Chapter 5

Full-Length Single-Cell RNA-Sequencing with FLASH-seq

Vincent Hahaut and Simone Picelli

Abstract

The single-cell RNA-sequencing (scRNA-seq) field has evolved tremendously since the first paper was published back in 2009 (Tang et al. Nat Methods 6:377–382, 2009). While the first methods analyzed just a handful of cells, the throughput and performance rapidly increased over a very short time span. However, it was not until the introduction of emulsion droplets methods, such as the well-known kits commercialized by 10x Genomics, that the robust and reproducible analysis of thousands of cells became feasible (Zheng et al Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049, 2017). Despite generating data at a speed and a cost per cell that remains unmatched for full-length protocols like Smart-seq (Hagemann-Jensen et al Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol 38:708–714, 2020; Picelli et al Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098, 2013), scRNA-seq in droplets still comes with the drawback of addressing only the terminal portion of the transcripts, thus lacking the required sensitivity for comprehensively analyzing the entire transcriptome.

Building upon the existing Smart-seq2/3 workflows (Hagemann-Jensen et al Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol 38:708–714, 2020; Picelli et al Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098, 2013), we developed FLASH-seq (FS), a new full-length scRNA-seq method capable of detecting a significantly higher number of genes than previous versions, requiring limited hands-on time and with a great potential for customization (Hahaut et al. Lightning Fast and Highly Sensitive Full-Length Single-cell sequencing using FLASH-Seq. http://biorxiv.org/lookup/doi/10.1101/2021.07.14.452217. https://doi.org/10.1101/2021.07.14.452217, 2021). Here, we present three variants of the FS protocol.

Standard FLASH-seq (FS), which builds upon Smart-seq2 developed in the past, is non-stranded and does not use unique molecular identifiers (UMIs) but still remains the easiest method to measure gene expression in a cell population.

FLASH-seq low-amplification (FS-LA) represents the fastest method, which generates sequencing-ready libraries in 4.5 h, without sacrificing performance.

FLASH-seq with UMIs (FS-UMI) builds upon the same principle as Smart-seq3 and introduces UMIs for molecule counting and isoform reconstruction. The newly designed template-switching oligonucleotide (TSO) contains a 5-bp spacer, which allows the generation of high-quality data while minimizing the amount of strand-invasion artifacts.

Key words FLASH-seq, FLASH-seq low-amplification, FLASH-seq with UMI, RNA-seq, single cell, full-length, automation, high-throughput

Raffaele A. Calogero and Vladimir Benes (eds.), *Single Cell Transcriptomics: Methods and Protocols*,
Methods in Molecular Biology, vol. 2584, https://doi.org/10.1007/978-1-0716-2756-3_5,

1 Introduction

Performing scRNA-seq in emulsion droplets by using 10X Genomics technology is nowadays routine practice for many research groups around the world. The system is robust, reproducible, and offers an unparalleled throughput at a very low cost per cell. The main limitation remains that available kits enable the characterization of either the 3′- or the 5′-end of each RNA, at least when using short read sequencing. This is generally sufficient for the investigation of cellular heterogeneity and the identification of population sub-structures. However, sequencing full-length transcripts would also allow the analysis of single nucleotide polymorphisms (SNPs) and enable splice isoform reconstruction, the characterization of transcriptional start sites (TSSs), as well the identification of monoallelic and imprinted genes [6–8].

Smart-seq is probably the most well known and robust among the full-length RNA-seq methods and has now reached the third generation with the recent publication of Smart-seq3 [3]. Although differences exist between Smart-seq workflows, they all rely on the SMART technology (switching mechanism at the 5′-end of the RNA transcript), a peculiar property of the Moloney murine leukemia virus reverse transcriptase (MMLV-RT), which couples reverse transcription (RT) and template switching (TS) in a single reaction [9]. Template switching represents the ability of the MMLV-RT to introduce a few untemplated nucleotides, often 2-5 cytosines, upon reaching the 5′-end of the RNA template during RT reaction. These extra nucleotides provide an entry point for a TSO carrying three riboguanosines at its 3′-end. The bond cytosines-riboguanosines (deoxyribonucleotide-ribonucleotide) is key for a stable annealing and ensures that the MMLV-RT can "switch template," synthesizing a complementary DNA strand using the TSO as template. Therefore, TS and oligo-dT priming introduce known sequences at both ends of each transcript, allowing PCR amplification with a single pair of primers.

In FLASH-seq, we combined these two steps in a single RT-PCR reaction, which resulted in a faster, more efficient, and easy-to-use protocol, also thanks to a novel combination of enzymes, buffers, and additives [5]. Like Smart-seq2/3, FLASH-seq can be performed entirely with off-the-shelf reagents keeping the costs around $1/cell. We have shown that FLASH-seq can be easily miniaturized and automated by using the I.DOT (Dispendix) nanodispenser and the Fluent 780 workstation (Tecan) [5]. Several other instruments are most likely going to result in high-quality data; therefore, the choice is rather based on what is available in each lab. We always recommend choosing an automated solution whenever possible, in order to improve reproducibility, increase processing speed, minimize reagent waste and, ultimately, reduce

personnel and reagents' cost. However, if automated solutions are not available in the lab, then all FLASH-seq steps can be carried out in larger volumes by using manual pipettes and still returning data of satisfactory quality.

Like all Smart-seq methods, FLASH-seq also requires the cDNA to be further prepared for sequencing, a step carried out by a hyperactive Tn5 transposase. We leave it to the reader to choose which Tn5 (commercial or in-house produced) better suits the experimental needs. Commercial alternatives like the Nextera® XT kit (Illumina) are a good option for inexperienced users due to their robustness and consistency but come with a higher price tag. Alternatively, the plexWell kit (seqWell) offers the same ease of use and performance as the Nextera® XT kit but is significantly faster (especially when processing multiple plates at once) and has a lower price tag. Finally, the cheapest alternative remains the in-house Tn5 transposase, with all its advantages and disadvantages. Protein production might be challenging to optimize at first, but once it has been mastered and tagmentation protocols have been optimized [10, 11], it offers unparalleled flexibility and virtually unlimited room for customization.

Both the commercial and in-house Tn5 can be interchangeably used in the protocols below, as they will generate data of comparable quality. However, in the case of FS-UMI, we recommend using a commercial alternative to ensure higher reproducibility due to the lower batch-to-batch variation, making experiments more comparable even if performed months apart from each other.

A common limitation shared among Smart-seq2/3 and FLASH-seq is that all use an oligo dT-based strategy for priming exclusively polyadenylated RNAs, thus neglecting other potentially relevant RNA species such as microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and non-polyadenylated long non-coding RNAs (lncRNAs). If retaining strand-specificity is key for the experiment at hand, like when analyzing overlapping genes transcribed by opposite strands, then FS-UMI should be preferred, as the generated UMI reads will be stranded. Although technically easier and faster, both FS and FS-LA will not retain strand information, similar to what happens with Smart-seq2.

We believe that FLASH-seq has the potential to become the tool of choice when looking for an efficient, robust, modular, affordable, and automation-friendly full-length scRNA-seq protocol.

2 Materials

- All solutions should be prepared by using RNase- and DNase-free water and analytical grade reagents.
- It is very important to work as cleanly as possible, especially when the samples are still at the RNA stage.
- All working surfaces, racks, and pipettes should be wiped with 0.5% sodium hypochlorite (NaClO) followed by a rinse with nuclease-free water. Do not neglect this last step for pipettes, to avoid corrosion and damage over time.
- Use separate pre- and post-PCR working areas, in order to avoid contaminations. Be especially careful when handling the index primers used in the final enrichment PCR step.
- All reagents except enzymes can be thawed at room temperature. All master mixes can be briefly vortexed after preparation. However, do not vortex the enzyme stock solutions but rather mix them by (gentle) inversion.

Standard FLASH-seq (FS)

2.1 Cell Lysis Mix

1. 10% v/v Triton X-100. Store at +4 °C (*see* **Note 1**).
2. Pre-mixed dNTP solution (25 mM each). Store at −20 °C.
3. Recombinant RNase inhibitor (40 U/μL). Store at −20 °C.
4. Betaine (5 M). Store at +4 °C.
5. Dithiothreitol (DTT, 100 mM; part of the SuperScript™ IV kit). Store at −20 °C.
6. dCTP (100 mM). Store at −20 °C.
7. Template-switching oligonucleotide, FS-TSO (5′ Bio- AAG CAGTGGTATCAACGCAGAGTAC rGrGrG-3′, 100 μM). "Bio" = biotin. Store working aliquots at −20 °C and stock tubes at −80 °C (*see* **Note 2**).
8. SMART dT_{30}VN oligonucleotide (5′Bio-AAGCAGTGGTAT CAACGCAGAGTACT $_{30}$VN-3′, 100 μM). "Bio" = biotin. Store at −20 °C.
9. Nuclease-free water. Store at room temperature.

2.2 RT-PCR Mix

1. Superscript™ IV kit: dithiothreitol (DTT, 100 mM). Store at −20 °C.
2. Superscript™ IV kit: Superscript™ IV reverse transcriptase (200 U/μL). Store at −20 °C (*see* **Note 3**).
3. Betaine (5 M). Store at +4 °C.
4. Magnesium chloride ($MgCl_2$, 1 M). Store at +4 °C.

5. Recombinant RNase inhibitor (40 U/μL). Store at −20 °C.
6. KAPA HiFi HotStart ReadyMix (2×). Store at −20 °C.
7. Nuclease-free water. Store at room temperature.

2.3 Magnetic Beads Preparation

Here, we exclusively use a Sera-Mag SpeedBeads™ solution containing 18% w/v polyethylene glycol (PEG), prepared as described below (*see* **Note 4**). For 50 ml of working bead solution:

1. Withdraw 1 ml of Sera-Mag SpeedBeads™ suspension (carboxyl magnetic beads, hydrophilic, 5% suspension) and transfer it into a 1.5 ml tube.
2. Pellet the beads by placing the tube on a magnetic stand, wait until the solution is clear, and discard the supernatant.
3. Add 1 ml 10 mM Tris–HCl pH 8.0 and 1 mM EDTA (TE buffer) and resuspend the beads off the magnet. Pellet the beads again, wait until the solution is clear, discard the supernatant, and repeat one more time.
4. Pellet the beads once more, wait until the solution is clear, discard the supernatant, and resuspend off the magnet with 0.9 ml TE buffer.
5. In a beaker, mix 2.92 g NaCl, 500 μL Tris–HCl pH 8.0 (1 M), 100 μL EDTA (500 mM), and 9.5 g polyethylene glycol (PEG, MW = 8000).
6. Dissolve completely the reagents with the help of a magnetic stirrer and by heating the mix to 37 °C.
7. Add 50 μL Tween-20 (10% v/v, (*see* **Note 5**)), 250 μL sodium azide (NaN_3, 10% w/v (*see* **Note 6**)), and the cleaned-up beads prepared in **step 4** to the PEG solution.
8. Adjust the volume to 50 ml with nuclease-free water. Store at +4 °C. Do not freeze (*see* **Note 7**).

2.4 Library Preparation with Commercial Illumina Reagents

1. Nextera® XT DNA sample preparation kit. Store at −20 °C.
2. Nextera® XT index kit v2. Store at −20 °C (*see* **Note 8**).

2.5 Library Preparation with In-House Tn5 Transposase

In-house Tn5 transposase (~20 μM stock solution, diluted 1:10 with storage buffer according to [10]), pre-loaded with the following oligonucleotides: Tn5MErev: 5′-[phos]CTGTCTCTTATACACATCT-3′. "Phos" = phosphate; this oligonucleotide should be annealed either with Tn5ME-A: 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′ or with Tn5ME-B: 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′.

Store working aliquots at −20 °C. Store stock solution at −80 °C to preserve transposase activity in the long term (*see* **Note 9**).

1. 5× TAPS-$MgCl_2$ buffer: 50 mM TAPS-NaOH pH 7.3 at 25 °C, 25 mM $MgCl_2$. Adjust the pH as indicated above with sodium hydroxide (NaOH). Filter with 0.22 μm filters and store at +4 °C.
2. Dimethylformamide (98–100% v/v): store at room temperature in a ventilated cabinet and away from light (*see* **Note 10**).
3. 0.2% w/v SDS: store at room temperature (*see* **Note 11**).
4. KAPA HiFi non-HotStart kit: includes KAPA HiFi DNA polymerase (1 U/μL), KAPA HiFi high-fidelity buffer (5×), and dNTP mix (10 mM each). Store at −20 °C.
5. Nuclease-free water. Store at room temperature.
6. Additional index primer sets. Store at −20 °C (*see* **Note 12**).

2.6 Sample QC and Sequencing

1. Qubit™ dsDNA high-sensitivity assay kit. Store at +4 °C protected from light.
2. Qubit™ assay tubes.
3. Quant-iT™ PicoGreen™ dsDNA assay kit. Store at +4 °C protected from light.
4. Nunc™ F96 MicroWell™ polystyrene plate, black.
5. Agilent high-sensitivity DNA assay. Store at +4 °C.
6. Sera-Mag SpeedBeads™ working solution (for details about preparation, *see* Subheading 2.3). Store at +4 °C.
7. 80% v/v ethanol. Store at room temperature.
8. Nuclease-free water. Store at room temperature.
9. NextSeq™ 550 high output v2.5 kit (75 or 150 cycles), NextSeq™ 1000/2000 (100 or 200 cycles), or NovaSeq 6000 (100 or 200 cycles), depending on sample throughput, desired yield, and required read mode. Other instruments and kits can also be used. Follow the manufacturer's instructions regarding storage.

2.7 Instruments and Consumables

Depending on the throughput, the protocol can be carried out in single PCR tubes, 96- or 384-well plates. Below we list the consumables required for an experiment in 384-well plates.

1. Fully skirted 384-well plates (*see* **Note 13**).
2. Aluminum seal and adhesive plastic foil (*see* **Note 14**).
3. Plate centrifuge.
4. Thermocycler with 384-well block and heated lid set to 105 °C for all programs.

5. I.DOT (Dispendix) nanodispenser. Other (contact as well as non-contact) nanodispenser can also be used.
6. Fluent 780 workstation (Tecan). Other liquid handling robots can also be used.
7. 384 post-magnet plate (Alpaqua).
8. FACSAria™ Fusion (BD Biosciences). Other cell dispensers can also be used.
9. Agilent 2100 Bioanalyzer.
10. Nextseq™ 550/1000/2000 or Novaseq™ 6000 Sequencing Systems, depending on sample throughput and desired read output. Other Illumina instruments can also be used.

3 Methods

3.1 Cell Lysis Mix Preparation

1. For each sample, prepare 1 μL of the following cell lysis mix: 0.020 μL Triton X-100, 0.240 μL dNTP Mix, 0.018 μL SMART $dT_{30}VN$ oligonucleotide, 0.030 μL recombinant RNase inhibitor, 0.012 μL DTT, 0.092 μL FS-TSO, 0.090 μL dCTP, 0.200 μL betaine, Add nuclease-free water reach the final indicated volume, see above.
2. Mix well and gently spin down.
3. Dispense the cell lysis mix into a 384-well plate.
4. If used immediately, keep the plate on ice until needed. The cell lysis mix can be prepared and dispensed on plates several months in advance without any performance loss. The plates can be stored at −20 °C until needed. Avoid multiple freezing-thawing cycles.
5. Collect the cells by FACS using a FACSAria™ Fusion, preferably with a larger nozzle (85 μm or, even better, 100 μm) to minimize shearing forces upon sorting (*see* **Note 15**).
6. Once the sorting is completed, seal the plate with aluminum seal and snap-freeze it at −80 °C or by placing it on dry ice, especially if not proceeding immediately with RT-PCR. Lysed cells can be stored in these conditions for several months without appreciable decrease in RNA quality. The plate should never undergo freeze-thaw cycles for any reason.

3.2 RT-PCR Mix Preparation and RT-PCR Reaction

1. For each sample, prepare 4 μL of the following RT-PCR mix: 0.238 μL DTT, 0.800 μL betaine, 0.046 μL $MgCl_2$, 0.096 μL recombinant RNase inhibitor, 0.050 μL Superscript™ II reverse transcriptase, 2.500 μL KAPA HiFi HotStart Ready-Mix, Add nuclease-free water reach the final indicated volume, see above.
2. Briefly vortex and gently spin down the mix. Keep it on ice until needed.

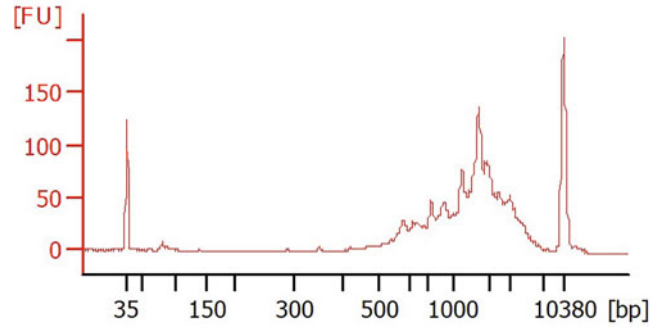
3. Take the plate out of the −80 °C freezer and leave it for a couple of minutes at room temperature to thaw.
4. Briefly spin down to collect eventual drops that might have condensed on the lid.
5. Place the plate on a thermocycler block and perform cell lysis and mRNA denaturation at 72 °C for 3 min.
6. Remove from the thermocycler and keep it on ice for a couple of minutes to cool. Briefly spin down to collect the lysate at the bottom.
7. Dispense 4 μL of RT-PCR Mix into each well. The final volume is now 5 μL.
8. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the RT-PCR reaction: 50 °C for 60 min, 98 °C for 3 min, then cycles of (98 °C for 20 s, 67 °C for 20 s, 72 °C for 6 min), 15 °C hold. The number of cycles should be adjusted according to the cell type used in the experiment. We recommend 18–19 cycles for HEK 293 T cells and 21 cycles for hPBMC. As a rule of thumb, start with 1–2 fewer PCR cycles compared to Smart-seq2. Avoid over-amplification of the cDNA.
9. It is safe to stop here and store the PCR product in a −20 °C freezer until needed.

3.3 Magnetic Beads Cleanup After Preamplification

Remove the Sera-Mag SpeedBeads™ working solution from the +4 °C storage and equilibrate it at room temperature for 15 min.

1. **OPTIONAL**: We generally add 10 μL of nuclease-free water into each well with the I.DOT to increase the reaction volume and make sample handling a bit easier when the cleanup is performed on the Fluent 780 workstation.
2. Add a 0.8× ratio of Sera-Mag SpeedBeads™ working solution (0.8:1 = beads: cDNA) and pipet up and down 20–30 times to mix.
3. Incubate off the magnetic stand for 5 min at room temperature.
4. Place the plate on the magnetic stand and leave it there for 5 min or until the solution appears clear.
5. Carefully remove the supernatant without disturbing the beads. Proceed immediately to the next step.
6. Remove the plate from the magnetic stand, add 15 μL nuclease-free water, and mix well by pipetting or vortexing to resuspend the beads.
7. Incubate 2 min off the magnetic stand.

8. Place the plate back on the magnetic stand and incubate for 2 min or until the solution appears clear.
9. Carefully remove 14 µL of the supernatant trying to minimize the bead carryover and transfer it to a new plate. It is safe to stop here and store the cDNA in a −20 °C freezer until needed (*see* **Note 16**).
10. Check the cDNA quality on the Agilent Bioanalyzer instrument. Follow the instructions as described in the high sensitivity DNA chip user manual. FLASH-seq yields generally fragments with an average size of 1.8–2.2 Kb and a near-to-complete absence of fragments <500 bp.



Example of amplified cDNA from a single hPBMC (21 PCR cycles)

3.4 cDNA Quantification and Sample Normalization

The preamplified cDNA is generally too concentrated and needs to be diluted before proceeding with tagmentation.

1. Allow the Quant-iT PicoGreen reagent to reach room temperature before opening the vial. PicoGreen is light sensitive; while thawing, wrap in aluminum foil.
2. Prepare a 1× working solution of TE using 20× TE (supplied) and nuclease-free water.
3. Prepare a 1:400 dilution of PicoGreen solution and always use a plastic vessel. Do not use glass, as PicoGreen may adsorb to glass.
4. Prepare the standard curve using the Lambda DNA standard (supplied with the PicoGreen kit at a concentration of 100 ng/µL) and 1× TE in 8 tubes, as below. The stock tubes can be used multiple times, so keep any leftover in the fridge at +4 °C between experiments.
5. Vortex well and spin down the DNA standards before every use. Not vortexing thoroughly the standards is going to negatively affect the standard curve and your readings! Serial dilutions should be prepared as shown in the table below.

Tube no.	Content	Concentration	Final volume
1	90 µL TE + 10 µL Lambda DNA stock	10 ng/µL	100 µL
2	50 µL from Tube 1 + 50 µL TE	5 ng/µL	100 µL
3	50 µL from Tube 2 + 50 µL TE	2.5 ng/µL	100 µL
4	50 µL from Tube 3 + 50 µL TE	1.25 ng/µL	100 µL
5	50 µL from Tube 4 + 50 µL TE	0.625 ng/µL	100 µL
6	50 µL from Tube 5 + 50 µL TE	0.3125 ng/µL	100 µL
7	50 µL from Tube 6 + 50 µL TE	0.15625 ng/µL	100 µL
8	TE only	blank	–

6. Prepare the PicoGreen solution by pipetting 0.5 µL PicoGreen dye +99.5 µL 1× TE for each sample. Vortex to mix.
7. Pipette 1 µL of each of the 7 standards +1 Blank into a black, flat-bottom Nunc™ F96 MicroWell™ plate. Place the standards on one column.
8. Pipet 1 µL of your samples into the center of each well of the Nunc™ F96 MicroWell™ polystyrene plate.
9. Add 99 µL PicoGreen + TE mix into every well. There is no need to mix.
10. Cover the plate with the provided plastic (transparent) lid to prevent possible contaminations.
11. Allow 2 min for the dye to bind the DNA. Protect from light. For optimal results, read the plate within 1 h.
12. Use a plate reader to measure fluorescence (excitation: 485 nm; emission: 530 nm; read from top; endpoint reading).
13. Prepare a normalization plate by adding 1 µL purified cDNA and nuclease-free water to a final concentration of 150 pg/µL using the PicoGreen readings.

3.5 Library Preparation

This step can be carried out by using either the commercially available Nextera XT kit or the in-house Tn5 transposase. Both protocols are described below.

Indexing primers can be purchased from Illumina or ordered from your local oligo manufacturer.

3.5.1 Library Preparation with the Nextera® XT Kit

1. Remove the Amplicon Tagment Mix (ATM) and the Nextera® PCR mix (NPM) from the −20 °C storage and place them on ice. Remove the Tagment DNA buffer (TD) and the adaptor index plate from the −20 °C storage and keep them at room

temperature. Remove the neutralization buffer (NT) from the +4 °C storage and keep it at room temperature.

2. For each sample, prepare 1.25 μL tagmentation mix: 0.25 μL ATM and 1 μL TD.
3. Briefly vortex and gently spin down the mix. Keep it on ice until needed.
4. Transfer 1 μL of normalized cDNA into a new 384-well plate.
5. Add 1.25 μL of tagmentation mix to each well.
6. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the tagmentation reaction: 55 °C for 8 min, 4 °C hold. Once the reaction is completed, proceed immediately to the next step.
7. Add 0.5 μL NT buffer to each well.
8. Seal the plate with adhesive transparent film, vortex, spin down, and incubate for 5 min at room temperature. Do not put the plate back on ice.
9. Add 1 μL N7xx + S5xx indices from the adapter index plate (5 μM each).
10. Add 1.5 μL NPM to each well.
11. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the enrichment PCR reaction: 72 °C for 3 min, 95 °C for 30 s, then N cycles of (95 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min, 4 °C hold. The number of cycles "N" should be adjusted according to the amount of cDNA used in the tagmentation and the number of cells that are going to be pooled before cleanup. When starting from 150 pg input cDNA, 14 cycles are sufficient even if pooling only 32–48 cells.
12. It is safe to stop here and store the final library in a −20 °C freezer until needed.

3.5.2 Library Preparation with In-House Tn5 Transposase

Please note that the Tn5 amount indicated below is merely a starting point for the tagmentation of 150 pg cDNA. Optimization might be necessary, depending on the specific activity of each batch of Tn5. If needed, reactions can be miniaturized, paying attention to keep the same ratio of cDNA and Tn5, in order to avoid under- or over-tagmentation.

1. Remove the Tn5 transposase and the KAPA HiFi polymerase from the −20 °C storage and place them on ice. Remove the dNTP mix, high-fidelity KAPA HiFi buffer and the adapter index plate from the −20 °C storage, and keep them at room temperature. Remove the TAPS-$MgCl_2$ buffer from the +4 °C storage and keep it at room temperature. Get the SDS and the DMF solutions.

2. For each sample, prepare 3 μL tagmentation mix: 0.8 μL TAPS-$MgCl_2$ buffer, 0.8 μL DMF, 0.125 μL in-house Tn5 transposase, nuclease-free water to volume.
3. Briefly vortex and gently spin down the tagmentation mix. Keep it on ice until needed.
4. Transfer 1 μL normalized cDNA in a new 384-well plate.
5. Add 3 μL tagmentation mix to each well.
6. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the tagmentation reaction: 55 °C for 8 min, 4 °C hold. Once the reaction is completed, proceed immediately to the next step.
7. Add 1 μL 0.2% SDS solution to each well.
8. Seal the plate with adhesive transparent film, vortex, spin down, and incubate for 5 min at room temperature. Do not put the plate back on ice.
9. Add 2 μL N7xx + S5xx indices from the adaptor index plate (5 μM each).
10. Add 3 μL enrichment PCR mix: 0.2 μL KAPA HiFi DNA polymerase, 0.3 μL dNTP mix, 2 μL high-fidelity buffer, nuclease-free water to volume.
11. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the enrichment PCR reaction: 72 °C for 3 min, 98 °C for 30 s, then N cycles of (98 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min, 4 °C hold. The number of cycles "N" should be adjusted according to the amount of cDNA used for the tagmentation reaction. When starting from 150 pg input cDNA, 14 cycles are sufficient even if pooling only 32–48 cells.
12. It is safe to stop here and store the final library in a −20 °C freezer until needed.

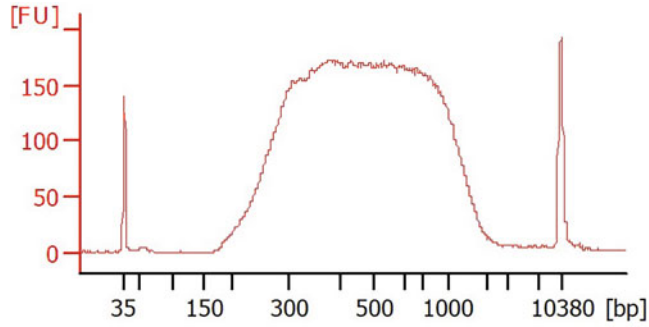
3.6 Pooling and Bead Cleanup of the Final Library

1. Remove the Sera-Mag SpeedBeads™ working solution from the +4 °C storage and equilibrate it at room temperature for 15 min.
2. Pool the entire volume or just part of the tagmentation reaction from each well into a 2 ml LoBind tube. Vortex to mix and briefly spin down.
3. Add a 0.8× ratio of Sera-Mag SpeedBeads™ working solution, vortex thoroughly, and incubate for 5 min off the magnetic stand.
4. Place the tube on the magnetic stand and leave it there for 5 min or until the solution appears clear.
5. Carefully remove the liquid paying attention not to disturb the beads.

6. Add 1 ml 80% w/v ethanol and incubate for 1 min without removing the tube from the magnet.
7. Remove any trace of ethanol and let the bead pellet dry for 2 min (tube always on the magnetic stand and with the lid open!).
8. Remove the tube from the magnetic stand; add 100 μL nuclease-free water and mix well to resuspend the beads.
9. Incubate for 2 min off the magnetic stand.
10. Place the tube back on the magnetic stand and leave it there for 2 min or until the solution appears clear.
11. Carefully remove the supernatant trying to minimize the bead carryover and place it in a new 1.5 ml LoBind tube. This is the final pool that is going to be used for sequencing.

3.7 Sequencing Library QC and Sequencing

1. Use 1 μL of the purified sequencing pool to assess the concentration on a Qubit instrument.
2. Use 1 μL to assess the average size on a high-sensitivity DNA chip.



Example of sequencing-ready library (pool of 384 HEK 293 T cells)

3. Single-end 75 or 100 bp mode is generally sufficient but longer read modes or paired-end sequencing can be an option, depending on the question at hand.

4 Data Preprocessing

The data processing will depend on the final goal. The following description provides an example of data preprocessing. We assume that the data were generated using single-end sequencing (75 bp) on an Illumina instrument and are processed on a unix-based machine.

4.1 Prerequisites

Download the following tools. Versions indicated are the ones tested in *Hahaut* et al. [5].

- bcl2fastq (Illumina, v2.20)
- STAR [12] (v2.7.3)
- FeatureCounts [13] (v1.6.5)
- BBMAP [14] (v38.86)
- samtools [15] (v1.9)
- IGV

When working on hundreds of samples, it is recommended to use virtual machines or cloud platforms to process the data, as it will typically require >48 Gb of RAM, >12 cores, and >500 Gb of free space. The footprint of the most time- and resource- consuming step (= read alignment) can be reduced using pseudoalignment tools (i.e., salmon [16] or kallisto [17]), not described in this manuscript.

4.2 Optional: Demultiplexing

Skip this step if the sequencing data are directly provided as separated FASTQ files.

In case samples are sequenced on an illumina sequencer and only the raw sequencing files are provided (.blc2), bcl2fastq can be used for demultiplexing.

```
# 0. Define paths
BASECALL_DIR="/path/to/folder/Data/Intensities/BaseCalls/"
OUTPUT_DIR="/path/to/output_directory/"
SAMPLESHEET="/path/to/demultiplexing_sampleSheet.csv"

# 1. Run bcl2fastq
ulimit -n 10000
bcl2fastq --input-dir $BASECALL_DIR --output-dir $OUTPUT_DIR
--sample-sheet $SAMPLESHEET --create-fastq-for-index-reads --
no-lane-splitting
```

The sample sheet description is provided in the bcl2fastq manual. Alternatively, the Illumina Experiment Manager (IEM) software can be used to generate it.

After demultiplexing, check if any index combinations were missed using:

```
# 2. What are the remaining undemultiplexed indexes?
# Returns the number of reads associated with each index
combination in the Undetermined fastq file.
zcat "$OUTPUT_DIR"/Undetermined_S0_I1_001.fastq.gz | awk -F'
1:N:0:' 'NR%4==1{print $2}' | sort | uniq -c > left_index.txt
sort -k1,1 left_index.txt
```

The top I5 + I7 combinations should be associated with a limited number of unassigned reads. When sequencing on a NextSeq500 instrument, the undetermined index combinations often contain "GGGGGGGG" I5 or I7 indexes. These represent indexes that could not be read by the sequencer (= dark cycles).

Finally, the number of reads per sample can be obtained with:

```
for file in "$OUTPUT_DIR"/*R1*
do
echo $(zcat $file|wc -l)/4|bc
done
```

The read distribution should be relatively uniform. Large variations can indicate sample or library pooling issues.

4.3 Optional: Remove Sequencing Adapter Leftovers

In case sequencing adapter leftovers are observed, bbduk [14] or trimmomatic [18] can be used to trim them.

```
# BBDUK sequencing adapters
# Provided with bbmap
BBDUK_REF="/path/to/BBDUK/adapters.fa"

# Remove left-over adapter (left-side)
$BBDUK -Xmx12g in=sample.fastq.gz out=cleaned.left.fastq t=32
ktrim=l ref="$BBDUK_REF" k=23 mink=7 hdist=1 hdist2=0 tbo
# Remove left-over adapter (right-side)
$BBDUK -Xmx12g in=cleaned.left.fastq out=cleaned.fastq t=32
ktrim=r ref="$BBDUK_REF" k=23 mink=7 hdist=1 hdist2=0 tbo
```

4.4 Data Alignment

4.4.1 Genome Indexing

Read aligners (e.g., STAR or HISAT2 [19]) or pseudo-aligners (e.g., salmon or kallisto) are used to align the data against a reference genome/transcriptome. The following example is demonstrated using STAR.

The first step of the data alignment process is to generate an index for the reference genome. This procedure is generally performed only once.

```
# 0. Define paths
OUTPUTREF="/path/to/output_reference_index/"
FASTA="/path/to/fasta_reference_genome.fa"
GTF="/path/to/genome_annotation.gtf"

# 1. Create the index repository
mkdir $OUTPUTREF
cd $OUTPUTREF

# 2. OPTIONAL: Save the fasta/GTF files to the reference folder
scp $FASTA .
scp $GTF .
```

```
# 3. Genome indexing
# runThreadN: Adapt the number of threads to your device
# sjdbOverhang: Usually set to read length -1
STAR --runThreadN 30 \
--runMode genomeGenerate \
--genomeDir $OUTPUTREF \
--genomeFastaFiles $FASTA \
--sjdbGTFfile $GTF \
--sjdbOverhang 74
```

The reference genome/transcriptome and gene annotation files can be obtained from ENSEMBL, UCSC, or gencode.

4.4.2 Align Reads

```
MYFASTQ=”/path/to/fastq_file.fq.gz”

mkdir STAR
STAR --runThreadN 10 --limitBAMsortRAM 20000000000 --genome-
Load LoadAndKeep --genomeDir "$OUTPUTREF" --readFilesIn "$MY-
FASTQ" --readFilesCommand zcat --limitSjdbInsertNsj 2000000 --
outFilterIntronMotifs RemoveNoncanonicalUnannotated --outSAM-
type BAM SortedByCoordinate --outFileNamePrefix STAR/"$ID"_
```

This command will output an alignment file (.bam) and a summary of the alignment results (.log.final.out)

4.4.3 Optional: Filter Reads

Even though they should not influence the rest of the pipeline, unmapped reads (sam flag 4) and secondary alignments (sam flag 256) can be discarded using samtools.

```
samtools view -@ 10 -Sb -F 260 STAR/"$ID"_Aligned.sortedBy-
Coord.out.bam > STAR/"$ID"_Aligned.sortedByCoord.filtered.bam
samtools index STAR/"$ID"_Aligned.sortedByCoord.filtered.bam
```

4.5 Assign Reads to Feature Using Featurecounts

Sequencing libraries are generated by tagmentation, which randomly cut and tag cDNA molecules. The reads generated in FLASH-seq are therefore unstranded (-s 0). Given the mild non-ionic detergent used during cell lysis, nuclear unspliced mRNAs molecules are not expected to be efficiently captured by FLASH-seq. We therefore advise counting reads assigned to each feature using only the exonic features (-t exon).

```
mkdir FEATURECOUNTS

featureCounts -T 1 -t exon -g gene_name --fracOverlap 0.25 -a
"$GTF" -o FEATURECOUNTS/"$ID"_ReadCount.featureCounts.txt
STAR/"$ID"_Aligned.sortedByCoord.filtered.bam
```

4.6 Recommended: Quality Checkups

FLASH-seq is a full-length RNA-sequencing protocol. The gene-body coverage should be uniformly distributed from 5′ to 3′. Deviations can, among others, indicate RNA degradations or reverse transcription issues.

4.6.1 Gene-Body Coverage (ReSQC)

```
# BED12 Reference file.
# This file should contain the same information as the GTF file
used for read alignment / feature assignment.
ReSQCBED="/path/to/annotation.bed"

mkdir ReSQC

geneBody_coverage.py -r "$ReSQCBED" -i STAR/"$ID"_Aligned.
sortedByCoord.filtered.bam -o ReSQC/"$ID"_geneBody.all

Warning: This step is single-threaded (= time-consuming) and
requires a lot of RAM.
```

4.6.2 Read Distribution (ReSQC)

The majority of the reads should be assigned to the gene coding sequences followed by 5′ / 3′ UTRs and intronic sequences. Fewer reads should be assigned to intergenic regions (<10% read tags).

```
read_distribution.py -i STAR/"$ID"_Aligned.sortedByCoord.fil-
tered.bam -r "$ReSQCBED" > ReSQC/"$ID"_readDistribution.txt
```

4.6.3 STAR Mapping Statistics (STAR)

We highly recommend exploring the STAR "Log.final.out" file. It contains a summary of the mapping statistics of each sample.

Important values to monitor include the following:

- Uniquely mapped reads % [70–90%]
- Average mapped length [≃ average read length]
- Mismatch rate per base, % [<1.5%]
- Percentage of reads mapped to multiple loci + too many loci [<20%]
- Percentage of reads unmapped: too many mismatches + too short [<10%]

The thresholds are given as examples and can highly vary from one cell type to another. The best indicator of a good quality sample is a high percentage of uniquely mapped reads.

4.7 Recommended: Processing Multiple Samples

When dealing with multiple samples, it is advisable to run the steps described above in a loop and/or in parallel to speed up the analysis process. Dozens of alternative approaches can be used to perform this operation. The following line describes one of them:

```
# 1. Create a sample sheet containing the prefix of each sample
- one per line. As an example:

cd "$OUTPUT_DIR"
ls *R1*fastq.gz | perl -pe 's|_S.{1,3}_R1_001\.fastq\.gz||' >
mysamplesheet.txt
```

which results in:

```
head -n 3 mysamplesheet.txt

sample_1
sample_2
sample_3
```

This sample sheet can then be used to serially process each sample.

```
# Load the reference genome into memory to greatly improve
speed
STAR --genomeLoad LoadAndExit --genomeDir $OUTPUTREF --run-
ThreadN 12

cat mysamplesheet.txt |
while IFS=$'\t' read -r -a myArray; do

 # 0. Get the sample ID and create a repository
 ID="${myArray[0]}"
 echo "$ID"
 mkdir "$ID"
 cd "$ID"

 # 1. Get the FASTQs associated with this sample
 FASTQ=`ls "$OUTPUT_DIR"/${ID}*R1*fastq.gz`
 cp $FASTQ .
 cat *fastq.gz > sample.fastq.gz

 # 2. Align the data
 mkdir STAR
STAR --runThreadN 10 --limitBAMsortRAM 20000000000 --genome-
Load LoadAndKeep --genomeDir "$OUTPUTREF" --readFilesIn sam-
ple.fastq.gz --readFilesCommand zcat --limitSjdbInsertNsj
2000000 --outFilterIntronMotifs RemoveNoncanonicalUnannotated
--outSAMtype BAM SortedByCoordinate --outFileNamePrefix STAR/
"$ID"_
```

```
 # 3. Filter BAM file
samtools view -@ 10 -Sb -F 260 STAR/"$ID"_Aligned.sortedBy-
Coord.out.bam > STAR/"$ID"_Aligned.sortedByCoord.filtered.bam
 samtools index STAR/"$ID"_Aligned.sortedByCoord.filtered.bam

 # 4. Assign Reads to features
 mkdir FEATURECOUNTS
featureCounts -T 1 -t exon -g gene_name --fracOverlap 0.25 -a
"$GTF" -o FEATURECOUNTS/"$ID"_ReadCount.featureCounts.txt
STAR/"$ID"_Aligned.sortedByCoord.filtered.bam

 # 5. Quality check-ups
 mkdir ReSQC

geneBody_coverage.py -r "$ReSQCBED" -i STAR/"$ID"_Aligned.
sortedByCoord.filtered.bam -o ReSQC/"$ID"_geneBody.all

read_distribution.py -i STAR/"$ID"_Aligned.sortedByCoord.fil-
tered.bam -r "$ReSQCBED" > ReSQC/"$ID"_readDistribution.txt

done

# Unload STAR Reference
STAR --genomeLoad Remove --genomeDir "$OUTPUTREF"
```

5 Data Postprocessing

The postprocessing steps will vary depending on the question at hand. In our experience, most tools working for Smart-seq2 can be used as well with FLASH-seq data. For instance, differential expression can be performed with Seurat [20] (R), scanpy [21] (python), or scatter/scran [22] (R) packages/modules.

As FLASH-Seq does not contain UMI sequences, the sample's library size must be normalized.

6 FLASH-seq Low-Amplification (FS-LA)

We have previously shown that FLASH-seq guarantees a higher gene detection than Smart-seq2/3, thanks to its novel combination of buffers, enzymes, and additives [5]. By combining RT and PCR in a single step, FLASH-seq significantly shortens the reaction and hands-on time required for an experiment. In an attempt to further streamline the original protocol, we developed FLASH-seq low-amplification (FS-LA) which, to our knowledge, is the fastest full-length RNA-seq protocol ever designed.

We reasoned that the entire amount of cDNA generated after 20 or more PCR cycles is never entirely used in the tagmentation (even if the reaction has to be repeated multiple times!), causing unnecessarily longer waiting times and generating a higher number of PCR duplicates.

Therefore, we decreased the number of PCR cycles, aiming to obtain sufficient cDNA for direct tagmentation without intermediate steps. This amount of cDNA remains well below the detection limit of most instruments and cannot be reliably assessed, making intermediate QC unfeasible.

Another factor to take into account is that skipping the cleanup step requires a predilution of the cDNA ahead of tagmentation. Residual salt and primers might otherwise completely inhibit Tn5 transposase activity.

This can be achieved in two ways:

- Either by taking directly an aliquot of the unpurified cDNA and performing the tagmentation reaction in a larger volume, as we chose to do in FS-LA. When using our in-house Tn5 transposase, the costs associated with a larger reaction volume are negligible.
- Alternatively, the cDNA can be prediluted [23] and the tagmentation carried out in a smaller volume, which makes the use of commercial kits (more) financially sustainable. The use of in-house transposase would of course remain an option also in this case.

The removal of QC and normalization steps allowed us to carry out the entire FS-LA protocol, from cell sorting to sequencing-ready libraries, in just 4.5 h, of which <1 h represents hands-on time.

Data analysis of FS-LA samples shows an increasing number of detected genes associated with the lower number of PCR cycles as well as excellent gene coverage [5].

Our experiments showed that once a critical amount of cDNA has been generated, high-quality libraries can be prepared using the RT-PCR product directly. This threshold has to be determined empirically for each cell type in a pilot experiment. Elsewhere we provided general guidelines for avoiding under- or overamplification of the cDNA, which might negatively impact the tagmentation and, ultimately, data quality [5].

As a general guideline, the cell RNA content determines the number of PCR cycles that are required. Large cells, such as HEK 293T, contain about ~16 pg of total RNA, and 10 to 12 PCR cycles are sufficient to generate enough cDNA for direct tagmentation and high-quality libraries. On the other hand, human peripheral blood mononuclear cells (hPBMCs) typically contain 2–10 times

less RNA compared to HEK 293T cells and might require a higher number of PCR cycles.

6.1 Cell Lysis Mix

Same reagents as the standard FLASH-seq protocol (Fig. 1).

6.2 RT-PCR Mix

Same reagents as the standard FLASH-seq protocol.

6.3 Magnetic Beads Preparation

Same reagents as the standard FLASH-seq protocol.

6.4 Library Preparation with Commercial Illumina Reagents

Same reagents as the standard FLASH-seq protocol.

6.5 Library Preparation with In-House Tn5 Transposase

Same reagents as the standard FLASH-seq protocol.

6.6 Sample QC and Sequencing

1. Qubit™ dsDNA high-sensitivity assay kit. Store at +4 °C protected from light.
2. Qubit™ assay tubes.
3. Agilent high-sensitivity DNA assay. Store at +4 °C.
4. Sera-Mag SpeedBeads™ working solution (for details about preparation, *see* Subheading 2.3). Store at +4 °C.
5. 80% v/v ethanol. Store at room temperature.
6. Nuclease-free water. Store at room temperature.
7. NextSeq™ 550 high output v2.5 kit (75 or 150 cycles), NextSeq™ 1000/2000 (100 or 200 cycles), or NovaSeq 6000 (100 or 200 cycles), depending on sample throughput, desired yield, and required read mode. Other instruments and kits can also be used. Follow the manufacturer's instructions regarding storage.

6.7 Instruments and Consumables

To appreciate the speed and throughput that FS-LA can offer, we recommend performing the experiment only in 384-well plates.

1. Fully skirted 384-well plates (*see* **Note 13**).
2. Aluminum seal and adhesive plastic foil (*see* **Note 14**).
3. Plate centrifuge.
4. Thermocycler with 384-well block.
5. I.DOT (Dispendix) nanodispenser.
6. Fluent 780 workstation (Tecan).

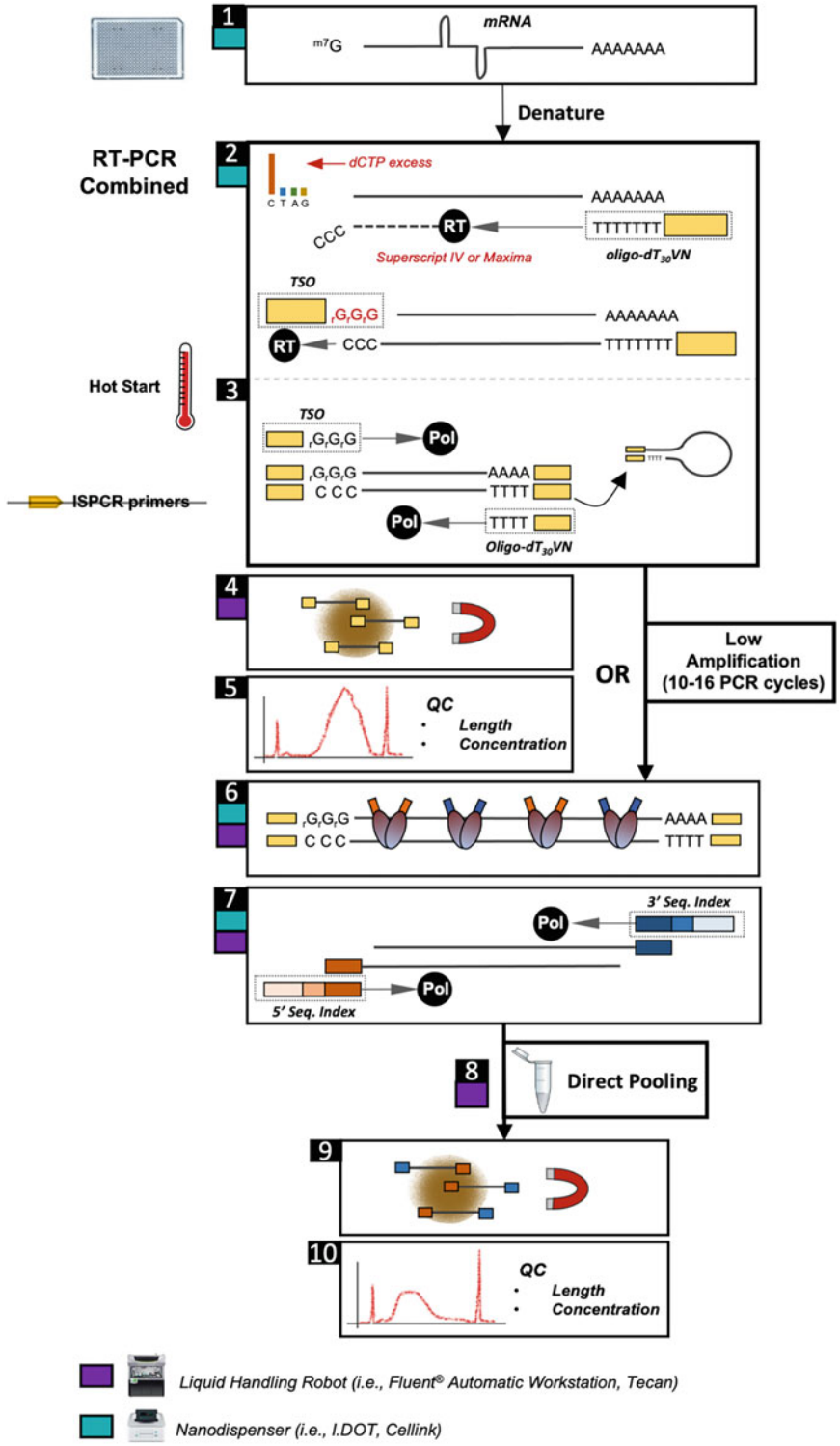


Fig. 1 Standard FLASH-seq workflow. Steps automated with a liquid-handling robot or nanodispenser are marked by purple and teal squares, respectively. **Step 1**, cells are sorted by FACS in wells of a plate containing a lysis buffer. **Step 2**, after cell lysis, the secondary structure of mRNAs is resolved by heating before priming with an oligodT primer and reverse transcription with a template-switching reverse transcriptase (RT) in the

7. FACSAria™ Fusion (BD Biosciences). Other cell dispensers can also be used.
8. Agilent 2100 Bioanalyzer.
9. Nextseq™ 550/1000/2000 or Novaseq™ 6000 Sequencing Systems, depending on sample throughput and desired read output. Other Illumina instruments can also be used.

7 Methods

7.1 Cell Lysis Mix Preparation

Follow the standard FLASH-seq protocol.

7.2 RT-PCR Mix Preparation and RT-PCR Reaction

Follow the standard FLASH-seq protocol.

IMPORTANT - Adjust the number of cycles according to the cell type. We recommend 10–12 cycles for HEK 293T cells and 14–16 cycles for hPBMC.

IMPORTANT - Do not perform QC (Bioanalyzer/PicoGreen) or sample normalization!

7.3 Library Preparation

7.3.1 Library Preparation with In-House Tn5 Transposase

Please note that the Tn5 transposase amount is a suggested starting point only. Optimizations might be necessary, depending on the specific batch activity and desired library size.

1. Remove the in-house Tn5 transposase (2 μM working dilution) and the KAPA HiFi polymerase from the −20 °C storage and place them on ice. Remove the dNTP mix, high-fidelity KAPA HiFi buffer, and the index adapter plate from the −20 °C storage and keep them at room temperature. Remove the TAPS-$MgCl_2$ buffer from the +4 °C storage and keep them at room temperature. Get the SDS and the DMF solutions.
2. For each sample, prepare 9 μL tagmentation mix: 2 μL TAPS-$MgCl_2$ buffer, 2 μL DMF, 0.025 μL in-house Tn5 transposase, nuclease-free water to volume.
3. Briefly vortex and gently spin down the tagmentation mix. Keep it on ice until needed.

Fig. 1 (continued) presence of an excess of dCTP. **Step 3**, cDNA is amplified by semi-suppressive PCR. For Steps 2 and 3, only one master mix is required. **Step 4**, cDNA is cleaned up using magnetic beads. **Step 5**, cDNA concentration and fragment size are measured and the samples are normalized to the desired concentration for the tagmentation reaction. **Step 6**, cDNA is tagmented using a Tn5, to make it amenable to Illumina sequencing. **Step 7**, tagmented cDNA is amplified by PCR and sequencing indices are added. **Step 8**, samples are pooled together to facilitate handling. **Step 9**, sequencing-ready library is cleaned up with magnetic beads. **Step 10**, concentration and average fragment size are measured in preparation for the sequencing. **Steps 4** and **5** can be skipped when using FS-LA

4. Transfer 1 μL unpurified cDNA into a new 384-well plate.
5. Add 9 μL tagmentation mix to each well.
6. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the tagmentation reaction: 55 °C for 8 min, 4 °C hold. Once the reaction is completed, proceed immediately to the next step.
7. Add 2.5 μL 0.2% SDS solution to each well.
8. Seal the plate with adhesive transparent film, vortex, spin down, and incubate for 5 min at room temperature. Do not put the plate back on ice.
9. Add 2.5 μL N7xx + S5xx indices from the adapter index plate (5 μM each).
10. Add 10 μL enrichment PCR mix: 0.5 μL KAPA HiFi DNA polymerase, 0.75 μL dNTP mix, 5 μL high-fidelity buffer, nuclease-free water to volume.
11. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the enrichment PCR reaction: 72 °C for 3 min, 98 °C for 30 s, then N cycles of (98 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min, 4 °C hold. The number of cycles "N" should be adjusted according to the number of preamplification cycles used in the experiment. As a general guideline we found that 14–16 cycles are sufficient for most applications, especially when pooling 384 cells before the library QC.
12. It is safe to stop here and store the final library in a −20 °C freezer until needed.

7.3.2 Library Preparation with the Nextera® XT Kit

1. Remove the Amplicon Tagment Mix (ATM) and the Nextera® PCR mix (NPM) from the −20 °C storage and place them on ice. Remove the Tagment DNA buffer (TD) and the adaptor index plate from the −20 °C storage and keep them at room temperature. Remove the neutralization buffer (NT) from the +4 °C storage and keep it at room temperature.
2. Take an aliquot of the preamplified and unpurified cDNA and transfer it into a new 384-well plate.
3. Add nuclease-free water to obtain a 1:10 dilution. This represents the cDNA that is going to be used for tagmentation.
4. Prepare the tagmentation mix: 1 μL TD and 0.02–0.25 μL ATM. Refer to the table below for estimating the most appropriate amount of ATM to use.

No. pre-amplification cycles	Amount ATM used (μl)	Average library size Bioanalyzer (bp)
10	0.25	460
10	0.125	498
10	0.0625	657
10	0.03125	794
10	0.015625	931
10	0.0078125	998
10	0.00390625	1069
10	0.001953125	1020
12	1	448
12	0.5	507
12	0.25	567
12	0.125	644
12	0.0625	735
12	0.03125	822
12	0.015625	1028
12	0.0078125	1085
12	0.00390625	1027

The table shows that the approximate library size is to be expected when using 1:10 diluted cDNA after 10 or 12 preamplification cycles in HEK 293T cells

5. Briefly vortex and gently spin down the mix. Keep it on ice until needed.
6. Transfer 1 μL diluted cDNA into a new 384-well plate.
7. Add 1 μL tagmentation mix to each well.
8. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the tagmentation reaction: 55 °C for 8 min, 4 °C hold. Once the reaction is completed, proceed immediately to the next step.
9. Add 0.5 μL NT buffer into each well.

10. Seal the plate with adhesive transparent film, vortex, spin down, and incubate for 5 min at room temperature. Do not put the plate back on ice.
11. Add 1 μL N7xx + S5xx indices from the adapter index plate (5 μM each).
12. Add 1.5 μL NPM into each well.
13. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the enrichment PCR reaction: 72 °C for 3 min, 95 °C for 30 s, then N cycles of (95 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min, 4 °C hold. The number of cycles "N" can vary according to several parameters (number of preamplification cycles used, RNA content, number of cells that are being processed), but we found that 12–14 cycles are generally more than sufficient, especially when pooling 384 cells before the final cleanup.
14. It is safe to stop here and store the final library in a −20 °C freezer until needed.

7.4 Pooling and Bead Cleanup of the Final Library

Follow the standard FLASH-seq protocol, regardless of which Tn5 was used for tagmentation in the steps above.

7.5 Sequencing Library QC and Sequencing

Follow the standard FLASH-seq protocol. Single-end 75 or 100 bp is generally sufficient but longer read modes or paired-end sequencing can be an option, depending on the question at hand.

8 Data Processing

Refer to the standard FLASH-seq protocol for the data processing. Particular attention must be given to the data quality checkups. As described in *Hahaut* et al. [5], performing a limited number of PCR cycles can influence data quality. We observed that an insufficient number of PCR cycles increases the percentage of intergenic and intronic reads as well as the percentage of multimapped/unmapped reads.

This is likely due to the balance between amplified cDNA, tagmented genomic DNA, and ribosomal RNAs, which are always present in the reaction. In Smart-seq2, the cDNA overamplification and the intermediate cleanup masked these contaminants. However, in FS-LA, the right number of PCR cycles must be chosen carefully, in order to obtain a sufficient amount of cDNA and avoid compromising the final data.

To ensure that the data generated are of good quality, we advise the user to compare FS-LA data with a control FS sample done with the standard number of PCR cycles.

Insufficient cDNA amplification can be identified by:

- Higher intergenic/intronic read tag percentages (ReSQC)
- Lower exonic read tag percentages (ReSQC)
- Lower uniquely mapping read percentages (STAR)
- Increased multimapped/unmapped read percentages (STAR)

As the baseline of these mapping statistics vary from one cell type to another, the thresholds may vary as well. We typically tolerate a variation of 2–5% in the percentage of uniquely mapped reads and <2% in the intergenic reads between standard FS and FS-LA values.

In addition, we recommend exploring the aligned data with a genome browser viewer such as Integrated Genome Viewer (IGV). Given the nature of the FS protocol (polyA-based), most of the reads should be confined to exons and introns of protein coding genes, with clear exon/intron boundaries. High levels of single reads in intergenic, intronic, or centromeric regions may indicate that the cDNA was not sufficiently amplified. We also observed that in lower quality samples, the intergenic reads tended to be regularly spaced, at least to some extent.

We do not recommend doing isoform reconstruction/quantification in the presence of genomic DNA contamination. However, the tagmentation is a random process which means that low levels of genomic DNA contamination are unlikely to influence differential expression results at the gene/exon level as the same DNA regions will not be systematically sequenced in every cell.

9 FLASH-seq with Unique Molecular Identifiers (FS-UMI)

To exploit the robustness and flexibility of FLASH-seq even further, we modified the original protocol by including UMIs into the TSO sequence, as originally introduced by Smart-seq3 [3], thus generating both 5′ UMI-containing reads and internal reads without UMI. However, based on existing literature, we also realized that the eight random nucleotides in the UMI followed by three ribo-G (Smart-seq3 TSO design) might increase the risk of strand invasion [24].

Strand invasion is a phenomenon occurring when the TSO anneals to an internal sequence of the cDNA or mRNA molecule that, by chance, displays sequence complementarity [5]. The end result is a molecule that is shorter and with a chimeric 5′-end. Although not always easy to exactly quantify, these artifacts should be minimized whenever possible, as they affect isoform detection and bias gene counts.

In a similar fashion to the solution implemented in the nanoCAGE protocol and the 10x Genomics Single Cell 5′ v1/v2 kits, we tested seven new TSO where the UMI and the ribo-Gs were

separated by a 5-bp spacer sequence. These new TSOs were tested in combination with the STRT-seq2i-oligo-dT [25], with the pre-amplification reaction deviating from the suppression PCR used by Smart-seq2 and the standard FLASH-seq protocol.

In our experiments we observed that the cDNA yield was strongly influenced by the TSO sequence. Interestingly, most conditions displayed great mapping statistics and coverage [5]. The addition of a spacer sequence increased by up to ~15% (TSO-CATCA) the number of detected genes compared to SS3. Six out seven TSOs with a spacer even outperformed the standard FS in respect of the number of detected genes. We eventually settled on a spacer carrying the sequence "CTAAC," as it also showed excellent results and because the sequence "CTAACGGG" had the lowest likelihood to match any transcriptome or genome sequences among all tested TSO.

Our experiments indicate how robust and versatile FLASH-seq is and confirm that UMIs can also be used in our RT-PCR setup when isoform reconstruction is the primary goal of the experiment.

9.1 Cell Lysis Mix

1. 10% v/v Triton X-100. Store at +4 °C (*see* **Note 1**).
2. Premixed dNTP solution (25 mM each). Store at −20 °C.
3. Recombinant RNase inhibitor (40 U/μL). Store at −20 °C.
4. Betaine (5 M). Store at +4 °C.
5. Dithiothreitol (DTT, 100 mM; part of the SuperScript™ IV kit). Store at −20 °C.
6. dCTP (100 mM). Store at −20 °C.
7. STRT-dT_{31} oligonucleotide (5′Bio-AATGATACGGCGAC-CACCGATCGT$_{31}$-3′, 100 μM). "Bio" = biotin. Store at −20 °C.
8. Nuclease-free water. Store at room temperature.

9.2 RT-PCR Mix

1. Superscript™ IV kit: dithiothreitol (DTT, 100 mM). Store at −20 °C.
2. Superscript™ IV kit: Superscript™ IV reverse transcriptase (200 U/μL). Store at −20 °C (*see* **Note 3**).
3. Betaine (5 M). Store at +4 °C.
4. Magnesium chloride ($MgCl_2$, 1 M). Store at +4 °C.
5. Recombinant RNase inhibitor (40 U/μL). Store at −20 °C.
6. Template-switching oligonucleotide with UMIs, TSO-UMI (5′ Bio-AAGCAGTGGTATCAACGCAGAGTNNNNNNNN CTAACrGrGrG-3′, 100 μM). "Bio" = biotin; NNNNNNNN = UMI; CTAAC = spacer. Store aliquots in use at −20 °C and stock tubes at −80 °C (*see* **Note 2**).

7. Forward PCR primer, Tn5-ISPCR-F (5′-TCGTCGGCAGCG TCAGATGTGTATAAGAGACAGAAGCAGTGGTAT-CAACGCAGAGT-3′, 100 μM). Store at −20 °C.
8. Reverse PCR primer, DI-PCR-P1A-R (5′-AATGATACGGCG ACCACCGA-3′, 100 μM). Store at −20 °C.
9. KAPA HiFi HotStart ReadyMix (2×). Store at −20 °C.
10. Nuclease-free water. Store at room temperature.

9.3 Magnetic Beads Preparation Follow the standard FLASH-seq protocol.

9.4 Library Preparation with Commercial Illumina Reagents Follow the standard FLASH-seq protocol.

9.5 Sample QC and Sequencing Follow the standard FLASH-seq protocol.

9.6 Instruments and Consumables Follow the standard FLASH-seq protocol.

10 Methods

10.1 Cell Lysis Mix Preparation

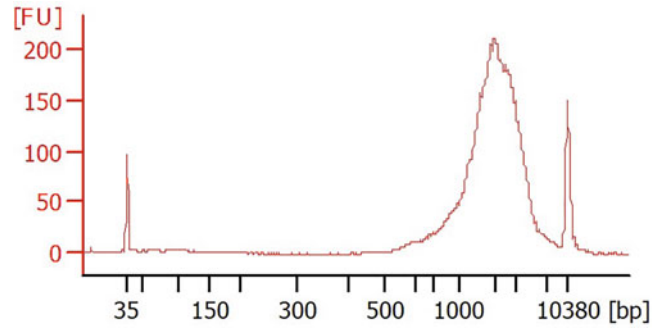
1. For each sample, prepare 1 μL of the following cell lysis mix: 0.020 μL Triton X-100, 0.240 μL dNTP mix, 0.018 μL STRT-dT_{31} oligonucleotide, 0.030 μL recombinant RNase inhibitor, 0.012 μL DTT, 0.090 μL dCTP, 0.200 μL betaine, nuclease-free water to volume.
2. Mix well and gently spin down.
3. Dispense the cell lysis mix into a 384-well plate.
4. If used immediately, keep the plate on ice until needed. The cell lysis mix can be prepared and dispensed on plates several weeks or months in advance. The plates are then stored at −20 °C until needed. Avoid multiple freezing-thawing cycles.
5. Collect the cells by FACS using a FACSAria™ Fusion, preferably with a larger nozzle (85 μm or, even better, 100 μm) to minimize shearing forces upon sorting (*see* **Note 15**).
6. Once the sorting is completed, seal the plate with aluminum seal and snap-freeze it at −80 °C or by placing it on dry ice, especially if not proceeding immediately with RT-PCR. Lysed cells can be stored in these conditions for several months without appreciable decrease in RNA quality. The plate should never undergo freeze-thaw cycles for any reason.

10.2 RT-PCR Mix Preparation and RT-PCR Reaction

1. For each sample, prepare 4 μL of the following RT-PCR mix: 0.238 μL DTT, 0.800 μL betaine, 0.046 μL $MgCl_2$, 0.096 μL recombinant RNase inhibitor, 0.050 μL Superscript™ II reverse transcriptase, 2.500 μL KAPA HiFi HotStart Ready-Mix, 0.092 μL TSO-UMI oligonucleotide, 0.025 μL Tn5-ISPCR- F primer, 0.005 μL DI-PCR-P1A-R primer, nuclease-free water to volume.
2. Briefly vortex and gently spin down the mix. Keep it on ice until needed.
3. Take the plate out of the −80 °C freezer and leave it for a couple of minutes at room temperature to thaw.
4. Briefly spin down to collect eventual drops that might have condensed on the lid.
5. Place the plate on a thermocycler block and perform cell lysis and mRNA denaturation at 72 °C for 3 min.
6. Remove it from the thermocycler and keep it on ice for a couple of minutes to cool. Briefly spin down to collect the lysate at the bottom of each well.
7. Dispense 4 μL RT-PCR mix in each well. The final volume is now 5 μL.
8. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the RT-PCR reaction: 50 °C for 60 min, 98 °C for 3 min, then N cycles of (98 °C for 20 s, 65 °C for 20 s, 72 °C for 6 min), 4 °C hold. The number of cycles "N" should be adjusted according to the cell type used in the experiment. We recommend 20–21 cycles for HEK 293T cells and 23–24 cycles for hPBMC. The addition of UMI and the lack of semi-suppressive PCR decrease reaction yield. As a rule of thumb, use the same number of PCR cycles as in the Smart-seq2 protocol.
9. It is safe to stop here and store the PCR product in a −20 °C freezer until needed.

10.3 Magnetic Beads Cleanup After Preamplification

Follow the standard FLASH-seq protocol but use a 0.6:1 ratio of beads:cDNA. Check the cDNA quality on the Agilent Bioanalyzer instrument.



Example of amplified cDNA from a single human retinal organoid cell (24 PCR cycles)

10.4 cDNA Quantification and Sample Normalization

Follow the standard FLASH-seq protocol. In the last step, prepare a normalization plate by adding 1 μL purified cDNA and nuclease-free water to a final concentration of 100 pg/μL using the Pico-Green readings.

10.5 Library Preparation with the Nextera XT Kit

Please note that the Tn5 transposase amount indicated below is a suggested starting point for tagmenting 100 pg cDNA. Optimization might be necessary, depending on the size of the sequencing libraries that need to be obtained.

To ensure better reproducibility between experiments, we recommend using the Nextera XT kit. However, the in-house Tn5 will provide data of comparable quality.

If the cDNA quantification was accurate, then 0.1–0.2 μL ATM will give sequencing-ready libraries in the range of 700–1000 bp. In our experience, longer libraries gave higher percentages of UMI reads.

The balance between UMI- and internal reads might be complex to achieve and does not always relate to the library size. Increasing the amount of forward primer in the PCR reaction (up to 2–4-times) may sometimes help increase the proportion of UMI reads. However, it has to be taken into account that fragments of >1000 bp are not expected to efficiently bind to NextSeq or NovaSeq flow cells and should therefore be avoided.

For simplicity, here we use 0.2 μL throughout the text.

1. Remove the Amplicon Tagment mix (ATM) and the Nextera® PCR mix (NPM) from the −20 °C storage and place them on ice. Remove the Tagment DNA buffer (TD) and the adaptor index plate from the −20 °C storage and keep them at room temperature. Remove the neutralization buffer (NT) from the +4 °C storage and keep it at room temperature.
2. For each sample, prepare 1.2 μL tagmentation mix: 0.2 μL ATM and 1 μL TD.
3. Briefly vortex and gently spin down the mix. Keep it on ice until needed.
4. Transfer 1 μL normalized cDNA into a new 384-well plate.
5. Add 1.2 μL tagmentation mix to each well.
6. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the tagmentation reaction: 55 °C for 8 min, 4 °C hold. Once the reaction is completed proceed immediately to the next step.
7. Add 0.5 μL NT buffer to each well.
8. Seal the plate with adhesive transparent film, vortex, spin down, and incubate for 5 min at room temperature. Do not put the plate back on ice.
9. Add 1 μL N7xx + S5xx indices from the adaptor index plate.

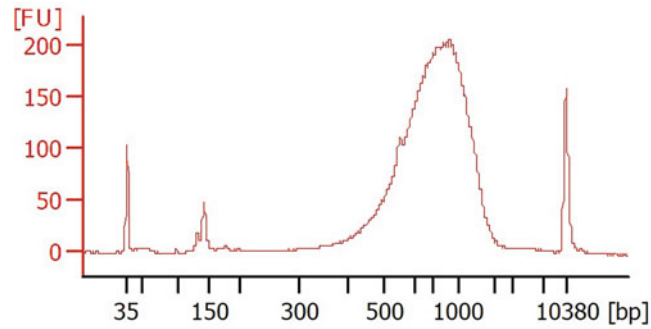
10. Add 1.5 μL NPM into each well.
11. Seal the plate with adhesive transparent film, vortex, spin down, place it in a thermocycler, and carry out the enrichment PCR reaction: 72 °C for 3 min, 95 °C for 30 s, then N cycles of (95 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min, 4 °C hold. The number of cycles "N" should be adjusted according to the amount of cDNA used for the tagmentation reaction. When pooling at least 32–48 cells (or more), 14 cycles are sufficient.
12. It is safe to stop here and store the final library in a −20 °C freezer until needed.

10.6 Pooling and Bead Cleanup of the Final Library

Follow the standard FLASH-seq protocol.

10.7 Sequencing Library QC and Sequencing

1. Use 1 μL of the purified sequencing pool to assess the concentration on a Qubit instrument.
2. Use 1 μL to assess the average size on a High Sensitivity DNA chip.



Example of a sequencing-ready library from a pool of HEK 293T cells. Average size around 800 bp

3. The purified library can be sequenced on any Illumina sequencer. Follow the specifications reported for each instrument. Paired-end sequencing is recommended to reap the most benefits in terms of isoform reconstruction. Read 1 length should not be <75 bp and preferably ≥100 bp. We regularly sequence FS-UMI libraries on a NextSeq500 using 100-8-8-50 read mode but preliminary data indicate that 90-8-8-60 or 80-8-8-70 read modes might lead to better results.

11 Data Preprocessing

The data processing will depend on the final goal. The following description provides guidelines to separate and independently preprocess UMI and internal reads.

In specific cases (e.g., isoform detection), using both internal and UMI reads together is required to properly cover the full body of the transcripts. Trimming of adapters, spacer, and UMI sequences should still be performed in the first step. Failing to properly remove these sequences can result in a decreased percentage of reads uniquely mapped/assigned to a feature.

Here we assume that the data were generated using paired-end sequencing (100 × 50) on an Illumina instrument and are processed on a unix-based machine.

11.1 Prerequisites

Download the following tools. Versions indicated are the ones tested in *Hahaut* et al. [5].

- bcl2fastq (Illumina, v2.20)
- STAR [12] (v2.7.3)
- FeatureCounts [13] (v1.6.5)
- BBMAP [14] (v38.86)
- samtools [15] (v1.9)
- umi_tools [26] (v1.1)
- IGV

When working on hundreds of samples, it is recommended to use virtual machines or cloud platforms to process the data as it will typically require >48 Gb of RAM, >12 cores, and >500 Gb of free space. The footprint of the most time and resource consuming step (= read alignment) can be reduced using pseudoalignment tools (i.e., salmon [16] or kallisto [17]).

11.2 Demultiplexing

Refer to FLASH-Seq data preprocessing.

11.3 Separate UMI/Internal Reads and Extract UMI Sequences

Umi_tools can be used to remove the TSO adapter and spacer sequences and extract the UMI sequences, which are typically located in read 1 (R1-UMI). The UMI sequence can also sometimes be found in read 2 (R2-UMI) instead of read 1 when a tagmentation event occurs upstream of the UMI in the TSO adapter sequence. For this reason, the UMI search must be performed in both R1 and R2 reads.

After obtaining the UMI reads, the internal reads can be extracted from the raw fastq file.

```
# 1. Extract R1-UMI sequences
# The barcode pattern is made of all tagmented variations of
the TSO adapter sequence
umi_tools extract --bc-pattern="^(?P<discard_1>AAGCAGTGGTAT-
CAACGCAGAGT|AGCAGTGGTATCAACGCAGAGT|GCAGTGGTATCAACGCAGAGT|
CAGTGGTATCAACGCAGAGT|AGTGGTATCAACGCAGAGT|GTGGTATCAACGCAGAGT|
TGGTATCAACGCAGAGT|GGTATCAACGCAGAGT|GTATCAACGCAGAGT|ATCAACGCA-
GAGT|CAACGCAGAGT|AACGCAGAGT|ACGCAGAGT|GCAGAGT|CAGAGT|AGAGT|
GAGT|AGT|GT)(?P<umi_1>.{8})(?P<discard_2>CTAACGG)(?P<dis-
card_3>G{0,4})" --stdin=sample.R1.fastq.gz --stdout=umi.
UMIinR1.R1.fq --read2-in=sample.R2.fastq.gz --read2-out=umi.
UMIinR1.R2.fq --extract-method=regex

# 2. Extract R2-UMI sequences
# The barcode pattern is made of the known most common
tagmented variations of the TSO adapter sequence
umi_tools extract --bc-pattern="^(?P<discard_1>GAGT|AGT|GT)(?
P<umi_1>.{8})(?P<discard_2>CTAACGG)(?P<discard_3>G{0,4})" --
stdin=sample.R2.fastq.gz --stdout=umi.UMIinR2.R2.fq --read2-
in=sample.R1.fastq.gz --read2-out=umi.UMIinR2.R1.fq --ex-
tract-method=regex

# 3. Find reads with a UMI in both R1 and R2
# In very rare cases (<0.0001%) the UMI is found in both R1 and
R2
# This can be due to short insert size (= R1 invades R2) or
retrotranscription errors when two TSO are used
# These must be removed or they will create duplicated read ID
entries

# 3.1. Get the R1-UMI / R2-UMI read IDs
cat umi.UMIinR1.R1.fq | uniq | awk 'NR%4==1{print}' | sed 's/
\@//g' > names.R1umi.txt
cat umi.UMIinR2.R1.fq | uniq | awk 'NR%4==1{print}' | sed 's/
\@//g' > names.R2umi.txt
# 3.2. Remove the UMI info from their name
# Umi_tools stores the UMI sequence in the read ID
cat names.R1umi.txt | sed 's/_........ .*$//g' > names.R1umi.
cleaned
cat names.R2umi.txt | sed 's/_........ .*$//g' > names.R2umi.
cleaned
# 3.3. Get the read IDs found in both R1-UMI and R2-UMI
comm -12 <(sort names.R1umi.cleaned) <(sort names.R2umi.
cleaned) > R1R2.toFilterOut

echo "===> Number of R1-R2 with both a UMI: $(wc -l R1R2.
toFilterOut) <==="
echo "===> Number of R1 UMI before cleanup: $(wc -l names.
```

```
R1umi.txt) <==="
echo "===> Number of R2 UMI before cleanup: $(wc -l names.
R2umi.txt) <==="
# 4. Filter out readw with dual UMI sequences

# 4.1. Get the actual read ID of dual-UMI reads
# ID + UMI sequences
grep -f R1R2.toFilterOut names.R1umi.txt > R1.toFilterOut
grep -f R1R2.toFilterOut names.R2umi.txt > R2.toFilterOut

# 4.2. Filter them out using BBMAP filterbyname.sh

filterbyname.sh -Xmx6g in=umi.UMIinR1.R1.fq in2=umi.UMIinR1.
R2.fq  out=umi.UMIinR1.R1.tmp  out2=umi.UMIinR1.R2.tmp
names=R1.toFilterOut include=f overwrite=t
filterbyname.sh -Xmx6g in=umi.UMIinR2.R1.fq in2=umi.UMIinR2.
R2.fq  out=umi.UMIinR2.R1.tmp  out2=umi.UMIinR2.R2.tmp
names=R2.toFilterOut include=f overwrite=t

# 5. Rename the files
mv umi.UMIinR1.R1.tmp umi.UMIinR1.R1.fq
mv umi.UMIinR2.R1.tmp umi.UMIinR2.R1.fq
mv umi.UMIinR1.R2.tmp umi.UMIinR1.R2.fq
mv umi.UMIinR2.R2.tmp umi.UMIinR2.R2.fq

echo "===> Number of R1 UMI after cleanup: $(grep -c \@ umi.
UMIinR1.R1.fq) <==="
echo "===> Number of R2 UMI after cleanup: $(grep -c \@ umi.
UMIinR2.R1.fq) <==="

# 6. Internal reads

# 6.1. UMI read IDs
cat umi.UMIinR1.R1.fq | uniq | awk 'NR%4==1{print}' | sed 's/
\@//g' | sed 's/_........//g' > names.R1umi.txt
cat umi.UMIinR2.R1.fq | uniq | awk 'NR%4==1{print}' | sed 's/
\@//g' | sed 's/_........//g' > names.R2umi.txt
sed -i 's/_........//g' R1.toFilterOut
sed -i 's/_........//g' R2.toFilterOut
cat names.R1umi.txt names.R2umi.txt R1.toFilterOut R2.toFil-
terOut > names.umi.txt

# 6.2. Extract internal reads
filterbyname.sh -Xmx6g in=sample.R1.fastq.gz in2=sample.R2.
fastq.gz out=internal.R1.fq out2=internal.R2.fq names=names.
umi.txt include=f overwrite=t

echo "===> Number of Internal Reads after cleanup: $(grep -c
```

```
\@ internal.R1.fq) <==="

# 7. Combine R1/R2-UMI reads
# UMI reads are stranded (= same orientation as the original
molecule)
# R2-UMI are in the opposite direction compared to the feature
due to the sequencing process
# To reconcile them use read 2 R2-UMI as read 1.
cat umi.UMIinR1.R1.fq umi.UMIinR2.R2.fq > umi.R1.fq
cat umi.UMIinR1.R2.fq umi.UMIinR2.R1.fq > umi.R2.fq

# 8. Clean-up
rm R1.toFilterOut R2.toFilterOut toFilterOut.txt names.umi.txt
rm names.* umi.UMIinR*.R*.fq
```

11.4 Optional: Remove Sequencing Adapter Leftovers

Refer to FLASH-seq data pre-processing.

11.5 Recommended: Discarding Too Short Reads

UMI extraction and trimming can result in short reads (<50 bp). Discard too short reads using trimmomatic.

This step can be skipped when using longer R2 reads (>70 bp).

```
java -jar trimmomatic-0.39.jar PE -threads 2 -phred33 FASTQ/
umi.R1.trim.fq  FASTQ/umi.R2.trim.fq  FASTQ/umi.R1.trim2.R1.fq
FASTQ/umi.R1.trim2.unpaired.fq  FASTQ/umi.R2.trim2.R1.fq
FASTQ/umi.R2.trim2.unpaired.fq CROP:100 MINLEN:29

mv FASTQ/umi.R1.trim2.R1.fq FASTQ/umi.R1.trim.fq
mv FASTQ/umi.R2.trim2.R1.fq FASTQ/umi.R2.trim.fq

gzip -c FASTQ/umi.R1.trim.fq > FASTQ/umi.R1.fq.gz
gzip -c FASTQ/umi.R2.trim.fq > FASTQ/umi.R2.fq.gz
```

11.6 Data Mapping

11.6.1 Genome Indexing

Read aligners (e.g., STAR or HISAT2) or pseudo aligners (e.g., salmon or kallisto) can be used to map the data. The following example is demonstrated using STAR.

The first step of the data mapping process is to generate an index for the reference genome/transcriptome onto which the read will be aligned. This procedure is generally performed only once. The same index is reused for every sample.

The presence of reads of varying lengths represents an additional challenge, as it requires adapting the sjdbOverhang values accordingly. Large values should be privileged in most cases, setting sjdbOverhang to the maximal mate read length - 1. If reads <50 bp are expected after trimming, the default seedSearchStartLmax must be decreased during read alignment.

```
# 0. Define paths
OUTPUTREF="/path/to/output_reference_index/"
FASTA="/path/to/fasta_reference_genome.fa"
GTF="/path/to/genome_annotation.gtf"

# 1. Create the index repository
mkdir $OUTPUTREF
cd $OUTPUTREF

# 2. OPTIONAL: Save the fasta/GTF files to the reference folder
scp $FASTA .
scp $GTF .

# 3. Genome indexing
# runThreadN: Adapt the number of threads to your device
# sjdbOverhang: Usually set to read length -1
STAR --runThreadN 30 \
--runMode genomeGenerate \
--genomeDir $OUTPUTREF \
--genomeFastaFiles $FASTA \
--sjdbGTFfile $GTF \
--sjdbOverhang 99
```

11.6.2 Align Reads

```
mkdir STAR

# 1. UMI Reads
MYFASTQ_R1=”/path/to/FASTQ/umi.R1.fq.gz”
MYFASTQ_R2=”/path/to/FASTQ/umi.R2.fq.gz”

STAR --runThreadN 10 --limitBAMsortRAM 20000000000 --seed-
SearchStartLmax 30 --genomeLoad LoadAndKeep --genomeDir "$OUT-
PUTREF" --readFilesIn "$MYFASTQ_R1" “MYFASTQ_R2” --
readFilesCommand zcat --limitSjdbInsertNsj 2000000 --outFil-
terIntronMotifs RemoveNoncanonicalUnannotated --outSAMtype
BAM SortedByCoordinate --outFileNamePrefix STAR/"$ID"_UMI_

# 2. Internal Reads
MYFASTQ_R1=”/path/to/FASTQ/internal.R1.fq.gz”
MYFASTQ_R2=”/path/to/FASTQ/internal.R2.fq.gz”STAR --run-
ThreadN 10 --limitBAMsortRAM 20000000000 --genomeLoad LoadAnd-
Keep --genomeDir "$OUTPUTREF" --readFilesIn "$MYFASTQ_R1"
“MYFASTQ_R2” --readFilesCommand zcat --limitSjdbInsertNsj
2000000 --outFilterIntronMotifs RemoveNoncanonicalUnannotated
--outSAMtype BAM SortedByCoordinate --outFileNamePrefix STAR/
"$ID"_INTERNAL_
```

11.7 Optional: Filter Reads

Refer to FLASH-seq data preprocessing.

11.8 Optional: Remove UMI Invasion Events

As previously described [5, 24], TSO invasions can happen in SMART-seq methods and can be favored by the presence of a random sequence in the TSO (e.g., UMI). The addition of a spacer in FLASH-seq TSO significantly decreases the probability of such events. However, it can be interesting to evaluate the degree of strand-invasion and remove putative events before performing isoform reconstructions or counting UMI reads.

We created an R script that takes as input a BAM file of reads processed with umi_tools/STAR and filters out putative strand-invasion events. These events are identified based on the homology between the read's UMI sequence and the 20-bp sequence adjacent to the read starting position. Given the random nature of the UMI sequence, we do not expect an overlap between these two sequences. Partial homology can indicate that the UMI was used as an internal primer during RT/PCR and does not originate from an actual random distribution.

```
# Arguments:
# arg1: /path/to/mapped.umi_tools.bam
# arg2: sample ID
# arg3: method 'FLASH-Seq' or 'zUMIs'
# arg4: Maximum number of mismatches (recommended: 1)
# arg5: path/to/genome.fa - genome used for the mapping
# arg6: Minimum number of total reads to process the sample
(recommended: 10000)
# arg7: Minimum mapq to look for a match between UMI and
adjacent sequence (recommended: 5)
# arg8: /path/to/output/folder/"

Rscript filterINvasionEventsfromBAM.R "$ID"_"$TYPE"_"$DOW-
N"_Aligned.sorted.bam "$ID"_"$TYPE"_"$DOWN" 'FLASH-Seq'
1 $FASTA 1000 5 ./
```

This script will return two files:

/path/to/output/folder/ID_filteredInvasion.bam

/path/to/output/folder/ID_filteredInvasion.log.out

The first one is a bam file where the putative invasion events have been removed. The second one is a log file containing some filtering information.

The filterINvasionEventsfromBAM.R script can be found in the FLASH-Seq Github repository:

https://github.com/Radek91/FLASH-Seq

11.9 Assign Reads to Feature Using Featurecounts

UMI read assignment to a feature differs from the one of internal reads. First, they should be assigned to a feature in a stranded manner. Second, similar UMI sequences assigned to the same feature should be collapsed (= deduplication), as they represent PCR duplicates.

```
mkdir FEATURECOUNTS

# 1. UMI Reads

# 1.1. Assign UMI reads to a feature
# Returns a BAM file
featureCounts -T 1 -p -t exon -g gene_name -s 1 --fracOverlap
0.25 -a "$GTF" -R BAM -o FEATURECOUNTS/"$ID".featureCounts.txt
STAR/"$ID"_UMI_Aligned.sortedByCoord.filtered.bam

# 1.2. Sort and index BAM
samtools sort -@ 10 FEATURECOUNTS/"$ID"_UMI_Aligned.sortedBy-
Coord.filtered.bam.featureCounts.bam -o FEATURECOUNTS/
"$ID"_UMI_Aligned.sortedByCoord.filtered.bam.featureCounts.
sorted.bam
samtools index FEATURECOUNTS/"$ID"_UMI_Aligned.sortedByCoord.
filtered.bam.featureCounts.sorted.bam

# 1.3. Count UMI reads
umi_tools count --per-gene --paired --gene-tag=XT --chimeric-
pairs=discard --unpaired-reads=discard --assigned-status-
tag=XS -I FEATURECOUNTS/"$ID"_UMI_Aligned.sortedByCoord.fil-
tered.bam.featureCounts.sorted.bam -S FEATURECOUNTS/"$ID".
umi.counts.tsv.gz

# 2. Internal Reads
featureCounts -T 1 -p -t exon -g gene_name --fracOverlap 0.25
-a "$GTF" -o FEATURECOUNTS/"$ID"_INTERNAL_featureCounts.txt
STAR/"$ID"_INTERNAL_Aligned.sortedByCoord.filtered.bam
```

11.10 Recommended: Quality Checkups

Refer to FLASH-seq preprocessing.

12 Data Postprocessing

The postprocessing steps will greatly vary depending on the question at hand. In our experience, most tools working for Smart-seq2/Smart-seq3 can be used as well with FLASH-seq-UMI data.

13 Notes

1. The concentration of Triton X-100 is important and should not be significantly increased. Although concentrations of 1% or higher still result in cDNA that is indistinguishable from that obtained with 0.2%, the amount of multi-mapped and unmapped reads increases significantly [5]. Other non-ionic surfactants such as Tween-20, Igepal® CA-630, NP-40, or similar could be used but we did not test them extensively. The use of anionic detergents such as SDS results in the complete inactivation of enzymes even at very low concentrations and should be avoided.
2. Primers should be biotinylated and, whenever possible, HPLC-purified. In the RT reaction, the biotinylation prevents secondary strand-switch events and creation of concatemers.
3. In our hands, there was no difference when replacing Superscript™ IV with Maxima H-RT while using the same reaction buffer and reaction conditions.
4. Commercial alternatives like Ampure XP beads or SPRI beads can be used without any negative impact on the reaction.
5. It is recommended to add Tween-20 at the end, as it tends to foam.
6. Due to its toxicity, it is recommended to add sodium azide at the very end.
7. It is a good practice to titrate every new batch of Sera-Mag SpeedBeads™ side by side against commercial products, such as AMPure XP or SPRI magnetic beads.
8. For all tagmentation protocols described here, indices from the Nextera® index kit can be diluted 1:5 with nuclease-free water or low-EDTA TE buffer (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0) before combining them on a plate. Each well will contain a unique combination of N7xx and S5xx (each 10 times diluted compared to their original concentration).
9. Tn5MErev primer needs to be pre-annealed in separate vials with Tn5ME-A and Tn5ME-B before loading the two partly double-stranded oligonucleotides on the Tn5 transposase. See [10] for a detailed protocol.
10. Dimethylformamide is a potent liver toxin that is readily absorbed through the skin. Avoid inhalation of vapors and always handle the solution under a fume hood.
11. SDS concentration is extremely important for the inactivation of the Tn5 transposase and for the enrichment of PCR reaction to take place afterward. Best results are obtained when the concentration of SDS is in the range 0.1–0.2%. Do not increase

the concentration further, as already 0.3% leads to a complete failure of the following PCR.

12. Additional sets of index adaptors guarantee higher multiplexing capabilities, thus maximizing the number of cells that can be pooled and sequenced in each run. All index adaptors should carry a 5′-biotin to minimize artifacts as well as a phosphorothioate bond between the last and second last nucleotide at the 3′-end to make them more resistant to exonucleases. IDT is our preferred vendor but several others can manufacture oligos of comparable quality. Please see [5] for an extensive list of additional index adaptors.
13. Always choose fully skirted plates when carrying out the protocol with liquid handling robots, which also have the additional advantage of offering increased rigidity, thus reducing warping during thermal cycling. Twin.tec® (Eppendorf) is our preferred choice.
14. The choice of a good seal that can resist long-term −80 °C storage is rather limited. We recommend AlumaSeal® 384 film (VWR International) for optimal results. For storage at −20 °C of preamplified cDNA or final libraries, use the adhesive plastic foil of your choice.
15. Whenever possible, use both a positive and a negative control on every plate. The negative control well contains only 1 μL cell lysis mix but no cell. The positive control well contains either high-quality total RNA (10–100 pg) or a pool of 10–20 cells.
16. For long-term storage of preamplified cDNA or sequencing-ready libraries, Lobind tubes and/or Lobind Twin.tec® plates should be preferred. This ensures a minimal adsorption of the DNA to the plastic and guarantees that little DNA is lost upon storage.

References

1. Tang F et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382
2. Zheng GXY et al (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049
3. Hagemann-Jensen M et al (2020) Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol 38:708–714
4. Picelli S et al (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098
5. Hahaut V, Pavlinic D, Cowan C, Picelli S (2021) Lightning Fast and Highly Sensitive Full-Length Single-cell sequencing using FLASH-Seq. http://biorxiv.org/lookup/doi/10.1101/2021.07.14.452217. https://doi.org/10.1101/2021.07.14.452217
6. Deng Q, Ramsköld D, Reinius B, Sandberg R (2014) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343:193–196
7. Petropoulos S et al (2016) Single-cell RNA--Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell 165:1012–1026

8. Reinius B et al (2016) Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. Nat Genet 48:1430–1435
9. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction. BioTechniques 30:892–897
10. Picelli S et al (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24:2033–2040
11. Chen W, Gardeux V, Meireles-Filho A, Deplancke B (2017) Profiling of single-cell transcriptomes. Curr Protoc Mouse Biol 7:145–175
12. Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
13. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930
14. Bushnell B BBMap – sourceforge.net/projects/bbmap/
15. Li H et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
16. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419
17. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527
18. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
19. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360
20. Hao Y et al (2021) Integrated analysis of multimodal single-cell data. Cell 184:3573–3587.e29
21. Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19:15
22. Lun ATL, McCarthy DJ, Marioni JC (2016) A step-by-step workflow for low-level analysis of singlecell RNA-seq data with Bioconductor. F1000 Res 5:1–71
23. Hagemann-Jensen M, Ziegenhain C, Sandberg R (2021) Scalable full-transcript coverage single cell RNA sequencing with Smart-seq3xpress. http://biorxiv.org/lookup/doi/10.1101/2021.07.10.451889, https://doi.org/10.1101/2021.07.10.451889
24. Tang DTP et al (2013) Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. Nucleic Acids Res 41:e44
25. Hochgerner H et al (2017) STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. Sci Rep 7:16327
26. Smith T, Heger A, Sudbery I (2016) UMI-tools: modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. http://biorxiv.org/lookup/doi/10.1101/051755 https://doi.org/10.1101/051755