Title

Author1[1], Author2[2], Author3[3,*]
1 Author1 Dept/Program/Center, Institution Name, City, State, Country
2 Author2 Dept/Program/Center, Institution Name, City, State, Country
3 Author3 Dept/Program/Center, Institution Name, City, State, Country
∗ **E-mail: Corresponding author@institute.edu**

Abstract

Author Summary

Introduction

Single cells measurements allow studying biological processes that cannot be observed at the level of unsynchronized populations. Some examples include intrinsic and extrinsic noise (Elowitz 2002), cycles and . Among the promising techniques for single cell measurements is florescent in situ hybridization (FISH) pioneered by Robert Singer () and Alexander van Oudenaarden().

A major appeal of FISH is the ability of count mRNAs corresponding to different genes without making it's scalability to medium and high-throughput To fully realize those promises, we need robust and reliable algorithms to ... as first developed by Singer at al. In this paper we build upon and those first steps, and extend them by adding new features, including automated cell identification and budding index quantification. All these features are integrated into user friendly software package that can be used both interactively (via GUI) and from scripts for high-throughput image analysis.

Results

Identifying Cell and Cell Buds

Singer at. al. identified cells manually (). We use a fully automated approach to cell identification based on DAPI staining of the nucleus and the autofluorescence of the cells.

Furthermore, for isolated cells (cells in whose vicinity there are no other cells on the image) we compute the probability of being budded based on 3 criteria: (1) cell ellipticity (the mother and daughter cells form more elongated body than non-budded cells); (2) concave curvature of the membrane at the division furrow (rather than convex everywhere else); (3) for late stage budded cells two DAPI stained regions connected by cell cytoplasm, an isthmus of high-autofluorescence. A short description how we do

it...

Fig.1 with 3 panels: 1) Cells, 2) Budded cell, 3) Scatter plot of correspondence between human counting and computer counting

Finding and quantifying spots

algorithm description We developed and tested two algorithms for identifying spots and quantifying their intensities:

Algorithm A: The two most notable features (and differences compared to previously employed approaches) are:

plots of the multi-probe labeled mRNAs

Robust Probabilistic Spot Analysis

Ideally, the separation between the mode corresponding to single probes (M1) and the mode corresponding to multi-labeled mRNAs (M2) should be complete. If the separation is not perfect, using a hard threshold is likly to introduce false positive and negative assigments (mRNAs will not be counted or single non-hybridized probes will be counted as mRNAs). The bigger the overlap between the two modes (M1 & M2), the bigger the error in mRNA quantification. A second problem with a hard threshold on the intensity of spots is that the position of the threshold can depend strongly on numerous parameters such as incident light intensity, efficiency of probe labeling, spectral filters, fluorophore quantum efficiency, and even sample preparation. Therefore, establishing a good threshold might require a set of control experiments specific to the equipment and every set of samples, or human decision (and the potential bias) about the threshold position on every single experiment.

To mitigate those problems, we develop a simple approach based upon the conditional probability that the j^{th} spot is mRNA $p(X_j = 1)$ given its intensity, I_j. The key assumption behind our approach is that there are no mRNAs (or very few mRNAs) outside of cells. This assumption is strongly supported by the data in most experiments and in the experiments when it is violated (because cell bursting during cell wall digestion and immobilization) can be avoided by using extracelluar spots from the experiments that worked well. When the assumption is correct, all spots outside of cells correspond to single probes and the empirical cumulative distribution of their intensities characterizes the probability for a spot with a given intensity to correspond to a single-probe. For example, a spot within a cell whose intensity is higher than the intensities of all spots outside of cells has a probability of being a single-probe equal to $1/N$, where N is the number of spots outside of cells. If many experiments are performed using the same equipment

and sample preparation, all extracellular spots (from all experiments) for a dye (such as cy3) can be pulled together and used as the null distribution of intensities of single probes. Formally, the conditional probability $p(X_j = 1|I_j)$ for the j^{th} spot to be a mRNA can be written as $p(X_j = 1|I_\omega) = \mathcal{D}(I_j|I_\omega)$. Here $\mathcal{D}(I_\omega)$ is the empirical cumulative distribution for the set of spots (ω) that are outside of cell boundaries.

Using $p(X_j = 1|I_j)$ we compute both Bonferroni corrected *p–values* and *q–values* that can be used to select the spots likly to correspond to mRNAs while keeping the false discovery rate (FDR) below a defined level, such as 5%.

Quantifying the Numper of mRNAs per Cell

In the previous subsection, we outlined an approach for quantifying the probability for the j^{th} spot to be a mRNA, $p(X_j = 1|I_j)$. Next, we want to use these probabilities for each spot ($p(X_j = 1|I_j)$) for computing the marginal probabilities for the distribution of the k^{th} gene in the i^{th} cells, that is probability that the i^{th} cell contains n mRNAs from the k^{th} gene, $p(Y_{ik} = n)$. Assuming that the $p(X_j = 1|I_j)$ are independent from each other, $p(Y_{ik} = n)$ follows a multinomial distribution whose expectations are $p(X_j = 1|I_j)$. Givven idependence of the error in indentifying the mRNAs for different genes, the joing probabilities for the k^{th} and the l^{th} mRNAs can be computted as the product of the correponding marginal probabilities, $p(Y_{ik} = n, Y_{il} = n) = p(Y_{ik} = n)p(Y_{il} = n)$.

Discussion

Materials and Methods

Acknowledgments

References

Figure Legends

Tables