407–1414, doi:10.4161/rna.25758.
62. Wilkening, S.; Pelechano, V.; Jarvelin, A.I.; Tekkedil, M.M.; Anders, S.; Benes, V.; Steinmetz, L.M. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* **2013**, *41*, e65, doi:10.1093/nar/gks1249.
63. Fu, Y.; Sun, Y.; Li, Y.; Li, J.; Rao, X.; Chen, C.; Xu, A. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.* **2011**, *21*, 741–747, doi:10.1101/gr.115295.110.
64. Fu, Y.; Ge, Y.; Sun, Y.; Liang, J.; Wan, L.; Wu, X.; Xu, A. IVT-SAPAS: Low-Input and Rapid Method for Sequencing Alternative Polyadenylation Sites. *PLoS ONE* **2015**, *10*, e0145477, doi:10.1371/journal.pone.0145477.

65. Hwang, H.W.; Park, C.Y.; Goodarzi, H.; Fak, J.J.; Mele, A.; Moore, M.J.; Saito, Y.; Darnell, R.B. PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage. *Cell Rep.* **2016**, *15*, 423–435, doi:10.1016/j.celrep.2016.03.023.
66. Zawada, A.M.; Rogacev, K.S.; Müller, S.; Rotter, B.; Winter, P.; Fliser, D.; Heine, G.H. Massive analysis of cDNA Ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. *Epigenetics* **2014**, *9*, doi:10.4161/epi.26931.
67. Moll, P.; Ante, M.; Seitz, A.; Reda, T. QuantSeq 3'mRNA sequencing for RNA quantification. *Nat. Methods* **2014**, *11*, i–iii, doi:10.1038/nmeth.f.376.
68. Zhou, Y.; Li, H.R.; Huang, J.; Jin, G.; Fu, X.D. Multiplex analysis of polyA-linked sequences (MAPS): An RNA-Seq strategy to profile poly(A+) RNA. *Methods Mol. Biol.* **2014**, *1125*, doi:10.1007/978-1-62703-971-0_15.
69. Pallares, L.F.; Picard, S.; Ayroles, J.F. TM3'seq: A tagmentation-mediated 3' sequencing approach for improving scalability of RNAseq experiments. *G3 Genes Genomes Genetics* **2020**, *10*, 143–150.
70. Routh, A.; Ji, P.; Jaworski, E.; Xia, Z.; Li, W.; Wagner, E.J. Poly(A)-ClickSeq: click-chemistry for next-generation 3-end sequencing without RNA enrichment or fragmentation. *Nucleic Acids Res.* **2017**, *45*, e112, doi:10.1093/nar/gkx286.
71. Welch, J.D.; Slevin, M.K.; Tatomer, D.C.; Duronio, R.J.; Prins, J.F.; Marzluff, W.F. EnD-Seq and AppEnD: Sequencing 3' ends to identify nontemplated tails and degradation intermediates. *RNA* **2015**, *21*, doi:10.1261/rna.048785.114.
72. Lee, J.Y.; Yeh, I.; Park, J.Y.; Tian, B. PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res.* **2007**, *35*, doi:10.1093/nar/gkl870.
73. Derti, A.; Garrett-Engle, P.; Macisaac, K.D.; Stevens, R.C.; Sriram, S.; Chen, R.; Rohl, C.A.; Johnson, J.M.; Babak, T. A quantitative atlas of polyadenylation in five mammals. *Genome. Res.* **2012**, *22*, 1173–1183, doi:10.1101/gr.132563.111.
74. Hwang, H.W.; Darnell, R.B. Comprehensive Identification of mRNA Polyadenylation Sites by PAPERCLIP. *Methods Mol. Biol.* **2017**, *1648*, 79–93, doi:10.1007/978-1-4939-7204-3_7.
75. Ziegenhain, C.; Vieth, B.; Parekh, S.; Reinius, B.; Guillaumet-Adkins, A.; Smets, M.; Leonhardt, H.; Heyn, H.; Hellmann, I.; Enard, W. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **2017**, *65*, 631–643.e4, doi:10.1016/j.molcel.2017.01.023.
76. Camp, J.G.; Wollny, D.; Treutlein, B. Single-cell genomics to guide human stem cell and tissue engineering. *Nat. Methods* **2018**, *15*, 661–667, doi:10.1038/s41592-018-0113-0.
77. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome. Res.* **2015**, *25*, 1491–1498, doi:10.1101/gr.190595.115.
78. Stegle, O.; Teichmann, S.A.; Marionni, J.C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **2015**, *16*, 133–145, doi:10.1038/nrg3833.
79. Islam, S.; Zeisel, A.; Joost, S.; Manno, G.L.; Zajac, P.; Kasper, M.; Lönnerberg, P.; Linnarsson, S. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **2014**, *11*, doi:10.1038/nmeth.2772.
80. Klein, A.M.; Mazutis, L.; Akartuna, I.; Tallapragada, N.; Veres, A.; Li, V.; Peshkin, L.; Weitz, D.A.; Kirschner, M.W. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **2015**, *161*, doi:10.1016/j.cell.2015.04.044.
81. Patrick, R.; Humphreys, D.T.; Janbandhu, V.; Oshlack, A.; Ho, J.W.; Harvey, R.P.; Lo, K.K. Sierra: Discovery of differential transcript usage from polyA-captured single-cell RNA-seq data. *Genome. Biol.* **2020**, *21*, doi:10.1186/s13059-020-02071-7.
82. Wu, X.; Liu, T.; Ye, C.; Ye, W.; Ji, G. scAPATrap: Identification and quantification of alternative polyadenylation sites from single-cell RNA-seq data. *Briefings Bioinform.* **2020**, doi:10.1093/bib/bbaa273.
83. Kim, N.; Chung, W.; Eum, H.H.; Lee, H.O.; Park, W.Y. Alternative polyadenylation of single cells delineates cell types and serves as a prognostic marker in early stage breast cancer. *PLoS ONE* **2019**, *14*, e0217196,
84. Park, M.; Lee, D.; Bang, D.; Lee, J.H. MAPS-seq: Magnetic bead-assisted parallel single-cell gene expression profiling. *Exp. Mol. Med.* **2020**, *52*, 804–814, doi:10.1038/s12276-020-0433-x.
85. Zheng, G.X.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **2017**, *8*, 14049, doi:10.1038/ncomms14049.
86. Hashimshony, T.; Wagner, F.; Sher, N.; Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* **2012**, *2*, doi:10.1016/j.celrep.2012.08.003.
87. Yanai, I.; Hashimshony, T. CEL-Seq2—Single-cell RNA sequencing by multiplexed linear amplification. *Methods Mol. Biol.* **2019**, *1979*, doi:10.1007/978-1-4939-9240-9_4.
88. Hashimshony, T.; Senderovich, N.; Avital, G.; Klochendler, A.; de Leeuw, Y.; Anavy, L.; Gennert, D.; Li, S.; Livak, K.J.; Rozenblatt-Rosen, O.; et al. CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **2016**, *17*, doi:10.1186/s13059-016-0938-8.
89. Keren-Shaul, H.; Kenigsberg, E.; Jaitin, D.A.; David, E.; Paul, F.; Tanay, A.; Amit, I. MARS-seq2.0: An experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nat. Protoc.* **2019**, *14*, doi:10.1038/s41596-019-0164-4.
90. Macosko, E.Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A.R.; Kamitaki, N.; Martersteck, E.M.; Trombetta, J.J.; Weitz, D.A.; Sanes, J.R.; Shalek, A.K.; Regev, A.; McCarroll, S.A. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **2015**, *161*, doi:10.1016/j.cell.2015.05.002.
91. Soumillon, M.; Cacchiarelli, D.; Semrau, S.; van Oudenaarden, A.; Mikkelsen, T. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* **2014**, doi:10.1101/003236.

92. Velten, L.; Anders, S.; Pekowska, A.; Jarvelin, A.I.; Huber, W.; Pelechano, V.; Steinmetz, L.M. Single-cell polyadenylation site mapping reveals 3' isoform choice variability. *Mol. Syst. Biol.* **2015**, *11*, 812, doi:10.15252/msb.20156198.
93. Pesole, G.; Liuni, S.; Grillo, G.; Saccone, C. UTRdb: A specialized database of 5'- and 3'-untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* **1998**, *26*, 192–195, doi:10.1093/nar/26.1.192.
94. Wang, R.; Nambiar, R.; Zheng, D.; Tian, B. PolyA-DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.* **2018**, *46*, doi:10.1093/nar/gkx1000.
95. Brockman, J.M.; Singh, P.; Liu, D.; Quinlan, S.; Salisbury, J.; Graber, J.H. PACdb: PolyA Cleavage Site and 3'-UTR Database. *Bioinformatics* **2005**, *21*, 3691–3693, doi:10.1093/bioinformatics/bti589.
96. You, L.; Wu, J.; Feng, Y.; Fu, Y.; Guo, Y.; Long, L.; Zhang, H.; Luan, Y.; Tian, P.; Chen, L.; et al. APASdb: a database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. *Nucleic Acids Res.* **2015**, *43*, D59–D67, doi:10.1093/nar/gku1076.
97. Muller, S.; Rycak, L.; Afonso-Grunz, F.; Winter, P.; Zawada, A.M.; Damrath, E.; Scheider, J.; Schmah, J.; Koch, I.; Kahl, G.; et al. APADB: a database for alternative polyadenylation and microRNA regulation events. *Database* **2014**, *2014*, doi:10.1093/database/bau076.
98. Herrmann, C.J.; Schmidt, R.; Kanitz, A.; Artimo, P.; Gruber, A.J.; Zavolan, M. PolyASite 2.0: A consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Res.* **2020**, *48*, D174–D179, doi:10.1093/nar/gkz918.
99. Feng, X.; Li, L.; Wagner, E.J.; Li, W. TC3A: The Cancer 3'UTR Atlas. *Nucleic Acids Res.* **2018**, *46*, doi:10.1093/nar/gkx892.
100. Frankish, A.; Diekhans, M.; Ferreira, A.M.; Johnson, R.; Jungreis, I.; Loveland, J.; Mudge, J.M.; Sisu, C.; Wright, J.; Armstrong, J.; et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **2019**, *47*, doi:10.1093/nar/gky955.
101. Harrow, J.; Frankish, A.; Gonzalez, J.M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B.L.; Barrell, D.; Zadissa, A.; Searle, S.; et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **2012**, *22*, 1760–1774, doi:10.1101/gr.135350.111.
102. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, doi:10.1038/ng.2653.
103. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Wspolczesna Onkol.* **2015**, *1A*, doi:10.5114/wo.2014.47136.
104. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The Human Genome Browser at UCSC. *Genome Res.* **2002**, *12*, doi:10.1101/gr.229102.
105. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, doi:10.1038/nbt.1754.
106. Zhang, H.; Hu, J.; Recce, M.; Tian, B. PolyA_DB: A database for mammalian mRNA polyadenylation. *Nucleic Acids Res.* **2005**, *33*, D116–D120.
107. Katz, Y.; Wang, E.T.; Airoidi, E.M.; Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **2010**, *7*, doi:10.1038/nmeth.1528.
108. Grassi, E.; Mariella, E.; Lembo, A.; Molineris, I.; Provero, P. Roar: Detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinform.* **2016**, *17*, 423, doi:10.1186/s12859-016-1254-8.
109. Pera, L.L.; Mazzapioda, M.; Tramontano, A. 3USS: A web server for detecting alternative 3'UTRs from RNA-seq experiments. *Bioinformatics* **2015**, *31*, 1845–1847, doi:10.1093/bioinformatics/btv035.
110. Fahmi, N.A.; Chang, J.W.; Nasserreddeen, H.; Ahmed, K.T.; Fan, D.; Yong, J.; Zhang, W. APA-Scan: Detection and Visualization of 3'-UTR APA with RNA-seq and 3'-end-seq Data. *bioRxiv* **2020**, doi:10.1101/2020.02.16.951657.
111. Guvenek, A.; Tian, B. Analysis of alternative cleavage and polyadenylation in mature and differentiating neurons using RNA-seq data. *Quant. Biol.* **2018**, *6*, doi:10.1007/s40484-018-0148-3.
112. Wang, R.; Tian, B. APAlzyer: A bioinformatics package for analysis of alternative polyadenylation isoforms. *Bioinformatics* **2020**, *36*, doi:10.1093/bioinformatics/btaa266.
113. Arefeen, A.; Liu, J.; Xiao, X.; Jiang, T. TAPAS: Tool for alternative polyadenylation site analysis. *Bioinformatics* **2018**, *34*, 2521–2529, doi:10.1093/bioinformatics/bty110.
114. Kim, M.; You, B.H.; Nam, J.W. Global estimation of the 3' untranslated region landscape using RNA sequencing. *Methods* **2015**, *83*, 111–117.
115. Shenker, S.; Miura, P.; Sanfilippo, P.; Lai, E.C. IsoSCM: improved and alternative 3' UTR annotation using multiple change-point inference. *RNA* **2015**, *21*, 14–27, doi:10.1261/rna.046037.114.
116. Bicknell, A.A.; Cenik, C.; Chua, H.N.; Roth, F.P.; Moore, M.J. Introns in UTRs: Why we should stop ignoring them. *BioEssays* **2012**, *34*, doi:10.1002/bies.201200073.
117. Barrett, L.W.; Fletcher, S.; Wilton, S.D. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **2012**, *69*, doi:10.1007/s00018-012-0990-9.
118. Ye, C.; Long, Y.; Ji, G.; Li, Q.Q.; Wu, X. APAtrop: Identification and quantification of alternative polyadenylation sites from RNA-seq data. *Bioinformatics* **2018**, *34*, doi:10.1093/bioinformatics/bty029.
119. Wang, W.; Wei, Z.; Li, H. A change-point model for identifying 3'UTR switching by next-generation RNA sequencing. *Bioinformatics* **2014**, *30*, doi:10.1093/bioinformatics/btu189.

120. Harrison, P.F. Tools for Matrices with Precision Weights, Test and Explore Weighted or Sparse Data. 2020. Available online: [doi:10.18129/B9.bioc.weitrix](https://doi.org/10.18129/B9.bioc.weitrix) (accessed on 8 April 2021).
121. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, doi:10.1093/nar/gkv007.
122. Harrison, P.F.; Pattison, A.D.; Powell, D.R.; Beilharz, T.H. Topconfects: A package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome Biol.* **2019**, *20*, doi:10.1186/s13059-019-1674-7.
123. Ensembl insights: How are UTRs annotated? Ensembl Blog. Available online: <https://www.ensembl.info/> (accessed on 8 April 2021).
124. Cass, A.A.; Xiao, X. mountainClimber Identifies Alternative Transcription Start and Polyadenylation Sites in RNA-Seq. *Cell Syst.* **2019**, *9*, 393–400.e6, doi:10.1016/j.cels.2019.07.011.
125. Ye, W.; Liu, T.; Fu, H.; Ye, C.; Ji, G.; Wu, X. movAPA: modeling and visualization of dynamics of alternative polyadenylation across biological samples. *Bioinformatics* **2020**, doi:10.1093/bioinformatics/btaa997.
126. Zhu, S.; Ye, W.; Ye, L.; Fu, H.; Ye, C.; Xiao, X.; Ji, Y.; Lin, W.; Ji, G.; Wu, X. PlantAPAdb: A comprehensive database for alternative polyadenylation sites in plants. *Plant Physiol.* **2020**, *182*, doi:10.1104/pp.19.00943.
127. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Baren, M.J.V.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, doi:10.1038/nbt.1621.
128. Shulman, E.D.; Elkon, R. Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data. *Nucleic Acids Res.* **2019**, *47*, doi:10.1093/nar/gkz781.
129. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, doi:10.1093/bioinformatics/btp352.
130. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, doi:10.1093/bioinformatics/btq033.
131. Smith, T.; Heger, A.; Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **2017**, *27*, doi:10.1101/gr.209601.116.
132. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, doi:10.1093/bioinformatics/btt656.
133. Anders, S.; Reyes, A.; Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **2012**, *22*, doi:10.1101/gr.133744.111.
134. Ye, C.; Zhou, Q.; Wu, X.; Yu, C.; Ji, G.; Saban, D.R.; Li, Q.Q. ScDAPA: Detection and visualization of dynamic alternative polyadenylation from single cell RNA-seq data. *Bioinformatics* **2020**, *36*, doi:10.1093/bioinformatics/btz701.
135. Masamha, C.P.; Wagner, E.J. The contribution of alternative polyadenylation to the cancer phenotype. *Carcinogenesis* **2018**, *39*, 2–10, doi:10.1093/carcin/bgx096.
136. Kataoka, K.; Shiraishi, Y.; Takeda, Y.; Sakata, S.; Matsumoto, M.; Nagano, S.; Maeda, T.; Nagata, Y.; Kitanaka, A.; Mizuno, S.; et al. Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature* **2016**, *534*, 402–406.
137. Xiang, Y.; Ye, Y.; Lou, Y.; Yang, Y.; Cai, C.; Zhang, Z.; Mills, T.; Chen, N.Y.; Kim, Y.; Ozguc, F.M.; et al. Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. *J. Natl. Cancer Inst.* **2018**, *110*, 379–389, doi:10.1093/jnci/djx223.
138. Wang, L.; Hu, X.; Wang, P.; Shao, Z.M. The 3'UTR signature defines a highly metastatic subgroup of triple-negative breast cancer. *Oncotarget* **2016**, *7*, 59834–59844, doi:10.18632/oncotarget.10975.
139. Gillen, A.E.; Brechbuhl, H.M.; Yamamoto, T.M.; Kline, E.; Pillai, M.M.; Hesselberth, J.R.; Kabos, P. Alternative Polyadenylation of PRELID1 Regulates Mitochondrial ROS Signaling and Cancer Outcomes. *Mol. Cancer Res.* **2017**, *15*, 1741–1751, doi:10.1158/1541-7786.MCR-17-0010.
140. Schwab, M. MammaPrint Test. *Encycl. Cancer* **2015**, doi:10.1007/978-3-642-27841-9_7029-2.
141. Jensen, M.B.; Lænkholm, A.V.; Nielsen, T.O.; Eriksen, J.O.; Wehn, P.; Hood, T.; Ram, N.; Buckingham, W.; Ferree, S.; Ejlertsen, B. The Prosigna gene expression assay and responsiveness to adjuvant cyclophosphamide-based chemotherapy in premenopausal high-risk patients with breast cancer. *Breast Cancer Res.* **2018**, *20*, doi:10.1186/s13058-018-1012-0.
142. Andres, S.F.; Williams, K.N.; Plesset, J.B.; Headd, J.J.; Mizuno, R.; Chatterji, P.; Lento, A.A.; Klein-Szanto, A.J.; Mick, R.; Hamilton, K.E.; et al. IMP1 3'UTR shortening enhances metastatic burden in colorectal cancer. *Carcinogenesis* **2019**, *40*, doi:10.1093/carcin/bgy153.
143. Triple-Negative Breast Cancer: Overview, Treatment, and More. Available online: [Breastcancer.org](https://www.breastcancer.org/) (accessed on 8 April 2021).
144. Chou, J.; Quigley, D.A.; Robinson, T.M.; Feng, F.Y.; Ashworth, A. Transcription-Associated Cyclin-Dependent Kinases as Targets and Biomarkers for Cancer Therapy. *Cancer Discov.* **2020**, *10*, 351–370, doi:10.1158/2159-8290.CD-19-0528.
145. Ogorodnikov, A.; Levin, M.; Tattikota, S.; Tokalov, S.; Hoque, M.; Scherzinger, D.; Marini, F.; Poetsch, A.; Binder, H.; Macher-Göppinger, S.; et al. Transcriptome 3' end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat. Commun.* **2018**, *9*, 5331, doi:10.1038/s41467-018-07580-5.