OXFORD

Briefings in Bioinformatics, 2023, **24(5)**, 1–14
https://doi.org/10.1093/bib/bbad249
Advance access publication date 18 July 2023
Review

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

Jun Wang, Marc Horlacher, Lixin Cheng and Ole Winther

Corresponding author. Jun Wang, Bioinformatics Centre, Department of Biology, University of Copenhagen, København Ø 2100, Denmark.
E-mail : wangjun19950708@gmail.com

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.

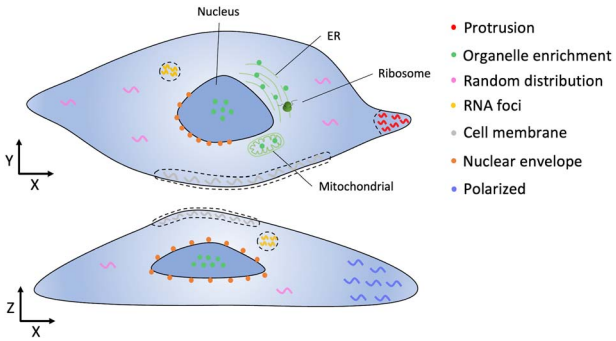

Figure 1. Spatial distribution of RNAs among various cell compartments. The cellular plot is divided into two sets for different perspectives. The top set represents the x–y axis, while the lower set represents the x–z axis. Squiggles and dots denote different RNAs, and distinct colors indicate different compartments.

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

Jun Wang is a PhD student at the University of Copenhagen, Denmark. His research focuses on transfer learning and its application in spatial omics data.
Marc Horlacher is a PhD student at the Computational Health Center of the Helmholtz Center Munich. His research focuses on interpretable machine learning methods for biomolecular sequences.
Lixin Cheng, PhD, is a principal investigator of bioinformatics at Shenzhen People's Hospital, China. His research interests include biomarker identification and omics data analysis.
Ole Winther is a professor at the Department of Biology, University of Copenhagen and DTU Compute, The Technical University of Denmark. Ole Winther's research focuses on machine learning methodology, bioinformatics and natural language processing.
Received: March 17, 2023. **Revised:** May 30, 2023. **Accepted:** June 16, 2023

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 maturation [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 methanism [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 RNAs 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, RBPDB 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

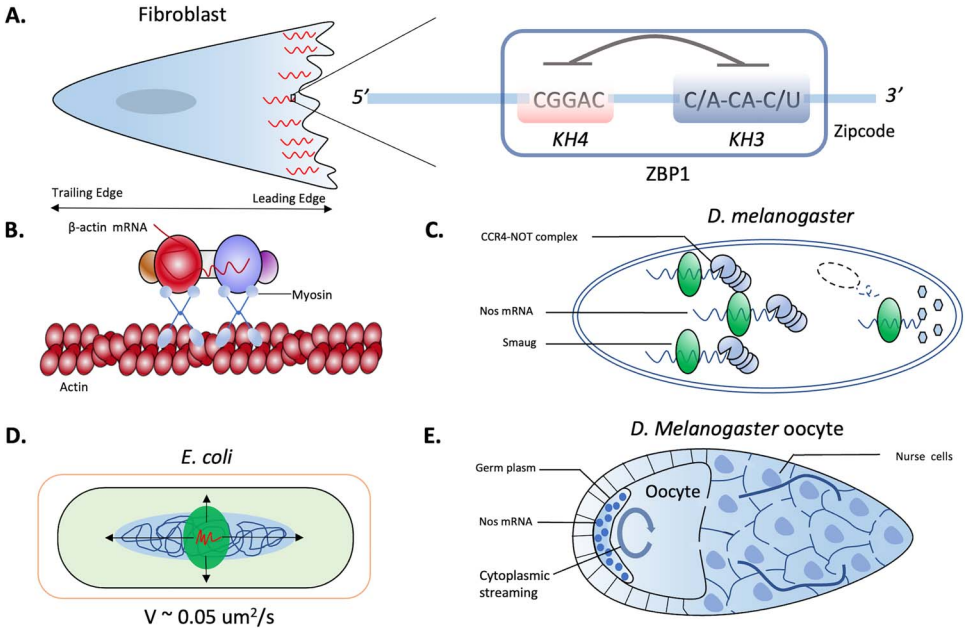

Figure 2. The introduction of the molecular mechanism of localization through three mainly investigated manners. (**A**) In fibroblasts, β-actin mRNAs are localized to the leading edge of cell with the RBP KH3 domain binding to the 3′ end of the sequence and KH4 domain binding to the 5′ end of sequence. Interestingly, these two domains can identify specific sequence motif. For example, KH4 recognizes CGGAC and KH3 for C/A-CA-C/U, where the first base can be substitute as C or A, and the fourth base can be C or U. (**B**) The transport of β-actin mRNAs to the leading edge of fibroblasts can be achieved by binding the mRNPs with multiple motors. For example, β-actin mRNAs have two binding domains, which mediate the binding of ZBP1, and in turn, the binding of myosin. Afterward, this mRNP-motor complex can be transported via actin filaments. (**C**) In *Drosophila melanogaster* embryo, CCR4–NOT complex was recruited to decay the mRNA when RBP Smung bind to the Nos mRNAs. However, if the Nos mRNAs are localized to the posterior pole of the embryo, Smung was detached from Nos mRNA and was replaced by Oskar proteins, results in the protection from degradation and improves the concentration of Nos mRNA at the posterior site. (**D**) In *Escherichia coli*, nascent mRNAs can localize to the ribosome-rich poles or to the membrane from the nucleoid by random diffusion with the speed of 0.05 $\mu m^2/s$. (**E**) During the oogenesis of *D. melanogaster*, hundreds of mRNA, including Nos mRNA, were squeezed from the nurse cells into the oocyte, where they are entrapped in the germ plasm followed by the cytoplasmic streaming. mRNP, mRNA-protein.

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 $\mu m^2/s$ within the 1–2 μm long cell [69] (Figure 2D). In eukaryotes, Nos mRNA can be squeezed from 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

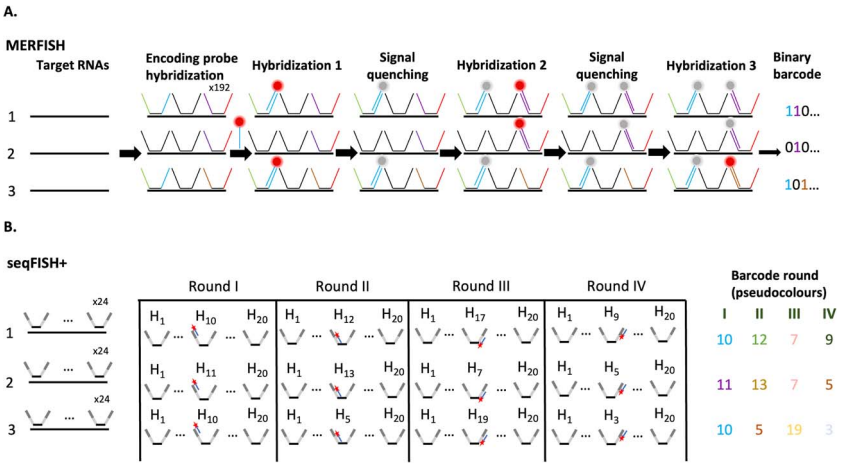

Figure 3. The basic principles of MERFISH (multiplexed error-robust RNA fluorescence *in situ* hybridization) and seqFISH+ are outlined. (**A**) MERFISH uses a two-step hybridization process. Initially, multiple nonfluorescent encoding probes are hybridized to RNA molecules, providing several readout hybridization sequences for rapid binding of fluorescently labeled readout probes. The fluorescent dyes attached to the readout probes can be chemically deactivated, enabling multiple cycles of hybridization with different readout probes, fluorescence imaging and signal quenching. This results in a sophisticated error-robust binary barcode readout for each target transcript. (B) In seqFISH+, mRNAs are hybridized with 24 primary probes, which can be imaged after 20 cycles of readout hybridization. The *in situ* mRNAs are then imaged using fluorescence, with each localized mRNA assigned a pseudocolor. After four rounds of imaging hybridization, genes can be encoded using different imaging channels, typically 640 nm, 561 nm and 488 nm.

studies make use of smFISH-based techniques followed by epi-fluorescence 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 fluorescence *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 organelles, 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. lncLocator [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, lncLocator 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, miRGOFS 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 (PseEIIP), 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 lncLocator 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, miRLocator [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 1. Summary of current sequence and image-based and machine learning-based predictive tools

	Tools	RNA types	Modality	Models	Features	Localizations	Multi-label prediction	Ref
1	CNN workflow (Dubois, R. *et al.* 2019)	All	Simulated image	CNN[a]	Auto	Cell edge; cell extension; foci; intranuclear; nuclear edge; polarized; random	No	[123]
2	RFclassifier (Samacoits, A. *et al.* 2018)	All	Simulated image	RF[b]	23 curated localization features	Foci; extension; nuclear envelope 3D; nuclear envelope 2D; random	Yes	[104]
3	Bento (Mah, C. K. *et al.* 2022)	All	Simulated image	RF	13 curated localization features	Cell edge; cytoplasmic; nuclear; nuclear edge; random	Yes	[105]
4	EL-RMLocNet (Asim, M. *et al.*2022)	mRNA, miRNA, snoRNA, lncRNA	RNA sequence	LSTM[c]	Auto	Pseudopodium; nucleolus; nucleus; cytosol; mitochondrion; ribosome; endoplasmic reticulum; exosome; microvesicle; cytoplasm	Yes	[124]
5	iLoc-mRNA (Zhang, Z.-Y. *et al.* 2021)	mRNA	RNA sequence	SVM[d]	*k*-Mer	Exosome; cytosol; endoplasmic; reticulum; ribosome; mitochondrion; nucleus	No	[107]
6	DM3Loc (Wang, D. *et al.* 2021)	mRNA	RNA sequence	CNN	Auto	Exosome; membrane; cytoplasm; ribosome; endoplasmic reticulum; nucleus	Yes	[119]
7	mRNALoc (Garg, A. *et al.* 2020)	mRNA	RNA sequence	SVM	pseKNC	Mitochondria; extracellular region; cytoplasm; nucleus; endoplasmic reticulum	No	[111]
8	RNATracker (Yan, Z. *et al.* 2019)	mRNA	RNA sequence	CNN + LSTM	Auto	Cytosol; nuclear; membranes; insoluble; endoplasmic reticulum; mitochondria	No	[120]
9	mRNALocater (Tang, Q. *et al.* 2021)	mRNA	RNA sequence	Ensemble model (LightGBM, XGBoost, CatBoost)	PseEIIP, PseKNC	Mitochondria; endoplasmic reticulum; extracellular region; nucleus; cytoplasm	No	[115]
10	SubLocEP (Li, J. *et al.* 2021)	mRNA	RNA sequence	Ensemble model of LightGBM integrated by sequence-based model and physicochemical-based model	Sequence based (PseEIIP, TNC, DNC, CKSNAP) Physicochemical properties (PCPseDNC, PCPseTNC, SCPseDNC, SCPseTNC, DACC)	Cytoplasm; endoplasmic reticulum; extracellular location; mitochondria; nucleus	No	[113]
11	MIRGOFS (Yang, Y. *et al.* 2018)	microRNA	RNA sequence	SVM	Correlation scores obtained from BP, CC and MF DAGs	Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus	Yes	[112]
12	miRNALoc (Meher, P. K. *et al.* 2020)	microRNA	RNA sequence	SVM	PseDNC DiPro	Axon; circulating; cytoplasm; exosome; extracellular; vesicle; microvesicle; mitochondrion; nucleus	No	[110]
13	L2S-MirLoc (Asim, M *et al.* 2021)	microRNA	RNA sequence	A RF + NB[e] + SVM model pool	PseEIIP	Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus	Yes	[114]
14	MirLocPredictor (Asim, M *et al.* 2020)	microRNA	RNA sequence	CNN	Auto	Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus	No	[118]
15	miRLocator (Xiao, Y. *et al.* 2018)	microRNA	RNA sequence	Encoder-decoder model	Auto	Exosome; cytoplasm; mitochondrion; microvesicle; circulating; nucleus	No	[122]
16	iLoc-lncRNA (Su, Z.-D. *et al.* 2018)	lncRNA	RNA sequence	SVM	*k*-Mer	Nucleus; exosome; ribosome; cytoplasm	No	[125]

(Continued)

Table 1. Continued.

	Tools	RNA types	Modality	Models	Features	Localizations	Multi-label prediction	Ref
17	DeepLncRNA (Gudenas, B. *et al.* 2018)	lncRNA	RNA sequence	DNN[f]	Auto	Nuclear; cytosolic	No	[126]
18	lncLocator (Cao, Z. *et al.* 2018)	lncRNA	RNA sequence	Ensemble model (SVM + RF)	*k*-Mer + latent space from AE[g]	Cytoplasm; nucleus; cytosol; ribosome; exosome	No	[106]
19	lncLocator 2.0 (Lin, Y. *et al.* 2021)	lncRNA	RNA sequence	CNN, BiLSTM[h], MLP[i]	Auto	Cytoplasm/nucleus relative concentration index (CNRCI)	NA	[121]
20	lncLocation (Feng, S. *et al.* 2020)	lncRNA	RNA sequence	SVM	Selected *k*-mer and biological features	Cytoplasm; exosome; nucleus; ribosome	No	[116]
21	Locate-R (Ahmad, A. *et al.* 2020)	lncRNA	RNA sequence	LD-SVM	n-Gapped l-mers; l-mers	Cytoplasm; nucleus; ribosome; exosome	No	[127]
22	KD-KLNMF (Zhang, S. *et al.* 2020)	lncRNA	RNA sequence	SVM	*k*-Mer; dinucleotide-based spatial autocorrelation (DSA)	Nucleus; cytoplasm; ribosome; exosome	No	[128]
23	DeepLncLoc (Zeng, M. *et al.* 2022)	lncRNA	RNA sequence	CNN	Auto	Nucleus; cytosol; ribosome; cytoplasm; exosome	No	[117]
24	TACOS (Young-Jun. et al. 2022)	lncRNA	RNA sequence	6 tree-based classifiers	Primary sequence; composition and physicochemical properties	Nucleus and cytoplasm	No	[129]
25	RNAlight (Guo-Hua et al. 2022)	lncRNA; mRNA	RNA sequence	lightGBM[k]	*k*-Mers	Nucleus and cytoplasm	No	[130]

(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.

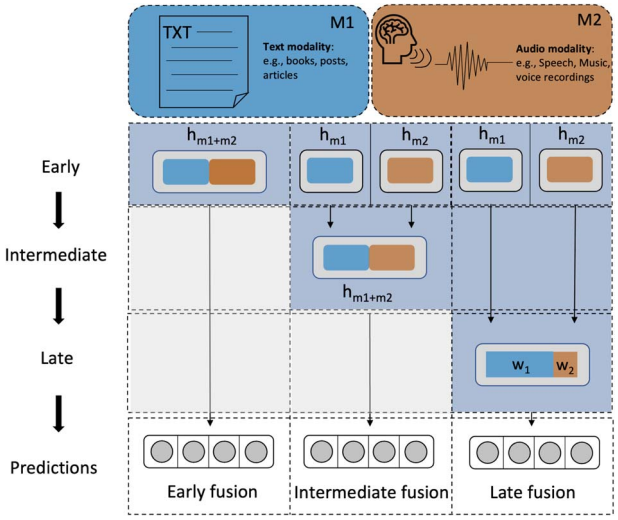

Figure 4. The design of a bimodal fusion incorporates data from multiple sources. There are three primary strategies for fusing data from different modalities. Early fusion: This strategy involves encoding each modality and projecting them into a high-dimensional representation, which is then used as a single input to construct the prediction model. Intermediate fusion: Instead of fusing data through voting or average-based methods, encoded unimodal representations are integrated through a separate model to form multimodal representations. These representations are then used as inputs to binary or multi-class classifiers. Late fusion: In this approach, each modality is encoded into its own unimodal representation, which is then processed using simple calculations such as average, weighted vote or majority vote.

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 architectures 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 need to be investigated. To direct the implementation of different methodologies according to variant 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 data 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 thoughtful 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

1. Lawrence JB, Singer RH. Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* 1986;**45**:407–15.
2. Knowles RB, Sabry JH, Martone ME, *et al.* Translocation of RNA granules in living neurons. *J Neurosci* 1996;**16**:7812–20.
3. Lécuyer E, Yoshida H, Parthasarathy N, *et al.* Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 2007;**131**:174–87.
4. Claußen M, Lingner T, Pommerenke C, *et al.* Global analysis of asymmetric RNA enrichment in oocytes reveals low conservation between closely related *Xenopus* species. *Mol Biol Cell* 2015;**26**:3777–87.
5. Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O. Translation-independent localization of mRNA in *E. coli*. *Science* 2011;**331**:1081–4.
6. Cajigas IJ, Tushev G, Will TJ, *et al.* The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* 2012;**74**:453–66.
7. Mili S, Moissoglu K, Macara IG. Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature* 2008;**453**:115–9.
8. Wojciechowska M, Krzyzosiak WJ. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet* 2011;**20**:3811–21.
9. Irastortza-Olaziregi M, Amster-Choder O. RNA localization in prokaryotes: where, when, how, and why. *WIREs RNA* 2021;**12**:e1615.
10. Cornelison GL, Levy SA, Jenson T, Frost B. Tau-induced nuclear envelope invagination causes a toxic accumulation of mRNA in *Drosophila*. *Aging Cell* 2019;**18**:e12847.
11. Butler MT, Wallingford JB. Planar cell polarity in development and disease. *Nat Rev Mol Cell Biol* 2017;**18**:375–88.
12. Xia C, Fan J, Emanuel G, *et al.* Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc Natl Acad Sci* 2019;**116**:19490–9.
13. Gadir N, Haim-Vilmovsky L, Kraut-Cohen J, Gerst JE. Localization of mRNAs coding for mitochondrial proteins in the yeast *Saccharomyces cerevisiae*. *RNA* 2011;**17**:1551–65.
14. van Heesch S, van Iterson M, Jacobi J, *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* 2014;**15**:R6.
15. Clark MB, Johnston RL, Inostroza-Ponta M, *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res* 2012;**22**:885–98.
16. Didiot M-C, Ferguson CM, Ly S, *et al.* Nuclear localization of huntingtin mRNA is specific to cells of neuronal origin. *Cell Rep* 2018;**24**:2553–2560.e5.
17. Chrisafis G, Wang T, Moissoglu K, *et al.* Collective cancer cell invasion requires RNA accumulation at the invasive front. *Proc Natl Acad Sci* 2020;**117**:27423–34.
18. Fallini C, Zhang H, Su Y, *et al.* The survival of motor neuron (SMN) protein interacts with the mRNA-binding protein HuD and regulates localization of poly(a) mRNA in primary motor neuron axons. *J Neurosci* 2011;**31**:3914–25.
19. Akten B, Kye MJ, Hao LT, *et al.* Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. *Proc Natl Acad Sci* 2011;**108**:10337–42.
20. Rossoll W, Jablonka S, Andreassi C, *et al.* Smn, the spinal muscular atrophy–determining gene product, modulates axon growth and localization of β-actin mRNA in growth cones of motoneurons. *J Cell Biol* 2003;**163**:801–12.
21. Fallini C, Rouanet JP, Donlin-Asp PG, *et al.* Dynamics of survival of motor neuron (SMN) protein interaction with the mRNA-binding protein IMP1 facilitates its trafficking into motor neuron axons: SMN controls IMP1 axonal localization. *Dev Neurobiol* 2014;**74**:319–32.
22. Fallini C, Donlin-Asp PG, Rouanet JP, *et al.* Deficiency of the survival of motor neuron protein impairs mRNA localization and local translation in the growth cone of motor neurons. *J Neurosci* 2016;**36**:3811–20.
23. Baleriola J, Walker CA, Jean YY, *et al.* Axonally synthesized ATF4 transmits a neurodegenerative signal across brain regions. *Cell* 2014;**158**:1159–72.
24. Bassell GJ, Warren ST. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron* 2008;**60**:201–14.

25. Dictenberg JB, Swanger SA, Antar LN, *et al.* A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. *Dev Cell* 2008;**14**:926–39.
26. Ivy AS, Rex CS, Chen Y, *et al.* Hippocampal dysfunction and cognitive impairments provoked by chronic early-life stress involve excessive activation of CRH receptors. *J Neurosci* 2010;**30**:13005–15.
27. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell* 2009;**136**: 777–93.
28. Cody NAL, Iampietro C, Lécuyer E. The many functions of mRNA localization during normal development and disease: from pillar to post. *Wiley Interdiscip Rev Dev Biol* 2013;**2**:781–96.
29. Chin A, Lécuyer E. RNA localization: making its way to the center stage. *Biochim Biophys Acta BBA - Gen Subj* 2017;**1861**: 2956–70.
30. Medioni C, Mowry K, Besse F. Principles and roles of mRNA localization in animal development. *Development* 2012;**139**: 3263–76.
31. Liu-Yesucevitz L, Bassell GJ, Gitler AD, *et al.* Local RNA translation at the synapse and in disease. *J Neurosci* 2011;**31**:16086–93.
32. Jung H, Yoon BC, Holt CE. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat Rev Neurosci* 2012;**13**:308–24.
33. Smith R. Moving molecules: mRNA trafficking in mammalian oligodendrocytes and neurons. *Neuroscientist* 2004;**10**:495–500.
34. Prasanth KV, Prasanth SG, Xuan Z, *et al.* Regulating gene expression through RNA nuclear retention. *Cell* 2005;**123**:249–63.
35. Kloc M, Dougherty MT, Bilinski S, *et al.* Three-dimensional ultrastructural analysis of RNA distribution within germinal granules of *Xenopus*. *Dev Biol* 2002;**241**:79–93.
36. Jenny A, Hachet O, Závorszky P, *et al.* A translation-independent role of *oskar* RNA in early *Drosophila* oogenesis. *Development* 2006;**133**:2827–33.
37. Tautz D, Pfeifle C. A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* 1989;**98**:81–5.
38. Macdonald PM, Struhl G. Cis- acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. *Nature* 1988;**336**:595–8.
39. Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. *Cell* 2009;**136**:719–30.
40. Buxbaum AR, Wu B, Singer RH. Single β-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* 2014;**343**:419–22.
41. Cabili MN, Dunagin MC, McClanahan PD, *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* 2015;**16**:20.
42. Liu X, Xu Y, Wang R, *et al.* A network-based algorithm for the identification of moonlighting noncoding RNAs and its application in sepsis. *Brief Bioinform* 2021;**22**:581–8.
43. Cheng L, Leung K-S. Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome. *Bioinformatics* 2018;**34**:3519–28.
44. Benoit Bouvrette LP, Cody NAL, Bergalet J, *et al.* CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in *drosophila* and human cells. *RNA* 2018;**24**:98–113.
45. Hansen TB, Jensen TI, Clausen BH, *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* 2013;**495**: 384–8.
46. Memczak S, Jens M, Elefsinioti A, *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013;**495**: 333–8.
47. Tripathi V, Ellis JD, Shen Z, *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* 2010;**39**:925–38.
48. McDermott SM, Davis I. Drosophila Hephaestus/polypyrimidine tract binding protein is required for dorso-ventral patterning and regulation of signalling between the germline and soma. *PloS One* 2013;**8**:e69978.
49. Clouse KN, Ferguson SB, Schüpbach T. Squid, cup, and PABP55B function together to regulate gurken translation in Drosophila. *Dev Biol* 2008;**313**:713–24.
50. Sladewski TE, Bookwalter CS, Hong M-S, Trybus KM. Single-molecule reconstitution of mRNA transport by a class V myosin. *Nat Struct Mol Biol* 2013;**20**:952–7.
51. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In *International conference on machine learning*. PMLR, 2017, 3319–28.
52. Engel KL, Arora A, Goering R, *et al.* Mechanisms and consequences of subcellular RNA localization across diverse cell types. *Traffic* 2020;**21**:404–18.
53. Patel VL, Mitra S, Harris R, *et al.* Spatial arrangement of an RNA zipcode identifies mRNAs under post-transcriptional control. *Genes Dev* 2012;**26**:43–53.
54. Oleynikov Y, Singer RH. Real-time visualization of ZBP1 association with β-actin mRNA during transcription and localization. *Curr Biol* 2003;**13**:199–207.
55. Ferrandon D, Elphick L, Nüsslein-Volhard C, St Johnston D. Staufen protein associates with the 3′ UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* 1994;**79**:1221–32.
56. Ferrandon D, Koch I, Westhof E, Nüsslein-Volhard C. RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFEN ribonucleoprotein particles. *EMBO J* 1997;**16**:1751–8.
57. Takizawa PA, Sil A, Swedlow JR, *et al.* Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. *Nature* 1997;**389**:90–3.
58. Goering R, Arora A, Taliaferro M. RNA localization mechanisms transcend cell morphology. *Elife* 2023;**12**:e80040.
59. Liao J-Y, Yang B, Zhang YC, *et al.* EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Res* 2020;**48**:D307–13.
60. Giudice G, Sánchez-Cabo F, Torroja C, Lara-Pezzi E. ATtRACT—a database of RNA-binding proteins and associated motifs. *Database* 2016;**2016**:baw035.
61. Cook KB, Kazan H, Zuberi K, *et al.* RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res* 2011;**39**:D301–8.
62. Macdonald PM, Kerr K, Smith JL, Leask A. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development* 1993;**118**:1233–43.
63. Savulescu AF, Jacobs C, Negishi Y, *et al.* Pinpointing cell identity in time and space. *Front Mol Biosci* 2020;**7**:209.
64. Moor AE, Harnik Y, Ben-Moshe S, *et al.* Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell* 2018;**175**:1156–1167.e15.
65. Das S, Vera M, Gandin V, *et al.* Intracellular mRNA transport and localized translation. *Nat Rev Mol Cell Biol* 2021;**22**:483–504.
66. Liao Y-C, Fernandopulle MS, Wang G, *et al.* RNA granules hitchhike on lysosomes for long-distance transport, using Annexin A11 as a molecular tether. *Cell* 2019;**179**:147–164.e20.
67. Baumann S, Ko J. Endosomal transport of septin mRNA and protein indicates local translation on endosomes and is required for correct septin filamentation. *EMBO Rep* 2013;**15**: 94–102.

68. Cohen B, Altman T, Golani-Armon A, *et al.* Co-transport of the nuclear-encoded *Cox7c* mRNA with mitochondria along axons occurs through a coding-region-dependent mechanism. *J Cell Sci* 2022;**135**:jcs259436.
69. Castellana M, Hsin-Jung Li S, Wingreen NS. Spatial organization of bacterial transcription and translation. *Proc Natl Acad Sci* 2016;**113**:9286–91.
70. Forrest KM, Gavis ER. Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. *Curr Biol* 2003;**13**:1159–68.
71. Yoon YJ, Wu B, Buxbaum AR, *et al.* Glutamate-induced RNA localization and translation in neurons. *Proc Natl Acad Sci* 2016;**113**:E6877–86.
72. Park HY, Lim H, Yoon YJ, *et al.* Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* 2014;**343**:422–4.
73. Turner-Bridger B, Jakobs M, Muresan L, *et al.* Single-molecule analysis of endogenous β-actin mRNA trafficking reveals a mechanism for compartmentalized mRNA localization in axons. *Proc Natl Acad Sci* 2018;**115**:E9697–706.
74. Baumann S, Komissarov A, Gili M, *et al.* A reconstituted mammalian APC-kinesin complex selectively transports defined packages of axonal mRNAs. *Sci Adv* 2020;**6**:eaaz1588.
75. Beach DL, Salmon ED, Bloom K. Localization and anchoring of mRNA in budding yeast. *Curr Biol* 1999;**9**:569–S1.
76. Bertrand E, Chartrand P, Schaefer M, *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol Cell* 1998;**2**:437–45.
77. Hocine S, Raymond P, Zenklusen D, *et al.* Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat Methods* 2013;**10**:119–21.
78. Wu B, Chen J, Singer RH. Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Sci Rep* 2014;**4**:3615.
79. Long RM, Gu W, Lorimer E, *et al.* She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to ASH1 mRNA. *EMBO J* 2000;**19**:6592–601.
80. Haim-Vilmovsky L, Gadir N, Herbst RH, Gerst JE. A genomic integration method for the simultaneous visualization of endogenous mRNAs and their translation products in living yeast. *RNA* 2011;**17**:2249–55.
81. Raj A, van den Bogaard P, Rifkin SA, *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* 2008;**5**:877–9.
82. Lohoff T, Ghazanfar S, Missarova A, *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat Biotechnol* 2022;**40**:74–85.
83. Eng C-HL, Lawson M, Zhu Q, *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019;**568**:235–9.
84. Mas-Ponte D, Carlevaro-Fita J, Palumbo E, *et al.* LncATLAS database for subcellular localization of long noncoding RNAs. *RNA* 2017;**23**:1080–7.
85. Wang Y, Zhu W, Levy DE. Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR. *Methods* 2006;**39**:356–62.
86. Zhang T, Tan P, Wang L, *et al.* RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res* 2017;**45**:D135–8.
87. Cui T, Dou Y, Tan P, *et al.* RNALocate v2.0: an updated resource for RNA subcellular localization with increased coverage and annotation. *Nucleic Acids Res* 2022;**50**:D333–9.
88. Yue L, Liu F, Hu J, *et al.* A guidebook of spatial transcriptomic technologies, data resources and analysis approaches. *Comput Struct Biotechnol J* 2023;**21**:940–55.
89. Shah S, Lubeck E, Zhou W, Cai L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* 2016;**92**:342–57.
90. Shah S, Takei Y, Zhou W, *et al.* Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* 2018;**174**:363–376.e16.
91. Shi H, He Y, Zhou Y, *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *BioRxiv* 2022.06.20.496914. 2022.
92. He S, Bhatt R, Brown C, *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol* 2022;**40**:1794–806.
93. Su J-H, Zheng P, Kinrot SS, *et al.* Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* 2020;**182**:1641–1659.e26.
94. Chen A, Liao S, Cheng M, Brown C, *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022;**185**:1777–92.e21. http://biorxiv.org/lookup/doi/10.1101/2021.01.17.427004.
95. Goh JJL, Chou N, Seow WY, *et al.* Highly specific multiplexed RNA imaging in tissues with split-FISH. *Nat Methods* 2020;**17**:689–93 .
96. Zeng H, Huang J, Zhou H, *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat Neurosci* 2023;**26**(3):430–46.
97. Borm LE, Mossi Albiach A, Mannens CCA, *et al.* Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat Biotechnol* 2023;**41**(2):222–31.
98. Holt CE, Bullock SL. Subcellular mRNA localization in animal cells and why it matters. *Science* 2009;**326**:1212–6.
99. Paquin N, Chartrand P. Local regulation of mRNA translation: new insights from the bud. *Trends Cell Biol* 2008;**18**:105–11.
100. Shestakova EA, Singer RH, Condeelis J. The physiological significance of β-actin mRNA localization in determining cell polarity and directional motility. *Proc Natl Acad Sci* 2001;**98**:7045–50.
101. Adereth Y, Dammai V, Kose N, *et al.* RNA-dependent integrin α3 protein localization regulated by the muscleblind-like protein MLP1. *Nat Cell Biol* 2005;**7**:1240–7.
102. Imbert A, Ouyang W, Safieddine A, *et al.* FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* 2022;**28**:786–95.
103. Walter FC, Stegle O, Velten B. FISHFactor: a probabilistic factor model for spatial transcriptomics data with subcellular resolution. *Bioinformatics* 2023;**39**:btad183.
104. Samacoits A, Chouaib R, Safieddine A, *et al.* A computational framework to study sub-cellular RNA localization. *Nat Commun* 2018;**9**:4584.
105. Mah CK, *et al.* Bento: a toolkit for subcellular analysis of spatial transcriptomics data. *BioRxiv*2022.06. 10.495510 2022.
106. Cao Z, Pan X, Yang Y, *et al.* The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* 2018;**34**:2185–94.
107. Zhang Z-Y, Yang YH, Ding H, *et al.* Design powerful predictor for mRNA subcellular location prediction in *Homo sapiens*. *Brief Bioinform* 2021;**22**:526–35.
108. Chen W, Lei T-Y, Jin D-C, *et al.* PseKNC: a flexible web server for generating pseudo K-tuple nucleotide composition. *Anal Biochem* 2014;**456**:53–60.
109. Chou K-C. Prediction of protein cellular attributes using pseudo-amino acid composition. *Proteins Struct Funct Bioinforma* 2001;**43**:246–55.

110. Meher PK, Satpathy S, Rao AR. miRNALoc: predicting miRNA subcellular localizations based on principal component scores of physico-chemical properties and pseudo compositions of di-nucleotides. *Sci Rep* 2020;**10**:14557.
111. Garg A, Singhal N, Kumar R, Kumar M. mRNALoc: a novel machine-learning based in-silico tool to predict mRNA subcellular localization. *Nucleic Acids Res* 2020;**48**:W239–43.
112. Yang Y, Fu X, Qu W, *et al.* MiRGOFS: a GO-based functional similarity measurement for miRNAs, with applications to the prediction of miRNA subcellular localization and miRNA–disease association. *Bioinformatics* 2018;**34**:3547–56.
113. Li J, Zhang L, He S, *et al.* SubLocEP: a novel ensemble predictor of subcellular localization of eukaryotic mRNA based on machine learning. *Brief Bioinform* 2021;**22**:bbaa401.
114. Asim MN, Ibrahim MA, Zehe C, *et al.* L2S-MirLoc: a lightweight two stage MiRNA sub-cellular localization prediction framework. In: *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, SZ, China: IEEE, 2021, pp. 1–8. https://doi.org/10.1109/IJCNN52387.2021.9534015.
115. Tang Q, Nie F, Kang J, Chen W. mRNALocater: enhance the prediction accuracy of eukaryotic mRNA subcellular localization by using model fusion strategy. *Mol Ther* 2021;**29**:2617–23.
116. Feng S, Liang Y, Du W, *et al.* LncLocation: efficient subcellular location prediction of long non-coding RNA-based multi-source heterogeneous feature fusion. *Int J Mol Sci* 2020;**21**:7271.
117. Zeng M, Wu Y, Lu C, *et al.* DeepLncLoc: a deep learning framework for long non-coding RNA subcellular localization prediction based on subsequence embedding. *Brief Bioinform* 2022;**23**:bbab360.
118. Asim MN, Malik MI, Zehe C, *et al.* MirLocPredictor: a ConvNet-based multi-label microRNA subcellular localization predictor by incorporating k-mer positional information. *Genes* 2020;**11**:1475.
119. Wang D, Zhang Z, Jiang Y, *et al.* DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism. *Nucleic Acids Res* 2021;**49**:e46–6.
120. Yan Z, Lécuyer E, Blanchette M. Prediction of mRNA subcellular localization using deep recurrent neural networks. *Bioinformatics* 2019;**35**:i333–42.
121. Lin Y, Pan X, Shen H-B. lncLocator 2.0: a cell-line-specific subcellular localization predictor for long non-coding RNAs with interpretable deep learning. *Bioinformatics* 2021;**37**:2308–16.
122. Xiao Y, Cai J, Yang Y, *et al.* Prediction of microRNA subcellular localization by using a sequence-to-sequence model. In: *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 1332–7. IEEE, Singapore, 2018.
123. Dubois R, Imbert A, Samacoïts A, *et al.* A deep learning approach to identify mRNA localization patterns. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice, Italy: IEEE, 2019:1386–90.
124. Asim MN, Ibrahim MA, Malik MI, *et al.* EL-RMLocNet: an explainable LSTM network for RNA-associated multi-compartment localization prediction. *Comput Struct Biotechnol J* 2022;**20**:3986–4002.
125. Su ZD, Huang Y, Zhang ZY, *et al.* iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics* 2018;**34**(24):4196–204.
126. Gudenas BL, Wang L. Prediction of LncRNA subcellular localization with deep learning from sequence features. *Sci Rep* 2018;**8**(1):16385.
127. Ahmad A, Lin H, Shatabda S. Locate-R: subcellular localization of long non-coding RNAs using nucleotide compositions. *Genomics* 2020;**112**(3):2583–89.
128. Zhang S, Qiao H. KD-KLNMF: identification of lncRNAs subcellular localization with multiple features and nonnegative matrix factorization. *Anal Biochem* 2020;**610**:113995.
129. Jeon YJ, Hasan MM, Park HW, *et al.* TACOS: a novel approach for accurate prediction of cell-specific long noncoding RNAs subcellular localization. *Brief Bioinform* 2022;**23**(4):bbac243.
130. Yuan GH, Wang Y, Wang GZ, *et al.* RNAlight: a machine learning model to identify nucleotide features determining RNA subcellular localization. *Brief Bioinform* 2023;**24**(1):bbac509.
131. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020;**21**:6.
132. Savulescu AF, Bouilhol E, Beaume N, Nikolski M. Prediction of RNA subcellular localization: learning from heterogeneous data sources. *iScience* 2021;**24**:103298.
133. Ardila D, Kiraly AP, Bharadwaj S, *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019;**25**:954–61.
134. Fujioka T, Kubota K, Mori M, *et al.* Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. *Jpn J Radiol* 2019;**37**:466–72.
135. van Grinsven MJJP, van Ginneken B, Hoyng CB, *et al.* Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. *IEEE Trans Med Imaging* 2016;**35**:1273–84.
136. Burlina PM, Joshi N, Pekala M, *et al.* Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol* 2017;**135**:1170–6.
137. De Fauw J, Ledsam JR, Romera-Paredes B, *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;**24**:1342–50.
138. Peng Y, Dharssi S, Chen Q, *et al.* DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology* 2019;**126**:565–75.
139. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 2016;**6**:26094.
140. Wickramasinghe N. A convolutional net for medical records. *Eng Med Biol Soc* 2017;**21.1**:22–30.
141. Tran T, Nguyen TD, Phung D, Venkatesh S. Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). *J Biomed Inform* 2015;**54**:96–105.
142. Kwon BC, Choi MJ, Kim JT, *et al.* RetainVis: visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Trans Vis Comput Graph* 2019;**25**:299–309.
143. Boehm KM, Khosravi P, Vanguri R, *et al.* Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer* 2022;**22**:114–26.
144. Cui C, Yang H, Wang Y, *et al.* Deep multi-modal fusion of image and non-image data in disease diagnosis and prognosis: a review. *Prog Biomed Eng* 2023.
145. Ngiam J, Khosla A, Kim M, *et al.* Multimodal deep learning. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Bellevue, WA, USA, 2011, pp. 689–96.

146. Zhang L, Lv C, Jin Y, *et al.* Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. *Front Genet* 2018;**9**:477.
147. Huang Z, Zhan X, Xiang S, *et al.* SALMON: survival analysis learning with multi-omics neural networks on breast cancer. *Front Genet* 2019;**10**:166.
148. Chen RJ, Lu MY, Wang J, *et al.* Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imaging* 2020;**41**:757–70.
149. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE J Biomed Health Inform* 2019;**23**: 538–46.
150. Holste G, Partridge SC, Rahbar H, *et al.* End-to-end learning of fused image and non-image features for improved breast cancer classification from MRI. In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, IEEE, 2021, pp. 3287–96. https://doi.org/10.1109/ICCVW54120.2021.00368.
151. Wang W, Tran D, Feiszli M. What makes training multi-modal classification networks hard? In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2020, pp. 12692–702. https://doi.org/10.1109/CVPR42600.2020.01271.
152. Jaegle A, Gimeno F, Brock A, *et al.* Perceiver: general perception with iterative attention. In: *International Conference on Machine Learning*, pp. 4651–64 (PMLR, 2021).
153. Fayek HM, Kumar A. Large scale audiovisual learning of sounds with weakly labeled data. ArXiv Prepr. ArXiv200601595. 2020.
154. Alwassel H, Mahajan D, Korbar B, *et al.* Self-supervised learning by cross-modal audio-video clustering. *Adv Neural Inf Process Syst* 2020;**33**:9758–70.
155. Aytar Y, Vondrick C, Torralba A. Soundnet: learning sound representations from unlabeled video. *Adv Neural Inf Process Syst* 2016;**29**.892–900.
156. Chen F-L, Zhang DZ, Han ML, *et al.* VLP: a survey on vision-language pre-training. *Mach Intell Res* 2023;**20**:38–56.
157. Hosny A, Schwier M, Berger C, *et al.* ModelHub.AI: dissemination platform for deep learning models. *arXiv* 2019;**12**:1911.13218.
158. Thumuluri V, Almagro Armenteros JJ, Johansen AR, *et al.* DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res* 2022;**50**: W228–34.
159. Kobayashi H, Cheveralls KC, Leonetti MD, Royer LA. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat Methods* 2022;**19**:995–1003.
160. Wu Q, Zheng X, Leung KS, *et al.* meGPS: a multi-omics signature for hepatocellular carcinoma detection integrating methylome and transcriptome data. *Bioinformatics* 2022;**38**:3513–22.