

REVIEW

Population-scale genetic control of alternative polyadenylation and its association with human diseases

Lei Li^{1*}, Yumei Li², Xudong Zou¹, Fuduan Peng², Ya Cui², Eric J. Wagner³, Wei Li²

¹ Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China

² Division of Computational Biomedicine, Department of Biological Chemistry, School of Medicine, University of California, Irvine, CA 92697, USA

³ Department of Biochemistry & Molecular Biology, University of Texas Medical Branch, Galveston, TX 77555, USA

* Correspondence: lilei@szbl.ac.cn

Received November 10, 2020; Revised January 6, 2021; Accepted January 25, 2021

Background: Genome-wide association studies (GWAS) have identified thousands of genomic non-coding variants statistically associated with many human traits and diseases, including cancer. However, the functional interpretation of these non-coding variants remains a significant challenge in the post-GWAS era. Alternative polyadenylation (APA) plays an essential role in post-transcriptional regulation for most human genes. By employing different poly(A) sites, genes can either shorten or extend the 3'-UTRs that contain *cis*-regulatory elements such as miRNAs or RNA-binding protein binding sites. Therefore, APA can affect the mRNA stability, translation, and cellular localization of proteins. Population-scale studies have revealed many inherited genetic variants that potentially impact APA to further influence disease susceptibility and phenotypic diversity, but systematic computational investigations to delineate the connections are in their earliest states.

Results: Here, we discuss the evolving definitions of the genetic basis of APA and the modern genomics tools to identify, characterize, and validate the genetic influences of APA events in human populations. We also explore the emerging and surprisingly complex molecular mechanisms that regulate APA and summarize the genetic control of APA that is associated with complex human diseases and traits.

Conclusion: APA is an intermediate molecular phenotype that can translate human common non-coding variants to individual phenotypic variability and disease susceptibility.

Keywords: GWAS; eQTL; disease; alternative polyadenylation

Author summary: Genome-wide association studies (GWAS) have identified hundreds of inherited single-nucleotide polymorphisms (SNPs) associated with human diseases. However, they fall short of linking SNPs with dysregulation of specific genes and the molecular mechanisms responsible for disease/traits. By investigating the genetic basis of alternative polyadenylation we may gain insight into how GWAS non-coding variants influence complex diseases.

INTRODUCTION

With the exception of replication-dependent histone mRNA processing, all mRNA 3'-end processing events involve endonucleolytic cleavage at poly(A) sites (PAS) followed by the addition of a poly(A) tail. Because mRNAs, in general, have more than one PAS, alternative

polyadenylation (APA) can generate mRNAs of varying lengths through the selection of a specific cleavage site [1]. Besides these protein-coding mRNAs, long non-coding RNAs (lncRNAs) have also been regulated by APA. About 66% of lncRNAs undergo APA and mostly located within the upstream poly(A) exons [2]. Thus, APA is a widespread and evolutionarily conserved

mechanism in the regulation of mammalian genes and it is estimated that approximately 70% of human genes undergo APA [3].

APA can be generally classified into four different types. The two most common APA classes are tandem 3'-UTR APA, in which alternative PASs are located within the same terminal exon (Fig. 1A), and alternative terminal exon APAs (sometimes called 'splicing-APA') (Fig. 1B), in which PASs are located in distinct terminal exons. The less-frequent APA classes include intronic APA and internal exon APA (Fig. 1C, D). APA is tightly regulated via a combination of cis-regulatory sequences and APA regulators. Because the 3'-UTR region hosts many essential regulatory elements, such as AU-rich elements, microRNAs or RNA-binding protein sites, alteration of 3'-UTR length could alter function, stability, and translation efficiency of target mRNAs [4]. The consequence of APA on mRNA stability/degradation is primarily mediated through regulating the presence and accessibility of these regulatory elements in different cellular contexts [5]. APA can also impact other mRNA functions such as mRNA translation, mRNA nuclear export, and localization [4]. Not only on the mRNA level, but APA mediated 3'-UTR alteration can also regulate the protein localization that is independent of RNA localization [1]. For example, long 3'-UTR of CD47 localizes to the plasma membrane by recruiting the effect protein to the site of translation, while short 3'-UTR isoform of CD47 localizes to the ER and functions in the apoptosis regulation [6]. Recent studies also show that 3'-UTR can be cleaved off to form small non-coding RNAs which act independently as repressors or activators in prokaryotes [7–9] and polyadenylation near the end of 3'-UTR preserved that of the prokaryotic ancestor, which is also essential to the stability and degradation of mRNAs in human mitochondria [10]. Such events can have phenotypic impacts on both normal development and the progression of diseases, such as cancer [1,5,11].

Genome-wide association studies (GWAS) have significantly expanded our understanding of common inherited genetic variation effects on complex human diseases and traits, for which thousands of non-coding variants have been identified and associated with gene regulatory activities [12]. Despite massive experimental efforts, many GWAS loci cannot be rationalized as impacting normal mRNA expression levels. For example, Chun *et al.* found that only a small fraction of autoimmune disease risks are likely explained by impacting basal gene expression [12]. Recent studies also suggest that genetic variations in RNA processing, such as alternative splicing and APA, likely perform an independent but equally as important role as genetic variations impacting transcription [13] in associating with GWAS loci. Here we focused on the emerging evidence regarding the genetic role of APA in translating common genetic variations to phenotypes.

EXPERIMENTAL AND COMPUTATIONAL TOOLS TO DETECT GENETIC INFLUENCES OF APA EVENTS

RT-PCR/exon arrays

Quantitative RT-PCR analyses revealed that genetic variations can have critical regulatory consequences impacting APA events (Fig. 2, Table 1). For example, Yang *et al.* reported that a single-base change in the poly(A) signal could alter the polyadenylation pattern of *DHFR* [14]. APA of the *HLA-DQA1* gene was also associated with genetic variations [17]. A single-nucleotide polymorphism (SNP) (rs10954213) within the *IRF5* 3'-UTR region can alter both the length of 3'-UTR itself and mRNA stability [19]. In addition to these individual cases, a limited number of studies investigate the genetic influence on APA on a more global scale using exon

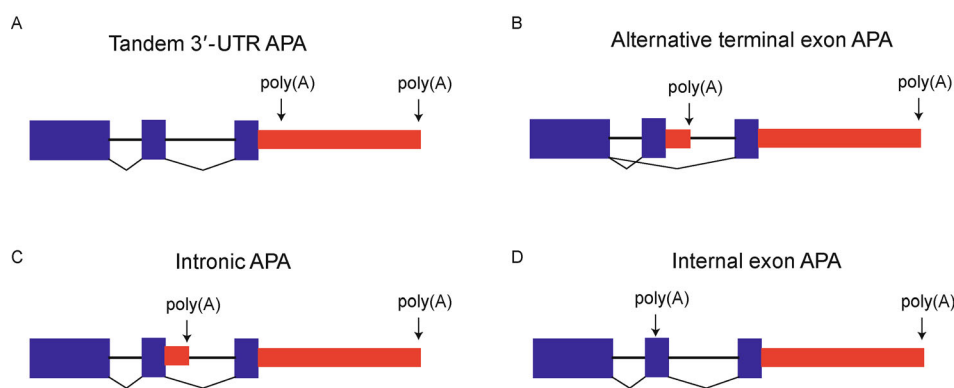


Figure 1. Four different APA types. (A) Tandem 3'-UTR APA that alternative poly(A) sites are located within the same terminal exon. (B) Alternative terminal exon APA that alternative poly(A) sites are located within distinct terminal exons. There are also two less-frequency APA class including (C) intronic APA and (D) internal exon APA.

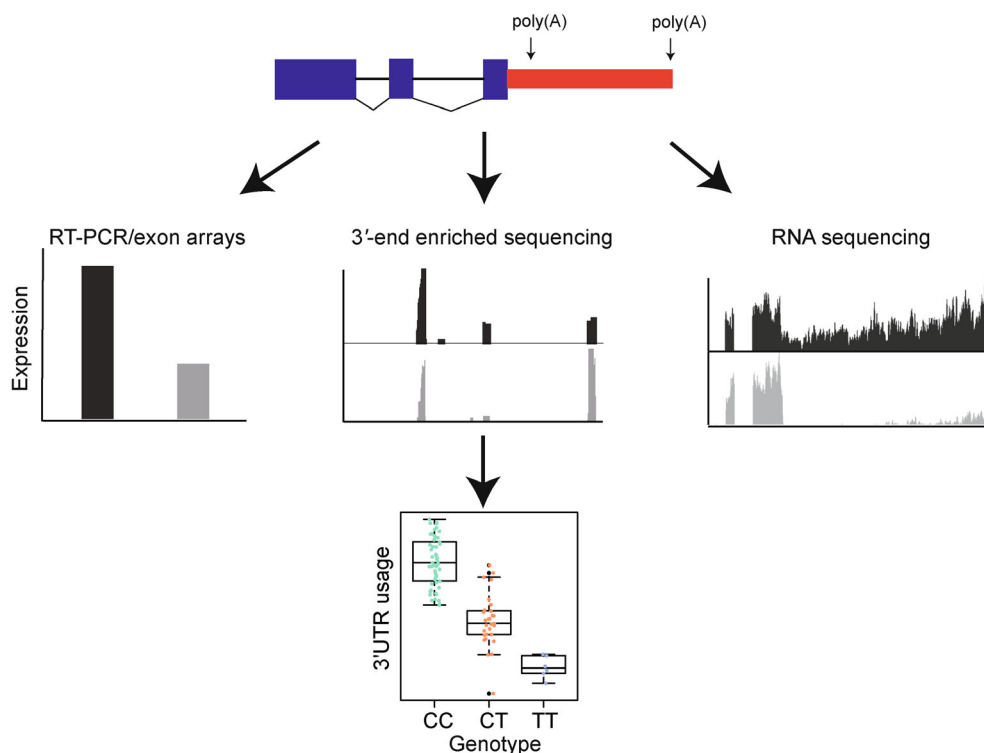


Figure 2. Experimental approaches to detect genetic control of APA. There are primarily three approaches including RT-PCR/exon arrays, 3'-end enriched sequencing approaches that detect 3'-ends using oligo(dT) primer based reverse transcription, and RNA sequencing approaches.

Table 1 Literature reported genetic-associated APA target genes

Year	PMID	Journal	Tissues	Genes	Refs.
1994	8041633	Nucleic Acids Research	Hamster lung cells	<i>DHFR</i>	[14]
1995	7585580	Cancer Research	Bladder and colon	<i>NAT1</i>	[15]
1999	10098839	Journal of Neurochemistry	Whole blood; human placental JAR cells	<i>SLAC6A4 (hSERT)</i>	[16]
2003	14651525	HLA	EBV-transformed lymphocytes	<i>HLA-DQA1</i>	[17]
2006	17189288	Human Molecular Genetics	EBV-transformed lymphocytes	<i>IRF5</i>	[18]
2007	17412832	PNAS	EBV-transformed lymphocytes	<i>IRF5</i>	[19]
2007	17220214	Journal of Medical Genetics	Monocyte cell lines	<i>GIMAP5</i>	[20]
2008	18193047	Nature Genetics	Lymphoblastoid cell line	<i>IRF5, ERAP1, TAP2, PPIL2, PTER, WARS2, MRP43</i>	[21]
2008	19067360	Human Mutation	Blood (mostly white cells)	<i>SLC7A1</i>	[22]
2012	22916029	PLoS Genetics	Lymphoblastoidcellline	<i>EIF2A, IRF5, DIP2B, NAB1</i>	[23]
2012	22431634	PNAS	Brain	<i>SLAC6A4(5-HTT)</i>	[24]
2013	23818875	PLoS Genetics	Blood	<i>ARNT, THEM4, F11R, ERAP1, DCP2, IRF5, STAT6, LPCA2, AMFR, CAPS</i>	[25]
2013	24098465	PLoS One	EBV-transformed lymphocytes	<i>ATP1B1</i>	[26]

arrays. Fraser *et al.* analyzed 176 human lymphoblastoid cell lines and reported that 37.9% exhibited changes in 3'-UTR constituency. They further investigated the influence of genetic polymorphisms on alternative transcript iso-

forms and found that 16 out of 20 significant genetic associations were APA-related [27]. Kwan *et al.* analyzed the 57 lymphoblastoid cell lines (LCL) samples sequenced by exon tiling arrays [21] and found that

55% of the genes were strongly associated with isoform changes. The selection of an alternative splice site could also result in differential stop codon usage and create further variability in 3'-UTR length, such as with the genes *ATPIF1* and *TAP2* [21]. However, potential weaknesses of all of these methods are that they depend on annotated poly(A) sites (Table 2) and are limited by inherent experimental caveats associated with microarrays such as poor or cross-hybridization [11].

3'-End enriched RNA sequencing (3'-Seq) has advanced the global discovery and functional characterization of APA sites

There are now nearly twenty 3'-end enriched sequencing approaches [14], including the QuantSeq 3'mRNA-Seq Library Prep Kit, Quantitative tag-based sequencing (3'-Seq) [17], Poly(A)-ClickSeq [19] and 3'READS+ [27]. These 3'-end enriched approaches, which primarily detect 3'-ends using oligo(dT) primer based reverse transcription, have been employed to efficiently investigate the global effects of variations on APA (Fig. 2). Yoon *et al.* performed 3'-end RNA sequencing in human B-lymphoblastoid cell lines from six individuals and identified the essential role of genetic variants by altering polyadenylation signals that could further lead to gene expression changes [23]. Mittleman *et al.* applied 3'-Seq to both the nuclear and total mRNA fractions of 52 lymphoblastoid cell lines. They identified 602 genetic variants that were associated with APA [28]. These 3'-end approaches have also been applied to studies of non-human model organisms. Using 80 inbred *Drosophila* wild isolates, Cannavo *et al.* identified 311 genes with genetic variations affecting their 3'-UTR length [29]. Although these APA protocols have increased the sensitivity of detecting precise locations of poly(A) sites, they are restricted by several technical issues, such as internal priming and general noise [30]. In addition, they have not been widely adopted due to the lack of a unified APA profiling method and the small sample sizes of previous studies (Table 2).

In addition to these sequencing technologies, analytical methods have been developed for these 3'-end enriched techniques. DPAC [31] is one of the first tools that streamlined the data analysis and includes preprocessing,

poly(A)-site identification, poly(A)-clustering and differential Poly(A) clusters usage. PolyA-miner [30] is another recently developed tool that enables *de novo* detection of differential APA using iterative consensus clustering and vector projections algorithms.

The RNA-seq approach can be used to detect genetic control of APA

RNA-seq is routinely used to measure gene expression in human genetic studies [32]. Capitalizing on the rich RNA-seq databases, the relative isoform abundance of individual APA events can also be captured by RNA-seq, which enables scientists to investigate genetic variants that act on APA on a population-scale. Early studies identified the genetic control of APA primarily through analyses of generic mRNA transcript structure with limited variants or coupling with splicing [33]. For example, DeepSAGE sequencing can target mRNA 3'-ends [25]. Using this technique to analyze 94 individual samples, Zhernakova *et al.* identified and validated SNPs that could affect APA usage, thereby potentially influencing mRNA stability. Lappalainen *et al.* analyzed the RNA-seq data from 462 LCLs cell lines and identified 639 genes in which genetic variants were associated with their altered mRNA transcript isoforms with 43% of these variants associated with an alternative 3'-end [32]. In another study, Mariella *et al.* analyzed Geuvadis RNA sequencing data for 373 European individuals and identified 2530 APA events associated with genetic variants [34]. More recently, Li *et al.* has constructed the first comprehensive atlas of human 3'-UTR alternative polyadenylation quantitative trait loci (3'aQTLs) using data from the Genotype-Tissue-Expression (GTEx) Project [35]: ~0.4 million genetic variants associated with APA of target genes across 46 tissues from 467 individuals [36]. Although the RNA-seq-based approaches may not be as accurate as 3'-end enriched approaches, they have tremendous upsides to allow for population-scale studies with larger sample sizes and broader conditions.

Along with these sequencing technology advances, several computational approaches have also been developed for genome-wide investigations of how genetic variations impact APA. These can be broadly classified as either annotation-based or *de novo* methods. The annota-

Table 2 Features of three experimental approaches to detect genetic influence of APA

Experiment	Known pAs annotation required	Throughput	Downstream <i>in silico</i> tools available	Population-scale samples available
qRT-PCR/Exon array	Yes	Relatively low	–	–
3'-seq	No	High	DPAC, PolyA-miner	No
RNA-seq	No	High	QAPA, Roar, GETUTR, TAPAS, APATrap, Dapars	Yes

tion-based methods such as QAPA [37] and Roar [38] rely on existing poly(A) sites that are annotated in the GENCODE, polyADB database [39], PolyAsite [40], or APASdb databases [41]. For example, QAPA employs sailfish to calculate isoform expression and then estimate the poly(A) usage by using the ratio of isoform expression to the sum of the expression of all detected 3'-UTR isoforms. These tools often provide greater accuracy and sensitivity. However, a potential limitation is that annotated transcript isoforms can only account for a small percentage of poly(A) sites [40]. These annotated poly(A) sites are compiled for various tissues and conditions, thus may lack the comprehensive information for a particular tissue or cell type. There are also several *de novo* analytical algorithms, such as GETUTR [42], TAPAS [43], APATrap [44] and DaPars [11,45]. DaPars is the first of its kind for the *de novo* identification of dynamic APA events based on localized changes in 3'-UTR RNA-seq read density and was used to identify differential APA events in tumors and normal tissues. For a given transcript, DaPars identifies the distal poly(A) site independent of the gene model, corrects for potential RNA-seq non-uniformity bias, and uses a linear regression model to infer the *de novo* proximal poly(A) site as an optimal fitting point that can best explain localized read density changes. APA usage differences can be quantified as the change in percentage of distal poly(A) site usage index. DaPars version 2 extended an earlier DaPars analysis of pairwise tumor/normal tissue comparisons [11,45] to include multiple RNA-seq joint analyses. The dynamics of APA genes can be determined based on a two-normal mixture model. Another important advantage of DaPars version 2 is that it supports multi-threading and requires significantly less processing time than other tools such as APATrap, which required an over 100-fold longer runtime than DaPars v2 to analyze the same 1,000 transcripts. Therefore, DaPars version 2 is quite suitable for running large-scale population RNA-seq analyses.

EMERGING MOLECULAR MECHANISMS OF APA REGULATION

As a crucial post-transcriptional regulation mechanism, APA is precisely controlled by cis regulatory elements and trans-acting factors [1,4,46]. Dysregulation of APA often results in hematological diseases or immunological diseases and cancer [11,47–50]. Several mechanisms underlying these dysregulated APA events have been investigated, such as loss or gain of individual poly(A) sites due to poly(A) signal (PAS) mutations and alterations in canonical cleavage and polyadenylation factors [51]. Recent studies aimed at discovering genetic variants affecting APA on the population level indicated a

key role for genetically controlled APA in human diseases [52]. Here we primarily discuss the emerging molecular mechanisms of APA regulations in human diseases.

Cis regulatory elements

Among these cis regulatory elements involved in RNA 3'-end processing, the AAUAAA hexamer is the core and canonical poly(A) signal recognized by the cleavage and polyadenylation specificity factor. Other non-canonical variants can also function as a PAS similar to AAUAAA but with a relative lower recognition efficiency [53]. In addition to hexamers, the upstream regulatory element UGUA motif and the downstream G/U-rich region can modulate the relative strength of the PAS in several specific conditions [4]. Perturbations of these cis-acting elements could lead to certain human diseases, especially variations in the PAS, which directly change PAS binding affinity of the cleavage and polyadenylation machinery [51]. In addition to altering the PAS, mutations can also affect the binding of miRNAs resulting in altered gene expression [54]. For example, mutants in the 3'-UTR of *ACTB* mRNA facilitate its interaction with miR-1 and miR-29a via AGO2, promoting hepatocellular carcinoma (HCC) cell migration and invasion [55]. Interestingly, mRNA 3'-UTR shortening can also play a key role in repressing tumor suppressor by disrupting competing endogenous RNAs (ceRNA) interactions in trans. Hyun *et al.* performed a model-based analysis of the trans effect of 3'-UTR shortening and predicted many trans-targets of 3'-UTR shortening, including PTEN, a tumor-suppressor gene involved in ceRNA crosstalk with other 3'-UTR shortening genes [56].

Trans factors

CFIm25

CFIm25 is encoded by the *Nudt21* gene and is an essential cleavage and polyadenylation factor that plays a crucial role in APA regulation. *CFIm25* knockdown can result in extensive 3'-UTR shortening of transcripts by increasing the proximal poly(A) site usage [49,57,58]. While 3'-UTR global shortening occurs in different cancers, CFIm25 is likely an essential factor in this process. In glioblastoma, *CFIm25* depletion causes the up-regulation of several known oncogenes, such as cyclin D1 and *Pak1* [49,57]. Also in HCC, knockdown of *CFIm25* promotes HCC cell proliferation and metastasis, in part by increasing the expression levels of *PSMB2* and *CXXC5* [58]. Recent findings showed that CFIm25 plays an essential role in bladder cancer progression through *ANXA2* and *LIMK2* by APA [59]. In addition to cancer [46], CFIm25 is also involved in other human diseases,

such as systemic sclerosis and idiopathic pulmonary fibrosis [60]. In both of these diseases, CFIm25 is down-regulated in key cell types present in the skin or lung promoting the 3'-UTR shortening of key TGF β -regulated fibrotic genes [61]. Finally, CFIm25 is critical to normal brain development as patients with copy number variation of *CFIm25* or mutations present significant intellectual disability [62,63].

PCF11

PCF11, a subunit of the CFIm complex, is involved in tumor progression, cell cycle regulation, cell proliferation, apoptosis, and neurodifferentiation [64]. Low expression of PCF11 in neuroblastoma is correlated with a favorable outcome and spontaneous tumor regression [64]. A recent study revealed that PCF11 can impact the expression of longer genes through regulating intronic polyadenylation [65].

MAGE-A11

The melanoma-associated antigen (MAGE) gene family is a large and conserved group of genes defined by a common MAGE homology domain [66]. One member of this family, MAGE-A11, was found to be aberrantly expressed in cancer [67,68]. By designing a computational approach from existing cancer “big-data”, Seung *et al.* recently reported that MAGE-A11-HUWE1 promotes APA and 3'-UTR shortening in cancer through ubiquitination of CFIm25 [69]. This finding provides new insights into the functions of MAGE genes regarding APA processing in cancer development.

HnRNPC

Heterogeneous nuclear ribonucleoprotein C (hnRNPC) is an RNA-binding protein that aberrantly up-regulated in multiple cancer types [70–73]. hnRNPC plays a critical role in regulating APA in metastatic colon cancer cells [74]. Mechanistically, hnRNPC regulates poly(A) site selection in a subset of genes implicated in cancer progression [74].

Long-distance APA regulation

In addition to cis regulatory elements and trans-acting factors, recent studies also demonstrated how functional elements outside of transcribed regions of genes can profoundly impact RNA processing within that gene. Nanavaty *et al.* elegantly demonstrates the DNA methylation patterns can have significant impact on APA via the Cohesin and CTCF factors [75]. In this instance, mutations/SNPs or changes in DNA methylation state

outside the transcribed region of genes indeed causes APA changes via disruption in gene looping. Xiong *et al.* demonstrate an unexpected finding that enhancers regulate alternative polyadenylation *in trans* [59]. Specifically, they reveal how enhancers can, independent of transcription output, specifically alter polyA site selection. In another study, Oktaba *et al.* demonstrate that in *Drosophila*, longer 3'-UTRs emanate from specific promoters active in neural tissue [76]. Their data reveal a connection between RNA Pol II pausing and loading of the ELAV RNA binding protein as a key factor to repress proximal polyA site usage thereby promoting long 3'-UTRs. Importantly, they show that swapping out these promoters causes significant changes in APA events.

GENETIC ARCHITECTURE OF APA VARIATIONS

Although recent population-scale RNA-seq data revealed associations between genetic variants and APA, characterization of genetic architecture remains a significant challenge. Mittleman *et al.* found many apaQTLs are intronic, and the genetic variants associated with increasing intronic poly(A) site usage tend to have lower gene expression levels [28]. In addition, Li *et al.* found that these genetic variants could alter poly(A) motifs and RNA-binding protein binding sites [35]. Another challenge is to identify the causal APA variants. Several innovative statistical fine-mapping algorithms have been proposed to identify these causal variants. CAVIAR is a method that models the association between the local linkage disequilibrium (LD) structure and effect sizes to quantify the posterior probability of causality for each variant [77]. The Sum of Single Effect (SuSiE) method operates on individual-level data to efficiently analyze loci with many independent effect variables [78]. SuSiE produces clusters of association signals, formally defined as 95% bayesian credible sets and each signal clusters are highly correlated due to LD.

In addition to the above, experimental approaches such as massive parallel reporter assays (MPRAs) [79] can test the effect of genetic variants on a tens-of-thousands scale for several target genes. These massive datasets can couple with machine learning approach to model APA patterns and predict the effects of genetic variants on APA. A recent study [80] generated mini gene libraries of over 3 million unique UTRs constructs. Each library varied in mRNA structure and 3'-UTR region. These constructs were then transiently transfected into HEK293 cells for high-throughput sequencing. The massive sequencing data were then trained using the deep learning method APARENT (APA REgression NeT) to predict the impact and putative casual mechanism of genetic variants on APA.

GENETIC CONTROL OF APA IS ASSOCIATED WITH COMPLEX HUMAN DISEASES AND TRAITS

Genetic variations impacting APA play essential roles in the dysregulation of RNA 3'-end processing and have been frequently associated with human diseases and phenotypic changes. In one common scenario, genetic variants can alter poly(A) site usage and thus influence the gene expression of key genes, leading to a series of different human complex diseases. For example, You-Jun *et al.* found an A-to-G mutation in the polyadenylation site of the $\alpha 2$ -hemoglobin gene *HBA2* (AAUAAA to AAUAAG) that causes α -thalassemia (hemoglobin H disease) by reducing the expression of the $\alpha 1$ -globin gene, resulting in a down-regulation of both the $\alpha 1$ -globin and $\alpha 2$ -globin genes [81,82]. In another thalassemia study, a T-to-C substitution within the 3'-end conserved sequence (AAUAAA to AACAAA) was shown to disrupt polyadenylation signals for APA, which results in an at least 900-bp extension of the human β -globin transcription and the subsequent β -thalassemia [83,84]. In addition, an A-to-C mutation in the canonical PAS within *TP53* (AAUAAA to AAUACT) was associated with the impaired 3'-end processing of *TP53* transcripts and increased susceptibility to multiple cancers, including cutaneous basal cell carcinoma, prostate cancer, glioma, and colorectal adenoma [85]. Bennett *et al.* reported a rare A-to-G PAS mutation of the *FOXP3* gene (AAUAAA to AAUGAA), which results in decreased expression of *FOXP3* due to degradation of the transcripts of *FOXP3*, leading to the immunodysregulation polyendocrinopathy enteropathy X-linked syndrome [86]. Besides these variations within poly(A) sites, a few studies have demonstrated that variants located within other cis regulatory elements are also associated with disease risk. For example, a single-nucleotide change within the GU-rich downstream element of *FGG* gene can lead to increased distal poly(A) site usage. These differential APA usages are strongly associated with disease risk of deep vein thrombosis [87].

SUMMARY AND OUTLOOK

Traditional RNA-seq reads cannot provide connectivity for most mRNA transcripts, but with the advance of long-read Oxford nanopore sequencing and PacBio isoform sequencing, we can rebuild the coordinate regulation of transcription initiation, alternative splicing, and APA. A recent study using MinION nanopore sequencing found the tight coordination between the *Dscam1* long 3'-UTR and skipping of exon 19 in neurons. This co-regulation generates a specific isoform essential for neural develop-

ment [88]. In another study, Anvar *et al.* also revealed an interdependent relationship in cultured human MCF-7 breast cancer cells [89]. We expect more coordination of the landscape of APA with other transcript elements with the expansion of long-read sequencing in coming years. In addition, current studies are primarily restricted to the LCL. LCLs is well-studied cell type for investigating the genetic influence of APA and have been widely employed in large consortia studies, including the HapMap and 1000 Genome projects. Expansion to multiple other tissues, as in the case of the GTEx Project, or other cell types, as in the case of DICE [90], will provide further insights into interpretation of GWAS non-coding variants.

Moreover, single-cell sequencing coupled with genetic information has opened a new era of population genetics [91]. A recent study used single-cell RNA sequencing to sequence ~25,000 peripheral blood mononuclear cells from 45 donors and detect many cell type-specific expression quantitative trait loci (eQTLs) and gene-gene interaction network [92]. More recently, several innovative computational tools such as scDaPars [93] can quantify and recover APA usage at single-cell and single-gene resolution. Thus, we expect more population-scale studies with these computational approaches to precisely define the cellular contexts in which GWAS variants affect polyadenylation usage from these emerging single-cell data. This will help better understand the molecular mechanisms by which GWAS variant is conferred and therapeutic design strategies.

APA is also a conserved phenomenon across species, and our recent work indicated that genetic influences on APA are mostly in a tissue-specific manner [36]. It would be interesting to investigate the tissue-dependent genetic influence of APA across species. It has been known that species type, rather than tissue type, is the primary determinant of the splicing patterns [94] and RNA-editing [95]. However, it is currently unclear the evolutionary forces of tissue-dependent APA patterns. These studies could provide important insight into understanding the relationship between evolutionary patterns of APA and variation of phenotype across species.

ACKNOWLEDGEMENTS

The authors apologize to those colleagues whose relevant studies were not cited due to spacing limits. We thank members of the Wei Li lab for helpful discussions. The Wagner lab acknowledges support from the National Institutes of Health grant R01-GM134539 (E.J.W).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Lei Li, Yumei Li, Xudong Zou, Fuduan Peng, Ya Cui, Eric J. Wagner and Wei Li declare no competing financial interests.

All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted,

and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Mayr, C. (2017) Regulation by 3'-untranslated regions. *Annu. Rev. Genet.*, 51, 171–194
- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J. Y., Yehia, G. and Tian, B. (2013) Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods*, 10, 133–139
- Derti, A., Garrett-Engele, P., Macisaac, K. D., Stevens, R. C., Sriram, S., Chen, R., Rohl, C. A., Johnson, J. M. and Babak, T. (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res.*, 22, 1173–1183
- Tian, B. and Manley, J. L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, 18, 18–30
- Mayr, C. (2019) What Are 3' UTRs Doing? *Cold Spring Harb. Perspect. Biol.*, 11, a034728
- Berkovits, B. D. and Mayr, C. (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, 522, 363–367
- Chao, Y., Li, L., Girodat, D., Förstner, K. U., Said, N., Corcoran, C., Šmiga, M., Papenfort, K., Reinhardt, R., Wieden, H. J., *et al.* (2017) *In vivo* cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol. Cell*, 65, 39–51
- Holmqvist, E., Li, L., Bischler, T., Barquist, L. and Vogel, J. (2018) Global maps of ProQ binding *in vivo* reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol. Cell*, 70, 971–982.e6
- Mercer, T. R., Wilhelm, D., Dinger, M. E., Soldà, G., Korbie, D. J., Glazov, E. A., Truong, V., Schwenke, M., Simons, C., Matthaei, K. I., *et al.* (2011) Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res.*, 39, 2393–2403
- Levy, S. and Schuster, G. (2016) Polyadenylation and degradation of RNA in the mitochondria. *Biochem. Soc. Trans.*, 44, 1475–1482
- Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J. and Li, W. (2014) Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.*, 5, 5274
- Chun, S., Casparino, A., Patsopoulos, N. A., Croteau-Chonka, D. C., Raby, B. A., De Jager, P. L., Sunyaev, S. R. and Cotsapas, C. (2017) Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.*, 49, 600–605
- Manning, K. S. and Cooper, T. A. (2017) The roles of RNA processing in translating genotype to phenotype. *Nat. Rev. Mol. Cell Biol.*, 18, 102–114
- Yang, H. and Melera, P. W. (1994) A genetic polymorphism within the third poly(A) signal of the DHFR gene alters the polyadenylation pattern of DHFR transcripts in CHL cells. *Nucleic Acids Res.*, 22, 2694–2702
- Bell, D. A., Badawi, A. F., Lang, N. P., Ilett, K. F., Kadlubar, F. F. and Hirvonen, A. (1995) Polymorphism in the N-acetyltransferase 1 (NAT1) polyadenylation signal: association of NAT1*10 allele with higher N-acetylation activity in bladder and colon tissue. *Cancer Res*, 55, 5226–5229
- Battersby, S., Ogilvie, A. D., Blackwood, D. H., Shen, S., Muqit, M. M., Muir, W. J., Teague, P., Goodwin, G. M. and Harmar, A. J. (1999) Presence of multiple functional polyadenylation signals and a single nucleotide polymorphism in the 3' untranslated region of the human serotonin transporter gene. *J. Neurochem.*, 72, 1384–1388
- Hoarau, J. J., Cesari, M., Caillens, H., Cadet, F. and Pabion, M. (2004) HLA *DQA1* genes generate multiple transcripts by alternative splicing and polyadenylation of the 3' untranslated region. *Tissue Antigens*, 63, 58–71
- Graham, D. S. C., Manku, H., Wagner, S., Reid, J., Timms, K., Gutin, A., Lanchbury, J. S. and Vyse, T. J. (2007) Association of IRF5 in UK SLE families identifies a variant involved in polyadenylation. *Hum. Mol. Genet.*, 16, 579–591
- Graham, R. R., Kyogoku, C., Sigurdsson, S., Vlasova, I. A., Davies, L. R., Baechler, E. C., Plenge, R. M., Koeuth, T., Ortmann, W. A., Hom, G., *et al.* (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci. USA*, 104, 6758–6763
- Hellquist, A., Zucchelli, M., Kivinen, K., Saarialho-Kere, U., Koskenmies, S., Widen, E., Julkunen, H., Wong, A., Karjalainen-Lindsberg, M. L., Skoog, T., *et al.* (2007) The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. *J. Med. Genet.*, 44, 314–321
- Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T. J., Sladek, R. and Majewski, J. (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.*, 40, 225–231
- Yang, Z. and Kaye, D. M. (2009) Mechanistic insights into the link between a polymorphism of the 3'UTR of the *SLC7A1* gene and hypertension. *Hum. Mutat.*, 30, 328–333
- Yoon, O. K., Hsu, T. Y., Im, J. H. and Brem, R. B. (2012) Genetics and regulatory impact of alternative polyadenylation in human B-lymphoblastoid cells. *PLoS Genet.*, 8, e1002882
- Hartley, C. A., McKenna, M. C., Salman, R., Holmes, A., Casey, B. J., Phelps, E. A. and Glatt, C. E. (2012) Serotonin transporter polyadenylation polymorphism modulates the retention of fear

- extinction memory. *Proc. Natl. Acad. Sci. USA*, 109, 5493–5498
25. Zhernakova, D. V., de Klerk, E., Westra, H. J., Mastrokolias, A., Amini, S., Ariyurek, Y., Jansen, R., Penninx, B. W., Hottenga, J. J., Willemsen, G., *et al.* (2013) DeepSAGE reveals genetic variants associated with alternative polyadenylation and expression of coding and non-coding transcripts. *PLoS Genet.*, 9, e1003594
 26. Prasad, M. K., Bhalla, K., Pan, Z. H., O'Connell, J. R., Weder, A. B., Chakravarti, A., Tian, B. and Chang, Y. P. (2013) A polymorphic 3'UTR element in ATP1B1 regulates alternative polyadenylation and is associated with blood pressure. *PLoS One*, 8, e76290
 27. Fraser, H. B. and Xie, X. (2009) Common polymorphic transcript variation in human disease. *Genome Res.*, 19, 567–575
 28. Mittleman, B. E., Pott, S., Warland, S., Zeng, T., Mu, Z., Kaur, M., Gilad, Y. and Li, Y. (2020) Alternative polyadenylation mediates genetic regulation of gene expression. *eLife*, 9, e57492
 29. Cannavò, E., Koelling, N., Harnett, D., Garfield, D., Casale, F. P., Ciglar, L., Gustafson, H. E., Viales, R. R., Marco-Ferreres, R., Degner, J. F., *et al.* (2017) Genetic variants regulating expression levels and isoform diversity during embryogenesis. *Nature*, 541, 402–406
 30. Yalamanchili, H. K., Alcott, C. E., Ji, P., Wagner, E. J., Zoghbi, H. Y. and Liu, Z. (2020) PolyA-miner: accurate assessment of differential alternative poly-adenylation from 3' Seq data using vector projections and non-negative matrix factorization. *Nucleic Acids Res.*, 48, e69
 31. Routh, A. (2019) DPAC: A tool for differential poly(A)-cluster usage from poly(A)-targeted RNAseq data. *G3 (Bethesda)*, 9, 1825–1830
 32. Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A., Monlong, J., Rivas, M. A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P. G., *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501, 506–511
 33. Monlong, J., Calvo, M., Ferreira, P. G. and Guigó, R. (2014) Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat. Commun.*, 5, 4698
 34. Mariella, E., Marotta, F., Grassi, E., Gilotto, S. and Provero, P. (2019) The Length of the Expressed 3' UTR Is an intermediate molecular phenotype linking genetic variants to complex diseases. *Front. Genet.*, 10, 714
 35. The GTEx Consortium, the Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, the Statistical Methods groups—Analysis Working Group, the Enhancing GTEx (eGTEx) groups, the NIH Common Fund, the NIH/NCI, the NIH/NHGRI, the NIH/NIMH, the NIH/NIDA, the Biospecimen Collection Source Site—NDRI, *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, 550, 204–213
 36. Li, L., Huang, K.L., Gao, Y., Cui, Y., Wang, G., Elrod, N.D., Li, Y., Chen, Y.E., Ji, P., Peng, F. *et al.* (2021) An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.*, doi: 10.1038/s41588-021-00864-5
 37. Ha, K. C. H., Blencowe, B. J. and Morris, Q. (2018) QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.*, 19, 45
 38. Grassi, E., Mariella, E., Lembo, A., Molineris, I. and Provero, P. (2016) Roar: detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinformatics*, 17, 423
 39. Wang, R., Nambiar, R., Zheng, D. and Tian, B. (2018) PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.*, 46, D315–D319
 40. Gruber, A. J., Schmidt, R., Gruber, A. R., Martin, G., Ghosh, S., Belmadani, M., Keller, W. and Zavolan, M. (2016) A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.*, 26, 1145–1159
 41. You, L., Wu, J., Feng, Y., Fu, Y., Guo, Y., Long, L., Zhang, H., Luan, Y., Tian, P., Chen, L., *et al.* (2015) APASdb: a database describing alternative poly(A) sites and selection of heterogeneous cleavage sites downstream of poly(A) signals. *Nucleic Acids Res.*, 43, D59–D67
 42. Kim, M., You, B. H. and Nam, J. W. (2015) Global estimation of the 3' untranslated region landscape using RNA sequencing. *Methods*, 83, 111–117
 43. Arefeen, A., Liu, J., Xiao, X. and Jiang, T. (2018) TAPAS: tool for alternative polyadenylation site analysis. *Bioinformatics*, 34, 2521–2529
 44. Ye, C., Long, Y., Ji, G., Li, Q. Q. and Wu, X. (2018) APAtrop: identification and quantification of alternative polyadenylation sites from RNA-seq data. *Bioinformatics*, 34, 1841–1849
 45. Feng, X., Li, L., Wagner, E. J. and Li, W. (2018) TC3A: The Cancer 3' UTR Atlas. *Nucleic Acids Res.*, 46, D1027–D1030
 46. Yuan, F., Hankey, W., Wagner, E. J., Li, W. and Wang, Q. (2019) Alternative polyadenylation of mRNA and its role in cancer. *Genes Dis.*, 8, 61–72
 47. Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A. and Moreira, A. (2014) Implications of polyadenylation in health and disease. *Nucleus*, 5, 508–519
 48. Chang, J. W., Yeh, H. S. and Yong, J. (2017) Alternative polyadenylation in human diseases. *Endocrinol. Metab. (Seoul)*, 32, 413–421
 49. Masamha, C. P., Xia, Z., Yang, J., Albrecht, T. R., Li, M., Shyu, A. B., Li, W. and Wagner, E. J. (2014) CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*, 510, 412–416
 50. Mayr, C. and Bartel, D. P. (2009) Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138, 673–684
 51. Gruber, A. J. and Zavolan, M. (2019) Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.*, 20, 599–614
 52. Mariella, E., Marotta, F., Grassi, E., Gilotto, S. and Provero, P. (2019) The length of the expressed 3' UTR is an intermediate molecular phenotype linking genetic variants to complex diseases. *Front. Genet.*, 10, 714

53. Sanfilippo, P., Wen, J. and Lai, E. C. (2017) Landscape and evolution of tissue-specific alternative polyadenylation across *Drosophila* species. *Genome Biol.*, 18, 229
54. Mayr, C. and Bartel, D. P. (2009) Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138, 673–684
55. Li, Y., Ma, H., Shi, C., Feng, F. and Yang, L. (2020) Mutant ACTB mRNA 3'-UTR promotes hepatocellular carcinoma development by regulating miR-1 and miR-29a. *Cell. Signal.*, 67, 109479
56. Park, H. J., Ji, P., Kim, S., Xia, Z., Rodriguez, B., Li, L., Su, J., Chen, K., Masamha, C. P., Baillat, D., *et al.* (2018) 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. *Nat. Genet.*, 50, 783–789
57. Chu, Y., Elrod, N., Wang, C., Li, L., Chen, T., Routh, A., Xia, Z., Li, W., Wagner, E. J. and Ji, P. (2019) Nudt21 regulates the alternative polyadenylation of Pak1 and is predictive in the prognosis of glioblastoma patients. *Oncogene*, 38, 4154–4168
58. Tan, S., Li, H., Zhang, W., Shao, Y., Liu, Y., Guan, H., Wu, J., Kang, Y., Zhao, J., Yu, Q., *et al.* (2018) NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. *Oncogene*, 37, 4887–4900
59. Xiong, M., Chen, L., Zhou, L., Ding, Y., Kazobinka, G., Chen, Z. and Hou, T. (2019) NUDT21 inhibits bladder cancer progression through ANXA2 and LIMK2 by alternative polyadenylation. *Theranostics*, 9, 7156–7167
60. Weng, T., Ko, J., Masamha, C. P., Xia, Z., Xiang, Y., Chen, N. Y., Molina, J. G., Collum, S., Mertens, T. C., Luo, F., *et al.* (2019) Cleavage factor 25 deregulation contributes to pulmonary fibrosis through alternative polyadenylation. *J. Clin. Invest.*, 129, 1984–1999
61. Weng, T., Huang, J., Wagner, E. J., Ko, J., Wu, M., Wareing, N. E., Xiang, Y., Chen, N. Y., Ji, P., Molina, J. G., *et al.* (2020) Downregulation of CFIm25 amplifies dermal fibrosis through alternative polyadenylation. *J. Exp. Med.*, 217, e20181384
62. Gennarino, V. A., Alcott, C. E., Chen, C. A., Chaudhury, A., Gillentine, M. A., Rosenfeld, J. A., Parikh, S., Wheless, J. W., Roeder, E. R., Horovitz, D. D., *et al.* (2015) NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *eLife*, 4, e10782
63. Alcott, C. E., Yalamanchili, H. K., Ji, P., van der Heijden, M. E., Saltzman, A., Elrod, N., Lin, A., Leng, M., Bhatt, B., Hao, S., *et al.* (2020) Partial loss of CFIm25 causes learning deficits and aberrant neuronal alternative polyadenylation. *eLife*, 9, e50895
64. Ogorodnikov, A., Levin, M., Tattikota, S., Tokalov, S., Hoque, M., Scherzinger, D., Marini, F., Poetsch, A., Binder, H., Macher-Göppinger, S., *et al.* (2018) Transcriptome 3'end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat. Commun.*, 9, 5331
65. Wang, R., Zheng, D., Wei, L., Ding, Q. and Tian, B. (2019) Regulation of intronic polyadenylation by PCF11 impacts mRNA expression of long genes. *Cell Rep.*, 26, 2766–2778.e6
66. Lee, A. K. and Potts, P. R. (2017) A comprehensive guide to the MAGE family of ubiquitin ligases. *J. Mol. Biol.*, 429, 1114–1142
67. Minges, J. T., Su, S., Grossman, G., Blackwelder, A. J., Pop, E. A., Mohler, J. L. and Wilson, E. M. (2013) Melanoma antigen-A11 (MAGE-A11) enhances transcriptional activity by linking androgen receptor dimers. *J. Biol. Chem.*, 288, 1939–1952
68. Xia, L. P., Xu, M., Chen, Y. and Shao, W. W. (2013) Expression of MAGE-A11 in breast cancer tissues and its effects on the proliferation of breast cancer cells. *Mol. Med. Rep.*, 7, 254–258
69. Yang, S. W., Li, L., Connelly, J. P., Porter, S. N., Kodali, K., Gan, H., Park, J. M., Tacer, K. F., Tillman, H., Peng, J., *et al.* (2020) A cancer-specific ubiquitin ligase drives mRNA alternative polyadenylation by ubiquitinating the mRNA 3' end processing complex. *Mol. Cell*, 77, 1206–1221.e7
70. Park, Y. M., Hwang, S. J., Masuda, K., Choi, K. M., Jeong, M. R., Nam, D. H., Gorospe, M. and Kim, H. H. (2012) Heterogeneous nuclear ribonucleoprotein C1/C2 controls the metastatic potential of glioblastoma by regulating PDCD4. *Mol. Cell. Biol.*, 32, 4237–4244
71. Pino, I., Pío, R., Toledo, G., Zabalegui, N., Vicent, S., Rey, N., Lozano, M. D., Torre, W., García-Foncillas, J. and Montuenga, L. M. (2003) Altered patterns of expression of members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family in lung cancer. *Lung Cancer*, 41, 131–143
72. Mulnix, R. E., Pitman, R. T., Retzer, A., Bertram, C., Arasi, K., Crees, Z., Girard, J., Uppada, S. B., Stone, A. L. and Puri, N. (2013) hnRNP C1/C2 and Pur-beta proteins mediate induction of senescence by oligonucleotides homologous to the telomere overhang. *Onco Targets Ther*, 7, 23–32
73. Wu, Y., Zhao, W., Liu, Y., Tan, X., Li, X., Zou, Q., Xiao, Z., Xu, H., Wang, Y. and Yang, X. (2018) Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response. *EMBO J.*, 37, e99017
74. Fischl, H., Neve, J., Wang, Z., Patel, R., Louey, A., Tian, B. and Furger, A. (2019) hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Res.*, 47, 7580–7591
75. Nanavaty, V., Abrash, E. W., Hong, C., Park, S., Fink, E. E., Li, Z., Sweet, T. J., Bhasin, J. M., Singuri, S., Lee, B. H., *et al.* (2020) DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol. Cell*, 78, 752–764.e6
76. Oktaba, K., Zhang, W., Lotz, T. S., Jun, D. J., Lemke, S. B., Ng, S. P., Esposito, E., Levine, M. and Hilgers, V. (2015) ELAV links paused Pol II to alternative polyadenylation in the *Drosophila* nervous system. *Mol. Cell*, 57, 341–348
77. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. and Eskin, E. (2014) Identifying causal variants at loci with multiple signals of association. *Genetics*, 198, 497–508
78. Wang, G., Sarkar, A., Carbonetto, P. and Stephens, M. (2020) A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.*, 82, 1273–1300
79. Starita, L. M., Ahituv, N., Dunham, M. J., Kitzman, J. O., Roth, F. P., Seelig, G., Shendure, J. and Fowler, D. M. (2017) Variant interpretation: functional assays to the rescue. *Am. J. Hum. Genet.*,

- 101, 315–325
80. Bogard, N., Linder, J., Rosenberg, A. B. and Seelig, G. (2019) A deep neural network for predicting and engineering alternative polyadenylation. *Cell*, 178, 91–106.e23
 81. Higgs, D. R., Goodbourn, S. E., Lamb, J., Clegg, J. B., Weatherall, D. J. and Proudfoot, N. J. (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature*, 306, 398–400
 82. Fei, Y. J., Oner, R., Bözkurt, G., Gu, L. H., Altay, C., Gurgey, A., Fattoum, S., Baysal, E. and Huisman, T. H. (1992) Hb H disease caused by a homozygosity for the AATAAA → AATAAG mutation in the polyadenylation site of the alpha 2-globin gene: hematological observations. *Acta Haematol.*, 88, 82–85
 83. Orkin, S. H., Cheng, T. C., Antonarakis, S. E. and Kazazian, H. H. Jr. (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J.*, 4, 453–456
 84. Rund, D., Dowling, C., Najjar, K., Rachmilewitz, E. A., Kazazian, H. H. Jr and Oppenheim, A. (1992) Two mutations in the beta-globin polyadenylation signal reveal extended transcripts and new RNA polyadenylation sites. *Proc. Natl. Acad. Sci. U.S.A.*, 89, 4324–4328
 85. Stacey, S. N., Sulem, P., Jonasdottir, A., Masson, G., Gudmundsson, J., Gudbjartsson, D. F., Magnusson, O. T., Gudjonsson, S. A., Sigurgeirsson, B., Thorisdottir, K., *et al.* (2011) A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.*, 43, 1098–1103
 86. Bennett, C. L., Brunkow, M. E., Ramsdell, F., O'Briant, K. C., Zhu, Q., Fuleihan, R. L., Shigeoka, A. O., Ochs, H. D. and Chance, P. F. (2001) A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA → AAUGAA) leads to the IPEX syndrome. *Immunogenetics*, 53, 435–439
 87. Uitte de Willige, S., Rietveld, I. M., De Visser, M. C. H., Vos, H. L. and Bertina, R. M. (2007) Polymorphism 10034C > T is located in a region regulating polyadenylation of FGG transcripts and influences the fibrinogen $\gamma/\gamma A$ mRNA ratio. *J. Thromb. Haemost.*, 5, 1243–1249
 88. Zhang, Z., So, K., Peterson, R., Bauer, M., Ng, H., Zhang, Y., Kim, J. H., Kidd, T. and Miura, P. (2019) Elav-mediated exon skipping and alternative polyadenylation of the *dscaml1* gene are required for axon outgrowth. *Cell Rep.*, 27, 3808–3817.e7
 89. Anvar, S. Y., Allard, G., Tseng, E., Sheynkman, G. M., de Klerk, E., Vermaat, M., Yin, R. H., Johansson, H. E., Ariyurek, Y., den Dunnen, J. T., *et al.* (2018) Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol.*, 19, 46
 90. Schmiedel, B. J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A. G., White, B. M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J. A., McVicker, G. *et al.* (2018) Impact of genetic polymorphisms on human immune cell gene expression. *Cell*, 175, 1701–1715
 91. van der Wijst, M., de Vries, D. H., Groot, H. E., Trynka, G., Hon, C. C., Bonder, M. J., Stegle, O., Nawijn, M. C., Idaghdour, Y., van der Harst, P., *et al.* (2020) The single-cell eQTLGen consortium. *eLife*, 9, e52155
 92. van der Wijst, M. G. P., Brugge, H., de Vries, D. H., Deelen, P., Swertz, M. A. and Franke, L., and the LifeLines Cohort Study, and the BIOS Consortium. (2018) Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.*, 50, 493–497
 93. Gao, Y., Li, L., Amos, C. I. and Li, W. (2020) Dynamic analysis of alternative polyadenylation from single-cell RNA-seq (scDaPars) reveals cell subpopulations invisible to gene expression analysis. *bioRxiv*, 310649
 94. Merkin, J., Russell, C., Chen, P. and Burge, C. B. (2012) Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science*, 338, 1593–1599
 95. Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., Liu, K. I., Zhang, R., Ramaswami, G., Ariyoshi, K., *et al.* (2017) Dynamic landscape and regulation of RNA editing in mammals. *Nature*, 550, 249–254