USING SINGLE-CELL RNA SEQUENCING TO STUDY HOST-PATHOGEN INTERACTIONS

Neus Mestre Farràs supervised by Jan-Willem Veening

Abstract

Populations of cells are heterogeneous. Despite single cells of a population present common gene expression regulation they also present small differences in the expression. Single cell methodologies are needed to study the impact of this heterogeneity in cellular processes. Single-cell RNA sequencing (scRNAseq) has already been used to study heterogeneity of eukaryotic cell populations at transcriptional level. The aim of this report is to review the current state of scRNAseq technology to apply it for the study of host-pathogen interactions. Studies that already used this technique to study immune cells response to lipopolysaccharides showed different activation thresholds between host cells under same stimuli and that heterogeneity in pathogenic cells influences heterogeneity in host cells response. Therefore, scRNAseq is a powerful tool to characterize expression heterogeneity during host-pathogen interactions.

Introduction

It is well known that cells in populations are heterogeneous. In stable environments, cells are maximally adapted to the environment conditions, and heterogeneity is reduced to fluctuations around the optimal phenotype (Altschuler & Wu, 2010; de Jong, Haccou, & Kuipers, 2011). On the contrary, if cells grow in changing environments, heterogeneity is favourable because it increases the change of population survival. An example of biological heterogeneity as survival strategy is persisters cells (Lewis, 2010). These cells are minority subpopulation that does not grow in presence of antibiotics, which confers the property to escape drug treatment. In order to maintain heterogeneity of isogenic populations in changing environmental conditions, there are transcriptional regulation mechanisms to generate distinct phenotypes. For instance, some highly expressed genes are expressed in stochastic bursts, which generates biological noise (Zong, So, Sepúlveda, Skinner, & Golding, 2010).

The reference method to study transcriptomics is RNA sequencing (RNAseq) of a bulk of cells. The results obtained with bulk RNAseq are the averaged expression of the cells in the populations under study. It is a powerful tool to capture the cell state of a population and serves to identify the common selective pressures. However, averaged results mask the heterogeneity in gene expression and usually it only represents the dominant subpopulation (Altschuler & Wu, 2010; Proserpio & Mahata, 2015). In order to study heterogeneity at transcriptional level, RNA transcripts have to be identified and quantified in single cells. Reference methods to study single-cell gene expression at RNA level are single-cell quantitative PCR (qPCR) or RNA fluorescence in situ hybridization (RNA FISH) (Raj, van den Bogaard, Rifkin, van Oudenaarden, & Tyagi, 2008; Taniguchi, Kajiyama, & Kambara, 2009). However, these methods have a very low throughput and are hypothesis driven, limiting their use to few cells and few genes depending on prior knowledge. Single-cell RNAseq (scRNAseq) has emerged as a technique to study the whole transcriptome of many individual cells, overcoming the limitations of qPCR and RNA FISH (Stegle, Teichmann, & Marioni, 2015).

During infection, pathogenic cells invade the host organism. Afterwards, host cells have to recognise the pathogen and initiate signalling cascades to activate the immune and inflammatory responses that leads to pathogen clearance. Despite pathogenic cells

present common features, the outcome of the infection present variability. Some pathogenic cells do not manage to infect any cell and die in the extracellular space while the ones that succeed in infecting the host cells can survive in the intracellular space or not. This variability does not only points out that some pathogenic cells present a better transcriptome to infect and survive inside the host, but also that some host cells are more prepared to face the infection. Therefore, host-pathogen interactions have to be seen as the confrontation of two groups of heterogenic cells. Cell heterogeneity is an important characteristic that confers plasticity and flexibility to survive to changing environmental conditions during infection.

The aim of this report is to review the current state of scRNAseq technology and to understand how it can be used to study the role of cell heterogeneity in host-pathogen interactions. First of all, methods to isolate single cells are explained followed by the approaches developed to sequence RNA of single cells. Then, some studies that were carried on with scRNAseq are exposed and discussed.

Single cell isolation

Before RNA sequencing, cells from a population have to be isolated. There are several methods to achieve it. They differ in the speed of sorting (throughput), the cell characteristics that the experimentalist can determine before isolation (visualization) and the sample volume. To identify rare phenotypes (expressed in <1 % cells) hundreds of single cells have to be analysed and a method of high throughput is thus preferred (Grün & van Oudenaarden, 2015). However, for some studies is more important to visualize cells with microscopy, which compromises the throughput. Low samples volumes are recommended in order to minimize degradation and increase the effective concentration (Jianbin Wang, Fan, Behr, & Quake, 2012). In the next paragraphs the isolation methods based on microfluidic devices, on micromanipulation and on flow cytometry are addressed (table 1).

Microfluidic devices can be used to sort and isolate cells based on the observed morphology under microscopy. Also, cells can be selected based on DNA expression using fluorescence reporters, as GFP (Gossett et al., 2012; Zhang et al., 2012). Electric and magnetic properties can also be evaluated (Huang et al., 2008; Vahey, Quiros Pesudo, Svensson, Samson, & Voldman, 2013). An advantage of this method is that the starting volume is low. A recent commercially available microfluidic system, the C1^{TM} single-cell mRNA sequencing, automates the sorting and also the RNA sequencing inside the device. This system increased the throughput to 96 cells, keeping the cell visualization. However, this system is only available for eukaryotic cells at the moment. Two other microfluidic systems that automatized sorting and RNA sequencing of single cells are Drop-seq (Macosko et al., 2015) and iDrop sequence (Klein et al., 2015). The sorting is based on droplets. However, these systems are not commercialized and the device design is laborious.

Another option is micromanipulation. Single cells can be aspired with micropipettes from a culture or complex mixtures with a joystick while visualizing them under the microscope. It is a suitable method when it is necessary to visualize cells before isolation. The drawback is that it is laborious and, consequentially, the throughput is really low (Fröhlich & König, 2000).

Laser based micromanipulation is also possible, without the need of using micropipettes. Optical tweezers have been used to trap individual cells in suspension

and laser capture microdissection for cell tissues (Fröhlich & König, 2000; Saliba, Westermann, Gorski, & Vogel, 2014).

Fluorescent activating cell sorting is currently the isolating method of choice for many scRNAseq studies (Grün & van Oudenaarden, 2015; Saliba et al., 2014). It can be used for both eukaryotic and prokaryotic cells. The main advantage is that it sorts and isolates hundreds of cells in a few minutes. The possibility to sort cells makes this method convenient to study only the subpopulation of interest. For example, it can isolate immune cells that present a specific combination of antigens. Subpopulations can also be sorted based on the presence of gene expression, pigment or biochemical activity (Verschoor, Lelic, Bramson, & Bowdish, 2015).

When cells are loaded in a flow cytometer a nozzle produces drops that contain one single cell. When cells pass through a laser beam a scattering light is generated, which depends on the cell characteristics. Currently, up to 20 parameters can be characterized for every cell. Two parameters are cell size and granularity of the cytoplasm. The other parameters depend on the fluorescence emitted by the cell, as there are flow cytometers with up to 18 colour filters (Chattopadhyay & Roederer, 2012). The fluorescence can reflect gene expression or cell-surface markers by the use of fluorescent gene reporters or labelled antibodies, respectively. When cells with the desired parameters are identified they are isolated into multi-well plates (96 or 384-well plates) (Saliba et al., 2014).

The main advantage of FACS is that it is a high throughput method. The only disadvantages are that cells cannot be visualized and it needs fluorescence induction for the sorting, that might alter the cell physiology. Raman-activated cell sorting (RACS) has been suggested as an alternative to FACS without the need of fluorescence induction. Raman single-cell spectroscopy applies a laser beam to each molecule. The scattering incident light generates a spectrum of more than 1000 Raman bands, which depends on the biomolecules vibration inside the cell. Consequently, cells containing different amount and kind of molecules can be sorted. Therefore, Raman spectrums can serve as a molecular fingerprint for single cells (M. Li, Xu, Romero-Gonzalez, Banwart, & Huang, 2012), which have been used for antibody characterization (A. E. Baker, Mantz, & Chiu, 2014).

Table 1. Isolation techniques.

	Throughput	Visualization	Sample volume
FACS (or RACS)	High	Not possible	High
Microfluidics	Medium	Yes	Low
Micromanipulation	Low	Yes	Low

RNA sequencing

Once the cells are isolated, RNA sequencing for these cells can be tackled. What makes scRNAseq more challenging than standard bulk RNAseq is the low amount of starting RNA that has to be reverse transcribed to cDNA and ultimately sequenced. Eukaryotic cells contain around 10-50 pg of total RNA per cell and prokaryotic cells around 0.0015-2 pg (Y. Kang et al., 2011; Jiangxin Wang, Chen, Chen, & Zhang, 2015). Therefore, scRNAseq has to be extremely sensitive, specially for prokaryotic samples. The key step to have enough amount of cDNA is the amplification. In traditional bulk RNAseq methods the RNA is first fragmented and then amplified

because polymerases processivity is limited and cannot completely extend long transcripts. However, for scRNAseq experiments samples are first amplified and then fragmented, as some RNA molecules are degraded during fragmentation. Several strategies have been developed for scRNAseq amplification, which try to avoid amplification bias, loss of strand information and low coverage.

In the last years several single-cell transcriptomic studies with eukaryotic cells have been published (Grün & van Oudenaarden, 2015; Saliba et al., 2014) but the little RNA amount of single bacterial cells is still a limitation, with only one published study (Jiangxin Wang et al., 2015).

The main steps for the RNA sequencing are RNA extraction, mRNA capture, reverse transcription to have the first strand of cDNA, second strand synthesis, amplification and finally cDNA library preparation for next-generation sequencing technologies. The first steps, from RNA extraction to reverse transcription are common among the studies. On the contrary, several strategies have been developed for the cDNA second strand synthesis and amplification of cDNA. The next-generation sequencing technology of choice in the scRNAseq community is Illumina sequencing (Trombetta et al., 2014).

First-strand cDNA synthesis

First of all, cells are lysed to extract the RNA. A hypotonic lysis buffer with low concentration of both RNase-inhibitor and surfactant is the preferred method because it does not need a post-lysis clean-up, which causes RNA loss and reduces the sensitivity of the assay (Trombetta et al., 2014). Second, the RNA of interest has to be captured. For both eukaryotic and prokaryotic cells, it has been calculated that the 80% of the total RNA is rRNA while the 15% corresponds to tRNA (Lodish et al., 2000; Rosenow, Saxena, Durst, & Gingeras, 2001). In order to optimize the use of next generation sequencing technologies they have to be removed. For eukaryotic cells, mRNA can be selectively amplified with poly(T) primers that anneal with the 3' polyadenilated tails of mRNA, despite this also causes the loss of non-coding RNAs. For prokaryotic cells this approach is not possible, as in these cells the polyadenilated RNA is a mark for degradation. Alternatively, rRNA and tRNA can be degraded with 5' monophosphate-dependent exonucleases. However, it can also degrade some mRNA, which introduces a bias in the relative transcript amount. Third, the captured mRNA has to be reverse transcribed to obtain the cDNA first-strand. The reverse polymerase of choice is the M-MuLV, an engineered version of the Moloney Murine Leukemia Virues reverse transcriptase that presents low RNase H activity and increased thermostability, resulting in higher yields. Poly(T) primers or DNA random primers can be used for the reverse transcription. This results in RNA:DNA hybrid molecules with an averaged size of 1.5-2 kb (Saliba et al., 2014).

Second cDNA strand and amplification

Several approaches have been used to synthesize the second cDNA strand and amplify the cDNA in order to have enough sensitivity for the sequencing. The four different approaches that have been developed are based on PCR amplification, *in vitro* transcription, isothermal amplification, and rolling circle amplification (figure 1 and table 2).

For PCR amplification two primers are needed. At 5' terminal of the cDNA, the same that was used for the first strand synthesis can be used (generally poly(T) primer). A way to extend from the 3' end is to introduce a homopolymer tailing, usually a tail of 30 poly(A) nucleotides, and ligate the tail with the complementary sequence of an universal primer. As a consequence, after sequencing the cDNA the strand information will be lost, as it is not possible to recognise which strand originally had the poly(A) at the 3' terminal end. Another problem of this approach is that during the reverse transcriptase the polymerase has terminations before reaching the RNA 5', and sequencing at this region is underrepresented (Grün & van Oudenaarden, 2015; Saliba et al., 2014; Sasagawa et al., 2013). In order to solve this problem and cover the sequence of the whole transcript, the template-switching mechanism property of the M-MuLV polymerase can be used. This polymerase introduces three or four cytosines at the 3' end of the first cDNA strand only when it reaches the end of the RNA template. Then, the complementary sequence of the universal primer for the second-strand synthesis is ligated after these cytosines. Therefore, just the whole transcribed transcripts will be amplified (Islam et al., 2011; Saliba et al., 2014). Another alternative to have full-length coverage was developed with the use of random primers that can anneal along the transcript. Each primer has a different combination of 9 nucleotides at 3' terminal and a universal sequence with a restriction site for BciVI at 5' terminal. This restriction site is surrounded by complementary sequences to form a hairpin and prevent self-annealing. First, overlapping amplicons, which together cover the whole transcript, are generated for three PCR cycles. Then, the random primers are substituted with universal 5' primers that contain the restriction site. Finally, BciVI digestion removes the primed sequence to introduce sequencing primers (Pan et al., 2013).

A common problem of PCR approaches is the amplification bias. Transcripts amplified with primers that bind with high affinity will be overrepresented. On the other side, if a transcript is amplified with a primer with some mismatches the amplification will be less efficient and consequently the transcript will be underrepresented. An approach to overcome this bias and have linear amplification is to amplify with *in vitro* transcription. In this approach, the primer for the first strand synthesis comprises a T7 promotor. Next, the synthesis of a second-strand is performed, which will have the T7 promoter at 3'. Then, the *in vitro* transcription is possible and an RNA strand complementary to the original one is amplified. Reverse transcriptase for all transcripts is done afterwards to have the cDNA. A drawback of this technique is that during the *in vitro* transcription the polymerase can have spontaneous terminations, which cause a loss of sequence information at the 5' of the original RNA transcript (Hashimshony, Wagner, Sher, & Yanai, 2012).

Isothermal amplification is another option to amplify the cDNA with high sensitivity. For this approach, chimeric primers, DNA polymerase with strand displacement activity and RNaseH endonucleases are needed. The first-strand generation is performed with a chimeric primer, which has an RNA sequence at 5' and a DNA poly(T) or random primers at 3'. The cDNA second strand is synthesized with a DNA polymerase and a DNA primer. Then, the amplification of this second strand is performed with the chimeric primer. After each cycle the RNA sequence at 5' of the new amplicon is removed with RNaseH and for the next cycle new chimeric primers are introduced (Kurn et al., 2005). By using random hexamer primers for the first-strand synthesis, this

approach was used for the only published study that performed scRNAseq in prokaryotes. However, in this study rRNA and tRNA were not fished out previously, which implies that it does not overcome the problem of having low amount of RNA in single cells after mRNA enrichment (Jiangxin Wang et al., 2015).

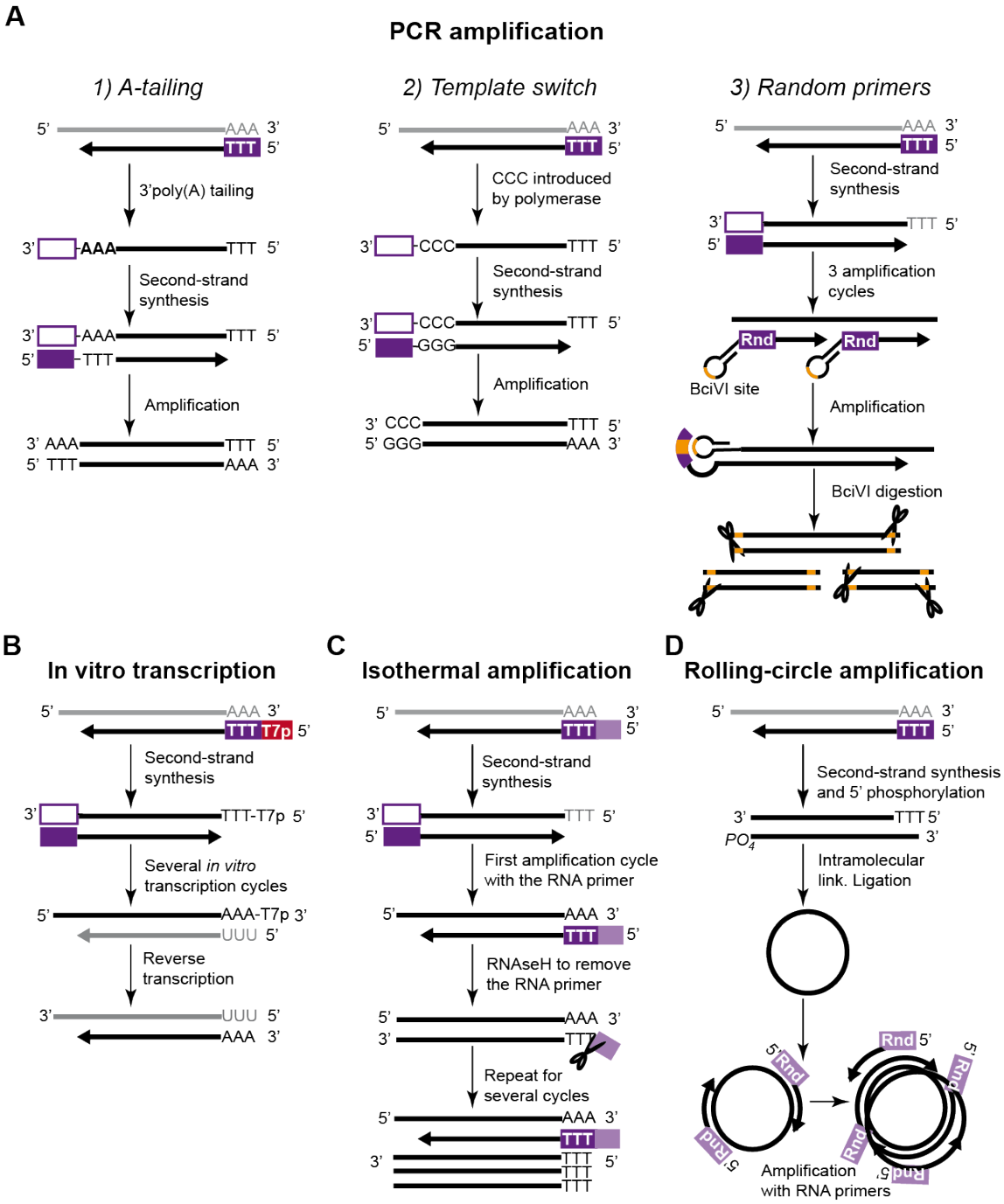

Figure 1. Amplification approaches. After cell lysis, RNA has to be amplified before scRNAseq. The three alternatives of PCR amplification (A), the amplification with *in vitro* transcription (B), the isothermal amplification (D) and the rolling-circle amplification (D) are depicted. For first strand synthesis (first step), random primers can also be used instead of poly(T) primers. Gray lines depict RNA and black lines cDNA. RNA primers are shown in light purple and DNA primers in dark purple, empty squares correspond to the primer sequence ligated to the template and full squares to the extension primer. In red the primer sequence that corresponds to the T7 promotor and, in orange, to the BciVI restriction site. Rnd stands for random and T7p for T7 promoter sequence.

Another approach uses the DNA polymerase of bacteriophage Φ29, which has strand displacement activity and high processivity for long DNA templates. This approach is called rolling circle amplification because the cDNA is ligated in order to have a "long" DNA template and take advantage of the polymerase high processivity. The ligation of 3' and 5' ends is achieved with an intramolecular link by an ATP-dependent ligase after 5' phosphorylation. Then, the cDNA circle is amplified with Φ29 polymerase using random primers. If DNA primers are used, the polymerase can also use them as a template, which causes undesired by-products. RNA primers are preferentially used because Φ29 polymerase can extend RNA primers but cannot use RNA as a template (Y. Kang et al., 2011; Yun Kang, McMillan, Norris, & Hoang, 2015; Pan et al., 2013). This approach was reported to have full-length coverage and it has been done for both eukaryotic (Pan et al., 2013) and prokaryotic cells (Y. Kang et al., 2011; Yun Kang et al., 2015). When this method was done for prokaryotic cells the mRNA enrichment step was followed. However, the studied organism, *Burkholderia thailendensis*, has 2 pg of RNA, several orders of magnitude higher than the standard prokaryotic cells.

Spike-in RNA, barcoding, and strand specificity strategies

In order to study gene expression with RNAseq technology is necessary to identify and quantify all transcripts. The sequencing results present biological and technical variability. In order to study biological variability in cell populations, the technical variability introduced mainly during amplification has to be removed. With this aim, spike-in RNA are unique molecular identifiers (UMIs) are used (Grün & van Oudenaarden, 2015; Islam et al., 2014; Stegle et al., 2015).

Spike-in RNA molecules are introduced during sequencing as a negative controls in order to normalize the expression values. The External RNA Control Consortium (ERCC) (S. C. Baker et al., 2005) designed a set of 92 synthetic RNA with several lengths of known concentration that should mimic the natural RNA molecules. They are mixed with the extracted RNA from cells, representing a 5-10% of the total sample. Differences in the quantity of spike-in molecules between samples indicate technical variability, as the starting spike-in molecules are identical in all samples (Stegle et al., 2015). For instance, if an RNA sample of one cell is not amplified neither its spike-in RNA. However, it is not clear if ERCC RNAs perfectly mimic natural RNAs and some authors proposed to use naturally expressed genes, for example housekeeping genes, for the normalization (Risso, Ngai, Speed, & Dudoit, 2014).

Another source of technical variability is the different amplification efficiency between genes. For example, transcripts with higher affinity for the primers will be more represented than transcripts with less affinity for the primers. Other factors as the GC content, the secondary structure, the size and the quantity of the transcripts are other sources of amplification bias (Kebschull & Zador, 2015). UMIs are barcodes to cope with these biases. A barcode is a short sequence that is used as identifier of one molecule or a group of them. A different UMI barcode is introduced for each different transcript from the same cell. During the analysis, UMIs are counted to know how many times each transcript was amplified (Islam et al., 2014).

Barcodes are also used to mark cell identity. Each RNA set from a single cell is tagged with a specific barcode either during first-strand synthesis or downstream to the sequencing primers. Then, transcripts from all cells can be pooled and sequenced

together, which reduces the cost of the experiment (Grün & van Oudenaarden, 2015). If the cell barcode is introduced during first strand synthesis, an early pooling is possible, reducing even more the experiment cost.

Another information that is important to keep is which RNA strand was extracted from the cell, as the sequencing is done with double stranded cDNA. In order to preserve this information, differential primers can be used to tag the original 3' or 5' end. However, if one of these ends are tagged and selected, the coverage of the opposite end is lost, compromising the full-length coverage. STRT is an example of 5' terminal end selection. During STRT, cDNA is immobilized on beads thorough the 5' end after amplification, then samples are fragmented loosing the information at the 3' end (Islam et al., 2011). A complementary method is CEL-seq. In it, the first primer sequencing is added during single cDNA strand synthesis, marking the 3' end of the RNA. Then, the second sequencing primer is introduced after fragmentation. Just fragments that have both sequencing primers are sequenced, loosing coverage at the 5' end (Hashimshony et al., 2012).

Table 2. Developed methods of scRNAseq and its characteristics.

	Amplification method	Coverage	Linearity	UMIS	Strand information	Early pooling	Poly(A) dependent[1]	Ref
STRT	PCR-template switch	Only 5'	No	Yes	Yes	Yes	Yes	Islam et al., 2011
SMART-seq2	PCR-template switch	Full, 3' bias	No	No	No	No	Yes	Picelli et al., 2014
Quartz	PCR-A tailing	Full, 3' bias	No	Possible[2]	No	Possible	Yes	Sasagawa et al., 2013
SMA	PCR-based	Full	No	No	No	No	No	Pan et al., 2013
CEL-seq /MARS-Cell[3]	In vitro transcription	Only 3'	Yes	Yes	Yes	Yes	Yes	Hashimshony et al., 2012; Jaitin et al., 2014
TIVA[4]	In vitro transcription	Full, 3' bias	Yes	Possible	Possible	Possible	Yes	Lovatt et al., 2014
SPIA	Isothermal amplification	Full	Yes	No	Possible	No	No	Kurn et al., 2005; Jiangxin Wang et al., 2015
PMA	Circle-rolling amplification	Full	No	No	No	No	No	Pan et al., 2013

[1]Methods that are not dependent on Poly(A) tailing can be used for prokaryotes. However, only SPIA and PMA have been used for bacterial cells.
[2]Possible means that the design of the method is compatible with the characteristic but that it has not been done.
[3]MARS-cell is the automated version based on CEL-seq.
[4]TIVA: is the only in vivo method for scRNAseq.

Data analysis

The aim of a single-cell RNAseq analysis is to identify which genes are being transcribed in each single cell and quantify its expression level. Using these parameters it is possible to group cells based on transcription expression and characterize subpopulation or different cell types.

First of all, mapping is done aligning sequences to a reference transcriptome. If it is not available, a reference genome can be used. Sequences corresponding to barcodes have to be removed before the mapping. These initial steps are used with the same tools for bulk cells RNAseq data. Raw data should be filtered in order to reduce the technical noise. In some cases, if few bases at the ends of the transcripts have low quality, they can be trimmed. (Grün & van Oudenaarden, 2015; P. Li, Piao, Shon, & Ryu, 2015; Stegle et al., 2015).

To quantify the expression, reads have to be assigned to each studied cell with the aid of cell barcodes. If cDNA was linearly amplified, the transcripts expression level for each cell can be estimated. On the contrary, UMIs barcodes are used to eliminate PCR bias. By counting the number of different UMIs for one transcript, the number of original copies of this transcript can be known. Cells presenting low transcript numbers can be eliminated of the study to avoid a biased quantification. The main causes of low transcript numbers are low amplification efficiency (when spike-in RNAs also have a low number of copies) or low input material (if spike-in RNAs are overrepresented in the amplification product) (Grün & van Oudenaarden, 2015).

Normalization of expression data is necessary to compare and distinguish differences in expression levels between transcripts and cells. When reads cover the whole transcript, the expression levels are usually expressed with transcripts per one million reads per kilobase of transcript (RPKM). However, there are several and more refined methods for normalization. Some methods used for bulk RNAseq, as RSEM (B. Li & Dewey, 2011) and Sailfish (Patro, Mount, & Kingsford, 2014), make abundance estimation before normalization, improving the normalized results (P. Li et al., 2015). Alternatively, sub-sampling of the same number of transcripts for each cell, a method called down-sampling also succeeds to reduce technical-noise but then some transcripts are not considered, losing sample complexity. (Grün, Kester, & van Oudenaarden, 2014).

After sequencing data analysis, further analysis to interpret the data can be done. Usually, clustering is performed in order to identify subpopulations or cell types. This classification is done based on the differential expression levels and the kind of expressed genes. A common unsupervised method is principal component analysis (PCA) but other methods as hierarchical clustering, k-means, network analysis or even combination of them (Grün & van Oudenaarden, 2015). If prior information is available, for example gene expression determined with reporter gens, this can be merged with the unsupervised methods (Stegle et al., 2015). It is important to identify and, if necessary, deconvolve the changes in expression due oscillatory cellular processes like cell cycle or metabolism, which can alter the results if cells do not grow synchronously (Francesconi & Lehner, 2015).

Single-cell RNAseq was developed to study molecular reprogramming

Tang et al were the first ones using RNAseq for single cells (F. Tang et al., 2009; F. Tang, Barbacioru, Nordman, et al., 2010). They developed this method to study changes in expression from internal cell mass (ICM) cells to embryonic stem cells (ESC) in mice (F. Tang, Barbacioru, Bao, et al., 2010). Cells were manually isolated with a mouth pipette. A PCR based amplification, using poly(A) tailing was performed and then the library prepared. They sequenced a total of 12 cells simultaneously, barcodes to identify cells were used and finally the selected sequencing technology was SOLiD.

During this study, they identified a total of 4,837 genes with a different expression between IMC and ESC cells. These genes are implicated in mRNA processing, splicing, protein modification, cytoskeleton, microtubules, carbohydrate metabolic processes and protein catabolic processes. Then, based on the gene expression dynamics during IMC outgrowth, four gene clusters were identified. One cluster of genes was highly upregulated in ESC, which means that they are specific for maintaining the self-renewal. Two other groups had a decrease of expression, one during early stages and the other during late stages. And finally another group of genes was identified as responsible of pluripotency shut down (F. Tang, Barbacioru, Bao, et al., 2010).

Single cells RNAseq analysis also revealed some mechanisms of cell reprogramming during development. New splicing variants were observed in ESC cells compared to IMC cells and expression level changes were observed for epigenetic regulators (F. Tang, Barbacioru, Bao, et al., 2010).

Therefore, this study proved that scRNAseq analysis is a suitable tool to identify different cell types. Moreover, they determined the genes that dynamically change its expression to reprogram ICM cells to ESC.

Similarly, during pathogen-host immune response, cells from both organisms undergo changes in gene expression in order to survive and colonize the host or alternatively to stop colonization of pathogen. Thus, scRNAseq analysis can be used to describe the initial heterogeneity of pathogen and host cells and determine the expression changes after their encounter. RNA sequencing in single mammalian host cells as macrophages or dendritic cells is possible and have been done (Avraham et al., 2015; Shalek et al., 2014). However, the study of bacterial single cells is still challenging.

Single-cellRNAseq analysis to study the influence of LPS on host-pathogen interactions

Recently, two studies (Avraham et al., 2015; Shalek et al., 2014) performed scRNAseq to investigate LPS response of immune cells, showing the importance of both host and pathogen cell heterogeneity. With the determination of single cell transcripts, genes that are just expressed only in few cells were identified. As a consequence, a better characterization of the temporal regulation and the identification of signals that trigger activation of pathways in response to LPS were possible.

Signalling pathways activated during LPS stimulation

During pathogen and host encounters, both organisms undergo changes in expression in order to invade host cells or eradicate the bacteria, respectively. Bacteria present patterns known as pathogen-associated molecular patterns (PAMP) that are recognized with extracellular receptors of host cells. Common PAMPS in bacteria are

lipoproteins, lipopolysaccharides (LPS), peptidoglycans, bacterial DNA or flagella. PAMPS are recognized by Toll-like receptors (TLR), which transduce the signal into the intracellular compartment and regulate gene expression. Each PAMP is recognised by a different TLR (Newton & Dixit, 2012; D. Tang, Kang, Coyne, Zeh, & Lotze, 2012).

LPS are an important PAMP of gram-negative bacteria to activate the immune response in macrophages, dendritic cells (DC), lympohcytes or other immune cells. LPS are recognised by TLR4. This receptor has been shown to regulate the hypersentivity response (Freudenberg et al., 2003). In *Salmonella*, lack of TLR2, TLR7 or TLR9 are compensated by TLR4 (Arpaia et al., 2011; Feuillet et al., 2006; Weiss, Raupach, Takeda, Akira, & Zychlinsky, 2004). Thus, TLR4 has a central role in immunity. Once TLR4 receptors recognize LPS, it can remain in the membrane, activating Myd88, or it can be internalized by endosomes, activating Trif. Activation of Myd88 is common with the other TLR, which results in Nuclear Factor κβ (NF-κβ) activation, which upregulates the transcription of pro-inflammatory cytokines. If TLR4 activates Trif, then the transcription factor Irf3 induces the transcription of Type 1 interferon (IFN) (figure 2). When type I IFNs are secreted, they can be recognised by their receptors. For example, it has been described that IFN-β is recognized by IFNAR, which activates the transcription factor Stat1 to express cytokines that activate the antiviral response of uninfected cells (Doyle et al., 2002).

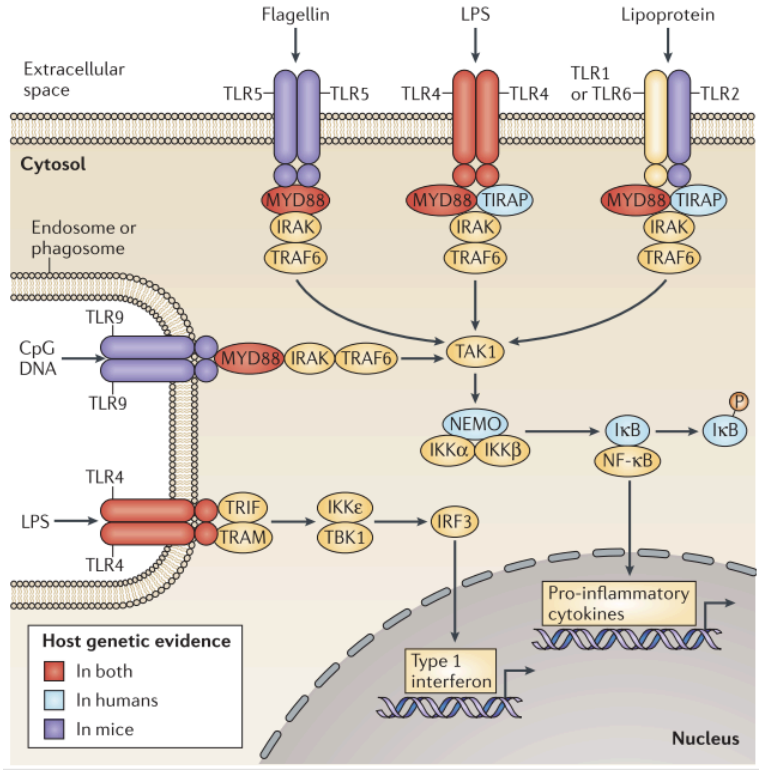

Figure 2. Signalling pathways activated as a response to PAMPS in immune cells. From Gilchrist, MacLennan, & Hill, 2015.

Influence of host and pathogen cell heterogeneity

Shalek et al. (2014) studied LPS stimulation on mice DC over time. To sort DC cells from bone marrow extracts, cells expressing CD11 antibody were selected with FACS. Then, these cells were isolated with the C1 microfluidics device from Fluidigim. The SMART-seq amplification method was followed with Nextera sequencing. To analyse the sequencing data, PCA analysis were carried out and then clusters were correlated

with core antiviral and inflammatory genes. Finally, in order to characterize changes in expression, they considered two types of heterogeneity: analogue and digital. Digital refers to presence or absence of transcription for a given gene (on/off) and analogue to changes in the expression levels (more/less). For all analysed gens, they calculated digital heterogeneity by measuring the fraction of detectable expressing cells (α) and the analogue heterogeneity with changes in the variance of the mean (σ) (figure 3).

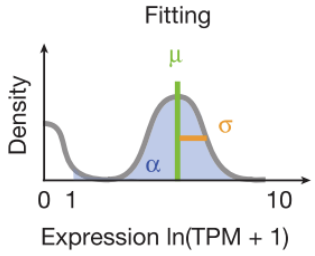

Figure 3. Determination of digital and analogue expression. For each gene, the fraction of detectable expressing cells was measured (α, in blue), which accounts for the digital expression. The expression values of the expressing cells were fitted in a normal distribution and then the variance of the mean (σ in orange) calculated, which accounts for the analogue expression. Figure from Shalek et al. (2014).

They observed that the cluster of core antiviral genes presented a bimodal response during early LPS stimulation response (after 2 h), with some cells having a low expression and others a high one. After 4 h, more cells had a high expression, and the expression distribution became unimodal. On the contrary, the cluster of peak inflammatory genes started with an unimodal high expression followed by a bimodal expression after 4 h, as some cells had an expression decreased. They identified IFN-β as the key regulator of these changes in the expression.

In this study it was possible to identify two cells out of 75 (2,7 %) that express the core antiviral genes after 1 h of infection. The authors described them as precocious cells and they hypothesized that these cells could trigger the other cells response through cell communication. In order to test this hypothesis, they performed the experiments sealing each chamber of the microfluidic device, impairing cell-to-cell communication. Without cell-to-cell communication, a large proportion of cells failed to express the core antiviral genes, maintaining the bimodal response for this cluster after 4 h. On the other hand, the impaired communication maintained the unimodal response of peak inflammatory genes. These results indicated that cell-to-cell communication is necessary to globally switch on the core antiviral genes and down-regulate peak inflammatory genes. Similar results were obtained using cells that lacked IFN-β or Stat1 genes (*Ifnar*$^{-/-}$ or *Stat1*$^{-/-}$). If secretion of IFN-β was impaired after one hour, only the peak inflammatory genes were altered, which means that peak inflammatory genes are regulated in a second wave of cytokines secretion.

In conclusion, this study supports a model of heterogenic response to LPS stimulation. All cells can respond to LPS, activating peak inflammatory genes. However, only a little percentage (around 2-3 %) of cells presents an early transcriptional change in response to IFN-β. The first effect is the expression of core antiviral genes. Then, these cells release more IFN-β, switching on the expression of neighbour cells with this paracrine signal. Lately, negative paracrine signals are secreted, inhibiting peak inflammatory genes expression (figure 4).

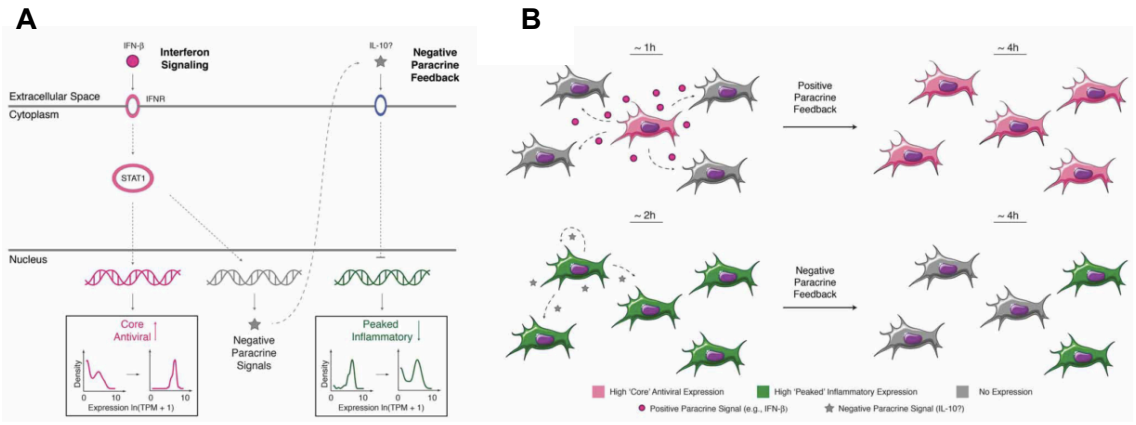

Figure 4. Conclusions of the study of Shalek et al. (2014). On the left, the gene network model showing how positive IFN-b signalling induces the antiviral response and reduces its heterogeneity, while simultaneously activates the negative paracrine feedback, which inhibits the peaked inflammatory cluster and increases its heterogeneity. On the right, cell population model showing how positive and negative paracrine signalling alter antiviral (magenta) and inflammatory (green) gene expression variability across cells. Grey denotes no expression. From supplementary material of Shalek et al. (2014).

Avraham et al. (2015) also studied LPS response in immune cells over time. They infected mice macrophages with *Salmonella typhimurium*. First, they sorted and isolated the cells with FACS based on the outcome of the infection. SMART-seq amplification method with Nextera sequencing was carried on. Sequencing data was clustered with PCA and weighted gene correlation network analysis (WGCNA).

When *S. typhimurium* infects macrophages cells, there is heterogeneity on the outcome. The three possible outcomes are no infection, infection with intracellular bacterial death or infection with intracellular survival of bacteria. These authors elaborated a system to distinguish the infection outcome with FACS. They used GFP-expressing *S. typhimurium* and stained its membrane with the red dye pHrodo. Therefore, red signal (pHrodo) means presence of infection and green signal (GFP) bacterial survival. Non-infected cells would not present fluorescence, infected macrophages with bacterial death only red fluorescence and infected macrophages with bacteria survival would have a yellow signal (presence of green and red fluorescence) (figure 5). The macrophages isolated with FACS were used for scRNAseq and also unexposed macrophages were isolated as a control.

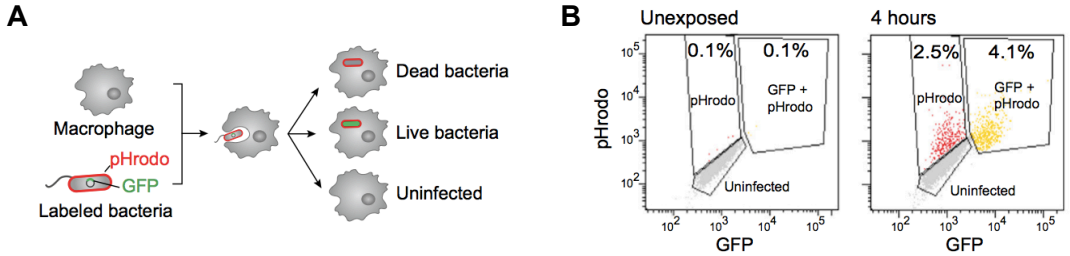

Figure 5. Cell sorting with FACS based on the infection outcome. The design of the approach used to sort the three possible outcomes of *S. typhimurium* infection based on GFP and pHrodo fluorescence (A). And the FACS analysis of fluorescently labeled populations for unexposed cells in left and for exposed cells after 4 h in right (B). Figures from Avraham et al. (2015).

The PCA identified two clusters. Genes in cluster I corresponded to extracellular bacteria exposure, which comprised genes activated in response of LPS; genes in cluster II corresponded to intracellular bacteria response. All exposed cells expressed genes of cluster I while just the infected cells expressed genes of cluster II. However, at later time points (8 h), uninfected cells also expressed genes from cluster II, suggesting

cell-to-cell communication. WGCNA indentified a third cluster that was enriched for the type I IFN response. Then, they compared the response of *Tlr4*$^{-/-}$, *Trif*$^{-/-}$, *Myd88*$^{-/-}$ and *Irf3*$^{-/-}$ mutants to determine through which signalling cascade the type IFN expression is induced. *Trif*$^{-/-}$ and *Irf3*$^{-/-}$ mice failed to expressed cluster III genes, indicating that TLR4 signalling through Trif and Irf3 activates type I INF response in macrophages after *S. typhimurium* infection.

Similar to Shalek et al. (2014) they identified a bimodal response for genes in cluster III. However, despite the cell-to-cell communication, cells with low gene expression were still present after 4 and 8 h, maintaining the bimodal response. To understand the differences between the two studies, Avraham et al. (2015) performed the same experiment using coated beads instead of bacteria. As a result, more cells had a high expression of cluster III genes at later time points, which resembles the previous results (Shalek et al., 2014). It was concluded that there is a bacterial factor that maintains the heterogeneity of host cells response.

In order to identify the bacterial factor responsible to maintain the bimodal response, a GFP reporter to determine cells with Irf3 activity was used. Infected cells with Irf3 activity, which means that they have the type I INF response activated, were sorted with FACS. For both, macrophages and bacteria, bulk RNAseq was performed for infected cells with Irf3 activity and for infected cells without Irf3 activity. Bacteria extracted from macrophages with Irf3 activity presented an increased PhoP expression compared with cells that did not have Irf3 activity. PhoP, together with PhoQ, are a two-component system that regulates intracellular bacterial survival (Groisman, 2001). Moreover, PhoP also controls the expression of genes that modify LPS (Ernst, Guina, & Miller, 2001). Macrophages stimulation with LPS extracted from constitutively expressing PhoP bacteria, resulted in an unimodal response of high expression type I IFN response (figure 6). Therefore, differential LPS modifications are the cause of macrophages differential responses of type I IFN activity.

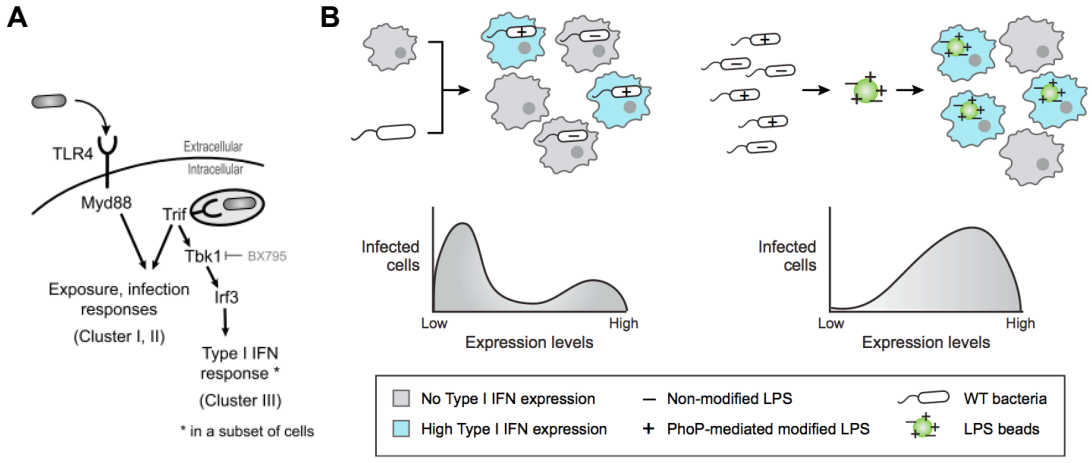

Figure 6. Conclusion of Avraham et al. (2015) study. Signalling cascade activated in macrophages expressing genes from cluster III as a response of S. typhimurium infection (A). Schematic representation of the differences in the responses of macrophages to infection with live bacteria and to stimulation with LPS-coated beads. Live bacteria are more heterogeneous than LPS-coated beads (B). Figures from Avraham et al. (2015).

In conclusion, in this study it was shown that heterogeneity of pathogenic cells induce a heterogenic host cell response. Concretely, macrophages infected with *S. typhimurium* bacteria expressing PhoP activated type I IFN response through Trif and Irf3. On the contrary, macrophages infected with bacteria without PhoP activity, did not activate

genes upregulated during type I IFN response. Despite evidences of cell-to-cell communication were observed, the response remained heterogenic, probably because the source of heterogeneity was present during the whole infection.

It is important to note that in both studies (Avraham et al., 2015; Shalek et al., 2014) the scRNAseq expression correlated with bulk RNAseq analysis, which shows that the low starting RNA sample did not biased the results (figure 7). Moreover, also in both studies, RNA FISH experiments were carried out for at least five genes to validate the results obtained with scRNAseq.

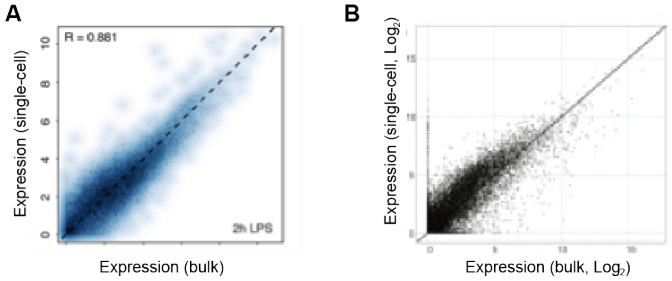

Figure 7. Correlation between averaged single-cell expression and bulk expression. Scatter plots from Shalek et al. (2014) (A) and Avraham et al. (2015) (B) that show the relation between the average single-cell expression (y axis) and bulk level expression (x axis) for a gene in a concrete time after infection.

To sum up, these studies showed that it is possible to perform scRNAseq studies of host cells during infection. Which permits to study heterogeneity in pathogenic and host cells and its influence during pathogen-host interaction. One study (Shalek et al., 2014) showed that host cells present heterogeneity in the activation threshold for IFN-β while the other study (Avraham et al., 2015) supports that heterogeneity in bacteria cells induce another level of heterogeneity in the host cells response. Therefore, scRNAseq studies can provide detailed information of how and when the pathways already described are regulated and the importance of other factors like paracrine signals.

Discussion

The aim of this essay is to review the technical advances of the scRNAseq technology and its possible application to study pathogen-host interactions. Concerning the technique, the main challenge of the technique is the low starting RNA amount from single cells. Compared with bulk RNAseq, moving the amplification step before fragmentation permitted to carry on scRNAseq with eukaryotic cells and with procariotic cells if the rRNA and tRNA is not fished out. With the development of several amplification approaches is possible to preserve full-length coverage, strand information and to either linearly amplify the cDNA or introduce unique molecular identifiers to correct the amplification bias. However, any amplification approach has been used to sequence transcripts of single prokaryotic cells after fishing out the rRNA and tRNA. New improvements to either make this technique even more sensitive or fish out the rRNA and tRNA with less mRNA looses have the potential to solve this limitation.

ScRNAseq was initially developed to study changes during development in order to distinguish transcriptional changes between cell types. Nowadays this technique has been also used to study splicing events, cell reprogramming during cancer progression and host-pathogen interactions (Saliba et al., 2014). In this essay, two papers to study host-pathogen interactions were reviewed in order to understand how this technique is used to study the role of cell heterogeneity during infections.

If bulk RNAseq is performed, is not possible to determine whether changes of expression are due digital expression regulation (change in number of cells expressing a gene) or due a global analogue expression regulation (changes in the number of transcripts per cell). Shalek et al. (2014) could differentiate these two kinds of expression changes by using scRNAseq. For instance, they could identify that only a 3 % of the total population started expressing a set of genes. On later time points, a greater percentage of cells expressed the same set of genes, establishing how the global expression is temporally regulated by paracrine signals (figure 8 A and B1).

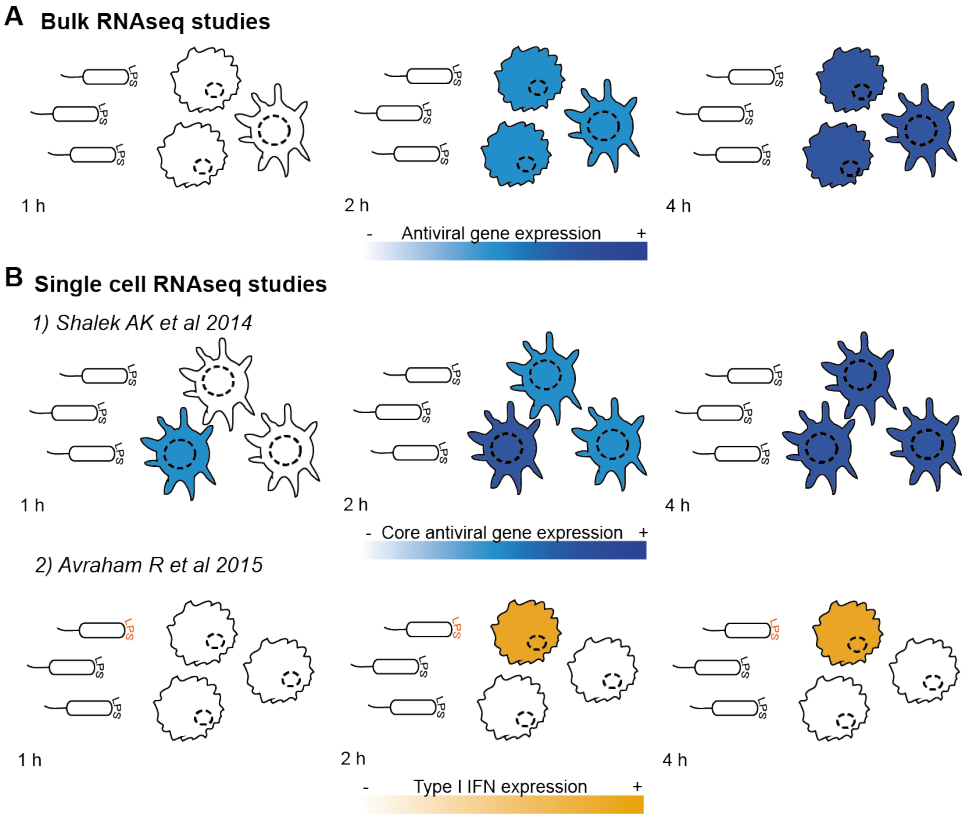

Figure 8. Comparison between insights from bulk RNAseq and scRNAseq studies in host-pathogen interactions. With bulk RNAseq is possible to determine the averaged expression of antiviral genes in host cell populations as a response of bacterial infection. This approach assumes that the expression levels are homogeneous in all cells (A). scRNAseq studies reveled heterogeneity in the expression (B). Shalek et al. (2014) showed that dendritic cells, in response to a homogenic LPS presence, some precocious cells switch on the core antiviral gene expression earlier by determining digital heterogeneity (in 1h) and analogue heterogeneity (in 2h). Avraham et al. (2015) could detect a cluster of cells that expressed Type I IFN response when they encountered a PhoP modified LPS. In blue, antiviral gene expression levels (lighter, less; darker, more). In orange, Type I IFN response expression levels (lighter, less; darker, more). LPS in orange depicts LPS that present a PhoP modification.

Using scRNAseq Avraham et al. (2015) could determine a cluster of cells (cluster III) that had the same phenotype to the other infected cells, but that presented an up-regulation of only around 100 genes. Cells in this cluster presented a differential expression because they were exposed to pathogenic cells that expressed a gene that modify LPS (the PhoP gene). Despite scRNAseq was not carried on for *S. typhimurium*, sorting out bacteria cells based on heterogenic host cells response showed this PhoP expression heterogeneity. In a previous population level transcriptional analysis of *S. typhimurium* response during macrophages infection with bulk RNAseq (Eriksson, Lucchini, Thompson, Rhen, & Hinton, 2003), it was not possible to determine any increased expression of PhoP, highlighting the need of scRNAseq (figure 8A and B2).

In summary, scRNAseq is a technique still under development. The studies that already used this technique show that it has the potential to characterize expression heterogeneity during host-pathogen interaction. With further improvements in sensitivity for scRNAseq technique, it would be possible to carry on scRNAseq in prokaryotic cells, expanding even more the insights of the role of transcription heterogeneity during infection.

References

Altschuler, S. J., & Wu, L. F. (2010). Cellular heterogeneity: do differences make a difference? *Cell*, *141*(4), 559–63. http://doi.org/10.1016/j.cell.2010.04.033

Arpaia, N., Godec, J., Lau, L., Sivick, K. E., McLaughlin, L. M., Jones, M. B., … Barton, G. M. (2011). TLR signaling is required for Salmonella typhimurium virulence. *Cell*, *144*(5), 675–88. http://doi.org/10.1016/j.cell.2011.01.031

Avraham, R., Haseley, N., Xavier, R. J., Regev, A., Hung, D. T., Avraham, R., … Trombetta, J. J. (2015). Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses Article. *Cell*, *162*(6), 1309–1321. http://doi.org/10.1016/j.cell.2015.08.027

Baker, A. E., Mantz, A. R., & Chiu, M. L. (2014). Raman spectroscopy characterization of antibody phases in serum. *mAbs*, *6*(6), 1509–17. http://doi.org/10.4161/19420862.2014.975100

Baker, S. C., Bauer, S. R., Beyer, R. P., Brenton, J. D., Bromley, B., Burrill, J., … Zadro, R. (2005). The External RNA Controls Consortium: a progress report. *Nature Methods*, *2*(10), 731–4. http://doi.org/10.1038/nmeth1005-731

Chattopadhyay, P. K., & Roederer, M. (2012). Cytometry: today's technology and tomorrow's horizons. *Methods (San Diego, Calif.)*, *57*(3), 251–8. http://doi.org/10.1016/j.ymeth.2012.02.009

De Jong, I. G., Haccou, P., & Kuipers, O. P. (2011). Bet hedging or not? A guide to proper classification of microbial survival strategies. *BioEssays*, *33*(3), 215–223. http://doi.org/10.1002/bies.201000127

Doyle, S. E., Vaidya, S. A., O'Connell, R., Dadgostar, H., Dempsey, P. W., Wu, T.-T., … Cheng, G. (2002). IRF3 Mediates a TLR3/TLR4-Specific Antiviral Gene Program. *Immunity*, *17*(3), 251–263. http://doi.org/10.1016/S1074-7613(02)00390-4

Eriksson, S., Lucchini, S., Thompson, A., Rhen, M., & Hinton, J. C. D. (2003). Unravelling the biology of macrophage infection by gene expression profiling of intracellular Salmonella enterica. *Molecular Microbiology*, *47*(1), 103–118. http://doi.org/10.1046/j.1365-2958.2003.03313.x

Ernst, R. K., Guina, T., & Miller, S. I. (2001). Salmonella typhimurium outer membrane remodeling: role in resistance to host innate immunity. *Microbes and Infection*, *3*(14-15), 1327–1334. http://doi.org/10.1016/S1286-4579(01)01494-0

Feuillet, V., Medjane, S., Mondor, I., Demaria, O., Pagni, P. P., Galán, J. E., … Alexopoulou, L. (2006). Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12487–92. http://doi.org/10.1073/pnas.0605200103

Francesconi, M., & Lehner, B. (2015). Reconstructing and analysing cellular states, space and time from gene expression profiles of many cells and single cells. *Molecular bioSystems*, *11*(10), 2690–8. http://doi.org/10.1039/c5mb00339c

Freudenberg, M. A., Kalis, C., Chvatchko, Y., Merlin, T., Gumenscheimer, M., & Galanos, C. (2003). Role of interferons in LPS hypersensitivity. *Journal of Endotoxin Research*, *9*(5), 308–312. http://doi.org/10.1179/096805103225002566

Fröhlich, J., & König, H. (2000). New techniques for isolation of single prokaryotic cells. *FEMS Microbiology Reviews*, *24*(5), 567–572. http://doi.org/10.1111/j.1574-6976.2000.tb00558.x

Gilchrist, J. J., MacLennan, C. A., & Hill, A. V. S. (2015). Genetic susceptibility to invasive Salmonella disease. *Nature Reviews Immunology*, *15*(7), 452–463. http://doi.org/10.1038/nri3858

Gossett, D. R., Tse, H. T. K., Lee, S. A., Ying, Y., Lindgren, A. G., Yang, O. O., … Di Carlo, D. (2012). Hydrodynamic stretching of single cells for large population mechanical phenotyping. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(20), 7630–5. http://doi.org/10.1073/pnas.1200107109

Groisman, E. A. (2001). The pleiotropic two-component regulatory system PhoP-PhoQ. *Journal of Bacteriology*, *183*(6), 1835–42. http://doi.org/10.1128/JB.183.6.1835-1842.2001

Grün, D., Kester, L., & van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nature Methods*, *11*(6), 637–40. http://doi.org/10.1038/nmeth.2930

Grün, D., & van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing Experiments. *Cell*, *163*(4), 799–810. http://doi.org/10.1016/j.cell.2015.10.039

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports*, *2*(3), 666–73. http://doi.org/10.1016/j.celrep.2012.08.003

Huang, R., Barber, T. A., Schmidt, M. A., Tompkins, R. G., Toner, M., Bianchi, D. W., … Flejter, W. L. (2008). A microfluidics approach for the isolation of nucleated red blood cells (NRBCs) from the peripheral blood of pregnant women. *Prenatal Diagnosis*, *28*(10), 892–9. http://doi.org/10.1002/pd.2079

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J. B., Lönnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, *21*(7), 1160–1167. http://doi.org/10.1101/gr.110882.110

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., … Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, *11*(2), 163–6. http://doi.org/10.1038/nmeth.2772

Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., … Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science (New York, N.Y.)*, *343*(6172), 776–9. http://doi.org/10.1126/science.1247651

Kang, Y., McMillan, I., Norris, M. H., & Hoang, T. T. (2015). Single prokaryotic cell isolation and total transcript amplification protocol for transcriptomic analysis. *Nature Protocols*, *10*(7), 974–84. http://doi.org/10.1038/nprot.2015.058

Kang, Y., Norris, M. H., Zarzycki-Siek, J., Nierman, W. C., Donachie, S. P., & Hoang, T. T. (2011). Transcript amplification from single bacterium for transcriptome analysis. *Genome Research*, *21*(6), 925–935. http://doi.org/10.1101/gr.116103.110

Kebschull, J. M., & Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing datasets Author summary. *Nucleic Acids Research*, *43*(21), e143–.

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., … Kirschner, M. W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*, *161*(5), 1187–1201. http://doi.org/10.1016/j.cell.2015.04.044

Kurn, N., Chen, P., Heath, J. D., Kopf-Sill, A., Stephens, K. M., & Wang, S. (2005). Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications. *Clinical Chemistry*, *51*(10), 1973–81. http://doi.org/10.1373/clinchem.2005.053694

Lewis, K. (2010). Persister cells. *Annual Review of Microbiology*, *64*, 357–72. http://doi.org/10.1146/annurev.micro.112408.134306

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. http://doi.org/10.1186/1471-2105-12-323

Li, M., Xu, J., Romero-Gonzalez, M., Banwart, S. A., & Huang, W. E. (2012). Single cell Raman spectroscopy for cell sorting and imaging. *Current Opinion in Biotechnology*, *23*(1), 56–63. http://doi.org/10.1016/j.copbio.2011.11.019

Li, P., Piao, Y., Shon, H. S., & Ryu, K. H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinformatics*, *16*(1), 347. http://doi.org/10.1186/s12859-015-0778-7

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Processing of rRNA and tRNA. W. H. Freeman.

Lovatt, D., Ruble, B. K., Lee, J., Dueck, H., Kim, T. K., Fisher, S., … Eberwine, J. (2014). Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature Methods*, *11*(2), 190–6. http://doi.org/10.1038/nmeth.2804

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., … McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, *161*(5), 1202–1214. http://doi.org/10.1016/j.cell.2015.05.002

Newton, K., & Dixit, V. M. (2012). Signaling in innate immunity and inflammation. *Cold Spring Harbor Perspectives in Biology*, *4*(3), a006049–. http://doi.org/10.1101/cshperspect.a006049

Pan, X., Durrett, R. E., Zhu, H., Tanaka, Y., Li, Y., Zi, X., … Weissman, S. M. (2013). Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(2), 594–9. http://doi.org/10.1073/pnas.1217322109

Patro, R., Mount, S. M., & Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnology*, *32*(5), 462–4. http://doi.org/10.1038/nbt.2862

Picelli, S., Faridani, O. R., Björklund, A. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, *9*(1), 171–81. http://doi.org/10.1038/nprot.2014.006

Proserpio, V., & Mahata, B. (2015). Single-cell technologies to study the immune system. *Immunology*, n/a–n/a. http://doi.org/10.1111/imm.12553

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., & Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, *5*(10), 877–9. http://doi.org/10.1038/nmeth.1253

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, *32*(9), 896–902. http://doi.org/10.1038/nbt.2931

Rosenow, C., Saxena, R. M., Durst, M., & Gingeras, T. R. (2001). Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. *Nucleic Acids Research*, *29*(22), E112.

Saliba, A.-E., Westermann, A. J., Gorski, S. A., & Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Research*, *42*(14), 8845–8860. http://doi.org/10.1093/nar/gku555

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T., & Ueda, H. R. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biology*, *14*(4), R31. http://doi.org/10.1186/gb-2013-14-4-r31

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., … Regev, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature*, *509*(7505), 363–9. http://doi.org/10.1038/nature13437

Stegle, O., Teichmann, S. A., & Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews. Genetics*, *16*(January 2014), 133–145. http://doi.org/10.1038/nrg3833

Tang, D., Kang, R., Coyne, C. B., Zeh, H. J., & Lotze, M. T. (2012). PAMPs and DAMPs: signal 0s that spur autophagy and immunity. *Immunological Reviews*, *249*(1), 158–75. http://doi.org/10.1111/j.1600-065X.2012.01146.x

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., … Surani, M. A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell*, *6*(5), 468–78. http://doi.org/10.1016/j.stem.2010.03.015

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., … Surani, M. A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols*, *5*(3), 516–35. http://doi.org/10.1038/nprot.2009.236

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., … Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–82. http://doi.org/10.1038/nmeth.1315

Taniguchi, K., Kajiyama, T., & Kambara, H. (2009). Quantitative analysis of gene expression in a single cell by qPCR. *Nature Methods*, *6*(7), 503–6. http://doi.org/10.1038/nmeth.1338

Trombetta, J. J., Gennert, D., Lu, D., Satija, R., Shalek, A. K., & Regev, A. (2014). Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, *107*, 4.22.1–4.22.17. http://doi.org/10.1002/0471142727.mb0422s107

Vahey, M. D., Quiros Pesudo, L., Svensson, J. P., Samson, L. D., & Voldman, J. (2013). Microfluidic genome-wide profiling of intrinsic electrical properties in Saccharomyces cerevisiae. *Lab on a Chip*, *13*(14), 2754–63. http://doi.org/10.1039/c3lc50162k

Verschoor, C. P., Lelic, A., Bramson, J. L., & Bowdish, D. M. E. (2015). An Introduction to Automated Flow Cytometry Gating Tools and Their Implementation. *Frontiers in Immunology*, *6*, 380. http://doi.org/10.3389/fimmu.2015.00380

Wang, J., Chen, L., Chen, Z., & Zhang, W. (2015). RNA-seq based transcriptomic analysis of single bacterial cells. *Integr. Biol.*, *7*(11), 1466–1476. http://doi.org/10.1039/C5IB00191A

Wang, J., Fan, H. C., Behr, B., & Quake, S. R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell*, *150*(2), 402–12. http://doi.org/10.1016/j.cell.2012.06.030

Weiss, D. S., Raupach, B., Takeda, K., Akira, S., & Zychlinsky, A. (2004). Toll-like receptors are temporally involved in host defense. *Journal of Immunology (Baltimore, Md. : 1950)*, *172*(7), 4463–9.

Zhang, W., Kai, K., Choi, D. S., Iwamoto, T., Nguyen, Y. H., Wong, H., … Qin, L. (2012). Microfluidics separation reveals the stem-cell-like deformability of tumor-initiating cells. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 18707–12. http://doi.org/10.1073/pnas.1209893109

Zong, C., So, L., Sepúlveda, L. A., Skinner, S. O., & Golding, I. (2010). Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Molecular Systems Biology*, *6*(1), 440. http://doi.org/10.1038/msb.2010.96