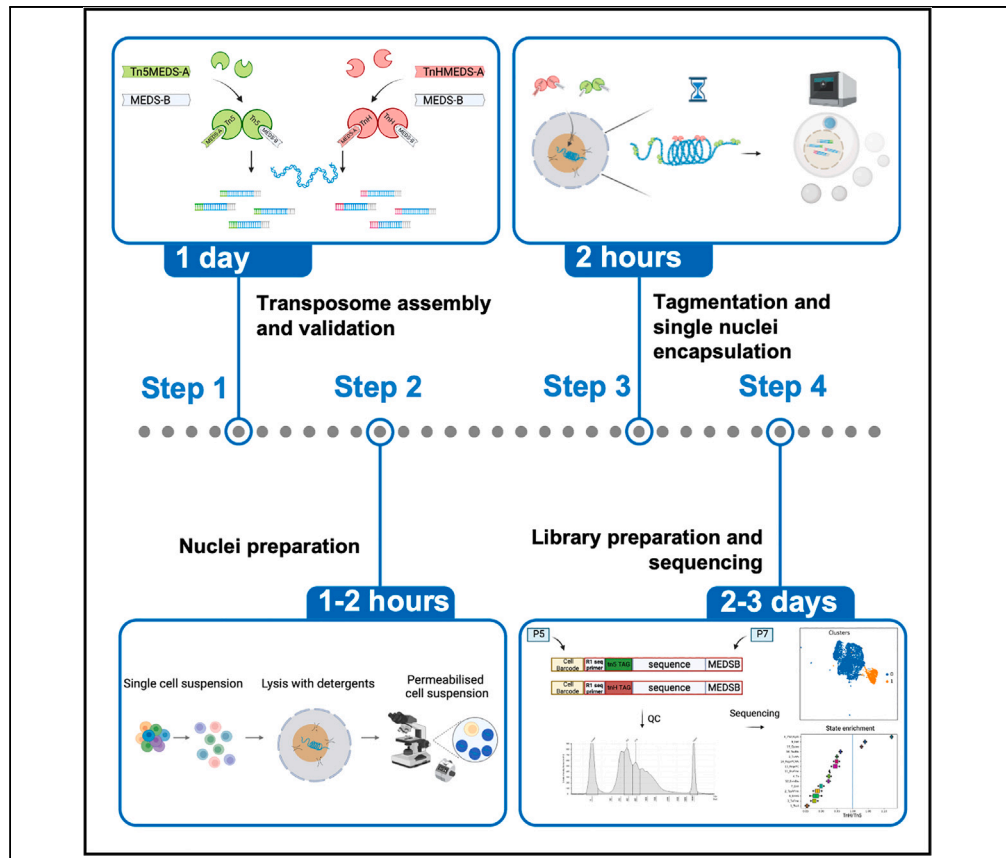


## Protocol

# Analyzing genomic and epigenetic profiles in single cells by hybrid transposase (scGET-seq)



scGET-seq simultaneously profiles euchromatin and heterochromatin. scGET-seq exploits the concurrent action of transposase Tn5 and its hybrid form TnH, which targets H3K9me3 domains. Here we present a step-by-step protocol to profile single cells by scGET-seq using a 10x Chromium Controller. We describe steps for transposomes preparation and validation. We detail nuclei preparation and transposition, followed by encapsulation, library preparation, sequencing, and data analysis.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Davide Cittaro,  
Dejan Lazarević,  
Giovanni Tonon,  
Francesca Giannese

cittaro.davide@hsr.it  
(D.C.)  
giannese.francesca@hsr.it  
(F.G.)

### Highlights

Expands standard ATAC-seq approach using hybrid TnH transposase with H3K9me3 tropism

scGET-seq queries accessible and compacted chromatin

scGET-seq allows definition of cell states by analysis of chromatin dynamics

Cittaro et al., STAR Protocols  
4, 102176  
June 16, 2023 © 2023 The  
Authors.  
<https://doi.org/10.1016/j.xpro.2023.102176>



## Protocol

## Analyzing genomic and epigenetic profiles in single cells by hybrid transposase (scGET-seq)

Davide Cittaro,<sup>1,4,\*</sup> Dejan Lazarević,<sup>1,2</sup> Giovanni Tonon,<sup>1,2,3</sup> and Francesca Giannese<sup>1,3,4,5,\*</sup><sup>1</sup>Center for Omics Sciences, IRCCS San Raffaele Hospital, Milano, Italy<sup>2</sup>Università Vita-Salute San Raffaele, Milano, Italy<sup>3</sup>These authors contributed equally<sup>4</sup>Technical contact<sup>5</sup>Lead contact\*Correspondence: [cittaro.davide@hsr.it](mailto:cittaro.davide@hsr.it) (D.C.), [giannese.francesca@hsr.it](mailto:giannese.francesca@hsr.it) (F.G.)  
<https://doi.org/10.1016/j.xpro.2023.102176>

## SUMMARY

scGET-seq simultaneously profiles euchromatin and heterochromatin. scGET-seq exploits the concurrent action of transposase Tn5 and its hybrid form TnH, which targets H3K9me3 domains. Here we present a step-by-step protocol to profile single cells by scGET-seq using a 10× Chromium Controller. We describe steps for transposomes preparation and validation. We detail nuclei preparation and transposition, followed by encapsulation, library preparation, sequencing, and data analysis.

For complete details on the use and execution of this protocol, please refer to Tedesco et al. (2022)<sup>1</sup> and de Pretis and Cittaro (2022).<sup>2</sup>

## BEFORE YOU BEGIN

## Introduction

Analysis of genome and epigenome at single-cell level could reveal complex cell hierarchies in heterogeneous tissues, as for example clonal structure in tumor evolution, especially as a result of treatment administration.<sup>3</sup> At the same time, cellular epigenetic configuration is crucial for the definition of cell states during differentiation,<sup>4,5</sup> tumor progression<sup>6</sup> and related regulatory programs.<sup>3</sup>

Analysis of the accessible chromatin landscape is typically performed by scATAC-seq.<sup>7</sup> More recently, a number of technologies have been introduced to expand the targeted epigenomic features, such as scNMT<sup>8</sup> and scCUT&TAG.<sup>9</sup> All these technologies are focused on the study of regulatory regions and do not target heterochromatin.

scGET-seq<sup>1</sup> makes use of a recombinant transposase (TnH) obtained by fusing Tn5 to the chromodomain of HP1- $\alpha$ , acquiring tropism toward H3K9me3. By using a mixture of Tn5 and TnH, scGET-seq allows the concurrent study of accessible and compacted chromatin and their dynamic changes. As heterochromatin covers a large part of the genome,<sup>10</sup> scGET-seq extends the fraction of genome that is sequenced compared to standard scATAC-seq, allowing a more precise genetic profiling of single cells.

We describe the specific steps to process adherent cell lines samples (Caki-1, HeLa cell lines). However, we have also used this protocol in cancer-derived organoids, in human-derived xenograft (PDX) models and in models of Induced Pluripotent Stem Cells (iPSC) differentiation.<sup>1</sup>

Success of the protocol is strictly related to the quality of nuclei preparation; therefore, researcher can combine this STAR protocol with any tested protocol for nuclei isolation from the specific tissue type under investigation. We successfully applied scGET-seq on 500 to 10 000 nuclei per sample.



### Preparation

If working with biological replicates, we recommend processing all samples at the same time to minimize batch effects. If this is not possible, process samples in separate batches each with a fraction of total sample number. While doing so, split sample in batches randomizing conditions across batches. Ideally, it is preferable to work from fresh samples, but is also possible to freeze cells in cryopreservative (e.g., containing 15% DMSO in cell culture media supplemented with 20% FBS) and process samples later. At the time of experiment, quickly thaw the cryovials in a water bath at 37°C for 2–3 min (or until there is only a tiny ice crystal visible). Transfer cryovial content to a 15-mL Falcon tube containing 10 mL prewarmed media. Centrifuge at 300×g, 5' at room temperature. Discard supernatant, being careful of not touching cell pellet. Add 1 mL prewarmed PBS+0,04% BSA and transfer cell suspension to a new 1,5 mL tube. Repeat this wash step for a total of two washes. Count cells, assess viability and proceed with the protocol.

### Part one: Wet-lab preparation

Check [materials and equipment](#) section before starting!

Save protocols on your PCR instrument.

Prepare required solutions (see [materials and equipment](#) section). Prepare an ice bucket. Clean the bench with available decontamination reagents. It is strongly recommended working in a dedicated pre-PCR bench for transposomes assembly and all library preparation steps before final amplification.

### *Transposomes assembly and validation*

⌚ Timing: ~ 1 day

1. Obtain plasmid or purified protein.

**Note:** Tn5 is commercially available from different providers, or you can obtain the plasmid from Addgene (ref #60240).

2. Thaw Tn5ME-A, TnHME-A, TnME-B, TnME-rev oligonucleotides.
3. Thaw Tn5 and TnH on ice.
4. Thaw genomic DNA (extracted from any source by using commercial DNA isolation kits or purchased from available vendors).
5. Perform tagmentation on gDNA and assess efficiency by comparison with results obtained from previous Tn5 batches or commercial Tn5.
6. Aliquot transposomes and keep them at –20°C, minimizing room temperature exposure.

### *Nuclei preparation*

⌚ Timing: 1–2 h

**CRUCIAL:** Check cell vitality (using dye exclusion test, such as diluting cells with Trypan blue dye and visually assess the percentage of cells which include dye using microscope) before performing Lysis step. Low viability will result in high background signal!

7. On the day of the experiment, prepare fresh solutions of 1× ATAC Lysis Buffer, 1× Wash Buffer and 1× Diluted nuclei buffer, and keep tubes on ice. Discard leftover solution after the experiment.
8. Precool a minicentrifuge to 4°C.
9. Prepare and label 1.5 mL tubes and 8-tubes 0.2 mL strips.

10. Prepare the required reagents for nuclei counting and viability assessment after permeabilization (trypan blue, Burker chamber or automated counters).

### *Transposition and GEM formation*

⌚ **Timing:** 2 h

Equilibrate reagents to recommended temperatures (as for manufacturer's instructions). For encapsulation and library preparation, we tested scGETseq with 10× Genomics Chromium Next GEM Single Cell ATAC Kit v1.1 kit version.

11. Have transposome complexes ready. Shortly before performing nuclei tagmentation, place Tn5 and TnH transposome aliquots on ice and minimize the exposure to room temperature.

**CRUCIAL:** We strongly recommend using only transposomes batches which have been previously tested for transposition activity *in vitro* (bulk tagmentation).

12. If working with less than 8 samples, have a 50% glycerol solution ready to fill empty wells on the Chromium chip.

### *Library preparation and sequencing*

⌚ **Timing:** 2–3 days

13. Follow 10× Genomics instructions for the most recent version of the Chromium Single Cell ATAC Reagent Kits User Guide on the 10× Genomics website (make sure to choose the User Guide that corresponds to your Reagent Kit version; we here refer to [V1.1](#)).
14. Sequencing requires custom primers; it is crucial to have those added to the Illumina cartridge.

## **Part two: Data analysis**

### *Hardware*

Running the analysis of scGET-seq data requires approximately 500 Gb disk space per sample, to store intermediate files, and a minimum of 64 Gb of RAM.

### *Software*

Download of `cellranger-atac` requires registration to 10× genomics website. Download scGET tools and scatACC tools from their github repositories.

Analysis of scGET-seq data requires alignment to reference genome. Download the appropriate fasta file and index it. Human genome hg38 is used along this manuscript.

The automated analysis pipeline is managed and executed by snakemake,<sup>11</sup> refer to its installation guide to set up a working environment ([https://snakemake.readthedocs.io/en/stable/getting\\_started/installation.html](https://snakemake.readthedocs.io/en/stable/getting_started/installation.html)).

## **KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Biological samples</i>		
Genomic DNA HMW	Any (in-house extracted or available from commercial supplier)	

(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, peptides, and recombinant proteins</b>		
Tn5 transposase (unloaded)	Diagenode	C01070010-10
Optional 1: Tagmentase (Tn5 transposase) – loaded Optional 2: Tagment DNA TDE1 Enzyme and Buffer Kits	Diagenode Illumina	C01070012 20034197
TnH transposase (unloaded)	Authors	Available upon request
Decontamination reagent PDS-250	Merck	BS-040107-DK
Tris-HCl, pH 7.4	Thermo Fisher	15567027
Trizma hydrochloride	Merck	93363
Sodium chloride (NaCl)	AppliChem	A1149
Magnesium chloride (MgCl <sub>2</sub> ) 1 M	Sigma	63069
TWEEN 20	Sigma	P7949
Bovine sodium albumin (BSA)	Sigma	A3059
Nonidet P-40 Substitute	Sigma	74385
Digitonin 2% solution	Promega	G9441
Poly(ethylene glycol) BioUltra, 8,000	Fluka	89510
[tris(hydroxymethyl)methylamino] propanesulfonic acid (TAPS)	Sigma	T5130
Sodium dodecyl sulfate (SDS) (10%)	Merck	71736
AMPure XP Reagent	Beckman Coulter	A63881
SPRIselect Reagent	Beckman Coulter	B23318
KAPA HiFi HotStart ReadyMix	Roche	KK2601
Ethanol absolute	Merck	107017
PhiX Control v3	Illumina	FC-110-3001
Nuclease free water	Thermo Fisher	4387936
Trypan blue solution (0.4%)	Gibco (Thermo Fisher)	15250061
Ethylenediaminetetraacetic acid (EDTA) (0.5 M)	Amresco (Thermo Fisher)	E177
Hydrochloric acid (37%)	Sigma	320331
NaOH (10 N)	Merck	72068
<b>Critical commercial assays</b>		
Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v1.1	10× Genomics	PN-1000175 or PN-1000176
Chromium Next GEM Chip H Single Cell Kit	10× Genomics	PN-1000161 (48 rxn) or PN-1000162 (16 rxn)
Single Index Kit N, Set A, 96 rxns	10× Genomics	PN-1000212
High Sensitivity D5000 ScreenTape	Agilent	5067- 5592
High Sensitivity D5000 Reagents	Agilent	5067- 5593
High Sensitivity D1000 ScreenTape	Agilent	5067-5582
High Sensitivity D1000 Reagents	Agilent	5067-5583
Qubit assay tubes	Thermo Fisher	Q32856
Qubit dsDNA HS Assay kit	Thermo Fisher	Q33230
DNA Clean & Concentrator-5	Zymo Research	D4013
<b>Deposited data</b>		
Raw data	Tedesco et al. <sup>1</sup>	Array Express ID: E-MTAB-9650
Human reference genome NCBI build 38, GRCh38	UCSC genome browser	<a href="https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz">https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz</a>
<b>Experimental models: Cell lines</b>		
Caki-1	American Type Culture Collection (ATCC)	<a href="http://www.atcc.org">www.atcc.org</a>
HeLa	American Type Culture Collection (ATCC)	<a href="http://www.atcc.org">www.atcc.org</a>
<b>Oligonucleotides</b>		
Tn5ME-A.1 5'-TCGTCGGCAGCGTC TCCGATCTCGTACTAGAGATGTGT ATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Tn5ME-A.2 5'-TCGTCGGCAGCGT CTCCGATCTTCTGAGCAGATGT GTATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
Tn5ME-A.7 5'-TCGTCGGCAGCGTC TCCGATCTTCATGAGCAGATGTGT ATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
Tn5ME-A.8 5'-TCGTCGGCAGCGT CTCCGATCTCCTGAGATAGATGT GTATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
TnHME-A.4 5'-TCGTCGGCAGCGTC TCCGATCTTAAGGCCAAGATGTG TATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
TnHME-A.5 5'-TCGTCGGCAGCG TCTCCGATCTGCTACGCTAGAT GTGTATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
TnHME-A.9 5'-TCGTCGGCAGCGTC TCCGATCTAGGCTCCGAGATGTGT ATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
TnHME-A.10 5'-TCGTCGGCAGCG TCTCCGATCTTGCCATAGATGT GTATAAGAGACAG-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
TnME-B 5'-GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG-3'	Picelli et al. <sup>12</sup>	N/A (recommended HPLC purification)
TnME-rev 5'-[phos]CTGTCTCT TATACATCT-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
CustomR1seqprimer 5'-TCGTCGGCAGCGTCTCCGATCT-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
Custom index_2_primer 5' – AGATC GGAGACGCTGCCGACGA-3'	Tedesco et al. <sup>1</sup>	N/A (recommended HPLC purification)
<b>Recombinant DNA</b>		
Plasmid: pTXB1-Tn5 vector	Addgene	60240
Plasmid: pTXB1-TnH vector	Tedesco et al. <sup>1</sup>	Contact authors
<b>Software and algorithms</b>		
Python version 3.8+	Python Software Foundation	<a href="https://www.python.org">https://www.python.org</a>
cellranger-atac	10X genomics	<a href="https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac">https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac</a>
scatACC tools	Tedesco et al. <sup>1</sup>	<a href="https://github.com/dawe/scatACC">https://github.com/dawe/scatACC</a>
bwa	Li <sup>13</sup>	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
samtools	Li et al. <sup>14</sup>	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
sambalster	Faust et al. <sup>15</sup>	<a href="https://github.com/GregoryFaust/sambalster">https://github.com/GregoryFaust/sambalster</a>
tagdust	Lassmann <sup>16</sup>	<a href="http://tagdust.sourceforge.net/">http://tagdust.sourceforge.net/</a>
scGET tools	Tedesco et al. <sup>1</sup>	<a href="https://github.com/leomorelli/scGET/">https://github.com/leomorelli/scGET/</a>
umi_tools	Smith et al. <sup>17</sup>	<a href="https://umi-tools.readthedocs.io/en/latest/">https://umi-tools.readthedocs.io/en/latest/</a>
scanpy	Wolf et al. <sup>18</sup>	<a href="https://scanpy.readthedocs.io/en/stable/">https://scanpy.readthedocs.io/en/stable/</a>
tensorly	Kossaifi et al. <sup>19</sup>	<a href="http://tensorly.org">http://tensorly.org</a>
snakemake	Köster and Rahmann <sup>11</sup>	<a href="https://snakemake.readthedocs.io/en/stable/">https://snakemake.readthedocs.io/en/stable/</a>
bedtools	Quinlan <sup>20</sup>	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
cellrank	Lange et al. <sup>21</sup>	<a href="https://cellrank.readthedocs.io/en/stable/">https://cellrank.readthedocs.io/en/stable/</a>
schist	Morelli et al. <sup>22</sup>	<a href="https://schist.readthedocs.io/en/latest/">https://schist.readthedocs.io/en/latest/</a>

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Other</i>		
Chromium Controller & Next GEM Accessory Kit	10× Genomics	1000204
Chromium accessories (10× vortex adapter, Chromium Next GEM Secondary Holder, 10× magnetic separator)	10× Genomics	120251, 1000195, 120250
Thermal cyclers (should support uniform heating of 100 μL emulsion volumes). We routinely use Bio-Rad C1000 Touch Thermal Cycler with 96-Deep Well Reaction Module	Bio-Rad	PN-1851197
Standard laboratory vortex	Any	
Qubit fluorimeter (several versions available)	Thermo Fisher	<a href="#">Q33238</a>
IDT for Illumina DNA/RNA UD Indexes	Illumina	20027213
1.5 mL polypropylene conical tubes	Eppendorf or other vendors	0030121872
15 mL polypropylene conical tubes	Corning or other vendors	CLS430829
0.2 mL 8-tube PCR strips	Eppendorf or other vendors	0030124359
Burker chamber		
Syringe 0.22-μm filter	Merck	<a href="#">SLGL0250S</a>
OPTIONAL Automated cell counter such as Luna-FL or TC20	Logos Biosystems, Bio-Rad	L20001 1450102

## MATERIALS AND EQUIPMENT

### 1 M Tris-HCl, pH 7.4 – 8.0 – 8.5 (500 mL stock)

Dissolve 78.8 g of Tris-HCl (molecular weight 157.6 g/mol) in 350 mL of nuclease-free water. Adjust the pH to the correct pH by slowly adding 37% HCl solution drop-wise. Bring solution to 500 mL with nuclease-free water. Sterile filter using 0.22-μm filter and store at + 4°C up to 1 year.

### 5 M NaCl (100 mL stock)

Weight 29.2 g of NaCl. Dissolve powder in 80 mL nuclease-free water. Adjust the volume to 100 mL with nuclease-free water. Autoclave sterilize. Store the NaCl solution at room temperature. Solution is stable for at least 1 year.

### 10% Tween 20 (50 mL stock)

Mix 10 mL of 50% Tween 20 with 40 mL of nuclease-free water. Stir to mix. Store at 4°C covered with aluminum foil. Solution is stable up to 6 months.

### 10% NP-40 (100 mL stock)

Combine 10 mL of NP-40 with 90 mL of nuclease-free water. Stir to mix. Store at 4°C. Solution is stable up to 6 months.

### 80% EtOH solution (100 mL stock)

Dilute EtOH with distilled water: add 20 mL distilled water to 80 mL EtOH. Store at 4°C for few days or at –20°C for 2–3 weeks. Always equilibrate at room temperature 30' before purification steps.

### 1 M TAPS-NaOH at pH 8.5 solution (100 mL stock)

To achieve 1 M molar concentration in 100 mL final volume, dissolve 24.3 g of TAPS (MW = 243.28 g/mol) in 75 mL of distilled H<sub>2</sub>O. Adjust pH using 10 N NaOH and bring to 100 mL with distilled H<sub>2</sub>O. Filter and store at 4°C. Stable up to 1 year.

### 10% BSA (100 mL stock)

Prepare 10% BSA solution from powder by dissolving 10 g molecular biology grade BSA in 100 mL of distilled H<sub>2</sub>O. Filter, prepare 5 mL aliquots and store at –20°C for up to 1 year. Once thawed, store at 4°C for <6 months. Visually inspect before use.

### 0.2% SDS solution (50 mL)

Dilute 1:50 SDS stock solution (10%): combine 49 mL nuclease-free water with 1 mL stock SDS. Stir to mix. Store at room temperature. Stable for at least 6 months.

### Resuspension Buffer (RSB)

Reagent	Final concentration	Amount
Tris-HCl, pH 8.5 (1 M)	10 mM	100 $\mu$ L
Tween 20 (10%)	0.1%	100 $\mu$ L
Nuclease free water		9.89 mL
Total volume		10 mL

Sterile filter, aliquot and store at 4°C. Stable up to 1 year. Bring to room temperature upon use.

### TE buffer

Reagent	Final concentration	Amount
Tris-HCl, pH 8.0 (1 M)	10 mM	50 $\mu$ L
EDTA (0.5 M)	1 mM	10 $\mu$ L
Nuclease-free water		4.94 mL
Total volume		5 mL

Sterile filter, aliquot and store at 4°C. Stable up to 1 year. Bring to room temperature upon use.

### 1 $\times$ ATAC Wash buffer

Reagent	Final concentration	Amount
Tris-HCl, pH 7.4 (1 M)	10 mM	50 $\mu$ L
NaCl (5 M)	10 mM	10 $\mu$ L
MgCl <sub>2</sub> (1 M)	3 mM	15 $\mu$ L
Tween 20 10%	0.1% (vol/vol)	50 $\mu$ L
BSA (10%)	1%	500 $\mu$ L
Nuclease-free water		4.37 mL
Total volume		5 mL

Prepare fresh the day of the experiment, keep on ice and discard leftovers.

### 1 $\times$ ATAC LYSIS BUFFER

Reagent	Final concentration	Amount
Tris-HCl, pH 7.4 (1 M)	10 mM	50 $\mu$ L
NaCl (5 M)	10 mM	10 $\mu$ L
MgCl <sub>2</sub> (1 M)	3 mM	15 $\mu$ L
Tween 20 (10%)	0.1% (vol/vol)	50 $\mu$ L
NP-40 (10%)	0.1% (vol/vol)	50 $\mu$ L
Digitonin (2%)	0.01%	25 $\mu$ L
BSA (10%)	1%	500 $\mu$ L
Nuclease-free water		4.3 mL
Total volume		5 mL

Prepare fresh the day of the experiment, keep on ice and discard leftovers.

### 5 $\times$ Tagmentation buffer

Reagent	Final concentration	Amount
1 M TAPS-NaOH at pH 8.5 [RT]	50 mM	250 $\mu$ L
MgCl <sub>2</sub> (1 M)	25 mM	125 $\mu$ L

(Continued on next page)



**Continued**

Reagent	Final concentration	Amount
PEG8000 (50%)	40%	4 mL
Nuclease-free water		625 $\mu$ L
Total volume		5 mL

Store at RT, stable >6 months.

**1  $\times$  Diluted Nuclei Buffer**

Reagent	Final concentration	Amount
Nuclei Buffer (PN-2000153, PN-200020, part of the Single Cell ATAC Library and Gel Bead Kit) [20 $\times$ ]	1 $\times$	50 $\mu$ L
Nuclease free water		950 $\mu$ L

Prepare fresh the day of experiment, keep on ice and discard leftovers.

Prepare oligonucleotides stock solution (see [key resources table](#)) by diluting in TE to 100  $\mu$ M. Aliquot and store at  $-20^{\circ}\text{C}$  for >6 months.

- $\Delta$  CRITICAL: NP-40 (Sigma, 74385) can cause skin and eye irritation. Wear appropriate safety equipment while handling.
- $\Delta$  CRITICAL: Ethanol (Merck, 107017) is flammable. Handle it inside a hood or in a well-ventilated room using appropriate safety equipment.
- $\Delta$  CRITICAL: Digitonin (Promega, G9441) is toxic if swallowed or by contact. Handle it using appropriate safety equipment.
- $\Delta$  CRITICAL: SDS (Merck, 71736) can be harmful if swallowed or inhaled and cause skin/eye irritation. Wear protective equipment while handling.
- $\Delta$  CRITICAL: Trypan blue (Gibco, 15250061) may be carcinogenic. Always use protective equipment during use and dispose accordingly.
- $\Delta$  CRITICAL: EDTA (Amresco, E177) is harmful if swallowed or if inhaled and cause skin irritation. Avoid breathing vapors. Wear protective equipment while handling.
- $\Delta$  CRITICAL: HCl (Sigma, 320331) causes skin and eye damage, corrosion or irritation Do not breathe mist, vapors or spray. Handle under chemical hood and wear protective equipment.

For commercial kits used in this protocol, please refer to safety instructions provided by manufacturer.

**Alternatives:** The standard reagents (e.g., Ethanol, Trypan blue, etc) and plasticware described in this protocol can be replaced by equivalent products from other manufacturers.

**STEP-BY-STEP METHOD DETAILS**

The GET-seq protocol can be applied to bulk or single cell analysis. It has been validated on several sample types such as cell lines and cancer organoids.<sup>1</sup> We successfully applied this protocol on a input cell number ranging from 50 000 to 200 000, and a target nuclei number of 500 to 10 000.

In this protocol, we demonstrate the workflow using Caki1 cell line and describe scGET-seq method steps.

### Transposome assembly and validation

⌚ Timing: 1 day

The following steps describe the preparation of the enzymatic complexes which will be exploited for tagmentation and DNA tagging. The validation of transposomes by bulk tagmentation on genomic DNA will also be described. The latter step is necessary to proceed with the GET-seq protocol.

*Optional:* If starting from plasmid, follow instructions for Tn5 expression and purification as described in<sup>12</sup> for either Tn5 and TnH protein expression and purification.<sup>1</sup>

1. Anneal modified Mosaic End Double-Stranded (MEDS) oligonucleotides.  
To distinguish tagmentation products derived from Tn5 and TnH transposomes, the two enzymes are loaded with MEDS encoding transposase-specific barcodes (referred to as modified MEDS).  
Tn5 and TnH enzymes are used at a 1,62  $\mu\text{M}$  concentration.
  - a. Prepare equimolar mixture of Tn5ME-A and TnHME-A:
    - i. Prepare Tn5ME-A: in a 0.2 mL tube, add 10  $\mu\text{L}$  of each of the four Tn5ME-A oligo (Tn5ME-A.1, Tn5ME-A.2, Tn5ME-A.7 and Tn5ME-A.8, see [key resources table](#)) to 60  $\mu\text{L}$  H<sub>2</sub>O, final volume 100  $\mu\text{L}$ .
    - ii. Prepare TnHME-A: Add 10  $\mu\text{L}$  of each of the four TnHME-A oligo (TnHME-A.4, TnHME-A.5, TnHME-A.9 and TnHME-A.10, see [key resources table](#)) to 60  $\mu\text{L}$  H<sub>2</sub>O, final volume 100  $\mu\text{L}$ .
  - b. Prepare transposons Tn5MEDS and TnHMEDS:
    - i. Prepare Tn5MEDS-A: in a new 0.2 mL tube, mix 10  $\mu\text{L}$  Tn5ME-A (from step 1.a.i.) to 10  $\mu\text{L}$  ME-rev.
    - ii. Prepare TnHMEDS-A: in a new 0.2 mL tube, mix 10  $\mu\text{L}$  TnHME-A (from step 1.a.ii) to 10  $\mu\text{L}$  TnME-rev.
    - iii. Prepare ME-B: mix 50  $\mu\text{L}$  TnME-B with 50  $\mu\text{L}$  TnME-rev.
    - iv. Anneal transposons. Incubate the microcentrifuges tubes prepared in previous steps containing TnHMEDS-A, Tn5MEDS-A, ME-B in a thermocycler<sup>23</sup>: denature 3' at 95°C, ramp down in 30'' steps to 26°C at a rate of  $-1^\circ\text{C}/\text{step}$  (total of 46 cycles). Store transposons at  $-20^\circ\text{C}$ ; transposons are stable for at least 1 year.
    - v. Mix equal molar annealed ME-B with Tn5MEDS-A to obtain Tn5MEDS transposon. Mix equal molar annealed ME-B with TnHMEDS-A to obtain TnHMEDS transposon.
2. Assemble Tn5 and TnH transposome.
  - a. Dilute Tn5MEDS and TnHMEDS transposon to 10  $\mu\text{M}$  in nuclease-free water.
  - b. In two 1.5 mL tube mix 50  $\mu\text{L}$  of purified Tn5 and purified TnH with 7,15  $\mu\text{L}$  (0.143 vol) Tn5MEDS and TnHMEDS transposon, respectively.
  - c. Incubate 1 h at room temperature, then keep complexes on ice if proceeding with next step, or return tubes to  $-20^\circ\text{C}$ .
3. Validation of transposomes activity.
  - a. Bulk tagmentation of genomic DNA.
    - i. Set thermocycler for 55°C incubation.
    - ii. Thaw gDNA at room temperature. Dilute gDNA to 0.5 ng/ $\mu\text{L}$  with TE.
    - iii. Assemble tagmentation reaction, using two 0.2 mL tubes. Of those, use one for TnH and the other for Tn5 tagmentation:  
Tagmentation reaction mix

Reagent	Amount
5× Tagmentation buffer	4 μL
gDNA 0.5 ng/μL	1 μL
Tn5/TnH transposome	1 μL
ddH <sub>2</sub> O	14 μL

Mix by repeated pipetting, then briefly spin down.

**CRUCIAL:** add a positive control reaction. It is possible to use a previously validated in-house produced transposomes or TDE1 Enzyme (formerly Nextera) (Illumina) enzyme. Expect resulting libraries to have the same distribution profile of the positive control.

- iv. Incubate 7' at 55°C in the thermocycler.
- v. Bring the tube to room temperature and stop tagmentation reaction by adding 5 μL of 0.2% SDS solution. Incubate 5' at room temperature.

**CRUCIAL:** This step allows transposase detachment from DNA.<sup>12</sup> If SDS is not added, this will result in failure of the subsequent amplification step. Note that also chaotropic agents, such as guanidinium thiocyanate contained in the column-based purification kit will result in protein detachment from DNA and fragment releasing.

▮▮ **Pause point:** Tagmented DNA can be stored at –20°C before purification for at least 1 week.

- b. **Purification of tagmentation product.** For this purification we routinely use the DNA Clean & Concentrator- 5 kit (Zymo Research), but it is possible to use other column-based purification kit (e.g., Qiagen MinElute Reaction Cleanup Kit).
  - i. Add 5× vol of Zymo binding buffer (20 μL) to sample, mix by pipetting and add to the filter column. Centrifuge to ≥ 10'000 rcf.
  - ii. Add 200 μL Zymo Wash Buffer (prior first use, add absolute ethanol as recommended). Centrifuge for 30 s. Repeat this step once (total of two washes). Discard flowthrough and collection tube. Put the column filter onto a new labeled 1.5 mL microcentrifuge tube.
  - iii. Elute by adding 15 μL of Zymo Elution Buffer directly to the filter. Incubate 2' at room temperature, then centrifuge to ≥ 10'000 rcf.

▮▮ **Pause point:** Store eluted purified tagmentation product at –20°C if not proceeding immediately to the next step.

- c. **PCR amplification.**

Before starting, thaw IDT for Illumina UD Indexes plate at room temperature, spin at 1,000 × g for 1 min and keep it on ice.

Load the amplification program on the thermocycler.

Have the AmpureXP beads equilibrated at room temperature for at least 30'. Prepare fresh 80% ethanol.

Prepare 8-tube strips (0.2 mL).

  - i. Assembly amplification mastermix for the number of samples prepared plus a negative control (no template, substituting template volume with water) as follows:

<b>PCR mastermix</b>	
Reagent	Amount
DNA template	5 μL
2× KAPA HiFi HS ReadyMix	12.5 μL
IDT for Illumina DNA/RNA UD Indexes	1 μL
ddH <sub>2</sub> O	6.5 μL

ii. Use the following PCR program:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Gap Filling	72°C	3'	1
Initial Denaturation	98°C	30 s	1
Denaturation	98°C	45 s	15 cycles
Annealing	55°C	30 s	
Extension	72°C	1 min	
Final extension	72°C	5 min	1
Hold	4°C	forever	

**CAUTION:** The gap filling step is crucial to create an amplifiable template, as the tagmentation reaction result in a 9 bp gap between the nontransferred strand and the target DNA.<sup>24</sup>

d. Post-PCR purification.

Resuspend AmpureXP beads by vortexing.

- i. Add 0.8× AmpureXP beads to each PCR-amplified sample, mix by pipetting and incubate 5' at room temperature.
- ii. Perform a quick spin and put the 8-tube strip on a 96-well plate magnetic stand for 5' or until solution clears. Discard supernatant while avoid disrupting bead pellet.
- iii. Using a multichannel pipette, add 200 µL 80% ethanol to each well. Wait 30" and discard ethanol. Repeat this step once for a total of two washes.
- iv. Transfer the 8-tube strip to a tube rack, wait 3' to achieve ethanol evaporation but be careful to not overdry beads pellet. This would result in beads cracking and affect recovery yield.
- v. Add 20 µL of nuclease-free water or Resuspension Buffer. Pipette mix and put strip back to magnet. Wait 2'. Meanwile, prepare a new 8-tubes strip.
- vi. Transfer supernatant without disrupting bead pellet.
- vii. Quantify product using fluorimetric method such as Qubit (ThermoFisher). Negative control should have a non-detectable concentration (typically <0.5 µg/mL).

e. Assessment of tagmentation product fragment size.

We routinely utilize a digital capillary electrophoresis platform to evaluate fragment distribution, such as TapeStation (Agilent) or Bioanalyzer (Agilent).

**Alternatives:** Other systems can be used to analyze fragments, e.g., LabChip (PerkinElmer). For genomic DNA tagmentation is possible to use D1000 Tape Station assay. Sample concentration should be adjusted according to quantitative chip range (in case of D1000, is 0.1–50 ng/µL; in case of High Sensitivity D1000HS is 10–1,000 pg/µL).

### Nuclei preparation

⌚ Timing: 1–2 h

In the following steps we describe cell permeabilization to allow nuclei transposition. We routinely observe a 50% cell loss during nuclei preparation; therefore, we suggest to always start from an excess cell number. We routinely start from 100,000 cells.

For this step, precool fixed-angle centrifuge; have freshly prepared (<24 h) Lysis, Wash and Diluted nuclei buffers, keep those refrigerated on ice.

**△ CRITICAL:** For any successful single cell protocol, high quality preparation with viable cells is required. Dead or dying cells can affect cell recovery and release genomic material that

will drain the sequencing efficiency for cells selected for the analysis (see Problems and Solutions). For any new sample type, we therefore suggest performing beforehand pilot experiments to assess i) the best protocol for sample dissociation and ii) always thoroughly check vitality using Trypan Blue stain and a manual (Burker chamber) or automated bright-field/fluorescence cell counter (e.g., Luna, Logos Biosystem; TC20, Bio-Rad). In case of a vitality <80% we do recommend performing live cell enrichment by means of FACS-sorting, enrichment kits such as Dead Cell Removal kit (Mylteni), or microfluidic-based separation approaches such as LeviCell (LevitasBio).

**Note:** You can start from a fresh sample, cryopreserved nuclei (e.g., obtained after tissue disaggregation) or cryopreserved cells or cryopreserved tissues. Whole cells can be frozen in a 15% DMSO-FBS solution and stored at  $-80^{\circ}\text{C}$  for short-term storage, or liquid nitrogen for extended periods, then thawed and processed as fresh cells.

- Starting from cell culture suspension, wash cells in PBS+0.04%BSA two times, then assess cell concentration and vitality.

**Note:** If cell suspension vitality is equal to or higher than 80% proceed with the preparation, otherwise see recommendation in the CRITICAL section.

- Prepare 200,000 cells aliquots in a 1.5 mL tube.

**Note:** It is possible to start from even lower cell numbers. We tested as low as 50,000 cells. In these cases, cell pellets may not be evident during washes, so particular care should be taken during pipetting steps, removing the supernatant gently on the tube side opposite of where the pellet is expected.

**△ CRITICAL:** From now on, always work on ice and use precooled centrifuge.

- Centrifuge tubes at 300 g, 5' at  $4^{\circ}\text{C}$ . Turn all the tubes hinges to point outward centrifuge axle. Expect the pellet to be on the hinge side.
- With a p1000 pipette, gently aspirate 80% of the supernatant volume. With a p200 pipette try to aspirate as much supernatant as possible without disturbing the pellet (e.g., if starting with 1,000  $\mu\text{L}$  sample, aspirate 800  $\mu\text{L}$  with a p1000 pipette, and 190  $\mu\text{L}$  with a p200).
- Keeping the tube on ice, add 100  $\mu\text{L}$  Lysis Buffer and pipette mixing 3 times. Start the timer as soon as you mix the reaction components.

**Note:** permeabilization time must be assessed for each cell type. Start from 1 min with 30 s increases to find the permeabilization time yielding >90% trypan-blue positive nuclei.

For Caki-1 and HeLa cells, incubate cells 3 min in Lysis buffer on ice.

**△ CRITICAL:** optimize incubation time in Lysis Buffer for each new sample type. We suggest starting from 1' incubation and increase in 30'' steps. Before starting the incubation, we suggest having a 96-well plate to dilute cells for vital counting nearby in which you already added 5  $\mu\text{L}$  of Trypan Blue in the first row. Then, for each timing, 10'' before the time is up, take 5  $\mu\text{L}$  of cell suspension in Lysis Buffer and quickly dilute in the plate containing Trypan blue, then load 8  $\mu\text{L}$  on your Burker Chamber or automated counter slide. Check the lysis efficacy (percentage of Trypan Blue positive cells) as function of incubation time.

**Note:** If processing multiple samples in parallel, make sure to have all tubes incubated for the same amount of time to avoid different lysis efficacy.

9. Add 1 mL cold Wash Buffer to stop the lysis by dilution. Mix by inversion.
10. Centrifuge at 500\*g 5' at 4°C.
11. Discard the supernatant, being careful of not perturbing the nuclei pellet. With a p1000 pipette, gently aspirate 80% of the supernatant volume. With a p200 pipette try to aspirate as much supernatant as possible without disturbing the pellet (e.g., if starting with 1,000 µL sample, aspirate 800 µL with the p1000 and 190 µL with p200).
12. Resuspend nuclei pellet in Diluted Nuclei Buffer (DNB) to obtain the final Nuclei Stock. The volume is computed according to target concentration which is function of target nuclei recovery (as described in the 'Nuclei concentration guidelines' guide in the [Chromium Next GEM Single Cell ATAC Reagent Kits manual](#), 10x Genomics).

For example, if we aim at 5,000 nuclei target recovery, and we start from 200,000 cells, we suggest a target concentration of 3,000 nuclei/µL. In this case you will resuspend nuclei pellet in 30 µL of DNB. After adding DNB, mix by pipetting 3–4x.

13. Double-check actual concentration by manual counting using a Burkner chamber (we suggest diluting the sample 10 times using DNB): this represents the Nuclei Stock Concentration. Calculate the volume of nuclei stock to be used in the following step with this formula (as stated in Chromium Next GEM Single Cell ATAC Reagent Kits v1.1/2 User Guide):

$$\text{Vol of Nuclei Stock } (\mu\text{L}) = \frac{\text{Target nuclei recovery} \times 1.53 \text{ (recovery efficiency factor)}}{\text{Nuclei Stock Concentration}}$$

Following the previous case example, if we aim at analyzing 5,000 cells, and our nuclei stock concentration is 3,000 nuclei/µL, the nuclei stock volume will be:

$$\text{Vol of Nuclei Stock } (\mu\text{L}) = \frac{5000 \times 1.53}{3000} = 2.55 \mu\text{L}$$

### Transposition and GEM formation

⌚ Timing: 2 h

Prepare the PCR thermocycler at 37°C.

14. In a new tube, prepare the transposition mix:

Transposition mix	
Reagent	Amount for 1 rxn
ATAC Buffer B	7 µL
Tn5 transposome	1.5 µL
TOTAL	8.5 µL

If performing multiple reactions, prepare a mastermix accounting for a 10% volume excess. Mix by pipetting; spin down briefly to collect any drops on tube walls. Aliquot 8.5 µL for each experimental condition in a new 8-tube strip.

15. Add Diluted nuclei buffer to the nuclei stock, with a final total volume of 5 µL. In our example, 5 µL total volume – 2.55 µL nuclei stock = 2.45 µL diluted nuclei buffer. Gently pipette mix. Add to the transposition mix tube.
16. Transfer the tube-strip to the thermocycler set at 37°C for 30'.

17. After 30', pause incubation timer, open tubes cap and add 1.5  $\mu$ L TnH transposome stock to each reaction tube. Gently pipette mix 3 $\times$ , very briefly spin down and start last 30' incubation.

#### Transposition protocol

Steps	Temperature	Time
Tn5 incubation	37°C	30'
TnH incubation	37°C	30'
hold	4°C	forever

During this time, equilibrate at room temperature the following reagents from Chromium Next GEM Single Cell ATAC Reagent Kits:

Reagent	Temperature
Single Cell ATAC Gel Beads	RT
Reducing Agent B	RT
Barcoding Reagent B	RT
Barcoding Enzyme	ice

#### Library preparation and sequencing

⌚ Timing: 2–3 days

In this section we provide instructions and recommendation for library preparation and sequencing on Illumina platforms.

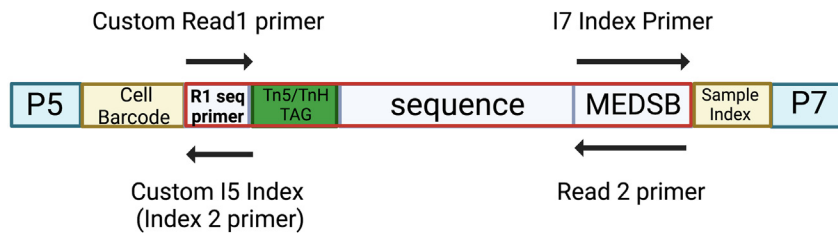
Prepare sequencing reagents and custom primers for Read1 and Index2. Primers resuspended to 100  $\mu$ M can be aliquoted and stored at  $-20^{\circ}\text{C}$  for  $>6$  months.

18. At the end of the incubation time, proceed with GEM generation and library preparation according to [10X Genomics protocol](#) (here we refer to V1.1 chemistry).
19. Assess library concentration using a fluorimetric method such as Qubit dsDNA assay. Evaluate fragment distribution by digital electrophoresis platform such as TapeStation or Bioanalyzer (Agilent). We use the D5000 HS reagents (Agilent). Final library structure is reported in [Figure 1](#).

**Note:** Expected library profile will be slightly different from the standard 10 $\times$  genomics scATAC library profile. We routinely observe a higher representation of the lower library peak (approx. 250 bp).

**Note:** We observed that library fragmentation profile is not always a directly measure of tagmentation performance and specificity. Indeed, in some cases we observed traces with less pronounced library peaks yielding proper relative enrichment over chromatin states. In such cases, we suggest to perform a preliminary library quality check by bulk sequencing aiming at 50M reads, with paired-end sequencing protocol and single index read. In this way you will not read Cell Barcode (i5 read) but obtain a pseudo-bulk tagmentation profile to assess library quality (details in steps 21–26).

**Note:** If your library has  $>50\%$  of fragments higher than 800 bp, we expect this to be due to errors in the last step of library purification and/or to undertagmentation. Under-tagmentation can be related to poor permeabilization, especially when working with complex tissue for which sometimes is difficult to achieve a proper disaggregation resulting in a single cell



**Figure 1. Final scGET-seq library structure**

Position of sequencing primer is reported. Please note use of custom Read 1 and Index 2 primers.

suspension. This high MW fragment will impact sequencing as those fragments will hybridize to the flow-cell as well but with low sequencing efficiency. It's possible to eliminate or decrease the amount of >800 bp library by performing a right-side clean up using Ampure XP or SPRI beads (See [troubleshooting](#) Section, [problem 3](#)).

20. Calculate library molar concentration: use ng/μL concentration assessed using fluorimetric assay (e.g., Qubit) and mean library size (bp) assessed by digital electrophoresis (e.g., TapeStation):

$$nM = \frac{\text{library concentration } \left(\frac{\text{ng}}{\mu\text{L}}\right)}{660 \left(\text{average MW of dsDNA bp, } \frac{\text{g}}{\text{mol}}\right) * \text{mean fragment size (bp)}} \times 10^6$$

To assess mean library size, we routinely use TapeStation Analysis Software and select a Region which includes library up to 750 bp. Please refer to previous notes if your library distribution is significantly enriched in fragments >800 bp.

21. Compute target cluster number. We recommend a higher sequencing depth than for standard scATAC-seq, specifically 100,000 reads pairs per nucleus to account for increased genomic coverage. E.g., when targeting 5,000 nuclei, we recommend aiming at 100,000 reads pair /nucleus will result in a total of 500,000,000 read pairs (i.e., 500,000,000 clusters). Follow Illumina specification for diluting sample for loading and sequencing on selected platform.

**CRUCIAL:** Add 15% PhiX to the library pool to account for the low diversity in the first 27 bases of Read1.

**Note:** scGETseq libraries can be sequenced on Novaseq and Nextseq series platforms and require custom primers for Read1 and Index2 reads, as the tag included in the MEDS-A sequence disrupts native Illumina R1 and I2 primer binding site (see [key resources table](#) for sequences). Custom primers are to be added to the wells containing Illumina primers at a final concentration of 0.3 μM:

- a. NextSeq loading (reagents v2 and later): add Read1\_primer to well #20 and index\_2\_primer to well #22.
- b. NovaSeq (reagents V1.5): add Read1\_primer to well #24 and index\_2\_primer to well #23.

**Note:** Always refer to Illumina Custom Primer guide for updated instructions about Illumina primer positions and volumes (<https://support.illumina.com/bulletins/2016/04/spiking-custom-primers-into-the-illumina-sequencing-primers-.html>).

### Analysis of minicontrol -bulk data (transposomes validation)

© Timing: 1–2 days



In this section, we detail the analysis of bulk GET-seq data sequenced to evaluate the goodness of enrichment over heterochromatic regions. We exemplify the process assuming a single sample has been processed and it has been given the mock name “sample1”.

**Optional:** Merge sequencing lanes into single files. If demultiplex was performed allowing multiple lanes to be splitted (i.e., sample1\_L001\_R1\_001.fastq.gz, SEQ1\_L002\_R1\_001.fastq.gz...), these could be merged together to simplify the procedure.

**Note:** merging should be performed at read pair level, i.e., R1 and R2 files must be merged separately.

```
$ SAMPLE='sample1'
$ for r in 1 2
do
  cat ${SAMPLE}_*_R${r}*.fastq.gz > ${SAMPLE}_R${r}.fastq.gz
done
```

22. Separate reads generated by each transposase with tagdust<sup>16</sup> and compress resulting files.

```
$ tagdust -1 B:CGTACTAG,TCCTGAGC,TCATGAGC,CCTGAGAT,TAAGGCGA,GCTACGCT,AGGCTCCG,CTGCGCAT \
  -2 S:AGATGTGTATAAGAGACAG \
  -3 R:N ${SAMPLE}_R1.fastq.gz ${SAMPLE}_R2.fastq.gz \
  -o ${SAMPLE} \
  -t 8
$ gzip ${SAMPLE}_BC_*.fq
```

Repeat this step for each sample. This step generates 8 set of paired reads for each sample. Each file will have a structure like `${SAMPLE}_BC_${BARCODE}_READ[12].fq.gz`, where `$BARCODE` is one of the 8 transposase-specific sequences, also indicated in the box above. More in detail, the expected files are:

```
sample1_BC_CGTACTAG_READ1.fq.gz
sample1_BC_CGTACTAG_READ2.fq.gz
sample1_BC_TCCTGAGC_READ1.fq.gz
sample1_BC_TCCTGAGC_READ2.fq.gz
sample1_BC_TCATGAGC_READ1.fq.gz
sample1_BC_TCATGAGC_READ2.fq.gz
sample1_BC_CCTGAGAT_READ1.fq.gz
sample1_BC_CCTGAGAT_READ2.fq.gz
sample1_BC_TAAGGCGA_READ1.fq.gz
sample1_BC_TAAGGCGA_READ2.fq.gz
```

```
sample1_BC_GCTACGCT_READ1.fq.gz
sample1_BC_GCTACGCT_READ2.fq.gz
sample1_BC_AGGCTCCG_READ1.fq.gz
sample1_BC_AGGCTCCG_READ2.fq.gz
sample1_BC_CTGCGCAT_READ1.fq.gz
sample1_BC_CTGCGCAT_READ2.fq.gz
```

⚠ **CRITICAL:** this step generates uncompressed fastq files, be sure you have enough disk space available (e.g., 10× the size of compressed files at the end of step 1).

23. Align reads to reference genome. In principle any aligner could be used to achieve this, we only tested bwa.<sup>13</sup> We recommend default bwa options for alignment. Options impacting performance, such as number of threads, could be set matching the hardware in use.

```
$ bwa mem [BWA OPTIONS] ${SAMPLE}_BC_{BARCODE}_READ[12].fq.gz | \
  samblaster | \
  samtools sort -T ${SAMPLE}_BC_${BARCODE}_tmp -o ${SAMPLE}_BC_${BARCODE}.bam
```

This step must be repeated for each barcode and for each sample. We suggest samblaster<sup>15</sup> to remove duplicated reads as it can be piped, but alternative strategies could be used as well.

24. Merge alignments according to their transposase.

```
$ samtools merge -cp ${SAMPLE}_Tn5.bam \
  ${SAMPLE}_BC_CGTACTAG.bam \
  ${SAMPLE}_BC_TCCTGAGC.bam \
  ${SAMPLE}_BC_TCATGAGC.bam \
  ${SAMPLE}_BC_CCTGAGAT.bam
$ samtools index ${SAMPLE}_Tn5.bam
$ samtools merge -cp ${SAMPLE}_TnH.bam \
  ${SAMPLE}_BC_TAAGGCGA.bam \
  ${SAMPLE}_BC_GCTACGCT.bam \
  ${SAMPLE}_BC_AGGCTCCG.bam \
  ${SAMPLE}_BC_CTGCGCAT.bam
$ samtools index ${SAMPLE}_TnH.bam
```

This step is repeated for each sample.

25. Compute read counts over chromatin states using bedtools.<sup>20</sup> We provide chromatin segmentation derived from ENCODE project<sup>25</sup> for human (hg38) and mouse (mm10) in the scatACC github repository (see [key resources table](#)).

```
$ ls sample1*.bam | sort -u > files.txt
$ bedtools multicov -bed ${CHROMATIN_SEG} -bams 'cat files.txt' > state_counts.bed
```

26. Compute the relative enrichment of read counts in heterochromatic regions.

```
$ python enrich_gc.py state_counts.bed files.txt
```

This script calculates the relative enrichment of TnH over Tn5 for all chromatin states. It generates three files.

- State\_Enrichment.png: a boxplot representation of the relative enrichments over chromatin segmentation groups. The boxplot is ordered according to decreasing mean (see section “[expected outcomes](#)”).
- State\_Enrichment.txt: a text file containing the average enrichment over chromatin segmentation groups for each sample.
- State\_enrichment\_state.txt: a text file with descriptive statistics for enrichments.

### Automated processing of single cell data

⌚ Timing: 2–3 days

In this section, we describe the steps for automated data analysis of single cell data. It requires the scGET tools listed in the Resource Table. We exemplify the process assuming two samples have been processed and they have been given the mock names “sample1” and “sample2”.

- Demultiplex sequencing reads. Refer to cellranger-atac manual to perform this step. Only cellranger-atac mkfastq step is required.
- Define variables into config.yaml file.
  - genome: full path to genome indices for bwa. Must include genome prefix as well.
  - bed\_file: full path to a bed file containing the intervals to be analyzed. We suggest and provide 5kb fixed-size windows over the entire genome.
  - threads: number of computing threads.
  - cell\_number: the expected number of cells (see step 12).
  - scatacc\_path: the full path to the scatACC tools directory.
  - input\_path: path to the directory containing fastq files. Note that paths defined at step 3 will be relative to this path.
  - output\_path: path to the output directory.
  - input\_list: the path to input file containing the list of fastq files (see step 3).

```
—
barcodes:
  {'tn5': ['CGTACTAG', 'TCCTGAGC', 'TCATGAGC', 'CCTGAGAT'], 'tnh': ['TAAGGCGA', 'GCTACGCT', 'AGGCTCCG', 'CTGCGCAT']}
genome: PATH_TO_GENOME_FASTA_FILE
bed_file: PATH_TO_REGIONS_FILE
threads: 8
cell_number: 5000
scatacc_path: PATH_TO_scatacc_DIRECTORY
```

```
input_path: ROOT_PATH_TO_FASTQ_FILES
input_list: PATH_TO_INPUT_FILE_LIST
output_path: PATH_TO_OUTPUT_FOLDER
binary: False
tnh: True
tn5: True
atac: False
...
```

Optional values include the possibility to produce binary counts and to analyze only Tn5 or TnH barcodes. The pipeline could also be used to analyze scATAC-seq files, setting “atac” to True and both “tn5” and “tnh” to False.

29. Create a text file named `input_file.txt` with information about sequence reads separated by space. Each line contains three fields: the path to fastq file, the read number and the name of sample. The path to fastq files could be absolute or relative to step 2.f. The path of `input_file.txt` should be set in step 2.h.

```
path/to/sample1_file1_R1.fastq.gz 1 sample1
path/to/sample1_file1_R2.fastq.gz 2 sample1
path/to/sample1_file1_R3.fastq.gz 3 sample1
path/to/sample1_file2_R1.fastq.gz 1 sample1
path/to/sample1_file2_R2.fastq.gz 2 sample1
path/to/sample1_file2_R3.fastq.gz 3 sample1
path/to/sample2_file1_R1.fastq.gz 1 sample2
path/to/sample2_file1_R2.fastq.gz 2 sample2
path/to/sample2_file1_R3.fastq.gz 3 sample2
path/to/sample2_file2_R1.fastq.gz 1 sample2
path/to/sample2_file2_R2.fastq.gz 2 sample2
path/to/sample2_file2_R3.fastq.gz 3 sample2
```

30. Move to scGET tools directory where the github repository has been cloned (hereafter defined as `scGET_tools_DIR`) and run the workflow specifying the path to `config.yaml` file (step 2) and, possibly, the number of threads/processes to be spawned in parallel by the workflow manager (`SLURMTHREADS`). Currently only slurm workflow manager is fully supported.

```
$ cd ${scGET_tools_DIR}
$ snakemake -profile=slurm \
  -threads=${SLURMTHREADS} \
  -configfile=${PATHTO}/config.yaml
```

## Quality control

⌚ Timing: 15–45 min

In this section, we describe how to control if scGET-seq properly enriched reads on heterochromatic regions of the genome. This step can only be executed if epigenetic annotation is available. We provide annotation files, derived from the ENCODE project, for hg38 and mm10 genomes as part of scatACC tools. This step evaluates enrichment at “bulk level”, i.e., it doesn’t take into account information of single cells.

31. Create a list of files containing all deduplicated bam files and count the number of reads over intervals defining chromatin segmentation using bedtools.<sup>20</sup>

```
$ cd $OUTPUT_PATH
$ find . -name *_bcdedup.bam | sort -u > files.txt
$ bedtools multicov -bed ${CHROMATIN_SEG} -bams 'cat files.txt' > state_counts.bed
```

32. Run the QC script, specifying the output files of the previous step.

```
$ python enrich_qc_bc.py state_counts.bed files.txt
```

## Data preprocessing

⌚ Timing: 1–5 h

In this section, we describe the minimal code snippet used to read data and perform preprocessing. The entire workflow is executed in python. Duration of this step can be influenced by the number of samples and the computing power.

33. Import required libraries.

```
import scanpy as sc
import cellrank as cr
import scipy.sparse as ssp
import statsmodels.api as sm
import sklearn.preprocessing as skp
import tensorly as tl
import tensorly.decomposition as tld
import copy
```

34. Define statistical functions to fit Zero Inflated Poisson (ZIP) distributions. This function will be used to fit distributions of Tn5 and TnH read counts.

```
def fit_zip_mme(M):
    U = skp.StandardScaler(with_mean=False).fit(M.T)
    m = np.abs(U.mean_)
    s2 = U.var_
    mask = m >= s2
    l_m = np.abs((s2+m**2)/m - 1)
    p_m = (s2-m)/(s2 + m**2 - m)
    p_m[mask] = 0
    l_m[mask] = m[mask]
    return(l_m, p_m)
```

35. Read data and filter background entries. In the current example, the directory will contain files “sample1.h5ad” and “sample2.h5ad”. Different samples need to be imported and concatenated into a single data file. Mitochondrial chromosome information, which usually accumulate a large fraction of counts, should be removed to properly fit the ZIP distribution. Data are stored in a AnnData object.<sup>26</sup> As such it can contain multiple count matrices in slots named “layers”. The current workflows stores counts from both transposases in two layers named ‘Tn5’ and ‘TnH’. For more information about the AnnData object, refer to its official guide (<https://anndata.readthedocs.io>).

```
samples = ['sample1', 'sample2']
data = [None for x in samples]
for x in range(len(samples)):
    data[x] = sc.read(f"{samples[x]}/{samples[x]}.h5ad")
adata = data[0].concatenate(data[1:], batch_categories=samples)
autosomes = [x for x in adata.var_names if not x.startswith('chrM')]
adata = adata[:, autosomes]
cells = [x for x in adata.obs_names if not x.startswith('Back') and not x.starts
with('GGGGGGGGGGGGGGG-')]
adata = adata[cells]
```

**Optional:** regions that are blacklisted, such as genome gaps or centromeric repeats, could be removed at this stage. We provide blacklist overlap with 5kb windows in the same file used in step 27.b.

```
df = pd.read_table("hg38_5kbin_dhs_counts_blcunts.txt", header=None, index_col=0)
adata.var['DHS_Count'] = df[1]
adata.var['blacklist_coverage'] = df[2]
adata = adata[:, adata.var['blacklist_coverage'] == 0]
```

36. Fit Zero Inflated Poisson distributions over data.

```
x5 = adata.layers['Tn5']
xh = adata.layers['TnH']
l_h, p_h = fit_zip_mme(xh)
l_5, p_5 = fit_zip_mme(x5)
adata.obs['lambda_TnH'] = l_h.ravel()
adata.obs['lambda_Tn5'] = l_5.ravel()
adata.obs['p0_TnH'] = p_h.ravel()
adata.obs['p0_Tn5'] = p_5.ravel()
```

37. Filter cells with low coverage or excess of zero counts. The number of cells to retain should be roughly in the same order of magnitude with the number of cells defined at step 12. In our experience, a coverage not lower than 2,000 reads (as sum of Tn5 and TnH reads.) is suitable for most experiments (twice the coverage suggested elsewhere for analysis of scATAC-seq data<sup>27</sup>). In addition, we usually filter out cells with less than 90% of regions with zero counts.

```
nnz = (x5 + xh).nonzero()
u, c = np.unique(nnz[1], return_counts=True)
adata = adata[:, u[c > np.nanpercentile(c, 90)]]
mask = adata.obs['coverage'] > 2000
adata = adata[mask]
l_h = adata.obs['lambda_TnH']
l_5 = adata.obs['lambda_Tn5']
p_h = adata.obs['p0_TnH']
p_5 = adata.obs['p0_Tn5']
```

38. Define the support for ZIP distribution and the number of random samples to evaluate the empirical cumulative function.

```
dmin, dmax = np.min(x5-xh), np.max(x5-xh)
dmax += 1
Xr = np.arange(dmin, dmax)
ZIP = sm.distributions.zipoisson
ECDF = sm.distributions.empirical_distribution.ECDF
n_rvs = 100000
```

39. Calculate the accessibility according to ZIP distributions.

**Note:** Since probabilities are extracted from the empirical cumulative density function, computed on randomly generated numbers, accessibility values can fluctuate in multiple iterations of the same routine.

```
Diff = adata.layers['Tn5'] - adata.layers['TnH']
adata.layers['log_cdf'] = np.zeros(adata.shape)
for x in range(Diff.shape[0]):
    V = Diff[x].A.squeeze()
    zmask = (adata.layers['Tn5'][x].A.squeeze() == 0) & (adata.layers['TnH'][x].A.squeeze() == 0)
M = ZIP(l_5[x], p_5[x]).rvs(n_rvs) - ZIP(l_h[x], p_h[x]).rvs(n_rvs)
pr = ECDF(M, side='right')(Xr)
pl = ECDF(M, side='left')(Xr)
tail_thr = Xr[pr >= 0.5][0] # the thr for tails
adata.layers['log_cdf'][x][V < tail_thr] = np.log(pl[(V[V < tail_thr] + np.abs(dmin)).astype(int)])
adata.layers['log_cdf'][x][V >= tail_thr] = -np.log(1 - pr[(V[V >= tail_thr] + np.abs(dmin)).astype(int)])
adata.layers['log_cdf'][x][zmask] = 0
adata.layers['log_cdf'][adata.layers['log_cdf'] == -np.inf] = -np.ceil(-np.log(1/n_rvs))
adata.layers['log_cdf'][adata.layers['log_cdf'] == np.inf] = np.ceil(-np.log(1/n_rvs))
adata.layers['log_cdf'] = ssp.csr_matrix(adata.layers['log_cdf'])
```

**Optional:** Normalize and log transform data. This step affects the default count matrix in the object, which stores only Tn5 signals. Since the subsequent analysis of scGET-seq data rely on the difference between Tn5 and TnH data, this step is not necessary and it could be used only to look at scGET-seq data as they were produced by a standard scATAC-seq protocol.

```
sc.pp.normalize_total(adata)
sc.pp.log1p(adata)
```

40. Compute tensor decomposition of Tn5 and TnH data. Each region/cell data point in scGET-seq is represented by a tuple of two numbers, one from each enzyme. The two count matrices are then coupled and can be represented as a tensor. Similarly to what is done in scATAC-seq, where Singular Value Decomposition is applied to reduce the dimensionality, we here perform tensor decomposition.

```
pf = adata.layers['Tn5'].sum(1).A.ravel() + adata.layers['TnH'].sum(1).A.ravel()
x5 = np.log1p(ssp.diags(pf.mean()/pf)@adata.layers['Tn5'])
xh = np.log1p(ssp.diags(pf.mean()/pf)@adata.layers['TnH'])
GET = np.array([x5.A, xh.A])
GET = tl.tensor(GET)
rank = 200
ttd = tld.tensor_train(GET, rank=[1, 1, rank, 1])
```



```

adata.obsm['X_ttd'] = ttd.factors[1].reshape((GET.shape[1], rank))

adata.uns['ttd_weights'] = ttd.factors[0].squeeze()

adata.layers['NTn5'] = x5

adata.layers['NTnH'] = xh

```

41. Compute neighbors and UMAP embeddings. We use `bbknn`<sup>28</sup> for batch integrations, other batch integration strategies could be used. For a detailed overview on integration methods, refer to <sup>29</sup> and <sup>30</sup>.

```

n_samples = len(samples)

n_neighbors = int(np.sqrt(adata.shape[0])/2)

n_neighbors=n_neighbors//n_samples

sc.external.pp.bbknn(adata, use_rep='X_ttd', n_pcs=20,
                    neighbors_within_batch=n_neighbors,
                    metric='angular'
                    )

sc.tl.umap(adata)

```

In the following box we provide an example of integration using Harmony.<sup>31</sup>

```

n_neighbors = int(np.sqrt(adata.shape[0])/2)

sc.external.pp.harmony_integrate(adata, basis='X_ttd',
                                adjusted_basis='X_ttd_harmony',
                                max_iter_harmony=100
                                )

sc.tl.neighbors(adata, metric='cosine', use_rep='X_ttd_harmony',
               n_neighbors=n_neighbors)

sc.tl.umap(adata)

```

42. Compute cell clusters. We here use `schist` library<sup>22</sup> and select the first level of the hierarchy with more than one group defined.

```

import schist as scs

scs.inference.nested_model(adata)

modularity = adata.uns['schist']['nsbm']['stats']['modularity']

level = np.where(modularity > 0)[0][-1]

sc.pl.umap(adata, color=f'nsbm_level_{level}',
           title='Clusters', frameon=False)

```

*Optional:* Compute the global accessibility score.

```
n_open = np.sum(adata.layers['log_cdf'] > 0, 1).A.squeeze()
n_close = np.sum(adata.layers['log_cdf'] < 0, 1).A.squeeze()
adata.obs['r_open'] = n_open / (n_open + n_close)
```

*Optional:* Compute cell-wise accessibility moments, these will be used for cellrank analysis.

```
adata.layers['MTn5'] = ssp.csr_matrix.dot(adata.obsp['connectivities'], adata.X)
M = ssp.csr_matrix.dot(adata.obsp['connectivities'], adata.layers['log_cdf'])
adata.layers['Macc'] = M
```

*Optional:* Compute cellrank kernels. Use a velocity kernel to model accessibility and a connectivity kernel to regularize. Cellrank<sup>21</sup> is a toolkit to describe cellular dynamics in single-cell data. A cellrank kernel is a computational framework to compute cell-cell transition probabilities.

```
vk_cdf = cr.tl.kernels.VelocityKernel(adata, xkey='MTn5', vkey='Macc')
vk_cdf.compute_transition_matrix(scheme='cosine')
ck = cr.tl.kernels.ConnectivityKernel(adata)
ck.compute_transition_matrix()
combk = .8 * vk_cdf + .2 * ck
combk.compute_transition_matrix()
```

*Optional:* Compute macrostates. A macrostate is a coarse-grained group of cells having which is unlikely to transition to other macrostates. In this example we only compute two macrostates. Refer to cellrank manual pages on how to inspect cellrank results and calculate the optimal number of states.

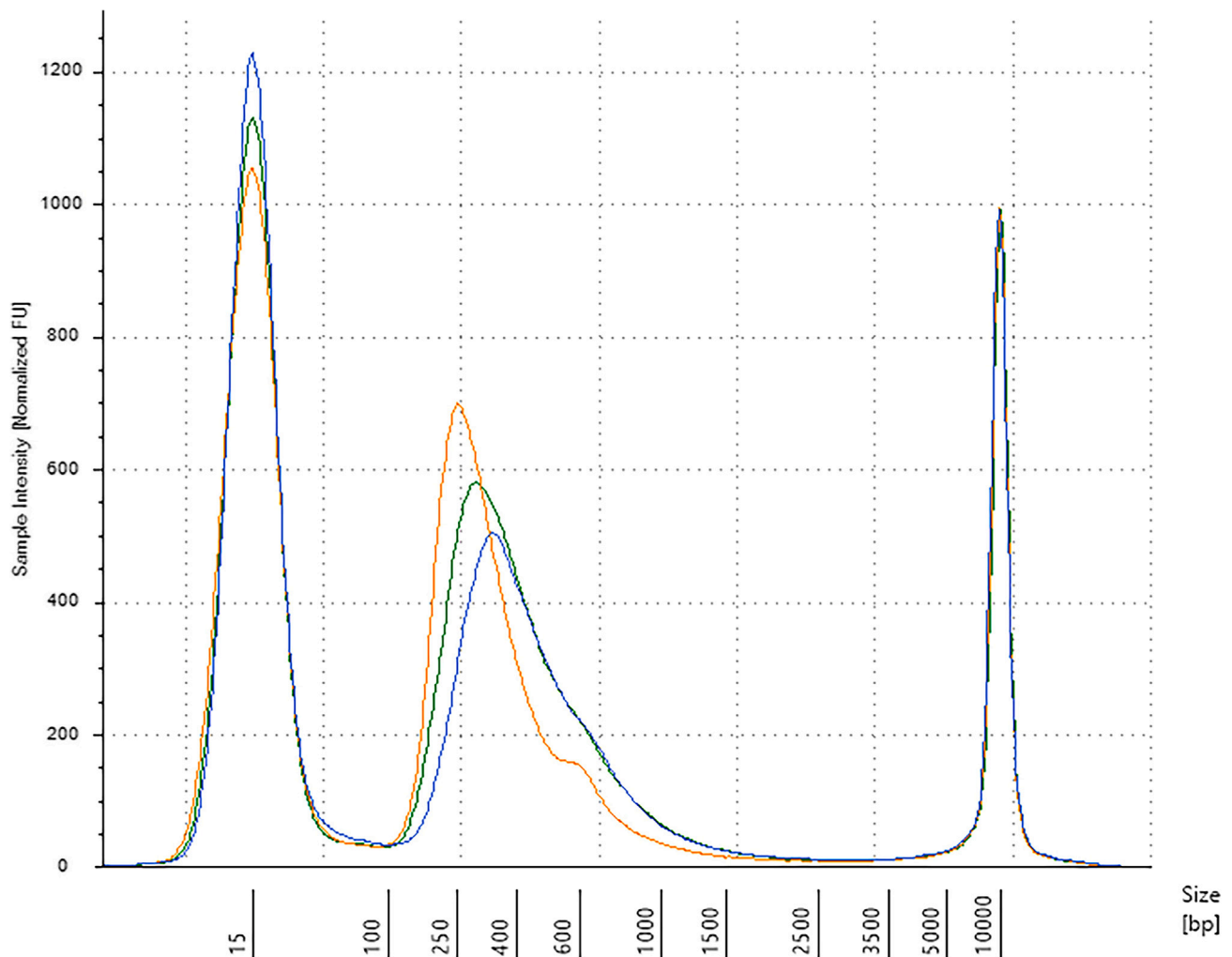
```
g_fwd = cr.tl.estimators.GPCCA(combk)
g_fwd.compute_schur()
g_fwd.compute_macrostates(n_states=2)
```

## EXPECTED OUTCOMES

### Transposome efficiency

Before starting with scGET-seq experiments, it is mandatory to check transposomes efficiency by performing bulk gDNA tagmentation and comparing resulting library distribution profiles with positive controls (previously validated transposomes complexes or commercially available alternatives, such as Nextera, Illumina; or Tagmentase (Tn5 transposase) – loaded, Diagenode).

Fragment distribution is expected to be centered in the 300–400 bp range (Figure 2).



**Figure 2.** High sensitivity D5000 ScreenTape assay showing fragment size distribution of gDNA tagmentation products obtained using Tn5 (blue trace), TnH (orange trace), vs Nextera tagmentation profile (green line)

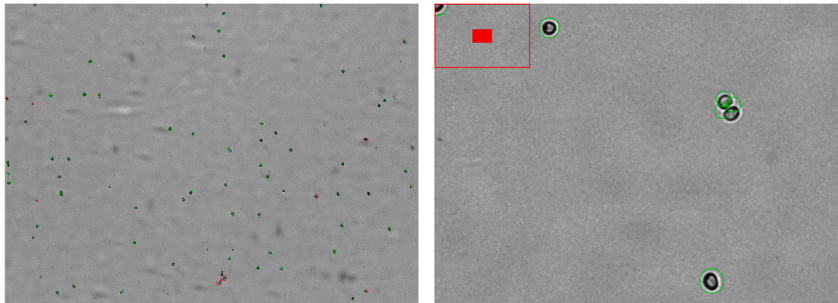
### Nuclei permeabilization

Starting from high quality single cell suspension with high vitality is crucial to achieve high quality scGET-seq data. Specifically, it is ideal to start from a >90% vitality cell suspension, assessed by trypan blue exclusion assay (Figure 3). If samples have a lower viability, we suggest to perform a dead cell removal step (e.g., depletion by magnetic labeling such as Milteny Dead Cell Removal kit, cat. No. 130-090-101; levitation technology such as LeviCell from LevitasBio, part number PN 1000001; or others).

Permeabilized nuclei should be almost 90% positive to vitality test. Monitor lysis efficacy microscopically. Nuclei margins should be round and well defined (Figure 4)

### Library profile

In a typical HeLa or Caki-1 scGET-seq experiment, aiming at 5,000 nuclei, we routinely expect a library concentration between 5 and 10 ng/ $\mu$ L. Library should result in a fragmentation profile similar to Figure 5. Please note variability in library profile across samples; this is not directly affecting data quality.



**Figure 3. Representative image of a Caki-1 cell suspension having >90% vitality.**

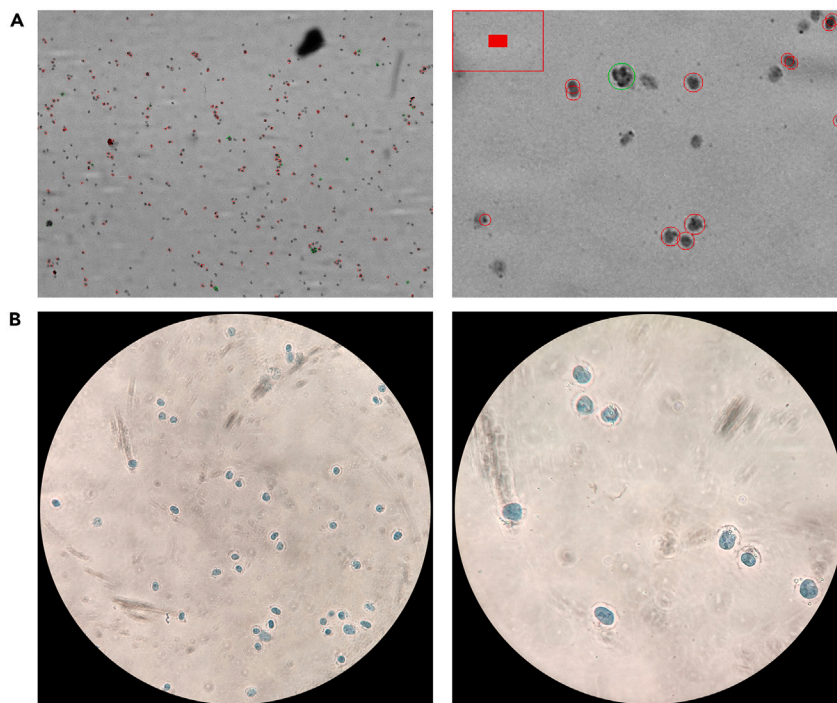
Caki-1 cells were stained with Trypan Blue and image taken using Logos Luna FL automated cell counter. Cells positive for trypan blue dye retention were automatically identified by the software and labeled with a red circle.

### Barcode demultiplexing

Steps 22 and 30 separate reads based on barcodes inserted in transposome-specific MEDS using tagdust<sup>16</sup> which generates a log file of the demultiplex process, one for each sample, reporting the rate of success of extraction. This rate is expected to be higher than 95% in a well-behaved experiment.

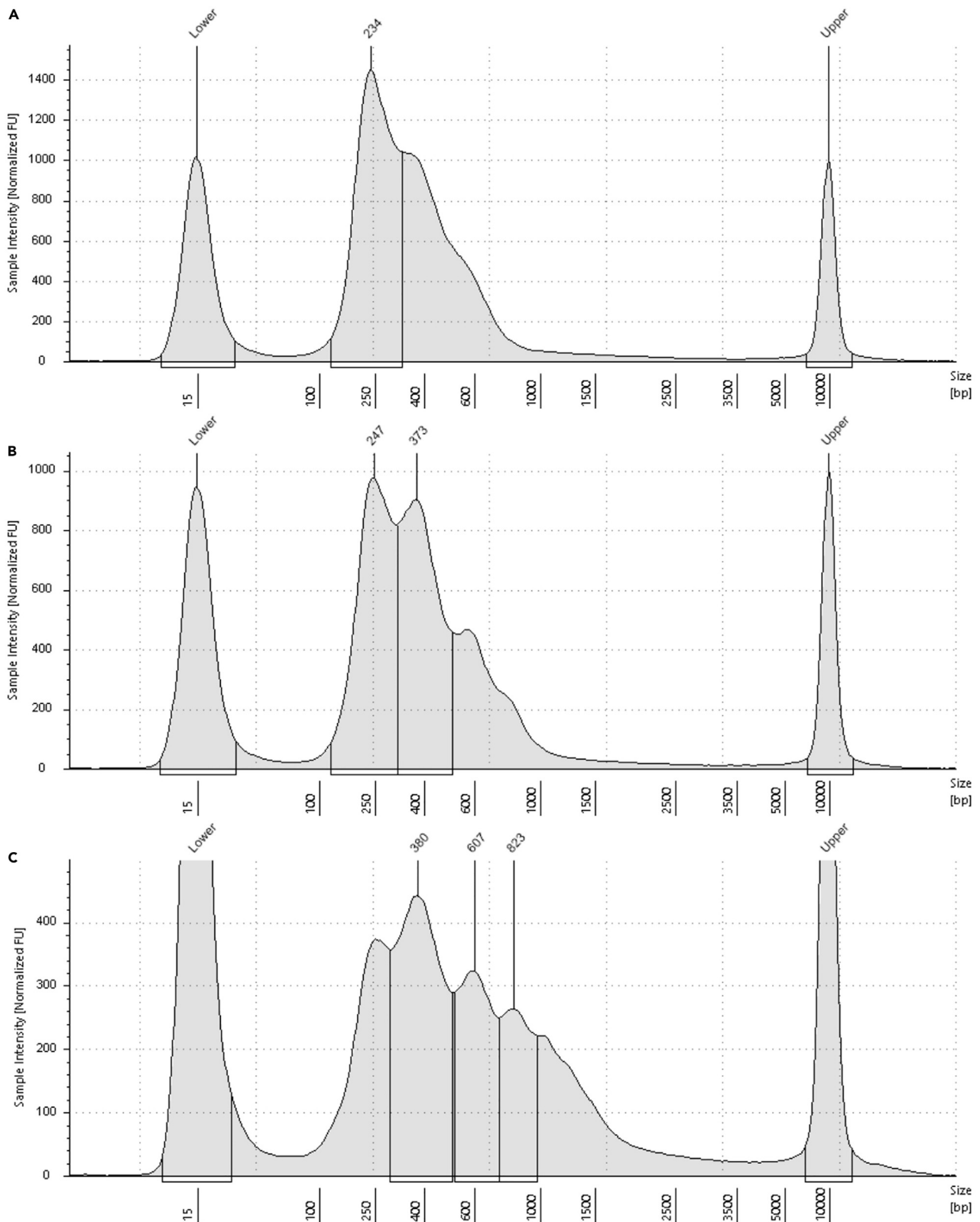
### Heterochromatin enrichment

Steps 26 and 32 generate a boxplot representing the relative enrichment over heterochromatic regions, in decreasing order (Figure 6). In a typical experiment, the states at the top of the plot are associated to heterochromatin and are characterized by an enrichment ratio higher than 1. At the



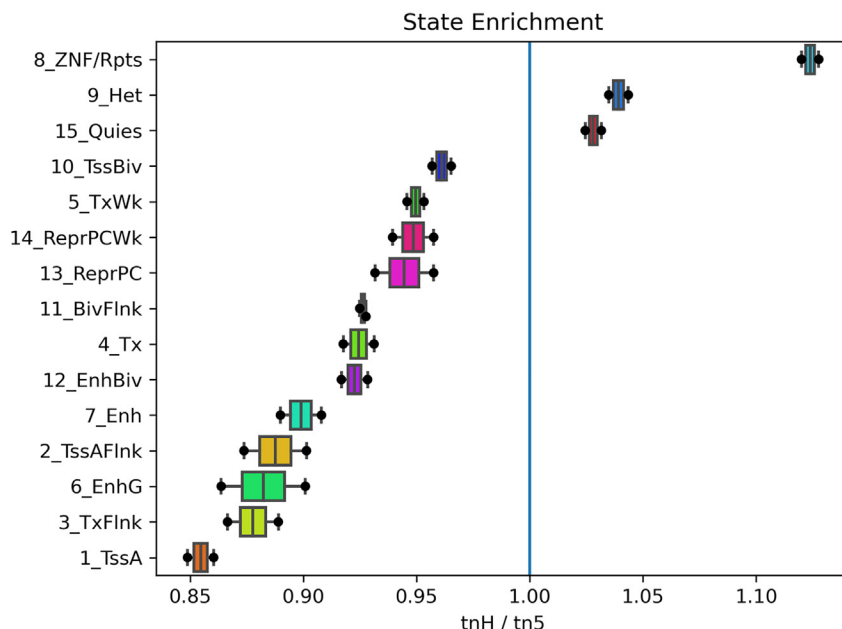
**Figure 4. Representative image of permeabilized Caki-1 cells after 3' in lysis buffer**

(A and B) Cells were stained with Trypan Blue and (A) image taken using Logos Luna FL automated cell counter. Cells positive for trypan blue dye retention were automatically identified by the software and labeled with a red circle. (B) images of permeabilized cells at the inverted microscope. Note nuclei shape with round, clean margins.



**Figure 5. Illustrative TapeStation analysis of scGET-seq library**

Panels A to C represent profiles obtained from the tagmentation of 5'000 nuclei derived from pancreatic cancer cells.



**Figure 6. The boxplots show the typical distribution of TnH/Tn5 enrichment over chromatin state segments**

Boxplots are sorted in decreasing order of average enrichment. A well-behaved experiment should have heterochromatic-related segments with enrichment ratio  $>1$ .

bottom of the plot it is expected to find states associated to accessible chromatin (i.e., TSS or enhancer associated). If two plots have been generated from bulk sequencing and single cell libraries, the enrichments should be consistent.

Depending on the cell type under investigation, the chromatin structure can be different and raise different enrichment profiles. In particular, we found that pluripotent cells have smaller enrichments, while retaining the same ordering of chromatin states (i.e., states marking TSSs and accessible regions at the bottom, states marking heterochromatic features at the top, [Figure 7](#))

### Barcode enrichment

For each sample defined at step 29, a directory containing output files will be created. Among these files, `$SAMPLE_cell_barcode_counts.png` shows the enrichment of unique cell barcodes found in the sample ([Figure 8](#)). The left part of the curve, highlighted in blue, indicates the barcodes that have been selected, the right part (in black) indicates the background signal. In a well-behaved experiment, the profile will show a bump for which there is evidence of a knee. The position of the knee may not correspond to the user-selected threshold, however it will be within a order of magnitude from it

### Alignment statistics

Automated analysis of single cell data (step 30) generates, for each sample, a text file with basic alignment statistics. Each file collects statistics from each barcode, so that there are 8 fields.

Alignment statistics. Each column refers to one transposase barcode. A statistics file is generated for each sample analyzed.

#### Alignment statistics

	AGGCTCCG	CCTGAGAT	CGTACTAG	CTGCGCAT	GCTACGCT	TAAGGCGA	TCATGAGC	TCCTGAGC
Total	122277866	69040434	53347578	74573820	69562908	136228140	73431050	81380144
Mapped	114467196	63018472	49594491	70049547	65253479	127386912	66656392	74200452

(Continued on next page)

Continued

	AGGCTCCG	CCTGAGAT	CGTACTAG	CTGCGCAT	GCTACGCT	TAAGGCGA	TCATGAGC	TCCTGAGC
Duplicated	79480600	38009634	29785022	48154944	43916444	90850385	40316608	45505311
In Cell	91746302	47216266	38201488	56465560	54537622	96348764	50077582	56428870
%Mapped	0.9361	0.9128	0.9296	0.9393	0.938	0.9351	0.9077	0.9118
%Duplicated	0.6944	0.6032	0.6006	0.6874	0.673	0.7132	0.6048	0.6133
%In Cell	0.7503	0.6839	0.7161	0.7572	0.784	0.7073	0.682	0.6934

Total number of reads is a function of the sequencing throughput. Values should be consistent across different barcodes, although some barcodes (CCTGAGAT and CTGCGCAT in particular) may result in smaller libraries. The rate of duplicates is a function of the library complexity, we usually observe high values which should be lower than 0.9. The fraction of reads in cell is an indirect measure of the tagmentation efficiency and single cell associated fragments barcoding, it should be higher than 0.5.

### Single cell data processing and cell embedding

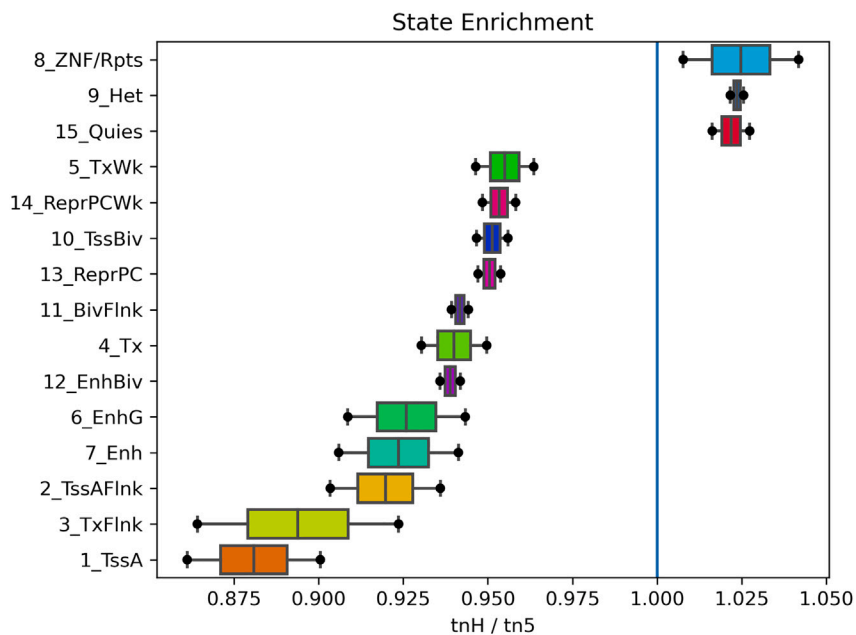
If count matrices have been computed over 5kb fixed-size windows, the raw data will contain approximately 600k features for a human genome. Depending on the thresholds set in step 37, more than 50k will be likely retained for further analysis.

scGET-seq is sensitive to cell types and states: analysis of chromatin accessibility should reveal the coarse-grained composition of cell types under investigation (Figure 9)

### LIMITATIONS

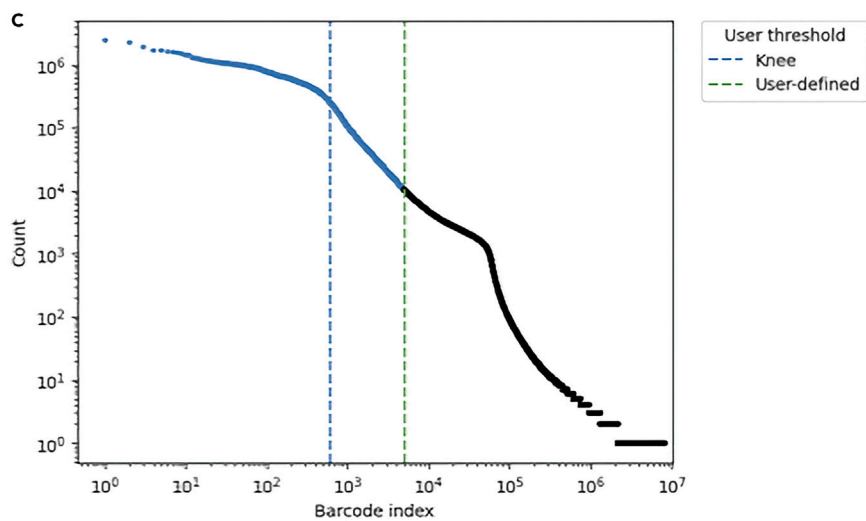
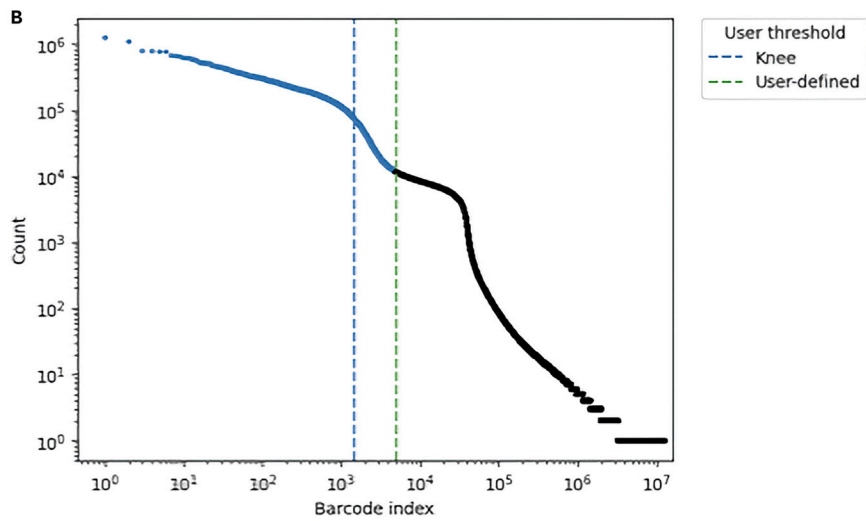
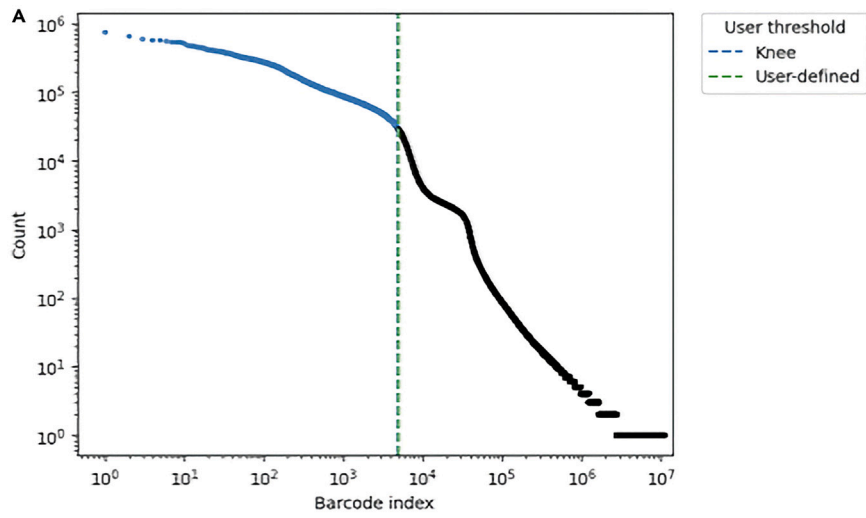
The present scGET-seq protocol has been thoroughly tested on fresh and cryopreserved samples (cell lines, organoids, dissociated tissues). It has not been tested on FFPE and flash-frozen samples. Those type of samples are already known to produce poor quality ATAC data, although several papers proposed modified ATAC approaches.<sup>32,33</sup>

Current version of scGET-seq is only suitable for probing H3K9me3 associated heterochromatin. Other heterochromatin makers, such as H3K27me3 or H4K20me3, are not probed.



**Figure 7. The boxplots show the distribution of TnH/Tn5 enrichment over chromatin state segments in pluripotent cells**

Cells with a higher rate of chromatin accessibility display milder enrichments over heterochromatic segments. However, these are properly sorted.





**Figure 8. Cell barcode abundance plots computed by UMItools for three well-behaved experiments**

The software is able to identify a knee in the left part of the curve (blue dashed line, panel A). While the location of the knee does not necessarily correspond to the selected number of cells (green dashed line, panels B and C), the difference is within an order of magnitude.

**TROUBLESHOOTING**

**Problem 1**

Poor nuclei permeabilization.

At step 8 (section “Nuclei preparation and tagmentation”) after performing permeabilization with Lysis Buffer on ice, visual evaluation of nuclei shows low number of trypan blue stained cells (<90%).

**Potential solution**

- **Incorrect Lysis buffer composition.** If detergents and digitonin were not added, or added in incorrect concentrations, this can lead to sub-optimal permeabilization efficiency. Always use fresh Lysis buffer. Prepare new buffer.
- **Incorrect Lysis time.** Always perform permeabilization test before performing actual experiment. Each sample type need careful evaluation of permeabilization condition. Work in 30” steps until you find the time required to achieve  $\geq 90\%$  permeabilization (estimated by trypan blue exclusion).

**Problem 2**

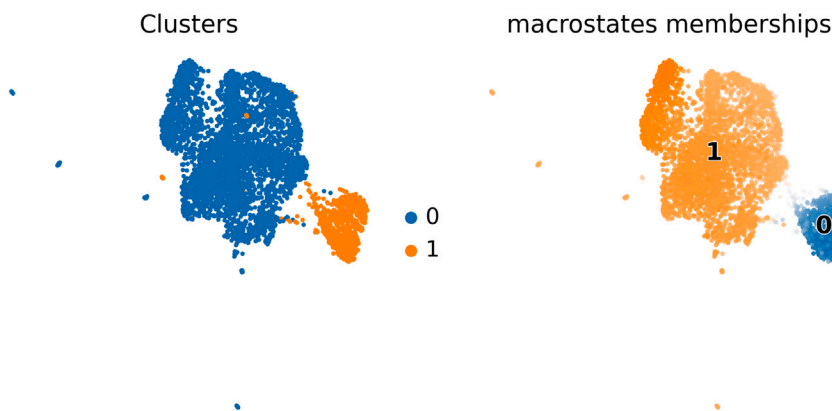
Library yield.

At step 19, section “Transposition, GEM formation and library preparation” library fluorometric quantification (e.g., Qubit) shows a concentration <0.5 ng/ $\mu$ L.

**Potential solution**

Library yield is a function of tagmentation efficiency, target nuclei number, library preparation efficiency. Although we do not recommend to proceed recovering current experiment, there are several points to check to avoid failure.

- **Transposomes are not functional.** See step 3 of “[transposome assembly and validation](#)” section. As scGET-seq requires in-house assembled transposomes complexes, make sure to use validated transposomes. Check transposomes activity by bulk tagmentation. Use new transposomes, prepare new MEDS.
- **Incorrect dilution of permeabilized nuclei in the transposition reaction** (step 15 of the “Transposition, GEM formation and library preparation” section). Make sure to perform accurate counting at step 12.



**Figure 9. UMAP embedding showing cell mixture (Caki + HeLa) colored by cell cluster computed by *schist* and by macrostate membership computed by *cellrank***

- **Errors in library preparation protocol.** For library preparation, make sure to follow the 10× Genomics User Guide corresponding to your Reagent Kit version (in the text we refer to [Version 1.1](#)). Main cause of library failure is use of incorrect index plate (therefore no amplification is expected) or error in PCR program in the GEM incubation (e.g., no gap filling step of 3' at 72°C).

### Problem 3

Incorrect library size distribution.

At step 19, section “Transposition, GEM formation and library preparation”, your library profile is different from expected. For example, there is an enrichment (>50%) in high MW fragments (>800 bp). Lower molecular weight (<150 bp) fragments can be visible in the TapeStation trace, but should not affect sequencing quality.

### Potential solution

Perform additional purification step using SPRIselect or AMPureXP beads cleanup (right-side clean up, 0.5× ratio). However, if you decide to proceed with sequencing, be aware that resulting data could possibly not represent the complexity of the original sample. We suggest repeating sample processing from Nuclei preparation.

### Problem 4

Low rate of transposase demultiplexing.

The expected success rate for tagdust in identifying transposase barcode is usually higher than 95%. A lower rate can be due to library contamination or sequencing problem, most probably due to incorrect sequencing primers (step 21).

### Potential solution

- If a sufficient number of reads has been processed, these can be used, otherwise library preparation and sequencing should be repeated from the beginning.
- Check if custom sequencing primers have been added in correct well position and at the correct concentration in the Illumina reagent cartridge. Failure in adding custom primer will result in no useful sequences (if sequencing only scGET-seq libraries, the PhiX read yield will be much higher than expected). Repeat sequencing on a low throughput sequencing cartridge/platform to assess failure is not due to custom primers.

### Problem 5

Lack of enrichment of TnH signal over heterochromatic features.

The enrichment plot generated at steps 26 and 32 should report heterochromatic features at the top of the figure, possibly with small variability across samples. While differences in enrichment can be due to chromatin context and are specific to cell types, high variability can be indicative of problematic samples ([Figure 10](#))

Identify the ill-behaved samples by inspecting the “State\_Enrichment.txt” file.

#### Samplewise TnH/Tn5 enrichments over chromatin states

	Sample A	Sample B	Sample C
1_TssA	0.83313687	0.84223407	0.64090355
2_TssAFlnk	0.91966016	0.93661643	0.69824365

(Continued on next page)

Continued

	Sample A	Sample B	Sample C
3_TxFlnk	0.89783742	1.02160831	0.48878109
4_Tx	0.94468709	0.93501682	0.70290321
5_TxWk	0.88032582	0.89134796	0.76050744
6_EnhG	0.93987416	0.92975318	0.65065172
7_Enh	0.95025836	0.97831866	0.69662704
8_ZNF/Rpts	1.27740049	1.13327923	0.77745233
9_Het	1.10811078	1.08030501	0.71561746
10_TssBiv	1.10278212	1.13401595	0.91990154
11_BivFlnk	0.95955523	1.00470737	0.80729406
12_EnhBiv	0.93547259	1.00528484	0.71975653
13_ReprPC	1.02093206	1.01676024	0.78374218
14_ReprPCWk	1.06005473	1.06387565	0.85020142
15_Quies	1.03159387	1.02889095	1.0970373

In the example provided, Sample C shows no enrichment over heterochromatin (9\_Het is the 9<sup>th</sup> most enriched feature), Sample B shows suspicious enrichment over putatively active features (3\_TxFlnk).

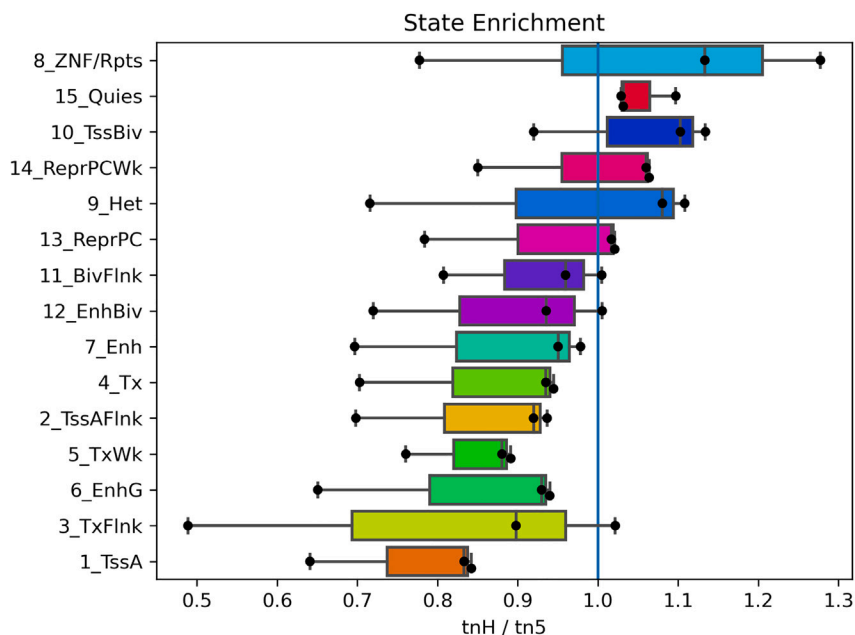
### Potential solution

Native chromatin structure was disrupted during nuclei processing. This can lead to altered TnH chromatin profiling, e.g., the transposomes will bind to euchromatin. Always work on ice, and work with high viability cell suspension. Repeat experiment.

### Problem 6

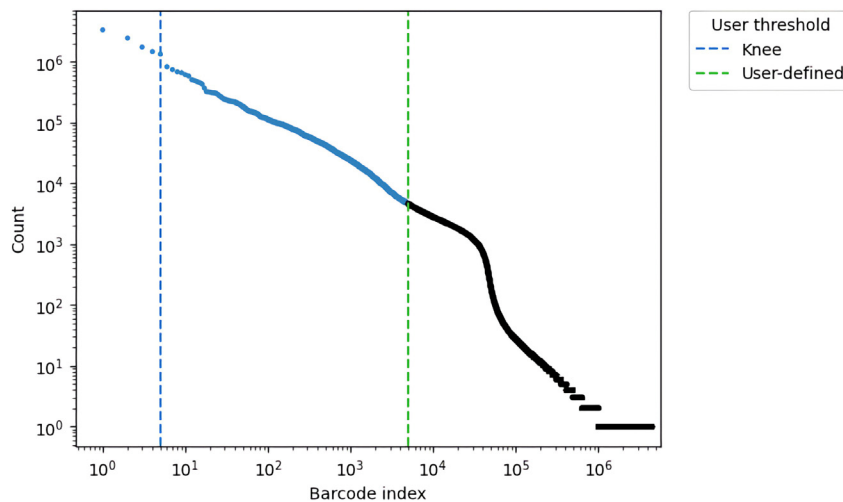
Low number of cells recovery rate.

When the rate of recovered cells is lower than expected, plot of counts of cell barcodes, generated after step 30 by UMI\_tools, will have a linear decay in the left part of the curve, with few cells (less



**Figure 10. Boxplots showing a ill-behaved experiment**

Some samples, represented by dots, do not show a proper enrichment over heterochromatin. The high variability affects the proper sorting.



**Figure 11. Cell barcode abundance plot computed by umi\_tools**

In a ill-behaved experiment, the software identifies the knee with difficulties, being the decay of enrichment almost linear (blue part of the line on the left).

than 100) accounting for the largest part of the sequenced reads (Figure 11). Under this condition, the rate of duplicated reads in the alignment statistics can also increase (>90%).

### Potential solution

- This result can be associated with low cell suspension vitality, leading to ambient DNA which is then fragmented during tagmentation step and encapsulated randomly in droplets creating background signal.
- This result can be also related to suboptimal cell concentration quantification. Work with high vitality cell preparation (live cells > 90%); consider removing dead cell from your cell suspension (e.g., depletion by magnetic labeling such as Milteny Dead Cell Removal kit; levitation technology such as LeviCell from LevitasBio or others).

### Problem 7

Low data complexity.

Low data complexity usually reflects into high data sparsity and low counts. Given the processing filters used in step 37, it is expected that about 10% of the genomic regions will be retained for analysis (e.g., ~50k regions for human genome). This percentage can be much smaller in case of low complexity (less than 5%, ~20k regions for human genome). In case of low complexity, high number of duplicates in alignment statistics (> 90%) could be found; however, the rate can be influenced also by other issues (see [problem 6](#)).

Such data should not be analyzed, as they do not reflect the biology of the system (Figure 12).

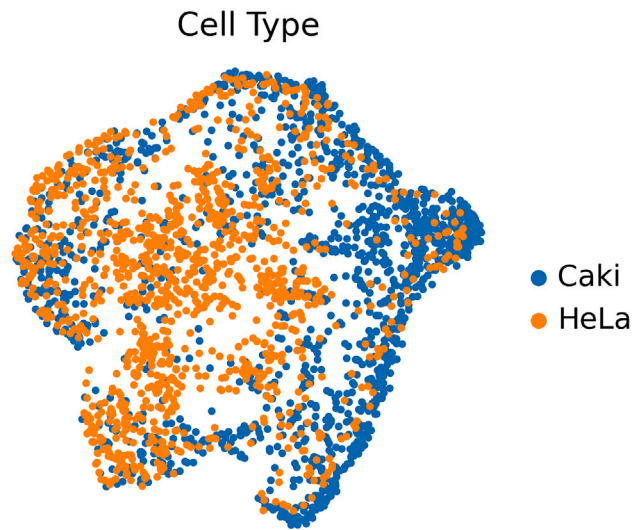
### Potential solution

Low complexity can be related to poor transposase tagmentation, due to improper nuclei permeabilization (see [problem 1](#)) or transposomes functionality (see [problem 2](#)).

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Francesca Giannese ([giannese.francesca@hsr.it](mailto:giannese.francesca@hsr.it)).



**Figure 12. UMAP embedding showing a cell mixture (Caki + HeLa) from an experiment with low library complexity**

In this case, scGET-seq data cannot discriminate between cell lines, this is reflected in almost completely overlapping cells in the embedding.

#### Materials availability

TnH transposase is covered by pending patent application (PCT WO2022167665). Purified protein can be obtained upon request to [lead contact](#).

#### Data and code availability

Raw data for the experiment described in the published article have been deposited to the Array Express platform (<https://www.ebi.ac.uk/arrayexpress/>) with the following ID: E-MTAB-9650. The published article includes all the code generated in this study.

#### ACKNOWLEDGMENTS

This work was supported by AIRC/CRUK/FC AECC Accelerator Award ‘Single Cell Cancer Evolution in the Clinic’ A26815 (AIRC number program 2279).

We thank all the members at the Center for Omics Sciences for discussions, support, and critical reading of the manuscript. We thank Dr. Timon Heide at the Human Technopole for his suggestions on QC procedure.

#### AUTHOR CONTRIBUTIONS

D.C. and F.G. wrote the manuscript. G.T. and D.L have reviewed and edited the manuscript.

#### DECLARATION OF INTERESTS

F.G., D.L., D.C., and G.T. have submitted a patent application (PCT WO2022167665), pending, covering TnH.

#### REFERENCES

1. Tedesco, M., Giannese, F., Lazarević, D., Giansanti, V., Rosano, D., Monzani, S., Catalano, I., Grassi, E., Zanella, E.R., Botrugno, O.A., et al. (2022). Chromatin velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. *Nat. Biotechnol.* 40, 235–244. <https://doi.org/10.1038/s41587-021-01031-1>.
2. de Pretis, S., and Cittaro, D. (2022). Dimensionality reduction and statistical modeling of scGET-seq data. Preprint at bioRxiv. <https://doi.org/10.1101/2022.06.29.498092>.
3. Greaves, M. (2015). Evolutionary determinants of cancer. *Cancer Discov.* 5, 806–820. <https://doi.org/10.1158/2159-8290.CD-15-0439>.

4. Desanlis, I., Kherdjemil, Y., Mayran, A., Bouklouch, Y., Gentile, C., Sheth, R., Zeller, R., Drouin, J., and Kmita, M. (2020). HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. *Nat. Commun.* *11*, 2491. <https://doi.org/10.1038/s41467-020-16317-2>.
5. Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., Kapourani, C.-A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., et al. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* *576*, 487–491. <https://doi.org/10.1038/s41586-019-1825-8>.
6. LaFave, L.M., Kartha, V.K., Ma, S., Meli, K., Del Priore, I., Lareau, C., Naranjo, S., Westcott, P.M.K., Duarte, F.M., Sankar, V., et al. (2020). Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* *38*, 212–228.e13. <https://doi.org/10.1016/j.ccell.2020.06.006>.
7. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* *109*, 21.29.1–21.29.9. <https://doi.org/10.1002/0471142727.mb2129s109>.
8. Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* *9*, 781. <https://doi.org/10.1038/s41467-018-03149-4>.
9. Bartosovic, M., Kabbe, M., and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* *39*, 825–835. <https://doi.org/10.1038/s41587-021-00869-9>.
10. Tatarakis, A., Behrouzi, R., and Moazed, D. (2017). Evolving models of heterochromatin: from foci to liquid droplets. *Mol. Cell* *67*, 725–727. <https://doi.org/10.1016/j.molcel.2017.08.022>.
11. Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* *28*, 2520–2522. <https://doi.org/10.1093/bioinformatics/bts480>.
12. Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* *24*, 2033–2040. <https://doi.org/10.1101/gr.177881.114>.
13. Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv.
14. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* *25*, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
15. Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* *30*, 2503–2505. <https://doi.org/10.1093/bioinformatics/btu314>.
16. Lassmann, T. (2015). TagDust2: a generic method to extract reads from sequencing data. *BMC Bioinf.* *16*, 24. <https://doi.org/10.1186/s12859-015-0454-y>.
17. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* *27*, 491–499. <https://doi.org/10.1101/gr.209601.116>.
18. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* *19*, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
19. Kossaifi, J., Panagakis, Y., Anandkumar, A., and Pantic, M. (2019). Tensorly: tensor learning in Python. *J. Mach. Learn. Res.* *20*, 1–6.
20. Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* *47*, 11.12.1–11.12.34. <https://doi.org/10.1002/0471250953.b1112s47>.
21. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. *Nat. Methods* *19*, 159–170. <https://doi.org/10.1038/s41592-021-01346-6>.
22. Morelli, L., Giansanti, V., and Cittaro, D. (2021). Nested Stochastic Block Models applied to the analysis of single cell data. *BMC Bioinf.* *22*, 576. <https://doi.org/10.1186/s12859-021-04489-7>.
23. Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R., et al. (2013). Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.* *8*, 2022–2032. <https://doi.org/10.1038/nprot.2013.118>.
24. Reznikoff, W.S. (2003). Tn5 as a model for understanding DNA transposition. *Mol. Microbiol.* *47*, 1199–1206. <https://doi.org/10.1046/j.1365-2958.2003.03382.x>.
25. Amin, V., Harris, R.A., Onuchic, V., Jackson, A.R., Charnecki, T., Paithankar, S., Lakshmi Subramanian, S., Riehle, K., Coarfa, C., and Milosavljevic, A. (2015). Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. *Nat. Commun.* *6*, 6370. <https://doi.org/10.1038/ncomms7370>.
26. Virshup, I., Rybakov, S., Theis, F.J., Angerer, P., and Wolf, F.A. (2021). anndata: annotated data. Preprint at bioRxiv. <https://doi.org/10.1101/2021.12.16.473007>.
27. Fang, R., Preissl, S., Li, Y., Hou, X., Lucero, J., Wang, X., Motamedi, A., Shiau, A.K., Zhou, X., Xie, F., et al. (2021). Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* *12*, 1337. <https://doi.org/10.1038/s41467-021-21583-9>.
28. Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.-E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* *36*, 964–965. <https://doi.org/10.1093/bioinformatics/bt2625>.
29. Tran, H.T.N., Ang, K.S., Chevrier, M., Zhang, X., Lee, N.Y.S., Goh, M., and Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* *21*, 12. <https://doi.org/10.1186/s13059-019-1850-9>.
30. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., and Theis, F.J. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* *19*, 41–50. <https://doi.org/10.1038/s41592-021-01336-8>.
31. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* *16*, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.
32. Zhang, H., Polavarapu, V.K., Xing, P., Zhao, M., Mathot, L., Zhao, L., Rosen, G., Swartling, F.J., Sjöblom, T., and Chen, X. (2022). Profiling chromatin accessibility in formalin-fixed paraffin-embedded samples. *Genome Res.* *32*, 150–161. <https://doi.org/10.1101/gr.275269.121>.
33. Yadav, R.P., Polavarapu, V.K., Xing, P., and Chen, X. (2022). FFPE-ATAC: a highly sensitive method for profiling chromatin accessibility in formalin-fixed paraffin-embedded samples. *Curr. Protoc.* *2*, e535. <https://doi.org/10.1002/cpz1.535>.