Video Article

A Combinatorial Single-cell Approach to Characterize the Molecular and Immunophenotypic Heterogeneity of Human Stem and Progenitor Populations

Mikael N.E. Sommarin[1], Rebecca Warfvinge[1], Fatemeh Safi[1], Göran Karlsson[1]

[1]Division of Molecular Hematology, Lund Stem Cell Center, Lund University

Correspondence to: Mikael N.E. Sommarin at mikael.sommarin@med.lu.se, Göran Karlsson at Goran.Karlsson@med.lu.se

URL: https://www.jove.com/video/57831
DOI: doi:10.3791/57831

Keywords: Genetics, Issue 140, Heterogeneity, single-cell, RT-qPCR, FACS, Index-sorting, immunophenotypic characterization, gene expression analysis

Date Published: 10/25/2018

Citation: Sommarin, M.N., Warfvinge, R., Safi, F., Karlsson, G. A Combinatorial Single-cell Approach to Characterize the Molecular and Immunophenotypic Heterogeneity of Human Stem and Progenitor Populations. *J. Vis. Exp.* (140), e57831, doi:10.3791/57831 (2018).

Abstract

Immunophenotypic characterization and molecular analysis have long been used to delineate heterogeneity and define distinct cell populations. FACS is inherently a single-cell assay, however prior to molecular analysis, the target cells are often prospectively isolated in bulk, thereby losing single-cell resolution. Single-cell gene expression analysis provides a means to understand molecular differences between individual cells in heterogeneous cell populations. In bulk cell analysis an overrepresentation of a distinct cell type results in biases and occlusions of signals from rare cells with biological importance. By utilizing FACS index sorting coupled to single-cell gene expression analysis, populations can be investigated without the loss of single-cell resolution while cells with intermediate cell surface marker expression are also captured, enabling evaluation of the relevance of continuous surface marker expression. Here, we describe an approach that combines single-cell reverse transcription quantitative PCR (RT-qPCR) and FACS index sorting to simultaneously characterize the molecular and immunophenotypic heterogeneity within cell populations.

In contrast to single-cell RNA sequencing methods, the use of qPCR with specific target amplification allows for robust measurements of low-abundance transcripts with fewer dropouts, while it is not confounded by issues related to cell-to-cell variations in read depth. Moreover, by directly index-sorting single-cells into lysis buffer this method, allows for cDNA synthesis and specific target pre-amplification to be performed in one step as well as for correlation of subsequently derived molecular signatures with cell surface marker expression. The described approach has been developed to investigate hematopoietic single-cells, but have also been used successfully on other cell types.

In conclusion, the approach described herein allows for sensitive measurement of mRNA expression for a panel of pre-selected genes with the possibility to develop protocols for subsequent prospective isolation of molecularly distinct subpopulations.

Video Link

The video component of this article can be found at https://www.jove.com/video/57831/

Introduction

Each individual blood cell is believed to reside in a cellular hierarchy, where stem cells form the apex on top of a series of increasingly committed intermediate progenitors that eventually terminally differentiate into the final effector cells carrying specific biological functions[1]. Much of the knowledge about how stem cell systems are organized has been generated in the hematopoietic system, largely because of the ability to prospectively isolate distinct hematopoietic populations highly enriched for stem cells or various progenitors[2] by FACS sorting. This has allowed for many of these populations to be analyzed functionally or molecularly, predominantly through gene expression profiling[3,4]. However when analyzing gene expression of bulk populations individual differences between cells are averaged out and lost[5]. Thus, incapacity to detect cell-to-cell variations within heterogeneous cell fractions may confound our understanding of critical biological processes if small subsets of cells account for the inferred biological function of that population[6,7]. Conversely, investigation of gene expression signatures at single-cell resolution offer a possibility to delineate heterogeneity and circumvent overshadowing influences from overrepresented subsets of cells[8].

To date many protocols for single-cell gene expression analysis have been developed; with each approach having its own caveats. The earliest method was RNA Fluorescent *in situ* hybridization (RNA-FISH), which measures a limited number of transcripts at a time but is unique in that it allows for investigation of RNA localization[9,11]. Early methods using PCR and qPCR to detect a single or very few transcripts were also developed[12]. However, these have lately been replaced by microfluidics-based methods which can simultaneously analyze the expression of hundreds of transcripts per cell in hundreds of cells through qPCR and thus allow for high-dimensional heterogeneity analysis using pre-determined gene panels[10,13]. Recently RNA sequencing-based technologies have become widely used for single cell analysis, as these theoretically can measure the entire transcriptome of a cell and thereby add an exploratory dimension to heterogeneity analysis[10,14]. Multiplexed qPCR analysis and single-cell RNA sequencing have different features, thus the rationale for using either of the methods depends on the question asked as well as the number of cells in the target population. The high-throughput and low cost per cell together with unbiased, exploratory characteristics of single-cell RNA sequencing are desirable when unknown cell or large populations are investigated. However,

single-cell RNA sequencing is also biased towards sequencing high abundant transcripts more frequently while transcripts with low abundance are prone to dropouts. This can lead to considerably complex data that puts high-demands on bioinformatic analysis to reveal important molecular signals that are often subtle or hidden in technical noise[15]. Thus, for well-characterized tissues, single-cell qPCR analysis using pre-determined primer panels selected for functionally important genes or molecular markers can serve as a sensitive, straightforward approach to determine the heterogeneity of a population. However, it should be noted that compared to single-cell RNA-seq, the cost per cell is generally higher for single-cell qPCR methods. Here, we describe an approach that combines single-cell RT-qPCR (modified from Teles J. *et al.*[16]), to FACS index sorting[17] and bioinformatics analysis[18] in order to simultaneously characterize the molecular and immunophenotypic heterogeneity within populations.

In this approach, the cell population of interest is stained, and single-cells are sorted by FACS directly into lysis buffer in 96-well PCR plates. Simultaneously, the expression levels of an additional set of cell-surface markers are recorded for each single cell during FACS-sorting, a method that is referred to as index-sorting. The lysed cell material is subsequently amplified and the gene expression of a selected set of genes analyzed with RT-qPCR, using a microfluidic platform. This strategy enables molecular analysis of the sorted single-cell as well as simultaneous characterization of each individual cell's cell-surface marker expression. By directly mapping molecularly distinct subsets of cells to the expression of the indexed sorted markers, the subpopulations can be linked to a specific immunophenotype that can be used for their prospective isolation. The method is outlined step by step in **Figure 1**. A pre-determined gene panel further contributes to a higher resolution of the targeted gene expression, since it circumvents measurement of irrelevant abundant genes that can otherwise occlude subtle gene expression signals. Moreover, the specific target amplification, one step reverse transcription, and amplification allows for robust measurement of low expressed transcripts, like transcription factors or non-poly-adenylated RNAs. Importantly, qPCR methods allow for measurement of mRNA from fusion proteins, which are important when investigating certain malignant diseases[19]. Finally, the focused number of genes investigated, low drop-out rates, and limited technical differences between cells make this method easily analyzed compared to higher dimensional methods, such as single-cell RNA-seq. By following the protocol, an entire experiment can be performed, from sorting cells to analyzed results, within three days, making this an uncomplicated and quick method for sensitive, high-throughput single-cell gene expression analysis.

Protocol

1. Preparation of Lysis Plates

1. Using a RNA/DNA free bench, prepare enough lysis buffer for 96 wells, with 10% extra, by mixing 390 µL nuclease free water, 17 µL of 10% NP-40, 2.8 µL 10 mM dNTP, 10 µL 0.1 M DTT and 5.3 µL RNAse inhibitor (see **Table of Materials**). Vortex and spin down.
2. Distribute 4 µL of lysis buffer to each well of a 96 well PCR plate and seal the plates with adhesive film. Spin down tubes to collect liquid at the bottom of the plates. Keep plates on ice until cell sorting (maximum 24 h).

2. Preparation of Cells for Cell Sorting

1. Thaw appropriate number of cells (here, CD34 enriched hematopoietic stem and progenitor cells) for the experiment. 1 x 10^6 cells are appropriate for sorting approximately three 96-well plates of single-cells with controls.
2. Transfer thawed cells to a 15 mL conical tube and add 1 mL FBS every 30 s until a total volume of 8 mL is reached. Spin cells in a centrifuge at 350 x *g* for 10 min at 4 °C and remove supernatant.
3. Resuspend cells in 8 mL staining buffer (PBS with 2% FBS) and centrifuge at 350 x *g* for 10 min at 4 °C and remove supernatant.
4. Resuspend cells in 200 µL staining buffer and remove cells for control stains.
5. Make Fluorescence minus one controls (FMOs) for each fluorophore, by staining a fraction of cells in 50 µL staining buffer. In this example, 6 microcentrifuge tubes with 20,000 cells are used as FMOSs. Note that the number of cells should be adjusted depending on the population investigated. Add all antibodies at the same concentration as in the sample stain except for one to each tube.
6. Make single stains for each fluorophore by staining a fraction of cells in 50 µL buffer for each fluorophore used. In this example 6 microcentrifuge tubes with 20,000 cells are used. Note that the target for each antibody needs to be expressed by the cells used for controls. Add each antibody at the same concentration as in the sample stain in individual tubes. Additionally keep 20,000 unstained cells in 50 µL as an unstained control.
7. To the cell sample, add antibodies at their appropriate concentration. Used here are CD34-FITC at a 1/100 concentration, CD38-APC 1/50, CD90-PE 1/10, CD45RA-bv421 1/50, CD49F-PECy7 1/50 and Lineage Mix: CD3-PECy5 1/50, CD2-PECy5 1/50, CD19-PECy5 1/50, CD56-PECy5 1/50, CD123-PECy5 1/50, CD14-PECy5 1/50, CD16-PECy5 1/50, and CD235a-PECy5 1/1000.
8. Incubate cells with antibodies for 30 min on ice in the dark.
9. Wash cells with 3 mL staining buffer. Centrifuge cells at 350 x *g* for 10 min at 4 °C and remove supernatant.
10. Resuspend cells and repeat step 2.9.
11. Resuspend sample in 500 µL and FMOs in 100 µL staining buffer with 1/100 7AAD and filter cells through a 50 µm filter to get a single-cell suspension.

3. Cell Sorting

1. Make sure that the FACS machine is set up correctly with drop delay and cytometer setup and tracking (CST) that have recently been performed according to manufacturer's instructions, to ensure that the appropriate cells are sorted. For hematopoietic cells, the use of the 85 micron nozzle and maximum speed of 4 is recommended, while the optimal event rate is between 800 and 2000 events/s.
2. Correct for spectral overlap by performing fluorescence compensation and set gates according to FMO controls or internal negative controls.
3. Perform reanalysis of the target population by sorting at least 100 target cells into a new microcentrifuge tube with 100 µL of stain buffer. FACS analyze the sorted cells by recording the sorted sample and make sure that they end up in the sort gate.
4. Set-up single cell plate sorting by centering the drop in well A1 in a 96 well plate. When it is centered, sort 50–100 6 µm particles into all wells around the edge of an empty 96 well plate to ensure that all wells will get a cell in the center of each well.

5. If possible, an additional control to ensure that viable cells are sorted can be added by sorting single-cells for *in vitro* growth. Haematopoietic cells can be grown in U-bottom 96 well plates in 100 µl SFEM with 1% penicillin streptomycin, 100 ng/mL FLT3L, TPO and SCF. Analyze each well after 3 days in culture for cell colonies using a microscope.
6. Remove adhesive film from plates. Sort a single cell of interest (here Lin-CD34+CD38- cells) into 92 out of the 96 wells, activate INDEX-sorting in the FACS sorting software to save the immunophenotypic profile for other markers of interest (here CD45RA, CD49f, and CD90) for each single cell.
7. Sort two wells with 10 and 20 cells respectively for linearity controls in the PCR amplification. Wells H1 and H2 are usually used.
8. Keep two wells without any cells as no-template controls, usually wells H3 and H4.
9. Seal the plates with clear adhesive film and spin the plates at 300 x *g* for 1 min before snap freezing on dry ice.
10. Store frozen plates at -80 °C.
    NOTE: Safe stopping point. Sorted and lysed cells can be kept at -80 °C for long-term storage.

4. Reverse Transcription and Specific Target Amplification

1. Prepare the primer mix for all 96 gene targets by adding 2 µL of each primer pair, including housekeeping gene primers and primers for spiked-in control RNA, in a 1.5 mL RNAse free tube on a RNA/DNA free bench. If using less than 96 primers, add an equivalent volume of nuclease free water for the missing primers. Primers are ordered separately to match the desired gene panel.
2. Make reverse transcription and specific target amplification mix by adding 632.5 µL 2x reaction mix, 101.2 µL Taq/SuperscriptIII, 151.8 µL Primer mix, and 0.7 µL spiked in control RNA. Preform this step on a DNA free bench. Mix by vortexing and spin down to collect the liquid in the bottom of the tube. Keep on ice until addition to the sample.
3. Make no-reverse transcription control mix for four wells by mixing 27.5 µL of 2x reaction mix, 1.76 µL of Taq enzyme, 6.6 µL of primer mix and 2.64 µL of nuclease free water. Spike in control RNA. Perform this step on a DNA free bench. Vortex and spin down to collect the liquid in the bottom of the tube and keep on ice until addition to sample.
4. Thaw lysate plates on ice. Add 8.75 µL of the previously prepared reverse transcription and specific target amplification mix to 92 wells, including the linearity and no-template controls. Add 8.75 µL of no-reverse transcription control mix to the four remaining wells. Seal plates with clear adhesive film and spin down to collect liquid at the bottom of the plates.
5. Preform reverse transcription and specific target amplification by running the plate in a PCR machine according to the preamp program; step 1: 50 °C for 60 min, step 2: 95 °C for 2 min, step 3: 95 °C for 15 s, step 4: 60 °C for 4 min, repeat steps 3–4 24 times and finally step 5: 8 °C forever.
6. After PCR is complete, keep the plate at 8 °C for short term storage and -20 °C for long term storage.
    NOTE: Safe stopping point. Amplified material can be kept at 8 °C for short-term storage and at -20 °C for long-term storage.

5. Preparation of Sample and Assay Plates for Multiplex Microfluidic Gene Expression Analysis

1. Prepare assay loading plate by pipetting 3 µL Assay loading reagent to each well of a 96 well plate. Add 3 µL of each primer to individual wells in the assay loading plate.
2. Seal plate with adhesive film and spin down to collect liquid at the bottom of the plates.
3. Prepare dilution plate by pipetting 8 µL of nuclease free water into all wells of a 96 well plate. Add 2 µL of amplified sample to the dilution plate, making a final dilution of 1:5.
4. Seal plate with adhesive film, mix by vortexing plate for 10 s, and finally spin down to collect liquid at the bottom of the plates.
5. Prepare sample loading mix by carefully mixing 352 µL of master mix with 35.2 µL of sample loading reagent. Prepare sample loading plate by aliquoting 3.3 µL of loading mix to each well of a 96 well plate.
6. Add 2.7 µL from the diluted sample into each well of the sample loading plate.
7. Seal plate with adhesive film and spin down to collect liquid at the bottom of the plates.

6. Loading of Microfluidic Chip

1. Take out a new 96 x 96 microfluidic chip. Prepare inlets by poking them with a syringe with cap on to make sure that they can be moved.
2. Remove bubbles from syringes. Add full volume of syringes to each valve while tilting the chip 45 degrees and pressing down the valve. Prime chip with the IFC controller.
3. Load each assay inlet with 4.25 µL from each of the wells in the assay loading plate. Avoid bubbles. If bubbles appear in the well, remove them with a pipette tip.
4. Continue loading each sample inlet with 4.25 µL from each of the wells in the sample loading plate, avoid bubbles, and if bubbles appear, remove them with pipette tip.
5. Load chip with the IFC controller.
6. Check that the chip looks even and that all chambers have been loaded. Remove dust from the chip surface by touching it with tape. Run the chip in the multiplex microfluidic gene expression platform.

7. Running Chip on Multiplex Microfluidic Gene Expression Platform

1. After loading chip into the multiplex microfluidic gene expression platform, name the sample.
2. Set ROX as passive dye. Set single probe and FAM-MGB as fluorescence. Use 96 x 96 standard v2 as protocol. Start run.
3. Remove chip when run is complete.

8. Preliminary Analysis of Chip Run

1. Load data into Real-Time PCR analysis software.
2. Load gene names and cell names by pasting cell and gene layouts from a tab delimited file.
3. Open image view and select ROX as dye. Check if all wells have ROX passive dye.
4. Investigate if all amplification plots look ok, with a smooth amplification curve with no spikes (similar to **Figure 3E**).
5. Ensure that all single-cells have expression of spiked-in control RNA to make sure that all have been loaded properly.
6. Ensure that all cells have housekeeping gene expression and thus have been sorted properly.
7. Ensure that 10 and 20 cell linearity controls have approximately 1 CT difference to validate linear amplification.
8. Check if there is expression in the noRT control samples. If expression is detected in noRT consider changing probes to probes which do not detect genomic DNA for subsequent runs.
9. Export data in csv files for further analysis.

9. Single-cell Analysis Using SCExV

NOTE: An introductory film is present[20] to introduce the tool. Here, a short recommendation of how to do analysis using the controls introduced in the protocol is presented.

1. Connect to the SCexV website[20].
2. Upload exported CSV files.
3. Chose the spiked-in control RNA as positive control. Remove any cell which has a control RNA CT above 25. Normalize the data to median expression of control RNA.
4. Click "Done here -> Analyze". Remove noRT, notemplate, 10 and 20 cell controls in the exclude cell option.
5. Cluster cells with the clustering approach of choice with the expected number of clusters. Export analyzed values.

10. Index-sorting Analysis

1. Open FACS analysis software and load the indexed samples.
2. Open the script editor and run the script available from Quinn J. *et al.*[21]
3. Now that the FACS analysis software should have made a gate for each of the single cells, open layout editor and color the cells according to the grouping from SCexV (available in the file named "Sample_complete_Data.xls").

Representative Results

The protocol described is quick, easily performed and highly reliable. An overview of the experimental set-up is presented in **Figure 1**. The entire protocol, from sorting of single-cells, to specific target amplification, gene expression measurements and preliminary analysis can be performed in three days. An example of analyzed results in the form of a heat map that represents preliminary analyzed data from single-cell gene expression analysis using 96 primers and 96 cells from either a chronic myeloid leukemia (CML) patient or 96 cells from an aged-matched healthy control is presented in **Figure 2**. Using hierarchical clustering, the analyzed cells can be divided into four subgroups based on their gene expression signatures. The heat map visualization is a convenient way to get an overview of the data as well as to control for wells that should be excluded from the analysis (*e.g.,* control wells). **Figure 2B** is a principal component analysis (PCA) visualizing how similar the cells of each group are to each other using dimension reduction. Here, outliers are easily distinguished from the rest of the cells. To analyze how individual genes are distributed among the clusters as well as differ between them, violin plots are useful. In **Figure 2C**, the expression of four representative genes are shown: the fusion gene BCR-ABL1, which marks all leukemic cells; the cell cycle marker mKI67, which is expressed in an actively dividing group; the myeloid differentiation marker MPO, which is restricted to the cycling subgroup; and finally the house keeping gene RPS18, which is expressed in all cells. Finally, in **Figure 2D** the molecular signatures have been correlated to immunophenotyped, as the identities of each cell in each cluster are used to color the individual cells in a FACS plot generated from the index-sorting. Shown is the cell surface marker expression for the hematopoietic stem cell markers CD90 and CD45RA that in this case could be used to discriminate between some of the clusters and prospectively isolate them for functional analysis.

The microfluidics used to perform the single-cell qPCR is very sensitive to introduction of air or particles into the system. Therefore, it is critical to do a post run quality check of the data. **Figure 3** shows representative data and the contrast between a successful and a failed run due to poor sample loading. **Figure 3A** shows how the Rox levels after a successful run spread evenly over all wells reassuring that the loading has been successful. In contrast, **Figure 3B** illustrates how sample and assay loading have failed as indicated by the lack of ROX in large parts of the chip. Once the quality of Rox has been assessed, the quality of the raw gene expression signals can be evaluated. In **Figure 3C**, a heat map of a successful run is presented, where expression is evenly spread out across samples with strong signal of around 7-25 Ct:s. In contrast, **Figure 3D** shows a heat map of a failed run where many samples have no signal or expression of a very limited amount of genes. The results in **Figure 3D** are likely due to an error in the FACS sorting prior to single-cell gene expression analysis, since many cells have clear expression signals while others lack expression completely. Finally, **Figure 3E** displays the expected amplification curve of spiked-in control RNA with all wells having clear expression of around 10 CT with little to no variation. This positive control measures the efficiency of all steps in the single-cell qPCR analysis. To ensure that amplification is within dynamic range, different cell numbers are included as linearity controls; here, 10- and 20-cell controls are used. The CT differences between linearity controls should reflect the differences in cell numbers. Finally, the no reverse transcription (noRT) and the no template negative controls ensure that no false positive signals are detected from genomic DNA or contaminations, respectively.

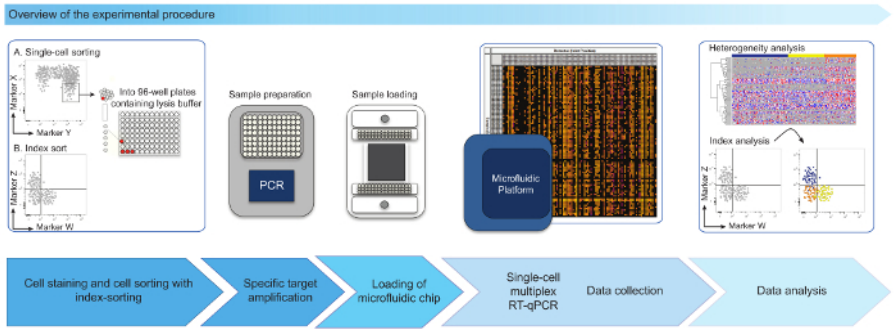

Figure 1: Schematic view of the protocol. Cells are stained, sorted with index-sort application activated to record surface marker expression, and lysed to release the mRNA. mRNA is subsequently reverse transcribed to cDNA and amplified using specific target amplification. Next, the samples are loaded together with individual primers into a microfluidic chip, where the gene expression profile of each cell is analyzed using RT-qPCR. After data collection, the data is pre-processed to remove low-quality cells and clustering is performed to define molecularly distinct cell populations. Finally, molecularly defined subpopulations are correlated to surface marker expression from FACS index sorting data to immunophenotypically characterize the heterogeneity of the populations. Please click here to view a larger version of this figure.

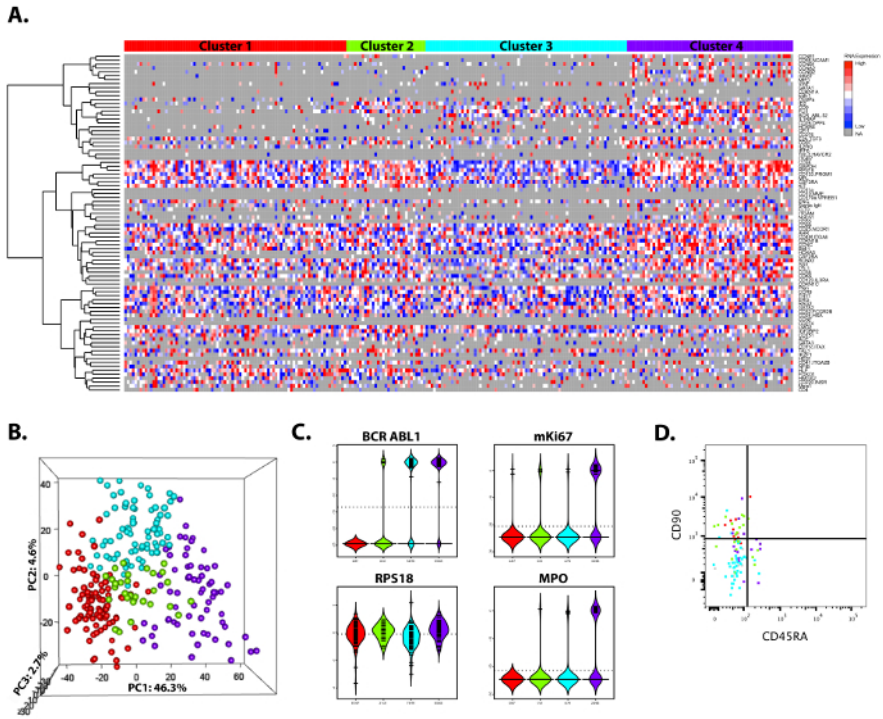

Figure 2: Representative results of three successful single-cell gene expression analyzes of normal and leukemic hematopoietic stem cells. (**A**) Heat map showing single-cell gene expression measurements of 270 single cells from one chronic myeloid leukemia patient as well as an age-matched healthy control analyzed by hierarchical clustering using SCExV [18], an analysis tool designed for this type of data. Red represents high expression, blue low expression and grey no expression. The cells were divided into four different clusters based on their gene expression signatures. It is clear from the data that the leukemic stem cell population from this patient can be divided in one quiescent population (cluster 3), with expression of only a few differentiation markers, and one actively dividing population that has initiated expression of genes associated with myeloid differentiation (cluster 4). The data used here is from a selected patient included in previously published work by Warfvinge *et al.*[22] (**B**) Principal component analysis (PCA) of the data displayed in A, where four separated groups can be shown. (**C**) Violin plots of four genes with cluster specific expression, BCR-ABL1, mKI67, RPS18 and MPO. Each violin plot represents the expression in each cluster and the dotted line across represents the mean value of expression for all cells. (**D**) Index-sorting analysis of one of the patient samples represented in the heat map, where each individual cell is marked with the color of its molecular subpopulation. Please click here to view a larger version of this figure.

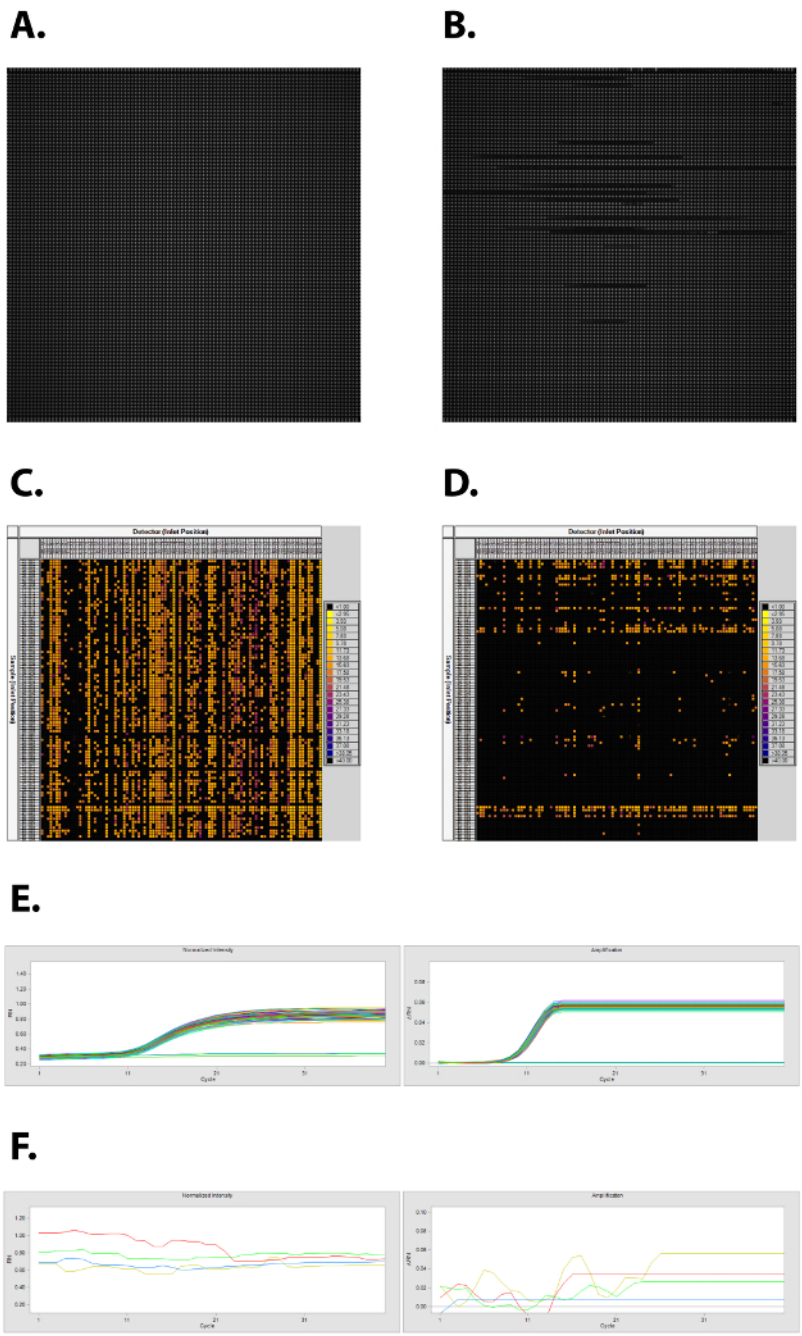

Figure 3: Quality controls of multiplex qPCR in real-time PCR software. (**A**) Rox loading image of successful sample and assay loading. (**B**) Rox loading image of unsuccessful sample and assay loading, where failed loadings of samples are shown by a lack of ROX in horizontal lines. Failed loading of assays would be shown in vertical lines. (**C**) Example of a successful run in Heat map view, where single-cells represent rows and genes represent columns. High gene expression is indicated in yellow, low expression in blue and no detected expression in black. (**D**) Example of a failed run in Heat map view, where single-cells represent rows and genes represent columns. High gene expression is represented by yellow, low expression is represented by blue, and no detected expression as black. (**E**) Normalized Intensity (left) and amplification plot (right) of a spiked-in RNA control. Ct-values of approximately 10 and curves with limited variation indicate successful and robust reverse transcription, pre-amplification, and qPCR analysis within all wells. (**F**) Normalized intensity (left) and amplification plot (right) of a failed gene. The normalized intensities do not form a curve and the amplification plot shows large spikes in fluorescence. Please click here to view a larger version of this figure.

Discussion

In recent years, single-cell gene expression analysis has become a valuable addition to define the heterogeneity of various cell populations[23]. The advent of RNA sequencing technologies theoretically provides a possibility to measure the entire transcriptome of a cell, however these methods are complicated by variations in cell-to-cell sequencing depths and drop-outs. Single-cell qPCR offers a sensitive and robust analysis

of the expression of hundreds of critical genes where all cells are treated similarly, reducing technical noise. The focused analysis of a limited number of transcripts additionally allows for simplified analysis without interference from highly expressed genes.

Since the approach only investigates a subset of genes expressed in the cells it is highly important to choose the correct set of genes when designing the gene panel. The choice of genes can be done in several ways; through literature, previous findings by bulk analysis, and bioinformatics analysis of publicly available data. This is not trivial; however, with the correct set of genes the approach can be very powerful. When the gene panel have been designed, it is important to validate that the primers are not interfering during multiplexing. Here we recommend analyzing if the primers are linear by preforming a dilution series with ample material; a dilution series of primers used in this study is shown in supplemental figure 2D in Warfvinge *et al.*[22]

The two main limitations of the protocol presented here are the limited number of cells and genes investigated. However, even though only 96 cells are investigated in each run, this relatively simple protocol can be completed in one day and thus hundreds of cells can be analyzed in one week. Additionally, the number of genes investigated could be increased if more primers are included in the target-specific pre-amplification step. In this case, several chips are needed to investigate each set of 96 primers.

If highly expressed genes or cells with high RNA-content are being analyzed, the number of cycles optimized for hematopoietic stem cells with low overall gene expression[23] should be modified. Changing amplification cycles is also important when investigating bulk populations. For bulk analysis of hematopoietic stem cells (100 cells) we recommend reducing the amount of pre-amplification cycles from 25 to 18.

The most common reasons for failed runs is due to unsuccessful sorting of single-cells into the well of the lysis plate, sorting of non-viable cells, suboptimal pre-amplification, or chip loading failure. The Rox level in each reaction chamber is an indicator of the loading of either sample or assay, and if these are low, it is recommended that the samples are re-run on a new chip, since low Rox levels are likely caused by introduction of bubbles during loading. Lack of spiked-in control RNA expression is an indicator of that the pre-amps have failed, likely due to issues with pre-amplification- or reverse transcription reagents. In this case, it is recommended that a new chip is run with newly sorted cells and new reactants. Detection of control RNA but no indication of other gene expression signals is an indication of unsuccessful single-cell FACS sorting prior to gene expression analysis. To avoid errors during FACS sorting it is important to ensure that the single-cells are sorted into the center of each well. This can be done by sorting beads into an empty plate and visually inspecting where the droplet lands. Make sure to test-sort into all wells in the edges of the plate to ensure that the set-up is not only correct for the first few wells. When applicable, single-cells can in parallel be sorted for *in vitro* growth and analyzed prior to qPCR analysis to control for successful sorting protocols. Index-sorting information does not only allow for investigating the immunophenotyped of a molecularly distinct subpopulation but can also provide useful insights when troubleshooting, if you are sorting an undefined population and many low-quality cells are present in the single-cell gene expression analysis. The index-sorting information can be used to optimize the gating strategy and avoid future sorting of low quality cells.

Single-cell gene expression analysis is revolutionizing how heterogeneous cell populations are investigated and paving the way for further understanding of cell differentiation. In haematopoiesis, single-cell gene expression analysis have been used to refine sorting strategies for subsets of HSCs[7], to resolve the heterogeneity of multipotent progenitor populations[6,24,25], and to identify therapy insensitive leukemic stem cell populations[22,26]. Combination of index sorting with both single-cell gene expression- and functional analysis has previously been shown to be reliable and efficient when investigating immunophenotype and heterogeneity of different populations[27,28,29]. Here, an approach for investigation of molecular and immunophenotypic heterogeneity of cell populations have been described where combining single-cell qPCR with index sorting enables fast and reproducible results without complicated analysis in a short timeframe.

Disclosures

The authors have nothing to disclose.

Acknowledgements

This work is supported by grants from the Swedish Cancer Society, The Swedish Research Council, The Swedish Society for Medical Research, The Swedish Childhood Cancer Foundation, The Ragnar Söderberg Foundation, and The Knut and Alice Wallenberg Foundation

References

1. Seita, J., & Weissman, I. L. Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med.* **2** (6), 640-653 (2010).
2. Orkin, S. H., & Zon, L. I. Hematopoiesis: An evolving paradigm for stem cell biology. *Cell.* **132** (4), 631-644 (2008).
3. Ye, F., Huang, W., & Guo, G. Studying hematopoiesis using single-cell technologies. *Journal of Hematology & Oncology.* **10** 27 (2017).
4. Hoppe, P. S., Coutu, D. L., & Schroeder, T. Single-cell technologies sharpen up mammalian stem cell research. *Nature Cell Biology.* **16** (10), 919-927 (2014).
5. Wills, Q. F. *et al.* Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nature Biotechnol.* **31** (8), 748-752 (2013).
6. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol.* **19** (4), 271-281 (2017).
7. Wilson, N. K., Nicola, K. *et al.* Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell.* **16** (6), 712-724 (2015).
8. Saliba, A. E., Westermann, A. J., Gorski, S. A., & Vogel, J. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Research.* **42** (14), 8845-8860 (2014).
9. Femino, A. M., Fay, F. S., Fogarty, K., & Singer, R. H. Visualization of Single RNA Transcripts in Situ. *Science.* **280** (5363), 585-590 (1998).
10. Kalisky, T. *et al.* A brief review of single-cell transcriptomic technologies. *Briefings in Functional Genomics.* elx019-elx019 (2017).
11. Crosetto, N., Bienko, M., & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nature Reviews Genetics.* **16** 57 (2014).

12. Bengtsson, M., Ståhlberg, A., Rorsman, P., & Kubista, M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research.* **15** (10), 1388-1392 (2005).
13. Bengtsson, M., Hemberg, M., Rorsman, P., & Ståhlberg, A. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Molecular Biology.* **9** (1), 63 (2008).
14. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods.* **10** 1096 (2013).
15. Tung, P.-Y. *et al.* Batch effects and the effective design of single-cell gene expression studies. *Scientific Reports.* **7** 39921 (2017).
16. Teles, J., Enver, T., & Pina, C. Single-cell PCR profiling of gene expression in hematopoiesis. *Methods in Molecular Biology.* **1185** 21-42 (2014).
17. Hayashi, T. *et al.* Single-cell gene profiling of planarian stem cells using fluorescent activated cell sorting and its "index sorting" function for stem cell research. *Development, Growth, & Differentiation.* **52** (1), 131-144 (2010).
18. Lang, S. *et al.* SCExV: A webtool for the analysis and visualisation of single cell qRT-PCR data. *BMC Bioinformatics.* **16** (1), 320 (2015).
19. de Klein, A. *et al.* A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature.* **300** (5894), 765-767 (1982).
20. Lang, S. *SCexV.*, <http://stemsysbio.bmc.lu.se/SCexV/> (2015).
21. Quinn, J. *indexed-sorting.* <https://github.com/FlowJo-LLC/indexed-sorting> (2016).
22. Warfvinge, R. *et al.* Single-cell molecular analysis defines therapy response and immunophenotype of stem cell subpopulations in CML. *Blood.* **129** (17), 2384-2394 (2017).
23. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood.* **128** (8), e20-e31 (2016).
24. Breton, G. *et al.* Human dendritic cells (DCs) are derived from distinct circulating precursors that are precommitted to become CD1c+ or CD141+ DCs. *The Journal of Experimental Medicine.* **213** (13), 2861-2870 (2016).
25. Alberti-Servera, L. *et al.* Single-cell RNA sequencing reveals developmental heterogeneity among early lymphoid progenitors. *The EMBO Journal.* **36** (24), 3619-3633 (2017).
26. Giustacchini, A. *et al.* Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nature Medicine.* **23** 692 (2017).
27. Hansmann, L., Han, A., Penter, L., Liedtke, M., & Davis, M. M. Clonal expansion and interrelatedness of distinct B-lineage compartments in multiple myeloma bone marrow. *Cancer Immunology Research.* **5** (9), 744-754 (2017).
28. Psaila, B. *et al.* Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biology.* **17** (1), 83 (2016).
29. Schulte, R. *et al.* Index sorting resolves heterogeneous murine hematopoietic stem cell populations. *Experimental Hematology.* **43** (9), 803-811 (2015).

