CIM-seq

Nathanael Andrews

Oncology and Pathology <https://orcid.org/0000-0003-4139-2904>

Martin Enge (✉ Martin.Enge@ki.se)

Oncology and Pathology

Method Article

Keywords: scRNA-seq, cell-cell interactions, cell communication

Posted Date: December 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.pex-1365/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

CIM-seq is a tool for deconvoluting RNA-seq data from cell multiplets (clusters of two or more cells) in order to identify physically interacting cell in a given tissue. The method requires two RNAseq data sets from the same tissue: one of single cells to be used as a reference, and one of cell multiplets to be deconvoluted. CIM-seq is compatible with both droplet based sequencing methods, such as Chromium Single Cell 3' Kits from 10x genomics; and plate based methods, such as Smartseq2. The pipeline consists of three parts: 1) Dissociation of the target tissue, FACS sorting of single cells and multiplets, and conventional scRNA-seq 2) Feature selection and clustering of cell types in the single cell data set - generating a blueprint of transcriptional profiles in the given tissue 3) Computational deconvolution of multiplets through a maximum likelihood estimation (MLE) to determine the most likely cell type constituents of each multiplet.

Introduction

Reagents

Lambda Exonuclease - New England Biolabs Catalog **#M0262L**

HotStart ReadyMix (KAPA HiFi PCR kit) - Kapa Biosystems Catalog **#KK2601**

psfTn5 Addgene - Catalog **#79107**

10% SDS solution - Teknova Catalog **#S0287**

SMARTScribe Reverse Transcriptase - Takarabio Catalog **#634888**

Magnesium chloride solution for molecular biology - Sigma Aldrich Catalog **#M1028**

Triton X-100 - Sigma Catalog **#93426**

KAPA HiFi PCR kit with dNTPs - Fisher Scientific Catalog **#NC0142652**

Recombinant RNase Inhibitor - Takarabio Catalog **#2313A**

Betaine 5M - Sigma Aldrich Catalog **#B0300**

ERCC RNA Spike-In Mix - Thermo Fisher Catalog **#4456740**

USB Dithiothreitol (DTT) Solution - Thermo Fisher Catalog **#707265ML**

DEPC-Treated Water - Thermo Fisher Catalog **#750023**

dNTP Mix - Thermo Fisher Catalog **#R0192**

Sera-Mag Speed Beads - GE Healthcare Catalog **#65152105050250**

Hard-Shell® 384-Well PCR Plates - BioRad Sciences Catalog **#HSP3801**

Qubit 1X dsDNA HS Assay Kit - Thermo Fisher Scientific Catalog **#Q33231**

PEG 8000 - Sigma Aldrich Catalog **#P2139**

TAPS - Sigma Aldrich Catalog **#T5130**

Oligonucleotides (all ordered from IDT using Standard desalting, except barcodes ordered in solution/plates):

Oligo-dT: AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT(N1:34333300)
(N2:25252525)

IS_PCR: 5'-AAGCAGTGGTATCAACGCAGAGT-3'

TSO: 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3'

ME-A: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'

ME-B: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'

ME-Rev: 5'-/5Phos/CTGTCTCTTATACACATCT-3'

Illumina-compatible barcodes used (Sxxx/Nxxx series, n=784) are available as a supplementary table in the manuscript.

Equipment

Procedure

A prerequisite for CIM-seq is the dissociation of cell lines or tissues into cell suspensions containing both single cells and multiplets. Dissociation procedure will have to be adjusted to the tissue and species of interest and a general method will therefore not be included here. For dissociation of murine lung, spleen, small intestine and colon as performed in the CIM-seq manuscript, please see the methods section of the CIM-seq manuscript.

We have divided this procedure section into three subsections:

- A) Cell sorting and library preparation
- B) Feature selection and clustering (Seurat)
- C) Multiplet deconvolution

The library preparation section will describe Smartseq2 (well based) library preparation. For 10x Genomics, consult the manufacturers instructions for library prep found at:

<https://support.10xgenomics.com/single-cell-gene-expression/library-prep/doc/user-guide-chromium-single-cell-3-reagent-kits-user-guide-v3-chemistry>.

A) Cell sorting and library preparation using Smartseq2:

1) Prepare lysis buffer mix (2µl per well):

NOTE: Reagents are prepared on ice, working quickly. ERCC is stored in single-use aliquots at -80 °C , thawed on ice and added last.

Reagent - µl per reaction

H2O - 1.31µl

RNase Inhibitor - 0.05µl

ERCC (1:600 000) - 0.05µl

Triton-X100 (10% solution) - 0.04µl

10mM dNTP - 0.5µl

100uM dT primer - 0.05µl

Total - 2µl

Add 2µl **lysis buffer mix** to each well. Cover with appropriate lids. Spin down.

Snap freeze on **dry ice**. Store until use at -80 °C

2) Sort single cells and multiplets (aggregates of multiple cells)

Sort into wells containing 2 µl **lysis buffer mix**.

Multiplets can be discerned from singlets by gating on the basis of FSC-W (Forward scatter - Width) and FSC-H (Forward scatter - Height) (see **Figure 1**).

Following sort, immediately seal with appropriate seals (approved for -80°C) and centrifuge at 2000 x g, 4°C, 5 min. Snap freeze on **dry ice**. Store until use at -80 °C.

3) Primer annealing

NOTE: Perform steps 3 and 4 as quickly as possible in order to prevent RNA degradation

Thaw plate and spin down briefly. Incubate in thermocycler at 72°C for 3 min to anneal primers. Place on ice immediately.

4) Reverse transcription

Prepare RT master-mix:

Reagent - Reagent volume

SmartScribe - 0.475µl

RNase Inhibitor - 0.125µl

5x First Strand buffer - 1µl

DTT (100mM) - 0.25µl

Betaine (5M) - 1µl

MgCl₂ (1M) - 0.03µl

TSO (100uM) - 0.05µl

H₂O - 0.07µl

Total - 3µl

Dispense 3µl per well, cover plate with new film and spin down. Incubate in thermocycler:

42°C for 90min.

70°C for 5 min.

4°C hold.

4) cDNA preamplification:

Prepare cDNA preamplification mix:

Reagents - Reagent volume

H₂O - 1.0688µl

Kapa HiFi HotStart ReadyMix (2x) - 6.25µl

IS_PCR primer (10uM) - 0.125µl

Lambda Exonuclease - 0.05625µl

Total - 7.5µl

Dispense 7.5 µl per well. Total reaction volume will be 12.5µl. Spin down and cover with new lid. Incubate in thermocycler with the following program:

1 cycle:

37°C for 30 min.

95°C for 3 min.

Repeat 18-24 cycles (depending on initial RNA concentration):

98°C for 20 sec.

67°C for 15 sec.

72°C for 4 min.

1 final elongation cycle:

72°C 5 min.

4°C min. hold

5) cDNA cleanup

For cDNA cleanup we prepare SPRI-beads in 20% PEG-8000 solution as described in:

https://openwetware.org/wiki/SPRI_bead_mix#Ingredients_for_50_mL_2

Using 20% SPRI-bead solution:

i. Add 0.7x the reaction volume of SPRI beads per well. Mix well by pipetting. (i.e 8.75 µl SPRI-bead solution for 12.5 µl reaction volume)

- ii. Incubate 5 min. Room temperature
- iii. Place on magnetic stand for 3 min.
- iv. Carefully remove supernatant
- v. Add 40 µl 80% EtOH and incubate 30 sec.
- vi. Remove EtOH (without disturbing the beads)
- vii. Wash again with EtOH. Make sure to remove as much ethanol as possible.
- viii. Allow beads to air-dry for 10 - 15 min.
- ix. Remove plate from magnetic stand
- x. Elute beads in 15 µl EB or TE buffer. Mix well by pipetting
- xii. Incubate 5 min. Room temperature
- xiii. Place on magnetic plate for 3 min.
- xiv. *Optional: Carefully transfer supernatant to a new plate*

6) cDNA quantification

In order to normalize cDNA amounts, measure concentration of randomly selected wells using Qubit HS dsDNA, adapted to a 96-well plate reader (May also be performed using standard Qubit protocol). Concentrations should be sampled both from wells containing single cells and multiplets in order to normalize separately.

- i. Add 97 µl of 1X Qubit HS dsDNA solution to a flat-bottom, black plate
- ii. Add 3 µl of cDNA sample
- iii. Add Standards (NOTE: We make a 8-step ladder from 0ng/µl → 10ng/µl Qubit Standard DNA in TE buffer)
- iv. Read in plate reader using 485nm excitation/528nm emission
- v. Calculate average well cDNA concentration for single cells and multiplets respectively

7) (optional) cDNA quality control

Using Agilent HS 5000 DNA chips (or equivalent) to profile cDNA fragment lengths (See example cDNA profile in Fig. 2.).

8) Dilute cDNA

Dilute cDNA in water based on average concentration from Qubit measurements.

Target concentration is 150pg/μl per well.

9) Tn5 digestion

Tn5 is produced from psfTn5 (Addgene #79107), purified to ~3mg/ml and assembled with Illumina Tn5 adapters (see oligos in Reagents) as in Picelli et al, 2014 (Ref. 1).

Prepare Tn5 master mix:

NOTE: TAPS-PEG Buffer contains PEG, which is viscous. Buffer should equilibrate to room temperature before use to allow proper mixing.

Reagent - Reagent volume

Nuclease free H₂O - 1.05μl

TAPS-PEG (50mM TAPS, 25mM MgCl₂, 40% PEG-8000) - 0.5μl

psfTn5 (loaded with 50μM MEDS-A/B) - 0.25μl

Total - 1.8µl

Dispense 1.8 µl per well in a new plate (tagmentation plate).

Add 0.7 µl cDNA (normalized to 150pg/µl) to tagmentation plate

Cover with new lid, mix well by vortexing plate and spin down.

Incubate in thermocycler at 55°C for 10 min.

Remove immediately and stop the reaction by adding 1µl of 0.1% SDS per well.

Vortex, spin down and incubate 7 min. at Room temperature

10) cDNA library PCR and barcoding

Make PCR master-mix:

Reagents - Reagent volume:

H2O - 13.25µl

5x buffer - 5µl

dNTPs - 0.75µl

KAPA - 0.5µl

Total - 19.50µl

Dispense 19.5 µl per well to tagmentation plate (containing 3.5 µl sample after step 9)

Add primers/barcodes:

Dispense 2 µl per well (from 384-well index plates, with 3.75µM/each forward/reverse primers; **see oligos in materials**).

Total reaction volume is 25 µl (3.5 µl sample + 19.5 µl PCR mix and 2 µl primers).

Cover, vortex and spin down. Incubate in thermocycler as follows:

1 cycle:

72°C 3 min.

95°C 30 sec

12 cycles:

95°C 15 sec

67°C 30 sec

72°C 45 sec

1 cycle:

72°C 4 min

4°C hold

Following barcoding, pool 2.5µl from each well to an 1.5ml Eppendorf tube.

11) Library cleanup

- i. Add 0.9x pooled library volume of SPRI-bead solution. Incubate for 5 min. at room temperature.
- ii. Place on magnetic rack for 3 min.
- iii. Remove supernatant without disturbing magnetic beads.
- iv. Add at least 1 mL 80% EtOH (fresh). Incubate for 30 sec.
- v. Remove supernatant.
- vi. Repeat EtOH wash.
- vii. Air dry for 10 - 15 min.
- viii. Re-suspend beads thoroughly in 100 µl EB or TE buffer.
- ix. Place eppendorf on magnetic rack for 3 min.
- x. Transfer supernatant to new 1.5ml Eppendorf tube.
- xi. Repeat cleanup (from step 1-7) and elute in 30µl EB or TE buffer.
- xii. *(Optional) Place eppendorf on magnetic rack for 3 min. and transfer supernatant to new tube.*

12) Library quality control

Use Agilent HS 5000 DNA chips (or equivalent) to profile cDNA fragment lengths (See example cDNA profile in Fig. 3).

13) Sequencing

Pool libraries and sequence on Illumina sequencing platform of choice according to manufacturers instructions (For data generated in CIM-seq manuscript barcoded libraries were subjected to 75 bp single read sequencing on the Illumina NextSeq 550 instrument).

B) Feature selection and clustering (Seurat)

1. Data alignment and pre-processing

A series of pre-processing steps must be performed in order to generate a counts file:

- i. Trim reads, remove adapter sequences and align RNAseq data to reference genome using STAR (Ref. 2)
- ii. Remove duplicate reads using Picard (Ref. 3)
- iii. Generate transcript counts file using HTSeq (Ref. 4)

2. Data preparation for CIM-seq

Once a counts file has been generated the data can be analyzed. CIM-seq requires five arguments in order to run:

- i. The raw counts data with gene IDs as row names and sample IDs as column names.
- ii. The ERCC spike-in counts data with gene IDs as row names and sample IDs as column names (Well based CIM-seq only).
- iii. The dimensionality reduced representation of the data.
- iv. A classification for each of the individual singlets.
- v. A set of marker genes to be used for deconvolution

In order to generate the last three of these as well as removing low quality cells from the data set we recommend using the Seurat package in R (Ref. 5), as CIM-seq is implemented in R as well. A number of tutorials for Seurat can be found on the Satijalab website:

<https://satijalab.org/seurat/index.html>

C) Multiplet deconvolution

CIM-Seq can be downloaded from:

<https://github.com/Engelab/CIMseq>

Or installed directly in R using the devtools package and running the command:

```
devtools::install_github("Engelab/CIMseq")
```

The CIM-Seq Vignette should be followed in order to get an understanding for how the method works and the available arguments:

<https://github.com/Engelab/CIMseq/tree/master/vignettes>

Troubleshooting

Time Taken

Anticipated Results

References

1. Simone Picelli et al., "Tn5 Transposase and Tagmentation Procedures for Massively Scaled Sequencing Projects," *Genome Research* 24, no. 12 (December 1, 2014): 2033–2040.
2. Alexander Dobin et al., "STAR: Ultrafast Universal RNA-Seq Aligner," *Bioinformatics* 29, no. 1 (January 1, 2013): 15–21.
3. Aaron McKenna et al., "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data," *Genome Research* 20, no. 9 (September 1, 2010): 1297–1303.
4. Simon Anders, Paul Theodor Pyl, and Wolfgang Huber, "HTSeq—a Python Framework to Work with High-Throughput Sequencing Data," *Bioinformatics* 31, no. 2 (January 15, 2015): 166–169.
5. Yuhan Hao et al., "Integrated Analysis of Multimodal Single-Cell Data," preprint at <https://www.biorxiv.org/content/10.1101/2020.10.12.335331v1>. (October 12, 2020)

Figures

Figure 1

Gating scheme and result of FSC-W and FSC-H based sort of multiplets (top) and singlets (bottom) using HCT116 cells.

Figure 2

cDNA profile of a single cell run on HS D5000 Agilent tapestation

Figure 3

cDNA profile of a library of 784 cells (both single cells and multiplets) on HS D5000 Agilent tapestation.