RNA trafficking and subcellular localization—a review of mechanisms, experimental and predictive methodologies

Jun Wang, Marc Horlacher, Lixin Cheng and Ole Winther

Abstract

RNA localization is essential for regulating spatial translation, where RNAs are trafficked to their target locations via various biological mechanisms. In this review, we discuss RNA localization in the context of molecular mechanisms, experimental techniques and machine learning-based prediction tools. Three main types of molecular mechanisms that control the localization of RNA to distinct cellular compartments are reviewed, including directed transport, protection from mRNA degradation, as well as diffusion and local entrapment. Advances in experimental methods, both image and sequence based, provide substantial data resources, which allow for the design of powerful machine learning models to predict RNA localizations. We review the publicly available predictive tools to serve as a guide for users and inspire developers to build more effective prediction models. Finally, we provide an overview of multimodal learning, which may provide a new avenue for the prediction of RNA localization.

Keywords: RNA, subcellular, localization, machine learning, multimodality

INTRODUCTION

The subcellular localization of mRNA was first reported in 1986 by Lawrence and Singer, who observed this phenomenon in chicken fibroblasts using in situ hybridization [1], and it was subsequently found in mammalian neurons [2]. Later, mRNA localization was found to encompass a large percentage of the transcriptome during Drosophila melanogaster [3] and Xenopus laevis [4] development as well as in Escherichia coli [5] and mammalian tissue [6]. Research has shown that RNAs can aggregate in distinct patterns within various cellular compartments, such as protrusions [7], RNA foci [8], cell membrane [9], nuclear envelope [10] and polarization [11]. RNAs can also assemble in multiple organelles such as the endoplasmic reticulum (ER) [12], mitochondrial [13], ribosome [14] and nucleus [15] (Figure 1). These observations suggested that mRNA localization contributes to the post-transcriptional fine-tuning of gene expression and the control of fundamental processes such as cell motility, polarity and differentiation.

Subcellular localization can be the potential treatment target in disease therapy. Most oligonucleotide therapies are highly efficient at silencing mRNA expressed in the cytoplasm, whereas HTT in neurons is difficult to silence because it is primarily concentrated in the nucleus, where HTT shows more resistance to silence the therapies. Therefore, subcellular localization is a promising target to develop and optimize the therapeutics to address the resistance of silencing in the nucleus [16]. In the context of cancer biology, protrusion-enriched RNAs, such as RAB13 and NET1 RNAs, are enriched specifically at the invasive front of leader cells in invasive cell strands, providing targeting opportunity for interfering with collective cancer cell invasion [17]. Furthermore, RNA localization can also give an implication for muscle and...
neurological disorders. For instance, in spinal muscular atrophy [18–22], the mutation of the SMN1 gene leads to the dysfunction of SMN proteins, resulting in the impaired trafficking and translation of axonal mRNA [18, 20]. Amyloid b-peptide (Ab) plaques in the affected region of the brain, the main reason of Alzheimer’s disease (AD), trigger the localized translation of mRNA coding for ATF4 transcription factor when exposed in axons, which is then moved to the nucleus and induce the neuron degeneration [23]. Fragile X syndrome is instigated by undesired CGG repeat expansion within the 5′ UTR of the FMR1 gene, which results in the loss-of-function of the encoded FMRP protein [24]. Strikingly, FMRP is an RBA-binding protein (RBP) involved in stimulus-induced dendritic mRNA transport, and its malfunctioning can cause mRNA trafficking defects and impaired synaptic maturaton [24–26]. Myotonic dystrophy (DM) is also caused by nucleotide repeat expansions at specific genomic loci, where CTG repeats within the 3′ UTR of the DMPK gene for DM type 1 and CCTG repeats within the first intron of the ZNF9 gene for DM type 2 [27]. The aberrance of these transcripts acts as sponges to bind and sequester RBPs in the nucleus, hindering these proteins to perform their biological functions. The aforementioned examples illustrate the significant role that abnormal RNA localization plays in disrupting cellular pathways, highlighting the potential benefits of regulating target-localized molecules to mitigate these perturbations. More comprehensive details about the relationship between RNA localization and disease formation can refer to the review of Cody et al. [28] and Ashley et al. [29].

In cells, localized RNA translation has several advantages. First, the localized translation can energetically benefit the translation process by encoding proteins locally, instead of synthesizing them from distant sites and exerting functions on the target compartments [30]. Second, it allows for a rapid response to extracellular stimuli if the sites of transcripts and their synthesized proteins coincide in a spatially and temporally restricted mechanism [31, 32]. Third, it allows proteins to disperse into specific cell compartments and avoid the ectopic expression of proteins in undesired compartments, where proteins may go to waste or result in deleterious effects [33]. Fourth, transcripts can provide a reservoir-like component, which can be released under a certain simulation [34]. Fifth, the distinct posttranslational modification owned by targeted translation could have profound influences on the activity of protein complexes at the target sites [30]. Finally, in addition to target translation, localized mRNAs may also carry out non-coding functions, e.g. by exerting catalytic or scaffolding activities [35, 36].

Early studies investigated the effects of subcellular localization of RNA on cell morphology gradients and cellular fates in the development models, such as the Drosophila embryo and Xenopus oocyte [37, 38], as well as polarized cells such as migrating fibroblasts [7], budding yeast [39] and neuronal cells [40]. It has been shown that RNA subcellular localization is a widespread phenomenon across several RNA families, including non-coding RNAs [41–44], whose localization is a critical determinant of function. They serve the competing endogenous RNA mechanism, inhibiting microRNAs [45, 46], and regulate the transcriptional process by recruiting chromatin-modifying complexes to control RNA splicing [47].

Localization of several RNAs cannot be easily detected and quantified using experimental methods due to their low abundance, rapid degradation and purification challenges. While experimental assays are the most straightforward way to characterize the subcellular localization of biomolecules, they are typically time consuming and costly. Furthermore, an mRNA may interact with several RBPs in different sequential components. For instance, dozens of RBPs have been found to interact with gurken mRNA [48, 49]. In the previous study, the RBP binding with multiple zip codes led to longer running distances and higher running frequency with the characteristics of recruiting motor more effectively [50]. Consequently, identifying which RBP and where the functional motif directs the mRNA localization is still challenging.

It is crucial to develop computational tools to effectively predict RNA subcellular localization, which can reduce costs and simultaneously predict multiple candidate biomolecules. RNA localization is known to be driven by the interaction of the RNA with a number of RNA-binding proteins (RBPs), often through the recognition of binding motifs. As computational methods commonly learn a function of RNA sequence to RNA localization, interpretation of method predictions, for instance through feature importance scores [51], may give insight into bound RBPs by revealing their binding motifs as predictive sub-sequences. Together, this may aid in elucidating pathways of RNA localization and enable the identification of sequence variants which may impact the localization of a given RNA. In this review, we focus on machine learning-based methods that have been shown to yield highly accurate predictions of RNA subcellular localization across different RNA species.

RNA subcellular localization

There are three major mechanisms to control the localization of mRNA to desired subcellular compartments: directed transport, protection from mRNA degradation and passive diffusion before local entrapment. The untranslated mRNAs are transported to target compartments to exert their biological functions according to different transport approaches, where cis-regulatory elements are bound by trans-regulatory factors, such as RBPs.

Cellular mechanisms of RNA localization

Molecular identification

The short sequence elements in the RNA sequence, primarily located in the 3′ UTR [52], can act in cis to control mRNA localization. For instance, ‘zip code’ sequences of β-actin mRNA, whose bipartite RNA element is in the 3′ UTR, are recognized by the third and fourth KH domains of zipcode-binding protein 1 (ZBP1) [53]. This recognition guides the sequences to the leading edge of fibroblasts along the cytoskeleton in a motor-dependent manner [54] (Figure 2A). In addition, the dimerization of the stem-loop structures localized in the 3′ UTR of Drosophila melanogaster bicoid mRNA and RBP Staufen activates the transport of bicoid mRNA to the anterior pole of the oocyte at the late stage of oogenesis along the microtubules [52, 54–56]. In yeast, in addition to one localization sequence in the 3′ UTR, three other elements in the coding sequence of ASH1 mRNA are also required for activating transport by the myosin motor Myo4 [57]. Motif binding can also regulate the localization across cell types with different morphologies. For example, the TOP motif, which was at the extreme 5′ end of the ribosomal proteins, and its regulatory activity was abolished upon perturbation of the TOP-binding protein LARP1, which was the RNA element regulator known to regulate RNA localization to the apicobasal axis of epithelial cells and neurites of mouse neuronal cells [58]. Comprehensive RBP binding events and their binding motifs can be accessed in several RBP databases [59–61]. For example, RBPD has curated 272 experimental validated RBPs with documented binding events, where 71 of them were in the form of position weight matrices that show the RBP binding motif
within various mRNA targets. However, further investigation is required to determine if these motifs directly contribute to RNA localization.

**Molecular transport**
mRNAs and ribonucleoproteins (RNPs) are organized in cellular units of diverse composition, structure, size and function, all of which are loosely termed RNA granules. These RNA granules are located in various subcellular locations and affect the RNA concentration [62]. Specific interactions between RNAs and proteins can affect cell functions and cell fates, together with cell–cell communication during the response to variants of environmental stimuli after being transported by variant shipping methods. Interestingly, the subcellular distribution of the transcripts may contribute to different cell states and types [63, 64].

**Directed transport**
Active transport is the most common mode of mRNA transportation in all eukaryotic cells via actin filaments or microtubules [65] (Figure 2B). The mRNA is transported unidirectionally along bidirectional microtubules with several cellular mechanisms of directed mRNA transport. This occurs by binding RBP with cis-elements to form a localization-competent transport granule. However, this motor-driven transport is energy consuming, especially for transporting in the long axon. To address this issue, mRNAs can hitch-hike on organelles such as lysosomes [66], endosomes [67] and mitochondria [68] for long-distance transport in the axon.

**Protection from mRNA degradation**
Furthermore, the mRNA encoding patterning can be determined by mRNA localization by different regulators. For example, in D. melanogaster embryos, RBP Smaug inhibits Nos mRNA translation and recruits the CCR4–NOT complex to trigger Nos mRNA decay in the cytoplasm. When Nos mRNA is transported to the posterior pole in a later development stage, Oskar replaces Smaug to interact with Nos mRNA, enabling the translation and protecting the mRNA from degradation [65] (Figure 2C).

**Diffusion and local entrapment**
mRNA diffusion in prokaryotes, such as bacteria, leads to the localization of mRNA to the ribosome-rich poles or membranes with a diffusion coefficient of 0.05 μm²/s within the 1–2 μm long cell [69] (Figure 2D). In eukaryotes, Nos mRNA can be squeezed from the nurse cells to the oocyte during D. melanogaster oogenesis. Following the cytoplasmic streaming that microtubules generate for transport [70], Nos mRNAs are entrapped in the germplasm in an actin-dependent manner once at the posterior pole (Figure 2E). In mammals, restricted diffusion was observed in fibroblasts and neurons [71–73], which was confirmed as faster and more efficient than directed transport along with the microtubules without binding with the motors in a short distance movement [74].

**Experimental assays for studying RNA subcellular localization**
To date, the most popular approach to studying subcellular localization is image based. Currently, two types of image-based methods are used to characterize RNA subcellular localization. Many...
studies make use of smFISH-based techniques followed by epifluorescence or confocal microscopy to visualize and quantify intracellular mRNAs. Alternatively, an MS2 tagging system, where the bacteriophage MS2 protein (MCP) was tagged to a unique RNA hairpin sequence [75, 76], followed by live cell imaging is applied [77]. Several novel MS2-like systems have been developed to investigate the localization of single mRNA molecules, which advances the usage of the MS2 system in more broad application paradigms [77–79], even enabling the simultaneous localization of the mRNA and its protein product [80].

With the development of single-cell spatial transcriptomics technologies, to visualize multiple transcripts in a single cell, a multiplexing technique was introduced to build the group of smFISH-based techniques. These advanced methods make use of multiple single-stranded DNA oligonucleotides, each labeled with a single fluorophore, tiling a specific RNA target; the signal obtained from multiple single fluorophores can be seen via confocal microscopy and can be easily quantified [81]. The most novel implementations, sequential fluorescent signal in situ hybridization (seqFISH+) [82, 83] and multiplexed error-robust fluorescence in situ hybridization (MERFISH) [12], allow for a cost-effective smFISH multiplexing and are characterized by high sensitivity (80–100%) and high scalability [12, 83] (up to more than 10 000 gene targets) (Figure 3).

Alternatively, the subcellular localization of specific RNAs can be experimentally established through biochemical fractionation followed by RT-qPCR or RNA-seq [84, 85]. This type of method requires the detergent lysis of tissue culture cells, breaking the intact cell and losing the overall distribution map of a target transcript, although it has the characteristics of extreme sensitivity and high sequence-specificity using versatile DNA binding dyes like SYBR Green I [85]. However, these methods are not suitable for investigating smaller organs, due to their vulnerability and insufficient abundance during homogenization and centrifugation. To summarize the experimental data, RNALocate [86, 87] was proposed as the most completed database to provide a comprehensive RNA localization resource for researchers to deconvolute the highly complex architecture of cells.

**Summarization of image collection at single-molecular resolution**

The advent of spatial transcriptomics techniques has revolutionized the study of gene expression by enabling the preservation of the native spatial context of RNA within tissues. These approaches can be broadly categorized into four distinct classes, namely, in situ hybridization (ISH) based, in situ sequencing based, next-generation sequencing based and spatial information reconstruction [88]. However, among these methods, the ISH-based approach is the only one that currently offers a subcellular resolution. This imaging-based protocol enables the investigation of RNA on the single-molecular level within cells, thereby providing valuable insights into the complex machinery of how RNA localization to subcellular regions is regulated and accomplished.

In recent years, image-based techniques have become a popular tool for counting RNAs in situ, with several studies applying them to human and mouse cell lines and tissues. To obtain a comprehensive overview of the available image-based datasets, we manually collected 40 publicly available datasets generated between 2015 and 2023, including those published by academic laboratories and commercial platforms with proof-of-concept (Supplementary Table 1). Currently, there are nine spatial transcriptomics techniques available for profiling transcripts at subcellular resolution, including MERFISH [12], seqFISH [89], seqFISH+ [83], intron seqFISH [90], splitFISH [72], STARmap PLUS [91], CosMx SMI [92], Xenium and EEL FISH. Notably, we found that more than half of these techniques are based on MERFISH, with gene panels ranging from 11 to 10 000. In addition, STARmap PLUS and CosMx SMI are two distinct methods that can simultaneously map single-cell transcripts and proteins, with CosMx SMI profiling a larger number of proteins than STARmap PLUS. These approaches enable the profiling of not only ubiquitous genes in the cell body but also the investigation of gene expression at their nascent transcriptome active sites [90, 93].

The availability of image data generated from spatial transcriptomics techniques provides a valuable resource for investigating the spatial localization of RNAs at subcellular resolution. These
datasets offer insights into the spatial expression patterns of genes in different cell types and in response to environmental stimuli. Although ISH-based techniques are highly efficient in terms of detection, they can be expensive to perform, resulting in low gene throughput. Nevertheless, recent advances in spatial transcriptomics techniques, such as large field-of-view [94], multi-omics [91, 92, 95, 96] and low-cost [97] approaches, have the potential to enable more accurate and comprehensive observation of the spatial distribution of RNAs. These advancements hold great promise for unlocking new discoveries in the field of spatial transcriptomics, ultimately leading to a better understanding of the complex regulation of gene expression within cells.

In silico prediction of RNA localization

The rise in studies emphasizes the importance of RNA spatial and temporal subcellular localization in biological processes [90, 98–101]. To avoid the inherent temporal and monetary defects of experimental methods, in silico computational frameworks are introduced to predict the potential subcellular compartments of RNAs. Machine learning-based methods, including traditional and deep learning methods, are well suited for the prediction of subcellular localization. Currently, there are several machine learning-based methods to predict RNA subcellular localizations. These methods are based on different approaches to extracting a high-level RNA feature representation (Table 1). Most of the efforts have been devoted to building classifiers to predict mRNA localization based on the sequences’ high-level representations. Model inputs can in principle be divided into two modalities, images and RNA sequences, which are followed by the process of different encoding strategies.

Image-based model

Generally, ground truth image data are less available than experimentally validated sequence data that use RNA-seq or RT-qPCR according to its fast turn-around time. However, with the in silico simulation techniques, variants of RNA localization patterns have been simulated [102, 103]. Samacoits et al. [104] designed a set of features to describe different RNA localization patterns based on simulated images, including polarized distribution, and accumulation in cell extensions or foci, at the cell membrane or nuclear envelope. Using 23 features including the ratio of mRNA inside the nucleus and outside the nucleus, they can precisely distinguish different localizations via supervised and unsupervised approaches. In addition, Clarence et al. [105] applied 13 features including point distribution, point density and symmetry about the center of mass to describe the RNA distribution in the synthetic images, building an effective random forest classifier to predict the subcellular localization. Notably, Samacoits et al. [104] assigned the 3D localization labels, such as nuclear envelope 3D, enabling the discrimination of intranuclear distribution and the nucleus membrane that are the same from the view of 2D. In recent years, deep neural networks have been applied to automatically extract features from simulated smFISH images, overcoming the time-consuming nature and difficulties of manual annotation, and yielding high classification performance on cell edge prediction according to the supervised model [82].

Sequence-based model

Sequence-based methods have been explored predominantly over image-based ones, as they predict RNA subcellular localization, and therefore transport mechanisms, based on RNA regulatory sequence elements, such as cis-elements located in the 3’ UTR region. Subsequent probing of the models may in silico explain the transport mechanisms through molecular binding between RNA elements and proteins. In comparison, smFISH images can only provide temporal visualization of transcript localization within the cell body. In addition, RNA sequences across RNA species can be easily obtained instead of requiring labor-intensive experiments to obtain RNA localization images. IncLocator [106] and iLoc-mRNA [107] take advantage of the sequence features using a k-mer sequence encoding strategy to predict the localization of lncRNA and mRNA. Concretely, IncLocator abstracts these k-mer features to high-level representation via stacked autoencoder (AE) and combines with an integrated model containing both random forest (RF) and support vector machine (SVM) followed by a fully connected neural network to build an integrated machine learning model.

Although the feature dimensions extracted by k-mer are fixed, discrete feature structures might completely lose all the sequence-order or pattern information. To explore the important information, pseudo-K-tuple nucleotide composition (pseKNC) was introduced [108], which is an extension of PseAC (Pseudo Amino Acid Composition) [109] that can be used to capture DNA/RNA sequences. The parameter k of pseKNC can be picked according to the biological function and specific structures need to be investigated, of which pseDNC is one of the variants considering dinucleotide and different tiers of dinucleotides along the DNA sequence. To the best of our knowledge, miRNAloc [110] and mRNALoc [111] implement the pseKNC technique to generate feature vectors to predict the microRNA and mRNA localization, respectively. In addition to nucleotide composition, mRGoFS use a similarity function to infer the correlation between the query miRNA and their six localizations, which can combine with GO terms of BP, MF and CC, resulting in the 18-dimensional feature vector [112]. Furthermore, physicochemical features [113], electronic properties [114, 115], e.g. electron–ion interaction pseudopotential values (PseEIP), and biological features [116] have been applied to characterize the sequences, followed by a rigorous feature selection scheme for selecting the optimal feature subsets to train the machine learning models.

Deep learning can significantly advance the model performance as it alleviates the need for manual feature engineering implemented in conventional machine learning models. Several researchers implemented convolutional neural network (CNN)-based structures independently to extract local information along the sequence [117, 118]. DM3Loc [119] used multiscale CNN filters and multi-head self-attention to make the model pay more attention to the influential sequence regions, of which zip code regions might be assigned with high attention weight. In contrast, RNATracker [120] and IncLocator 2.0 [121] couple a CNN with bidirectional LSTM, with the latter aggregating information from both directions while maintaining spatial order. To incorporate the inherent associations between locations, mRLoc [122] transforms the original multi-label problem of microRNA localization prediction as a sequence-to-sequence problem and trains the model end-to-end using an encoder-decoder architecture.

While there exists a range of computational approaches for predicting target RNA species, a comprehensive comparison among these approaches remains limited. Therefore, it is crucial to undertake a benchmark dataset that encompasses a wide range of well-defined cellular compartments and is meticulously categorized by different cell types. This comprehensive dataset will enable a more robust evaluation of the computational models’ performance. To assess model performance quantifiably, several confusion matrix-based evaluation methods can be employed, including area under the curve (AUC) and precision-recall curve.
<table>
<thead>
<tr>
<th>Tools</th>
<th>RNA types</th>
<th>Modality</th>
<th>Models</th>
<th>Features</th>
<th>Localizations</th>
<th>Multi-label prediction</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 CNN workflow</td>
<td>All</td>
<td>Simulated image</td>
<td>CNN(^{a})</td>
<td>Auto</td>
<td>Cell edge; cell extension; foci; intranuclear; nuclear edge; polarized; random</td>
<td>No</td>
<td>[123]</td>
</tr>
<tr>
<td>2 RF classifier</td>
<td>All</td>
<td>Simulated image</td>
<td>RF(^{b})</td>
<td>23 curated localization features</td>
<td>Foci; extension; nuclear envelope 3D; nuclear envelope 2D; random</td>
<td>Yes</td>
<td>[104]</td>
</tr>
<tr>
<td>3 Bento</td>
<td>All</td>
<td>Simulated image</td>
<td>RF</td>
<td>13 curated localization features</td>
<td>Cell edge; cytoplasmic; nuclear; nuclear edge; random</td>
<td>Yes</td>
<td>[105]</td>
</tr>
<tr>
<td>4 EL-RMLocNet</td>
<td>mRNA, miRNA, snRNA, IncRNA</td>
<td>RNA sequence</td>
<td>LSTM(^{c})</td>
<td>Auto</td>
<td>Pseudopodium; nucleolus; nucleus; cytoplasm; endoplasmic reticulum; exosome; microvesicle; cytoplasm</td>
<td>Yes</td>
<td>[124]</td>
</tr>
<tr>
<td>5 iLoc-mRNA</td>
<td>mRNA</td>
<td>RNA sequence</td>
<td>SVM(^{d})</td>
<td>k-Mer</td>
<td>Exosome; cytosol; endoplasmic reticulum; ribosome; mitochondrion; nucleus</td>
<td>No</td>
<td>[107]</td>
</tr>
<tr>
<td>6 DM3Loc</td>
<td>mRNA</td>
<td>RNA sequence</td>
<td>CNN</td>
<td>Auto</td>
<td>Exosome; membrane; cytoplasm; ribosome; endoplasmic reticulum; nucleus</td>
<td>Yes</td>
<td>[119]</td>
</tr>
<tr>
<td>7 mRNALoc</td>
<td>mRNA</td>
<td>RNA sequence</td>
<td>SVM</td>
<td>pseKNC</td>
<td>Mitochondria; extracellular region; cytoplasm; nucleus; endoplasmic reticulum</td>
<td>No</td>
<td>[111]</td>
</tr>
<tr>
<td>8 RNATracker</td>
<td>mRNA</td>
<td>RNA sequence</td>
<td>CNN + LSTM</td>
<td>Auto</td>
<td>Cytosol; nuclear; membranes; insoluble; endoplasmic reticulum; mitochondria</td>
<td>No</td>
<td>[120]</td>
</tr>
<tr>
<td>9 mRNALocator</td>
<td>mRNA</td>
<td>RNA sequence</td>
<td>Ensemble model (LightGBM, XGBoost, CatBoost)</td>
<td>PseEIIP, PseKNC</td>
<td>Mitochondria; endoplasmic reticulum; extracellular region; nucleus; cytoplasm</td>
<td>No</td>
<td>[115]</td>
</tr>
<tr>
<td>10 SubLocEP</td>
<td>mRNA</td>
<td>RNA sequence</td>
<td>Ensemble model of LightGBM integrated by sequence-based model and physicochemical-based model</td>
<td>Sequence based (PseEIIP, TNC, DNC, CKSNAP) Physicochemical properties (PCPseDNC, PCpseTNC, SCPseDNC, SCPseTNC, DACC)</td>
<td>Cytoplasm; endoplasmic reticulum; extra cellular location; mitochondria; nucleus</td>
<td>No</td>
<td>[113]</td>
</tr>
<tr>
<td>11 MIRGOFS</td>
<td>microRNA</td>
<td>RNA sequence</td>
<td>SVM</td>
<td>Correlation scores obtained from BP, CC and MF DAGs</td>
<td>Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus</td>
<td>Yes</td>
<td>[112]</td>
</tr>
<tr>
<td>12 mRNALoc</td>
<td>microRNA</td>
<td>RNA sequence</td>
<td>SVM</td>
<td>PseDNC</td>
<td>Axon; circulating; cytoplasm; exosome; extracellular; vesicle; microvesicle; mitochondrion; nucleus</td>
<td>No</td>
<td>[110]</td>
</tr>
<tr>
<td>13 L2S-MirLoc</td>
<td>microRNA</td>
<td>RNA sequence</td>
<td>A RF + NB(^{e}) + SVM model pool</td>
<td>PseEIIP</td>
<td>Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus</td>
<td>Yes</td>
<td>[114]</td>
</tr>
<tr>
<td>14 MirLocPredictor</td>
<td>microRNA</td>
<td>RNA sequence</td>
<td>CNN</td>
<td>Auto</td>
<td>Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus</td>
<td>No</td>
<td>[118]</td>
</tr>
<tr>
<td>15 miRLocator</td>
<td>microRNA</td>
<td>RNA sequence</td>
<td>Encoder-decoder model</td>
<td>Auto</td>
<td>Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus</td>
<td>No</td>
<td>[122]</td>
</tr>
<tr>
<td>16 iLoc-lncRNA</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>SVM</td>
<td>k-Mer</td>
<td>Nucleus; exosome; ribosome; cytoplasm</td>
<td>No</td>
<td>[125]</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Tools</th>
<th>RNA types</th>
<th>Modality</th>
<th>Models</th>
<th>Features</th>
<th>Localizations</th>
<th>Multi-label prediction</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepLncRNA (Gudenas, B. et al. 2018)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>DNN†</td>
<td>Auto</td>
<td>Nuclear; cytosolic</td>
<td>No</td>
<td>[126]</td>
</tr>
<tr>
<td>IncLocator (Cao, Z. et al. 2018)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>Ensemble model (SVM + RF)</td>
<td>k-Mer + latent space from AE§</td>
<td>Cytoplasm; nucleus; cytosol; ribosome; exosome</td>
<td>No</td>
<td>[106]</td>
</tr>
<tr>
<td>IncLocator 2.0 (Lin, Y. et al. 2021)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>CNN, BiLSTM§, MLP¶</td>
<td>Auto</td>
<td>Cytoplasm/nucleus relative concentration index (CNRCI)</td>
<td>NA</td>
<td>[121]</td>
</tr>
<tr>
<td>IncLocation (Feng, S. et al. 2020)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>SVM</td>
<td>Selected k-mer and biological features</td>
<td>Cytoplasm; exosome; nucleus; ribosome</td>
<td>No</td>
<td>[116]</td>
</tr>
<tr>
<td>Locate-R (Ahmad, A. et al. 2020)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>LD-SVM</td>
<td>n-Gapped l-mers; l-mers</td>
<td>Cytoplasm; nucleus; ribosome; exosome</td>
<td>No</td>
<td>[127]</td>
</tr>
<tr>
<td>KD-KLNMF (Zhang, S. et al. 2020)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>SVM</td>
<td>k-Mer; dinucleotide-based spatial autocorrelation (DSA)</td>
<td>Nucleus; cytoplasm; ribosome; exosome</td>
<td>No</td>
<td>[128]</td>
</tr>
<tr>
<td>DeepLncLoc (Zeng, M. et al. 2022)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>CNN</td>
<td>Auto</td>
<td>Nucleus; cytosol; ribosome; cytoplasm; exosome</td>
<td>No</td>
<td>[117]</td>
</tr>
<tr>
<td>TACOS (Young-Jun. et al. 2022)</td>
<td>IncRNA</td>
<td>RNA sequence</td>
<td>6 tree-based classifiers</td>
<td>Primary sequence; composition and physicochemical properties</td>
<td>Nucleus and cytoplasm</td>
<td>No</td>
<td>[129]</td>
</tr>
<tr>
<td>RNAlight (Guo-Hua et al. 2022)</td>
<td>IncRNA; mRNA</td>
<td>RNA sequence</td>
<td>lightGBMκ</td>
<td>k-Mers</td>
<td>Nucleus and cytoplasm</td>
<td>No</td>
<td>[130]</td>
</tr>
</tbody>
</table>

(a) CNN: convolutional neuronal network; (b) RF: random forest; (c) LSTM: long short-term memory; (d) SVM: support vector machine; (e) NB: naive Bayes; (f) DNN: deep neuronal network; (g) AE: autoencoder; (h) BiLSTM: bidirectional long short-term memory; (i) MLP: multilayer perceptron; (j) LD-SVM: locally deep SVM; (k) lightGBM: light gradient boosting machine.
which account for data imbalance. Additional metrics such as accuracy (ACC), F1 score and the more reliable Matthew’s correlation coefficient [131] can also be utilized. Furthermore, building a comprehensive computational framework to predict multiple RNA classes and species could be considered for developers in the future like what has been done by Muhamman et al. [124]. smFISH-based methods, which implement instantaneous imaging techniques that only capture the observation of transcripts, may be a great option to couple with the mechanism explained sequence-based model to comprehensively build a model to cover various RNA species. Currently, many prediction methods focus on identifying functional RNA localization elements solely based on primary sequences, which can be limited and biased in their explanations. In contrast, incorporating prior knowledge of RBPs, such as integrating reported sequence motifs associated with RNA-binding events, can provide a valuable guide for localization prediction, thereby enhancing the biological interpretability of the developed models.

Multimodal deep learning

While functional properties of RNA sequences can explain most of the localization mechanisms, they do not provide straightforward information regarding predicting subcellular localization. The variants of cell types, tissue types and stimulus conditions make it difficult for the invariant RNA sequence to perfectly match the dynamic cell prototypes. Fluorescent microscopy imaging data is the gold standard to determine RNA subcellular localization; however, they are costly and hard to reach the whole transcriptional scale. Therefore, the combination of these two modalities, as proposed by Flavia Savulescu et al. [132], may improve localization prediction, enabling better discrimination of the variations in a developmental, diseased and normal cellular context. In the following, we will explore multimodal modeling and its potential for subcellular localization. Our discussion will include various biomedical applications, methodologies and challenges associated with implementing multimodal integration.

Unimodal data, including image data [133–138] (e.g. radiological and histopathology images) and non-image data [139–142] (e.g. blood tests, structured genomics sequences or clinical reports), are commonly used to build the discriminative models separately. Ideally, these data can be integrated and used to advance model performance. Multimodal data integration makes it possible to investigate to what degree the independently derived data complement one another. By contrast, if two modalities have fully mutual information, the performance will not be improved when compared with a single modality alone [143].

Current methods alone exhibit limited performance, posing a significant challenge in achieving satisfactory results for all cellular compartments. The AUC typically falls within the range of 0.4 to 0.8 [119], indicating that relying solely on primary sequence, relevant physicochemical features and conserved motifs may not offer a comprehensive depiction of RNA translocation to target compartments. In contrast to the unimodal-based methods listed in Table 1, the integration of multimodal data, such as protein product information and smFISH-based images, holds the substantial potential to provide a more comprehensive understanding of the diverse factors influencing RNA trafficking. By representing these elements as vectors, it becomes possible to extract higher-level abstract features that capture the interactions among these factors, ultimately resulting in improved predictions for each subcellular compartment. Furthermore, the combination of complementary information from each modality can enhance their individual predictive capabilities. For example, smFISH images can enhance RNA localization prediction, while the sequence model can effectively identify distinctive RNA distribution patterns depicted in the images.

Multimodal machine learning has been a growing trend in machine learning research in the fields of clinical diagnosis and prognosis [144]. Ngiam et al. showed that better features of one modality can be learned if multiple modalities are present at training time [145], which creates the potential to use multimodal learning to facilitate clinical research on disease prognosis and diagnosis when heterogeneous clinical data are available [144]. For example, miRNA sequencing and promoter methylation status were integrated by early fusion autoencoder, enabling better stratifying patients with hepatocellular carcinoma by overall survival (OS) [146]. Huang et al. [147] integrated mRNA-seq and miRNA-seq data by a deep Cox proportional hazards model in an early fusion framework to stratify patients with breast cancer by OS, which proved the outperformance when compared with unimodal models. Image data also successfully integrate with molecular data. Specifically, histopathology images and their zoomed cell morphological features were used to combine with the genomic profile, and learned representative features were combined using tensor fusion networks to predict survival outcomes and grade classification [148].

A major design for multimodal approaches is the extent to which each data input should be modeled before joint representations are encoded. The strategies of fusing heterogeneous information from multimodal data are the key pursuit in multimodal learning. Based on the types of input for multimodal fusion, fusing strategies can be easily divided into feature-level fusion and decision-level fusion (Figure 4). Decision-level fusion is straightforward and can be achieved with simple calculation using average, weighted vote or majority vote [149, 150]. However, feature-level fusion, including early fusion and intermediate
fusion, fuses the original heterogeneous data by combining the informative multimodal hidden representation to construct the powerful model [151–153]. Broad categories have been proposed to capture the relationship of different modalities, including operation-based, subspace-based, attention-based, tensor-based and graph-based methods [144]. Although the integration of two or more modalities typically surpasses the unimodal in the downstream tasks such as disease diagnosis or prognosis, some research has proved the negative influence or no influence on multimodal fusion. Hence, for the high performance of multimodal, one should consider the modal capacity and data quality toward a specific task.

To fully implement the complementary information inherent in different modalities, several novel model structures have been proposed to take advantage of each modality. For example, Haytham et al. [153] presents an audiovisual fusion model that utilizes an attention mechanism to dynamically combine the processed visual and audio input and make the multi-label predictions. Interestingly, the representation of one modality can be learned from its complementary modality, such as visual and audio data of the same video clips, by using pseudo-labels generated from its counterpart to backpropagate error in a self-supervision manner [154], or a student-teacher training procedure to implement the learned discriminative knowledge from a pre-trained and fixed modality to improve the representation of the other [155]. There are several other significant advances in the field of visual-language pre-training (VLP) [156], which can be adapted and advanced with the release of several customized computational frameworks for the prediction of RNA localization.

Although training a model using heterogeneous data is promising and provides the opportunity to make use of complementary information among different modalities, there are three types of inherent challenges we face while using multimodal deep learning: (1) Data scarcity. Artificial intelligence (AI) has enormous potential to convert data into newly generated representations to facilitate diagnosis and prognosis in clinical research; however, the research datasets are sparse, for example, the genome profile data may lack paired histopathology image data when multimodal datasets are necessary for a multimodal model. (2) Overfitting. Most multimodal architectures have more parameters to fit than their unimodal counterparts, making them prone to overfitting. One of the strategies to address this is to weigh each unimodal contribution to the overall loss based on its estimated generalization error using the gradient blending technique [131]. (3) Reproducibility. As it occurred in major published biomedical AI studies, they failed to provide the source code, test data and both, making it hard to reproduce the results. To increase research transparency, investigators are encouraged to submit their new multimodal strategies and preprocessing regimes in standardized repositories such as modelhub.ai [157].

**CONCLUSION**

We presented a review of biological processes throughout RNA trafficking from the nascent RNA in the nucleus to their target localizations in variant cellular compartments, including the identification of RNA and its binding proteins before transport to remote localization through three different biological mechanisms. To determine the spatial localization of RNAs, experimental techniques based on the biochemical processes before sequencing and high-resolution imaging approaches have been proposed. To visualize RNAs of interest localized in variant compartments, there are many ISH-based approaches that enable reaching the subcellular resolution. With advances in ISH, an increasing amount of data had been generated when investigating different tissues in humans and mice, which may potentially offer direct transcript localization information in the cellular context within a specific field of view. However, spatial transcriptomics techniques are not designed to directly image RNA at the organelle level. To investigate RNA localization within organelles, a combination of spatial transcriptomics techniques with organelle-specific labeling or isolation methods may be necessary. Importantly, learning-based computational tools can dramatically extend the interpretability and accuracy when predicting the behavior of those RNAs, which are easily degraded and too rare to be detected, in a much more economical and effective manner. The development of multimodal natural language process and computer vision models has shown that learning complementary information from diverse modalities is possible. This provides opportunities to significantly improve the accuracy of RNA localization predictions.

The prediction of RNA localization has been well adapted using different types of machine learning-based tools. Potential localizations of RNAs of interest can be determined according to different customized model structures, which can predict not only mRNA, but also lncRNA and microRNA, allowing for comprehensive monitoring of RNA distribution at subcellular resolution. However, whether these tools can be applied at the single-cell level, such as in cell types with different morphologies, functions and metabolisms, still needs to be investigated. To direct the implementation of different methodologies according to different RNA species, a thorough benchmarking comparison needs to be conducted to decipher the suitable usage of these learning-based tools in different cellular settings. Furthermore, the sequence ontology, which was tokenized by almost all predictive methods, cannot fully capture the biological processes that occur during RNA trafficking. Inherent features, including gene ontology, RNA-binding proteins, sequence binding motif and physicochemical characteristics of sequence, can also influence the ultimate localization of RNAs. External environmental factors and cell–cell communication also significantly affect where RNAs should be located. These factors mentioned above should be taken into consideration when building strong and advanced computational tools in the future.

Moreover, we believe that leveraging language pre-trained models could be a promising approach in the context of predicting RNA localization. In the realm of protein research, large language pre-trained models have successfully been utilized to predict protein localization, as demonstrated in a study referencing DeepLoc2 [158]. In a similar vein, a visual model pre-trained on cell images could potentially capture organelle-related concepts, making it a suitable candidate for fine-tuning to identify RNAs that are enriched in specific organelles. For instance, utilizing pre-trained models that already extract highly representative features from proteins exhibiting diverse localizations [159] could prove advantageous.

Like multi-omics analysis [160], comprehensively using the complementary knowledge from at least two modalities to train multimodal models has been proven to be realistic and can significantly improve the model performance in VLP and biomedical diagnosis and prognosis. As the localization of RNAs is dynamic and complicated according to influential components, fluorescence-based imaging techniques provide an opportunity to snapshot where the transcripts should present after being influenced by various factors. Combining this information with influential sequence binding motifs and their functional protein
products may create a new era for RNA localization prediction as massive datasets are generated from the development of single-cell spatial transcriptomics technologies.

Key Points
• Our focus on biological mechanisms, experimental measurements and in silico modeling and prediction covers a broad range of topics related to the quickly developing field of RNA localization.
• Sequence-based and image-based models can be powerful in respective ways to decipher RNA subcellular localization in cellular ecosystems. In this review, we examine the current state of research on these models, which have traditionally been studied separately.
• Spatially resolved transcriptomics, named the method of the year in 2020 by Nature Methods, experienced a data explosion over the past few years. Hence, it is time to give thoughtfull consideration to how these image data can be utilized to elucidate the mechanisms underlying the trafficking and localization of RNAs to distinct cellular compartments.
• This manuscript provides a review of the wide-ranging methodological toolbox for people who are interested in predicting RNA localization and inspire computational biologists to develop state-of-the-art tools to predict and uncover RNA trafficking in an unbiased manner.
• Multimodal learning combines information from various sources and modalities for a comprehensive understanding of a task. We believe that this approach offers valuable insights and could be a promising direction for future research in RNA localization prediction.

SUPPLEMENTARY DATA
Supplementary data are available online at https://academic.oup.com/bib.

FUNDING
This work was supported by China Scholarship Council (CSC) with a 4-year PhD grant to J.W., Novo Nordisk Foundation [NNF20OC0062606], and Danish National Research Foundation [the Pioneer Centre for AI, grant number P1] to O.W.

REFERENCES


