Dissecting Transcriptional Heterogeneity in Pluripotency: Single Cell Analysis of Mouse Embryonic Stem Cells

Ana M.V. Guedes, Domingos Henrique, and Elsa Abranches

Abstract

Mouse Embryonic Stem cells (mESCs) show heterogeneous and dynamic expression of important pluripotency regulatory factors. Single-cell analysis has revealed the existence of cell-to-cell variability in the expression of individual genes in mESCs. Understanding how these heterogeneities are regulated and what their functional consequences are is crucial to obtain a more comprehensive view of the pluripotent state.

In this chapter we describe how to analyze transcriptional heterogeneity by monitoring gene expression of *Nanog*, *Oct4*, and *Sox2*, using single-molecule RNA FISH in single mESCs grown in different cell culture medium. We describe in detail all the steps involved in the protocol, from RNA detection to image acquisition and processing, as well as exploratory data analysis.

Keywords:: Stem cells, Pluripotency, Heterogeneity, Transcription, Single-molecule FISH, Stochastic gene expression

1 Introduction

Pluripotency is defined as the capacity of cells to differentiate into any cell type from each of the primary embryonic germ layers, while being able to maintain a pool of undifferentiated cells. The maintenance of pluripotent mouse Embryonic Stem cells (mESCs) in culture requires the coordinated activity of a complex Pluripotency Transcription Network in which the transcription factors Nanog, Oct4, and Sox2 play a central role [1–4].

Traditionally, the pluripotent state was viewed as being homogeneous and characterized by a specific genetic signature. This view has been replaced by the notion that the pluripotent state is highly heterogeneous and dynamic, comprising variable expression of both pluripotency genes and lineage-affiliated genes [5–8]. The discovery that expression of several pluripotency regulators fluctuates over time and the possibility that pluripotent stem cells might exist in multiple interconvertible states (associated with variable capacities of self-renewal and differentiation) [5, 9–11] highlight the dynamic nature of the pluripotent transcriptional program, which might be a fundamental feature of pluripotency. Moreover, it has been shown that different levels of expression of some

pluripotency genes, such as *Rex1*, *Dppa3*, *Prdm14*, or *Nanog*, has a functional impact in pluripotency and results in different propensities of cells to differentiate [12–14].

The variability observed in mESC cultures has been proposed to reflect two different features: (1) stochastic fluctuations or "noise" in gene expression, resulting from the inherent randomness of the biochemical processes associated with transcription and translation; and (2) the presence of multiple functional states that coexist and show different gene expression patterns and variable correlations between a set of genes [15]. However, the biological significance of this variability is yet to be understood.

In order to dissect mESC heterogeneity, single cell approaches are vital [16, 17], since analysis of individual cells usually leads to more accurate representations of variations between cells when compared with the averages typically obtained by bulk measurements. One of the most adequate methods to perform single cell analysis is single-molecule RNA FISH (smFISH) [18–21]. This technique is based on in situ hybridization followed by microscopy analysis, and has the advantage of providing accurate integer counts of mRNA molecule numbers in individual cells [17, 20]. smFISH uses a set of short fluorophore-labeled oligonucleotides that hybridize with the target mRNA, resulting in the presence of large number of fluorophores bound to an individual mRNA, providing enough fluorescent signal to be detected as a diffraction-limited spot [19, 22].

Several studies on single cell analysis of transcriptional heterogeneity in mESCs have been published recently. While some of them attempt to characterize gene expression profiles of several pluripotency markers, in cells subject to different environmental signals and perturbations [15, 23, 24], others focus on the analysis of specific genes, such as *Nanog* [15, 25, 26] and how their heterogeneity is modulated.

In this chapter, we describe how to perform single cell transcriptional analysis in mESCs by smFISH. We describe the workflow used to perform multiplex smFISH in mESCs, from cell culture and fixation (Section 3.1), to probe design and preparation (Sections 3.2 and 3.3), and hybridization with RNA in cultured cells (Section 3.4). In addition, we describe the procedures for image acquisition (Section 3.5) and processing (Section 3.6), as well as exploratory data analysis (Section 3.7). As an example, we characterize the expression of the pluripotency regulators *Nanog*, *Oct4*, and *Sox2* with the objective of getting a better understanding of their heterogeneity, dynamics and interactions. We show a comparative analysis in cells grown in ground state conditions (2i/LIF), in conventional Serum/LIF, and upon early events of non-directed differentiation induced by LIF withdrawal (Serum). This simplistic

analysis offers further evidence for the stochastic nature of gene expression in the pluripotent state, and indicates that mRNA transcription in mESCs is a noisy process, probably as a result of the uniquely permissive chromatin environment found in mESCs [27].

2 Materials

2.1 Equipment

1. Cell culture hood.
2. 37 °C CO_2 incubator.
3. Standard wide-field fluorescence microscope (in this case a Zeiss Axiovert) equipped with:
   (a) Mercury or metal-halide lamp;
   (b) Filter sets appropriate for the fluorophores selected (Table 1);
   (c) Cooled CCD or EMCCD camera;
   (d) High numerical aperture (>1.3) 100× objective;
   (e) Motorized XY stage to acquire multiple positions;
   (f) Motorized *Z*-stack acquisition;
   (g) Controlling computer and software able to do multi-position acquisition.
4. External storage for data files.
5. Computer for data processing analysis.
6. MATLAB software with image processing toolbox.
7. RStudio software for data analysis [28].

2.2 Reagents and Materials

Prepare all solutions under sterile conditions, in a laminar flow hood class II.

2.2.1 mESC Culture

All mESC culture procedures were performed according to what has been described by Pezzarossa et al. [29]. Here we provide a brief description of reagents and materials needed for mESC culture and advise consulting [29] for more details.

Table 1
Optical filters required for smFISH protocol

	Exitation	Beam splitter	Emission	Supplier
Alexa594	590DF10	610DRLP	630/DF30	Omega
Cy5	640/30	660	700/75	Chroma
TMR	546/10	560	580/30	Chroma
Dapi	365/12	395	>397	Zeiss

Details for excitation, beam splitter, and emission filters for all the channels needed in the protocol. Suppliers are also provided

1. Cell line: E14tg2a, a non-modified cell line derived from 129/Ola mice blastocysts (a kind gift from Austin Smith's lab, University of Cambridge, UK).

2. Cell culture medium: For the experiments described in this chapter we have used Serum/LIF, 2i/LIF, and Serum medium.

   (a) *Serum/LIF medium*: mix the following components to make 250 ml of medium; filter-sterilize and supplement with 2 ng/ml of Leukemia inhibitory factor (LIF) prior to use.

      • 200 ml of 1× Glasgow Modified Eagle's medium—GMEM (GIBCO);
      • 2 ml of 200 mM glutamine (100×, GIBCO);
      • 2 ml of 100 mM Na pyruvate (100×, GIBCO);
      • 2 ml of 100× nonessential amino acids (GIBCO);
      • 2 ml of 100× penicillin–streptomycin solution (GIBCO);
      • 200 µL of 0.1 M 2-mercaptoethanol (Sigma);
      • 20 ml of heat inactivated ES screened fetal bovine serum (Hyclone, cat.no. SH30070.03ES).

   (b) *2i/LIF medium*: commercial iStem medium [30] (Cellartis). Supplement with 2 ng/ml LIF prior to use (home-made iStem may also be used).

   (c) *Serum medium*: same composition as Serum/LIF medium but without supplementing with LIF prior to use.

3. Other reagents/materials:

   • 0.1 % gelatine diluted from 2 % (Sigma) in tissue culture grade H_2O.
   • 0.1× trypsin solution: mix 5 ml of 2.5 % trypsin (GIBCO), 0.5 ml of heat-inactivated chicken serum, 0.1 ml of 0.5 M EDTA, and PBS to 50 ml to prepare a 1× trypsin solution. Dilute the 1× trypsin solution with PBS to prepare 0.1× working solution.
   • 6-well multi-well tissue culture dish (Nunc, cat.no. 140675).
   • 60-mm tissue culture dishes (Nunc, cat.no. 150288).

2.2.2 Cell Fixation

Prepare solutions with ribonuclease free ultrapure water in a fume hood. Perform sample handling in a fume hood.

1. 37 % formaldehyde (Sigma, cat. no. 252549).

2. 1× PBS.

3. 70 % Ethanol.

2.2.3 Hybridization
and Washing

Prepare solutions with ribonuclease free ultrapure water in a fume hood. Perform sample handling in a fume hood.

1. Hybridization buffer: 1 g of dextran sulfate (Sigma, cat.no. D8906), 1 ml of formamide (Ambion, cat.no. AM-9342), 1 ml of 20× SSC and ultrapure ribonuclease free H_2O up to 10 ml. Store at −20 °C in 500 μl aliquots.

2. Probe sets specific for the genes to be detected, labeled with appropriate fluorophores (Stellaris RNA FISH Biosearch Technologies) (**Note 1**).

3. TE buffer: 10 mM Tris, 1 mM EDTA pH = 8.

4. Washing buffer: 5 ml of 20× SSC, 5 ml of formamide, 500 μl of 10 % Triton and ultrapure ribonuclease free H_2O up to 50 ml. Store at room temperature.

5. DAPI (Sigma).

2.2.4 Cell Mounting

1. Equilibration buffer: 850 μl of H_2O, 100 μl of 20× SSC, 40 μl of 10 % glucose, 10 μl of Tris 1 M pH 8, 10 μl of 10 % Triton. Prepare fresh.

2. Glucose oxidase (Sigma cat.no G2133): prepare aliquots of stock solution (37 mg/ml). Dilute stock to 3.7 mg/ml in 50 mM of Sodium Acetate. Store at −20 °C and use shortly after thawing.

3. Catalase (Sigma cat.no C-3515). Store at 4 °C. Vortex thoroughly before using.

4. Anti-fade buffer: 100 μl of Equilibration buffer, 1 μl of 3.7 mg/ml glucose oxidase, 1 μl of catalase. Always prepare fresh.

5. Glass slides 76 × 26 × 1 mm.

6. #1 glass coverslips 18 × 18mm or 50 × 24 mm.

7. Dow corning high vacuum silicone grease (Sigma cat.no Z273554).

8. Precision tweezers.

9. Stereoscope with bottom lighting.

3 Methods

3.1 Cell Culture

Perform all cell manipulations under sterile conditions in a laminar-flow cell culture hood (class II). All mESC culture procedures were performed according to what has been described by Pezzarossa et al. [29]. Here we provide a brief description and advise consulting [29] for a more detailed description.

1. Perform routine cell passaging and expansion in Serum/LIF medium, on gelatin-coated dishes, and incubate cells at 37 °C in a 5 % (v/v) CO_2 incubator.

2. Passage cells every other day at a cell density of 3×10^4 cells/cm^2.

3. Perform cell dissociation using $0.1\times$ trypsin and incubating for 2 min at 37 °C.

4. Plate mESCs in the conditions to be tested (Serum/LIF, 2i/LIF or Serum—as described in Section 2.2.1). Plating of one to three 60-mm Nunc dishes usually provides sufficient cells to perform several smFISH experiments (**Note 2**).

3.1.1 Cell Fixation

1. Two days after cell plating, harvest the cells cultured in Serum/LIF, 2i/LIF, and Serum medium (or other tested conditions).

2. Dissociate cells by trypsinization and ensure a good single resuspension.

3. Wash cells with $1\times$ PBS and spin cells down (4 min, $165 \times g$-force).

4. Resuspend thoroughly in 4.5 ml of $1\times$ PBS and add 500 µl of 37 % Formaldehyde, ensuring proper homogenization by pipetting up and down.

5. Incubate cells for 10 min at RT.

6. Spin cells down (4 min, $165 \times g$-force), remove supernatant and wash cells twice with $1\times$ PBS.

7. Resuspend in a desired volume of 70 % ethanol for permeabilization during at least 2 h and store at 4 °C (**Note 3**).

3.2 Probe Design (Custom Stellaris® FISH Probes)

The probe set sequences used in the described experiments (for Nanog, Oct4, and Sox2) are shown in Table 2.

1. Select the sense strand of the target sequence for the design of complementary oligonucleotide probes (**Note 4**).

2. Proceed to probe designing using Stellaris® FISH Probe Designer (Biosearch Technologies Inc., Petaluma, CA) available at www.biosearchtech.com/stellarisdesigner.

3. Select the fluorophores to be coupled with the probe set—Alexa594, Cy5 or TMR fluorophores (or the equivalents CAL Fluor® Red 610, Quasar® 670 or Quasar® 570, respectively) (Biosearch Technologies, Inc.) (**Note 5**).

4. Follow the website instructions for ordering the probe sets.

3.3 Preparation of Probe Stocks

1. Spin down the lyophilized pooled probes sets, blend and dissolve in the appropriate volume of TE buffer to create a probe stock of 100 µM.

Table 2
Oligo sequences for each probe set used (Nanog, Oct4, and Sox2)

Gene	Nanog	Oct4	Sox2
Probe set	aaatcagcctatctgaaggc	tgagaaggcgaagtctgaag	ccgtctccatcatgttatac
Oligos	cagaaagagcaagacaccaa	aggttcgaggatccacccag	tccgggctgttcttctggtt
	gaagtcagaaggaagtgagc	tggaggcccttggaagctta	ataccatgaaggcgttcatg
	actcagtgtctagaaggaaa	tgagcctggtccgattccag	ttctcctgggccatcttacg
	ggtttttaggcaacaaccaaa	acatggggagatccccaata	atctccgagttgtgcatctt
	cgagggaagggatttctgaa	tccctccgcagaactcgtat	tcggacaaaagtttccactc
	cacactcatgtcagtgtgat	aacctgaggtccacagtatg	ttataatccgggtgctcctt
	cagaactaggcaaactgtgg	aacttgggggactaggccca	tcatgagcgtcttggttttc
	ttcccagaattcgatgcttc	tcaggctgcaaagtctccac	ggaagcgtgtacttatcctt
	aaaaactgcaggcattgatg	tgctttccactcgtgctcct	tagctgtccatgcgctggtt
	agcaagaatagttctcggga	tcagaggaggttccctctga	ttgctccagccgttcatgtg
	cagagcatctcagtagcaga	ttctccaacttcacggcatt	tcctgcatcatgctgtagct
	gaagaggcaggtcttcagag	tttcatgtcctgggactcct	tgcatcggttgcatctgtgc
	tgggactggtagaagaatca	aactgttctagctccttctg	tcatggagttgtactgcagg
	tcaggacttgagagcttttg	tcttctgcttcagcagcttg	ttcatgtaggtctgcgagct
	cttgttctcctcctcctcag	tgggtgtaccccaaggtgat	agtaggacatgctgtaggtg
	gagaacacagtccgcatctt	aaagagaacgcccagggtga	ttgaccacagagcccatgga
	ctgtccttgagtgcacacag	tggtctggctgaacacctt	tgggaggaagaggtaaccac
	tgaggtacttctgcttctga	aaggcctcgaagcgacagat	aggtacatgctgatcatgtc
	gagagttcttgcatctgctg	catgttcttaaggctgagct	tgggccatgtgcagtctact
	atagctcaggttcagaatgg	ttcctccacccacttctcca	agtgtgccgttaatggccgt
	gaaaccaggtcttaacctgc	gaaggttctcattgttgtcg	aaaatctctccccttctcca
	ttgcacttcatcctttggtt	gtctccgatttgcatatctc	cccaattcccttgtatctct
	tcaaccactggttttctgc	tagttcgctttctcttccgg	tactctcctctttttgcacc
	ttctgaatcagaccattgct	cacctcacacggttctcaat	ctgcggagatttttttcct
	gatactccactggtgctgag	tcagaaacatggtctccaga	tttttccgcagctgtcgttt
	ggatagctgcaatggatgct	atctgctgtagggagggctt	aatttggatgggattggtgg
	cagatgcgttcaccagatag	aagctgattggcgatgtgag	tagtcggcatcacggttttt
	aagttgggttggtccaagtc	gaaccacatccttctctagc	gaagtcccaagatctctcat
	gtctggttgttccaagttgg	cgccggttacagaaccatac	ctgtacaaaaatagtcccc
	aaagtcctccccgaagttat	acttgatcttttgcccttct	tatacatggtccgattcccc

(continued)

Table 2
(continued)

Gene	Nanog	Oct4	Sox2
	ctgcaactgtacgtaaggct	cttctcgttgggaatactca	gcgtagtttttttcctccag
	caaatcactggcagagaagt	ggtgtccctgtagcctcata	cctaacgtaccactagaact
	tagtggcttccaaattcacc	agaggaaaggatacagcccc	aagacttttgcgaactccct
	ctaaaatgcgcatggctttc	atagcctggggtgccaaagt	ccggagtctagctctaaata
	ataattccaaggcttgtggg	gtgtggtgaagtggggggctt	ctgtacaaaagttgcttgca
	tggagtcacagagtagttca	tcaggaaaagggactgagta	gattgccatgtttatctcga
	agatgttgcgtaagtctcat	aacagagggaaaggcctcgc	caagaacccttтcctcgaaa
	gctttgccctgactttaagc	atgggagagcccagagcagt	aagctgcagaatcaaaaccc
	tttggaagaaggaaggaacc	gctggtgcctcagtttgaat	ccttgtttgtaacggtccta
	caaatcactggcagagaagt	ttgccttggctcacagcatc	ccagtacttgctctcatgtt
	tagtggcttccaaattcacc	aaagctccaggttctcttgt	aacaagaccacgaaaacggt
	ctaaaatgcgcatggctttc	ccctcctcagtaaaagaatt	acaatctagaacgtttgcct
	ataattccaaggcttgtggg	ccacccctgttgtgcttтta	gatatcaacctgcatggaca
	tggagtcacagagtagttca	agcttctttccccatcccac	gggtaggattgaacaaaagc
	agatgttgcgtaagtctcat	ctcctgatcaacagcatcac	cggaaaataaaagggggggaa
	gctttgccctgactttaagc	aatgatgagtgacagacagg	ccaataacagagccgaatct
	tttggaagaaggaaggaacc	gtgtgtcccagtctttattt	tatacatggattctcggcag

2. Rehydrate oligonucleotides for 30 min at room temperature, protect from light and vortex to ensure proper homogenization.

3. Store probe stocks at −20 °C, protected from light (**Note 6**).

4. Prepare working dilution, by adding 2 μl of probe stock to 38 μl of TE buffer for a final concentration of 5 μM (1:20 dilution) and mixing thoroughly (**Note 7**).

5. Store working stocks at −20 °C and protect from light when handling.

3.4 smFISH in Cells in Suspension

A schematic representation of the steps involved here is depicted in Fig. 1a.

1. Spin down 200 μl of fixed cells in ethanol (2 min, 165 × *g*-force) (**Note 8**).

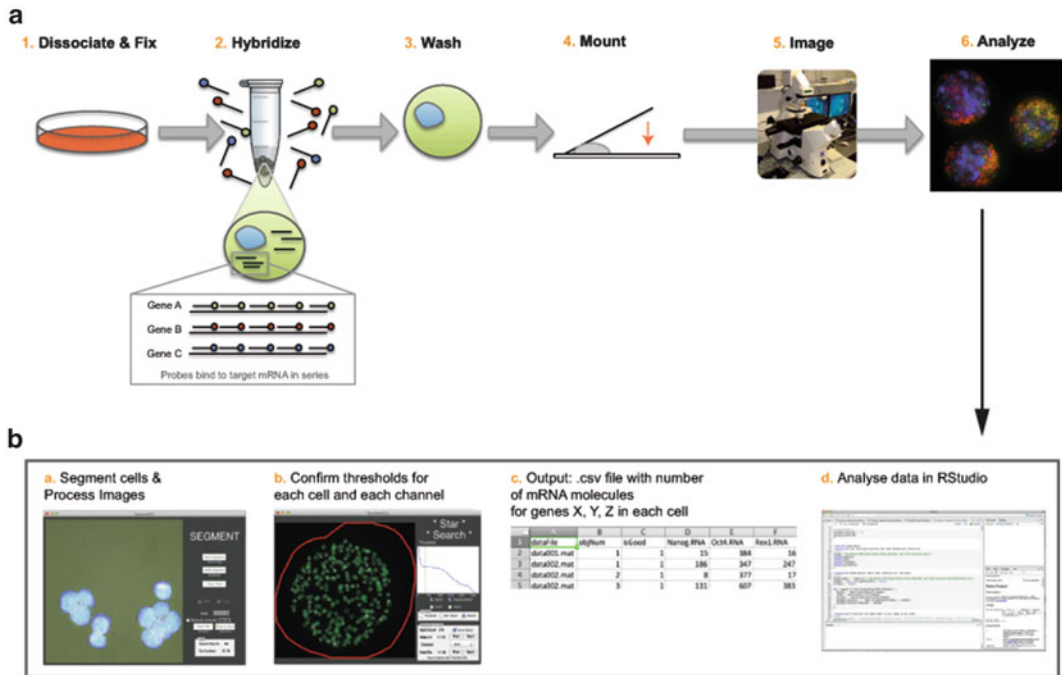2. Remove supernatant and wash in 850 μl of Wash buffer (**Note 9**).

Fig. 1 Representation of the smFISH protocol workflow. (**a**) Key steps involved in the protocol of smFISH for mESCs. (**b**) Key steps involved in image processing and data analysis of data generated by smFISH

3. Prepare the hybridization mix by adding 1 μl of the working dilutions of each of the desired probes to 100 μl of hybridization buffer (approximate final probe concentration of 50 nM).

4. Spin cells down (2 min, 165 × g-force), remove Wash buffer, and resuspend the pellet in 100 μl hybridization mix, mixing thoroughly and vortexing (**Note 10**).

5. Protect from light, seal the tube to prevent evaporation, and incubate overnight in a 37 °C oven.

6. On the following morning, add 850 μl of Wash buffer and spin down (2 min, 165 × g-force).

7. Remove supernatant, add 850 μl of Wash buffer and incubate for 30 min at room temperature.

8. Spin cells down (2 min, 165 × g-force).

9. Remove supernatant, add 850 μl of Wash buffer supplemented with 1 μl of 1 mg/ml DAPI solution, and incubate for 30 min at room temperature.

10. Prepare fresh Equilibration buffer.

11. Spin cells down (2 min, 165 × g-force), remove supernatant and wash with 850 μl of Equilibration buffer.

12. Spin cells down (2 min, 165 × g-force) and resuspend pellet in 10 μl of Anti-fade buffer (**Notes 11** and **12**).

13. Proceed to mounting (**Note 13**):

    (a) Place 5 µl of cell suspension on a clean glass microscope slide (**Note 14**).

    (b) Check cell density under the stereoscope and remove any cell clumps that will prevent proper smashing (**Note 15**). Add a few microliters of Anti-fade buffer whenever cell density is too high.

    (c) Use a #1 cover glass with the help of precision tweezers to spread cell solution and apply pressure gently over the surface of the cover glass (**Note 16**).

    (d) Wipe excess liquid with tissue paper while pressing gently with the tweezers.

    (e) While observing through the stereoscope lens, press the coverslip until cells are properly smashed (**Note 17**).

    (f) Seal cover glass using silicone grease to prevent liquid evaporation (**Note 18**).

    (g) Proceed to imaging.

3.5 Image Acquisition

1. At least 30 min before starting the acquisition turn ON the microscope setup.

2. Place a drop of immersion oil on the objective.

3. Position the slide in the microscope and allow the temperature to stabilize for 30 min (**Note 19**).

4. Choose an area with good cell density and focus using the DAPI channel (**Note 20**).

5. Test signal intensity for each channel and define the exposure time for each of the fluorophores (it ranges from 1 to 5 s) and for DAPI (usually few milliseconds) (**Notes 21** and **22**).

6. Select and record the positions to be imaged. Avoid marking the same position twice by starting from left and moving to right down (**Note 23**).

7. For each position, set the lower plane of the Z-stack, allow the software to define the distance between planes (usually 0.3 µM is enough) and the number of optical sections needed to span the vertical extent of the cell (usually 20–30 sections).

8. Use the acquisition software to prepare a protocol (*macros*) that allows automate running through the list of recorded positions and, in each position, goes to the lower plane, takes the defined optical sections with the defined exposure time for each channel and saves the Z-stack as a .tiff image format.

9. Begin imaging acquisition.

10. At the end of the acquisition, before beginning data analysis, review the images to check for the quality of the acquisition (**Note 24**).

3.6 Image Processing

Here we describe the procedures for Image Processing using custom software written in Matlab. This software has been developed in Arjun Raj's Laboratory [19, 22, 31] and is publicly available at https://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home, with a worked out example. A schematic representation of the workflow of Image processing is also available in Fig. 1b.

1. Segment cells using DAPI channel and a RNA fluorophore channel in order to be able to properly define cytoplasm and cell boundaries.

2. Process the images.

3. Review the automatic threshold performed by the software for each cell and each acquired RNA channel. Adjust if needed (**Note 25**).

4. Exclude cells that don't have proper RNA signal or that have autofluorescent dirt.

5. Save the output data as a .csv file, including the number of mRNA molecules in each channel for each cell.

6. Back up original raw data images and create a database with .csv files (**Note 26**).

7. Import .csv files for RStudio and proceed to statistical analysis and graphical representation of the observed data.

3.7 Data Analysis

A series of different statistical analysis and graphical representations can be performed with the acquired data using different software. Here we report a list of the most common statistics using R, a programming language for statistical computing and graphics. We have taken advantage of RStudio [32], an integrated development environment for R [28], and ggplot2, a data visualization and plotting package developed for R [33].

In this section, we use data obtained when performing triple labeling with Nanog, Oct4, and Sox2 probes in mESCs grown in Serum/LIF, 2i/LIF, or Serum conditions.

3.7.1 Distribution Analysis

A histogram graphical representation of the obtained data for each gene is crucial as the first approach during the exploratory analysis. It allows not only to understand the shape of data distribution but also to get an estimate of the probability distribution of mRNA molecules per cell (Fig. 2). Alternatively, boxplots or violin plots can be used.

Analysis of the histograms obtained in the described experiments shows that the distribution of *Nanog*, *Oct4*, and *Sox2* mRNA
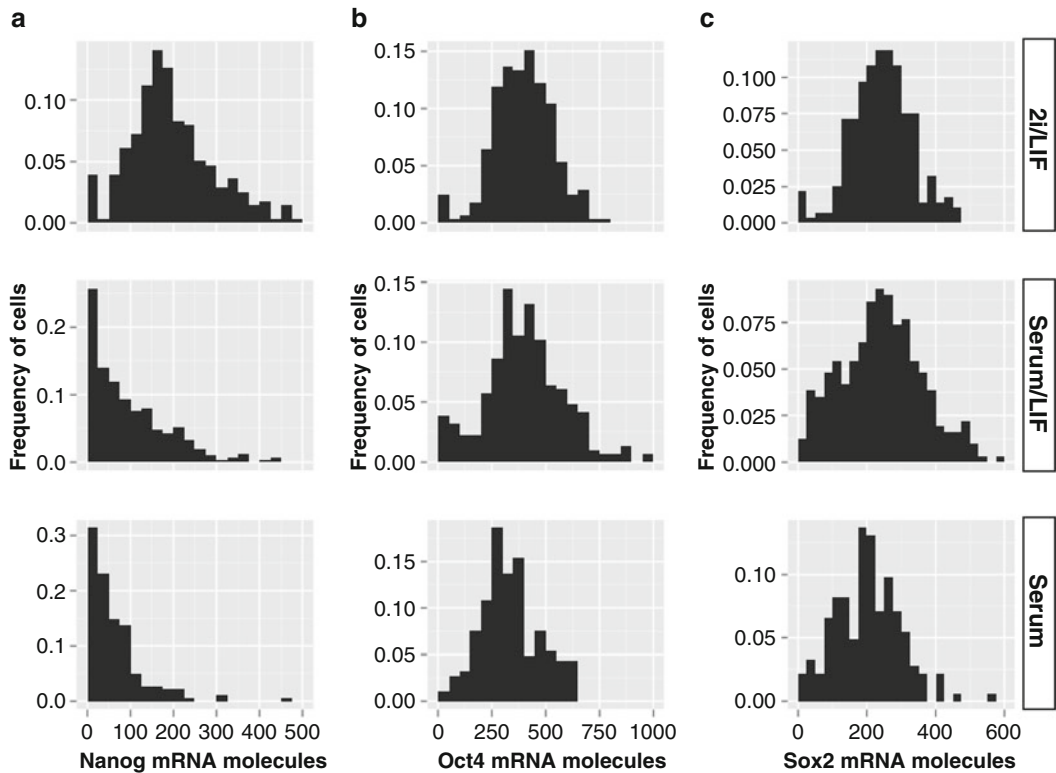
Fig. 2 Histogram representation of mRNA distributions from Nanog, Oct4, and Sox2 in single cells grown in different culture conditions (Serum/LIF, 2i/LIF, and Serum). (**a**) Histogram representation showing the distribution of Nanog mRNA molecules per cell for cells grown in 2i/LIF, Serum/LIF, and Serum. Bin-width = 25. (**c**) Same as (**a**) for Oct4 mRNA. Bin-width = 50. (**c**) Same as (**a**) for Sox2 mRNA. Bin-width = 25

molecules per cell is extremely variable, ranging from 0 molecules per cell to 500, 1000 and 600 mRNA molecules, respectively (Fig. 2). It can also be observed that the shape of the distributions is variable between genes and conditions. The shape of the distributions is a good indication of how the gene's transcription is regulated [17]. Poissonian distributions are suggestive of a one-state model, in which transcription is always ON. Non-Poissonian or long tailed distributions indicate a two-state model of transcription, in which the promoter alternates between ON and OFF states in a pulsatile manner [18, 19]. Regarding *Nanog*, the distribution of mRNAs/cell in 2i/LIF conditions shows a reasonable approximation to a Poisson, with few cells expressing very low or very high numbers of transcripts, whereas in Serum/LIF the distribution shows a long exponential tail, with a substantial number of cells having very few mRNA molecules. In Serum conditions, a reduction in the number of cells expressing high and intermediate levels of *Nanog* and an increase in the low-*Nanog* expressing cells is

observed, comparatively to the trend observed in Serum/LIF conditions. In the case of *Oct4* and *Sox2*, the shape of the mRNAs/cell distribution does not vary greatly within the different culture conditions. In all cases, the distributions show a good approximation to a bell-shaped distribution, with few cells expressing very low or high levels of *Oct4* or *Sox2*.

Overall, the high heterogeneity in *Nanog*, *Oct4*, and *Sox2* mRNA expression between mESCs points out to a bursty transcription, characterized by long periods of gene inactivity and sporadic pulses of transcription.

3.7.2 Descriptive Analysis

Several statistical measurements can be extracted from the distributions of mRNA molecules per cell that describe the distribution, and can give information about the transcription of the genes under study. In Table 3, we have summarized mean, standard deviation, median, minimum and maximum values, as well as Fano factor (FF), coefficient of variation (CV), and the total number of cells analyzed. Measurements such as FF (defined as the ratio between the variance and the mean) and CV (defined as the ratio between the standard deviation and the mean) are a good measure of heterogeneity and noise in gene expression [18, 19].

In the case of *Nanog*, it can be observed that when cells are passed from Serum/LIF to 2i/LIF there is an increase in the average number of transcripts/cell from 96 ± 89 to 195 ± 99. This increase is not due to a change in the distribution range of *Nanog* expression, reflecting instead an alteration in the balance between cells expressing high or low levels of *Nanog* mRNA (Fig. 2a). Upon LIF withdrawal (Serum conditions), the average value in transcripts per cell drops, as expected, from 96 ± 89 to

Table 3
Statistical measurements extracted from mRNA distributions

Gene	Medium	Mean $\pm$ sd	Median	Min–max	FF	CV	*N*
Nanog	2i/LIF	195 ± 99	178.5	4–504	50.2	0.5	278
	Serum/LIF	96 ± 89	70.5	0–443	81.7	0.9	312
	Serum	65 ± 73	44	0–511	81.9	1.1	183
Oct4	2i/LIF	395 ± 136	396	0–762	47.2	0.3	278
	Serum/LIF	398 ± 181	394	5–971	82.3	0.5	312
	Serum	335 ± 135	316	1–627	54.6	0.4	183
Sox2	2i/LIF	242 ± 87	238.5	0–466	31.4	0.4	278
	Serum/LIF	242 ± 117	240.5	0–599	57	0.5	312
	Serum	202 ± 94	203	7–561	44.1	0.5	183

Statistical measurements of Nanog, Oct4, and Sox2 mRNA distributions from cells grown in different cell culture medium (same data used to plot the histograms depicted on Fig. 2). Values are represented in the form of number of transcripts. FF stands for Fano factor, CV for coefficient of variation, and *N* for number of cells analyzed

65 ± 73, since *Nanog* expression is known to be downregulated upon differentiation. In the case of *Oct4* and *Sox2*, there is less variation between culture conditions. This is reflected in the CV values for *Oct4* and *Sox2*, which are close to 0.5 in all the analyzed conditions. On the contrary, the CV in *Nanog* expression is 0.5 only in 2i/LIF and increases to 0.9 in Serum/LIF and 1.1 in Serum, reflecting the increase in variability between cells in these two conditions.

Additionally, we can have an estimation of noise strength using the FF measure. High values indicate deviation from that predicted for a normal Poissonian distribution, in which FF is equal to 1. It can be observed that *Nanog*, *Oct4*, and *Sox2* have FF values much higher than 1 in all tested conditions, thus suggesting that transcription of these genes is noisy in pluripotent cells, occurring in transcriptional bursts.

3.7.3 Correlational Analysis

Correlations between mRNA molecule numbers in each cell can be calculated whenever double or triple stainings are performed. The best graphical representation of correlations between two genes is the scatter plot (Fig. 3). To obtain a statistical measure of the correlation, the Spearman correlation coefficient can be used, since is does not assume a normal distribution of the data. The correlation analysis between the three analyzed genes shows that *Sox2* and *Nanog* have a moderate to strong correlation in all culture conditions ($0.4 \leq r \leq 0.79$) (Fig. 3a), whereas *Oct4* and *Sox2* show moderate to strong correlation in 2i/LIF and Serum/LIF ($0.4 \leq r \leq 0.79$) and weak correlation in Serum ($0.2 \leq r \leq 0.39$) (Fig. 3b). The correlation between *Oct4* and *Nanog* expression is also variable between culture conditions (Fig. 3c). While in 2i/LIF most cells express high levels of both genes and the correlation is strong ($r = 0.57$), in Serum/LIF this correlation is weaker ($r = 0.4$) mostly due to the presence of two subpopulations of cells (one expressing high levels of both genes and another with low levels of *Nanog* and varying levels of *Oct4*), further decreasing in Serum conditions ($r = 0.27$).

To explore the correlation between the three genes, scatter plots can be used, coloring cells according to the expression levels of the third gene. This can be achieved by classifying cells as high- or low-expressing by the definition of a cut-off (Fig. 3c). The definition of a cutoff results from graphical interpretation and depends on the ability to distinguish two subpopulations of cells. The use of this graphical representation depicted in Fig. 3c shows that most low-*Sox2* cells in 2i/LIF also express low levels of *Oct4* and *Nanog*, while in Serum/LIF and Serum cells with low *Sox2* levels express low levels of *Nanog* but varying levels of *Oct4*.
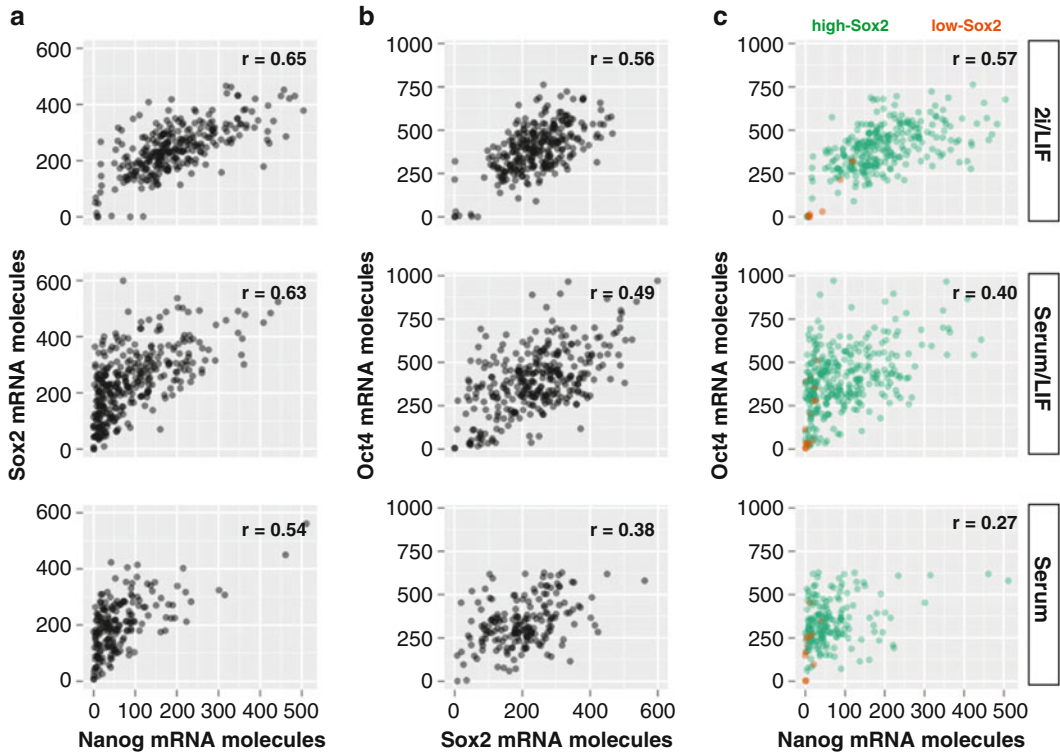
Fig. 3 Correlations between Nanog, Oct4, and Sox2 mRNA molecules in single cells. (**a**) smFISH analysis of Nanog versus Sox2 cells grown in 2i/LIF, Serum/LIF, and Serum. Each dot represents one single cell and the transparency of the dots was used to avoid over-plotting. For each graph, the Spearman correlation coefficient (*r*) is depicted. (**b**) Same as (**a**) for Sox2 versus Oct4. (**c**) Same as (**a**) for Nanog versus Oct4. Sox2 Level is color-coded as high (>50 transcripts—*green*) or low (<50 transcripts—*orange*)

4 Notes

1. Probe sets can be ordered directly labeled with the desired fluorophore or labeled in house and purified using high-pressure liquid chromatograph (HPLC) [34].

2. The amount of cells to fix is dependent on the experimenter objectives but should take into consideration that a considerable fraction of cells is lost during the protocol due to the high number of centrifugation steps.

3. Cells can be stored in 70 % ethanol, at 4 °C, for long periods of time without RNA degradation. The volume used to resuspend cells is dependent on the number of cells fixed and should ensure high concentration of cells (1 ml per 3×10^6 cells).

4. Go to the UCSC genome browser and select the species and gene of interest. Click on the RefSeq version of the gene and, if no preference exists for a specific isoform, select the one covering the largest shared common sequence between isoforms.

Finally, select the Genomic Sequence option and search for the coding sequence (CDS). If this region is not sufficient to design enough probes, the 3′ and 5′ UTR regions can be added.

5. Other fluorophores can be used and are available on the website. When using other fluorophores, compatibility has to be confirmed and proper filter sets have to be available.

6. Probe stocks should be exposed to a minimum of freeze/thaw cycles to prevent degradation.

7. Other dilutions can be tested if hybridization efficiency needs to be optimized (e.g., 1:10,1:20,1:50,1:100).

8. The volume of cells to use is variable. The volume is defined by the minimum needed to be able to have a detectable pellet and to account for cell loss during the various centrifugation steps of the protocol. Additionally, attention should be given to the type of microcentrifuge tubes used, because they can negatively influence the pellet formation and, consequently, the efficiency of the protocol.

9. In all centrifugations with Wash buffer, the pellets are very delicate and easily detach from the tubes, leading to loss of cells. In order to prevent this, do not remove all the supernatant.

10. When preparing the hybridization mix for several samples, take into account its high viscosity and always prepare extra volume.

11. The volume to resuspend cells is variable and depends on the pellet size. High concentration of cells is preferred for mounting.

12. Anti-fade solution is only needed when working with photosensitive-labeled probes (e.g., Cy5). This solution acts as an oxygen scavenging system, preventing photobleaching [35]. When only Alexa 594 and TMR labeled probes are being used, resuspend cells directly into $2\times$ SSC.

13. Alternatively, cells can be kept at 4 °C until mounting, but imaging should always be performed within 24–48 h, due to rapid signal degradation.

14. Make sure that the coverslips and slides are very clean since any dust will prevent proper cell mounting.

15. Be aware of the time spent in the mounting procedure. Long periods of time in the stereoscope may lead to heating of the sample and consequent evaporation of the small volume of cell suspension.

16. This is the most critical step in the protocol for signal quality. For acquisition of good fluorescence signal, cell thickness needs to be reduced (ideally to 2–5 μm) and the formation of air bubbles should be avoided. Practicing several times is strongly

advised in order to prevent over or under smashing. Over smashing of cells leads to loss of cell integrity and precludes proper cell segmentation, whereas under smashing leads to bad quality signal. Note that, when cells are not smashed, pressing the coverslip with the tweezers leads to an evident increase in cell surface area; when this increase stops, smashing is optimal.

17. Liquid evaporation leads to loss of signal quality, especially when using Cy5 labeled probes. The air that enters leads to increase in oxygen content and photobleaching, preventing glucose oxidase and catalase efficient action.

18. Do dot seal the cover glass with nail polish, since it leads to autofluorescence, thereby preventing proper RNA detection.

19. This step allows temperature to stabilize between all microscope components and sample, preventing thermal drift, which may consequently result on fluctuations in the Z-axis.

20. Cell density is an important parameter to consider in this step: cell density should be high in order to acquire several cells per image field, but not too high since this hampers the subsequent steps of image processing.

21. RNA signals are very dim and can only be seen through the CCD camera, and not directly in the eyepieces.

22. In some cases, acquisition of a channel with transmitted light may be important, since it facilitates identification of the borders of the cells in the posterior segmentation steps.

23. Before starting the full experiment, run a test acquisition with few positions in order to confirm that the setup is working correctly. Note that each position can be done individually, in the case of absence of XY stage. However this increases considerably the labor time spent at the microscope.

24. In this step, images that were acquired out of focus should be eliminated. Also, check if there are image fields with bad signal due to drying or presence of autofluorescence dirt and eliminate then from further analysis.

25. Cells with a well-defined threshold present a clear plateau between the background signal and the RNA signal in the histogram showing the cell's fluorescence intensity distribution (a representative histogram is shown in Fig. 1b, section b). In the case of a non-clear threshold, find a cell which has a clear signal and in which the threshold shows a defined plateau, and fix that contrast and threshold value (by not automatically adjusting the x axis) to all cells.

26. The database should include information on the quality of data in each channel for each experiment as well as the probe acquired in each channel. Always check for differences between same probe sets labeled with different fluorophores.

Acknowledgments

References

1. Niwa H (2007) How is pluripotency determined and maintained? Development 134:635–646

2. Boyer LA et al (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122:947–956

3. Young RA (2011) Control of the embryonic stem cell state. Cell 144:940–954

4. Kuijk EW et al (2008) Differences in early lineage segregation between mammals. Dev Dyn 237:918–927

5. Chambers I et al (2007) Nanog safeguards pluripotency and mediates germline development. Nature 450:1230–1234

6. Abranches E, Bekman E, Henrique D (2013) Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of Nanog expression. PLoS One 8: e59928

7. Abranches E et al (2014) Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. Development 141:2770–2779

8. Torres-Padilla M-EE, Chambers I (2014) Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development 141:2173–2181

9. Kalmar T et al (2009) Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol 7:e1000149

10. MacArthur BD et al (2012) Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity. Nat Cell Biol 14:1139–1147

11. Hayashi K, Lopes SM, Tang F, Surani MA (2008) Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell 3:391–401

12. Singh AM, Hamazaki T, Hankowski KE, Terada N (2007) A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells 25:2534–2542

13. Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H (2008) Identification and characterization of subpopulations in undifferentiated ES cell culture. Development 135:909–918

14. Yamaji M et al (2013) PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell Stem Cell 12:368–382

15. Singer Z et al (2014) Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol Cell 55:319–331

16. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467:167–173

17. Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216–226

18. Raj A, van Oudenaarden A (2009) Single-molecule approaches to stochastic gene expression. Annu Rev Biophys 38:255–270

19. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5:877–879

20. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4:e309

21. Etzrodt M, Endele M, Schroeder T (2014) Quantitative single-cell approaches to stem cell research. Cell Stem Cell 15:546–558

22. Raj A, Tyagi S (2010) Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. Meth Enzymol 472:365–386

23. Kumar R et al (2014) Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature 516:56–61

24. Faddah DA et al (2013) Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs. Cell Stem Cell 13:23–29

25. Hansen CH, van Oudenaarden A (2013) Allele-specific detection of single mRNA molecules in situ. Nat Methods 10:869–871

26. Miyanari Y, Torres-Padilla M-EE (2012) Control of ground-state pluripotency by allelic regulation of Nanog. Nature 483:470–473

27. Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M (2011) Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol 12:36–47

28. Core Team RCTR (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

29. Pezzarossa A, Guedes A, Henrique D, Abranches E (2015) Imaging pluripotency time-lapse analysis of mouse embryonic stem cells. Embryonic Stem Cell Protocols, Volume 1341 of the series Methods in Molecular Biology pp 87–100

30. Ying Q-LL et al (2008) The ground state of embryonic stem cell self-renewal. Nature 453:519–523

31. Raj A, Tyagi S (2009) Imaging individual mrna molecules using multiple singly labeled probes, US Patent App. 13/062. http://www.google.com/patents/US20120129165

32. RStudio Team (2015) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, http://www.rstudio.com/

33. Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer, New York

34. Batish M, Raj A, Tyagi S (2011) Single molecule imaging of RNA in situ. Methods Mol Biol 714:3–13

35. Aitken C, Marshall A, Puglisi J (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. Biophys J 94:1826–1835