

Single-cell, single-molecule RNA quantification of toxin-antitoxin gene expression in *E.coli*

Major research project
Bioinformatics and Systems Biology Master's programme

Author
Florence Berbain

Supervisors
Prof. Dr. Frank Bruggeman
Dr. Evelina Tutucci

Examiner
Dr. Matti Gralka

Abstract

Toxin-antitoxin (TA) systems are gene modules on the bacterial genome that are thought to be involved in triggering persistence, a transient phenotype switch in which cells enter a growth arrest state that allows them to survive catastrophic conditions. However, evidence for the role of toxin-antitoxin systems is conflicting, since the mechanisms that cause persistence, or even at what biological level they operate, are complex and currently quite unclear. In this project we investigated how a molecular level parameter, the expression level of toxin-antitoxin systems, is related to a physiological level parameter, growth rate. We adapted a single-molecule mRNA Fluorescence In Situ Hybridization protocol to quantify TA system expression for a range of growth rates in exponential growth. We also measured the expression of the gene coding for σ_{70} , the main sigma factor involved in exponential growth. Our results suggest that rpoD transcription is linked to growth, in contrast to that of the TA systems, and that some TA transcripts might exhibit some degree of spatial localization in the cell cytoplasm. Based on these preliminary results, more research is needed to further optimize the FISH protocol and the associated processing pipeline, as well as extend the investigation to the transition into stationary phase by measuring the expression of rpoS, the sigma factor that regulates the general stress response in *E.coli*.

April-December 2024 (48 EC)
Course code: XM_0104
Student number: 2785299
Vrije Universiteit Amsterdam

Contents

| | | |
|----------|---------------------------------|-----------|
| 1 | Introduction | 2 |
| 2 | Results | 5 |
| 2.1 | Growth rates | 5 |
| 2.2 | smFISH pictures | 5 |
| 2.3 | Quantitative analysis | 6 |
| 3 | Discussion | 15 |
| 4 | Outlook | 17 |
| 5 | Conclusions | 18 |
| 6 | Methods | 18 |
| A | smFISH E.coli protocol | 23 |
| B | Medium | 28 |

1 Introduction

Persisters are a subpopulation of dormant cells Subpopulations of stress-tolerant persister cells are found in many bacterial species. Persister cells are in a growth-arrested state, which allows them to evade the action of antibiotics, as antibiotics kill cells by interfering with a cellular growth process ([Lewis, 2010], [Page and Peti, 2016]). Persistence is different from resistance and tolerance: in resistance, bacteria acquire mutations that make them insensitive to the antibiotics action. A tolerant bacteria is sensitive to the antibiotics, but gets killed more slowly than the wild type. Persistence, in contrast, appears in an isogenic population and is the result of a transient phenotypic switch in a part of the population ([Balaban et al., 2004], [Rotem et al., 2010]). Figure 1 illustrates the difference in the killing curves of bacteria exhibiting resistance, tolerance, and persistence ([Balaban et al., 2019]). The phenotypic heterogeneity of a population containing persisters is reflected in a characteristic biphasic killing curve, corresponding to the two populations.

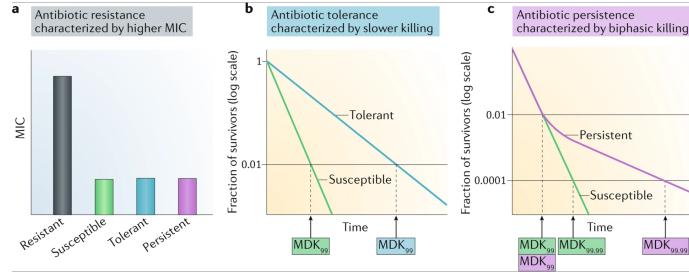


Figure 1: Killing curves for resistant, tolerant, and persistent bacteria ([Balaban et al., 2019])

Resistance is a serious threat because it can cause antibiotics treatment to fail. Persistence is now also increasingly recognized as playing a role in the recurrence of chronic infections in diseases like oral thrush, urinary tract infection, or lung infections in cystic fibrosis patients ([Lewis, 2010]). There is also evidence that persistence favors the evolution of bacteria towards resistance ([Lewis and Shan, 2017], [Levin-Reisman et al., 2017], [Barrett et al., 2019]). For all of these reasons, it is desirable to gain an understanding of persister cell formation. Explanations have been proposed at different levels of observation (evolutionary, phenotypic, and molecular scale) ([Berkvens et al., 2022]), but these different levels have not been consolidated into an overarching theory.

Persister cells are the result of phenotypic switching Persistence appears when some cells in an isogenic populations switch to the dormant state. What triggers the switch and why cells switch back to the growing phenotype is not well understood. Three mechanisms have been proposed ([Berkvens et al., 2022]): under any conditions, persisters might form by chance, at a very low, basal rate. From an evolutionary perspective, these cells could be seen as an insurance policy against the risk of a catastrophic event, because they can survive conditions that will kill the rest of the population and start growing again when conditions become favorable again. In contrast to this preventive mechanism, persisters could also form responsively, on an as-needed basis, as a response to stress conditions (e.g. low ATP levels, pH change, heat shock, nutrient depletion). A third hypothesis is that persisters form as a response to a reduction in growth rate, as part of the generalized stress response that is triggered by sudden growth rate reduction. In that case, persister formation would be associated with ppGpp increase and σ^s activation. It has been observed that stress can trigger persistence ([Lewis, 2010]), and toxin-antitoxin (TA) modules are also thought to possibly be involved ([Page and Peti, 2016], [Harms et al., 2018]).

Toxin-antitoxin systems might be involved in regulation of persistence TA systems are broadly conserved gene modules found on plasmids and on the chromosome of bacteria and archaea. They are composed of a pair of genes that encode a cognate set of toxin and antitoxin. The bacterial genome contains multiple TA systems, depicted in Figure 2. Toxins are stable proteins that interfere with a vital essential cellular process. Antitoxins can be RNAs or proteins. Antitoxins are unstable and under normal conditions must be continually expressed to compensate for their degradation and inhibit their cognate toxins. Within one system, the toxin and the antitoxin gene are contiguous, and both genes are co-transcribed, with the antitoxin and the bound TA complex

usually repressing their own transcription. The antitoxin is often located upstream of the toxin, resulting in higher transcription rates for the antitoxin, which under normal conditions is necessary for the unstable antitoxin to keep inhibiting the stable toxin.

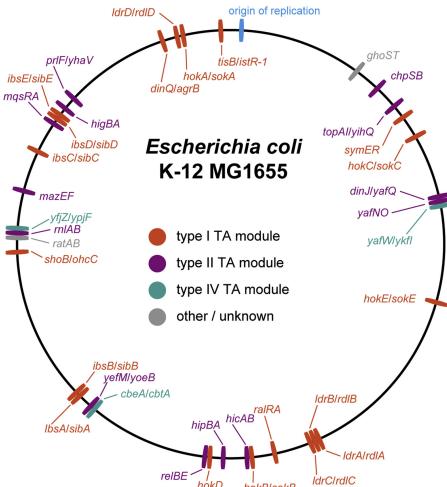


Figure 2: Genomic map of the toxin-antitoxin systems in *E.coli* ([Harms et al., 2018])

TA systems are classified according to whether the antitoxin is an RNA or a protein, and how it interacts with the toxin to interfere with its action. Type II systems are the most studied and best known TA systems. In type II systems, both toxin and antitoxin are proteins, and interact directly with each other.

Transcription regulation plays a key role in cell processes In the cell, mRNA molecules are the intermediary between DNA and proteins. While mRNA levels do not directly translate into instantaneous protein levels, understanding transcription patterns can give insights into protein composition over time ([Tunnacliffe and Chubb, 2020], [So et al., 2011]). Accurate quantification of mRNA expression is therefore needed to understand how genotype and phenotype are related. Fluorescence *in situ* hybridization is the technique of choice for quantifying and localizing RNA transcripts ([Rombouts and Nollmann, 2020], [Skinner et al., 2013], [Taniguchi et al., 2010]). In an earlier project carried out in this lab, expression of two different genes was measured in conditions differing by the stress *E.coli* cells were subjected to. Single molecule FISH (Fluorescence In Situ Hybridization) was used to measure the expression of two genes in the MG1655 K-12 sub-strain of *Escherichia coli*. The two target genes were *rpoD*, which codes for the exponential growth σ^{70} sigma factor, and the *hipBA* (High persister Protein A) toxin-antitoxin module. Conditions contrasted different media (rich versus minimal), pH and temperatures. Expression was found to be lower for both genes in minimal medium, and no significant correlation was found between *rpoD* and *hipBA* expression.

Despite extensive study over several decades, it has proved difficult to establish an undisputed causal link between TA systems and persistence ([Ronneau and Helaine, 2019]). [LeRoux et al., 2020] have found that even though stress increases TA system transcription, increased transcription does not necessarily cause an increase in toxin activity. Since evidence at the molecular level is conflicting, in this project we want to take a more integrative approach by investigating how the expression of some bacterial TA systems varies with a physiological level parameter, namely growth rate. More precisely, we want to compare the expression of two TA systems, *hipBA* and *rnlAB*, with that of *rpoD*, a gene that is constitutive of exponential growth. Our expectation is that a gene associated with growth will exhibit higher levels of expression at higher growth rates, whereas TA systems expression would remain unchanged.

Single-cell mRNA profiling in bacteria requires single-molecule sensitivity Because we are studying a heterogeneous population, we need to produce and analyze data at the single-cell level. This is especially true in the case of persisters, which have such low frequencies in the general population that any variation would be obscured by the majority of non-persistent cells.

In bacteria, we expect the transcripts we want to quantify to occur at low copy numbers ([Taniguchi et al., 2010]). To detect transcripts with single-molecule sensitivity, we used single-molecule RNA Fluorescence In Situ Hybridization (smRNA FISH) to tag and count the transcripts in fixed cells. In smRNA FISH, multiple DNA probes carrying a single fluorophore are hybridized to the mRNA of interest ([Raj et al., 2008]). The cells are then imaged using fluorescence microscopy, and the mRNA transcripts are detected as diffraction-limited fluorescent spots. [Skinner et al., 2013] have adapted smFISH protocols developed for eukaryotic cells to bacteria. An important difference is that, because bacterial cells are smaller than eukaryotic cells, individual spots can overlap in the output picture, and need to be separated algorithmically based on their fluorescence intensity.

Target genes For this project we have performed the [Skinner et al., 2013] protocol with three target genes to produce 3-color FISH images, which we have then analyzed using an in-house software pipeline build on the Python BigFish spot detection and analysis library ([Imbert et al., 2021]). The target genes for the FISH experiments were rpoD, hipBA, and rnlAB. One constraint for FISH probe design is that, for the fluorescent signal to be visible, the gene needs to be long enough to accommodate 48 probes of length 20 without causing non specific binding. This was the case for rpoD (1842 bp) and hipBA (1589 bp), and guided the choice of rnlAB (1437 bp) as second TA systems.

1. rpoD: rpoD codes for the σ^{70} sigma factor, the main σ factor during exponential growth, and is expected to be expressed in high numbers during exponential growth ([Pletnev et al., 2015]).
2. hipBA: the hipBA TA system was initially discovered by purposefully screening for high persistence mutants ('hip' stands for 'high persistence') ([Moyed and Bertrand, 1983], [Black et al., 1991]). These mutants were later shown to be indeed growth-arrested individuals ([Balaban et al., 2004]), establishing the link between growth arrest and persistence, while HipBA was also shown to be a toxin and HipB its cognate antitoxin ([Falla and Chopra, 1998], [Korch et al., 2003]), suggesting a link between persistence and TA systems. HipA is serine-threonine kinase that activates the production of the ppGpp alarmone by RelA and ultimately leads to growth arrest. HipA is toxic when HipB is not present. The hipBA operon is located at 50 min on the bacterial chromosome, right opposite the origin of replication. The hipB gene is located upstream of hipA, with the ORFs overlapping by one base pair. Both HipB and the HipA-HipB complex repress transcription (see Figure 3). [Rotem et al., 2010] found that cells enter the growth arrest state if the HipA level is higher than a threshold, that the amount by which the threshold was exceeded determines the duration of dormancy, and that the threshold level was determined by the level of the HipB antitoxin.
3. rnlAB: [Koga et al., 2011] first described rnlAB, another type II TA system. rnlA is a structural gene of RNase LS, an endoribonuclease that cleaves bacteriophage T4 mRNAs in T4 dmd-defective mutants. RNase LS also contributes to mRNA degradation. The rnlAB operon is unusual in that the toxin is located upstream of antitoxin gene, with a 7 base pair overlap. However, a promoter within the rnlA ORF was identified as an rnlB promoter in addition to an rnlA promoter upstream of rnlA. This promoter is eight-fold stronger than the rnlA promoter upstream of rnlA. In addition, transcription from the rnlA promoter is negatively regulated by IscR. The additional promoter and the IscR regulation might explain why the cell are able to produce enough antitoxin compared to the the toxin, despite rnlB behind located downstream from rnlA.

The goal of this project was to develop a method to quantify the expression of the rpoD gene and of the hipBA and rnlAB TA systems in *E.coli* at different growth rates using single molecule RNA FISH. Our expectation was that rpoD, a gene associated with exponential growth, would exhibit higher levels of expression at higher growth rates, whereas TA systems expression levels would not significantly change. When comparing different growth rates within the same growth medium, our results indeed suggest that rpoD is growth-regulated whereas the two TA systems are not. We found this to be valid for absolute expression numbers per cell, whether numbers of mRNA molecules or mRNA concentrations, as well as for expression normalized per gene copy number.

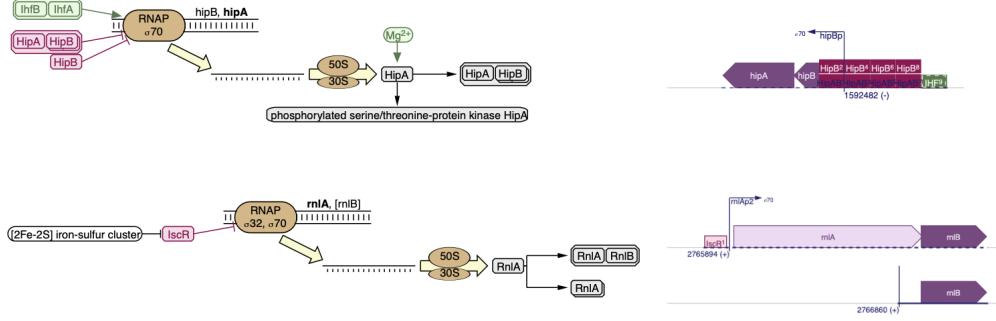


Figure 3: Regulatory influences and transcription units of *hipBA* (top) and *rnlAB* (bottom). Source: Ecocyc

2 Results

2.1 Growth rates

We measured the growth rates of *E.coli* MG1655 grown in AB minimal medium with four different carbon sources chosen to span a range of growth rates. We based our choice based on the growth rates reported in [Basan et al., 2020], though our strain is a different strain and appeared to grow significantly slower. The amount of carbon available in each condition was identical. Growth rates were measured in two different ways: automated measurements from a plate reader, and manual spectrophotometer measurements of shake flask cultures. The protocol for obtaining the cultures we measured on both types of devices was the same, but they were not necessarily performed on the same day the different replicates. Cultures had had at least 10 generations in exponential phase in the target medium for the condition before measurements were taken. We usually saw differences in growth rates between experiments done on different days, but within one experiment glucose was always the fastest carbon source, followed by glycerol, galactose, and mannose, with typical values of respectively 0.5, 0.3, 0.2, and 0.1. Example growth curves from the plate reader and fitting are shown in Figure 4 and Figure 5, Table 1 shows the numerical results.

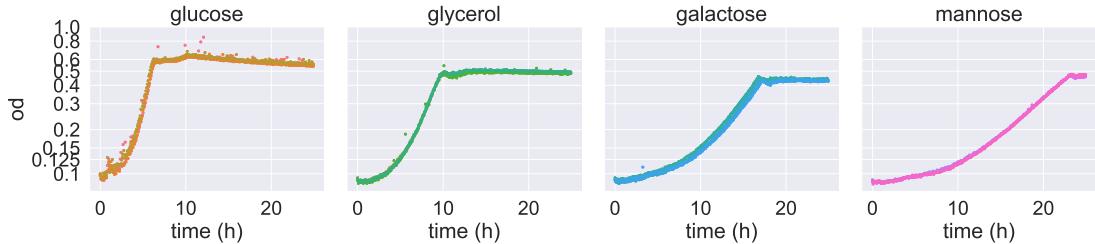


Figure 4: Growth curves. Four different carbon sources were used to perform FISH experiments at varying growth rates. In an initial experiment, growth in LB medium was measured at 2.3 h^{-1} . Growth rates were estimated by fitting a linear regression curve to the log of the OD in the exponential part of the curves. In the plate reader experiment shown here (AB medium + 4 different carbon sources), the estimated growth rates per hour were, from the fastest to the slowest carbon source: glucose: $\mu = 0.54$, glycerol: $\mu = 0.32$, galactose: $\mu = 0.17$, mannose: $\mu = 0.12$ (unit: 1/h).

2.2 smFISH pictures

We obtained FISH pictures for cultures grown in LB medium ($\mu = 2.3 \text{ h}^{-1}$), and for the four different carbon sources, approximately spanning a growth rate range between $\mu = 0.1$ and $\mu = 0.5$. Figure 6 shows a representative picture of cells grown in LB, whereas Figure 7 shows cells in the four different carbon sources (glucose, glycerol, galactose, mannose). Cells are larger at higher growth rate, which is expected ([Schaechter et al., 1958], [Bremer and Dennis, 2008]). Fluorescent

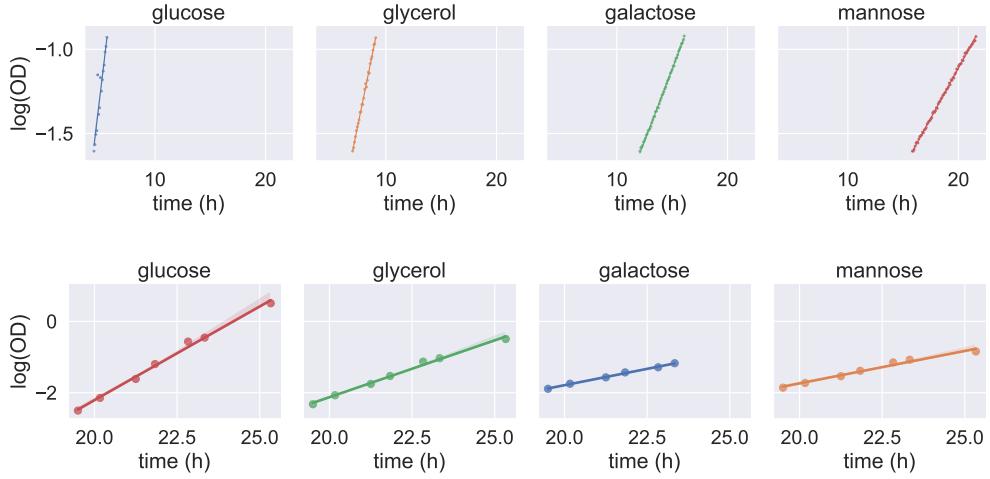


Figure 5: Linear regression fit, top: plate reader, bottom: shake flasks. Estimated growth rates are shown in Table 1.

| Carbon source | Device | $\mu_1(1/h)$ | $\mu_2(1/h)$ | $\mu_3(1/h)$ | $\mu_{mean}(1/h)$ | Replicates |
|---------------|-------------------|--------------|--------------|--------------|-------------------|------------|
| Glucose | Plate reader | 0.548 | 0.515 | - | 0.53 | 2 |
| Glycerol | Plate reader | 0.321 | 0.271 | - | 0.30 | 2 |
| Galactose | Plate reader | 0.170 | 0.132 | - | 0.15 | 2 |
| Mannose | Plate reader | 0.119 | 0.092 | - | 0.10 | 2 |
| Glucose | Spectrophotometer | 0.525 | 0.549 | 0.538 | 0.54 | 3 |
| Glycerol | Spectrophotometer | 0.319 | 0.423 | 0.416 | 0.39 | 3 |
| Galactose | Spectrophotometer | 0.183 | 0.219 | 0.225 | 0.21 | 3 |
| Mannose | Spectrophotometer | 0.182 | 0.131 | 0.134 | 0.15 | 3 |

Table 1: Estimated growth rates

spots are visible for all three probes, indicating mRNAs are expressed and the probes are able to bind to them in all five conditions. The frequency of TA system spots is significantly higher than the expected frequency of persister cells (from 10^{-6} under normal growth condition to 10^{-2} under stress conditions [Black et al., 1991]), underscoring that TA system expression is not a sufficient condition for persistence. In some cases, spots exhibit brighter intensities than the average and are colocated with the nucleoid, suggesting they might correspond to a transcription site, where more than just one isolated transcript are in close proximity to each other.

2.3 Quantitative analysis

Pipeline overview The FISH pictures were processed to detect spots corresponding to mRNA transcripts for each channel. Spots were detected using a software pipeline developed for the *Candida Albicans* fungal pathogen ([van Otterdijk et al., 2024]). This pipeline relies on the BigFish Python library ([Imbert et al., 2021] for spot detection, and consists of the following steps (illustrated in Figure 8): the image is first filtered to remove background noise, enhancing signal to noise ratio. Spots are detected on a z-section of the enhanced image determined to be in focus. Large, bright spots are decomposed into a set of individual spots by modeling their brightness as the superposition of the Gaussian signal that would be produced by the individual spots. To generate counts per cell, we first generated cell masks by segmenting the DIC picture using the `cyto2_omni` model of the Omnipose segmentation tool ([Cutler et al., 2022]), which we found to be the best-performing model on our DIC images. We also obtained DAPI masks from Omnipose by segmenting the maximal projection of the DAPI signal. While segmentation quality was acceptable to very good, depending on the picture, we improved the results after manual inspection by eliminating some cell segmentation masks that were visibly artifacts. We discarded masks that: 1. did not overlap with a DAPI mask (a "cell" with no nucleoid), 2. were too small to be a cell, and 3. masks that were both too big and too round (it is possible that some cells have very large areas, because they are hyphenated, but in that case they should be elongated). Finally, the decomposed spots are assigned to individual cell masks and counts per cell are generated for each fluorescent

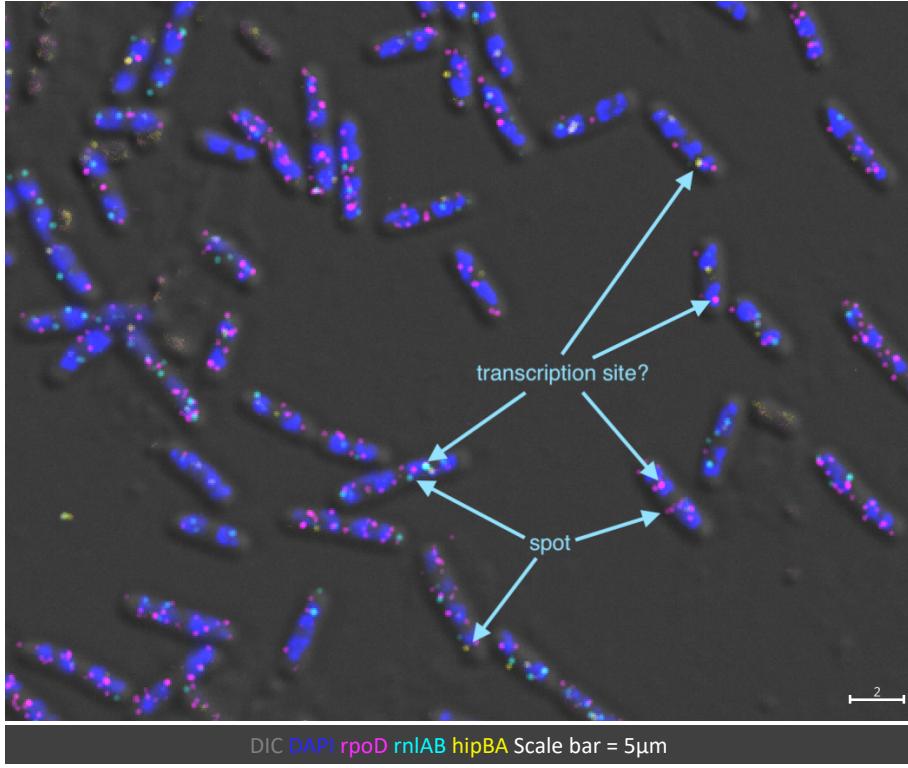


Figure 6: LB condition example FISH picture. Scale bar is 2 μm . Spots for all three genes are observed, indicating all three probes bind. Some spots have brighter intensities, suggesting they might be transcription sites, where several mRNAs are in close physical proximity.

channel.

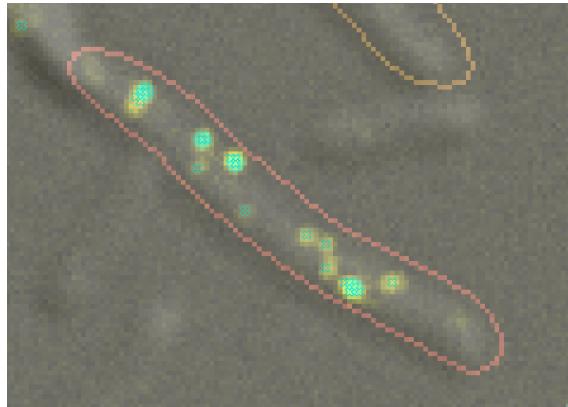


Figure 9: Spot decomposition example on the hipBA channel. The filtered hipBA signal is in yellow. Decomposed spots are marked by cyan crosses. Cyan disks denote bright intensity spots that were decomposed into several spots by the BigFish algorithm.

Number of cells After running the pipeline on pictures for all four AB medium conditions plus the LB medium condition, we obtained between 900 and 2500 cells per condition. It would make sense to aim for large number of cells, as the mRNAs we want to detect are likely quite rare, and our input pictures contained significantly more cells than these numbers. However, we had to discard part of them due to two reasons: first, the alignment between the DIC channel and the fluorescent channel was imperfect. This is known and expected, and is usually easily corrected by translating the DIC channel by an appropriate vector in the (x, y) plane. However, in our case, with cells than are smaller than the *Candida Albicans* cells the pipeline was initially developed for, we noticed that

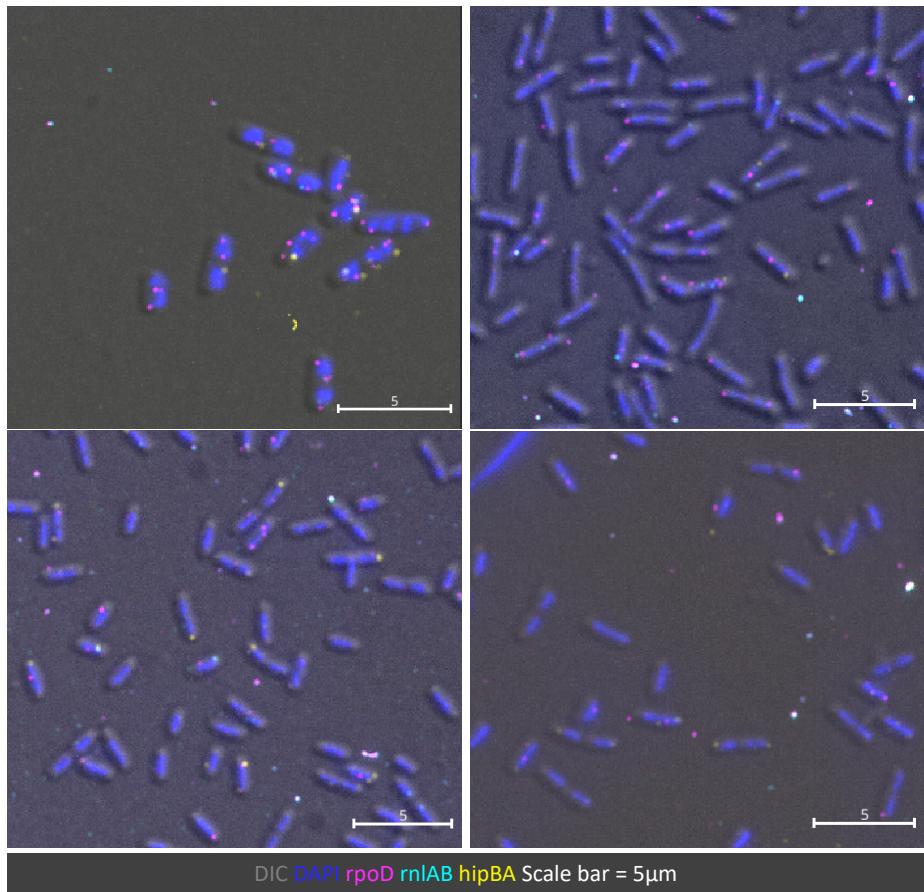


Figure 7: FISH pictures in 4 different carbon sources. Top left: glucose, top right: glycerol, bottom left: galactose, bottom right: mannose. Scale bar is 5 μ m. Spots for all three genes are observed in all conditions. Cells are smaller than in the rich medium LB. mRNA counts differ from cell to cell, which is expected as mRNA copy number vary stochastically within a population of genetically identical cells. Spot frequency for a given mRNA also seems to vary across conditions.

using the same translation vector across the whole picture was not enough to guarantee satisfactory alignment everywhere in the picture. This is important, as detecting segmentation masks with a correct shape but in a slightly erroneous location might impact mRNA counts for the cell. It is possible that using DIC as the cell shape channel is suboptimal. However, for this project, we decided to keep using it due to time constraints. To remedy the issue, we manually fine-tuned the translation for each image, possibly cropping out parts of the image where alignment was not good enough. We also slightly expanded the masks for the counting phase, since we noticed some spots on the cell border were not counted as in the cell when they should have been. This is ad-hoc processing and will need to be revisited in future to arrive at consistent, reliable results. The second step where we discarded cell masks was after segmentation: we removed cell masks that were smaller than 200 pixels (a pixel is 65 nm²), or larger than 1000 pixels and with an eccentricity below 0.8).

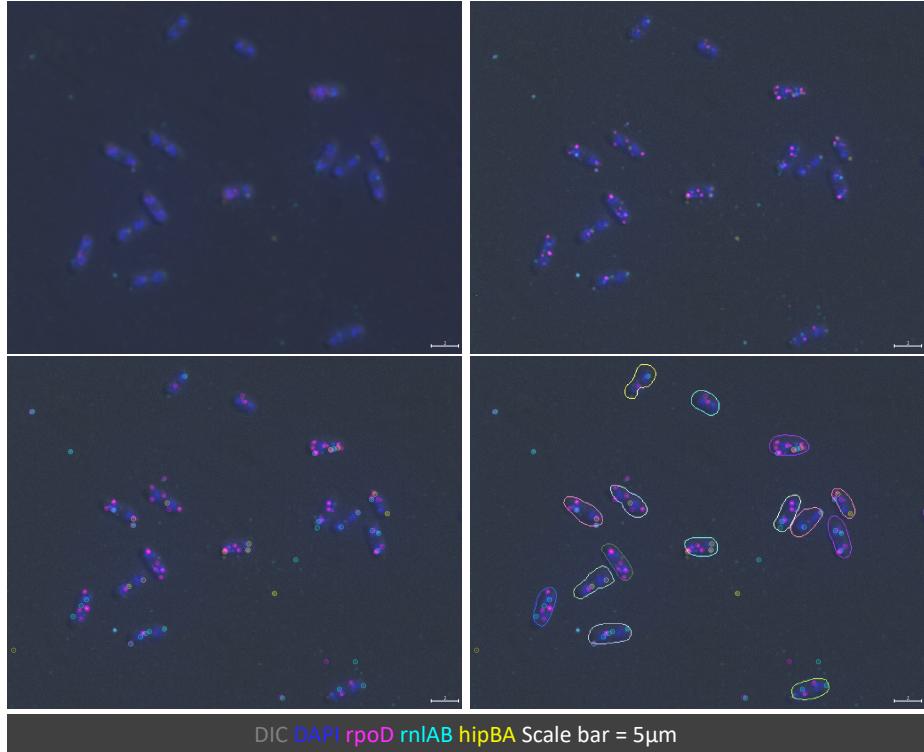


Figure 8: Analysis pipeline steps. Top left: raw input fluorescent signals for all three channels (magenta, cyan, yellow), DIC (grey), and DAPI (blue); top right: after filtering to improve signal to noise ratio; bottom left: spots detected from the enhanced image (spot decomposition not shown, see Figure 9), bottom right: adding segmentation masks allows to assign spots to individual cells and produce spot counts per cell for each mRNA ([Skinner et al., 2013]).

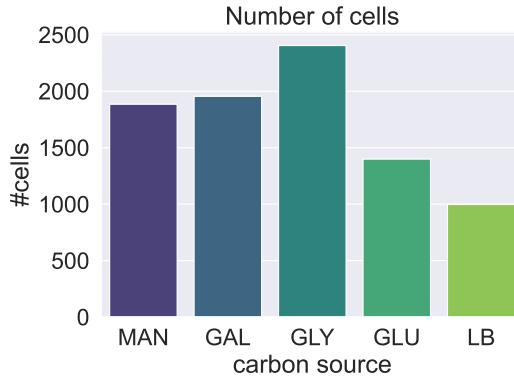


Figure 10: Final number of cells per condition after analysis (MAN: mannose, GAL: galactose, GLY: glycerol, GLU: glucose, LB: LB). Cells were discarded due to cropping and segmentation postprocessing.

Cell size increases with growth rate To estimate cell size, we used cell area, length, and volume estimate. Cell area is the area of the segmentation cell mask returned by Omnipose, before any cell mask expansion (expanded masks are exclusively used for spot assignment to cells). Cell length is the axis major length of an ellipse of equivalent shape and area to the cell mask, as computed by the `skimage.measure.regionprops` function. Cell volume is computed as the volume of a rod-shaped cell of length the axis major length and width the axis minor length, with spherical ends. The distribution of cell volume, area, length, and eccentricity is shown in Figure 11. While in our data cell size tends to increase with growth rate, as is expected in bacteria

([Bremer and Dennis, 2008], [Schaechter et al., 1958]), it is surprising that mean cell volume and area in LB are lower than in glucose, while at the same time, mean cell length is higher in LB than in glucose. When we inspected individual pictures, we saw that cells in LB looked substantially larger than cells in glucose, whereas cells in glucose were shorter and rounder (even shorter than in glycerol, agreeing with the plot). The exact cause of the issue is unclear, but one contributing factor is clearly the segmentation: the LB pictures were obtained from an early experiment, and suffer from artifacts that might cause noise in the data, making our LB data unreliable. Similarly, we found mean cell volume increases exponentially with growth rate for the AB medium conditions, but the relation does not hold for LB (Figure 12). It might be due to poor segmentation and volume estimation, or experimental variability, as the LB experiment was not performed on the same day as the experiment in AB.

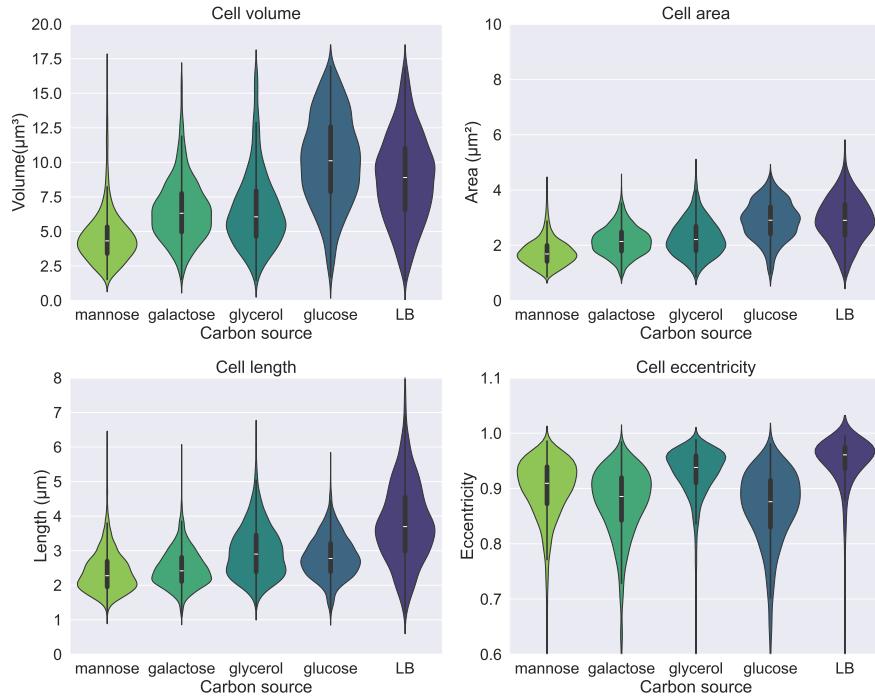


Figure 11: Distributions of cell volumes, area, length, and eccentricity in the different conditions, with growth rates increasing left to right. The estimated cell volume, length, and area tend to increase with growth rate, with cell volume and area in LB being surprisingly low. Cell eccentricity increases between mannose, glycerol and LB. Cells in galactose and glucose indeed tended to be rounder in our pictures, which results in lower eccentricity values.

rpoD expression increases with growth rate, whereas TA systems expression remain similar The distribution of expression (Figure 13) shows that expression at low growth rates is lower, both for rpoD and the two TA modules. The drop is sharper for rpoD. In rich medium, nearly all cells express rpoD, whereas only one in three cells express hipBA, and rnlAB is in-between the two. The trend is the same for mean expression per condition (Figure 14). Mean expression was computed as the mean number of mRNAs per cell, and also normalized per estimated cell length, cell area, and cell volume. Overall, for what concerns mean expression (Figure 14), we see that all four metrics show a similar trend of expression increasing with growth rate for rpoD, and staying roughly constant for hipBA and rnlAB.

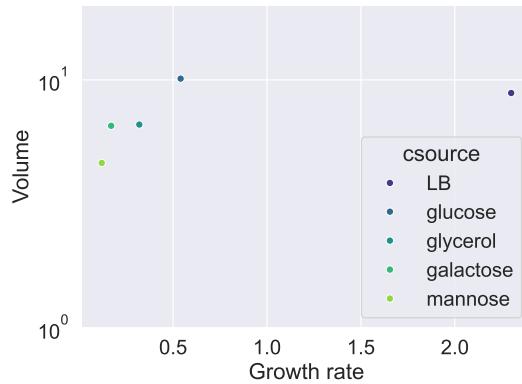


Figure 12: Cell volume as a function of growth rate. The estimated cell volume increases roughly exponentially with growth rate within the same medium (AB). LB medium is not on that line, pointing to either poor segmentation or volume estimation, or experimental variability issues

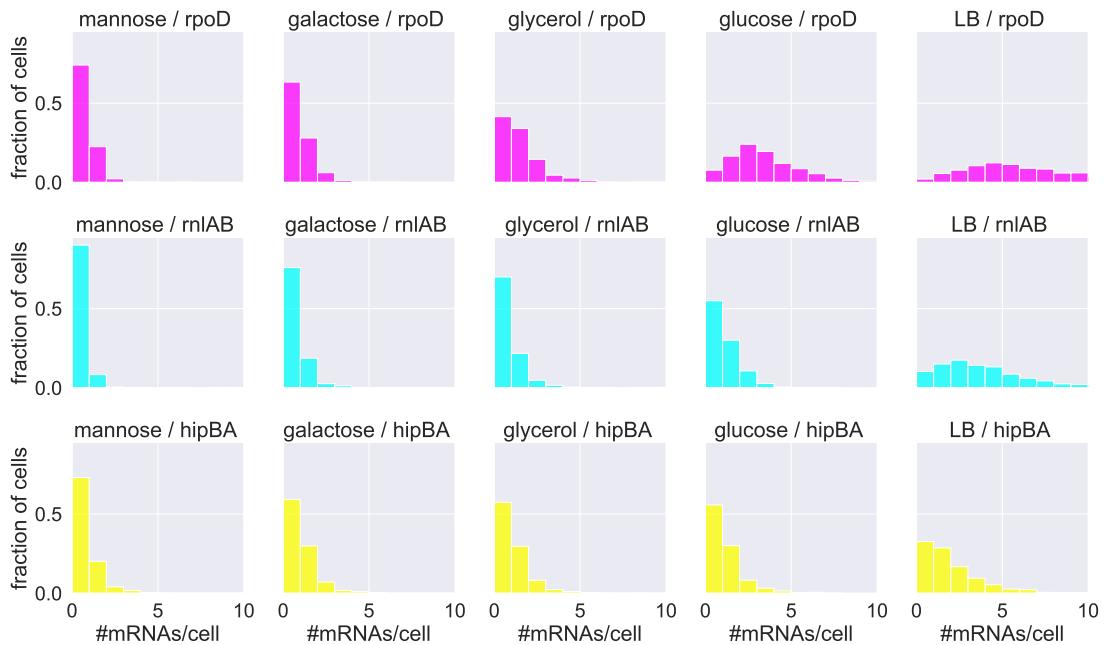


Figure 13: Distribution of expression per condition. Columns: condition (left to right: lower to higher growth rates), rows: gene. x-axis: number of detected mRNAs per cell, y-axis: fraction of cells with a given number of mRNAs. For all three RNAs, the fraction of non-expressing cells increases as the growth rate decreases. The increase is stronger for rpoD, which is expressed in most cells at higher growth rates. For the lower growth rates, most cells expressing one of the TA modules have only one mRNA, and this fraction is relatively constant for the middle range of growth rates (glucose, glycerol, galactose), before dropping for mannos. Even at the lowest growth rate, a fraction of cells keep expressing the three genes.

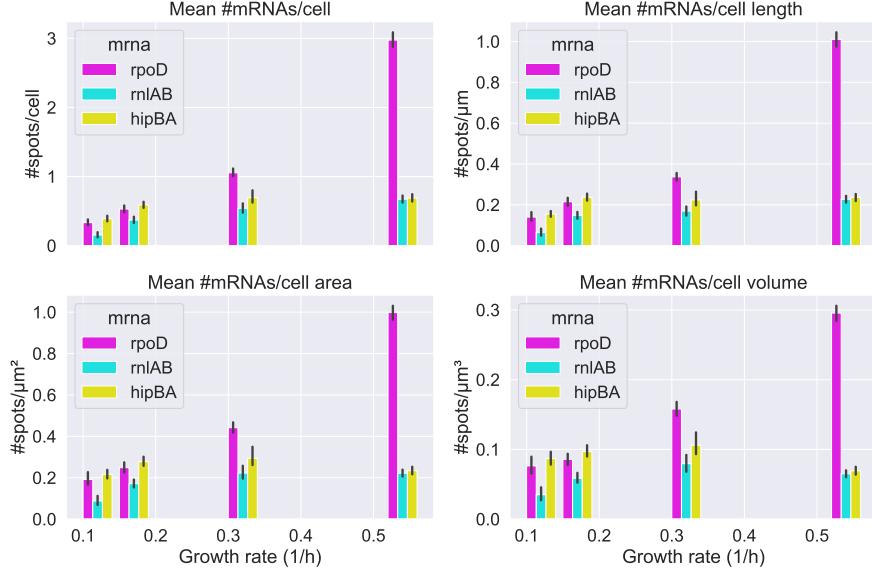


Figure 14: Mean expression per condition, Top left: number of mRNAs per cell, top right: number of mRNAs normalized per cell length, bottom left: number of mRNAs normalized by cell area, bottom right: number of mRNAs normalized by cell volume. For all four metrics, rpoD expression increases as growth rate increases, whereas rnlAB and hipBA remains roughly stable.

Data quality During spot assignment, spots that are detected but lie outside a cell mask are discarded as false positives. These spots may result from non specific binding, and constitute noise. To check data quality, we can compare the intensities of spots found inside and outside of cells. Ideally the intensities outside of cells should be lower than that inside the cells. Figure 15 shows the histograms for an example picture. In this picture, out-of-cells intensities are slightly lower than in-cell intensities, but not so much that the two distributions can be clearly separated. One issue with this analysis is that we did not aim for full recall when optimizing the segmentation, preferring high precision instead, so as to avoid noise in the mRNA counts. However, here, the false negatives from the segmentation, while of no influence on the counts provided they are uniformly distributed across cells, erroneously increase the out-of-cells counts. Another processing issue this plot shows is the selection of in-focus slices for the analysis is probably too broad, as a large number of out-of-cell spots tend to lie in one or two of the lower z-slices. Figure 16 show a side view of the slide, with the out-of-cell spots in red and in-cell spots in cyan. We inspected the picture with the DIC layer and found that most of the lower slice spots indeed lie out of the cells. Especially for rnlAB, it seems probes have a tendency to stick to the poly-L-lysine coated coverslip.

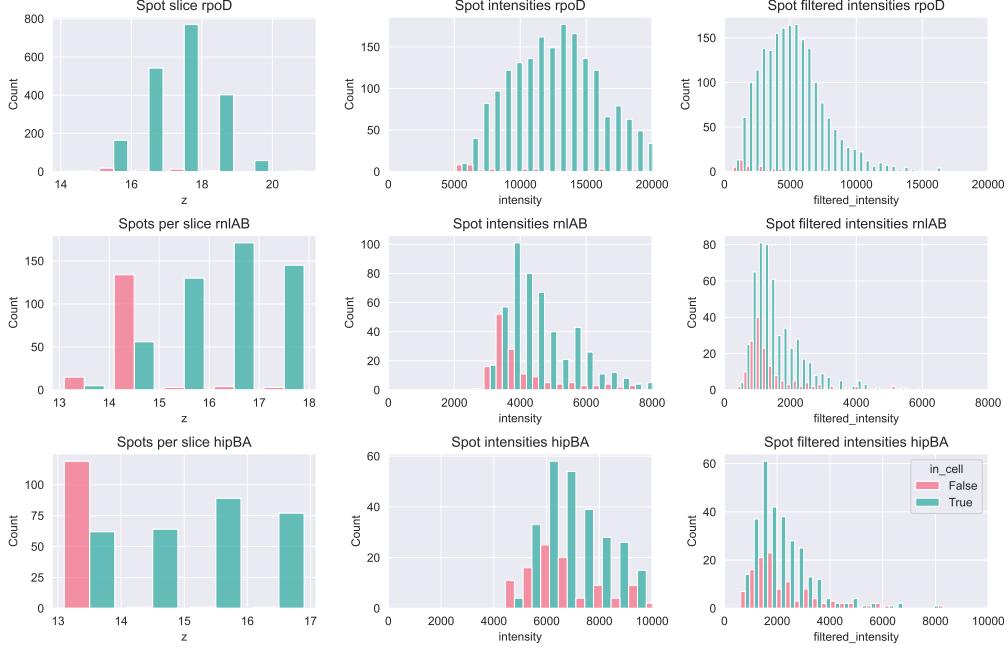


Figure 15: Spot intensities. Left column: z-slice where the spot was found, middle column: spot raw intensity, right column: filtered spot intensity. Top row: rpoD, middle row: rnlAB, bottom row: hipBA. For each plot, data is split into out-of-cell spots (red) and spots that are located in a cell (green). Out-of-cells spots are eliminated from the quantitative analysis. Too many out-of-cell spots will lead to a low signal to noise ratio, as the likelihood that a spot that was counted is actually a true spot decreases. Ideally we would like out-of-cell spots to have lower intensities overall than in-cell spots. For this picture, we do not observe full separation between the two populations based on the intensities. However, the number of spots outside the cells is significantly lower than inside. We also see that, especially for the two TA systems, out-of-cell spots tend to appear in slices that are at an extreme of the focus zone, pointing to a possible issue with the focus range chosen for analysis. The rnlAB histograms does exhibit two peaks, a large one at lower intensities and a smaller one at higher intensities. The presence of these two peaks is consistent with the second, stronger promoter on rnlAB the operon, targeting the antitoxin part of the gene exclusively: we would then expect a large peak for the shorter, rnlB transcripts, because we expect them to be more frequent due to their stronger promoter and their intensity would be lower because of their shorter length.

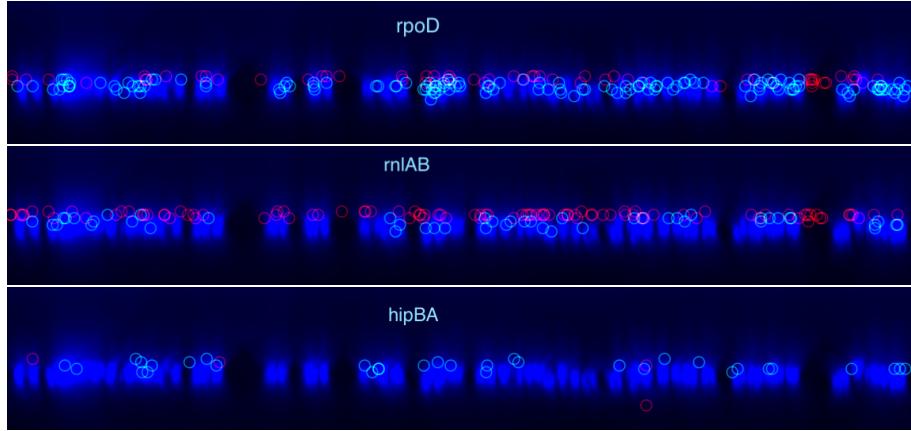


Figure 16: Sideway view of the slide with in- and out of cell spots. Detected spots that are deemed to be outside of cells are circled in red and discarded in the analysis, spots that are circled in cyan are deemed in-cell and counted for the analysis. The 3D DAPI signal is shown to hint at the approximate location of cells. The vertical dimension is the z-axis. Many red spots gather in the lower z-slices, as suggested by the Figure 14 plot. Manual examination of the picture, with DIC layer, suggests that these spots are often rightfully discarded.

hipBA spots seem to be often located close to the nucleoid tip From subjective examination it looks like hipBA transcripts are often located close to the distal tip of the nucleoid (see an example picture in Figure 17). This seems to be true in all conditions, and might be interesting to quantify by computing the distance to the center or tip of the cell.

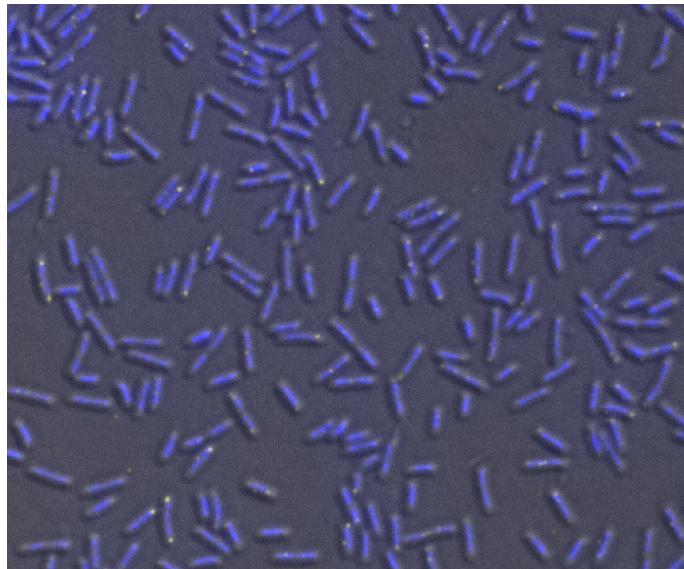


Figure 17: hipBA close to nucleoid tip

hipBA and rnlAB expression do not collocate in our data Figure 18 shows the distribution of cells expressing both hipBA and rnlAB. Our data does not suggest hipBA and rnlAB tend to be expressed in the same cells.

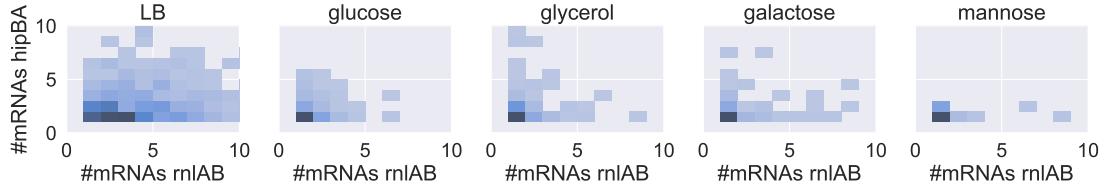


Figure 18: 2D histogram of cells expressing both *hipBA* and *rnlAB*.
Cells expressing neither one of the two TA systems were excluded.
Expression of the two gene does not correlate in our data.

3 Discussion

We expect volume to increase exponentially with growth rate ([Schaechter et al., 1958], [Neidhardt, 1999], [Bremer and Dennis, 2008], [Si et al., 2017]). This is approximately true for our AB medium experiment measurements, but is not valid on our data if we take LB into account (Figure 11). Both experiments, in LB and AB medium, were carried out on different days and possibly with slight differences in manipulations that could have influenced results ([Heinemann et al., 2020]). It is also possible that the segmentation masks underestimate cell area in LB, as we have optimized the segmentation postprocessing parameters mostly on AB medium pictures. Galactose cell volume is also quite high compared to glycerol. It was noticed during successive experiments that growth rates in galactose tended to exhibit the largest fluctuations. In a study over growth defects of *E.coli* strain MG1655, [Soupene et al., 2003] mentions that a mutation can turn the strain into a faster grower, which might explain some of the variation we saw.

While we have reported expression levels in absolute numbers of mRNAs per cell as function of growth rate, in a population of dividing cells, DNA replication is ongoing for most of the time, at least at high enough growth rates, and growth rate correlates with genome copy number ([Cooper and Helmstetter, 1968], [Mäkelä et al., 2024]). For slow growing cells, only one round of DNA replication occurs per cell cycle, and the genome copy number is 1. However, because DNA replication takes 40 minutes, at doubling times lower than 40 minutes the cell has to initiate replication more than once during one cell cycle. Using numbers from [Bremer and Dennis, 2008], we can estimate the genome copy numbers in our conditions to be 1 for the three slowest growth rates (mannose, galactose, glycerol), 1.8 for glucose, and 4 for LB. Even though our experimental method does not allow us to measure genome copy number accurately, we checked how many nucleoid masks were found by the segmentation of the DAPI layer (Figure 19), and found the agreement with the numbers from [Bremer and Dennis, 2008] to be surprisingly good.

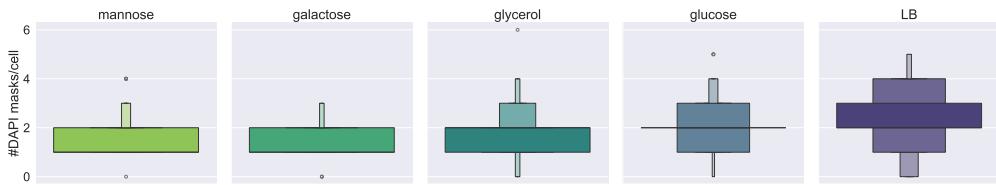


Figure 19: Distribution of the number of nucleoids per cells.

Our hypothesis was that *rpoD* transcription was regulated by growth, whereas TA systems were not. In our data, absolute numbers of mRNAs per cell are higher at higher growth rates. We can refine the analysis by taking gene dosage into account. How many copies a cell has of a gene depends on genome copy number and on the position of the gene on the bacterial chromosome: because replication always starts at the origin of replication (*oriC*), gene dosage for each chromosome depends on the distance between the gene and the origin of replication, with a gene closer to *oriC* having more copies present. The locations of all three genes are shown on Figure 20: *rpoD* is closest to *oriC*, at 69 min, *rnlAB* is further away at 59 min, and *hipBA* is opposite *oriC*, at 50 min, meaning it is one of the last genes to be copied in a replication round. At higher growth rates, we will thus have significantly more copies of *rpoD* than *hipBA*. Last, the concentration of RNA polymerase limits the rate of transcription in *E.coli*, and is also growth rate dependent. Here

again we used data from [Bremer and Dennis, 2008] to compute the expected mean copy number for rpoD, hipBA, and rnlAB at each growth rate, as well as the RNA polymerase concentration, resulting in the data shown in Figure 21. Plugging these numbers into the numbers of mRNAs per cell obtained from our experiments, we can compute the mean number of mRNAs per gene copy and the mean number of mRNAs per cell per RNAP molecule. The mean number of rpoD mRNAs per rpoD gene copy increases with growth rate, supporting the idea that rpoD is indeed growth regulated. Similarly, the ratio of rpoD mRNAs per RNA polymerase molecule remains constant as growth rate increases, suggesting that rpoD transcription increases with growth and the availability of RNA polymerase. For the TA modules, the number of mRNAs per gene copy remains approximately constant, whereas the number of mRNAs per RNAP decreases as growth rate and RNAP concentration increase. TA system transcription levels thus do not seem to depend strongly on growth rate in our data.

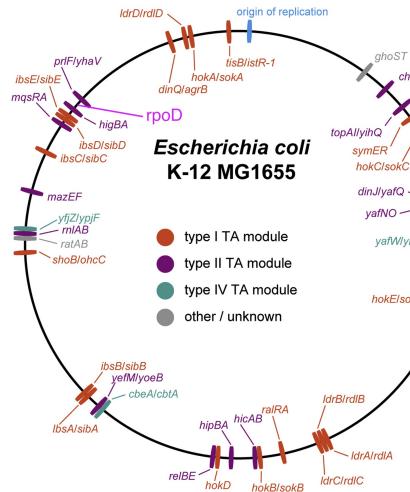


Figure 20: *rpoD* locus on the *E.coli* chromosome

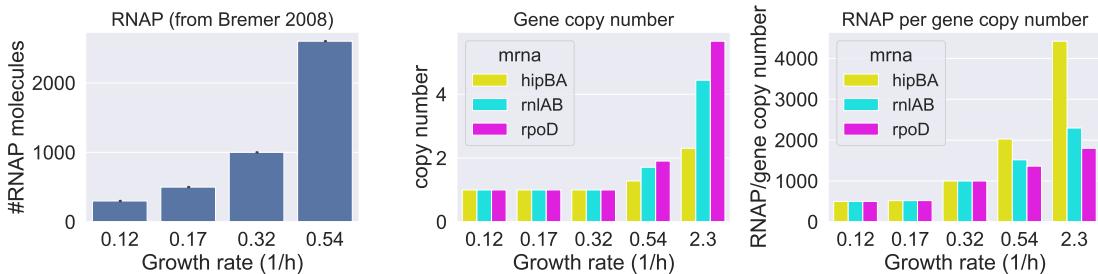


Figure 21: Left: RNA polymerase concentration. The transcription rate depends on the concentration of RNA polymerase, that increases with growth rate. Middle: gene copy number for all three genes increases with growth rate, with the gap between two genes increasing with growth rate and difference in distance from oriC. Right: RNA polymerase over gene copy number ratio. The ratio increases for all three genes, but less fast for rpoD, because of its higher copy number.

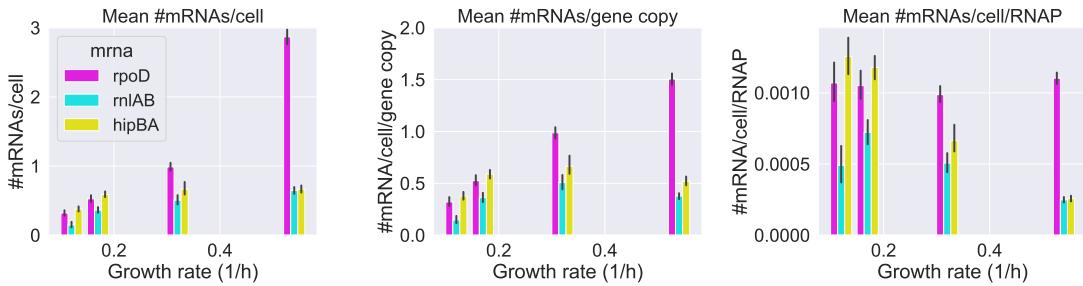


Figure 22: Mean normalized expression per gene copy and per RNAP molecule

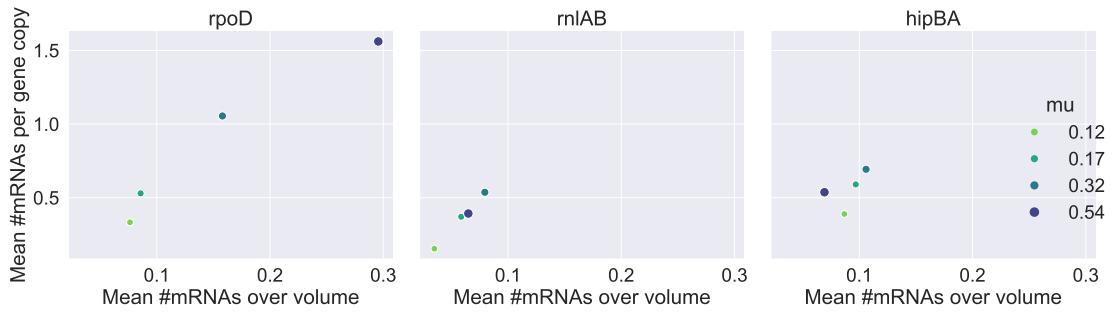


Figure 23: Mean expression per gene copy versus mean expression per volume

4 Outlook

For this project we extended an existing smFISH protocol ([Sorgucu, 2023]) to detect mRNAs from three genes instead of the previous two, and adapted a processing pipeline developed for *Candida Albicans* ([van Otterdijk et al., 2024]). While the results are interesting, the protocol should be optimized further to reduce the number of unbound rnlAB probes, which occurred in some of the experiments, and to make the pipeline more robust to segmentation errors and alignment issues between the DIC and fluorescent channels. One option would be to used brightfield or phase contrast images for the cell outlines, instead of DIC images. This would also make segmentation easier, as Omnipose offers a bacterial phase contrast model. This in turn would allow us to gather data on significantly more cells, which would be more reliable than the limited dataset we produced in this project. We build tooling to automate pipeline processing and record processing parameters to ensure repeatability, which was needed because we wanted to analyze large numbers of images, and some steps, like picking a suitable threshold for detecting spots, are time-consuming. We could improve the current tooling to make parameters more efficient and ergonomic, and also possibly automated threshold detection, for instance by using machine-learning models ([Xu et al., 2024]). Last, some of this tooling could be integrated into a Napari plugin for ease of use.

Part of the difficulties in processing the FISH pictures stems from the fact that we grew our cells in minimal media and as a consequence cells were relatively small. It would be valuable to redo the experiment with a medium supplemented with amino-acids, which would result in faster growth, larger cells, and would cover a wider range of growth rates.

It is by now acknowledged that bacteria can exhibit some degree of spatial localization of their transcriptome ([Moffitt et al., 2016], [Rombouts and Nollmann, 2020], [Das et al., 2021]), and that localization can help to fine-tune gene expression after transcription. We subjectively found hipBA spots to be frequently located at the tip of the nucleoid. This might be due to the spatial configuration of the chromosome during replication and the position of the hipBA gene. More research is needed to investigate whether there is indeed a pattern of hipBA transcript spatial localization, and if there is, what causes it and whether it has any consequences.

To further understand TA systems transcription regulation, we would like to investigate expression under stress conditions. To this end we have designed probes for the rpoS sigma factor that is characteristic of the general stress response in *E.coli* ([Landini et al., 2013], [Pletnev et al., 2015]), that could be used in the next phase of the project. In our FISH protocol this could be achieved for instance by fixing cells when they transition from exponential phase to stationary phase, experiencing nutrient stress.

Our approach suffers from inherent limitations because what we observe are mRNA transcripts in fixed cells. That means we do not have access to the dynamics of the transcription, nor to the post-transcriptional and translational steps that ultimately determine the relative concentrations of toxins and antitoxins that seem to determine entry into the persistence state. In addition, the TA systems transcripts are polycistronic, and it is unclear how much of the toxin or the antitoxin is made by a given transcript. It would probably make sense to try to broaden our methods to proteomics measurements in order to gain direct insights into the switching dynamics. Similarly, if localization plays a role, the resolution we have operated at is a clear limitation given the size of *E.coli* cells, and we should investigate whether using higher resolution approaches, like super-resolution microscopy ([Moffitt et al., 2016]), would be both feasible and fruitful.

5 Conclusions

In this project we produced smFISH pictures in which we detected mRNAs for the rpoD gene and the hipBA and rnlAB TA systems. After analysis, the available data supports the expectation that the transcription of rpoD is regulated by growth, in contrast to that of hipBA and rnlAB. At this stage, however, more work is needed to confirm this finding on a larger and more reliable dataset.

6 Methods

Cultures, medium, and carbon sources *Cultures:* we inoculated 50 mL Greiner tubes containing 10 mL Lennox-Broth medium with *E.coli* MG1655 cells grown on agar plates for fewer than 7 days. OD₆₀₀ was kept under 0.3 at all times to ensure cells were in exponential phase. After growing the cells in LB for at least 10 generations, we inoculated 250 mL shake flasks containing 20 mL of the target medium with the Greiner tube seed cultures and grew the cells at 37°C with 200rpm shaking for at least 10 generations. *Medium:* we used AB minimal medium (https://openwetware.org/wiki/AB_medium, see Annex ??) supplemented with trace elements, thiamine, and uracil to alleviate the pyrimidine deficiency of K-12 MG1655 ([Jensen, 1993]). *Carbon sources:* we used glucose, glycerol, galactose, and mannose as carbon sources to produce a range of growth rates. Carbon sources concentrations were based on the number of carbon atoms in the molecule: 10nM for glucose, galactose and mannose (C₆), and 20nM for glycerol (C₃). Plate reader experiments were carried out on Spectrostar machines with orbital shaking at 700 rpm, with OD₆₀₀ measurements taken every 5 minutes.

FISH experiments We followed the FISH protocol from [Tutucci and Singer, 2020], with three differences: 1. our cells are *E.coli* cells, for which we do not perform the lyticase digestion step, 2. we did not perform any immunofluorescence, 3. we have three probes, rpoD, hipBA, and rnlAB. The exact protocol we followed is attached in Annex A. For each FISH experiment, on day 1, cells growing in exponential phase were fixed in paraformaldehyde at an OD of 0.3, washed, attached to poly-L-lysine treated coverslips, permeabilized with ethanol, and stored at -20°C in ice-cold 70% ethanol. On day 2, cells were hybridized with the three probes. Cells were imaged on day 3. The full protocol is listed in Appendix A.

Probes Probes were designed using the Stellaris probe designer from Biosearch Technologies (<https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer>). Candidate probes returned by the Stellaris tool were checked for non-specific matches using Blastn.

rpoD Q570: taccacgggtgacaagaagt gaatcgacgatatcttccgg ggatgatgtctcgatctga cttcttcattcacacctaaatgc ttagttccacgctggaaag catttcacgcatacatgc gggtaacagttcaacagggtt ctttagcgtatgtcaatttcgc attcagcaacggcaggattga tgttccagcagataggtat tcttctgcttcaacacgatc gtgatcagatcggacagacg ctgcgttcgggtcaacaaag

Image acquisition Image acquisition was performed on an Olympus BX61 wide-field epifluorescence microscope with a 100x/1.35 NA Apo-chromatic objective, using the DAPI, CY3, CY3.5, and CY5 channels for the nucleoid, rpoD, rnlAB, and hipBA probes respectively. DIC pictures were acquired for the cell outlines. Pictures were shot using an sCMOS Hamamatsu Orca Fusion camera equipped with a 2304x2304 pixels sensor (sensor pixel size is $6.5 \mu\text{m} \times 6.5 \mu\text{m}$).

Image analysis Image analysis was performed using code based on the [van Otterdijk et al., 2024] pipeline. This code uses the BigFish library for spot detection ([Imbert et al., 2021]), and is available at https://github.com/LaToyaVanderB/Ecoli_smFISH

References

- [Balaban et al., 2019] Balaban, N. Q., Helaine, S., Lewis, K., Ackermann, M., Aldridge, B., Andersson, D. I., Brynildsen, M. P., Bumann, D., Camilli, A., Collins, J. J., Dehio, C., Fortune, S., Ghigo, J.-M., Hardt, W.-D., Harms, A., Heinemann, M., Hung, D. T., Jenal, U., Levin, B. R., Michiels, J., Storz, G., Tan, M.-W., Tenson, T., Van Melderen, L., and Zinkernagel, A. (2019). Definitions and guidelines for research on antibiotic persistence. *Nature Reviews Microbiology*, 17(7):441–448.
- [Balaban et al., 2004] Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625.
- [Barrett et al., 2019] Barrett, T. C., Mok, W. W. K., Murawski, A. M., and Brynildsen, M. P. (2019). Enhanced antibiotic resistance development from fluoroquinolone persisters after a single exposure to antibiotic. *Nature Communications*, 10(1).
- [Basan et al., 2020] Basan, M., Honda, T., Christodoulou, D., Hörl, M., Chang, Y.-F., Leoncini, E., Mukherjee, A., Okano, H., Taylor, B. R., Silverman, J. M., Sanchez, C., Williamson, J. R., Paulsson, J., Hwa, T., and Sauer, U. (2020). A universal trade-off between growth and lag in fluctuating environments. *Nature*, 584(7821):470–474.
- [Berkvens et al., 2022] Berkvens, A., Chauhan, P., and Bruggeman, F. J. (2022). Integrative biology of persister cell formation: molecular circuitry, phenotypic diversification and fitness effects. *Journal of The Royal Society Interface*, 19(194).
- [Black et al., 1991] Black, D. S., Kelly, A. J., Mardis, M. J., and Moyed, H. S. (1991). Structure and organization of hip, an operon that affects lethality due to inhibition of peptidoglycan or dna synthesis. *Journal of Bacteriology*, 173(18):5732–5739.
- [Bremer and Dennis, 2008] Bremer, H. and Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1).
- [Cooper and Helmstetter, 1968] Cooper, S. and Helmstetter, C. E. (1968). Chromosome replication and the division cycle of escherichia coli. *Journal of Molecular Biology*, 31(3):519–540.
- [Cutler et al., 2022] Cutler, K. J., Stringer, C., Lo, T. W., Rappez, L., Stroustrup, N., Brook Peterson, S., Wiggins, P. A., and Mougous, J. D. (2022). Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nature Methods*, 19(11):1438–1448.
- [Das et al., 2021] Das, S., Vera, M., Gandin, V., Singer, R. H., and Tutucci, E. (2021). Intracellular mrna transport and localized translation. *Nature Reviews Molecular Cell Biology*, 22(7):483–504.
- [Falla and Chopra, 1998] Falla, T. J. and Chopra, I. (1998). Joint tolerance to -lactam and fluoroquinolone antibiotics in escherichia coli results from overexpression of hipa. *Antimicrobial Agents and Chemotherapy*, 42(12):3282–3284.
- [Harms et al., 2018] Harms, A., Brodersen, D. E., Mitarai, N., and Gerdes, K. (2018). Toxins, targets, and triggers: An overview of toxin-antitoxin biology. *Molecular Cell*, 70(5):768–784.
- [Heinemann et al., 2020] Heinemann, M., Basan, M., and Sauer, U. (2020). Implications of initial physiological conditions for bacterial adaptation to changing environments. *Molecular Systems Biology*, 16(9).
- [Imbert et al., 2021] Imbert, A., Ouyang, W., Safieddine, A., Coleno, E., Zimmer, C., Bertrand, E., Walter, T., and Mueller, F. (2021). Fish-quant v2: a scalable and modular analysis tool for smfish image analysis.
- [Jensen, 1993] Jensen, K. F. (1993). The escherichia coli k-12 “wild types” w3110 and mg1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyre expression levels. *Journal of Bacteriology*, 175(11):3401–3407.
- [Koga et al., 2011] Koga, M., Otsuka, Y., Lemire, S., and Yonesaki, T. (2011). Escherichia coli rnlA and rnlB compose a novel toxin–antitoxin system. *Genetics*, 187(1):123–130.

- [Korch et al., 2003] Korch, S. B., Henderson, T. A., and Hill, T. M. (2003). Characterization of the hipA7 allele of escherichia coli and evidence that high persistence is governed by (p)ppgpp synthesis. *Molecular Microbiology*, 50(4):1199–1213.
- [Landini et al., 2013] Landini, P., Egli, T., Wolf, J., and Lacour, S. (2013). sigma σ scp σ scp σ , a major player in the response to environmental stresses in σ scp σ scp σ escherichia coli: role, regulation and mechanisms of promoter recognition. *Environmental Microbiology Reports*, 6(1):1–13.
- [LeRoux et al., 2020] LeRoux, M., Culviner, P. H., Liu, Y. J., Littlehale, M. L., and Laub, M. T. (2020). Stress can induce transcription of toxin-antitoxin systems without activating toxin. *Molecular Cell*, 79(2):280–292.e8.
- [Levin-Reisman et al., 2017] Levin-Reisman, I., Ronin, I., Gefen, O., Braniss, I., Shores, N., and Balaban, N. Q. (2017). Antibiotic tolerance facilitates the evolution of resistance. *Science*, 355(6327):826–830.
- [Lewis, 2010] Lewis, K. (2010). Persister cells. *Annual Review of Microbiology*, 64(1):357–372.
- [Lewis and Shan, 2017] Lewis, K. and Shan, Y. (2017). Why tolerance invites resistance. *Science*, 355(6327):796–796.
- [Mäkelä et al., 2024] Mäkelä, J., Papagiannakis, A., Lin, W.-H., Lanz, M. C., Glenn, S., Swaffer, M., Marinov, G. K., Skotheim, J. M., and Jacobs-Wagner, C. (2024). Genome concentration limits cell growth and modulates proteome composition in escherichia coli.
- [Moffitt et al., 2016] Moffitt, J. R., Pandey, S., Boettiger, A. N., Wang, S., and Zhuang, X. (2016). Spatial organization shapes the turnover of a bacterial transcriptome. *eLife*, 5.
- [Moyed and Bertrand, 1983] Moyed, H. S. and Bertrand, K. P. (1983). hipA, a newly recognized gene of escherichia coli k-12 that affects frequency of persistence after inhibition of murein synthesis. *Journal of Bacteriology*, 155(2):768–775.
- [Neidhardt, 1999] Neidhardt, F. C. (1999). Bacterial growth: Constant obsession with dn/dt. *Journal of Bacteriology*, 181(24):7405–7408.
- [Page and Peti, 2016] Page, R. and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature Chemical Biology*, 12(4):208–214.
- [Pletnev et al., 2015] Pletnev, P. I., Osterman, I. , Sergiev, P. V., Bogdanov, A. , and Dontsova, O. (2015). Survival guide: Escherichia coli in the stationary phase. *Acta Naturae*, 7(4):22–33.
- [Raj et al., 2008] Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mrna molecules using multiple singly labeled probes. *Nature Methods*, 5(10):877–879.
- [Rombouts and Nollmann, 2020] Rombouts, S. and Nollmann, M. (2020). Rna imaging in bacteria. *FEMS Microbiology Reviews*, 45(2).
- [Ronneau and Helaine, 2019] Ronneau, S. and Helaine, S. (2019). Clarifying the link between toxin–antitoxin modules and bacterial persistence. *Journal of Molecular Biology*, 431(18):3462–3471.
- [Rotem et al., 2010] Rotem, E., Loinger, A., Ronin, I., Levin-Reisman, I., Gabay, C., Shores, N., Biham, O., and Balaban, N. Q. (2010). Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proceedings of the National Academy of Sciences*, 107(28):12541–12546.
- [Schaechter et al., 1958] Schaechter, M., MaalOe, O., and Kjeldgaard, N. O. (1958). Dependency on medium and temperature of cell size and chemical composition during balanced growth of salmonella typhimurium. *Journal of General Microbiology*, 19(3):592–606.
- [Si et al., 2017] Si, F., Li, D., Cox, S. E., Sauls, J. T., Azizi, O., Sou, C., Schwartz, A. B., Erickstad, M. J., Jun, Y., Li, X., and Jun, S. (2017). Invariance of initiation mass and predictability of cell size in escherichia coli. *Current Biology*, 27(9):1278–1287.
- [Skinner et al., 2013] Skinner, S. O., Sepúlveda, L. A., Xu, H., and Golding, I. (2013). Measuring mrna copy number in individual escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nature Protocols*, 8(6):1100–1113.

- [So et al., 2011] So, L.-h., Ghosh, A., Zong, C., Septúlveda, L. A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in escherichia coli. *Nature Genetics*, 43(6):554–560.
- [Sorgucu, 2023] Sorgucu, M. (2023). Developing single molecule rna fish for the identification of antibiotics tolerant persister bacteria. *Master Thesis, VU Systems Biology group, internal*.
- [Soupene et al., 2003] Soupene, E., van Heeswijk, W. C., Plumbridge, J., Stewart, V., Bertenthal, D., Lee, H., Prasad, G., Paliy, O., Charernnoppakul, P., and Kustu, S. (2003). Physiological studies of escherichia coli strain mg1655: Growth defects and apparent cross-regulation of gene expression. *Journal of Bacteriology*, 185(18):5611–5626.
- [Taniguchi et al., 2010] Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538.
- [Tunnacliffe and Chubb, 2020] Tunnacliffe, E. and Chubb, J. R. (2020). What is a transcriptional burst? *Trends in Genetics*, 36(4):288–297.
- [Tutucci and Singer, 2020] Tutucci, E. and Singer, R. H. (2020). *Simultaneous Detection of mRNA and Protein in S. cerevisiae by Single-Molecule FISH and Immunofluorescence*, page 51–69. Springer US.
- [van Otterdijk et al., 2024] van Otterdijk, S., Motealleh, M., Wang, Z., Visser, T. D., Savakis, P., and Tutucciu-fish, E. (2024). *Single-Molecule Fluorescent In Situ Hybridization (smFISH) for RNA Detection in the Fungal Pathogen Candida albicans*, page 25–44. Springer US.
- [Xu et al., 2024] Xu, W., Cai, H., Zhang, Q., Mueller, F., Ouyang, W., and Cao, G. (2024). U-fish: a universal deep learning approach for accurate fish spot detection across diverse datasets.

Acknowledgments

I would like to address deep thanks to Evelina and Frank for giving me the opportunity to work on this project and supporting me along all meanders, Johan, Philipp, Sander and Wang for much invaluable practical help and stimulating discussions, Kelly for taking to me from zero to FISH over the course of two months, Teresa for taking our first FISH baby steps together, and the whole technicians team for their generous availability and constant cheerfulness, and generally remaining cool in the face of cluelessness.

— ”Nevertheless, they persisted”

A smFISH E.coli protocol

E.coli Single mRNA Resolution Fluorescent In Situ Hybridization

Singer lab, updated by Evelina Tutucci May 8th, 2018
Student version – by Kelly van Rossum (updated May 28th, 2024)

Cell Fixation, Preparation, and Storage

Materials

- Paraformaldehyde 32% solution, EM grade (Electron Microscopy Science #15714)
- Ribonucleoside–vanadyl complex (VRC; NEB #S1402S)
- Sorbitol 3M
- Potassium phosphate buffer (KHPO4; Gomori buffer), pH 7.5
- 70% ethanol
- Non-coated coverslips (Fisherbrand Borosilicate Glass Circle Coverslips No. 1: 0.16– 0.19 mm thick; size: 16 mm (#12323148) **or** Marienfeld Coverslips thickness 1.5H (0.17 ± 0.005 mm) round, size: 18 mm (#LH23.2))
- Poly- L-lysine (#P8920)
- 12-well cell culture plates

Solutions to be prepared:

Buffer B

1.2 M sorbitol,
100 mM KHPO4, pH 7.5

Resuspension buffer (Optional)

1.2 M sorbitol
100 mM KHPO4, pH 7.5
20 mM ribonucleoside–vanadyl complex

Protocol:

1. Grow 25 ml cells in LB, M9 or other medium at 37°C to an OD₆₀₀ of 0.3-0.4 (**Do not overgrow**).
 - a. When overgrown: dilute the cells to OD₆₀₀ of 0.1 and let them grow until OD₆₀₀ = 0.3-0.4 to ensure cells are in exponential phase
2. Prepare a 50 mL Falcon tube containing 3.15 ml of 32% (v/v) paraformaldehyde.
3. Fix cells by transferring 21.85 ml of culture to a 50 mL tube containing the paraformaldehyde (final concentration of 4%, v/v) and mix.
4. Incubate cells for 30 min at room temperature on a tabletop shaker. Pre-cool the centrifuge 15 mins prior to the end of the incubation.
5. Collect cells at 3500-4000 rpm for 3 mins at 4°C.
6. Wash cells two times with 10 mL of cold Buffer B.
7. Resuspend cells in max. 1 mL* Buffer B and transfer cells to a 1.5 mL Eppendorf tube, keep on ice. (*for 6 coverslips; otherwise use approximately 100-150 µL per coverslip)
8. Attaching cells to coverslips: Place poly-L-lysine treated 18 mm round coverslips face up into 12-well tissue culture dishes, one coverslip per well (for 25 mL culture use 6 coverslips).
9. Drop ~125 µL of cells to the center of a coated coverslip.
10. Let cells settle for 30 min-1 h at 4°C.
11. **Slowly** add 2 ml of buffer B to each well, then remove buffer B using a vacuum aspirator. This will remove cells not attached to the coverslip and leave a monolayer of immobilized cells.
12. **Slowly** add 2 ml of ice-cold 70% ethanol of each well.
13. Store cells for at least overnight at -20°C. Cells can be stored at -20°C for at least 6 months.

Prepare poly-L-lysine coverslips

- Carefully put one box of 18 mm round coverslips into 500 ml 0.1 M HCl and boil for 10 min.
- Rinse extensively with H₂O
- Autoclave and store in 70% ethanol at 4°C
- When necessary: take the coverslips out of the ethanol with new gloves
- Dry completely on chromatography paper
- Wash once with H₂O
- To coat coverslips with poly-L-lysine, place 100 µL of a 0.01% (w/v) poly-L-lysine solution onto a coverslip, incubate at room temperature for at least 15 min
- Remove the solution using a vacuum pump and **let the remaining liquid dry completely**. Then wash once with H₂O and **let air dry completely**.

Hybridization

Only very low probe concentrations are needed in the hybridization reaction to allow single mRNA detection. Generally, 5 ng per dye-conjugated DNA probes per hybridization reaction is sufficient. To block nonspecific binding of the probes, competitor DNA and RNA is added in large excess to the hybridization solution.

The formamide concentration in the hybridization mix and the subsequent wash steps is critical to get optimal hybridization specificity. Generally, we use 40% formamide for standard probes (50 nt/50% CG), but if high background is observed, increasing the formamide concentration from 40% to 50% can reduce background. To detect the entire pool of polyA, mRNAs in the cell can be detected using a 50-nt poly-dT probe, but the formamide concentration has to be reduced to 15%. For 20nt probes from Biosearch, we use standard concentration suggested by the producer (usually 100-150 mM) and 10% formamide.

Materials

- Glass plate, about 20x20 cm
- Parafilm
- Cardboard spacers
- 12-well cell culture plates
- Glass slide
- Formamide for analysis ACROS organics 99.5% code #205821000 (100 ml)

Solutions to be prepared:

- 10% formamide/2xSSC
- 2xSSC/0.1% Triton X-100
- 2x SSC and 1xSSC
- 1xPBS
- *Solution F* (10% formamide, 10 mM NaHPO₄, pH 7.5)

- *Solution H* (2xSSC, 2 mg/ml BSA, 10 mM VRC)
- Escherichia coli tRNA 10mg/ml (Roche # 10 109 541 001)
- ssDNA 10mg/ml (deoxyribonucleic acid, single stranded from salmon testes, Sigma #D9156)
- tRNA/ssDNA mix 10mg/ml final (mix ssDNA 10mg/ml and ssDNA 10mg/ml 1:1)
- Ethanol absolute
- DAPI solution (0.5 µg/ml DAPI (Sigma #D9564) in 1xPBS. Store at 4C in the dark)
- Mounting solution (ProLong Gold antifade reagent (Invitrogen #P36934)).

Please note: the following solutions are for 8 coverslips. Adjust the amounts if necessary.

Solution F 20 µL Formamide (10% final after mixing F+H)
 5 µL 200 mM NaHPO₄ pH7.5
75 µL H₂O
 100 µL

Solution H

| Final formamide concentration in hybridization | 10% | 50% |
|--|--------|--------------|
| H ₂ O | 50 µL | 30 µL |
| 20xSSC | 20 µL | 20 µL |
| BSA (20mg/ml) | 20 µL | 20 µL |
| VRC (200 mM) | 10 µL | 10 µL |
| Formamide | ----- | <u>20 µL</u> |
| | 100 µL | 200 µL |

Protocol

1. Put heat block at 95°C
2. Remove the ethanol from the 12-well plate using a vacuum pump and rehydrate samples by adding 2 mL 2xSSC at RT for 5 min. Do this twice.
or
 Transfer the necessary amount of coverslips to a new 12 wells plate and then rehydrate the samples twice by adding 2 mL 2xSSC at RT for 5 min.
3. Wash cells once with 10% formamide/2xSSC at RT for 30 min. During washes, prepare the hybridization mix:
Steps 4 – 7 are for one coverslip. Adjust amounts if necessary. Remember when working with light-sensitive probes, cover the probes as much as possible with aluminum foil (or any equivalent).
4. Mix 5-10 ng of each probe mix per coverslip with 5µL of the 10mg/ml E. coli tRNA/ssDNA mix (10 ng of probe mix when using four probes with different dyes against one gene). For Biosearch probes: stock 25 µM add probes mix to obtain a final concentration of 125 nM in 25 µl (F+H buffer)
5. Lyophilize using a SpeedVac (1 – 2 mins per µL)
6. Add 12.5 µL of solution F to probe tube, heat at 95°C for 2 min, cool at RT.
7. Add 12.5 µL of solution H to the hybridization mix.
8. Put a drop of 20 µL of hybridization mix onto the Parafilm stretched out in the hybridization chamber. Avoid bubbles in the hybridization mix. (Use the back of a forceps to scratch the edges of the Parafilm so that the Parafilm sticks to the hybridization chamber.)

9. Using forceps, place the coverslip with cells facing down on the hybridization mix. No bubbles should form. Multiple coverslips can be placed next to each other onto a single glass plate, but leave about 1.5 cm space between coverslips. Put a small cap of a Greiner tube with H₂O in the corner of the hybridization chamber.
10. To seal the “hybridization chamber,” parafilm the sides of the hybridization chamber. Cover with aluminum foil.
11. Incubate at 37°C overnight in the dark for 50mer probes mix. For 20mer probes, with a 10% formamide mix incubate 3-5 hrs at 37°C. Put the remaining 10% formamide/2xSSC at 37°C for subsequent washes.
12. Place cover slips back in 12-well tissue culture dish containing 10% formamide/2xSSC, cells facing up; remove ‘old’ 10% formamide/2xSSC and replace with ‘new’ 10% formamide/2xSSC & incubate 15 min at 37°C (incubator).
13. Wash once more with 10% formamide/2xSSC at 37°C (2 mL, 15 min).
14. Wash once with 2xSSC at RT (2 mL, 5 min).
15. Wash once with 2xSSC/0.1% Triton X-100 at RT (2 mL, 5 min).
16. Wash once with 1xSSC at RT (2 mL, 10 min).
17. ((IF DOING Immuno fluorescence **STOP** here and start IF)
18. Wash 1x with 1xPBS (2 ml, 10 min).
19. Before mounting, dip coverslip in 100% EtOH, let them dry.
20. Remove first air bubbles of the mounting solution on a paper tissue. Invert cells facing down onto a drop of mounting solution with DAPI placed on a glass slide. Allow the mounting solution to polymerize over night at room temperature in the dark.
21. Seal coverslips with nail polish. Let nail polish dry before imaging, otherwise the objective may be damaged.
22. Go to the microscope and enjoy your images. Slides can be stored at 4°C for a few days and at -20°C for months in the dark.

B Medium

| Thiamine (1000x) | MW | g | Conc N (mM) | Final Volume (L) |
|-------------------------|-----------|----------|--------------------|-------------------------|
| Thiamine Hydrochloride | 337,27 | | 0,001 | 0,3 0,01 |
| Uracil (100x) | MW | g | Conc N (mM) | Final Volume (L) |
| Uracil | 112,0868 | | 1,12 | 20,0 0,5 |

Sauer Trace elements (4000x)

| | | | | |
|--------------|-------|--------|----|--------|
| FeCl3*6H2O | T1 | 2,7 g | | |
| ZnCl2 | T2 | 0,27 g | | |
| CoCl2*6H2O | T3 | 0,2 g | | |
| Na2MoO4*2H2O | T4 | 0,2 g | | |
| CaCl2*2H2O | T5 | 0,1 g | | |
| CuCl2*2H2O | T6 | 0,17 g | | |
| H3BO3 | T7 | 0,05 g | | |
| Concentrated | HCl | 37% T8 | | 10 mL |
| miliQ | water | up | to | 100 ml |

AB medium base

| | | |
|-------------|------------|---------------|
| Na2(PHO4) | 6 g/l | Source of P |
| K(PH2O4) | 3 g/l | Source of P |
| NaCl | 3 g/l | |
| (NH4)2(SO4) | 2 g/l | Source of N S |
| MgCl2 | 0.2 g/l | |
| Na2(SO4) | 0.011 g/l | |
| CaCl2 | 0.01 g/l | |
| FeCl3 | 5.0e-4 g/l | |