RNA imaging in bacteria
Sara Rombouts, Marcelo Nollmann

To cite this version:
Sara Rombouts, Marcelo Nollmann. RNA imaging in bacteria. FEMS Microbiology Reviews, Wiley-Blackwell, 2020, 10.1093/femsre/fuaa051. hal-02991916

HAL Id: hal-02991916
https://hal.umontpellier.fr/hal-02991916
Submitted on 6 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
RNA imaging in bacteria

Sara Rombouts¹, Marcelo Nollmann¹,²

¹ Centre de Biochimie Structurale, CNRS UMR 5048, INSERM U1054, Université de Montpellier, 60 rue de Navacelles, 34090, Montpellier, France

² To whom correspondence should be addressed:

Marcelo Nollmann: marcelo.nollmann@cbs.cnrs.fr
Abstract

The spatiotemporal regulation of gene expression plays an essential role in many biological processes. Recently, several imaging-based RNA labeling and detection methods, both in fixed and live cells, were developed and now enable the study of transcript abundance, localization and dynamics. Here, we review the main single-cell techniques for RNA visualization with fluorescence microscopy, and describe their applications in bacteria.

One sentence summary: Recent advances in RNA labeling made it possible to study transcript abundance, localization and dynamics in bacteria.

Introduction

RNA is one of the main actors in the cell, allowing the flow of information between DNA, the carrier of genetic material, and proteins (Crick 1970). RNA transcripts are produced from double-stranded DNA by RNA polymerases in the process of transcription (Figure 1). However, not all transcripts are messenger RNA (mRNA) and translated into proteins (Wagner and Romby 2015). In fact, most of the transcribed RNAs in bacteria are ribosomal RNAs (rRNA), transfer RNA (tRNA) and other non-coding or regulatory small RNAs (Figure 1) (Argaman et al. 2001; Rivas et al. 2001; Wassarman et al. 2001). These RNAs do not directly result into proteins, however they influence the composition of the transcriptome and proteome of a cell by fine-tuning the expression of specific genes in response to environmental cues (Wagner and Romby 2015; Waters and Storz 2009; Storz, Vogel, and Wassarman 2011; Papenfort and Vogel 2009; Gottesman and Storz 2011; Gottesman 2004). Regulation of transcription plays a crucial role in many biological processes such as development, cell differentiation, and cell homeostasis (Browning and Busby 2016; Pope and Medzhitov 2018). Consequently, the detection and precise quantification of RNA levels in cells is important to understand how transcription is regulated, as well as to decipher the specific transcriptional patterns of different cell types.

Early pioneering techniques, such as northern blotting, qPCR and microarrays, enabled the detection and quantification of specific transcripts from a pool of RNA, and proved powerful in species identification and in differential gene expression studies (Strait et al. 2009; Jauregui et al. 2001; Sachse et al. 2005). Subsequently, high-density DNA microarrays and high-throughput RNA sequencing (RNAseq) were developed to detect whole transcriptomes (Lashkari et al. 1997; Bainbridge et al. 2006; Weber et al. 2007). These methods rely on the extraction of total RNA from a pool of cultured cells or a homogenized cell solution, and thus, result in an ensemble average for the population of cells under study. Recent advances in fractionation techniques and RNA sequencing now enable transcriptome profiling of single cells (scRNAseq), allowing further exploration of cell-to-cell variability and subpopulation identification (Xi Chen, Teichmann, and Meyer 2018; Montoro et al. 2018; Shalek et al. 2014). A major drawback of sequencing-based technologies still relies on their inability to monitor single-cell dynamics and sub-cellular localization. These measurements proved to be critical to understand the mechanism of transcriptional regulation and the functions of RNAs in the cell.
A major advantage of imaging-based methods is that cellular processes and their associated actors can be visualized in time and space within their native context. This ability, when combined with the development of a myriad of strategies to tag most macromolecules within a cell, has made fluorescence microscopy a popular tool for the study of many biological problems. Particularly, imaging-based methods have provided considerably new insight into the process of transcriptional regulation, and more generally into the cellular roles of different RNA species.

In eukaryotic cells, transcription occurs in the nucleus and translation in the cytoplasm. As a result, these processes are decoupled in space and time. The translocation and final positioning of RNA molecules within the cytoplasm plays a regulatory role in asymmetric cell division, cell migration and development (Chin and Lécuyer 2017; Buxbaum, Haimovich, and Singer 2015). In bacteria, on the other hand, the nucleoid and cytoplasm coexist within the same compartment. In addition, transcription and translation are often coupled and occur in the same nucleo-plasmic space (Bakshi, Choi, and Weisshaar 2015; Bakshi et al. 2012; Fan et al. 2017). This led to the early assumption that, in bacteria, RNA species lacked spatial compartmentalization or preferential sub-cellular localization. However, with the emergence of novel imaging-based techniques, several studies recently evidenced that RNA molecules as well as proteins involved in their processing can indeed exhibit specific spatial patterns of subcellular localization (Weng et al. 2019; Moffitt et al. 2016; Montero Llopis et al. 2010; Nevo-Dinur et al. 2011). In addition, the ability of imaging-based methods to monitor changes in protein and transcript abundance and localization was fundamental to characterize the dynamics of transcription. This ability was key to characterize the different sources of noise in gene expression (Ozbudak et al. 2002; Elowitz 2002; Rosenfeld et al. 2005) or to understand the mechanism of promoter specificity (Ferguson et al. 2012).

In this review, we will provide an overview of the most important and most recently developed single-cell approaches for the visualization of RNA with fluorescence microscopy, with a particular focus on those methods that were used in bacteria or that promise to be relevant to the study of transcriptional regulation in bacteria. We will describe techniques for the read-out of RNA molecules in fixed cells, which have mostly helped to understand the abundance and subcellular spatial organization of RNAs. In addition, we will highlight the most established techniques for imaging of RNA molecules in live cells, which enabled the description of gene expression and transcript dynamics in space and time. Ultimately, we will illustrate the recent advances made towards the simultaneous visualization of multiple RNA species, approaching whole-transcriptome imaging.

**Fixed-cell Imaging**

Fluorescence In Situ Hybridization (FISH) is one of the most powerful techniques for the visualization of nucleic acid molecules in fixed cells. This technique relies on the hybridization of single-stranded, fluorescently-labeled oligonucleotide probes to a complementary DNA or RNA sequence (Raj and van Oudenaarden 2009; Femino 1998). Although the technique was initially developed for DNA labeling (Langer-Safer, Levine, and
Ward 1982), it is now the golden standard for the imaging of native transcripts, especially for mRNA quantification and localization studies.

Initially, the detection of single mRNA molecules by single-molecule FISH (smFISH) was challenging due to the low signal to noise ratio of the labeled mRNA with respect to the background signal arising from freely-diffusing unbound probes and from non-specific binding or sticking of the probe to other cellular components (Femino 1998). To overcome this problem, the fluorescent signal emanating from a single RNA molecule needs to largely exceed the background signal. This problem was solved by increasing the number of fluorophores used to label a single mRNA molecule. Signal amplification can be achieved by the use of multiple, short, non-overlapping, fluorescently-labeled oligonucleotides probes (Raj et al. 2008) (Figure 2A). The image of this single transcript then appears as a diffraction-limited spot, because the multiple oligonucleotide probes remain closer than the diffraction limit of light (~250 nm). Thus, by counting the number of diffraction-limited spots per cell one can measure the number of single mRNA molecules in a given cell for a specific mRNA species. This approach has proven very powerful for the labeling of mRNA in eukaryotes and for the detection of different kinds of RNA in E. coli and B. subtilis (Maamar, Raj, and Dubnau 2007; Arbel-Goren et al. 2016; Fei et al. 2015; Skinner et al. 2013; So et al. 2011; Sepúlveda et al. 2016) (Table 1) (Figure 2B). However, bacterial genes are often short (~ 1kb or less), thus limiting the number of distinct fluorescent oligonucleotide probes that can encode a single RNA species. This results in decreased detection efficiencies, and calls for other experimental strategies.

Indirect labeling strategies allow for more complex read-out schemes to increase the number of fluorophores per oligonucleotide probe. Generally, a primary probe is designed to contain two essential regions: i) a complementary region to the RNA of interest, and ii) a tail sequence. For example, hybridization chain reaction (HCR) relies on the binding of the primary probe, containing an HCR initiator sequence in the tail, to the mRNA of interest (Choi et al. 2010; Choi, Beck, and Pierce 2014; R. M. Dirks and Pierce 2004; Shah et al. 2016). This initiator sequence triggers the self-assembly of two fluorescently-labeled metastable hairpins to a long polymer chain in which many fluorophores can be incorporated (Figure 2A). When using the HCR method, care needs to be taken with non-specific probe hybridization to limit artefacts (Shah et al. 2016; Choi et al. 2018). In fact, off-target binding of probes to sequences containing the initiator sequence will also trigger the HCR, thus amplifying the background. Another strategy, based on the concept of rolling circle amplification (RCA), uses a padlock probe to initiate the amplification reaction (Larsson et al. 2010, 2004). In short, an LNA primary probe is used to target an mRNA and initiate a reverse transcriptase reaction resulting in an RNA-cDNA hybrid. This hybrid will be partially digested by Rnase H, allowing the binding of a padlock probe to the single-stranded cDNA. The padlock probe is ligated after hybridization and serves as a primer for the RCA reaction. Then, multiple fluorescently-labeled oligonucleotide probes can hybridize to sequence repeats within the RCA product (Figure 2A). A major advantage of both HCR and RCA is that the signal can be tuned and amplified as much as desired (Shah et al. 2016; Larsson et al. 2010, 2004). This, however, requires the precise modulation of self-assembly or polymerization times.
A more straightforward approach is branched DNA amplification (Xia et al. 2019; Player et al. 2001). Here, the primary probe has a tail on both sides of the complementary sequence, each of which binds a primary amplifier. Such an amplifier sequence is made out of a sequence hybridizing to the tail of the primary probe and to several repeats, each of which can bind a secondary amplifier (Figure 2A). The secondary amplifier also contains several repeats that can bind a tertiary amplifier and so on, or a fluorescently-labeled oligonucleotide probe. Thus, in this method the degree of signal amplification is defined by the bDNA design.

Signal amplification methods, such as bDNA, HCR and RCA, are powerful strategies to upscale the fluorescent signal emitted from the labeled target transcript and thus, to increase signal to noise ratios. However, care needs to be taken when using amplified signals for the localization and quantification of transcripts (Shah et al. 2016). As signal amplification generates foci of increased size, caution is required to avoid compromising the localization accuracy. Additionally, non-linear signal amplification can complicate the quantification of RNA copy number as the brightness of a fluorescent spot no longer correlates with the number of RNA molecules localized in that spot.

Although the above mentioned amplification approaches enable the detection of short RNA species, the applications in prokaryotes so far remain limited. Only HCR has been successfully implemented in bacteria for species detection in complex samples and for the study of gene expression (Table 1) (Nikolakakis et al. 2015; Yamaguchi et al. 2015).

An additional complication to the accurate counting of single RNA molecules appears for highly expressed transcripts. In conventional fluorescence microscopy, the signal of one transcript labeled with multiple oligonucleotide probes results in a diffraction-limited spot. When transcripts are spatially well separated, they can be imaged, localized and quantified as distinct single molecules. However, when the RNA density increases, the fluorescence signal from several RNA molecules can overlap in space, making it impossible to resolve them as single transcripts. This often represents a major problem in bacteria, as their reduced cellular volume leads to considerably high global RNA densities. For this reason, signal overlapping occurs for many abundant RNA species, limiting detection with diffraction-limited microscopies to low abundance RNA species (G. Wang, Moffitt, and Zhuang 2018). Several groups have circumvented this issue by estimating mRNA copy numbers from the total fluorescence signal for highly expressed target mRNA molecules in E. coli (Taniguchi et al. 2010; Skinner et al. 2013). Here, the fluorescence brightness of a single molecule is used to estimate the number of mRNA molecules from the total fluorescence signal in a given cell. Alternatively, this issue can be circumvented by using super-resolution techniques able to resolve smaller detection volumes and thus enable detection of single mRNA molecules in crowded environments (Lubeck and Cai 2012; G. Wang, Moffitt, and Zhuang 2018). Both Structured Illumination Microscopy (SIM) (Gustafsson 2000) and Stimulated-Emission Depletion Microscopy (STED) (Hell and Wichmann 1994) have been employed to resolve RNA species labeled by smFISH in eukaryotic cells (Mito et al. 2016; W. I. Zhang et al. 2014). In bacteria, however, localization methods such as Stochastic Optical Resolution Microscopy (STORM) were more often used (Rust, Bates, and Zhuang 2006). By combining three-dimensional (3D)-STORM and whole-transcriptome FISH labeling, Moffitt and coworkers were able to demonstrate that the bacterial transcriptome in E. coli is spatially organized (Table 1) (Figure 2C) (Moffitt et al.
2016). Similarly, Fei and colleagues combined 3D-STORM with smFISH to label the bacterial small RNA (sRNA) SgrS and showed that the localization of SgrS depends on its expression level (Table 1) (Fei et al. 2015). More recently, a novel technique called Expansion Microscopy (ExM) was developed (F. Chen, Tillberg, and Boyden 2015). The ExM approach relies on the embedding of the biological sample in a polymer hydrogel and the subsequent swelling of the hydrogel by immersion in a liquid. This results in a physical magnification of the embedded sample, making it possible to image cellular components at high resolution. ExM has been proven successful for imaging FISH-labeled RNA species in eukaryotes (F. Chen et al. 2016; G. Wang, Moffitt, and Zhuang 2018; Asano et al. 2018; Tsanov et al. 2016). In bacteria, the technique has been validated, but no study has reported the imaging of RNA content with ExM yet (Lim et al. 2019).

Even though techniques for fixed cell imaging enable the quantification and study of gene expression in space, the dynamics of gene expression remain unresolved with these approaches. Excitingly, a number of alternative technologies have been developed to enable transcript imaging in live cells.

**Live cell imaging**

Live cell imaging of RNA requires the labeling of transcripts *in vivo*. Many different strategies have been developed over the years to target specific mRNA molecules, involving the binding of either a protein or a fluorogenic dye to an RNA-aptamer. The first of such strategies was developed in 1998 and makes use of the MS2 phage system to target a selected transcript (Bertrand et al. 1998). In short, a dimer of the MS2 phage coat protein (MCP) binds to one hairpin loop on the phage RNA (Bertrand et al. 1998; Beach, Salmon, and Bloom 1999; Peabody 1993; Lowary and Uhlenbeck 1987). The high specificity of MCP binding to these hairpins was exploited to label other RNA species. Bertrand and coworkers fused multiple repeats of the MS2 RNA hairpin to the 3' UTR of a plasmid-encoded ASH1 gene from *Saccharomyces cerevisiae* (Bertrand et al. 1998). A GFP fusion of the MS2 protein was then used to detect the ASH1 transcript (Bertrand et al. 1998) (Figure 3A) and in later studies to follow it in space and time (Ido Golding et al. 2005) (Figure 3B). Genomic MS2-tagging of transcripts is now more regularly used (Tutucci et al. 2018). Aside from its application to yeast, the MS2 system was further developed for other systems, including bacteria (Fusco et al. 2003; Rook, Lu, and Kosik 2000; F. Zhang and Simon 2003; I. Golding and Cox 2004; Forrest and Gavis 2003; Campbell et al. 2015; Ido Golding et al. 2005; So et al. 2011). The first study in bacteria used 96 MS2 binding sites fused to the target RNA to investigate transcript kinetics in *E. coli* (Table 1) (I. Golding and Cox 2004). The results of this study uncovered cellular fluctuations in gene expression in single cells, supporting the hypothesis of stochastic transcription (Ido Golding et al. 2005). The MS2 system has also proven to be a valuable tool for transcript localization in *E. coli* and other bacteria such as *Bacillus subtilis* and *Lactococcus lactis* (Table 1) (Nevo-Dinur et al. 2011; van Gijtenbeek et al. 2016; Santos et al. 2012). A major drawback of this approach to study transcript diffusion and localization is, however, that transcript lengthening can decrease the diffusion rate of the targeted RNA species, and ultimately limit the applicability of this approach to highly dynamic transcripts (I. Golding and Cox 2004; Ido Golding et al. 2005; Montero Llopis et al. 2010; Garcia and Parker 2015). Additionally, the high levels of expression of the MS2-GFP
fusion lead to high levels of background signal (from freely diffusing MS2-GFP), making it difficult to distinguish RNA-specific signals from background fluorescence. In eukaryotes, this problem can be mitigated by targeting the unbound protein to other cellular compartments (Bertrand et al. 1998). Unfortunately, this approach is not feasible in bacteria, because bacteria lack canonical membrane-bound organelles. Another way of circumventing this issue, which was successfully used in both prokaryotic and eukaryotic systems, is to tweak the expression ratio of the reporter protein to its binding sites on the RNA (Fusco et al. 2003; Wu, Chao, and Singer 2012; Nevo-Dinur et al. 2011; van Gijtenbeek et al. 2016; Santos et al. 2012).

An orthogonal approach, involving the use of fluorogenic dyes for the labeling of specific RNA-aptamers was developed to circumvent the problem of background fluorescence. Fluorogenic dyes are molecules that display low levels of fluorescence when unbound and undergo a shift in fluorescence intensity upon binding to the RNA-aptamer (Bouhedda et al. 2020; Wirth et al. 2019). This aptamer sequence is then tethered to the RNA of interest as a genetically-encoded tag. With this method, several RNA aptamers were developed, such as Spinach, Spinach2, and Broccoli, all binding 3,5-difluoro-4-hydroxybenzylidene (DFHBI), a non-toxic, cell-permeable dye (Table 1) (Paige, Wu, and Jaffrey 2011; Strack and Jaffrey 2015; Strack, Disney, and Jaffrey 2013; Filonov et al. 2014) (Figure 3A). Both aptamer-dye systems were validated in E. coli (Filonov and Jaffrey 2016; J. Zhang et al. 2015) (Figure 3C). In addition, Mango and Corn are aptamers that bind either thiazole orange or DFHO, respectively (Table 1) (Dolgosheina et al. 2014; Autour et al. 2018; Song et al. 2017). The Mango and Corn RNA aptamers were also validated in E. coli (Dolgosheina et al. 2014; Song et al. 2017). In addition, the Mango RNA aptamer was used in combination with the Spinach RNA aptamer in an interesting method called apta-FRET (Table 1) (Jepsen et al. 2018). Here, the Mango and Spinach RNA aptamers were placed in close proximity on a tRNA scaffold and the FRET-signal generated when these aptamers are in close proximity were used to report on tRNA conformational changes. The concept of apta-FRET is an interesting development for the field of nanotechnology, as it allows to study the specific folding and dynamics of RNA nanostructures in living cells.

To find the appropriate sequences that will bind the dye molecule of interest, a technique called Systematic Evolution of Ligands by Exponential Enrichment (SELEX) was developed (Holeman et al. 1998; Stoltenburg, Reinemann, and Strehlitz 2007; Ellington and Szostak 1990). This method allows for the selection, identification and amplification of the most optimal binding RNA aptamer from a random library. A drawback of SELEX is the suboptimal binding of target RNA sequences. This was partially solved by microfluidics-assisted in vitro compartmentalization (micro-IVC), a technique relying on the use of selection pressure to find high-affinity aptamer sequences (Ryckelynck et al. 2015; Autour, Westhof, and Ryckelynck 2016; Autour et al. 2018). Aside from the use of fluorogenic dyes, photostable organic dyes that are conditionally quenched were also used. Examples are the DNB and SRB-2 aptamers that bind a myriad of fluorophores quenched by dinitroaniline, and aptamers binding Black Hole Quenchers (BHQ) attached to fluorophores (Sunbul and Jäschke 2018; Arora, Sunbul, and Jäschke 2015; Sunbul and Jäschke 2013a; Sato et al. 2015; Murata et al. 2011) (Figure 3A). In E. coli, both the DNB and SRB-2 aptamers were validated (Table 1) (Arora, Sunbul, and Jäschke 2015; Sunbul and Jäschke 2018, [a] 2013). (Arora, Sunbul, and Jäschke 2015; Sunbul and Jäschke 2018, [b]...
2013) simultaneous labeling and detection of two different RNA species (Arora, Sunbul, and Jäschke 2015).

Recent advances in the RNA aptamer field have enabled the development of aptamer-binding silicon rhodamine-derived fluorophores in the far-red and near-infrared (Autour et al. 2018; Yerramilli and Kim 2018; Braselmann et al. 2018). These new dyes promise a significant reduction in phototoxicity and autofluorescence for live-cell imaging (Li et al. 2019; Umezawa, Citterio, and Suzuki 2014), and allowed super-resolution imaging of RNAs in *E. coli* by STED microscopy (Table 1) (Wirth et al. 2019). Furthermore, spontaneously blinking rhodamine-derived fluorophores were used in other super-resolution microscopy techniques such as dSTORM (Lukinavičius et al. 2013; Uno et al. 2014). Unfortunately, such fluorophores were not yet reported in bacterial systems. A disadvantage of organic dyes is that, unlike fluorescent proteins, they are exogenous to cells, and rely mainly on the diffusion through the cell membrane to enter the cell. This diffusion can be slow and challenging for certain bacterial species. Additionally, many dye molecules are toxic to cells and exhibit strong phototoxic effects (Bouhedda, Autour, and Ryckelynck 2017; Hilderbrand and Weissleder 2010; Martynov et al. 2016).

Avoiding the modification of the RNA template enables the study of native transcripts and thus, the endogenous behavior of transcription. This was achieved by resorting to the use of the Pumilio homology domain (PumHD), a protein that binds mRNA in a sequence-specific manner (Ozawa et al. 2007; Xiaojiang Wang et al. 2002; Filipovska et al. 2011; M. Chen et al. 2017). Recently, PumHD was engineered to generate four protein modules, each recognizing a different RNA base. These building blocks can then be concatenated to form a chain of desired composition and length which is expressed from a plasmid. Eventually, the PumHD protein chain binds a specific sequence on the target mRNA. When fusing a fluorescent protein to the PumHD chain, the native target mRNA can be visualized. An additional advantage of this method is that PumHD modules can be displaced by ribosomes, thus they can be exploited to study translation (Adamala, Martin-Alarcon, and Boyden 2016).

In *E. coli*, an elegant tetramolecular fluorescence complementation (TetFC) system was used to label a specific mRNA target (Table 1) (Kellermann and Rentmeister 2017). The TetFC system contains 2 variants of the Pumilio protein and a three partite split GFP. Each Pumilio variant contains one β-strand of the GFP fluorescent protein and thus, a fluorescence signal is generated only when the target mRNA is bound by both Pumilio variants and the detector GFP. This strategy was combined with FACS to sort *E. coli* cells in which the target mRNA was expressed from cells lacking the target mRNA.

Alternatively, endogenous RNA, such as mRNA and microRNA (miRNA), can be labeled *in vivo* with molecular beacons (Guk et al. 2019; Bratu et al. 2003; Sanjay Tyagi and Alsmadi 2004; Matsuo 1998; Deborah L. Sokol et al. 1998). Molecular beacons are probe-like structures made up of a binding domain that is complementary to the RNA of interest and of two flanking regions that form a hairpin structure when unbound to RNA (S. Tyagi and Kramer 1996; Zheng et al. 2015; M. Chen et al. 2017). At the 3’ and 5’ regions of the molecular beacon probe, a quencher and fluorophore are attached, respectively. When the probe forms the hairpin through intermolecular base-pairing, the fluorophore and quencher are in close proximity, rendering the probe non-fluorescent. Upon binding to the target
sequence, the hairpin opens and the quencher-fluorophore conjugate is broken, resulting in the emission of a fluorescence signal (Figure 3A). In addition to molecular beacons, several other constructs, such as forced intercalation (FIT) or quenched autoligation (QUAL) probes, exist (Köhler, Jarikote, and Seitz 2005; Felix Hövelmann et al. 2013; F. Hövelmann et al. 2016; Sando, Abe, and Kool 2004). For example, QUAL probes were used to discriminate between morphologically similar bacterial species (Table 1) (Silverman and Kool 2005) that hybridized in vivo to species-specific 16S rRNA sequences of either E. coli, S. enterica or P. putida and the detection using conventional fluorescence microscopy for identification.

Probe systems, such as the aforementioned molecular beacons, FIT probes and QUAL probes, share the same advantage. They are in the dark-state when unbound and only generate a fluorescent signal upon binding to the RNA target, solving the problem of background fluorescence. However, these probes can be degraded by endogenous nucleases. Backbone modifications, such as 2′-O-methylation, or synthetic probes, such as PNA or LNA probes, can be used to stabilize probes in the native environment of a cell (R. W. Dirks, Molenaar, and Tanke 2003; Fontenete et al. 2013; Wiegant et al. 2010). These synthetic probes are exogenous and therefore have to be introduced into cells. Generally this is achieved by microinjection or by the formation of transient pores in the cell membrane with pore-forming agents (S. Tyagi and Kramer 1996; Matsuo 1998; D. L. Sokol et al. 1998; Mhlanga 2005; Bratu et al. 2003). These applications have been validated in eukaryotic cultures, however applications in bacteria for in vivo RNA imaging have not been reported.

Live cell RNA imaging approaches have been proven powerful for the study of transcription dynamics and transcript localization. When combined with the detection of translating transcripts, the intertwining between transcription and translation can be studied. Recently, the MS2 system was combined with SunTag labeling to follow both the formation of the transcript and its translation (Tanenbaum et al. 2014; Yan et al. 2016; Wu et al. 2016; C. Wang et al. 2016). The gene encoding the transcript is modified to include a tandem array of the sequence coding for GCN4 peptide epitopes in the open reading frame. Upon translation of the transcript, short peptides are produced that act as epitopes for a single-chain variable fragment (scFv) antibody fused to a fluorescent protein coexpressed in the cell. Additionally, the transcript of interest is labeled by the MS2 system with a different fluorescent protein. In this way, translating transcripts will display a fluorescence signal for both fluorescent proteins, while transcribed but untranslated transcripts will only exhibit a fluorescence signal originating from the MS2 system. Thus, the combination of these two systems in single eukaryotic cells allowed the simultaneous study of both transcription and translation dynamics.

Alternative approaches focussed on the labelling and imaging of other actors involved in gene expression, such as RNA polymerase (RNAP), ribosomes and tRNA, and enabled the detection of the dynamics and localization of the transcription and translation machineries, as well as their interaction with RNA or DNA, respectively (Volkov et al. 2018; Sanamrad et al. 2014; Stracy et al. 2015). In bacteria, super-resolved single-particle tracking (SPT) is a particularly suitable tool for distinguishing the diffusive states of a molecule within a 3D space of a cell, thereby enabling the distinction between freely diffusing from bound/static molecules (Table 1) (Figure 3D) (Stracy et al. 2015). More specifically, SPT-PALM has
been used to track both RNAP and ribosomal subunits, through fusion of PAmCherry to the beta’ subunit of RNAP and the fusion of mEos to the 50S ribosomal protein L1 and 30S ribosomal protein S2, respectively (Sanamrad et al. 2014; Stracy et al. 2015). In these studies, it was shown that both bound RNAP and bound ribosomes are mostly excluded from the nucleoid, while unbound RNAP and ribosomal subunits can diffuse through the nucleoid space. A similar technology was more recently introduced to study translation kinetics. Codon specific, in vitro dye-labelled tRNA molecules were introduced into E. coli cells through electroporation and tracked with super-resolved SPT to measure the dwell times of tRNAs on ribosomes as a proxy for translation rates (Volkov et al. 2018).

**Multiplexed imaging**

RNA aptamer-based approaches and smFISH have proven to be very powerful tools for the study of transcript localization and abundance. However, the detection of multiple RNA species at the same time in the same single cell would additionally enable the characterization of transcriptional patterns with spatial resolution.

For live cell imaging of eukaryotic systems, the multicolor read-out of several different mRNAs was reported using RNA-binding protein systems. For instance, the MS2 system was combined with either the PP7 system, derived from the coat protein of bacteriophage PP7, or with the LambdaN system, derived from the lambda bacteriophage antiterminator protein N (Lange et al. 2008; Hocine et al. 2013). The main disadvantages of this approach are the extensive genetic/transcriptomic modifications required and the limited number of colors that can be detected at once. Both disadvantages limit the number of mRNA species that can be visualized simultaneously to only a few (2-3).

Recently, advances in labeling strategies, imaging and automated analysis led to the development of several in situ Sequencing (ISS) and in situ Hybridization (ISH) techniques for the read-out of tens to thousands of RNAs in fixed cells. The first multiplexing ISH strategies that enabled the simultaneous measurement of tens of mRNA species relied on direct transcript labeling schemes where target mRNAs were encoded with spectral barcodes and read-out by sequential hybridization and imaging rounds (Levsky 2002; Lubeck and Cai 2012; Lubeck et al. 2014; Levesque and Raj 2013). SeqFISH, the first protocol to realize sequential FISH (Lubeck et al. 2014), can suffer from high background autofluorescence when imaging thick and opaque tissue samples, thus limiting the accurate detection of mRNA molecules (Shah et al. 2016). This problem was circumvented by combining sequential imaging with single molecule HCR (smHCR) for signal amplification (Shah et al. 2016, 2017). Here, the complementary region of the primary probe hybridizes to the mRNA molecule of interest in each hybridization round and the HCR initiator sequence in the probe tail triggers the HCR reaction. After imaging, the probes are stripped off the transcripts by DNase treatment before a new round of hybridization is started. Both seqFISH and smHCR seqFISH enable the detection of tens to hundreds of mRNA species. Alongside seqFISH, another approach, called ouroboros single molecule FISH (osmFISH), was developed to address the issue of spatial overlapping of different RNA species in a diffraction-limited volume (Codeluppi et al. 2018). OsmFISH employs a
direct labeling scheme, similar to that of seqFISH, in which the number of target molecules is defined by the number of hybridization rounds and channels that are imaged. Even though the multiplexing capabilities of these approaches are limited, the advantage of seqFISH and osmFISH is that only one target RNA species is detected in each image, thus considerably reducing the signal overlapping of closely positioned mRNA molecules and therefore greatly facilitating the precise detection and localization of these mRNA molecules, and limiting undercounting.

Alternatively, approaches using combinatorial encoding schemes allow the increase of the number of detected RNA species in a manner that is not proportional to the number of hybridization cycles. The specific combinatorial scheme implemented in Multiplexed error-robust FISH (MERFISH) is able to detect and correct registration errors (K. H. Chen et al. 2015). The original publication reported the detection of 140 RNA species with error-correction. The same study also reported the detection of 1001 RNA species but with no error correction (K. H. Chen et al. 2015). RNA species were encoded with a 16-bit binary word where each bit represents the on (1) or off (0) signal of the RNA molecule in each hybridization round. Each RNA species is designed to be labeled with a read-out probe and detected in only 4 out of 16 hybridization rounds. The use of 16-bit binary words that are separated from each other by a Hamming distance of four enabled robust error-correction, as mutations of a single digit could be corrected without ambiguity. Transcript identities were then decoded by reconstruction of the 16-bit words through the sequential imaging and localization of the 16 read-out probes. After decoding the transcripts, the abundance and localization of each RNA species in the sample was obtained. MERFISH has single-molecule detection capabilities, however, care must be taken when imaging highly abundant species, as transcripts located closer to each other than the diffraction limit and sharing one readout probe cannot be decoded. The spatial overlap of RNA molecules in MERFISH was recently addressed by combining MERFISH with ExM to increase the RNA density limit (Wang et al. 2018). In a complementary approach, abundant species were decoded using a sequential -instead of a combinatorial- encoding scheme (Moffitt et al. 2018). More recently, the multiplexing capabilities of MERFISH were increased by one order of magnitude to detect approximately 10 000 RNA species (Xia et al. 2019).

Alternatively, a second combinatorial approach was developed (seqFISH+) that makes it possible to image and detect up to approximately 10 000 mRNA molecules (Eng et al. 2019). SeqFISH+ employs an indirect labeling scheme in which barcodes are sequentially read-out using combinatorial encoding combined with multicolor imaging. Each target RNA species was encoded with a unique sequence of 4 pseudocolors. In SeqFISH+, the 60 pseudocolors were generated by including in each barcoding round 20 hybridization cycles in which probes are imaged in three color channels. This means that the labeling and detection of a specific RNA species in a given hybridization round and in a given color channel represents the pseudocolor of that barcoding round. Transcript identities were decoded by reconstructing the pseudocolor barcode through the labeling and detection of RNA molecules for each of the four barcoding rounds. This labeling and imaging strategy results in 60 unique images for each barcoding round, each representing a specific pseudocolor. Single molecule detection efficiency in seqFISH+ is thus achieved by detecting only a subset of the total ensemble of targeted mRNA molecules in each image, which avoids the spatial overlapping of different RNA species and ultimately allows for the precise
localization of single target mRNAs by Gaussian fitting without the use of super-resolution methods. Despite the successful application of these FISH-based methods to a wide variety of eukaryotic systems, their use in bacterial systems is still to be reported.

A second family of technologies perform the optical read-out of different RNA species by in situ sequencing (ISS). ISS methods allow the direct linking of the sequencing output information to the spatial localization of target RNAs. Generally, the ensemble of target transcripts are amplified in situ by using RCA-based methods to create DNA amplicons containing a detection target (i.e. the mRNA transcript of interest) and a sequencing target consisting of either the detection target itself or a barcode incorporated during RCA. The sequencing target is sequenced by “Sequencing By Ligation” (SBL) or by “Sequencing By Synthesis” (SBS) using fluorescently labeled probes or nucleotides, respectively. The first ISS development reported used reverse transcription of mRNA to cDNA, and hybridization of padlock probes to cDNA followed by RCA for amplification, ultimately creating a DNA amplicon called a “Rolling Circle Product” (RCP) (Ke et al. 2013). An RCP contains multiple repeats of the sequencing target and is thus essential for signal amplification. Depending on the design of the padlock probe, the detection target is directly sequenced or a sequencing target is incorporated in the RCA which functions as an indirect barcode. In the original method, SBL is used to reconstruct a four base-long sequence and to identify the target transcript. Recently, the throughput of this approach has reached a multiplexing capability of approximately 100 target RNAs (Qian et al. 2020). A variation on this original work, called “Barcode in situ Targeted Sequencing” (BaristaSeq), employs the same strategy to generate RCPs of target transcripts and 15 nucleotide-long sequences by SBS (Chen et al. 2018). The increase in read length in BaristaSeq requires an increase in the number of imaging cycles to 15. To ensure the spatial stability of the RCP during imaging, the RCP is anchored to the cellular matrix. Concomitantly, Spatially-resolved Transcript Amplicon Readout Mapping (starMAP) (Wang et al. 2018) avoids cDNA synthesis by using a “Specific amplification of Nucleic Acids via Intramolecular Ligation” (SNAIL) probe system, containing a padlock carrying a unique five nucleotide-long barcode and a primer that hybridizes partially to the target transcript and partially to the padlock probe. The primer of the SNAIL probe can then be used to initiate RCA. The RCPs are embedded into a hydrogel, which has the advantage that unbound protein and lipids can be washed out of this gel, resolving any background fluorescence (Wang et al. 2018). Finally, the barcode or sequencing target is reconstructed by SBL. A disadvantage of both in situ hybridization and in situ sequencing methods is that they require a priori knowledge of the RNA target sequences to be detected. Generally, this implies that only well-annotated transcripts can be targeted. An interesting development, relying on the use of Fluorescence in situ Sequencing (FISSEQ) for DNA amplicon sequencing (Shendure et al. 2005; Kim et al. 2007; Mitra et al. 2003), circumvents this issue by allowing the targeting and sequencing of transcripts for which no a priori sequence knowledge is available as well as of non-coding RNA species (Lee et al. 2014; Lee et al. 2015). Unbiased labeling of RNA species is achieved by using random-hexamer primer sequences which initiate reverse transcription of the primer tagged RNA species (Lee et al. 2014). Each of these primer sequences carries a tag, which later functions as the sequencing primer. After cDNA synthesis, the cDNA molecules are anchored to the cellular matrix, circularized by ligation and amplified by RCA. Then SBL results in a read length of 30 bases. By employing a partition sequencing strategy where the RCPs are randomly sampled
and only a subset is sequenced, up to approximately 8000 different target RNA species can be discriminated. Despite the powerful multiplexing capabilities of ISS methods and their application to cultured eukaryotic cells and complex tissue samples, ISS methods are yet to be reported in bacteria.

Conclusion

The large palette of imaging-based methods described in this review allowed researchers to gain incremental insights into the mechanisms involved in the regulation of RNA biosynthesis as well as in cellular RNA functions. The ability to either follow RNA molecules in space and time with single molecule sensitivity in live cells or the read-out of many RNA species simultaneously in fixed cells allowed the investigation of RNA dynamics, abundance and localization at the single cell level in their native context. These methods led to several pioneering discoveries. For example, both the MS2 system and smFISH were used to examine the stochastic nature of bacterial gene activity in *E. coli*, which underlies transcriptional bursting (I. Golding and Cox 2004; Ido Golding et al. 2005; Zong et al. 2010; So et al. 2011; Skinner et al. 2013). Furthermore, smFISH has been used in *B. subtilis* to show that stochastic gene expression can drive cell states transitions resulting in cell competence (Maamar, Raj, and Dubnau 2007).

Additionally, imaging-based methods have expanded our understanding of the subcellular spatial organization of RNA within bacterial cells, which lack membrane-bound organelles. For instance, smFISH enabled the discovery of the sub-cellular localization of small bacterial RNAs depending on their expression level (Fei et al. 2015). The use of whole-transcriptome FISH labeling contributed to the recent understanding that mRNAs in bacteria can display specific patterns of subcellular localizations (Weng et al. 2019; Moffitt et al. 2016; Montero Llopis et al. 2010; Nevo-Dinur et al. 2011). Finally, labeling of pre-rRNA using smFISH and RNA polymerases using fluorescent protein tagging was used to show that the genomic organization within the bacterial nucleoid greatly influences the spatial organization of RNA polymerase clusters and of rRNA transcription sites (Weng et al. 2019).

Conversely, these important findings have also triggered novel biological questions regarding the mechanisms underlying RNA localization and its possible regulatory roles. One major question is whether the classical hypothesis of co-transcription and co-translation in bacteria holds true. On the one hand, localization studies of transcribed gene loci, RNAP and ribosomal subunits support a model in which active transcription seems to drive transcription and translation machineries towards the periphery of the nucleoid and the cytoplasmic space (Sanamrad et al. 2014; Stracy et al. 2015; Yang et al. 2019). On the contrary, imaging-based localization studies, cell fractionation, and RNAseq based studies are rather consistent with translation-independent patterning of a subset of mRNA species in several bacteria (Kannaiah, Livny, and Amster-Choder 2019; Nevo-Dinur et al. 2011; Benhalevey et al. 2017). Aside from its implications for translation, the function of subcellular RNA enrichment remains largely unexplained. Deeper insights into the relationship between gene expression, transcriptome localization and the phenotype of the cell might uncover co-regulated and co-localized transcripts, ultimately leading to a better understanding of the molecular pathways involved and their implications to cell physiology.
Despite these many important applications of imaging-based RNA detection methods, many challenges remain to be addressed. For instance, in the development of new methods for tagging RNAs without altering their regulatory function, diffusion behavior or localization (Garcia and Parker 2016; Heinrich et al. 2017; Garcia and Parker 2015; Haimovich et al. 2016; Ido Golding et al. 2005; Tutucci et al. 2018). Diffraction of light within a microscope, and the intrinsically high densities of RNA in bacteria, can both degrade our ability to discriminate and accurately quantify abundance and localization of large numbers of RNA species in bacteria. Approaches combining novel RNA labeling and sample preparation strategies with super-resolution fluorescence microscopies will likely help solve this issue in future.

Finally, while a large number of technologies to detect and quantify RNA in single cells was developed in the past two decades, only a subset of them was so far reported in bacteria. Established bacterial model systems can be genetically manipulated rather easily, and have proven to play a key role in the molecular dissection of important cellular processes, such as transcription, translation, or DNA replication and repair. Thus, we envision that application of novel RNA imaging technologies to bacterial model systems will likely lead to new important discoveries, relevant to the study of RNA biology in general.

Acknowledgments

We acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No 724429) to M.N and from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No 721874 (SPM2.0) to S.R.
<table>
<thead>
<tr>
<th>Method</th>
<th>Selected applications to bacteria</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>smFISH</td>
<td>Localization of mRNA and sRNA, in <em>E. coli</em> and <em>B. subtilis</em>. Examples: spatial localization of sRNA SgrS depends on its expression level, and transcriptome localization in <em>E. coli</em>, gene expression regulation in <em>E. coli</em>, noise in gene expression in <em>B. subtilis</em></td>
<td>(Fei et al. 2015; Moffitt et al. 2016; Maamar, Raj, and Dubnau 2007; Sepúlveda et al. 2016; Arbel-Goren et al. 2016; So et al. 2011)</td>
</tr>
<tr>
<td>MS2 system</td>
<td>Dynamics of transcription. Examples: measurement of stochastic gene expression in <em>E. coli</em>, or transcript localization in several species, including <em>E. coli</em>, <em>Lactococcus lactis</em> and <em>Bacillus subtilis</em>.</td>
<td>(Ido Golding et al. 2005; Nevo-Dinur et al. 2011; van Gijtenbeek et al. 2016; Santos et al. 2012)</td>
</tr>
<tr>
<td>HCR</td>
<td>Detection of bacterial species. The study of gene expression.</td>
<td>(Yamaguchi et al. 2015; Nikolakakis et al. 2015)</td>
</tr>
<tr>
<td>3D-STORM-FISH</td>
<td>Transcriptome localization in <em>E. coli</em>.</td>
<td>(Moffitt et al. 2016)</td>
</tr>
<tr>
<td>RNA aptamers</td>
<td>Study of conformational changes in RNA molecules. Apta-FRET in <em>E. coli</em> where Spinach and Mango are expressed in close proximity on a tRNA scaffold to study the conformational changes in RNA nanostructures. Spinach, Broccoli, Mango, Corn, DNB and SRB-2 RNA aptamers were used in bacteria.</td>
<td>(Jepsen et al. 2018) (Filonov and Jaffrey 2016; J. Zhang et al. 2015; Dolgosheina et al. 2014; Song et al. 2017; Arora, Sunbul, and Jäschke 2015; Sunbul and Jäschke 2013b, 2018)</td>
</tr>
<tr>
<td>fluorogenic dyes</td>
<td>Fluorogenic near-red and infra-red dyes used in <em>E. coli</em> for confocal imaging of tRNA/mRNA and STED imaging of mRNA.</td>
<td>(Wirth et al, 2019)</td>
</tr>
<tr>
<td>Pumilio-based system</td>
<td>The tetramolecular fluorescence complementation (TetFC) system was used to label specific RNA targets in <em>E. coli</em>.</td>
<td>(Kellermann and Rentmeister 2017)</td>
</tr>
</tbody>
</table>
**Figure 1**

**Synthesis of coding and non-coding RNA species in bacteria during transcription.**

In bacteria, both coding and non-coding RNA is synthesized from double-stranded DNA by RNA polymerase, presumably in the nucleoid periphery, in the process of transcription. Coding RNA or messenger RNA (mRNA) is translated to proteins by ribozymes, presumably in the nucleoid periphery as well. Non-coding RNA, such as transfer RNA (tRNA), ribosomal RNA (rRNA) and other small regulatory RNAs (sRNA), will fold into distinct nanostructures and will play a role in downstream processes such as translation and gene expression regulation.
Figure 2: Methods for fixed cell imaging of RNA species at single-molecule resolution.

A) Signal amplification methods. Single molecule FISH (smFISH) (Raj et al. 2008), in which several fluorescently-labeled oligonucleotides bind to the transcript. Hybridization Chain Reaction (HCR) (Choi et al. 2010; Choi, Beck, and Pierce 2014; R. M. Dirks and Pierce 2004; Shah et al. 2016), in which a primary probe containing a homology region to the transcript of interest and a HCR initiator sequence binds to the transcript. Subsequently, the HCR initiator sequence triggers the self-assembly of the two fluorescently-labeled hairpins into a long polymer chain. Padlock probes and Rolling Circle Amplification (RCA) (Larsson et al. 2010), in which an LNA primer sequence hybridizes to the transcript of interest and initiates a reverse transcriptase reaction. The RNA-cDNA hybrid is then partially digested by RNase H, which in turn allows the hybridization of a padlock probe. Finally, this padlock probe is used as a template for the RCA. The amplified product contains binding sites for fluorescently labeled oligonucleotides. Branched DNA (bDNA) Hybridization (Player et al. 2001; Xia et al. 2019; Kishi et al. 2019), in which an amplifier sequence is bound to the primary probe and provides a multitude of additional hybridizing sites for fluorescently-labeled oligonucleotides.

B) Examples smFISH in *B. subtilis* (left) and *E. coli* (right). Adapted from (Maamar, Raj, and Dubnau 2007) and (Fei et al. 2015), respectively.

C) Example of fluorescent labeling and simultaneous imaging of well-defined populations of RNAs in *E. coli*. Phase contrast and STORM cross-section images of cells stained with FISH probes against inner-membrane-protein mRNAs (left) and cytosolic-protein mRNAs (right). Adapted from (Moffitt et al. 2016).
Figure 3:
Methods for live cell imaging of RNA species at single-molecule resolution.
A) Genetically-encoded RNA aptamers (top) tethered to the transcript of interest bind either an RNA binding protein fused to a reporter protein (MS2 system) (Bertrand et al. 1998) or fluorogenic dyes (Paige, Wu, and Jaffrey 2011). Strategies for the study of native transcripts (bottom), based on conditionally quenched fluorophores. Opening of the hairpin structure upon binding of the molecular beacon (S. Tyagi and Kramer 1996) or by stable binding of the quencher to the RT aptamer (Sunbul and Jäschke 2013a; Arora, Sunbul, and Jäschke 2015) disables the quenches activity of the beacon and the dye molecule fluoresces.

B) Example of the MS2 system in E. coli with the protein mRFP1 (red) and its corresponding transcripts (green). MS2 aptamers are added to the mRFP1 transcript enabling binding of MS2 fused to GFP. Adapted from (Ido Golding et al. 2005).

C) Example of the Broccoli aptamer binding the fluorogenic dye DFHBI in E. coli. An empty plasmid and a plasmid carrying the Broccoli aptamer were transformed into E. coli cells. The empty plasmid does not give a fluorescence signal, and functions as the control, while the Broccoli aptamer binds the dye molecules and fluoresces. Adapted from (Filonov and Jaffrey 2016).

D) Example of SPT-PALM of RNAP in E. coli. A photoactivatable fluorescent protein PAmCherry is fused to the beta’ subunit of RNAP and tracked over time.
trajectories of individual bound (red) and mobile (blue) RNAPs (top). Localization of RNAP trajectories in representative cells, with bound or mobile RNAPs according to their diffusion coefficient (middle). Localization of RNAP trajectories in representative rifampicin-treated cells, with bound or mobile RNAPs according to their diffusion coefficient (bottom). Adapted from (Stracy et al. 2015).
References


Braselmann, Esther, Aleksandra J. Wierzbä, Jacob T. Polaski, Mikolaj Chromiński,


Moffitt, Jeffrey R., Shristi Pandey, Alistair N. Boettiger, Siyuan Wang, and Xiaowei Zhuang.


