# Targeted Single-Cell Transcription Factor Profiling Outperforms mRNA-seq for In-Depth Characterization of Cancer Cells

Camila Egidio, Robert Durruthy-Durruthy, Manisha Ray, Jason McKinney
Fluidigm Corporation

AACR 2017 LB-076

## Introduction

Single-cell mRNA sequencing (mRNA-seq) has enabled scientists to gain new insight into the diversity of cell populations. However, only approximately 20% of all transcripts in a single cell are detected using such methodology, with detection failure highest in low-expressed genes.[1]

Transcription factors (TF) are an example of a gene class that typically is found to be low-expressed in most human cells[2] and yet has a critical role in gene regulation and cell signaling. For accurate detection and reliable quantitation of such low-expressed genes, alternative techniques that provide higher sensitivity than mRNA-seq are required.

Here we sought to systematically investigate the differences in gene expression sensitivity between a targeted gene expression profiling approach using high-throughput qPCR readout and a nontargeted approach using mRNA sequencing. Quantitation of endogenously expressed genes as well as synthetic control genes of varying concentrations across four distinct cell types demonstrates that qPCR outperforms mRNA-seq. Particularly low-expressed transcripts are exclusively detected when using the targeted approach, with implications for accurate cell type classification and correct gene regulatory identification.

## Methods and Materials

Cell suspensions from a human induced pluripotent stem cell line (iPSC) and from lung (A549), breast (SKBR3) and blood (K562) cancer cell lines were processed independently on the C1™ system (Fluidigm, Figure 1A). Two integrated fluidic circuits (IFCs) were run for each cell line, one using STA chemistry (Ambion® Single Cell-to-CT™, Thermo Scientific™) and another using poly-A mRNA detection chemistry (SMART-Seq™ v1, Clontech®) (Figure 1B). ERCC RNA Spike-In Mix (Thermo Scientific) was used in each reaction chamber as control.

Two primer panels were used in the STA workflow: one including 92 assays that target ERCC RNA controls and one including 93 assays targeting TF and housekeeping targets. Primers were designed by the custom assay design service at Fluidigm (www.fluidigm.com/assays). The STA workflow on the C1 IFC was followed by qPCR using the Biomark™ system and both primer panels (Figure 1). Poly-A amplified cDNA was subject to library preparation and sequenced with an average 2 million 75 bp paired-end reads per single cell (Figure 1).

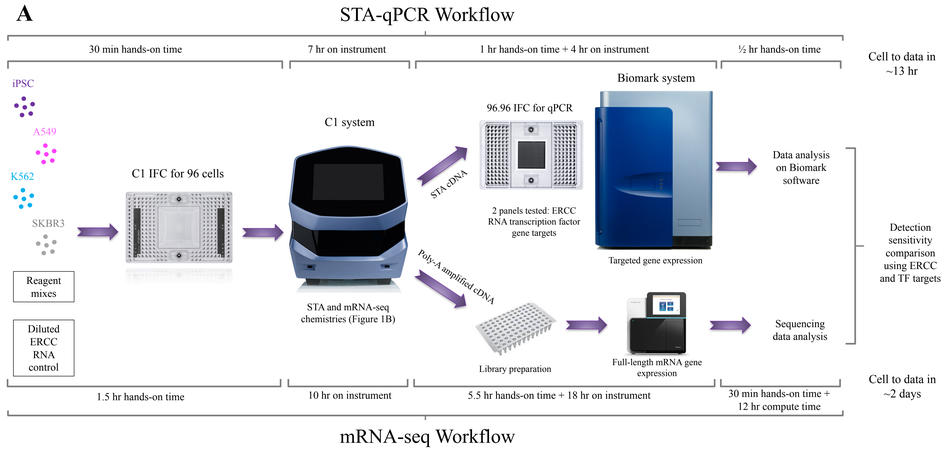
## Conclusions

- Specific target amplification (STA) plus qPCR has more than 4x increased sensitivity over mRNA-seq when detecting low-to-mid-expressing genes in single cells as measured with ERCC RNA controls.
- Accurate and quantitative detection of expression levels for a curated set (93) of transcription factors enables successful classification of distinct cell types when using STA-qPCR. mRNA-seq fails to correctly categorize different cell types based on the same curated set of TFs, likely due to a high rate of stochastic dropouts.
- Low-expressed genes may be critical to decipher molecular mechanisms in biological processes and can be effectively used to characterize and classify different cell subtypes.
- STA-qPCR has lower cost and faster turnaround time to data than mRNA-seq. Its use should be considered as a complement or alternative to mRNA-seq, depending on the context of the biological question in the study.

## References

1. Shalek et al. *Nature*, 2014.
2. Vaquerizas et al. *Nature Genetics Reviews*, 2009.

## Experimental Workflow



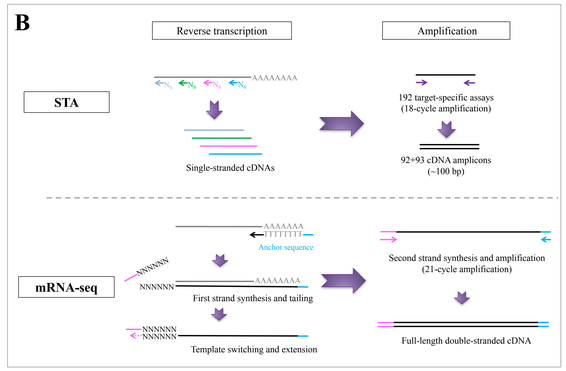




**Figure 1. Overview of the experimental workflow performed to compare the efficiency of detecting mRNA in single-cell reactions using qPCR versus mRNA-seq chemistries.** A) Both experiments were performed in parallel using identical cell suspension aliquots. The targeted approach used qPCR as readout while the mRNA-seq approach used sequencing. The targeted approach generates data in approximately 13 hours while the mRNA-seq approach takes at least two days from start to end. B) Differences during reverse transcription and cDNA amplification between STA and mRNA-seq chemistries for single-cell processing on the C1 system. All reactions are automated and performed in nanoliter reaction volumes inside a C1 IFC. After completion, the cDNA targets are harvested from the C1 IFC and processed as shown in panel A.

## Results

### qPCR ERCC RNA Detection Rates Are Over 4x Higher Compared to Sequencing Data

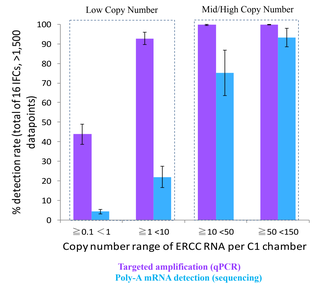



**Figure 2. ERCC RNA detection comparison between qPCR and sequencing data.** For each dataset, the percentage of reactions in which each ERCC RNA molecule was successfully detected was calculated. Input number of individual ERCC molecules ranges from 0.1 to 150 copies per reaction. 92 ERCC were binned according to absolute copy number/reaction and categorized as low copy number or mid/high copy number. ERCC RNA signal was detected in over 90% of the reactions tested when only 1 to 10 molecules were present in the reaction chambers as shown by the qPCR data (purple bars). Comparatively, sequencing data performance for the same ERCC RNA input is 20% (blue bars).

### Expression Levels of Transcription Factor Genes Obtained From qPCR Data Can Accurately Reconstruct Cell Types


**Figure 3. Hierarchical clustering of qPCR and mRNA-seq data.** A) Expression levels of 89 TFs and 4 housekeeping genes across 610 cells are shown in heat map format (columns = samples, rows = genes). 323 mRNA-seq-associated cells (blue bar, chemistry) cluster separately from 287 qPCR-associated cells (purple bar, chemistry). The increase in sensitivity observed in the detection of ERCC RNA targets in the qPCR versus the sequencing dataset is also applicable to the TF targets. The sequencing data showed a higher number of stochastic dropouts as compared to the qPCR data, which affected the reconstruction of the cells tested into their correct groups.

B) Expression levels of 100 most variable genes (based on first three principal components) across 323 mRNA-seq-associated cells. When independently analyzed based on this gene set, the mRNA-seq data successfully reconstructs the cell types based on their gene expression levels. This confirms that the sequencing data generated was able to detect the biological diversity of the different cell types tested, as expected. However, it is limited by genes that are detected as expressed at high levels. Not surprisingly, there was no TF gene included in this top 100 most variable gene set identified.

### Stochastic Gene Expression Dropout Affects Housekeeping and TF Targets Affecting Biological Conclusions

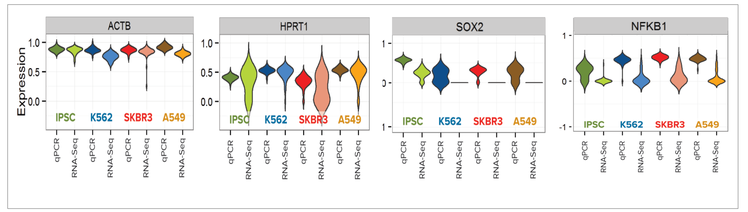



**Figure 4. Violin plots showing dropout of expression levels in the sequencing dataset (RNA-seq) for the following targets: housekeeping HPRT1, in all cell types; SOX2, pluripotent stem cell marker, in iPSC; NFKB1, molecular marker for cancer, in K562, SKBR3 and A549.** ACTB is an example of a housekeeping gene in which consistent expression is detected across both qPCR and RNA-Seq datasets since it is generally expressed at high levels.

Fluidigm Corporation
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080 USA
+1 650 266 6100 • Toll-free: 866 359 4354 in the US and Canada
fluidigm.com