

1 **i. Chapter Title**

2 Determination of the chromatin openness in bacterial genomes.

3
4 **ii. Abstract**

5 Abstract: The hyperactive Tn5 transposase in the ATAC-seq method has been widely used to
6 determine the open DNA regions and understand the overall epigenomic regulation in the
7 chromatins of eukaryotic cells. Here, we describe POP-seq (**P**rokaryotic **c**hromatin **O**penness
8 **P**rofiling **s**equencing), an adaptation of the ATAC-seq method, to interrogate changes in the
9 openness of prokaryotic nucleoids.

10 **iii. Keywords**

11 Nucleoid structure, Tn5 transposase, Nucleoid-associated proteins, Chromatin structure, H-NS,
12 HiC, Transcription factor binding sites, POP-seq

13 **1. Introduction**

14 In eukaryotes, histone oligomers organize the chromosomal DNA into nucleosomes of defined
15 sizes, which are the basic building blocks of the highly ordered chromosome. In addition to their
16 role in organizing the structure of the eukaryotic genome, histones and other structuring proteins
17 play a pivotal role in the regulation of gene expression and the functional differentiation of cells
18 [1]. Many tools have been developed, including MNase-seq [2], FAIRE-seq [3], and DNase-seq
19 [4], which showed that regions of DNA not occupied by histones (open chromatin) are highly
20 associated with active transcription. The ATAC-seq method, developed by the Greenleaf lab [5],
21 utilizes the hyperactive Tn5 transposase and represent a major improvement to the previous
22 methods. ATAC-seq is simpler, can be applied at high-throughput, and is highly sensitive, thus
23 allowing the study of the open chromatin at a single cell level [6]. Unlike in eukaryotes, well-
24 defined nucleosome structures are lacking in bacteria. The bacterial nucleoid is organized by
25 conserved nucleoid-associated proteins (NAPs), including HU and SMC [7,8], as well as poorly
26 conserved NAPs. The dynamic organization of the nucleoid directly affects how genetic
27 information is accessed, interpreted, and implemented [9–11]. The lack of methods to study the
28 open chromatin in prokaryotes has motivated us to develop POP-seq, which probes the open
29 DNA in the bacterial nucleoid that is prefixed with formaldehyde. POP-seq enables the
30 determination of the nucleoid openness and provides new insights into understanding nucleoid
31 structure, gene regulation, and phenotype.

32
33
34 **2. Materials**

- 35 2.1 Zr BashingBead Lysis Tubes (0.1 & 0.5 mm) Zymo Research.
36 2.2 Vortex Genie 2 (Scientific Industries Inc.) with horizontal shaker 24-tube adapter (Orbital
37 Shakers).
38 2.3 Nextera XT DNA library preparation kit (Illumina FC_131_1096). DNA/RNA UD
39 Indexes Set A, Tagmentation (96 Indexes, 96 Samples, Illumina 20027214).
40 2.4 Ampure XP DNA beads for PCR cleanup (Beckman Coulter A63880).
41 2.5 Nuclease-free water (Millipore-Sigma W4502-1L).

- 42 2.6 Qubit DNA high sensitivity kit (ThermoFisher)
43 2.7 Agilent High Sensitivity D1000 ScreenTape System (Agilent).
44 2.8 Hardware: 64-bit computer running Linux or Mac OS with at least 8 GB of RAM.
45 2.9 Software: Trim galore! A wrapper around Cutadapt [12] and FastQC.
46 2.10 Software: Bowtie2 [13] for short read sequencing alignment.
47 2.11 Software: SAMtools [14] for downstream analysis of alignment files.
48 2.12 Software: deepTools [15] for the calculation of sequencing depth and generation
49 of dense and continuous data tracks.
50 2.13 Software: Integrated Genome Browser (IGB) [16] for viewing BigWig files
51 generated by deepTools.
52 2.14 Software: The bioinformatic tools required for the basic analysis of the POP-seq
53 data are found in <https://github.com/maxmicrobe/POP-seq>.
54

55 3. Methods

56 3.1 Preparation of the crosslinked cell lysate

- 57
- 58 1. Grow bacterial cultures to mid exponential phase ($OD_{600} = 0.3-0.5$ for *Escherichia coli*) (**See note 1**). The volume of the culture depends on the density of the culture
59 (**see note 2**).
 - 60 2. Crosslink the cells by treatment with 1% formaldehyde for 10 minutes (**see note 3**)
 - 61 3. Quench the crosslinking reaction with 250 mM glycine (final concentration) for five
62 minutes. Centrifuge at maximum speed (i.e. 13,000 x g), remove the supernatant, and
63 place pellet on ice.
 - 64 4. Resuspend cell pellets in 300 μ l lysis buffer (75 mM NaCl, 25 mM EDTA pH 8, 20
65 nM Tris-HCl pH 8) and transfer the solution to the Zr BashingBead tube.
 - 66 5. Place the BashingBead tubes into the horizontal tube shaker pre-incubated at 4 °C and
67 run at maximum speed for 10 minutes.
 - 68 6. Centrifuge the lysate for 10 min at 14,000 rpm and 4 °C to remove the debris.
69 Incubate on ice.
 - 70 7. Take 1 μ l of the lysate and measure the DNA concentration by Qubit DNA high
71 sensitivity kit (ThermoFisher) (**see note 4**).
 - 72 8. Dilute the DNA to 0.7 ng/ μ l. Remeasure the DNA concentration to determine the
73 exact concentration using Qubit (**see note 5**). Do not exceed 0.7 ng/ μ l which is the
74 total amount of DNA input required for the library preparation kit.

75 3.2 Tagmentation and preparation of the PCR mix

- 76 1. In a 0.2 ml PCR tube, mix 10 μ l of Tagmentation DNA Buffer (TD), 5 μ l of
77 Amplicon Tagment Mix (ATM) and 5 μ l of the DNA input (0.7 ng total in 5 μ l). (**see**
78 **note 6**).
- 79 2. Incubate the mix at 55 °C for 7 minutes in a preheated PCR machine with the lid
80 heated at 100 °C. Incubate on ice immediately and add 5 μ l of Neutralize Tagment
81 buffer (NT) to stop the Tn5 reaction.
- 82 3. Add 4.5 μ l of the i7 adaptor and 4.5 μ l of the i5 adaptor.
- 83 4. In a separate tube make a master mix containing 15 μ l of Nextera PCR Master Mix
84 (NPM) and 1 μ l of 50x SYBR Green. Add 16 μ l to each sample and mix by pipetting.
85

86 3.3 Amplification of the POP-seq libraries

- 87 1. Program the qPCR machine according to the following:

- 88 a. 72 °C for 3 minutes
89 b. 95 °C for 30 seconds
90 c. 95 °C for 10 seconds
91 d. 60 °C for 30 seconds
92 e. 72 °C for 15 seconds + read tubes
93 f. 72 °C for 15 seconds
94 g. Return to (c) for 20 times
95 2. Transfer the 50 µl amplification mix into a qPCR tube and start the amplification.
96 3. Watch the progress of the amplification and stop the reaction at the end of the
97 exponential phase when the curve starts to plateau by simply pausing the program at
98 72 °C after the reading in step “e” and removing the tube from the qPCR machine
99 (**see note 7**). Incubate on ice.

100

101 3.4 Purification of the libraries

- 102 1. Mix 1.8 x Ampure XP beads with PCR product in either 0.2- or 1.5-ml tubes
103 (depending on the size of the magnet available in your lab). For example, 90 µl of
104 Ampure XP beads + 50 µl of PCR product. Incubate the mix at least 5 minutes at
105 room temperature.
106 2. Place on magnet till the solution is completely clear. While on the magnet, remove
107 supernatant and add 170 µl of freshly prepared 80% ethanol. Incubate for at least 30
108 seconds.
109 3. Remove the 80% ethanol and add another 170 µl of 80% ethanol. Incubate for at least
110 30 min and remove supernatant. Try to remove as much as possible of the ethanol,
111 remove the tubes from the magnet and incubate at room temperature and allow to dry
112 for 3-4 minutes to remove any traces of ethanol.
113 4. Finally, resuspend the dried beads with 25 µl of DNase-free water, mix well, incubate
114 at room temperature for 2 minutes. Place on magnet.
115 5. Once the solution is clear, take 23 µl of the purified library and transfer into 1.5 ml
116 tube.

117 3.5 Measuring the library concentration and checking the quality of the libraries

- 118 1. Measure the concentration of the library of each sample using the Qubit high
119 sensitivity DNA kit as described in section 2.1 step 3.
120 2. Use a TapeStation (Agilent) to determine the average size and the quality of the
121 libraries. Both the Qubit concentration and the library size are required to determine
122 the molar concentration of each library. See **Fig. 1** for an example of the POP-seq
123 library.

124 3.6 Alignment of the library sequencing reads to the reference genome.

- 125 1. Trim the libraries using specialized trimming software such as cutadapt [12].
126 2. Determine the quality of the libraries by using FastQC.
127 3. The trimmed fastq files can be used as input for fastq2wig2.pl customized script written
128 in Perl language that outputs “.wig” files with the genome coverage normalized as
129 counts per million (CPM).
130 a. Download the script from [https://github.com/maxmicrobe/POP-](https://github.com/maxmicrobe/POP-seq/blob/main/fastq2wig2.pl)
131 [seq/blob/main/fastq2wig2.pl](https://github.com/maxmicrobe/POP-seq/blob/main/fastq2wig2.pl)
132 b. To run the script, first generate a Bowtie2 index file specific for the genome
133 of interest, using this command:

- 134 i. bowtie2-build genome.fasta name_of_index.
135 c. The fasta file should contain the entire genome sequence in one entry. The
136 header should be the accession number of the genome.
137 d. If the fasta file is not available, download the functions.py tools from
138 <https://github.com/maxmicrobe/POP-seq/blob/main/functions.py> and run the
139 “write_full_seq()” function which takes two arguments: The genbank
140 file path and the output path.
141 e. The output of the fastq2wig2.pl script, a “.wig” file, can be readily viewed via
142 Integrated Genome Browser (IGB) [16]. See **Fig. 2** for an example of the
143 output on IGB.
144 f. A “.bed” file is typically required to view the starts and the ends of genes or
145 any other features. To generate the bed file, download the gbk2bed.pl script
146 from <https://github.com/maxmicrobe/POP-seq/blob/main/gbk2bed.pl>. This
147 script only takes the genbank file as input and prints the output on the screen.
148 To save the output into a file add “> file_name” at the end of the
149 command.

150 Notes

- 151 1. Harvesting cells at exponential phase is likely to give better results because the cell
152 growth is more synchronized. However, stationary phase POP-seq has been performed in
153 *E. coli* and was equally successful.
154 2. In rich media, culture volumes can be as low as 200 μ l. 5 μ l of 0.16 ng/ μ l input is
155 required for the Nextera XT sequencing library preparation step.
156 3. The final concentration of formaldehyde can vary between 1-3%. The fixation time can
157 also vary between 10-30 minutes depending on the bacterial strain. A good starting point
158 is 1% formaldehyde for 10 minutes.
159 4. We do not recommend using Nanodrop due to the lack of sensitivity. The exact DNA
160 concentration is very important for reproducible results and successful library
161 preparations.
162 5. We recommend using 5 μ l of input into the Qubit reaction as follows:
163 a. 5 μ l diluted DNA
164 b. 1 μ l Qubit dsDNA HS reagent
165 c. 194 μ l Qubit dsDNA HS buffer
166 6. It is well documented that Tn5 has increased accessibility towards high A/T regions.
167 Therefore, including a pure DNA negative control is highly recommended to normalize
168 the accessibility bias.
169 7. If you miss the 72 $^{\circ}$ C step, an extra cycle of amplification will not harm, but avoid over
170 amplification as this will cause uneven amplification of the amplicons and could affect
171 the overall outcome. Usually, the PCR ends by cycle 15-16, however, amplification to
172 cycle 20 is still acceptable if the plateau is not reached.

173 References

- 174 [1] A.J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications, Cell
175 Res. (2011). <https://doi.org/10.1038/cr.2011.22>.

- 176 [2] A.N. Schep, J.D. Buenrostro, S.K. Denny, K. Schwartz, G. Sherlock, W.J. Greenleaf,
177 Structured nucleosome fingerprints enable high-resolution mapping of chromatin
178 architecture within regulatory regions, *Genome Res.* 25 (2015) 1757–1770.
179 <https://doi.org/10.1101/gr.192294.115>.
- 180 [3] P.G. Giresi, J. Kim, R.M. McDaniel, V.R. Iyer, J.D. Lieb, FAIRE (Formaldehyde-
181 Assisted Isolation of Regulatory Elements) isolates active regulatory elements from
182 human chromatin, *Genome Res.* (2007). <https://doi.org/10.1101/gr.5533506>.
- 183 [4] A.P. Boyle, S. Davis, H.P. Shulha, P. Meltzer, E.H. Margulies, Z. Weng, T.S. Furey, G.E.
184 Crawford, High-Resolution Mapping and Characterization of Open Chromatin across the
185 Genome, *Cell.* (2008). <https://doi.org/10.1016/j.cell.2007.12.014>.
- 186 [5] J.D. Buenrostro, P.G. Giresi, L.C. Zaba, H.Y. Chang, W.J. Greenleaf, Transposition of
187 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
188 binding proteins and nucleosome position, *Nat. Methods.* (2013).
189 <https://doi.org/10.1038/nmeth.2688>.
- 190 [6] J.D. Buenrostro, B. Wu, U.M. Litzenburger, D. Ruff, M.L. Gonzales, M.P. Snyder, H.Y.
191 Chang, W.J. Greenleaf, Single-cell chromatin accessibility reveals principles of regulatory
192 variation., *Nature.* 523 (2015) 486–90. <https://doi.org/10.1038/nature14590>.
- 193 [7] P. Stojkova, P. Spidlova, J. Stulik, Nucleoid-associated protein Hu: A lilliputian in gene
194 regulation of bacterial virulence, *Front. Cell. Infect. Microbiol.* (2019).
195 <https://doi.org/10.3389/fcimb.2019.00159>.
- 196 [8] S. Nolivos, D. Sherratt, The bacterial chromosome: Architecture and action of bacterial
197 SMC and SMC-like complexes, *FEMS Microbiol. Rev.* (2014).
198 <https://doi.org/10.1111/1574-6976.12045>.
- 199 [9] R.T. Dame, F.-Z.M. Rashid, D.C. Grainger, Chromosome organization in bacteria:
200 mechanistic insights into genome structure and function, *Nat. Rev. Genet.* (2019).
201 <https://doi.org/10.1038/s41576-019-0185-4>.
- 202 [10] M. McArthur, M. Bibb, In vivo DNase I sensitivity of the *Streptomyces coelicolor*
203 chromosome correlates with gene expression: Implications for bacterial chromosome
204 structure, *Nucleic Acids Res.* (2006). <https://doi.org/10.1093/nar/gkl649>.
- 205 [11] N.T. Tran, M.T. Laub, T.B.K. Le, SMC Progressively Aligns Chromosomal Arms in
206 *Caulobacter crescentus* but Is Antagonized by Convergent Transcription, *Cell Rep.*
207 (2017). <https://doi.org/10.1016/j.celrep.2017.08.026>.
- 208 [12] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads,
209 *EMBnet.Journal.* (2011). <https://doi.org/10.14806/ej.17.1.200>.
- 210 [13] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2., *Nat. Methods.*
211 (2012). <https://doi.org/10.1038/nmeth.1923>.
- 212 [14] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis,
213 R. Durbin, The Sequence Alignment/Map format and SAMtools, *Bioinformatics.* 25
214 (2009) 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- 215 [15] F. Ramírez, D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F.
216 Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data
217 analysis, *Nucleic Acids Res.* (2016). <https://doi.org/10.1093/nar/gkw257>.

218 [16] N.H. Freese, D.C. Norris, A.E. Loraine, Integrated genome browser: Visual analytics
219 platform for genomics, *Bioinformatics*. 32 (2016) 2089–2095.
220 <https://doi.org/10.1093/bioinformatics/btw069>.

221

222 Figure legends

223

224 Figure 1. A snapshot from the TapeStation analysis software showing the lower and upper sizes
225 of a typical pop-seq library. The average library size is automatically calculated by the software.

226

227 Figure 2. A snapshot from the IGB software showing a section of the *E. coli* BW25113 genome.
228 The three tracks with black signals represent biological replicates of POP-seq experiments in the
229 wild type. The three tracks with red signals are three biological replicates of congenic *ihfB*
230 deletion mutants. The experiments are highly reproducible for both strains. The genes colored in
231 green have significantly higher Tn5 accession in the wild type strain compared to *ihfB* (p-value
232 <0.05). The locus_tags are shown for four genes.

233