Apoptotic Cell Exclusion and Bias-Free Single-Cell Selection Are Important Quality Control Requirements for Successful Single-Cell Sequencing Applications

Diana Ordoñez-Rueda,[1†] Bianka Baying,[2†] Dinko Pavlinic,[2†] Luca Alessandri,[3†] Yvonne Yeboah,[1] Jonathan J.M. Landry,[2*] Raffaele Calogero,[3*] Vladimir Benes,[2*] Malte Paulsen[1*]

[1]Flow Cytometry Core Facility, European Molecular Biology Laboratory, Heidelberg, Germany

[2]Genomics Core Facility, European Molecular Biology Laboratory, Heidelberg, Germany

[3]Bioinformatics and Genomics Unit, MBC Molecular Biotechnology Center, Torino, Italy

*Correspondence to: Jonathan J.M. Landry, Genomics Core Facility, European Molecular Biology Laboratory, Heidelberg, Germany. Email: landry@embl.de; Raffaele Calogero, Bioinformatics and Genomics Unit, MBC Molecular Biotechnology Center, Torino, Italy. Email: raffaele.calogero@unito.it; Vladimir Benes, Genomics Core Facility, European Molecular Biology Laboratory, Heidelberg, Germany. Email: benes@embl.de; Malte Paulsen, Flow Cytometry Core Facility, European Molecular Biology Laboratory, Heidelberg, Germany. Email: malte.paulsen@embl.de

†These authors contributed equally to the study.

ISAC
INTERNATIONAL SOCIETY FOR ADVANCEMENT OF CYTOMETRY

● **Abstract**

Single-cell sequencing experiments are a new mainstay in biology and have been advancing science especially in the biomedical field. The high pressure to integrate the technology into daily laboratory live requires solid knowledge with respect to potential limitations and precautions to be taken care of before applying it to complex research questions. In the past, we have identified two issues with quality measures neglected by the growing community involving SmartSeq and droplet or micro-well-based scRNASeq methods (1) how to ensure that only single cells are introduced without biasing on light scattering when handling complex cell mixtures and organ preparations or (2) how best to control for (pro-)apoptotic cell contaminations in single-cell sequencing approaches. Sighting of concurrent literature involving single-cell sequencing technologies revealed that these topics are generally neglected or simply approached in silico but not at the bench before generating single-cell data sets. We fear that those important quality aspects are overlooked due to reduced awareness of their importance for guaranteeing the quality of experiments. In this Cytometry rigor issue, we provide experimentally supported guidance on how to circumvent those critical shortcomings in order to promote a better use of the fantastic single-cell sequencing toolbox in biology. © 2019 The Authors. *Cytometry Part A* published by Wiley Periodicals, Inc. on behalf of International Society for Advancement of Cytometry.

● **Key terms**

single-cell Sequencing; cell sorting; apoptosis; in silico analysis; 10×Genomics; SmartSeq2; quality controls

T advent of single-cell sequencing technologies created a very productive and collaborative environment for genomics and flow cytometric methods in the last 5 years (1). Flow Cytometry by nature was in the past the best method to analyze quickly thousands, if not millions, of cells for their expression of proteins, peptides, single, or few mRNAs, or analyzing the metabolic state at single-cell resolution. Sequencing technologies have closed the gaps improving the quality and general robustness of sequencing DNA and RNA from single cells (2). This finally led to the development of multiple different, easy-to-use approaches in single-cell generation and subsequent sequencing methods. Single-cell sorting by fluorescence assisted cell sorting (FACS) and/or Fluidigm C1 capture paired with sensitive library preparation methods paved the way (3,4). Nowadays, complete kit-solutions from 10×Genomics and their likes have rendered the initial resistance—or better put the initial need for intense technical skill—to enter this method field to almost zero. Automation of analysis processes (5,6) and almost kit-like scientific methods are probably the biggest driver for the high-pace science that we are currently enjoying (7–9). As core facilities, we are very avid readers of papers and posters using current methods,

which we are also supporting in our laboratories and we see a steep increase in studies that involve to a significant level single-cell sequencing data. What strikes us is the apparent lag using known and standardized quality measures in the experiments at the level of the most basic quality features, like (1) clearly motivated gating strategies when using FACS to only introduce true and (2) fully viable single cells into downstream scRNAseq studies. The frequent lack in simple quality assurances is slowly becoming a source of argument with our own user who select studies that have questionable methodology with respect to cell selection criteria for single-cell sequencing assays and indicate that we are too critical while suggesting use of proper settings, controls or additional quality features in their experimental design.

We contacted authors of papers or approached researchers at their posters when their experimental strategy was missing quality assurance steps, such as guaranteeing that only single cells or truly healthy cells were used in their study. It is generally very confusing to see the large number of publications based on single-cell sorting of complex tissues for sequencing purposes, which come without respective cytometry data or at least Supporting Information figures displaying the results and gating strategies of the corresponding FACS sorts. Such omissions deprive the reader of the chance to evaluate the quality of the material used and reduce the ease to reproduce faithfully interesting findings within our laboratories, because essential information is not transparently shared. Some researchers we contacted were aware of their lapse in designing or reporting their methodology within single-cell study, but tried to argue that their clear gaps in controlling their experiments technically was intentional—almost selling it as feature of their study and not as an omission of long standing gold standards in the respective method. It was apparent that a fear of biasing the input of cells by gating for cells and singlets from tissue preparations based on FSC/SSC signals during cell sorts and the question of how and when to best assess cell health during the single-cell RNAseq experimental workflow was controversial. As core facilities offering access to scRNAseq technology, we are worried that ambiguity in method rigor during the wet-lab quality assurance stage could ultimately lead to complications with reproducibility. We would therefore like to present and suggest (1) an alternative strategy for cell and singlet selection from tissue preparations free of FSC/SSC bias for cell sorting and (2) to promote the removal of pro- and late-apoptotic cells prior to introducing any samples into any single-cell RNAseq workflow. The latter is important due to the here presented observation that pro- and late-apoptotic cells are very inefficiently removed from single-cell RNAseq data by purely computational means.

MATERIAL AND METHODS

Mouse Lung Cell Preparation and FACS
Mice were deeply anesthetized by intraperitoneal injection of 120 mg/kg ketamine and 16 mg/kg xylazin (Sigma-Aldrich, Munich, Germany) and exsanguinated. Lungs were perfused with phosphate-buffered saline (PBS) 1× and injected with dispase (BD Biosciences) for tissue digestion. After the lysis of red blood cells with ACK lysing buffer (ThermoFisher Scientific), cells were strained through a 70 µm filter (BD Biosciences) to obtain single-cell suspension. Cells were suspended in 1%FCS, 2 mM EDTA PBS at pH 7.4, and stained with 5 µM/ml DRAQ5 (ThermoFisher Scientific) and 1 µg/ml 4′,6-diamidino-2-phenylindole (DAPI; ThermoFisher Scientific) for live DNA staining and live-dead exclusion. Cells were measured on a BD LSRFortessa™ and BD FACSAria™ Fusion cell sorter (BD Biosciences).

Bone Marrow Preparation and Apoptosis Staining
Bone marrow (BM) cells were harvested from femurs of C57BL6/J mice as described in the literature (10). Briefly, back legs were removed from the mouse, muscle and tissue removed carefully with forceps, and intact bones were clean with 70% ethanol for few seconds. Both ends of the bones were cut and cells were flushed out using a 1 ml syringe filled with complete RPMI 1640 media (GIBCO, Life Technologies). Cell clusters were dissolved by pipetting and red blood cells were lysed with 2 ml of ACK lysing buffer (ThermoFisher Scientific). After cell counting, 1×10^6 cells were resuspended in 2.5 ml of PBS-2% FCS and stained with 2.5 µl of DRAQ5 (ThermoFisher Scientific) for 20 min at 4°C. After incubation, DAPI (ThermoFisher Scientific) was added at a final concentration of 1 µg/ml and samples were acquired on a BD LSRFortessa™ (BD Biosciences).

HEK293 Staurosporine Stressing and Cell Sorting
Cell culture and Apoptosis induction
HEK 293 cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 2 mM glutamine, 1% penicillin/streptomycin, and 10% fetal calf serum (GIBCO, Life Technologies). For apoptosis, induction cells were treated for 2 h with 1 µM Staurosporine (Sigma-Aldrich).

Annexin V binding
Staining for Annexin V binding was performed using FITC Annexin V Apoptosis Detection Kit I (BD Biosciences). Briefly, cells (10×10^6) were washed twice with cold PBS and resuspended in 1× binding buffer at a concentration of 2×10^6 cells/ml. Cells were stained with 250 µl of FITC-conjugated Annexin V (BD Biosciences), and DAPI (Thermo Fisher Scientific) at a final concentration of 1 µg/ml. Cells were then incubated for 15 min at room temperature in the dark.

Cell Analysis and Sorting
Cells were analyzed based on Annexin V binding and DAPI staining and four populations were identified: AnnexinV negative/DAPI negative (live cells), Annexin V positive/DAPI negative (early apoptotic), Annexin V positive/DAPI positive (late apoptotic), and Annexin V positive /DAPI positive (necrotic). Of these four populations, three (live, early apoptotic and late apoptotic) were sorted in a BD FACSAria™ Fusion cell sorter (BD Biosciences) using a 100 µm nozzle. Doublets were carefully excluded by plotting FSC-height

versus FCS-area and SSC-height versus SSC-area, cells with increased area were not consider.

Single-Cell Sequencing Using 10×Genomics

Library preparation was done with the Chromium Single Cell 3'Reagent Kits v3 from 10× Genomics according to manufacturer's protocol. Briefly, libraries were prepared from three different cell states; healthy, pro-apoptotic and apoptotic, which were FACS sorted prior. A fourth library was generated from a not sorted cell suspension to analyze FACS sorting effects. The cells from each sample were partitioned into thousands of nanoliter-scale Gel Beads-in-Emulsion (GEMs). The targeted cell recovery was 7,000 cells for each cell health state. After RT the cDNA was amplified in 12 PCR cycles. Fragmentation, End Repair/A-tailing, and Adapter Ligation were done according to the protocol. The libraries were finalized in 11 PCR cycles using unique Sample Index Primers for each sample. This allowed equimolar pooling of all four generated libraries for sequencing. The purity and library size were validated by capillary electrophoresis using 2,100 Bioanalyzer (Agilent Technologies). The quantity was measured fluorometrically using Qubit dsDNA HS Assay Kit from Invitrogen. Sequencing was done on Illumina NextSeq500 platform reading 16 bp 10× Cell-Barcode and 12 bp UMI in read 1, 8 bp for the sample index read and 130 bp into the insert of interest in read2.

Single-Cell Sequencing Using SmartSeq2

A modified smart-seq2 protocol (11) and tagmentation procedure(12) were used to prepare single-cell sequencing libraries. Briefly, HEK293 cells were FACS sorted directly into 96-well plates containing 4.4 μl of lysis buffer per well at 4°C, and then snap frozen and stored at −80°C until processing. Ten cycles of 50°C and 42°C were omitted from the RT thermal profile and the RT mix was as follows: 2 μl SSRT II 5× buffer; 0.5 μl 100 mM DTT; 2 μl 5 M betaine; 0.1 μl 1 M MgCl2; 0.25 μl 40 U/μl RNAse inhibitor; 0.5 μl SSRT II; 1 μl 10 mM TSO. In Situ PCR primers were omitted and replaced with water and 18 PCR cycles were used for cDNA amplification. The SPRI cleanup (0.6× ratio) was done omitting the ethanol wash step. After the final elution in 13 μl of H_2O, a subset of samples was quality controlled using Bionalyzer High Sensitivity DNA chip and the median cDNA concentration was determined to be used for dilution calculation to normalize the input to 0.2 ng/μl for tagmentation. Before the last cleanup after tagmentation and PCR 1 μl of each sample was pooled together and the pool cleaned up using 0.9× SPRI ratio.

Single-Cell Sequencing Data Analysis

After generating 10×Genomics fastqs and counts tables using cellranger version 3.0, Seurat R package (version 2.3.4) was used to generate quality control plots. Expressed genes in at least three cells and cells expressing at least 200 genes were selected for plotting. After de-multiplexing counts, tables were generated using cellRanger version 3.0.0 implemented in rCASC package (https://github.com/kendomaniac/rCASC,

preprint: https://www.biorxiv.org/content/10.1101/430967v1). Mapping was done using 10×genomics preassembled refdata-cellranger-GRCh38-3.0.0 index.

The late-apoptotic (APO), pro-apoptotic (PRO), and healthy (H) set tables resulted to contain, respectively, 2,261, 6,556 and 5,569 cells. All cells supported by less than 250 detected genes, that is, a gene is called detected is supported by at least three UMIs, were removed using the rCASC function scannobyGtf. After filtering APO, PRO and H tables resulted to contain, respectively, 989, 2081, and 2,260 cells. The 989 APO cells were combined with 989 cells randomly selected from the PRO and H set (apohpro set).

This set was further divided in apohpro_ribomito, that is, containing only ribosomal and mitochondrial protein genes, apohpro_apoptosis, that is, containing only apoptosis-related genes derived (QIAGENset https://www.qiagen.com/ch/resources/download.aspx?id=e5252c51-7513-44a0-b66d-927c53e0eeb2&lang=en), apohpro_celcycle, that is, containing only cell cycle related genes derived (QIAGENset https://www.qiagen.com/us/resources/resourcedetail?id=0ee18e97-d445-4fd7-9aa4-0ef4bece124f&lang=en) and the apohpro_sub set, that is, containing all genes but ribosomal/mitochondrial protein genes. Clustering of the apohpro_sub, apohpro_ribomito, and apohpro_apoptosis sets, in log10 format, was done with tSNE implemented in RtSNE version 0.15, using the following parameters pca = TRUE, perplexity = 50 and theta = 0.

tSNE analysis was done in R version 3.5.1, using as seed 111.

Flow Cytometry and Sequencing Data Deposition

Flow Cytometry files for Figures 1, 2, and 3A can be found at http://flowrepository.org/id/FR-FCM-Z28E.

Sequencing data from both 10×Genomics and SmartSeq2 experiments were deposited at the ENA repository: https://www.ebi.ac.uk/ena with the study accession number PRJEB33078.

Results and Discussion

The two frequently underestimated quality issues during sample preparation for scRNAseq experiments using complex tissue samples or frozen material are a lack of a clear strategy to ensure that only single cells are introduced into scRNAseq assays and making sure that only viable, non-apoptotic cells enter the downstream workflow—if scientifically prudent (e.g., when not studying apoptosis or cell death). Considering the major conclusions drawn from recent single-cell studies, we think it is necessary to explain and exemplify with a few experiments why we believe that researchers and especially core facility personnel should be conscious about these simple and easy to implement quality steps as they can raise awareness within their user-community and provide guidance on how to ensure sample quality.

Gating cells in a forward- and side-scatter (FSC/SCC) plot can be a challenge when dealing with complex tissue homogenates. Cells from tissues generally have a broader scatter profile compared to PBMCs and are often overlapping.
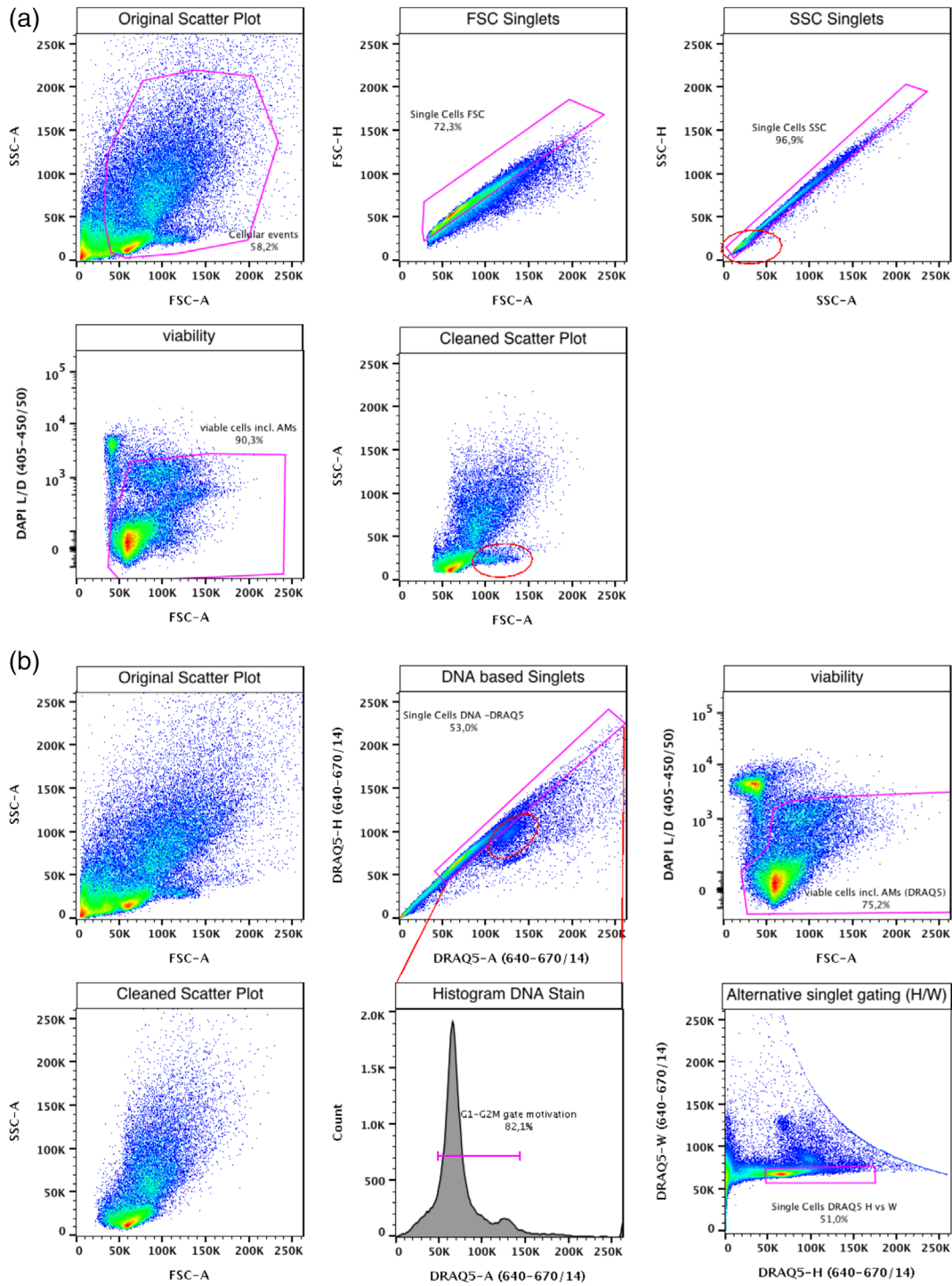
Figure 1. Alternative to FSC/SSC cell and singlet gating to reduce scatter bias. (**A**) Mouse lung tissue digested into single cells with conventional gating strategy using FSC/SSC scattering for cell identification and singlet gating. Red circle indicates a doublet contamination due to often observed scatter spread in complex tissue samples. (**B**) Alternative gating of the same sample after DNA staining with DRAQ5 to provide a precise and FSC/SSC independent singlet gate (DRAQ5 gate set on lower G1 DNA content boundary—see Histogram DNA Stain). Note that the clear detection of doublets (red circles) and the still complex, but uncut FSC/SSC profile when using the DRAQ5 cell and doublet identification strategy. Singlet gating based on H/A or H/W give the same results and can be used interchangeable.

Looking at a whole mouse lung digestion, drawing a gate into the FSC/SSC clouds could potentially induce a bias by excluding very low or very strongly scattering cells (Fig. 1A). Lung

tissue has a high complexity as it contains a number of different cell types with varying cell sizes and autofluorescence along with considerable debris (13,14). Gating cells and
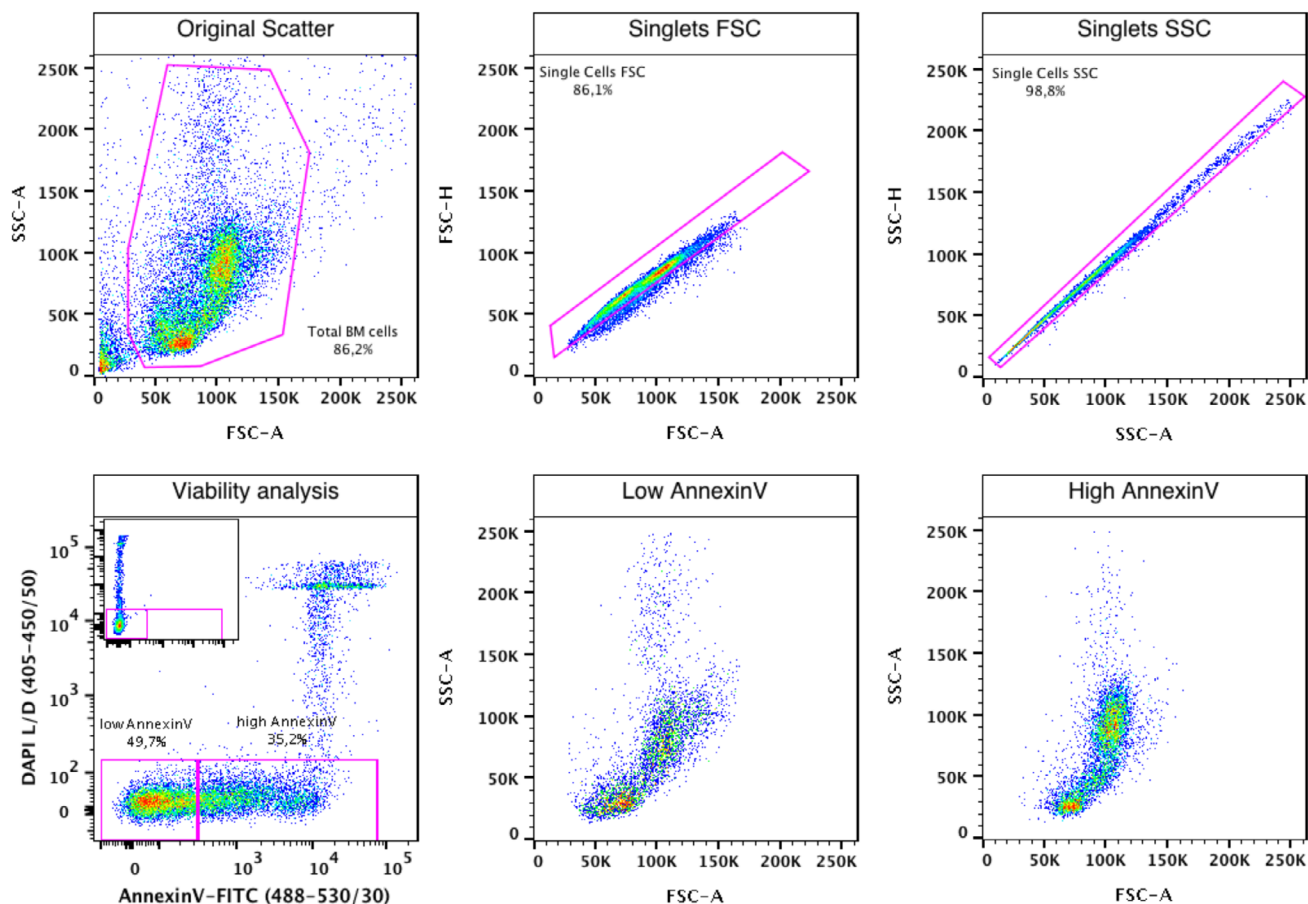
Figure 2. Complex tissue separations show increased occurrence of pro-apoptotic cells as a source for unwanted contamination for single-cell sequencing. Mouse Bone Marrow cells were prepared for cell sorting. FACS analysis of the tissue sample showed significant occurrence of pro-apoptotic cells that would be missed by simple DNA intercalation live-dead staining or relying on a change of FSC/SSC scatters (Plot inlay in the DAPI vs AnnexinV plot shows AnnexinV unstained bone marrow cells). The overlooked pro-apoptotic cells, when using only cell-impermeable live-dead stains, would be introduced into sequencing experiments as an unknown contaminant.

singlets on scatter signals will ultimately result in an arbitrary cut-off (Fig. 1A). If one were to omit a gate defining "cell-like" events in the FSC/SSC plot, in order to avoid a bias by selecting on scatter, one still needs to include a singlet-gating strategy to ensure that predominantly only single-cells are sorted for single-cell sequencing approaches. Some researchers have rightfully argued that gating singlets by FSC and SSC pulse-shape (A/H or H/W, etc) analysis will again reintroduce a bias because cells may have different scatter characteristics and laser-immersion periods due to their size difference causing a larger area and width signal that could be wrongly interpreted as doublet event. Indeed, singlet-gating plots from tissues (Fig. 1A) appear broader and not as defined as those from PBMCs, for example, making it difficult for an inexperienced researcher to limit the gate to avoid doublets and not to introduce a potential bias to larger cells. This argument has been made in a few studies, which completely avoided any gating on cellular scatters, including singlet gates. Those hopefully "single cells" were then sorted for SmartSeq2 sequencing claiming that the cells would be single cells and that any doublets could be identified by detecting sequences

of two different cell types stemming from the same "single-cell" event. As for droplet or micro-well-based scRNASeq based methods, a too high cell-concentration can lead to double loading of a droplet or micro-well with two cells, similar to droplet based cell sorting (15). Yet countering such Poisson statistic limitation by lowering the input with lower cell-concentrations will not change the frequency of true doublet cell–cell events, which are still bound by cell–cell contacts or extracellular matrix. It is of course possible to call a doublet if one has a doublet formed of clearly different cell types (15,16) but not applicable for doublets formed from the same cell type.

This problem can be addressed by using a suitable approach available to Flow Cytometry, as one could use a cell-permeable DNA dye to avoid the whole scattering bias problem and tackle singlet selection more efficiently. Basing the cell identification and singlet gating strategy on the DNA content signal allows to completely avoid FSC/SSC signals— the DNA content will be the same in all cells from the same organism, allowing for G1, S, and G2/M cell-cycle progression. For example, applying a simple DRAQ5 staining (17) to
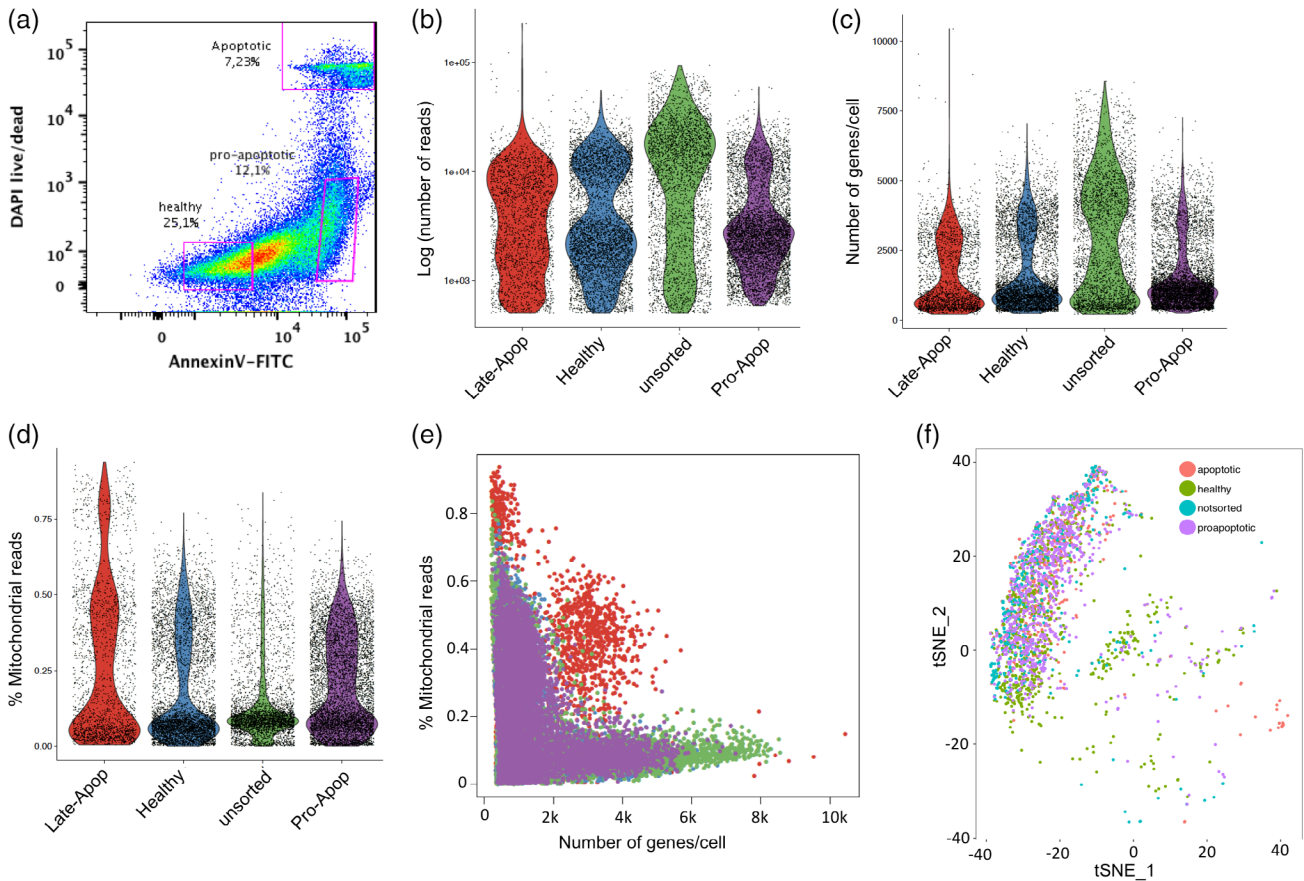
Figure 3. Dead and pro-apoptotic cell exclusion by single-cell sequencing is not efficient enough to guarantee removal from data set—10×Genomics data. (**A**) 70% confluent HEK293 cells were treated for 2 h with 1 μM Staurosporine, then harvested from the 10 cm dish, and stained with AnnexinV-FITC and DAPI to identify healthy, pro-apoptotic, and apoptotic cells. Each population was sorted and from 50.000 cells received approximately 7,500 cells were loaded onto a 10×Genomics cartridge—see Methods for details. (**B,C**) Quality analysis of the 10×Genomics and Illumina sequencing run checking overall number of reads and genes detected in the cells—no bias indicated, display type violin plots. (**D**) Analysis of %mitochondrial reads shows in general less cells with high %mitochondrial reads in healthy or non-sorted, but a significant number of cells reporting low %mitochondrial reads in the pro- and apoptotic populations which would escape this cutoff. (**E**) Plotting %mitochondrial reads versus number of genes detected per cell reveals that a majority of the apoptotic cells is clearly identifiable but that a large portion of the cells escapes independent of the total number of genes detected. (**F**) tSNE clustering of the fully bio-informatically cleaned sample set demonstrates that pro-apoptotic cells and escaping apoptotic cells cannot be separated and cluster with the healthy cells. Nonsorted cells cluster all over with the other populations recapitulating the input. Note that the small cluster of exclusively apoptotic cells might be misleading when analyzing the data.

the murine lung tissue homogenate allows for FSC/SSC free cellular event identification and singlet gating at the same time (Fig. 1B, circle gates highlight doublets identified by DRAQ5 pulse-shape analysis; gating DNA content based singlets can be done either by H/A or H/W pulse shape analysis) and will provide a simple, clear and solid motivation for setting sort gates. In this example, the cutoff between debris and cells was based on the very clearly defined G1 population, which ensures that particles with intact nuclei are used as a base gate in combination with the common singlet gating for cell-cycle experiments. This strategy was in our hands superior as it also picked up challenging doublets, using well-established cell-cycle gating strategies (18), which were difficult to gate without introducing arbitrary cutoffs in the FSC/SSC scatter and singlet plots (Fig. 1A,B). Importantly, several studies observed no adverse effects of DRAQ5 and

DAPI staining on library preparation or sequencing for many different downstream applications (scRNAseq, ATAC, Hi-C, SmartSeq2, CI-Seq, etc.) (19–22).

The second important quality factor, cell viability, came to our attention because we observed ourselves lower quality scRNAseq results stemming from complex tissue separations or tissue cultures treated with cytokines, viruses, or compounds. Because our users lost a significant number of cells with FACS-coupled single-cell CI-Seq (23) and SmartSeq2 (11) sequencing assays, we took a step back and evaluated our sorting strategy. Exploring our generally simple live-dead gating strategy using only DAPI, Hoechst or PI, we noticed that the pragmatic approach of live-dead staining with cell-impermeable DNA dyes was insufficient to guarantee a fully healthy and viable subgating strategy. DNA intercalation based on cell-impermeable live/dead staining approaches fail

to label pro-apoptotic cells, which we observed in treated tissue cultures, brain tissue samples, or bone marrow preparations (Fig. 2). Including an apoptosis staining like AnnexinV(24) or Caspase3/7 (25) probes, when experimentally possible, overall improved our operations toward single-cell sequencing. Our omission to suggest additional apoptotic staining to our users worried us that we were not following the standards in the field.

Our literature research and active inquiries gave a mixed picture with a strong tendency to avoid looking at the viability/apoptotic state of the sample at the stage of harvest and to limit the exclusion of potentially apoptotic or dead cells to bioinformatics after sequencing. This was of course more predominant with researchers using droplet or micro-well-based scRNASeq methods, like 10×Genomics. Further literature search on studies looking at the identification of pro-apoptotic and dying cells in single-cell sequencing approaches let us to a study using the C1-system and describing in essence the increase of mitochondrial reads paired with the abundance of ribosomal genes as a potential identifier for challenged cells(26).

Given the huge amount of single-cell studies currently undertaken from complex tissues and the rather vague definition for identifying challenged or dead cells by computational means, we decided to explore whether FACS-coupled SmartSeq2 or 10×Genomics and experimentally guided approaches would actually result in clearly identifiable transcriptomic signatures for healthy, pro-apoptotic, and apoptotic cells stemming from a single, staurosporine-treated HEK293 cell culture(27). We treated HEK293 cells for 2 h with 1 μM staurosporine and harvested the cells by collecting the media and still attached cells from the flask. The cells were stained with AnnexinV-FITC plus DAPI and then sorted into healthy, pro-apoptotic, and apoptotic fractions using low-pressure settings and a 100 μm nozzle (Fig. 3A). Approximately, 7,000 cells were targeted for library preparation from each fraction for standard 3' RNA sequencing with a version3 10×Genomics kit and complemented it with an unsorted sample from the same source. Additionally, we sorted for each health state 144 cells into 96-well qPCR plates for standard SmartSeq2 sequencing(11) and each 50,000 cells for bulk RNA analysis. We induced apoptosis by staurosporine as it provided us with a fast and global induction of caspase activity avoiding endogenous cues coming from cell cycle checkpoints or other extracellular sources (27).

Analyzing the 10×Genomics data set, we were able to recover 3,700 non-sorted, 5,547 healthy sorted, 6,644 pro-apoptotic sorted, and 2,295 apoptotic sorted cells, which underwent computational analysis. In a first instance, we observed no difference in the number of reads obtained from all four sample sets (Fig. 3B) indicating that our treatment did not result in a bias of the 10×Genomics protocol. Comparing the number of genes detected in all cells, we observed that our experiment showed a rather large spread from approximately 250 up to 5,000 genes per cell (Fig. 3C). The overall spread between low and high gene content cells was similar between the sorted fractions with healthy cells

showing slight increase of cells with more than 3,000 genes detected. We then analyzed the whole data set for the percentage of mitochondrial reads detected in each cell no matter the absolute number of genes revealing the expected larger proportion of mitochondrial reads in the apoptotic population (Fig. 3D). The healthy, pro-apoptotic and non-sorted populations reported in general an overall lower number of cells with <15% mitochondrial reads, but we observed a similar spread in apoptotic cells which we would not expect to detect. Plotting the mitochondrial reads versus the gene content of each cell revealed that an increase of mitochondrial reads content is indeed identifying most of the late-apoptotic cells (Fig. 3E), but there are a significant number of late-apoptotic cells that report a good gene content and low mitochondrial reads. Pro-apoptotic cells have very low number of mitochondrial reads, similar to healthy cells, and would substantially escape a harsh mitochondria read cutoff.

tSNE analysis using only apoptosis or cell cycle gene signature was not sufficiently informative to allow to discriminate between pro-apoptotic and healthy cells (Fig. 4A,B). Furthermore, differential expression analysis between pro-apoptotic and healthy cells do not detected any differential expressed gene. Thus suggesting that the transcriptome of pro-apoptotic and healthy cells does not retain sufficient information to allow a bioinformatic separation of the above-mentioned groups.

Many researchers are using the popular SEURAT package (5), which combines a QC matrix using (among other factors) a low threshold of mitochondrial reads, abundance of ribosomal genes, the number of total reads per cell, and the number of genes detected. Applying SEURAT to our data set removed the majority of potentially called cells from the 10×Genomics run penalizing low quality "droplets." Confirming our initial suspicion that biocomputational methods would probably be imperfect in removing dying or dead cells completely from a sequencing-based approach, we observed a significant number of late-apoptotic (approximately 25%) and pro-apoptotic cells (60%) escaping the SEURAT QC and clustering with healthy cells on our tSNE plot (Fig. 3F). Although this indicates that at least healthy and pro-apopotitc cells still have at large the same expression profile, it also concludes that generally used bioinformatic QC tools are not the best choice to remove health-challenged cells from a data set. It is noted that some of the escaping late-apoptotic cells appear to form a loose cluster on their own.

Apart from using SEURAT, we also used a tailored approach to further analyze this data set. Specifically, we used rCASC package (28) to evaluate, for each cell, the fraction of total cell counts associated with mitochondrial and ribosomal genes (Fig. 5A–C). We observed that the apoptotic cells (Fig. 4C) were characterized by a different distribution of cells expressing high number of genes, that is, > 250 genes (a gene is called detected if supported by at least three UMIs), with respect to healthy (Fig. 6A) and pro-apoptotic (Fig. 5B) cells. Specifically, apoptotic cells expressing high number of genes were mainly associated with high fraction of counts associated with mitochondrial genes and low fraction of counts
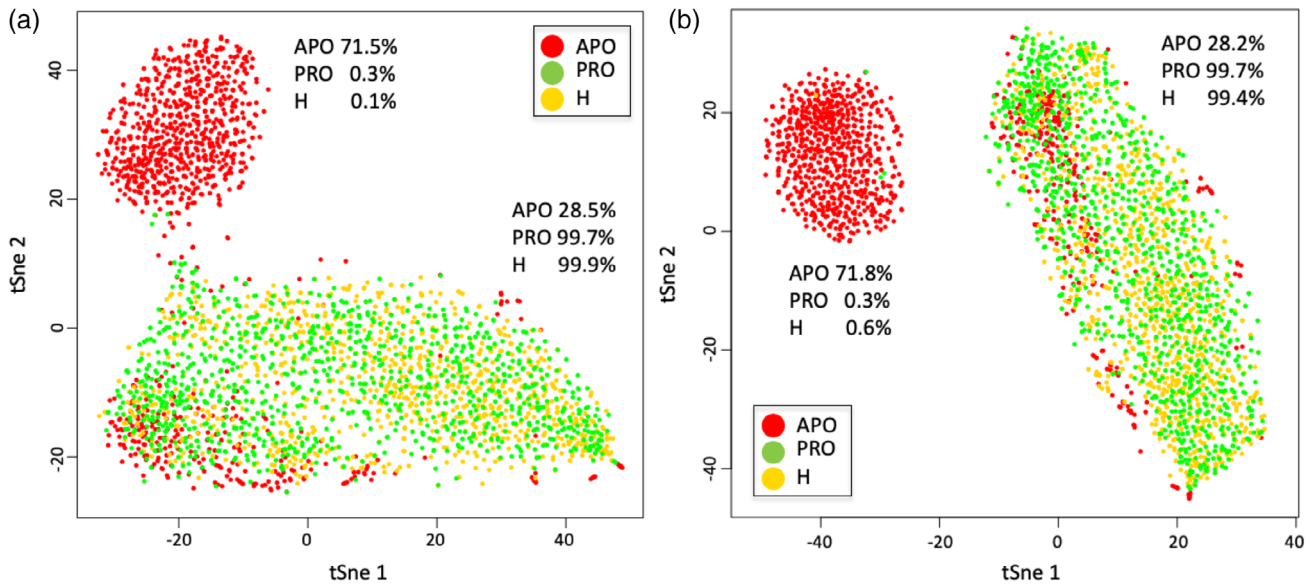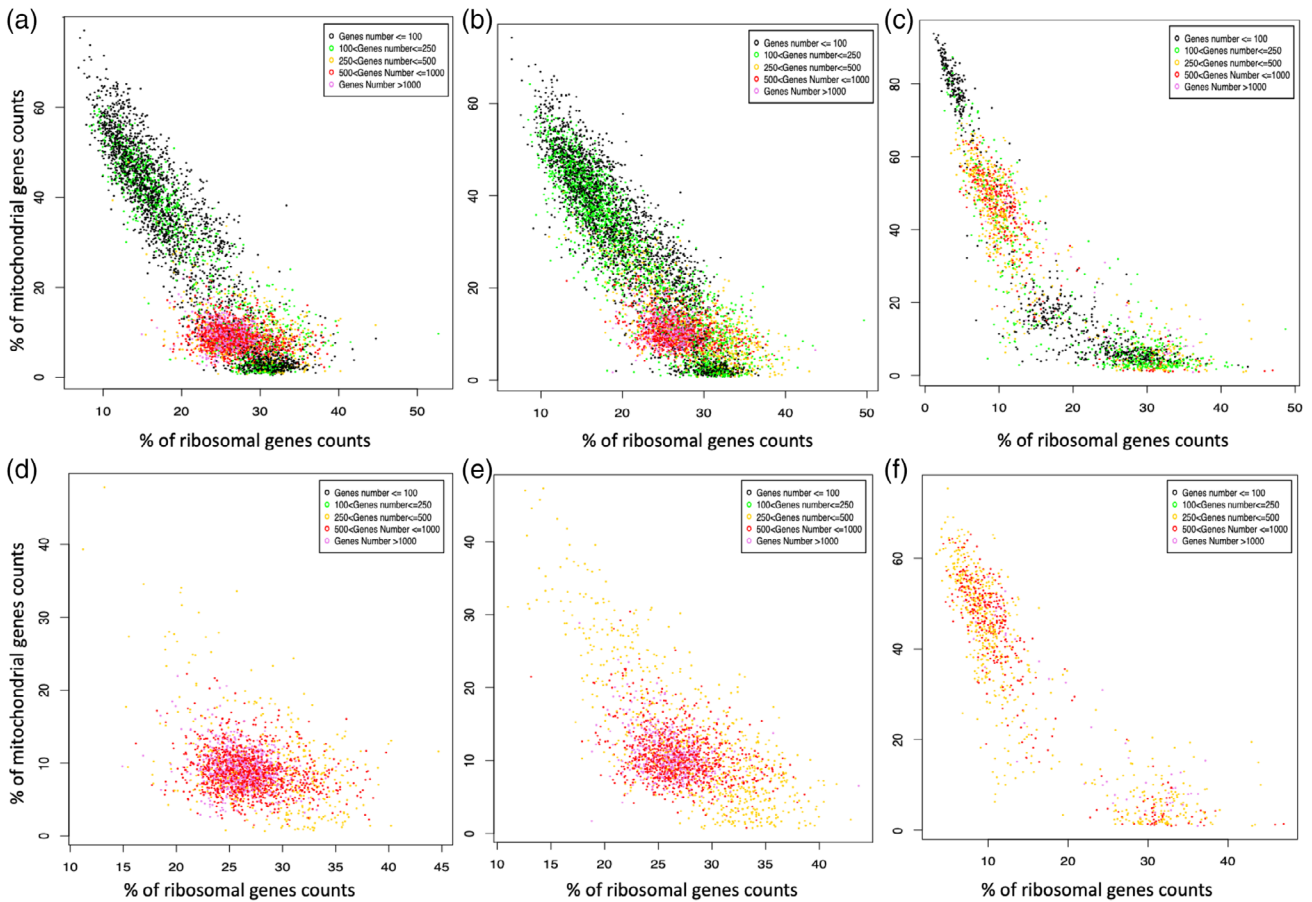
Figure 4. tSNE analysis of that data set made of healthy (gold), pro-apoptotic (green), and apoptotic cells (red), 989 cells for each data set after removal of cells with less than 250 detected genes (A) tSNE plot using all genes but mitochondrial and ribosomal genes using Qiagen apoptosis gene signature. (B) tSNE plot using all genes but mitochondrial and ribosomal genes using Qiagen cell cycle gene signature.

associated with ribosomal genes (Fig. 5C). In healthy (Fig. 5A) and pro-apoptotic (Fig. 6B) cells instead, cells expressing high number of genes were associated with low fraction of counts associated with mitochondrial genes and high fraction of counts associated with ribosomal genes. The different distribution of cells in the apoptotic group is even more evident if cells supported by less than 250 genes are removed (Fig. 5D–F). We assembled a data set made of healthy (H), pro-apoptotic (PRO) and apoptotic (APO) cells, made only of cells supported by more than 250 genes/cell, that is, 989 cells for each group. We analyzed this data set using tSNE (https://cran.r-project.org/web/packages/Rtsne/index.html). tSNE analysis was performed on all genes but mitochondrial and ribosomal genes (Fig. 6A) and only on mitochondrial and ribosomal genes (Fig. 6B). The analysis done on all genes but mitochondrial and ribosomal genes (Fig. 6A) generated three clusters, one made by 75% of apoptotic cells and a negligible fraction of healthy and pro-apoptotic cells, respectively, 0.2% and 0.5%. Another cluster was largely made of heathy and pro-apoptotic cells, respectively, 81.4% and 72%, and a small fraction of apoptotic cells, 3.1%. The third cluster was instead made of a similar fraction of the three cell populations: 23.3% apoptotic, 27.8% pro-apoptotic, and 18.3% healthy cells (Fig. 6A). tSNE analysis done on the data set made only of mitochondrial and ribosomal genes generated two clusters, one made mainly by apoptotic cells (74% APO, 0.5% PRO, and 0.2% H) and another cluster containing nearly all pro-apoptotic and healthy cells (99.4% APO and 99.7% H) and 25.9% of apoptotic cells (Fig. 6B). tSNE analysis done on a set of apoptotic genes (https://www.qiagen.com/ch/resources/resourcedetail?id=e5252-c51-7513-44a0-b66d-927c53e0eeb2&lang=en) produced results superimposable on those obtained using only

mitochondrial and ribosomal genes (not shown). In any case, there was never an efficient removal of all apoptotic cells from the data set and very much none for pro-apoptotic cells similar to the SEURAT approach.

Our results of largely ineffective in silico removal of (pro-)apoptotic cells from the 10×Genomics data set could be due to the lower sequencing depth of this approach. We therefore analyzed in parallel cells from the sample with SmartSeq2 which allows for deeper sequencing (2,15,29). We sequenced approximately 144 cells per condition paired with an independent bulk for each health status. Out of curiosity, we included a small set of 16 "doublets" made of healthy and apoptotic cells. SmartSeq2 allows for very deep sequencing and has a higher chance to resolve changes in low-expressed genes or possibly apoptosis regulated mRNA instabilities which might reveal subtle makers for identifying (pro-)apoptotic cells otherwise lost by the shallower sequencing depth in 10×Genomics runs(15,29). The sequencing run resulted on average 500,000 sequence reads and roughly 10,000 genes detected per cell—that is about two to three times as many as with 10×Genomics, although with a significant reduction in throughput. This was the same for all health conditions allowing us to exclude a bias of the data based on cell health similar to our 10×Genomics run (Fig. 7A, B). Generally using filters on the proportion of mitochondrial reads and on the number of detected genes per cell are even in deep-sequencing approaches insufficient in separating faithfully (pro-)apoptotic cells from the sample set (Fig. 7C). This is also impacting our downstream analysis, namely PCA. Unsupervised clustering analysis by PCA of any high-quality cell in the SmartSeq2 run confirms the nonsignificant separation of pro-apoptotic cells from healthy cells as they co-cluster extensively (Fig. 7D). We would like to postulate that

Figure 5. %of UMIs associated with mitochondrial/ribosomal genes (**A**, **D**) Healthy cells. (**B**, **E**) Pro-apoptotic cells. (**C**, **F**) Apoptotic cells. (A–C) All sequenced cells. (D–F) Cells left after removing cells with less than 250 detected genes, that is, a gene is called detected if supported by at least three UMIs.
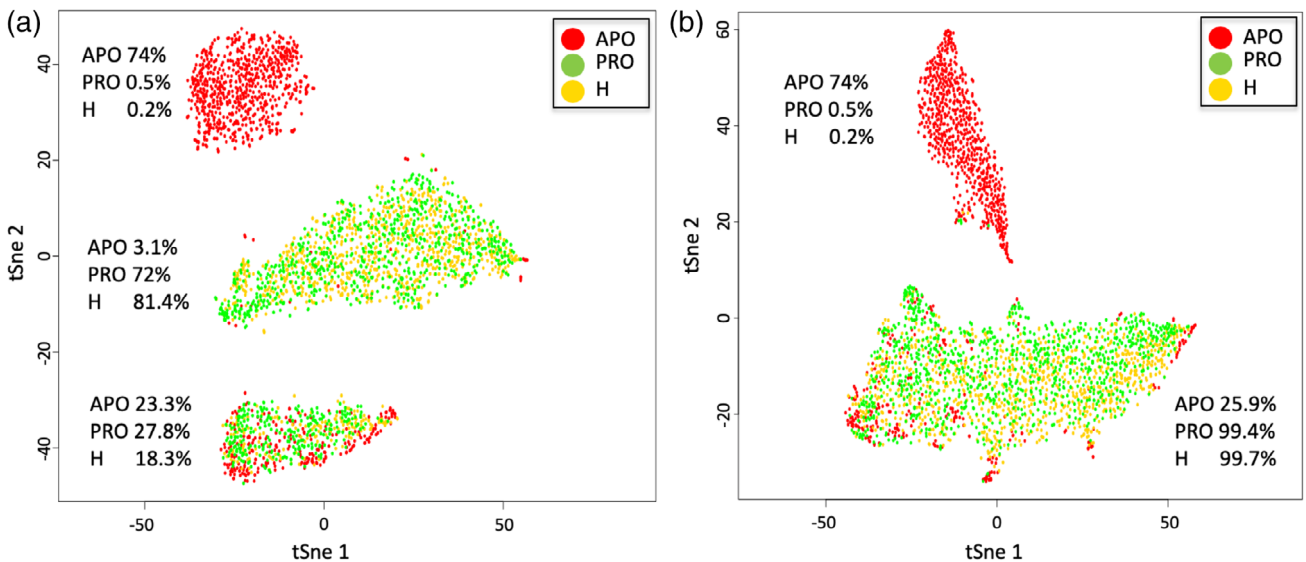

Figure 6. tSNE analysis of that data set made of healthy (gold), pro-apoptotic (green), and apoptotic cells (red), 989 cells for each dataset after removal of cells with less than 250 detected genes (**A**) tSNE plot using all genes but mitochondrial and ribosomal genes. (**B**) tSNE plot using only ribosomal and mitochondrial genes. % of each cell type in the tSNE clusters are indicated in the pictures.

the often-used quality filter based on mitochondrial reads, ribosomal sequences and total detected genes per cell might actually flag poor reverse transcription (RT) performance rather than "viability" of single cells (if applied to validated healthy cells)—this might be good news as it could be a perfect tool for deciding when to discard a data set if a high load of poor RT QC events is detected.

In parallel to the main apoptosis question, we tried to address doublet detection sensitivity—to this end, we sorted a healthy and a pro-apoptotic cell together into the same micro-well. Noteworthy, with regards to this artificial "doublet" subset (Fig. 7), we observed a significantly lower number of genes detected in those SmartSeq2 reactions (2,000–3,000

genes/cell). Although we cannot explain why two cells in the same reaction volume gave a lower detected gene count, we chose to report the data, because they still cluster very well with all other single cells. Those cells could be included in analyses where the filters for gene/cell expression are generally broader, for example, Fig. 3C with 200–7,000 genes/cell.

As this Cytometry Part A issue is about rigor in cytometry, we need to look at the results presented in this article from a couple of different angles: Cytometry is broadly pushed to new heights by immunological research handling mostly easily accessible single-cell material, for example, PBMCs. However, there is a large proportion of researchers applying single-cell sequencing methodology developed on
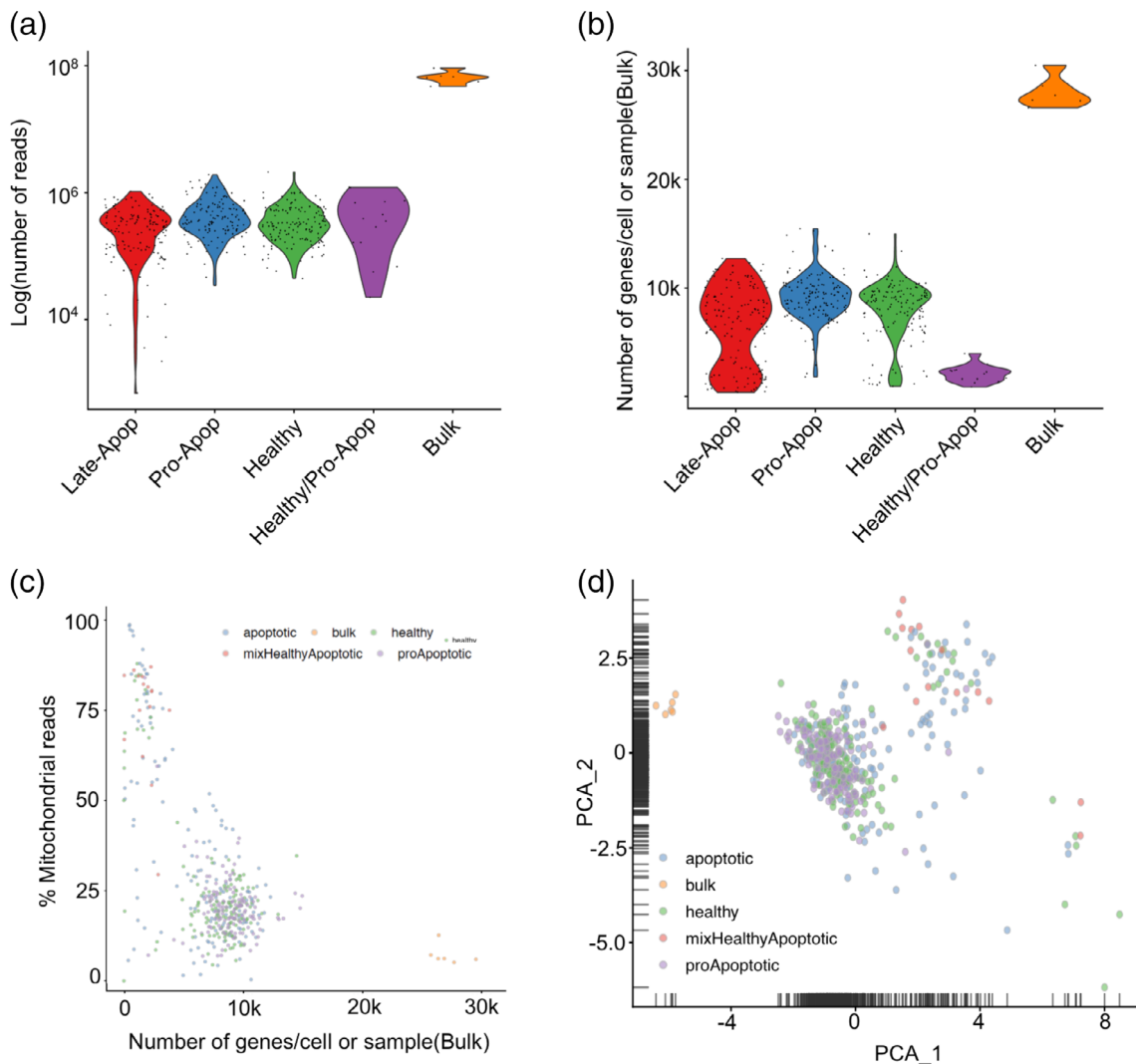
Figure 7. Dead and pro-apoptotic cell exclusion by single-cell sequencing is not efficient enough to guarantee removal from data set—SmartSeq2. (**A** and **B**) Violin plots showing the overall number of reads and number of genes detected for each cell and health state. (**C**) Quality assessment based on %mitochondrial sequences and gene count per cell is insufficient in completely removing (pro-)apoptotic cells from the sample set. (**D**) PCA clustering of the fully bio-informatically cleaned sample set demonstrates that pro-apoptotic cells and escaping apoptotic cells do not separate and cluster with the healthy cells. The bulk duplicates of the different health states cluster very closely with each other indicating that there is no underlying visibly different gene-expression profile during the mapped progression of apoptosis. Note that the overall broader clustering of apoptotic cells in the PCA analysis similar to the observation in the corresponding 10×Genomics experiment.

almost perfectly single cell and extremely low apoptotic blood samples to very complex and heterogeneous solid-tissue samples. Most tissue types require digestion and gauging to turn them into a potential single-cell suspension with a variable outcome in quality. Blindly using these preparations for downstream sequencing approaches without assessing viability or avoiding singlet gating during sorting, due to fear of biasing the input, is falling short of available and standardized quality assurance steps that would allow for an optimal assay setup. There are ready to use alternative strategies avoiding a bias based on light scattering when performing cell sorts—we suggest a simple and sequencing compatible strategy for singlet identification based on viable DNA staining with DRAQ5. At the same time, viability of the input material appears to be an area, which requires more attention in order to assure reasonable quality during sample preparation. The inability of bio-computational methods to faithfully remove all late-apoptotic cells from a single-cell sequencing experiment in silico (or better post mortem!), with a striking inability to detect a low percentage of pro-apoptotic cells, calls for a mandatory wet-lab step prior to loading droplet or micro-well-based scRNASeq chips or PCR-plates with single cells.

Presenting our data, we noticed that many researchers using a FACS-coupled SmartSeq2 scRNAseq approach were arguing that they use 7AAD or Sytox probes to label their dead cells and believed that they hence circumvent the issue of pro-apoptotic contaminants—but this is not the case: any cell membrane-non-impermeable DNA intercalating dye (like DAPI, Hoechst, PI, 7AAD, DRAQ7) or fixable amine-reactive live-dead staining will only report "dead cells" if the cellular membrane is broken up. This is not the case in pro-apoptotic cells and none of the classic live-dead stains will resolve the early apoptotic stages, which require alternative staining methods using Caspase probes or AnnexinV labeling (24,25).

There is a potential to argue that pro-apoptotic contaminants might not be an issue for single-cell sequencing approaches as they still cluster predominantly with healthy cells. Yet we would like to highlight that the more careful and prudent approach would be to better remove any non-healthy cells (pro- and apoptotic cells) when applying analysis methods based on mRNA undergoing constant turnover, like scRNAseq. This quality argument is especially prominent in single-cell preparations stemming from solid tissues and also valid for frozen PBMC samples as they tend to contain increased levels of (pro-)apoptotic cells(30) contaminating any following single-cell sequencing analysis that omitted a quality improving removal step of "apoptotic" cells. In order to execute a full circle on the subject, pro-apoptotic cells have been shown to display a fast decay of mRNAs (31). The reverse-transcription step in any analysis of cDNA-based experiment (RT-PCR, qPCR, single-cell sequencing, etc) is known to be rather inefficient and highly variable (32,33), depending upon single molecule/transcript abundance and template availability due to a secondary structure formation. Therefore, we believe that there is a likely possibility for

introducing unnecessary extrinsic noise into the experiment (34) when using cells that are undergoing random but efficient RNA degradation, like pro–/late-apoptotic cells (31).

We are fully aware that the spike in single-cell sequencing studies is driven by the excitement of being able to explore heterogeneity at high resolution. We are trying to highlight potential pitfalls to our users in our own institutions, although we fail to persuade everyone to adopt a more careful strategy toward potentially (pro-)apoptotic contaminants—we are indeed glad to see that the message is spreading and more researchers are approaching us with the will to optimize their sampling strategy with regards to enhancing quality, reproducibility, and comparability. As scientists serving the community in core facilities, we should embrace our responsibility to apply the best suitable strategy to guarantee that results are based on input material of the highest possible quality even if it means to introduce an additional time-consuming step into existing pipelines and continuously lobby for best practices.

LITERATURE CITED

1. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet 2019;2:666.
2. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell 2017;65:631–643.e4.
3. Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 2013;10:1093–1095.
4. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 2014;11:163–166.
5. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–420.
6. Tian L, Su S, Dong X, Amann-Zalcenstein D, Biben C, Seidi A, Hilton DJ, Naik SH, Ritchie ME. scPipe: A flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data. Pertea M, editor. PLoS Comput Biol. 2018;14:e1006361.
7. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature 2017;541:331–338.
8. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 2018;173:1307.
9. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 2015;161:1202–1214.
10. Liu X, Quan N. Immune Cell Isolation from Mouse Femur Bone Marrow. Bio Protoc 2015;5.
11. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 2014;9:171–181.
12. Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, Besir H, Remans K, Steinmetz LM. Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. G3 (Bethesda) 2018;8:79–89.
13. Iwasaki A, Foxman EF, Molony RD. Early local immune defenses in the respiratory tract. Nat Rev Immunol. 2017;17:7–20.
14. Lloyd CM, Marsland BJ. Lung Homeostasis: Influence of Age, Microbes, and the Immune System. Immunity 2017;46:549–561.
15. Lafzi A, Moutinho C, Picelli S, Heyn H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. Nat Protoc 2018;13:2742–2757.
16. DePasquale EAK, Schnell DJ, Valiente I, Blaxall BC, Grimes HL, Singh H, Salomonis N. DoubletDecon: Cell-State Aware Removal of Single-Cell RNA-Seq Doublets. bioRxiv 2018:364810.
17. Smith PJ, Blunt N, Wiltshire M, Hoy T, Teesdale-Spittle P, Craven MR, Watson JV, Amos WB, Errington RJ, Patterson LH. Characteristics of a novel deep red/infrared fluorescent cell-permeant DNA probe, DRAQ5, in intact human cells analyzed by flow cytometry, confocal and multiphoton microscopy. Cytometry 2000;40:280–291.
18. Wersto RP, Chrest FJ, Leary JF, Morris C, Stetler-Stevenson MA, Gabrielson E. Doublet discrimination in DNA cell-cycle analysis. Cytometry 2001;46:296–306.
19. Lother A, Bergemann S, Deng L, Moser M, Bode C, Hein L. Cardiac Endothelial Cell Transcriptome. Arterioscler Thromb Vasc Biol 2018;38:566–574.
20. Beetz N, Rommel C, Schnick T, Neumann E, Lother A, Monroy-Ordonez EB, Zeeb M, Preissl S, Gilsbach R, Melchior-Becker A, et al. Ablation of biglycan attenuates cardiac hypertrophy and fibrosis after left ventricular pressure overload. J Mol Cell Cardiol 2016;101:145–155.

21. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 2017;357:661–667.

22. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, Pliner HA, Christiansen L, Qiu X, Steemers FJ, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 2018;555:538–542.

23. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 2015;348:910–914.

24. Vermes I, Haanen C, Steffens-Nakken H, Reutelingsperger C. A novel assay for apoptosis. Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labeled Annexin V. J Immunol Methods 1995;184:39–51.

25. Belloc F, Belaud-Rotureau MA, Lavignolle V, Bascans E, Braz-Pereira E, Durrieu F, Lacombe F. Flow cytometry detection of caspase 3 activation in preapoptotic leukemic cells. Cytometry 2000;40:151–160.

26. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, Teichmann SA. Classification of low quality cells from single-cell RNA-seq data. Genome Biol. 2016;17:29.

27. Belmokhtar CA, Hillion J, Ségal-Bendirdjian E. Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms. Oncogene 2001;20:3354–3362.

28. Alessandri L, Beccuti M, Arigoni M, Olivero M, Romano G, De Libero G, Pace L, Cordero F, Calogero R. rCASC: reproducible Classification Analysis of Single Cell sequencing data. bioRxiv 2018:430967.

29. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, Hughes TK, Wadsworth MH, Burks T, Nguyen LT, et al. Systematic comparative analysis of single cell RNA-sequencing methods. bioRxiv. 2019;10:439.

30. Wang L, Hückelhoven A, Hong J, Jin N, Mani J, Chen B-A, Schmitt M, Schmitt A. Standardization of cryopreserved peripheral blood mononuclear cells through a resting process for clinical immunomonitoring–Development of an algorithm. Cytometry A 2016;89:246–258.

31. Thomas MP, Liu X, Whangbo J, McCrossan G, Sanborn KB, Basar E, Walch M, Lieberman J. Apoptosis Triggers Specific, Rapid, and Global mRNA Decay with 3' Uridylated Intermediates Degraded by DIS3L2. Cell Rep 2015;11:1079–1089.

32. Chen H-IH, Jin Y, Huang Y, Chen Y. Detection of high variability in gene expression from single-cell RNA-seq profiling. BMC Genomics 2016;17 Suppl 7:508.

33. Dueck HR, Ai R, Camarena A, Ding B, Dominguez R, Evgrafov OV, Fan J-B, Fisher SA, Herstein JS, Kim TK, et al. Assessing characteristics of RNA amplification methods for single cell RNA sequencing. BMC Genomics 2016;17:966.

34. Schwaber J, Andersen S, Nielsen L. Shedding light: The importance of reverse transcription efficiency standards in data interpretation. Biomol Detect Quantif 2019;17:100077.