Methods 85 (2015) 44–53

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Computer vision for image-based transcriptomics

Thomas Stoeger [a,b,1], Nico Battich [a,b,1], Markus D. Herrmann [a,b], Yauhen Yakimovich [a], Lucas Pelkmans [a,*]

[a] Faculty of Sciences, Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[b] Life Science Zurich Graduate School, Ph.D. program in Systems Biology, Switzerland

ARTICLE INFO

Article history:
Received 19 February 2015
Received in revised form 13 April 2015
Accepted 17 May 2015
Available online 23 May 2015

Keywords:
Image-based transcriptomics
Single-molecule
Single-cell
Segmentation
Localization
Subcellular
High-throughput
FISH
In situ hybridization

ABSTRACT

Single-cell transcriptomics has recently emerged as one of the most promising tools for understanding the diversity of the transcriptome among single cells. Image-based transcriptomics is unique compared to other methods as it does not require conversion of RNA to cDNA prior to signal amplification and transcript quantification. Thus, its efficiency in transcript detection is unmatched by other methods. In addition, image-based transcriptomics allows the study of the spatial organization of the transcriptome in single cells at single-molecule, and, when combined with superresolution microscopy, nanometer resolution. However, in order to unlock the full power of image-based transcriptomics, robust computer vision of single molecules and cells is required. Here, we shortly discuss the setup of the experimental pipeline for image-based transcriptomics, and then describe in detail the algorithms that we developed to extract, at high-throughput, robust multivariate feature sets of transcript molecule abundance, localization and patterning in tens of thousands of single cells across the transcriptome. These computer vision algorithms and pipelines can be downloaded from: https://github.com/pelkmanslab/ImageBasedTranscriptomics.

1. Image-based transcriptomics is unique in several ways

In the past few years a wealth of techniques have been developed to study genome-wide transcriptional output at the single-cell level [1–7]. In contrast to methods relying on sequencing or PCR, image-based transcriptomics visualizes single transcripts in a population of single cells *in situ*. This allows not only the absolute quantification of transcript copy numbers, but also the spatial mapping of transcript molecules to the sub-cellular microenvironment [4]. Being an *in situ* technology, it does not require homogenization of cells and therefore minimizes the loss of material, thus achieving very high detection efficiency [4]. Another advantage of image-based transcriptomics is that it can be combined with the phenotypic characterisation of each single cell and its context within a population of cells or tissue, by microscopic assays and stainings commonly used in cell and developmental biology. This makes image-based transcriptomics of particular interest when studying the localization dynamics of the transcriptome in response to stimuli or perturbations and to identify sources of cell-to-cell variability in these processes [8,9]. While establishing image-based transcriptomics, we soon realized that a robust computer vision pipeline was as important as the experimental platform for accurately identifying and characterizing each single transcript molecule within a cell. Therefore, we here describe in detail our recent computer vision algorithms that result in accurate detection of objects in spinning disk confocal microscopy images. Besides providing a robust guide for identifying billions of individual transcript molecules with little hands-on user time, we describe how to unlock functionally important parameters of gene expression, which are impossible to grasp without the power of computer vision. For instance, multivariate descriptors of the position of each single transcript molecule enable an unsupervised characterization of the localization of transcripts of every cell.

1.1. General outline

Image-based transcriptomics employs multi-well plates to stain cells in parallel with specific probes against a transcript of interest (Fig. 1). Within single wells of a multi-well plate, the transcripts of different genes are stained by an automated experimental procedure. Each single transcript molecule is detected by high-throughput microscopy and computer vision. Experimental and computational steps can be performed with equipment that is commonly used for image-based high-throughput assays.

* Corresponding author.
E-mail address: lucas.pelkmans@imls.uzh.ch (L. Pelkmans).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ymeth.2015.05.016
1046-2023/

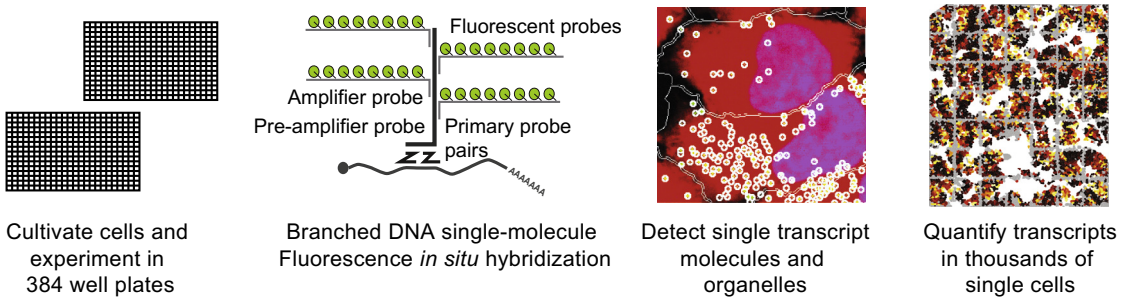

Fig. 1. Outline of image-based transcriptomics using bDNA sm-FISH.

Each single transcript molecule is stained by branched DNA single-molecule *in situ* hybridization (bDNA sm-FISH). This technology, which is commercially available from Affymetrix and Advanced Cell Diagnostics, applies a series of consecutive *in situ* hybridizations, which visualize each single transcript molecule as a bright fluorescent spot. In a first round of *in situ* hybridization, two epitope-specific primary probes bind next to each other on the same transcript molecule. While it is technically possible to implement bDNA FISH with only one epitope-specific probe [10], requiring the simultaneous binding of two probes in direct spatial adjacency should reduce unspecific signal [11]. Targeting 15 different epitopes of each transcript in a single hybridization reaction ensures that at least one epitope is accessible to the detection reagents without the need to denature the specimen. The subsequent rounds of *in situ* hybridization create a docking platform for ~500 fluorescently labelled probes per single epitope. This level of fluorescence is sufficiently high to enable the specific, rapid and robust detection of single transcript molecules by high-throughput imaging.

1.2. Alternative methods for RNA detection in imaging

Another method for directly visualizing single transcript molecules *in situ* is oligonucleotide-based single molecule FISH (o-nuc sm-FISH). This approach targets individual transcripts by up to 40 different oligonucleotides, which are directly conjugated to fluorophores. While a recent study achieved to monitor 61 different ncRNAs, it had to restrict itself to "a few dozen cells ... due to limited imaging throughput" [12]. Possibly, this reflects the lower signal-to-noise ratio of single fluorescent spots of o-nuc sm-FISH and their need for a 600 times longer illumination time [4].

Alternatively, transcripts can be visualized indirectly via reverse transcription to cDNA that can be sequenced *in situ* by padlock probes [13] or oligonucleotide ligation and detection [14,15]. While the former sequencing approach can presently detect 31 different genes simultaneously in thousands of single cells within a tissue slide [13], the latter approach can currently read around 200 mRNAs simultaneously for 40 different cells [15]. The efficiency for detecting single transcript molecules has been estimated to be 30% [13,16] and 3% [15] respectively, which is much lower than the 85% of hybridization efficiency in sm-FISH [4,17]. Such low efficiencies currently prevent these alternative methods from surveying the transcriptome with single-molecule sensitivity and resolution *in situ* [18,19].

2. Establishing image-based transcriptomics with single molecule resolution

The detailed experimental protocol for high-throughput bDNA sm-FISH has been published previously [4] and therefore, we here mainly provide additional assistance for setting up a robust automated experimental platform. As a general introduction to high-throughput image-based assays and the infrastructure and software supporting such experiments we highly recommend the excellent essay by Buchser and colleagues [20].

Table 1 contains an overview of potential problems occurring during the detection of single transcripts. The most critical factor in getting reliable results is to use an automated incubator that contains rotating towers for the individual storing of multi-well plates during hybridization reactions. This prevents the occurrence of different hybridization efficiencies in different wells of a multi-well plate (data not shown). Table 2 highlights potential pitfalls, which could affect the biological interpretation of accurate single-molecule measurements. We recommend repeating the control experiments suggested in Table 1 and Table 2 in different weeks to ensure that your setup of image-based transcriptomics functions robustly.

3. Establishing the image analysis pipeline

A robust image analysis pipeline is required for accurate measurements of absolute transcript levels as well as measurements of transcript localization in the cytoplasm of single cells, and extraction of features that describe the cellular phenotype. First, homogeneous intensity values throughout the images in all channels must be ensured, and then object segmentation must be performed minimizing errors. To ensure this, we developed four algorithms to perform high-throughput illumination correction of raw images, robust nuclei and cell segmentation, and robust spot detection. They can be downloaded from https://github.com/pelkmanslab/ImageBasedTranscriptomics and applied on an example dataset available on https://image-based-transcriptomics.org. The algorithms presented in this manuscript do not intend to replace single-cell quality control. For the latter we recommend interactive user-guided supervised machine learning, which has been implemented before by our group [23] and others [24]. Supervised machine learning not only readily identifies rare cells that have not been correctly segmented, but also allows the selection of a group of cellular objects that is relevant for a specific biological question (e.g., interphase cells).

The algorithms presented in this manuscript intend to reduce human hands on time and increase the amount of high-quality primary data after computational image-analysis (Table 3). Computational running time has not emerged as a practical issue for image-based transcriptomics. The algorithms are robust in the sense that their input parameters rarely have to be adjusted for individual experimental plates.

While the principles of the algorithms presented in this manuscript have been sketched in one of our earlier publications, the description beneath provide a detailed guide for using those algorithms. Moreover we here include implementations of these