



REVIEW

# Advances in analyzing RNA diversity in eukaryotic transcriptomes: peering through the Omics lens [version 1; referees: 3 approved]

Sushant Bangru<sup>1</sup>, Auinash Kalsotra  <sup>1-3</sup>

<sup>1</sup>Department of Biochemistry, University of Illinois at Urbana-Champaign, Illinois, USA

<sup>2</sup>Institute of Genomic Biology, University of Illinois at Urbana-Champaign, Illinois, USA

<sup>3</sup>College of Medicine, University of Illinois at Urbana-Champaign, Illinois, USA

**v1** First published: 14 Nov 2016, 5(F1000 Faculty Rev):2668 (doi: [10.12688/f1000research.9511.1](https://doi.org/10.12688/f1000research.9511.1))

Latest published: 14 Nov 2016, 5(F1000 Faculty Rev):2668 (doi: [10.12688/f1000research.9511.1](https://doi.org/10.12688/f1000research.9511.1))

**Abstract**

Alternative splicing, polyadenylation, and chemical modifications of RNA generate astonishing complexity within eukaryotic transcriptomes. The last decade has brought numerous advances in sequencing technologies that allow biologists to investigate these phenomena with greater depth and accuracy while reducing time and cost. A commensurate development in biochemical techniques for the enrichment and analysis of different RNA variants has accompanied the advancement of global sequencing analysis platforms. Here, we present a detailed overview of the latest biochemical methods, along with bioinformatics pipelines that have aided in identifying different RNA variants. We also highlight the ongoing developments and challenges associated with RNA variant detection and quantification, including sample heterogeneity and isolation, as well as ‘Omics’ big data handling.

**Open Peer Review**

Referee Status: 

	Invited Referees		
	1	2	3
version 1 published 14 Nov 2016			

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty. In order to make these reviews as comprehensive and accessible as possible, peer review takes place before publication; the referees are listed below, but their reports are not formally published.

- 1 **Graziano Pesole**, Università degli Studi di Bari Italy
- 2 **Emanuele Buratti**, International Center for Genetic Engineering and Biotechnology, University of Trieste Italy
- 3 **Gordon Carmichael**, University of Connecticut Health Center USA

**Discuss this article**

Comments (0)

**Corresponding author:** Auinash Kalsotra ([kalsotra@illinois.edu](mailto:kalsotra@illinois.edu))

**How to cite this article:** Bangru S and Kalsotra A. **Advances in analyzing RNA diversity in eukaryotic transcriptomes: peering through the Omics lens [version 1; referees: 3 approved]** *F1000Research* 2016, 5(F1000 Faculty Rev):2668 (doi: [10.12688/f1000research.9511.1](https://doi.org/10.12688/f1000research.9511.1))

**Copyright:** © 2016 Bangru S and Kalsotra A. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** AK is supported by grants from the US National Institutes of Health (R01HL126845), the March of Dimes (5-FY14-112), and the Center for Advanced Study at the University of Illinois.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Competing interests:** The authors declare that they have no competing interests.

**First published:** 14 Nov 2016, 5(F1000 Faculty Rev):2668 (doi: [10.12688/f1000research.9511.1](https://doi.org/10.12688/f1000research.9511.1))

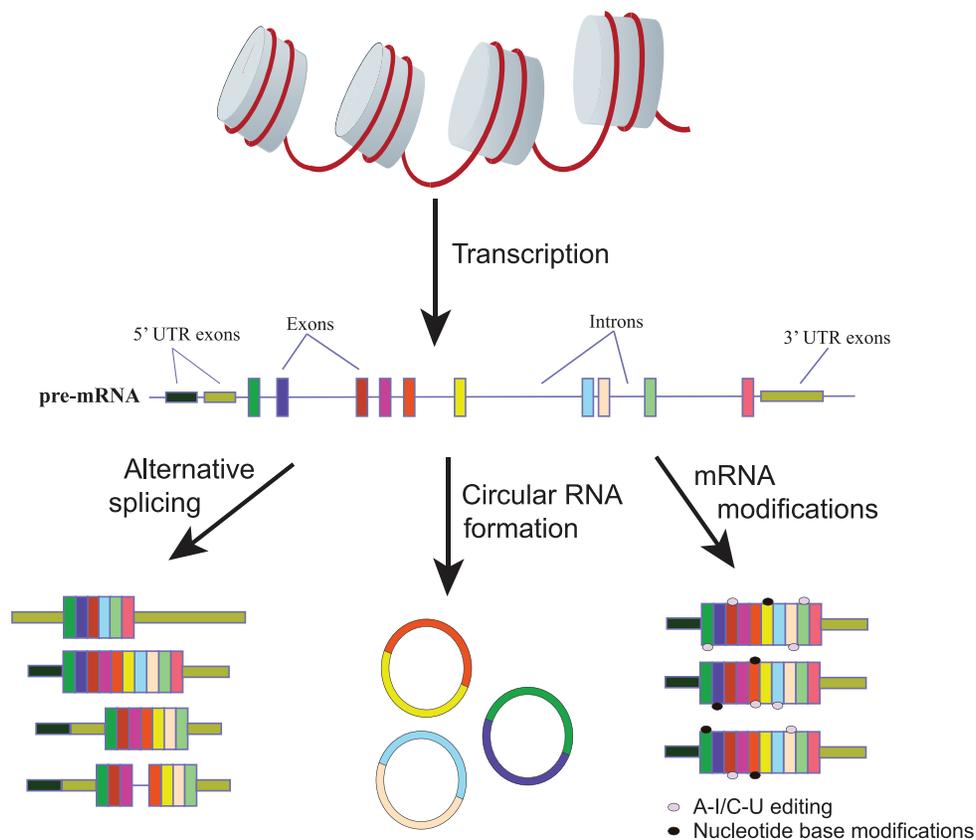
## How did we get here?

The RNA world hypothesis, a widely prevailing idea among molecular biologists, describes self-replicating RNA as the precursor to all modern life forms. Besides recognizing its primordial origin, we have come to realize that RNA is a highly complex and diverse macromolecule, which plays central roles in protein synthesis by serving as the encoder (mRNA), the decoder (tRNA), and the catalyst (rRNA). But our understanding of RNA diversity, form, and function continues to evolve as we keep discovering new and exciting classes of RNAs that carry out unexpected functions. Recent transcriptome studies, which capture nearly every transcript in a cell, have revealed an overwhelming number of RNA variants resulting from multiple transcription initiation sites, alternative pre-mRNA splicing and polyadenylation, post-transcriptional editing, and direct chemical modifications of RNA (Figure 1). These studies have demonstrated that the variants play key roles in generating mRNA diversity by altering their coding sequences, half-lives, and translation efficiencies, allowing complex organisms to control temporal and tissue-specific transcriptome patterns.

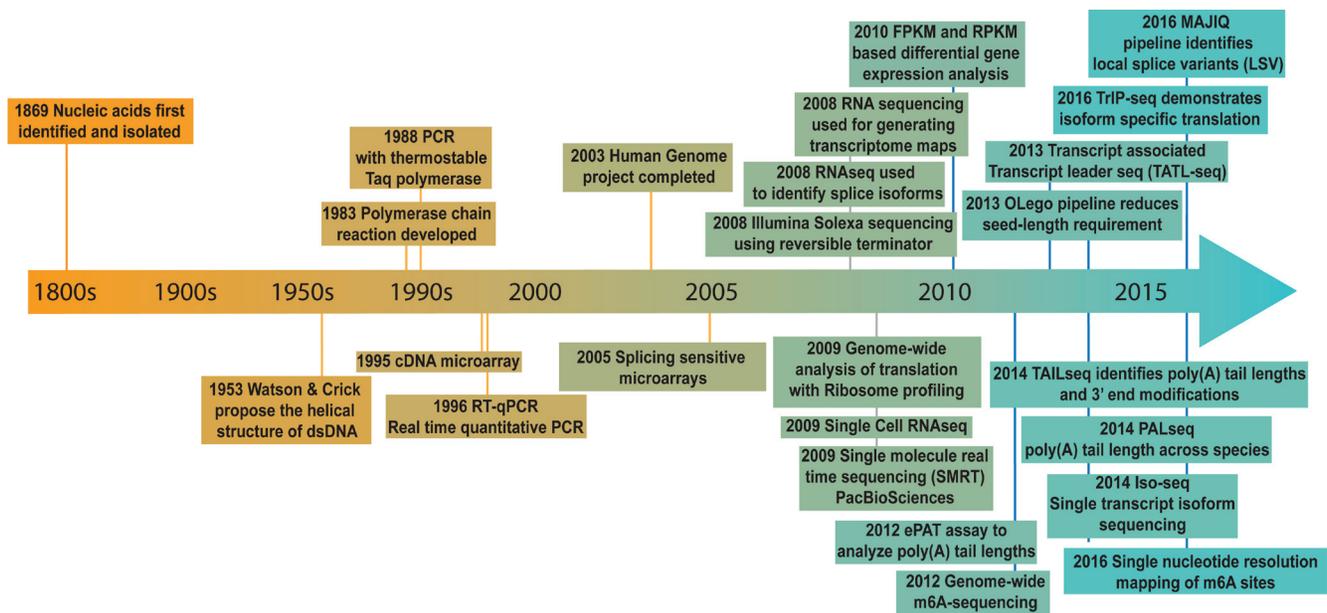
The systematic study of RNA variants began in the 1990s when cloning of expressed sequence tags revealed the diversity of mRNA

populations expressed in different cells and tissues of metazoans<sup>1</sup>. The advent of microarrays in the early 2000s expedited the discovery of new RNA variants<sup>2,3</sup>, but the technology had several limitations, including (i) high background levels due to dependence on hybridization, (ii) reliance on existing sequence knowledge, and (iii) a limited range of detection. In the last decade, the field has taken a significant leap forward with the development of next-generation sequencing. The timeline of major milestones that facilitated the discovery of novel RNA variants is provided in Figure 2.

First-generation sequencing required separate reactions for individual nucleotides and was limited by a gel electrophoresis step, which was further complicated by the need to use radioactive reagents<sup>4,5</sup>. Next-generation sequencing, however, uses advanced optics and flow cells to identify nucleotide sequences rapidly, making the analysis and sequence determination quicker and more accurate<sup>6-8</sup>. A number of next-generation sequencing platforms have been developed, and the most popular ones are Illumina Solexa<sup>9</sup> (relies on fluorescence imaging post-nucleotide addition every cycle, approximately 100 nucleotide reads, 0.1% error rate), Roche 454 sequencing<sup>10</sup> (long reads and short homopolymeric stretches sequenced in single-cycle, approximately 500 nucleotide reads, 1%



**Figure 1. Alternative splicing, circular RNA formation, and mRNA modifications are the major mechanisms generating much of the mRNA diversity.** Alternative splicing results in the selective inclusion or skipping of particular exons as well as alternate 5' and 3' untranslated region (UTR) selection within a processed mRNA transcript. In some cases, alternative splicing also results in intron retention events. Circular RNA formation occurs because of back splicing of the 3' end of a downstream exon to the 5' end of an upstream exon. Lastly, mRNAs can be modified through editing mechanisms converting A and C nucleotides to I and U nucleotides, respectively, as well as through the addition of functional groups on either the base or the sugar moiety of a nucleotide.



**Figure 2. Timeline outlining the recent advances in analyzing RNA variants.**

error rate), Ion Torrent<sup>11</sup> (based on detection of H<sup>+</sup> ion release due to polymerization of dNTP probes, approximately 200 nucleotide reads, 1% error rate), and SOLiD sequencing<sup>12</sup> (relies on sequencing by ligation over synthesis, 35 nucleotide reads, less than 0.1% error rate). Of the above methods, the first three are classified as sequencing by synthesis methods whereas SOLiD is under the sequencing-by-ligation category. The field of sequencing by ligation has made rapid strides with the development of complete genomics technology that uses combinatorial synthesis methods, involving the use of multiple probe-anchor combinations to determine sequence<sup>13</sup>. Additionally, Pacific Biosciences has recently introduced single-molecule real-time sequencing<sup>14</sup> of full-length transcripts, and the average read length is 2.5 kb (non-stop sequencing and real-time imaging, more than 1% error rates).

Perhaps the most significant breakthrough in the field was the development of massively parallel cDNA sequencing, commonly referred to as RNA-seq<sup>15</sup>. It involves sequencing of cDNA fragments generated from total RNA after oligo-dT or random priming. For most transcriptome studies, mRNA is enriched from the total RNA pool by using either a poly(A) pull-down or selective depletion of rRNA. Following sequencing, the reads are aligned to a pre-annotated reference genome or assembled *de novo* to provide both the qualitative and the quantitative information of entire transcriptomes. This technique now can be used to identify new transcripts without having *a priori* knowledge of all gene models<sup>16</sup>. Recent mapping tools can also identify splice junctions by using reads spanning across two exons, bolstering identification of alternative splice isoforms. Initially, RNA sequencing (RNA-seq) studies examined only the steady-state levels of RNA populations; however, the relative contribution of variable transcription (synthesis) and RNA degradation (stability) in those studies was unclear. New techniques such as Net-Seq<sup>17,18</sup> and BRU-seq<sup>19</sup> address this problem.

Net-Seq employs DNA-RNA-RNA polymerase complex stability to isolate 3' ends of nascent transcripts. Bru-Seq uses pulse-chase labeling of newly transcribed RNA with bromouridine to determine rates of RNA synthesis and stability.

The past five years have witnessed a surge of new techniques employing RNA-seq to identify many new RNA variants, and isolation/enrichment methods are based on each variant type. In parallel, significant advances in computing power have expanded our ability to annotate and quantify these variants with high confidence from a single cell to complex tissues, along with the advent of consortium projects that have given rise to databases for their comparative and meta-analyses<sup>20-22</sup>. The increasing sensitivity of detection, wherein extremely minute quantities (picograms) of RNAs can be amplified and sequenced, has also spurred development of single-cell RNA-seq (scRNAseq)<sup>23</sup> for identifying heterogeneity within cell populations and uncovering transcriptome differences that may be hidden in bulk RNA-seq<sup>24</sup>. Some of the commonly used sequencing technologies and bioinformatics packages for the detection of different RNA variants are provided in [Table 1](#).

In this review, we discuss different types of mRNA variants commonly found in metazoans, consider the latest advances in techniques to identify them, and provide a brief overview of the recent developments in bioinformatics pipelines for their quantitative analyses.

### Decoding the splice variants

Pre-mRNA splicing is an essential step for the expression of most genes in metazoans. Splice sites within a pre-mRNA facilitate the splicing and ligation of introns and exons, respectively, producing a mature mRNA transcript. Alternative splicing involves differential usage of splice sites to yield multiple mRNAs from a single

**Table 1. Summary list of bioinformatics pipelines/tools for analyzing different RNA variants.**

Variant analysis classification	Biochemical methods (methods for feature enrichment)	Bioinformatics tools	
		Pre-processing tools	Analysis tools
Alternative splice variants	a. RNAseq post Poly(A) enrichment for mRNA pulldown b. Isoform sequencing (Iso-Seq) and SLR-RNA-seq <sup>95</sup> (synthetic long-read) for full-length transcript sequencing c. RNaseR enrichment for circular RNA	- Quality check: fastQC and HTSeq - Processing: Fastx, Trimmomatic, and cutadapt - Mapping: STAR <sup>96</sup> , OLego <sup>97</sup> , and TopHat <sup>98</sup>	rMATS <sup>99</sup> , MISO <sup>53</sup> , IRCall, CIRCexplorer <sup>44</sup> , MAJIQ <sup>37</sup> , and DEXSeq
5' and 3' end variants	a. 3' end enrichment libraries: 3P-Seq, PAPERCLIP, and 3'-Seq b. poly(A) tail length: TAIL-seq, PAL-seq, and LM-PAT c. 5' end enrichment libraries: CAGE <sup>100</sup> and deepCAGE	- Quality check: fastQC and HTSeq - Processing: Fastx, Trimmomatic, and cutadapt - Mapping: STAR, OLego, and TopHat	MISO and DaPars
RNA modification variants	a. antibody-based IP: m6A, m1A, m5C, and hm5C b. Selective RNA chemistry: m5C and $\Psi$	- Processing: flexbar and Fastx - Mapping: STAR, OLego, and TopHat	MACS <sup>101</sup> , HOMER <sup>102</sup> , CIMS <sup>103</sup> , and CITS <sup>104</sup>
Translational variants	a. Ribosome profiling b. TATL-seq for alternative transcript leader sequences c. Poly(A)-primed sequencing (2P-seq) d. Frac-Seq and TrIP-seq	- Quality check: fastQC and HTSeq - Processing: Fastx, Trimmomatic, and cutadapt - Mapping: STAR, OLego, and TopHat	Plastid, RiboTaper <sup>105</sup> , Cufflinks/Cuffdiff <sup>106</sup> , and DESeq

gene. The functional outcomes are increased proteome diversity and the introduction of premature termination codons to degrade mRNAs by nonsense-mediated decay<sup>25</sup> and to generate variability in untranslated regions (UTRs) which modulate mRNA translation efficiency, stability, and localization<sup>26,27</sup>. Given the direct and widespread impact of alternative splicing on eukaryotic gene expression, cataloging the full repertoire of splice variants within individual cell types, developmental stages, and different human diseases has been a focus of many research laboratories.

The initial RNA-seq studies for detecting splicing patterns suffered from short read lengths, limiting the reliable genomic alignment of the alternative exon-exon junctions. Clever computational approaches, however, overcame these limitations through the addition of all potential splice junctions to the reference genome and accounting for them within the alignment mix<sup>28-30</sup>. The results transformed our view of the extent and complexity of eukaryotic transcriptomes, as more than 95% of human multi-exon genes were found to have alternatively spliced isoforms<sup>31</sup>. Importantly, calculating the number of mapped reads against individual exons plus each splice junction provided a direct measure of splicing efficiency that allowed precise quantification of isoform ratios.

The development of strand-specific and paired-end sequencing led to improved mapping, the discovery of pervasive anti-sense transcription<sup>32,33</sup>, and many novel splice variants in vertebrate species<sup>34,35</sup>. Emerging technologies are geared toward sequencing

the full-length transcripts (for example, Iso-seq<sup>36</sup>) such that variably spliced regions from different areas of a single gene can be accurately connected to provide information about the entire mRNA isoforms (that is, from the 5' end to the poly[A] tail). There has also been a strong push to improve the existing algorithms to capture the full complexity of splicing patterns within existing RNA-seq data. For instance, Vaquero-Garcia *et al.* have recently developed a new computational pipeline to accurately detect, visualize, and quantify complex local splice variants that are tissue-specific or change dynamically across different experimental conditions<sup>37</sup>.

In addition to the discovery of variably spliced regions, RNA-seq has been an incredible asset for the investigation of splicing intermediates and novel splice products. For instance, it led to the identification of recursive splicing where long introns require specific exon contexts and are spliced out in phases rather than the usual two-step process<sup>38,39</sup>. Recursive splicing uses cryptic sites within the intron, known as recursive splice (RS) site/ratchet points, which are surprisingly more conserved than the regular 5' and 3' splice sites. However, the identification of these RS sites required extensive deep sequencing with over a billion paired-end reads, without poly(A) selection in an attempt to capture pre-mRNA and nascent RNA transcripts<sup>40,41</sup>. Likewise, RNA-seq has renewed interest in studying circular RNAs (circRNAs). The discovery of circRNAs, though known for over two decades, remained mostly serendipitous<sup>42,43</sup>. Typical RNA-seq experiments fail to conserve RNA circularity because fragmentation and the absence of a

poly(A) tail or free 5' or 3' end cause de-enrichment of circRNAs during most library preparations. However, the lack of exposed 5' or 3' ends also stabilizes the circRNAs, making them resistant to exonuclease-mediated degradation. The recent development of biochemical methods to enrich for circRNAs (for example, RNaseR treatment to digest single-stranded RNA) along with new computational strategies (CIRCexplorer<sup>44</sup>) to mine the existing Encyclopedia of DNA Elements (ENCODE) Ribozero RNA-seq data has aided their *de novo* detection in humans, mice, and *Caenorhabditis elegans*. The results have revealed that most circRNAs are non-coding splice variants containing multiple exons located in the middle of genes that are often expressed in a tissue- and developmental stage-specific manner and that these variants may act as sponges to antagonize microRNA activity<sup>45,46</sup>.

### Cracking the terminal variants

Whereas the gene body variants in mRNAs promote proteome diversity, variability at the terminal ends is known to serve more of a regulatory role in protein synthesis. This realization has led to a strong surge in the development of new methodologies to identify all the 5' and 3' end variants within eukaryotic transcriptomes. For instance, recent methods have revealed that alternative transcription start sites (TSSs) are prevalent in the human genome and that approximately 30 to 50% of human genes have more than one promoter<sup>47</sup>. A substantial number of these alternative TSSs alter the 5' UTRs, in turn modulating the translational output of mRNAs<sup>48</sup>. 5' UTRs are known to comprise many regulatory features—5' cap structure, internal ribosome entry sites, G-quadruplexes, and so on—that can drastically affect the ability of ribosomes to assemble and initiate translation. Apart from forcing the gain or loss of such regulatory features, alternative TSSs generate RNA variants with different upstream open reading frames (uORFs). Ribosome footprinting studies have uncovered a significant fraction of such short translated uORFs within 5' UTRs that potently repress the translation of downstream ORFs<sup>49–51</sup>.

Apart from providing proper signals for mRNA 3' end formation and poly(A) site selection, the 3' UTR is considered to be a regulatory hub for gene regulation. Proper 3' end formation is essential for nuclear export and stability of mature mRNAs and their efficient translation<sup>52</sup>. Analysis of existing RNA-seq data with new computational pipelines (MISO and DaPars)<sup>53,54</sup> or sequencing of the libraries enriched for 3' ends of mRNAs (3P-Seq, 3'-seq, 3'READS, and PAPERCLIP) have documented that over 50% of human mRNAs express variable 3' UTRs—generated by alternative cleavage, splicing, and polyadenylation. Importantly, these 3' UTR variants are differentially expressed during differentiation, development, and disease<sup>55–58</sup>, and they often influence the accessibility of microRNAs and RNA-binding proteins (RBPs) to the 3' UTR *cis*-regulatory elements affecting mRNA stability, localization, and translation efficiency<sup>59</sup>.

Poly(A) tail length is another important 3' UTR feature that impacts the post-transcriptional fate of mRNAs. Although some biochemical methods (LMPAT<sup>60</sup> and ePAT<sup>61</sup>) can accurately measure the poly(A) tail lengths of individual genes, precise measurements at a genome-wide scale are hampered by difficulties in sequencing homopolymeric sequences longer than 30 nucleotides. However,

new advances in poly(A) tail length profiling methods (TAIL-seq and PAL-seq) use specialized 3' adapters, RNase T1 treatment (digests the mRNA body but not the poly[A] tail), and the spike in controls in library preparation. These biochemical improvements in conjunction with refined machine-learning algorithms to analyze the intensity of raw fluorescence signals allow the precise estimation of poly(A) tail lengths. These new methodologies have started to provide the first high-resolution, global views of poly(A) tail length variants and their relevance to mRNA stability and translational efficiency in cells and tissues from a variety of species, including yeast, *Drosophila*, zebrafish, *Xenopus*, mouse, and human<sup>62–64</sup>.

### Marking the variants

Though originally discovered in highly abundant rRNAs, tRNAs, and snRNAs, newly tailored mRNA sequencing methods have revealed that base modifications and editing sites are also highly prevalent in mRNAs<sup>65</sup>. The study of mRNA modifications (methylation, hydroxy-methylation, and pseudouridylation) plus editing (A-I or C-U) is now commonly referred to as “epitranscriptomics”<sup>66–68</sup>. Much like the epigenome, the epitranscriptome expands the information content of nucleic acids, rendering them additional structural and functional flexibility. Recent advances have led to the profiling of millions of modification sites for each of the known mRNA modifications, including methylation, as well as base editing (such as A-I or C-U)<sup>69,70</sup>. Whereas most of these have been observed across all genes, some are limited to specific classes of transcripts.

Remarkably, the genome-wide mapping of modifications and editing sites has not only expanded the landscape of this new class of mRNA variants within eukaryotic transcriptomes but also unraveled their dynamics in abundance and site selection in response to intra- and extra-cellular stimuli<sup>71–74</sup>. The impact of these post-transcriptional marks is now recognized in almost every step of gene regulation, including mRNA processing, stability, and translation<sup>75</sup>. An interesting example is the recent work on ADAR (adenosine deaminase acting on RNA) enzymes, showing their role in protecting double-stranded RNA (dsRNA) against recognition by dsRNA response pathways<sup>76</sup>. Importantly, the writers, erasers, and reader proteins for some of these modifications have been identified, and the phenotypic and molecular consequences of their gain or loss of function have been demonstrated<sup>77</sup>.

Today, the transcriptome-wide mapping of modified nucleosides relies on either antibody-based immunoprecipitation (m6A, m1A, m5C, and hm5C)<sup>66,78</sup> or a modification-selective RNA chemistry approach (m5C and Ψ)<sup>79,80</sup> to enrich for the target modification during library preparation. However, further optimization of these methods is needed to simultaneously obtain single-nucleotide resolution and stoichiometry of RNA modifications, as fractional modification represents another mechanism to generate functional diversity within individual mRNA pools. Current epitranscriptome profiling techniques accurately identify the modification sites within RNA transcripts but fail to provide information about the modified fraction for each site. Given that a modification may alter the mRNA secondary structure or binding site for RBPs, quantitative fractional information about each modified site in the transcriptome is highly desired.

## Translating the variants

Early dogma predicted that the amount of mRNA for a given gene directly correlated with the amount of protein made. However, we now know that through sequence variations discussed above and along with other RNA control elements, the translation of specific transcripts can be fine-tuned for the purpose of responding to cellular needs in a variety of contexts<sup>81</sup>. Therefore, estimating average gene-level translation by ribosome profiling<sup>82</sup>—a technique that uses short, approximately 28 nucleotide ribosome footprints on mRNA as a means of quantifying protein output—may not provide an accurate measure of the translational potential for individual RNA variants or isoforms. This is especially true for splice variants, which can exhibit variable coding capacities because of (i) sequence differences, (ii) altered mRNA half-lives, or (iii) differential nucleo-cytoplasmic export. Indeed, several recent efforts to measure isoform-specific translation have revealed discrete effects of 5' end<sup>83</sup>, 3' end<sup>84</sup>, and coding sequence<sup>85,86</sup> diversity.

Using translation-associated transcript leader sequencing (TATL-seq), Arribere and Gilbert successfully mapped transcript leader boundaries across the yeast genome and discovered hundreds of new alternative transcript leader variants, the majority having differential translation efficiencies<sup>83</sup>. Complementing this work, Spies *et al.* acquired decay and translation rates for proximal and distal 3' UTR variants using poly(A)-primed sequencing (2P-seq) but found only a modest regulatory influence for alternative 3' UTR sequences<sup>84</sup>. By comparing isoform ratios in cytoplasmic and polyribosomal extracts from human cells (Frac-seq), Sterne-Weiler *et al.* observed that not all splice isoforms are similarly loaded into polyribosomes and that the presence of microRNA binding sites or premature termination codons can greatly affect the loading of individual isoforms into polyribosomes<sup>85</sup>. Similarly, a major impact of mRNA isoforms on protein output was observed by Floor and Doudna, who recently adapted the classic approach of polysome profiling to measure isoform-specific translation (TriP-seq)<sup>87</sup>.

## Where are we headed?

Emerging technologies have given us a powerful toolbox to detect and quantify RNA variants from simple unicellular organisms to complex vertebrates. These new techniques allow for previously unimagined access to compare transcriptome patterns across species, tissues, developmental stages, and disease states. We now can isolate and sequence picogram amounts of RNA from fresh and formalin-fixed paraffin-embedded tissue specimens and even identify variants within single cells that are isolated from a wide range of biological samples. Moreover, the ENCODE project has generated unprecedented information about the many different RNA variants present in model organisms and human cells. Systematic discovery of *cis*-acting RNA elements, *trans*-acting RBPs that bind to these elements to control the production of variants, and their expression and subcellular localization is paving the way to uncovering the underlying code for generating much of the RNA diversity<sup>88,89</sup>. If sequencing costs continue to decline as anticipated, clinical measurements of normal and aberrant RNA variants in body fluids or tissue biopsies in the not-so-distant future may become part of personalized medicine, making for more accurate diagnoses and offering better-informed treatment plans.

However, as with most opportunities, certain challenges must be overcome to realize their full potential. What are the common challenges facing global profiling of RNA variants? First, biological specimens are usually heterogeneous, composed of multiple cell types. Since most RNA variants are expressed in a single cell type-specific<sup>90</sup>, developmental stage-specific<sup>91</sup>, or disease-specific<sup>92</sup> manner, particular attention must be paid when selecting optimal biological specimens for sequencing. For instance, individual cell types from most tissues now can be isolated by serial dilution, fluorescence-activated cell sorting, or laser-capture micro-dissection. Therefore, developing compound protocols that combine cell type-specific purification strategies with sequencing may provide more reliable information on RNA variants within particular cell types under normal or disease states. Second, identifying RNA variants from single-cell RNA-seq often requires multiple amplification steps to produce a sufficient quantity of nucleic acid for sequencing. These amplification steps introduce stochastic bias for certain sequences, which may cause dropout of certain low-abundance variants while amplifying others, thereby reducing the accuracy, reproducibility, and quantification power of detecting variants within single cells<sup>93</sup>. Third, there has been a rapid surge in the deposition of transcriptome data in public repositories in recent years. This has resulted in new challenges about how to effectively integrate the RNA variant information with other “Omics” datasets and build streamlined platforms for metagene analyses. Finally, perhaps the biggest challenge facing the field is the storage, distribution, and analysis of massive amounts of sequencing data. Multiple consortium projects like ENCODE, TCGA (The Cancer Genome Atlas), ExAC (Exome Aggregation Consortium), and other individual centers generate sequencing data at an astounding pace, doubling their numbers every seven months<sup>94</sup>. As the world moves toward a future of personalized genomic diagnostics and medicine, Stephens *et al.* propose that real-time transcript expression level analysis and the development of genomic variant databases against reference genomes will help solve the data storage problem<sup>94</sup>. With growing data, there is a need for platform development for data distribution. Cloud computing-based analysis may serve as the most practical option. While sequencing-based variant analysis provides a big boost for the future of precision medicine, handling genomics data is one of the most pressing challenges for the next decade.

---

## Competing interests

The authors declare that they have no competing interests.

## Grant information

AK is supported by grants from the US National Institutes of Health (R01HL126845), the March of Dimes (5-FY14-112), and the Center for Advanced Study at the University of Illinois.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Acknowledgments

We thank Sayeepriyadarshini Anakk and the members of the AK lab for constructive criticism and suggestions.

## References



1. Adams MD, Kelley JM, Gocayne JD, *et al.*: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science.* 1991; 252(5013): 1651–6.  
[PubMed Abstract](#) | [Publisher Full Text](#)
2. **F** Clark TA, Sugnet CW, Ares M Jr: **Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays.** *Science.* 2002; 296(5569): 907–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
3. Castle JC, Zhang C, Shah JK, *et al.*: **Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines.** *Nat Genet.* 2008; 40(12): 1416–25.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci U S A.* 1977; 74(12): 5463–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Sanger F, Air GM, Barrell BG, *et al.*: **Nucleotide sequence of bacteriophage phi X174 DNA.** *Nature.* 1977; 265(5569): 687–95.  
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Prober JM, Trainor GL, Dam RJ, *et al.*: **A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides.** *Science.* 1987; 238(4825): 336–41.  
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Beck S, Pohl FM: **DNA sequencing with direct blotting electrophoresis.** *EMBO J.* 1984; 3(12): 2905–9.  
[PubMed Abstract](#) | [Free Full Text](#)
8. Smith LM, Sanders JZ, Kaiser RJ, *et al.*: **Fluorescence detection in automated DNA sequence analysis.** *Nature.* 1986; 321(6071): 674–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Bentley DR, Balasubramanian S, Swerdlow HP, *et al.*: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature.* 2008; 456(7218): 53–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. **F** Margulies M, Egholm M, Altman WE, *et al.*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature.* 2005; 437(7052): 376–80.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
11. **F** Rothberg JM, Hinz W, Rearick TM, *et al.*: **An integrated semiconductor device enabling non-optical genome sequencing.** *Nature.* 2011; 475(7356): 348–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
12. Valouev A, Ichikawa J, Tonthat T, *et al.*: **A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning.** *Genome Res.* 2008; 18(7): 1051–63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Goodwin S, McPherson JD, McCombie WR: **Coming of age: ten years of next-generation sequencing technologies.** *Nat Rev Genet.* 2016; 17(6): 333–51.  
[PubMed Abstract](#) | [Publisher Full Text](#)
14. **F** Eid J, Fehr A, Gray J, *et al.*: **Real-time DNA sequencing from single polymerase molecules.** *Science.* 2009; 323(5910): 133–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
15. Morin R, Bainbridge M, Fejes A, *et al.*: **Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing.** *Biotechniques.* 2008; 45(1): 81–94.  
[PubMed Abstract](#) | [Publisher Full Text](#)
16. **F** Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet.* 2009; 10(1): 57–63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
17. **F** Nojima T, Gomes T, Grosso AR, *et al.*: **Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing.** *Cell.* 2015; 161(3): 526–40.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
18. **F** Churchman LS, Weissman JS: **Nascent transcript sequencing visualizes transcription at nucleotide resolution.** *Nature.* 2011; 469(7330): 368–73.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
19. **F** Paulsen MT, Veloso A, Prasad J, *et al.*: **Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA.** *Methods.* 2014; 67(1): 45–54.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
20. Giulietti M, Piva F, D'Antonio M, *et al.*: **SpliceAid-F: a database of human splicing factors and their RNA-binding sites.** *Nucleic Acids Res.* 2013; 41(Database issue): D125–31.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. **F** D'Antonio M, Castrgnano T, Pallocca M, *et al.*: **ASPicDB: a database web tool for alternative splicing analysis.** *Methods Mol Biol.* 2015; 1269: 365–78.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
22. **F** Picardi E, D'Erchia AM, Lo Giudice C, *et al.*: **REDportal: a comprehensive database of A-to-I RNA editing events in humans.** *Nucleic Acids Res.* 2016; pii: gkw767.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
23. **F** Tang F, Barbacioru C, Wang Y, *et al.*: **mRNA-Seq whole-transcriptome analysis of a single cell.** *Nat Methods.* 2009; 6(5): 377–82.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
24. **F** Wills QF, Livak KJ, Tipping AJ, *et al.*: **Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments.** *Nat Biotechnol.* 2013; 31(8): 748–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
25. McGlincy NJ, Smith CW: **Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?** *Trends Biochem Sci.* 2008; 33(8): 385–93.  
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Hughes TA: **Regulation of gene expression by alternative untranslated regions.** *Trends Genet.* 2006; 22(3): 119–22.  
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Kalsotra A, Cooper TA: **Functional consequences of developmentally regulated alternative splicing.** *Nat Rev Genet.* 2011; 12(10): 715–29.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. **F** Mortazavi A, Williams BA, McCue K, *et al.*: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods.* 2008; 5(7): 621–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
29. **F** Wang ET, Sandberg R, Luo S, *et al.*: **Alternative isoform regulation in human tissue transcriptomes.** *Nature.* 2008; 456(7221): 470–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
30. **F** Sultan M, Schulz MH, Richard H, *et al.*: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science.* 2008; 321(5891): 956–60.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
31. Pan Q, Shai O, Lee LJ, *et al.*: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.** *Nat Genet.* 2008; 40(12): 1413–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
32. **F** Seila AC, Calabrese JM, Levine SS, *et al.*: **Divergent transcription from active promoters.** *Science.* 2008; 322(5909): 1849–51.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
33. Faghihi MA, Wahlestedt C: **Regulatory roles of natural antisense transcripts.** *Nat Rev Mol Cell Biol.* 2009; 10(9): 637–43.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Merkin J, Russell C, Chen P, *et al.*: **Evolutionary dynamics of gene and isoform regulation in Mammalian tissues.** *Science.* 2012; 338(6114): 1593–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. **F** Barbosa-Morais NL, Irimia M, Pan Q, *et al.*: **The evolutionary landscape of alternative splicing in vertebrate species.** *Science.* 2012; 338(6114): 1587–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
36. **F** Sharon D, Tligner H, Grubert F, *et al.*: **A single-molecule long-read survey of the human transcriptome.** *Nat Biotechnol.* 2013; 31(11): 1009–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
37. **F** Vaquero-Garcia J, Barrera A, Gazzara MR, *et al.*: **A new view of transcriptome complexity and regulation through the lens of local splicing variations.** *eLife.* 2016; 5: e11752.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
38. **F** Burnette JM, Miyamoto-Sato E, Schaub MA, *et al.*: **Subdivision of large introns in *Drosophila* by recursive splicing at nonexonic elements.** *Genetics.* 2005; 170(2): 661–74.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
39. Hatton AR, Subramaniam V, Lopez AJ: **Generation of alternative *Ultrabithorax* isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions.** *Mol Cell.* 1998; 2(6): 787–96.  
[PubMed Abstract](#) | [Publisher Full Text](#)
40. **F** Sibley CR, Emmett W, Blazquez L, *et al.*: **Recursive splicing in long vertebrate genes.** *Nature.* 2015; 521(7552): 371–5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
41. **F** Duff MO, Olson S, Wei X, *et al.*: **Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*.** *Nature.* 2015; 521(7552): 376–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
42. Burd CE, Jeck WR, Liu Y, *et al.*: **Expression of linear and novel circular forms of an *INK4/ARF*-associated non-coding RNA correlates with atherosclerosis risk.** *PLoS Genet.* 2010; 6(12): e1001233.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Capel B, Swain A, Nicolis S, *et al.*: **Circular transcripts of the testis-determining gene *Sry* in adult mouse testis.** *Cell.* 1993; 73(5): 1019–30.  
[PubMed Abstract](#) | [Publisher Full Text](#)
44. **F** Zhang XO, Wang HB, Zhang Y, *et al.*: **Complementary sequence-mediated**

- exon circularization. *Cell*. 2014; **159**(1): 134–47.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
45. **F** Guo JU, Agarwal V, Guo H, *et al.*: Expanded identification and characterization of mammalian circular RNAs. *Genome Biol*. 2014; **15**(7): 409.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
46. **F** Memczak S, Jens M, Elefsinioti A, *et al.*: Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013; **495**(7441): 333–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
47. Davuluri RV, Suzuki Y, Sugano S, *et al.*: The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 2008; **24**(4): 167–77.  
[PubMed Abstract](#) | [Publisher Full Text](#)
48. Rojas-Duran MF, Gilbert WV: Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA*. 2012; **18**(12): 2299–305.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. **F** Johnstone TG, Bazzini AA, Giraldez AJ: Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J*. 2016; **35**(7): 706–23.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
50. **F** Bazzini AA, Johnstone TG, Christiano R, *et al.*: Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*. 2014; **33**(9): 981–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
51. **F** Ingolia NT, Lareau LF, Weissman JS: Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; **147**(4): 789–802.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
52. Proudfoot NJ: Ending the message: poly(A) signals then and now. *Genes Dev*. 2011; **25**(17): 1770–82.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Katz Y, Wang ET, Airolidi EM, *et al.*: Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*. 2010; **7**(12): 1009–15.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
54. **F** Xia Z, Donehower LA, Cooper TA, *et al.*: Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun*. 2014; **5**: 5274.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
55. **F** Jan CH, Friedman RC, Ruby JG, *et al.*: Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*. 2011; **469**(7328): 97–101.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
56. **F** Lianoglou S, Garg V, Yang JL, *et al.*: Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev*. 2013; **27**(21): 2380–96.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
57. Hoque M, Ji Z, Zheng D, *et al.*: Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods*. 2013; **10**(2): 133–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. **F** Hwang H, Park CY, Goodarzi H, *et al.*: PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage. *Cell Rep*. 2016; **15**(2): 423–35.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
59. Elkon R, Ugalde AP, Agami R: Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet*. 2013; **14**(7): 496–506.  
[PubMed Abstract](#) | [Publisher Full Text](#)
60. Sallés FJ, Richards WG, Strickland S: Assaying the polyadenylation state of mRNAs. *Methods*. 1999; **17**(1): 38–45.  
[PubMed Abstract](#) | [Publisher Full Text](#)
61. Jänicke A, Vancuylenberg J, Boag PR, *et al.*: ePAT: a simple method to tag adenylated RNA to measure poly(A)-tail length and other 3' RACE applications. *RNA*. 2012; **18**(6): 1289–95.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
62. **F** Chang H, Lim J, Ha M, *et al.*: TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell*. 2014; **53**(6): 1044–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
63. **F** Subtelny AO, Eichhorn SW, Chen GR, *et al.*: Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*. 2014; **508**(7494): 66–71.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
64. **F** Lim J, Lee M, Son A, *et al.*: mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes Dev*. 2016; **30**(14): 1671–82.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
65. **F** Licht K, Jantsch MF: Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. *J Cell Biol*. 2016; **213**(1): 15–22.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
66. **F** Meyer KD, Saleore Y, Zumbo P, *et al.*: Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012; **149**(7): 1635–46.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
67. Jaffrey SR: An expanding universe of mRNA modifications. *Nat Struct Mol Biol*. 2014; **21**(11): 945–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
68. **F** Schwartz S, Bernstein DA, Mumbach MR, *et al.*: Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*. 2014; **159**(1): 148–62.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
69. **F** Blanc V, Park E, Schaefer S, *et al.*: Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. *Genome Biol*. 2014; **15**(6): R79.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
70. **F** Bazak L, Haviv A, Barak M, *et al.*: A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res*. 2014; **24**(3): 365–76.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
71. Meyer KD, Jaffrey SR: The dynamic epitranscriptome: N<sup>6</sup>-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol*. 2014; **15**(5): 313–26.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
72. Karjane J, Yi C, Yu YT: Transcriptome-wide dynamics of RNA pseudouridylation. *Nat Rev Mol Cell Biol*. 2015; **16**(10): 581–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
73. Roundtree IA, He C: RNA epigenetics—chemical messages for posttranscriptional gene regulation. *Curr Opin Chem Biol*. 2016; **30**: 46–51.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
74. **F** Han L, Diao L, Yu S, *et al.*: The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer Cell*. 2015; **28**(4): 515–28.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
75. Gilbert WV, Bell TA, Schaenig C: Messenger RNA modifications: Form, distribution, and function. *Science*. 2016; **352**(6292): 1408–12.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
76. Liddicoat BJ, Piskol R, Chalk AM, *et al.*: RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science*. 2015; **349**(6252): 1115–20.  
[PubMed Abstract](#) | [Publisher Full Text](#)
77. Liu N, Pan T: N<sup>6</sup>-methyladenosine-encoded epitranscriptomics. *Nat Struct Mol Biol*. 2016; **23**(2): 98–102.  
[PubMed Abstract](#) | [Publisher Full Text](#)
78. **F** Dominissini D, Moshitch-Moshkovitz S, Schwartz S, *et al.*: Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature*. 2012; **485**(7397): 201–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
79. **F** Cartile TM, Rojas-Duran MF, Zinshteyn B, *et al.*: Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*. 2014; **515**(7525): 143–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
80. Squires JE, Patel HR, Nousch M, *et al.*: Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res*. 2012; **40**(11): 5023–33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
81. Brockmann R, Beyer A, Heinisch JJ, *et al.*: Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol*. 2007; **3**(3): e57.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
82. **F** Ingolia NT, Ghaemmaghami S, Newman JR, *et al.*: Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; **324**(5924): 218–23.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
83. Arribere JA, Gilbert WV: Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res*. 2013; **23**(6): 977–87.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
84. Spies N, Burge CB, Bartel DP: 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res*. 2013; **23**(2): 2078–90.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
85. **F** Sterne-Weiler T, Martinez-Nunez RT, Howard JM, *et al.*: Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res*. 2013; **23**(10): 1615–23.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
86. **F** Maslon MM, Heras SR, Bellora N, *et al.*: The translational landscape of the splicing factor SRSF1 and its role in mitosis. *eLife*. 2014; **3**: e02028.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
87. **F** Floor SN, Doudna JA: Tunable protein synthesis by transcript isoforms in human cells. *eLife*. 2016; **5**: pii: e10921.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
88. **F** van Nostrand EL, Pratt GA, Shishkin AA, *et al.*: Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 2016; **13**(6): 508–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
89. **F** Sundararaman B, Zhan L, Blue SM, *et al.*: Resources for the Comprehensive Discovery of Functional RNA Elements. *Mol Cell*. 2016; **61**(6): 903–13.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)

90. Bhatte A, Parker DJ, Bebee TW, *et al.*: **ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation.** *Nat Commun.* 2015; **6**: 8768.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
91. Giudice J, Xia Z, Wang ET, *et al.*: **Alternative splicing regulates vesicular trafficking genes in cardiomyocytes during postnatal heart development.** *Nat Commun.* 2014; **5**: 3603.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
92. Ly H, Calado RT, Allard P, *et al.*: **Functional characterization of telomerase RNA variants found in patients with hematologic disorders.** *Blood.* 2005; **105**(6): 2332–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
93. Liu S, Trapnell C: **Single-cell transcriptome sequencing: recent advances and remaining challenges [version 1; referees: 2 approved].** *F1000Res.* 2016; **5**: pii: F1000 Faculty Rev-182 .  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
94. **F** Stephens ZD, Lee SY, Faghri F, *et al.*: **Big Data: Astronomical or Genomical?** *PLoS Biol.* 2015; **13**(7): e1002195.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
95. **F** Tilgner H, Jahanbani F, Blauwkamp T, *et al.*: **Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events.** *Nat Biotechnol.* 2015; **33**(7): 736–42.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
96. **F** Dobin A, Davis CA, Schlesinger F, *et al.*: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013; **29**(1): 15–21.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
97. Wu J, Anczuków O, Krainer AR, *et al.*: **OLeGo: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds.** *Nucleic Acids Res.* 2013; **41**(10): 5149–63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
98. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics.* 2009; **25**(9): 1105–11.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
99. Shen S, Park JW, Huang J, *et al.*: **MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data.** *Nucleic Acids Res.* 2012; **40**(8): e61.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
100. **F** de Hoon M, Hayashizaki Y: **Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference.** *Biotechniques.* 2008; **44**(5): 627–8, 630, 632.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
101. Zhang Y, Liu T, Meyer CA, *et al.*: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol.* 2008; **9**(9): R137.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
102. **F** Heinz S, Benner C, Spann N, *et al.*: **Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.** *Mol Cell.* 2010; **38**(4): 576–89.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
103. Zhang C, Darnell RB: **Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data.** *Nat Biotechnol.* 2011; **29**(7): 607–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
104. **F** Weyn-Vanhenyck SM, Mele A, Yan Q, *et al.*: **HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism.** *Cell Rep.* 2014; **6**(6): 1139–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)
105. **F** Calviello L, Mukherjee N, Wyler E, *et al.*: **Detecting actively translated open reading frames in ribosome profiling data.** *Nat Methods.* 2016; **13**(2): 165–70.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [F1000 Recommendation](#)
106. **F** Trapnell C, Roberts A, Goff L, *et al.*: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc.* 2012; **7**(3): 562–78.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [F1000 Recommendation](#)

# Open Peer Review

Current Referee Status:



---

## Editorial Note on the Review Process

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

---

## The referees who approved this article are:

### Version 1

- 1 Gordon Carmichael**, Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT, USA  
**Competing Interests:** No competing interests were disclosed.
- 2 Emanuele Buratti**, Molecular Pathology, International Center for Genetic Engineering and Biotechnology, University of Trieste, Trieste, Italy  
**Competing Interests:** No competing interests were disclosed.
- 3 Graziano Pesole**, Dipartimento Di Bioscienze, Biotechnologie e Biofarmaceutica, Università degli Studi di Bari, Bari, Italy  
**Competing Interests:** No competing interests were disclosed.