

Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq

Cell Research (2016) 26:1345-1348. doi:10.1038/cr.2016.137; published online 25 November 2016

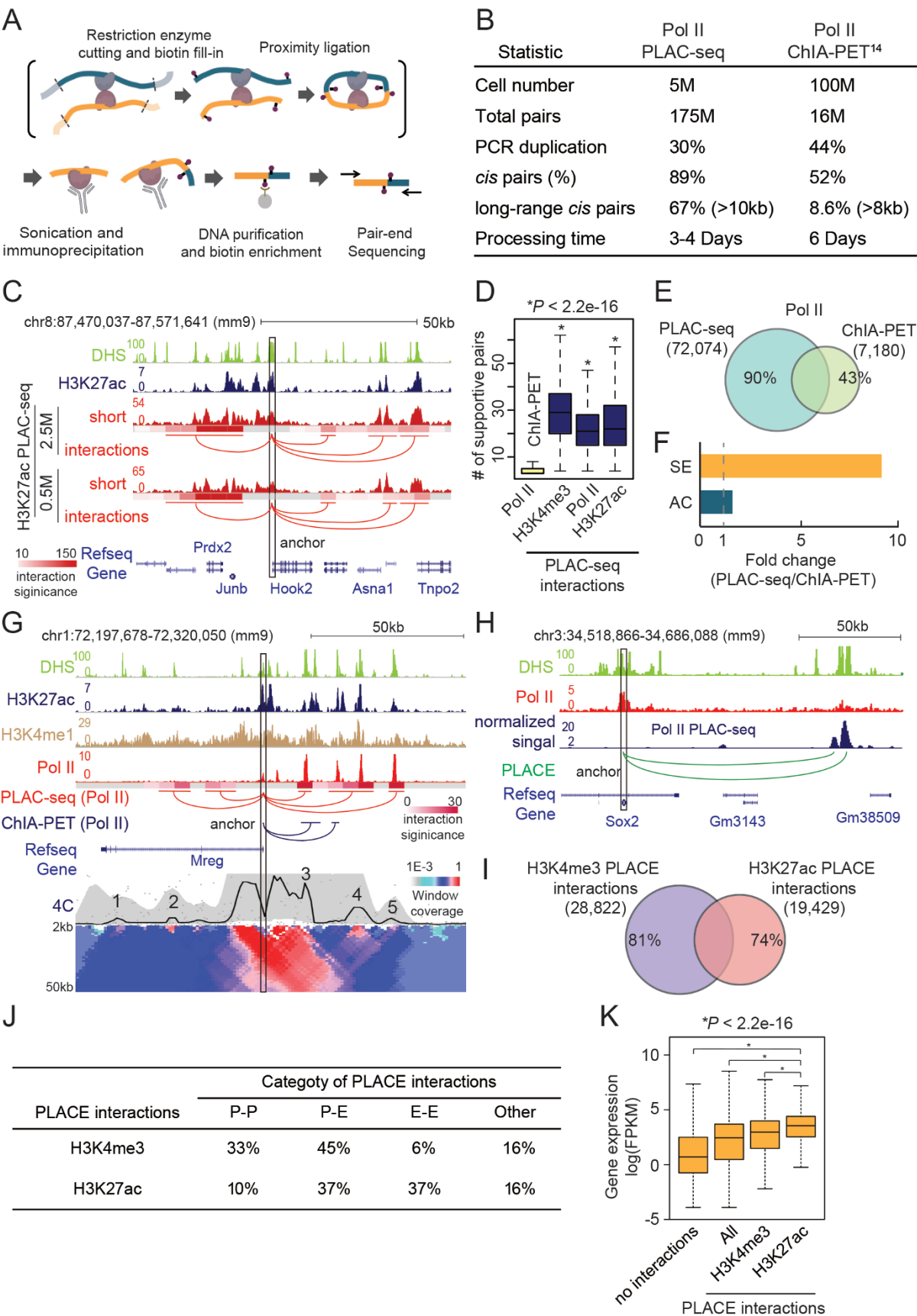
Dear Editor,

Formation of long-range chromatin loops is a crucial step in transcriptional activation of target genes by distal enhancers [1-2]. Mapping such structural features can help define target genes for enhancers and annotate non-coding sequence variants linked to human diseases [1-3]. Study of the higher-order chromatin organization has been facilitated by the development of chromosome conformation capture (3C)-based technologies [4, 5]. Among the commonly used high-throughput 3C approaches are Hi-C [6] and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [7]. Global analysis of long-range chromatin interactions using Hi-C has been achieved at kilobase resolution but requires billions of sequencing reads [8]. High-resolution analysis of long-range chromatin interactions at selected genomic regions can be attained cost-effectively through either ChIA-PET [7, 9] or targeted capture and sequencing of Hi-C libraries [10]. ChIA-PET has been used to identify long-range interactions at promoters and enhancers at high resolution in various cell types and species [11]. However, this procedure requires hundreds of million cells as starting materials, likely because chromatin immunoprecipitation and proximity ligation are performed after chromatin shearing, which potentially leads to great disruption of protein/DNA complexes. To reduce the amount of input materials and improve the sensitivity and robustness of the assay, we developed Proximity Ligation-Assisted ChIP-seq (PLAC-seq), in which proximity ligation is conducted in nuclei prior to chromatin shearing and immunoprecipitation (Figure 1A, Supplementary information, Figure S1A and Data S1). We demonstrated that by switching the order of proximity ligation and chromatin shearing steps, PLAC-seq greatly improves the efficiency and accuracy over ChIA-PET [7, 9] in detection of long-range chromatin interactions in mammalian cells.

We performed PLAC-seq in mouse embryonic stem (ES) cells using antibodies against RNA Polymerase II (Pol II), H3K4me3 and H3K27ac to determine long-range chromatin interactions at promoters and enhancers

in the genome (Supplementary information, Table S1). As shown in Figure 1B, PLAC-seq yielded libraries with higher number of unique read pairs compared with ChIA-PET. As expected, the sequencing reads were strongly enriched at the factor-binding sites detected by ChIP-seq analysis in the mouse ES cells [12] (Supplementary information, Figure S1B-S1D and S1F-S1H). Additionally, the PLAC-seq experiments generated long-range chromatin contacts that were highly reproducible between biological replicates (Pearson correlation > 0.90; Supplementary information, Figure S1E). To identify long-range chromatin interactions, we used 'FitHiC' [13] to analyze the combined datasets from two biological replicates (Supplementary information, Data S1). A total of 72 074, 273 145, and 155 545 chromatin loops (FDR < 0.01) were identified from the Pol II, H3K4me3, and H3K27ac PLAC-seq experiments, respectively. We found that PLAC-seq could be performed with much fewer cells than ChIA-PET. Even with 0.5 million (M) cells, a majority of strong long-range interactions could be detected (Figure 1C and Supplementary information, Figure S1I).

Several lines of evidence support the superior performance of PLAC-seq over ChIA-PET. First, PLAC-seq was nearly 100 times more cost-effective than ChIA-PET in generating long-range intra-chromosomal read pairs, which are typically used to infer chromatin loops. Using 20-fold fewer cells (5 M vs 100 M), Pol II PLAC-seq produced 10 times more reads (175 M vs 16 M) with lower PCR duplication rate (30% vs 44%) than a previously published Pol II ChIA-PET experiment [14]. In addition, PLAC-seq generated more long-range intra-chromosomal pairs (67% vs 9%) and fewer inter-chromosomal pairs (11% vs 48%) (Figure 1B). Second, PLAC-seq uncovered chromatin loops in the mouse ES cells with much higher sensitivity and specificity than ChIA-PET. Additionally, PLAC-seq chromatin interactions were typically supported by 24 unique read pairs (medium) compared to 3 PETs supporting ChIA-PET interactions [14] (Figure 1D). Pol II PLAC-seq analysis identified 57% of Pol II ChIA-PET interactions (FDR < 0.05 and PET count \geq 3, 10 kb to 3Mb) and a lot of additional



interactions (Figure 1E). PLAC-seq covered more regulatory elements, such as promoters and distal DNase I hypersensitive sites (DHSs), than ChIA-PET (Supplementary information, Figure S1J). As a reference, we performed *in situ* Hi-C with the mouse ES cell line and collected nearly 1.2 billion paired-end sequencing reads, from which we identified 68 781 long-range chromatin interactions (FDR < 0.01) using FitHiC [13]. Compared with chromatin interactions identified by *in situ* Hi-C, PLAC-seq is 8 times more sensitive than ChIA-PET and also more accurate (Figure 1F). Third, we performed 4C-seq analysis of four randomly selected genomic regions (Supplementary information, Table S1). Although both ChIA-PET and PLAC-seq identified many common chromatin interactions (Figure 1G and Supplementary information, Figure S2B–S2C), PLAC-seq uncovered seven additional strong interactions (marked 2, 4 and 5 in Figure 1G, and 1–4 in Supplementary information, Figure S2A–S2C) detected by 4C-seq. Taken together, the results above support the superior sensitivity and specificity of PLAC-seq over ChIA-PET.

We also developed a new computational algorithm to identify chromatin interactions at high resolution from PLAC-seq data. We used the binomial test (Supplementary information, Data S1) to determine the enrichment of read pairs for an interaction due to chromatin immunoprecipitation using *in situ* Hi-C analysis result as an estimation of background interaction frequency (Figure 1H). We termed this type of interactions as ‘PLACE’ (PLAC-Enriched) interactions. A total of 28 822 and

19 429 significant H3K4me3 and H3K27ac PLACE interactions (FDR < 0.05) in the mouse ES cells were identified, respectively. These corresponded to different sets of chromatin interactions, with 26% of H3K27ac PLACE interactions overlapping with 19% of H3K4me3 PLACE interactions (Figure 1I). A majority of H3K27ac PLACE interactions were enhancer-associated (74%) while H3K4me3 PLACE interactions were generally promoter-associated (78%) (Figure 1J). Genes involved in H3K27ac PLACE interactions had significantly higher expression levels than genes associated with H3K4me3 PLACE interactions ($P < 2.2 \times 10^{-16}$, Figure 1K), suggesting that H3K27ac PLAC-seq could be used to discover chromatin interactions at active enhancers and H3K4me3 PLAC-seq at active or poised promoters.

In summary, we developed a fast, sensitive and cost-effective method to map long-range chromatin interactions in mammalian cells. Using PLAC-seq, we obtained high-resolution maps of chromatin interactions at enhancers and promoters in the mouse ES cells. The ease of experimental procedure and small amount of input materials required will allow the mapping of long-range chromatin interactions in a broad set of species, cell types, and experimental settings. A similar method called HiChIP was recently reported by Mumbach *et al.* [15] when our manuscript was under review.

Accession Code

Raw and processed data have been deposited to NCBI Gene Expression Omnibus with the accession number GSE86150.

Figure 1 PLAC-seq reveals chromatin interactions in mammalian cells at high sensitivity and accuracy. **(A)** Overview of the PLAC-seq workflow. Formaldehyde-fixed cells were permeabilized and digested with a 4-bp cutter *MboI*, followed by biotin-tagged nucleotide fill-in and *in situ* proximity ligation. Nuclei were then lysed and the chromatin was sheared by sonication. The soluble chromatin fraction was then subjected to immunoprecipitation using specific antibodies against a transcription factor or a histone modification. Finally, after reverse-crosslinking the biotin-labeled DNA corresponding to ligation junctions was enriched followed by library preparation and paired-end DNA sequencing. **(B)** Comparison of the sequence outputs between PLAC-seq and ChIA-PET. **(C)** Comparison of short-range signals (short) and long-range chromatin interactions (interactions) identified by H3K27ac PLAC-seq using 2.5 M and 0.5 M cells in the indicated genomic region. Only the interactions with one end overlapping with a selected anchor point (chr8: 87 510 000–87 515 000, black rectangle) were shown. PLAC-seq interactions are marked by red arcs and interaction significance is denoted by $-\log$ (FDR). **(D)** Box plots of number of the unique read pairs supporting interactions identified by ChIA-PET and PLAC-seq. **(E)** Venn-diagram comparing the chromatin loops identified in Pol II PLAC-seq and Pol II ChIA-PET experiments. **(F)** Comparison of sensitivity (SE) and accuracy (AC) between PLAC-seq and ChIA-PET interactions using the loops detected by *in situ* Hi-C as a reference (SE = number of *in situ* HiC interactions overlapping with PLAC-seq or ChIA-PET interactions / total number of *in situ* HiC interactions; AC = number of PLAC-seq or ChIA-PET interactions overlapping with *in situ* HiC interactions / total number of PLAC-seq or ChIA-PET interactions). **(G)** Comparison of chromatin interactions identified by PLAC-seq, ChIA-PET and 4C-seq at the *Mreg* promoter (the anchor point is marked by a black rectangle, chr1: 72 255 000–72 260 000). PLAC-seq and ChIA-PET interactions were demonstrated by red and blue arcs, respectively; significance of interactions in PLAC-seq is denoted by $-\log$ (FDR). **(H)** Normalized Pol II PLAC-seq signals and PLACE (Supplementary information, Data S1) analysis revealed chromatin interactions between Sox2 and its super enhancer at nearly single-element resolution (anchor region, chr3: 34 546 927–34 553 382). **(I)** Overlap between H3K27ac and H3K4me3 PLACE interactions. **(J)** Distribution of promoter-promoter (P-P), promoter-enhancer (P-E), enhancer-enhancer (E-E) and other interactions for H3K27ac and H3K4me3 PLACE interactions. **(K)** Boxplot of expression of different groups of genes. H3K27ac PLACE interactions are associated with genes with significantly higher expression than other genes ($P < 2.2 \times 10^{-16}$). 2.5 M cells were used for H3K27ac PLAC-seq experiments in **D**, **J** and **K**.

Acknowledgments

The work is supported by funding from the Ludwig Institute for Cancer Research and NIH (1U54DK107977-01, 2P50 GM085764 and U54 HG006997).

Rongxin Fang^{1,2,*}, Miao Yu^{1,*}, Guoqiang Li¹,
Sora Chee¹, Tristin Liu¹, Anthony D Schmitt^{1,3},
Bing Ren^{1,4}

¹Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA; ²Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; ³Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; ⁴Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, and Moores Cancer Center, University of California at San Diego, La Jolla, CA 92093, USA

*These two authors contributed equally to this work.

Correspondence: Bing Ren
E-mail: biren@ucsd.edu

References

- 1 Gorkin DU, Leung D, Ren B. *Cell Stem Cell* 2014; **14**:762-775.
- 2 de Laat W, Duboule D. *Nature* 2013; **502**:499-506.
- 3 Sexton T, Cavalli G. *Cell* 2015; **160**:1049-1059.
- 4 Dekker J, Marti-Renom MA, Mirny LA. *Nat Rev Genet* 2013; **14**:390-403.
- 5 Denker A, de Laat W. *Genes & Development* 2016; **30**:1357-1382.
- 6 Lieberman-Aiden E, van Berkum NL, Williams L, et al. *Science* 2009; **326**:289-293.
- 7 Fullwood MJ, Liu MH, Pan YF, et al. *Nature* 2009; **462**:58-64.
- 8 Rao SSP, Huntley MH, Durand NC, et al. *Cell* 2014; **159**:1665-1680.
- 9 Tang Z, Luo OJ, Li X, et al. *Cell* 2015; **163**:1611-1627.
- 10 Mifsud B, Tavares-Cadete F, Young AN, et al. *Nat Genet* 2015; **47**:598-606.
- 11 Li G, Cai L, Chang H, et al. *BMC Genomics* 2014; **15 Suppl 12**:S11.
- 12 Shen Y, Yue F, McCleary DF, et al. *Nature* 2012; **488**:116-120.
- 13 Ay F, Bailey TL, Noble WS. *Genome Res* 2014; **24**:999-1011.
- 14 Zhang Y, Wong C-H, Birnbaum RY, et al. *Nature* 2013; **504**:306-310.
- 15 Mumbach MR, Rubin AJ, Flynn RA, et al. *Nat Methods* 2016; **13**:919-922.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)