Pore-C Simultaneously Captures Genome-wide Multi-way Chromatin Interaction and Associated DNA Methylation Status in Arabidopsis

1 Zhuowen Li^{1,2,3,#}, Yanping Long^{1,2,3,#}, Yiming Yu^{1,2,3}, Fei Zhang^{1,2,3}, Hong Zhang^{1,2,3},
 2 Zhijian Liu^{1,2,3}, Jinbu Jia^{1,2,3}, Weipeng Mo^{1,2,3}, Simon Zhongyuan Tian¹, Meizhen
 3 Zheng¹, and Jixian Zhai^{1,2,3,*}

4

5 Affiliations:

6 ¹ Department of Biology, School of Life Sciences, Southern University of Science and
 7 Technology, Shenzhen 518055, Guangdong, China;

8 ² Institute of Plant and Food Science, Southern University of Science and Technology,
 9 Shenzhen 518055, Guangdong, China;

10 ³ Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher
 11 Education Institutes, Southern University of Science and Technology, Shenzhen
 12 518055, Guangdong, China.

13

14 [#] Co-first authors

15 ^{*} Correspondence: zhaijx@sustech.edu.cn (J.Z.)

16

17 **Keywords:** Pore-C, multi-way interaction, DNA methylation, long-read sequencing

18

19

20 **Abstract**

21 In the past decade, genome-wide characterization of the three-dimensional
22 chromatin structure in plants using high-throughput methods has greatly advanced
23 our knowledge in plant genome architecture (Liu and Weigel, 2015; Ouyang et al.,
24 2020). However, due to the limitation of Illumina short-read sequencing, the
25 genome-wide contact map obtained by Hi-C/ChIA-PET is pairwise, and the
26 multi-way interaction can only be inferred from the two-way data. To directly
27 capture multi-way interaction in Arabidopsis, we applied a long-read-based method
28 called Pore-C that directly sequences the DNA fragments joined by
29 proximity-based ligation (Ulahannan et al., 2019).

30

31 Compared with Hi-C, Pore-C replaced the Illumina short-read sequencing with
32 Nanopore long-read sequencing, which drastically simplified the library construction
33 procedure by removing the biotin labeling, fragmentation and several
34 affinity-purification and DNA purification steps (Figure 1A). *In-situ* ligation makes
35 Pore-C reads to retain the interaction information within the individual nucleus. We
36 generated two Pore-C libraries using 12-day-old Arabidopsis seedlings of wildtype
37 Col-0 and analyzed the data according to the Pore-C pipeline (Ulahannan et al., 2019).
38 We obtained 6.5 million (11.7 Gb) and 3.5 million (7.9 Gb) high-quality reads (Q7
39 filter) with N50 up to 2654 bp and 3006 bp, respectively (Supplemental table 1). More
40 than 96% of the reads can be mapped to the Arabidopsis TAIR10 reference genome in
41 both libraries. From about 10 million Pore-C reads in total, we obtained 50.7 million
42 valid pairs of interaction, a similar amount but with fewer reads compared to previous
43 Hi-C data (59.1 million pairwise contacts from 239.5 million Illumina reads) (Wang
44 et al., 2015), demonstrating the efficiency of long-read sequencing in harvesting
45 chromosomal interaction. From the 2D-contact matrixes, the two Pore-C libraries and

46 the Hi-C data showed high consistency in their contact feature, suggesting high
47 reproducibility and accuracy of Pore-C (Figure 1B, C).

48

49 The Pore-C pipeline defined the number of genome-matched fragments on a Pore-C
50 read as the “Contact Order” (Ulahannan et al., 2019). Our results show that 44% reads
51 contain multi-way interactions (Contact Order ≥ 3), similar to Pore-C in human cells
52 (47.56%) (Ulahannan et al., 2019). Whereas in the Hi-C data (Wang et al., 2015), the
53 Contact Orders are mostly at two, consistent with its design (Figure 1D). Pore-C results
54 in Arabidopsis and animals show a higher efficiency in capturing multi-way interaction
55 compared to SPRITE (12%) (Quinodoz et al., 2018), demonstrating an advantage of the
56 long-read approach over the labeling approach in capturing multi-way interaction. The
57 contact frequency in relation to the genomic distance for our Pore-C data showed that
58 reads with higher contact order detected interactions across longer genomic regions
59 (Figure 1E), which can also give help in *de novo* genome assembly.

60

61 To validate the effectiveness of Pore-C in multi-way interaction in Arabidopsis, we
62 examined the interactions among KNOT ENGAGED ELEMENT (KEE) (Grob et al.,
63 2014) with our Pore-C data. In Arabidopsis, the interaction of ten KEEs (named
64 KEE1 to KEE10) can be visualized as discrete dots on the pairwise contact matrix of
65 Hi-C and their pairwise interactions were confirmed by light microscopy-based FISH
66 assay (Grob et al., 2014). However, it remains unclear if multiple KEEs can
67 simultaneously interact together in the same nucleus. In our Pore-C data, 1145 reads
68 simultaneously detect contact among three KEEs, higher than that of the 100
69 randomly selected control regions (mean 279.29 reads, *P*-value = 0.04), and 75 reads
70 capture the contact among four KEEs and more (Figure 1F). Interestingly, nearly half
71 of these high-order interactions involve both KEE3 and KEE4 (561 reads), suggesting
72 that these two KEEs serve as key hubs in the KEE interacting network (Figure 1F).
73 Direct interaction among four KEEs was also detected, including 9 reads for KEE3,

74 KEE4, KEE7 and KEE8 (referred to as KEE(3,4,7,8), [Figure 1G](#)), 4 reads for
75 KEE(1,3,4,6), and 11 reads for KEE(3,4,5,6). These results demonstrate the existence
76 of multi-way contacts of KEEs.

77

78 The PCR-free strategy of Pore-C enables to directly detect DNA methylation states
79 and higher-order chromatin interaction by long-read sequencing and thus helps to
80 reveal their coordination on the same read, without the requirement of bisulfite
81 conversion in the Methyl-HiC method (Li et al., 2019). The CG, CHG and CHH
82 methylation level we called from Pore-C reads are highly consistent with WGBS
83 (whole-genome bisulfite sequencing), the gold standard for DNA methylation ([Figure](#)
84 [1H](#)) (Ni et al., 2021; Stroud et al., 2012). We used nanopolish (Simpson et al., 2017)
85 to detect 5mCG related to each Pore-C read, and found a higher correlation of CG
86 methylation level among interacting fragments in Arabidopsis ([Figure 1I](#)), which is
87 consistent with the previous report in mouse ESCs (Li et al., 2019).

88

89 In summary, Pore-C efficiently captures the genome-wide multi-way chromatin
90 interaction landscape at a single-molecular level, and can reveal the epigenetic
91 modification and chromatin organization of the same group of interacting DNA
92 fragments, which may help to explore the heterogeneity of the chromatin interaction
93 and DNA methylation in nuclei of different cell types. With this method, we validated
94 the multi-way contacts among KEEs within nucleus, and found the CG methylated
95 regions on the genome tends to contact. Taken together, our results demonstrate that
96 Pore-C is a simple, accurate and effective method for exploring multi-way chromatin
97 interaction in plants.

98

99 **Acknowledgments**

100 Group of J.Z. is supported by the National Key R&D Program of China Grant

(2019YFA0903903), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2016ZT06S172), the Shenzhen Sci-Tech Fund (KYTDPT20181011104005) and a Stable Support Plan Program of Shenzhen Natural Science Fund Grant (20200925153345004), and Key Laboratory of Molecular Design for Plant Cell Factory of Guangdong Higher Education Institutes (2019KSYS006).

106

107 **Author Contributions**

108 Y.L. and Z.L. developed the method and performed the experiments. Z.L., Y.Y., F.Z.
109 and Y.L. analyzed the data. H.Z. set up the website, J.Z. conceived and oversaw the
110 study. Z.L., Y.L. and J.Z. wrote the manuscript. All authors revised the manuscript.

111

112 **Materials and Methods**

113 Please refer to Supplementary material.

114

115 **Data availability**

116 The Pore-C data generated can be found in the Genome Sequence Archive in National
117 Genomics Data Center under accession number CRA005105 (reviewer link:
118 <https://ngdc.cncb.ac.cn/gsa/s/INkjuqHu>).

119

120 **Competing Interests**

121 The authors declare no competing interests.

122

References

- Grob, S., Schmid, M.W. and Grossniklaus, U. (2014) Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Molecular cell* **55**, 678-693.
- Li, G., Liu, Y., Zhang, Y., Kubo, N., Yu, M., Fang, R., et al. (2019) Joint profiling of DNA methylation and chromatin architecture in single cells. *Nature methods* **16**, 991-993.
- Liu, C. and Weigel, D. (2015) Chromatin in 3D: progress and prospects for plants. *Genome biology* **16**, 170.
- Ni, P., Huang, N., Nie, F., Zhang, J., Zhang, Z., Wu, B., et al. (2021) Genome-wide detection of cytosine methylations in plant from Nanopore sequencing data using deep learning. *bioRxiv*, 2021.2002.2007.430077.
- Ouyang, W., Xiong, D., Li, G. and Li, X. (2020) Unraveling the 3D genome architecture in plants: present and future. *Molecular Plant* **13**, 1676-1693.
- Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., et al. (2018) Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744-757.e724.
- Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J. and Timp, W. (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nature methods* **14**, 407-410.
- Stroud, H., Hale, C.J., Feng, S., Caro, E., Jacob, Y., Michaels, S.D., et al. (2012) DNA methyltransferases are required to induce heterochromatic re-replication in *Arabidopsis*. *PLoS genetics* **8**, e1002808.
- Ulahannan, N., Pendleton, M., Deshpande, A., Schwenk, S., Behr, J.M., Dai, X., et al. (2019) Nanopore sequencing of DNA concatemers reveals higher-order features of chromatin structure. *bioRxiv*, 833590.
- Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., et al. (2015) Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome research* **25**, 246-256.

Figure1. Pore-C captures multi-way contacts and DNA methylation in Arabidopsis.

(A) Schematic of the Pore-C protocol.

(B,C) The consistency between two Pore-C libraries (B), between Hi-C and Pore-C (C), respectively. Left: contact maps across all chromosomes. Middle: contact maps on chromosome 3. Right: contact maps within chr3: 0-5 Mb. The interactions between KEEs are marked by black arrows.

(D) Comparison of contact order distribution between previous Hi-C data and Pore-C data in Arabidopsis.

(E) Contact frequency between different genomic distances of various order groups of Pore-C.

(F) Combination information of reads related to at least three different KEEs. Each row of the heatmap represents one read. The color intensities in the heatmap indicate the counts of interaction capture within one KEE on one read. KEE(*,*,*) means the interaction between at least three individual KEEs, “*” indicates at least one KEE, and “n” shows the read counts.

(G) Alignment locations of individual reads related to KEE3, KEE 4, KEE7 and KEE8. The red reads (n=9) show the direct interactions between all the four KEEs.

(H) The consistency of CG, CHG, CHH DNA methylation between Pore-C library and WGBS data.

(I) Pearson correlation coefficient (PCC) of CG methylation level between Pore-C contacts with genomic distance larger than 25 kb, 100 kb and 1 Mb, respectively.

A

Pore-C (input - 0.25 gram of fresh tissue)

Crosslinking isolated nuclei, in-nuclei ligation

B

D

C

E

F

G

H

I

Figure1. Pore-C captures multi-way contacts and DNA methylation in Arabidopsis.

(A) Schematic of the Pore-C protocol.

(B,C) The consistency between two Pore-C libraries (B), between Hi-C and Pore-C (C), respectively. Left: contact maps across all chromosomes. Middle: contact maps on chromosome 3. Right: contact maps within chr3: 0-5 Mb. The interactions between KEEs are marked by black arrows.

(D) Comparison of contact order distribution between previous Hi-C data and Pore-C data in Arabidopsis.

(E) Contact frequency between different genomic distances of various order groups of Pore-C.

(F) Combination information of reads related to at least three different KEEs. Each row of the heatmap represents one read. The color intensities in the heatmap indicate the counts of interaction capture within one KEE on one read. KEE(*,*,*) means the interaction between at least three individual KEEs, “*” indicates at least one KEE, and “n” shows the read counts.

(G) Alignment locations of individual reads related to KEE3, KEE 4, KEE7 and KEE8. The red reads (n=9) show the direct interactions between all the four KEEs.

(H) The consistency of CG, CHG, CHH DNA methylation between Pore-C library and WGBS data.

(I) Pearson correlation coefficient (PCC) of CG methylation level between Pore-C contacts with genomic distance larger than 25 kb, 100 kb and 1 Mb, respectively.