**METHODOLOGY ARTICLE**　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

# SALP, a new single-stranded DNA library preparation method especially useful for the high-throughput characterization of chromatin openness states

Jian Wu, Wei Dai, Lin Wu and Jinke Wang[*] (iD)

## Abstract

**Background:** Next-generation sequencing (NGS) is fundamental to the current biological and biomedical research. Construction of sequencing library is a key step of NGS. Therefore, various library construction methods have been explored. However, the current methods are still limited by some shortcomings.

**Results:** This study developed a new NGS library construction method, Single strand Adaptor Library Preparation (SALP), by using a novel single strand adaptor (SSA). SSA is a double-stranded oligonucleotide with a 3′ overhang of 3 random nucleotides, which can be efficiently ligated to the 3′ end of single strand DNA by T4 DNA ligase. SALP can be started with any denatured DNA fragments such as those sheared by Tn5 tagmentation, enzyme digestion and sonication. When started with Tn5-tagmented chromatin, SALP can overcome a key limitation of ATAC-seq and become a high-throughput NGS library construction method, SALP-seq, which can be used to comparatively characterize the chromatin openness state of multiple cells unbiasly. In this way, this study successfully characterized the comparative chromatin openness states of four different cell lines, including GM12878, HepG2, HeLa and 293T, with SALP-seq. Similarly, this study also successfully characterized the chromatin openness states of HepG2 cells with SALP-seq by using $10^5$ to 500 cells.

**Conclusions:** This study developed a new NGS library construction method, SALP, by using a novel kind of single strand adaptor (SSA), which should has wide applications in the future due to its unique performance.

**Keywords:** Single strand adaptor, Tn5, Library construction, Chromatin state, SALP-seq

## Background

Since next-generation sequencing (NGS) appeared on the marketplace in 2005 [1], this technology has changed the way we think about scientific approaches in basic, applied and clinical research [2]. With the innovative methodological and computational developments, NGS platforms have facilitated an explosion on biological knowledge over past years [3]. As the broadest application of NGS, the re-sequencing of human genomes dramatically enhance our understanding how genetic differences affect health and disease [2]. Compared with the Sanger sequencing, the key procedural difference in NGS methods is the construction of sequencing library [3]. As the productivity of NGS platforms continues to grow with hardware and software optimization, the sequencing library construction becomes the reamained restriction of the technology [4]. Therefore, various library construction methods have been developed, including several methods preparing NGS library with single strand DNA (ssDNA) [5–9]. However, some intrinsic shortcomings are still limiting the widely applications of these methods.

The current standard method of NGS library construction starts with double-stranded DNA (dsDNA). The main steps include DNA fragmentation, end-polishing, adaptor ligation, size selection, and PCR amplification [4]. This cumbursome procedure has reccently been greatly simplified by using a Tn5-based DNA tagmentation approach.

* Correspondence: wangjinke@seu.edu.cn
State Key Laboratory of Bioelectronics, Southeast University, Sipailou 2, Nanjing 210096, China

Wu *et al. BMC Genomics* (2018) 19:143

Page 2 of 12

In this approach, two mosaic end (ME) adaptors harboring the annealing sites of two primers are firstly complexed with a hyperactive derivative Tn5 transposase to form transposome, which then tagmented DNA into tagments with adaptors at their 5′ ends [4]. The DNA tagments were then amplified with PCR by using specific primers, which produces the DNA library compatible with massively parallel sequencing. However, when this method is performed to create dual-tagged libraries, only part of tagments can be correctly amplified and successfully sequenced by using two different library constructing primers, even suppression PCR is performed to increase the percentage of fragments with correct structure [10].

The processes of development and differentiation of mammals are regulated by the constant change of interaction between DNA-binding proteins and chromatin [11]. The chromatin provides the controlled access of transcription factors (TFs) to their DNA binding sites in a highly cell-type-specific manner [11–13]. Therefore, increasing researches investigated the chromatin openness state [14–17]. For this reason, as a rapid and sensitive method to capture the open chromatin sites, the transposase-accessible chromatin using sequencing (ATAC-seq) was developed and has been widely used to research the chromatin openness state [18–21]. Mapping the open chromatin regions is beneficial for finding the key cis elements and trans factors regulating a particular biological process. For example, the transcription factor AP1 was identified as a transcriptional regulator in oesophageal adenocarcinoma (OAC) by profiling open chromatins in the OAC-derived cell lines and patient-derived OAC material with ATAC-seq [22]. However, ATAC-seq was limited by the intrinsic shortcoming of its tagmentation-based library construction, that is, only part of the tagmentation fragments can be sequenced during the following sequencing process.

Here we report a novel and versatile library construction method that can be used to capture the chromatin openness as ATAC-seq and can construct the NGS library with DNA samples sheared by different fragmentation methods. We named the method as SALP, which stands for the Single strand Adaptor Library Preparation. When constructing the tagmentation-based NGS library, the method can pool the different tagmented chromatins of various cells together by using variant barcoded Tn5 adaptors (BTAs). The pooled DNA can then be treated as one DNA sample in the subsequent library construction steps. The DNA sample was successively undergone a single strand adaptor (SSA) ligation, an elongation, and T adaptor ligation. The dual adaptor-ligated DNA was finally amplified with PCR by using the NGS-compatible primers. With this method, we captured the high-quality chromatin openness states of GM12878, HepG2, HeLa, and 293T cell lines in an unbiased and high-throughput format, which is helpful for comparatively characterizing the chromatin openness states of various cells. With this method, we also characterized the chromatin openness states of HepG2 cells by using different amount of cells, from $10^5$ to 500 cells.

## Methods

### Cultivation of cells

HeLa, HepG2 and 293T cells were purchased from the China Center for Type Culture Collection (Shanghai, China). GM12878 cell were purchased from the GuangZhou Jennio Biotech Co.,Ltd. (GuangZhou, China). HeLa, HepG2 and 293T cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO), GM12878 cells were cultured in RPMI 1640, which were all supplemented with 10% fetal bovine serum (FBS, GIBCO) with 100 μg/mL streptomycin and 100 units/mL penicillin at 37 °C in 5% $CO_2$.

### Preparation of various adaptors

All oligonucleotides were synthesized by Sangong Biotech (Shanghai) (Additional file 1: Table S1). For preparing the barcoded Tn5 adaptors (BTAs), barcode and ME oligos were dissolved in dd$H_2O$ at the concentration of 20 μM, and then mixed in equimolar in PCR tube. For preparing single strand adaptors (SSAs), SSA-PN and SSA-PNrev oligos were dissolved in dd$H_2O$ at the concentration of 100 μM, and then mixed in equimolar in PCR tube. For preparing T adaptor, TOA and TOArev oligos were dissolved in dd$H_2O$ at the concentration of 100 μM, and mixed in equimolar in PCR tube. Finally, all oligo mixtures were denatured in the water bath for 5 min at 95 °C and gradually cooled to 25 °C for annealing into various adaptors.

### Preparation of Tn5 transposome

Briefly, 4 μL of BTA (10 μM) was mixed with 2 μL 10× Tn5 transposome assemble buffer (TPS), 1 μL Tn5 transposase and 13 μL $H_2O$ according the instruction of the Tn5 transposase (Robust Tn5 Transposase, Robustnique Corporation Ltd., Tianjin, China). The reaction was gently mixed at 25 °C for 30 min to generate Tn5 transposome. The transposome was stored at − 20 °C for later use.

### Fragmentation of genomic DNA

The genomic DNA (gDNA) was extracted with phenol:chloroform from HepG2 cells. For tagmentation, 50 ng of gDNA was tagmented in a 30-μL reaction containing 1 × Tn5 transposome reaction buffer (LM buffer), 3 μL Dimethylformamide (DMF) (Sigma) and 4 μL Tn5 transposome at 55 °C for 15 min. The tagmented gDNA was purified with the MinElute PCR Purification Kit (QIAGEN, 28004). For Hind III digestion, 1 μg of DNA was digested in a 50-μL reaction containing 1× FastDigest Buffer and 5 μL FastDigest Hind III (Thermo Fisher, ER0501)

Wu *et al. BMC Genomics* (2018) 19:143

Page 3 of 12

at 37 °C overnight. For sonication, 1 μg of gDNA was sonicated by using the BRANSON sonicator with 70% power, 20s on and 20 s off for 20 cycles. The fragmented gDNAs were denatured at 95 °C for 5 min and chilled on ice for 5 min, which was then run with 1.5% agarose gel. The gDNA fragments of 200–1000 bp were recovered with the QIAquick Gel Extraction Kit (Qiagen, 28704).

### Optimization of single strand adaptor of SALP

To generate Illumina compatible sequencing libraries and improve the ligation efficiency, we designed and used single strand adaptors (SSAs) with 3′ overhang of 4 different length (1–4 random nucleotides). For SSA ligation, 12.5 ng of tagmented HepG2 gDNA was denatured at 95 °C for 5 min and chilled on ice for 5 min. The denatured gDNA was ligated with variant SSAs in a 10-μL reaction containing 1 μL of T4 DNA ligase (NEB, M0202L), 1× T4 DNA ligase buffer and 1 μL of SSA (5 μM) at 16 °C for 60 min. Then the reaction was mixed with equal volume of 2× rTaq mix (Takara) and incubated at 72 °C for 15 min. Finally, the gDNA was purified with 1.2× Ampure XP beads (Beckman Coulter) and amplified in a 50-μL PCR reaction containing 25 μL of NEBNext® Q5® Hot Start HiFi PCR Master Mix (NEB, M0543S), 1 μL of NEBNext Universal PCR Primer (10 μM), and 1 μL of NEBNext Index Primers (10 μM). The PCR program was as follows: (i) 98 °C for 5 min; (ii) 98 °C for 10s; 65 °C for 30s; 72 °C for 1 min; 18 cycles; (iii) 72 °C for 5 min. The PCR products were run with agarose gel and the DNA fragments of 300–1000 bp were extracted with the QIAquick Gel Extraction Kit.

The prepared library was analyzed with cloning sequencing. The extracted DNA was mixed with the equal volume of 2× rTaq mix (TaKaRa, R001A) and incubated at 72 °C for 15 min for A tailing. The DNA was purified and cloned into the PMD19-T Simple vector (Takara, 6013). The cloned DNA was sequenced by the Sanger method. For each of SSAs, 10 clones were sequenced.

### Preparation of NGS libraries with tagmented chromatins of various cells by using SALP

We performed SALP-seq with 100,000 GM12878, HeLa, HepG2, and 293T cells, and various numbers of HepG2 cells (100,000, 50,000, 10,000, 5000, 2500, and 500). Cells were collected by spinning at 500 g for 5 min at 4 °C and washed once with 50 μL of cold phosphate buffered solution (PBS). Cells were lysed by resuspended in cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$ and 0.1% IGEPAL CA-630). Cells were then spun at 500 g for 10 min at 4 °C to collect the nuclei precipitate. For tagmentation of 100,000 cells, nuclei were tagmented in a 30-μL reaction containing 20 μL of Tn5 transposome, 3 μL of DMF and 1× LM buffer. For tagmentation of 50,000 and 10,000 cells, nuclei were tagmented in a 30-μL

reaction containing 4 μL of Tn5 transposome, 3 μL of DMF and 1× LM buffer. For tagmentation of 5000, 2500 and 500 cells, nuclei were tagmented in a 5-μL reaction containing 1 μL of Tn5 transposome, 0.5 μL DMF and 1× LM buffer. The tagmentation reactions were mixed gently and incubated at 37 °C for 30 min, during which the reactions were gently mixed per 10 min to improve the tagmentation efficiency. In tagmentation, different BTAs (Additional file 1: Table S1 and Additional file 2: Table S2) were used to tagment the different cell samples. After tagmentation, the chromatins of 100,000 cells of 4 cell lines and 5 cell numbers of HepG2 cell were separately pooled together to generate two mixtures. To mixture, 1% SDS and 20 mg/mL Proteinase K (Sigma) were add to final concentrations of 0.1% and 400 μg/mL, respectively. The mixture was incubated at 65 °C for 1 h and then added with 1× TE buffer to a final volume of 200 μL. The gDNA was then purified from the mixture by using standard phenol:chloroform extraction as described [23]. For preparing NGS library, the purified gDNA was successively ligated with SSA, elongated with rTaq, and amplified with Illumina compatible index primers as described above. Finally, the DNA fragments of 150–1000 bp were recovered to include more nucleosome-free sequences. In library preparation, the SSA with 3N overhang was used according to SSA optimization.

### Preparation of NGS libraries with Hind III-digested and sonicated DNA by using SALP

The Hind III-digested and sonicated gDNA was ligated with the SSA of 3N overhang and elongated with the same procedure as the tagmented DNA. Then the gDNA was ligated with a T adaptor in a 10-μL containing 1 μL of T adaptor (5 μM), 1× T4 DNA Ligase Reaction Buffer and 1 μL of T4 DNA ligase at 16 °C for 2 h. Finally, the gDNA was purified with 1.2× Ampure XP beads and amplified with different NEB index primers (Additional file 1: Table S1) in a 50-μL PCR reaction as described above. The PCR products were run with agarose gel and the gDNA fragments of 300–1000 bp were extracted with the QIAquick Gel Extraction Kit.

### Next-generation sequencing

After amplified with Illumina-compatible primers (Additional file 1: Table S1), four NGS libraries, including NGS-L1 (tagmented chromatins of four cell lines), NGS-L2 (tagmented chromatins of five cell numbers), NGS-L3 (HindIII-digested gDNA), and NGS-L4 (sonicated gDNA), were constructed by using SALP. The libraries were detected and quantified with Agilent Bioanalyzer 2100. Four libraries were mixed together according to the DNA quality (ng) at the ratio of 4:1:1:1 (NGS-L1:NGS-L2:NGS-L3:NGS-L4) and sequenced by a lane of Illumina Hiseq X Ten platform (Nanjing Geneseeq).

Wu *et al. BMC Genomics* (2018) 19:143

Page 4 of 12

## Analysis of SALP-seq data

The raw reads data were separated according to the index and barcode by using a homemade Perl scripts. Then the ME (19 bp) and barcode (6 bp) sequences were removed from the 5′ end of the pair-end sequencing reads 2. All reads were cut to 30 bp and aligned to the human genome (hg19) by using Bowtie program (version 1.1.2) [24], with the default settings, except that the parameter −X 2000 was used to ensure the long fragments could be aligned to the genome. Peak calling was performed with macs2 [25], with the following parameters: -f BEDPE −keep-dup = 2. Peak annotation was performed with Homer software [26]. Gene ontology analysis was performed with PANTHER by uploading the genes to the website (http://pantherdb.org/) [27]. De novo motif analysis was performed with MEME tool and the comparison between enriched motifs and HOCOMOCO human v10 database was performed with Tomtom tool from MEME software suite [28, 29]. The BEDTools intersect was used to detect the overlapped peaks with −wa −u parameters [30]. All peak tracks were displayed with UCSC genome browser and statistical analysis were performed with R program and homemade Perl scripts. The ATAC-seq reads data of GM12878 was downloaded from the GEO database (accession number: GSE47753). The ATAC-seq reads data were analyzed as the SALP-seq reads data to compare.

## Results

### Development of tagmentation-based SALP method

The DNA fragments tagmented by Tn5 transposase have recently been used to facilely construct the NGS library. We firstly constructed the NGS library by using Tn5-tagmented DNA fragments with SALP method (Fig. 1a and b). For this end, a new kind of barcoded Tn5 adaptor (BTA) was designed (Fig. 1a). The barcode can be used to discriminate the different samples pooled together after the tagmentation. By using the complex of Tn5 transposase and BTA, the gDNA from HepG2 cells was efficiently fragmented (Additional file 3: Figure S1A), based on the "cut and paste" mechanism of Tn5 [31]. The tagmented fragments were then denatured and a single strand adaptor (SSA) was ligated. The SSA was a double-stranded oligonucleotide with a 3′ overhang of 1–4 random nucleotides. After ligation, the gDNA was elongated with Taq polymerase and amplified with PCR by using a pair of primers that can anneal with BTA and SSA, respectively. The PCR amplification reveals that the SSA with 3-base overhang produced the highest ligation efficiency in all SSAs (Additional file 3: Figure S1B). To further confirm the constructed DNA library, the PCR products were cloned into T-vector and 40 clones identified by the colony PCR were sequenced by using Sanger method. The colony PCR detection indicated that the cloned DNA fragments had the size from 200 to 1000 bp (Additional file 3: Figure S1C). The sequencing revealed
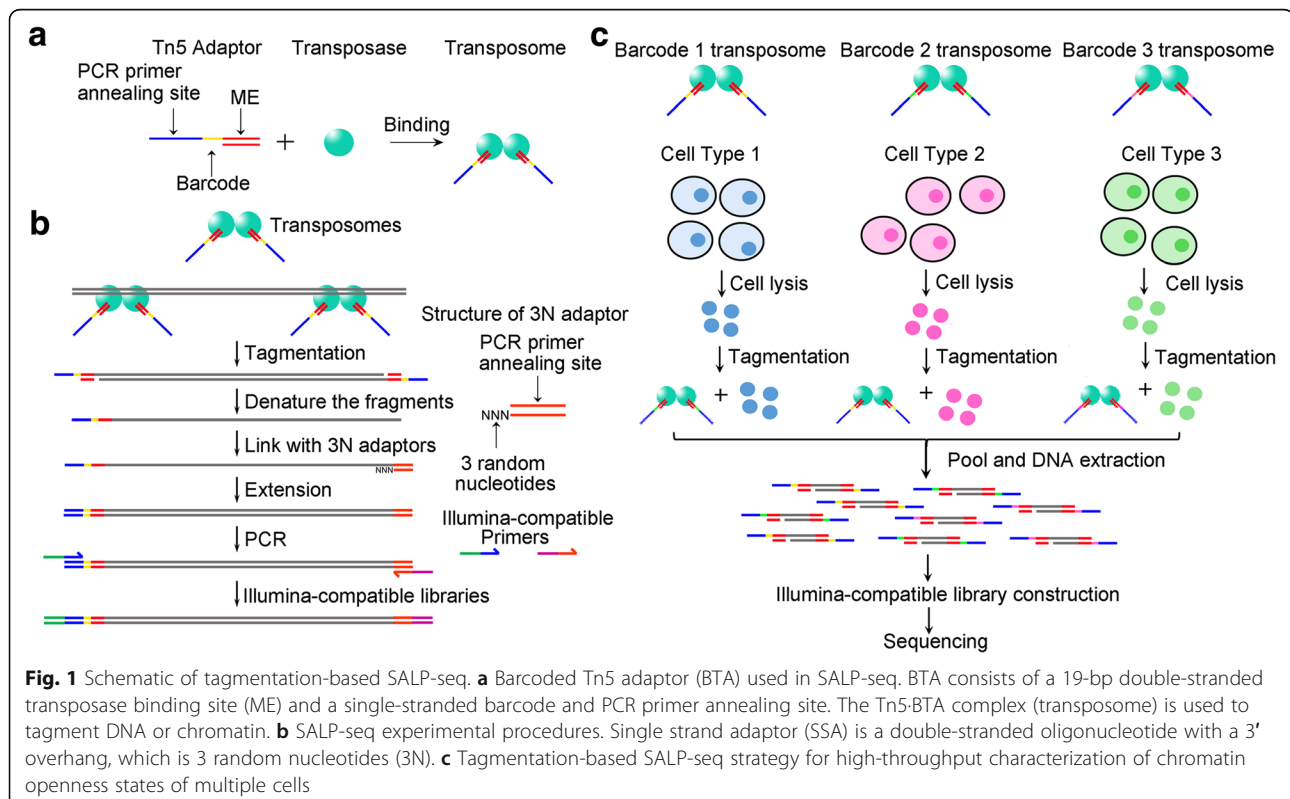


**Fig. 1** Schematic of tagmentation-based SALP-seq. **a** Barcoded Tn5 adaptor (BTA) used in SALP-seq. BTA consists of a 19-bp double-stranded transposase binding site (ME) and a single-stranded barcode and PCR primer annealing site. The Tn5·BTA complex (transposome) is used to tagment DNA or chromatin. **b** SALP-seq experimental procedures. Single strand adaptor (SSA) is a double-stranded oligonucleotide with a 3′ overhang, which is 3 random nucleotides (3N). **c** Tagmentation-based SALP-seq strategy for high-throughput characterization of chromatin openness states of multiple cells

Wu *et al. BMC Genomics* (2018) 19:143

Page 5 of 12

that the libraries were successfully constructed with 4 kinds of SSAs by using SALP method (Additional file 4), which were compatible with Illumina sequencing platform (Additional file 5: Figure S2). Based on these results, the SSA with 3-base overhang was subsequently employed to construct the following libraries. Because the ligation efficiency of SSA adaptors is critical for SALP-seq, we further validated the ligation efficiency of SSA adaptors. We found that all 4 kinds of SSA adaptors showed high ligation efficiency (Additional file 6: Figure S3).

### A high-throughput procedure for preparing tagmentation-based NGS library with SALP

Tn5 tagmentation has been recently used to profile the open chromatin regions in a technique named as ATAC-seq. Therefore, we then applied SALP in combination with NGS (SALP-seq) to the same field. Based on the BTA designed by this study, as many as 10 different BTAs (Additional file 1: Table S1) were used to prepare the NGS library for profiling the open chromatin regions in a facile high-throughput procedure (Fig. 1b). Four different cell lines including GM12878, HepG2, HeLa and 293T were respectively tagmented with the complexes of Tn5 and BTAs with different barcodes. The tagmented chromatins of four cell lines were then pooled together. Similarly, five different numbers of HepG2 cells were also respectively tagmented with the complexes of Tn5 and BTAs with other different barcodes. The tagmented chromatins of five cell numbers were then pooled together. DNA was then extracted from the pooled chromatins and two NGS libraries were prepared with the SALP procedure verified above. The two NGS libraries, together with two NGS libraries prepared with the enzymatically and physically fragmented gDNAs by using SALP, were then mixed together and sequenced by a lane of Illumina Hiseq X Ten platform. As a result, as many as 95,202,715 mappable reads were obtained (Additional file 7: Table S3).

### Characterization of chromatin openness state of lymphoblastoid with SALP-seq

The chromatin state of GM12878 cell had been widely explored by using various methods including DNase-seq, FAIRE-seq and ATAC-seq. We also explored the chromatin state of this cell line by using SALP-seq for comparing the SALP-seq results with those obtained by other methods. The genome-wide reads distribution of SALP-seq and ATAC-seq [32] was firstly compared, which revealed that the similar genome-wide reads distribution was obtained by two methods (Fig. 2a). Some reads highly enriched regions were commonly identified by two methods. However, some reads highly enriched regions were only identified by SALP-seq or ATAC-seq. Comparison of peaks generated by two methods showed

that SALP-seq found more peaks than ATAC-Seq according to the peak numbers normalized against sequencing depth (Fig. 2b). Calculation of reads density of overlapped peaks revealed that the reads density of SALP-seq peaks was higher than that of ATAC-seq peaks (Fig. 2c). Comparison of fold enrichment of peaks demonstrated that the peaks with low fold enrichment could be more easily identified by SALP-seq (Fig. 2d). These data indicate that SALP-seq could more sensitively find the open chromatin regions than ATAC-seq. The comparison of reads distribution demonstrated that the SALP-seq obtained the same reads density distribution as ATAC-seq (Fig. 2e), indicating the reliability of SALP-seq. For further confirming the reliability of SALP-seq in characterizing chromatin state, the peaks in a locus (Fig. 2f) was compared to those previously highlighted by other methods [33], which revealed that the same chromatin openness state was obtained by SALP-seq, ATAC-seq, FAIRE and DNase-seq. In addition, the SALP-seq peaks are well matched with the H3K27Ac track peaks and DNase clusters, further indicating the reliability of SALP-seq in characterizing chromatin state. Moreover, comparing with the H3K27Ac peaks, SALP-seq could identify many new peaks. It was found that these unique SALP-seq peaks had lower fold enrichment than those peaks commonly identified by both H3K27Ac ChIP-seq and SALP-seq (Additional file 8: Figure S4), indicating the high sensitivity of SALP-seq in identifying low-openness chromatin regions.

### Characterization of chromatin openness states of different cells with SALP-seq

Based on BTAs designed by this study, a NGS library was prepared with four cell lines including GM12878, HepG2, HeLa and 293T in a high-throughput approach by using SALP (Fig. 1c). The calculation of reads density around of transcription start sites (TSSs) revealed that the TSS regions showed the high reads density (Fig. 3a), indicating the high-level chromatin openness in these regions as previously reported [34]. In order to compare the chromatin openness states among different cell lines, reads density were calculated at the whole genome scale (Fig. 3b). It is clear that some regions showed as the common open chromatin regions in all cell lines, such as regions located in chromosome 5. As an example, a chromatin locus in chromosome 19 was shown, which reveals that the highly enriched SALP-seq peaks exist in all cell lines (Fig. 3c). These peaks also co-located with the H3K27Ac sites and DNase clusters generated by ENCODE Consortium (Fig. 3c). At the genome-wide scale, many co-located peaks were found in different cell lines, indicating that there are many common chromatin open regions among different cells.
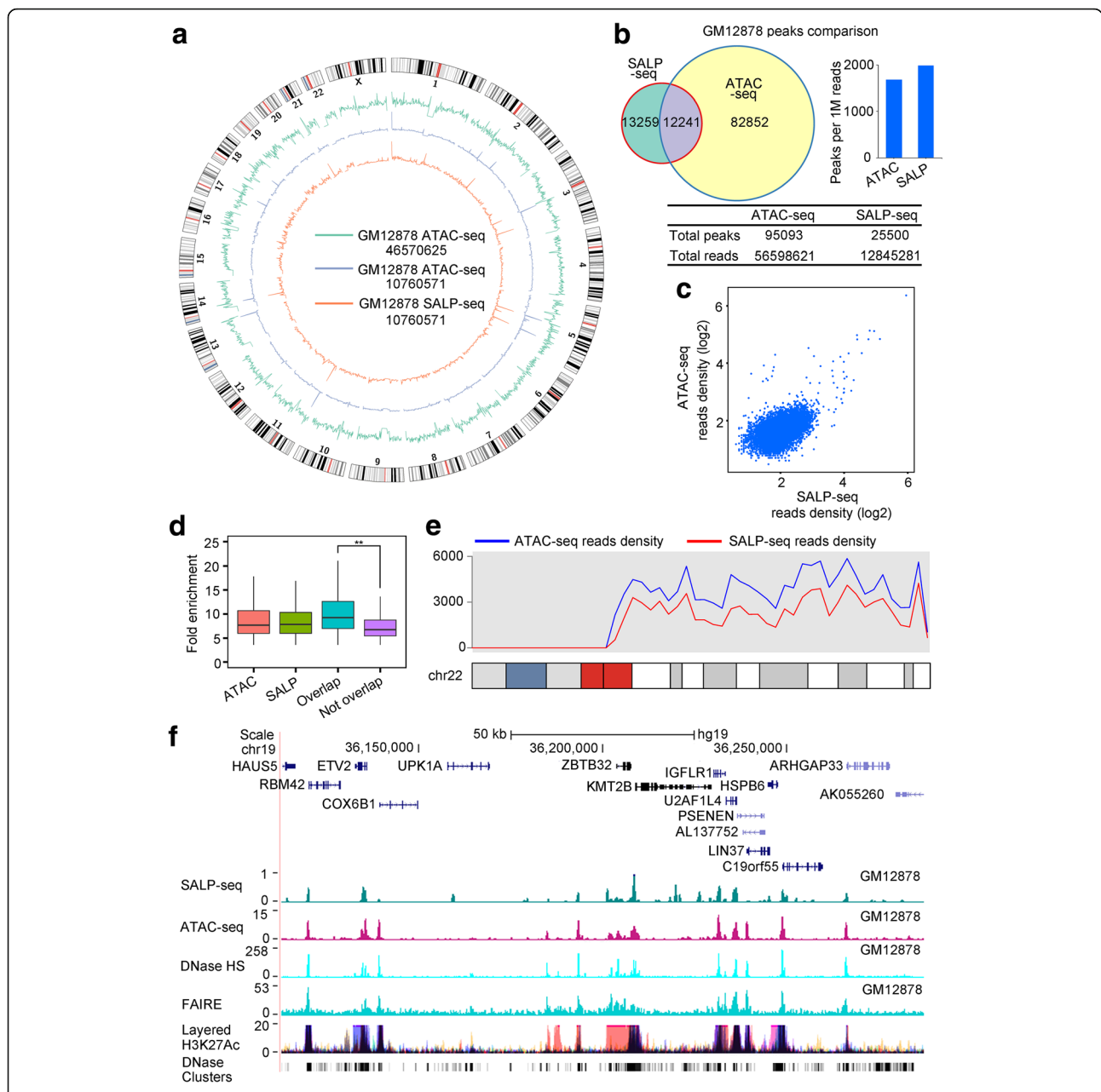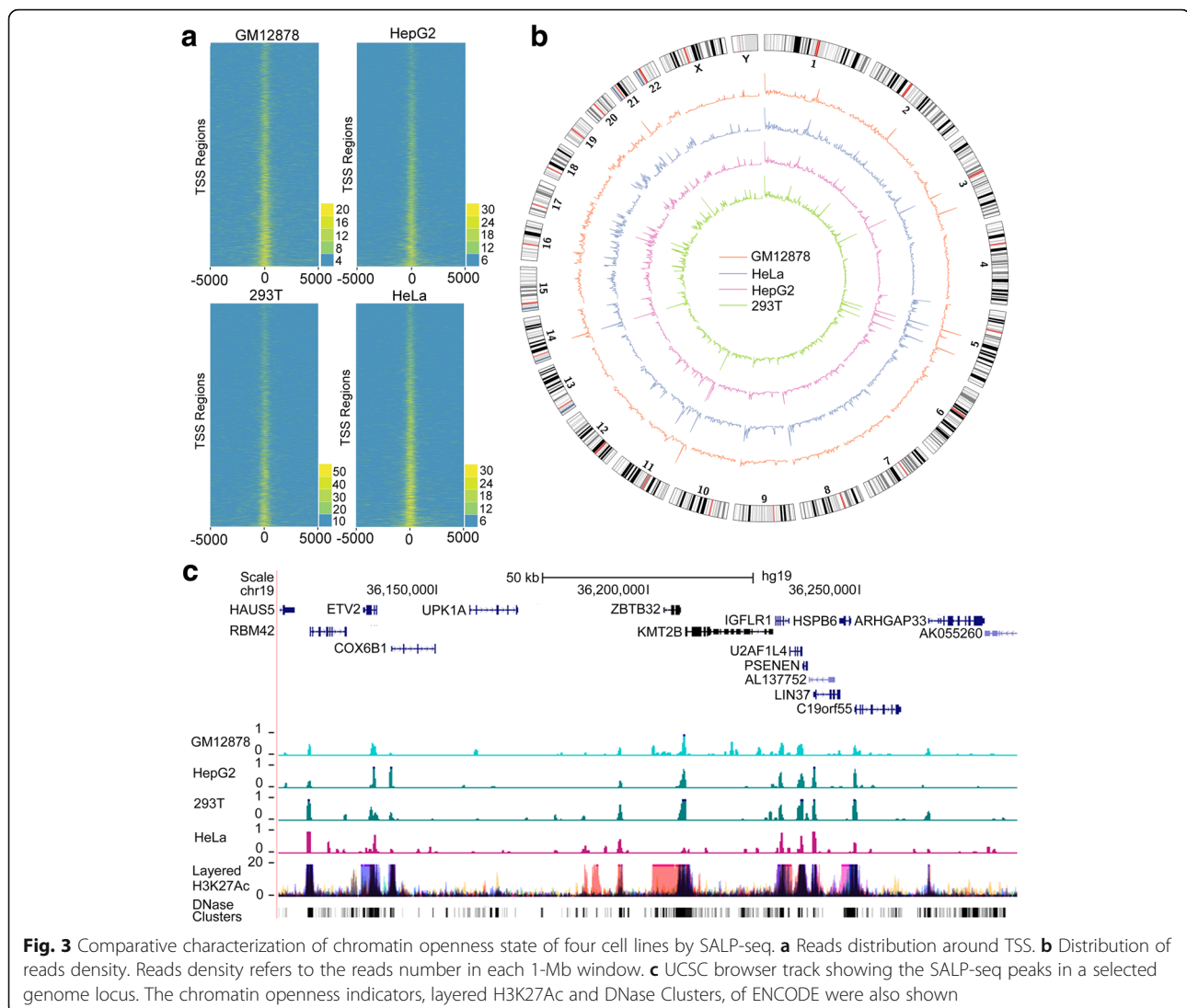
Wu *et al. BMC Genomics* (2018) 19:143

Page 6 of 12



**Fig. 2** Comparative characterization of chromatin openness states of GM12878 cell line by SALP-seq and ATAC-seq. **a** Distribution of reads density. The reads density refers to the reads number in each 1-Mb window. **b** Comparison of peaks identified by two methods. **c** Comparison of reads density of common peaks identified by two methods. **d** Fold enrichment of various peaks. ATAC, ATAC-seq peaks called with the same amount of randomly selected reads as total SALP-seq reads. SALP, SALP-seq peaks; Overlap, Overlapped ATAC-seq and SALP-seq peaks; Not Overlap, SALP-seq peaks not overlapped with ATAC-seq peaks. **e** Distribution of reads density on chromosome 22. **f** UCSC browser track showing chromatin openness state of a selected region characterized by SALP-seq and other methods

However, some common chromatin open regions showed the different open level. Additionally, each cell line contains its specific peaks, indicating the cell-type-specific chromatin openness states among different cells. These data reveal that SALP-seq provides an easy method to perform the comparative characterization of chromatin states of different cells.

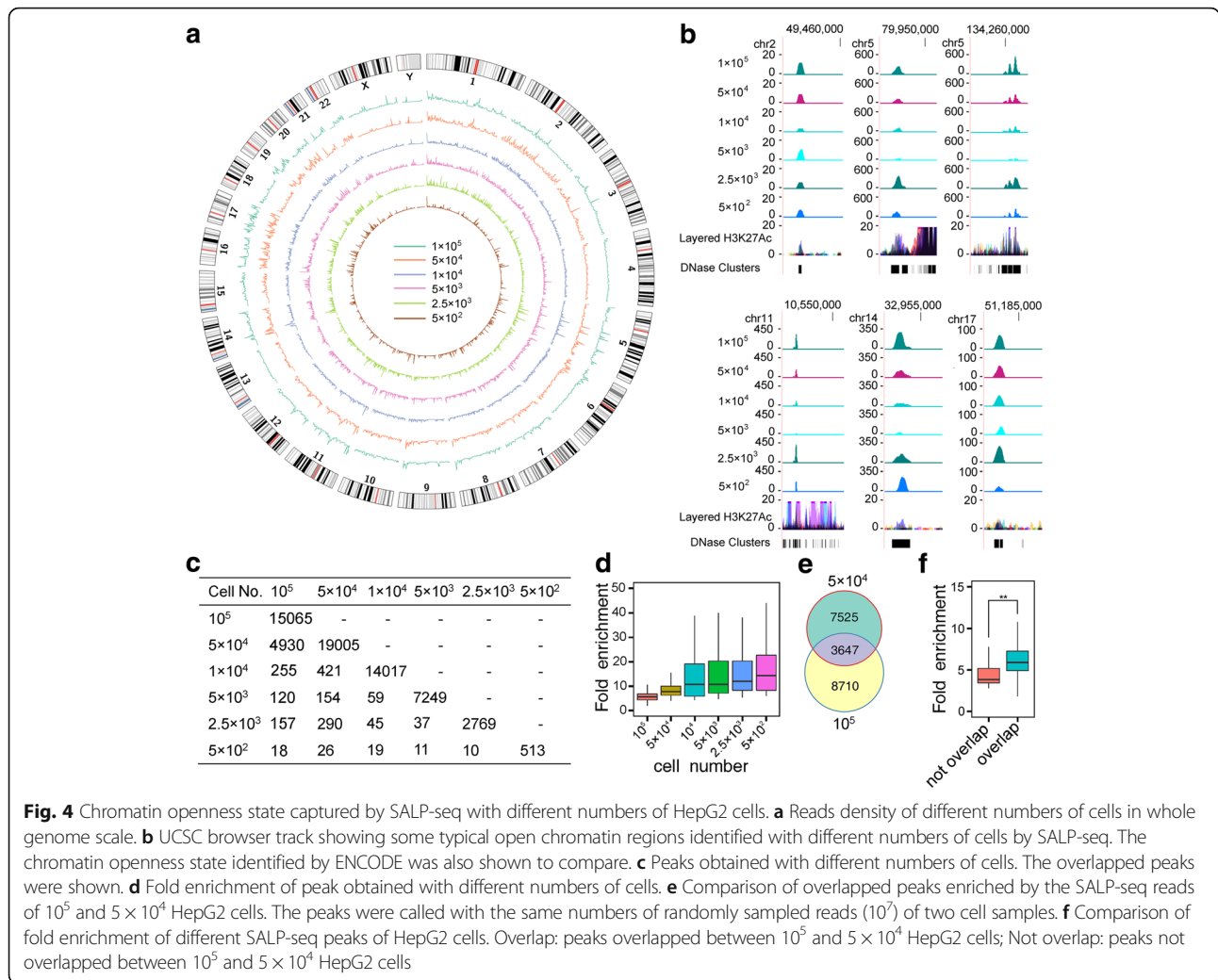**Characterization of open chromatin regions with different numbers of cells by SALP-seq**

As the methods to detect the chromatin openness, FAIRE-seq [35], and DNase-seq [36], usually requires 1–50 million cells, but ATAC-seq requires 50,000 to 500 cells [32]. To validate that the chromatin accessible regions could be identified with different numbers of cells by using SALP-

Wu *et al. BMC Genomics* (2018) 19:143

Page 7 of 12



**Fig. 3** Comparative characterization of chromatin openness state of four cell lines by SALP-seq. **a** Reads distribution around TSS. **b** Distribution of reads density. Reads density refers to the reads number in each 1-Mb window. **c** UCSC browser track showing the SALP-seq peaks in a selected genome locus. The chromatin openness indicators, layered H3K27Ac and DNase Clusters, of ENCODE were also shown

seq, we performed SALP-seq with different numbers of HepG2 cells, including 100,000, 50,000, 10,000, 5000, 2500, and 500 cells. The reads density of each number of cells was calculated at the genome scale, which revealed that the chromatin regions with high accessibility were commonly identified by 5 different numbers of cells (Fig. 4a and b). The open chromatin regions identified by SALP-seq are matched with the H3K27ac track and DNase Clusters of ENCODE (Fig. 4b). Nevertheless, the sensitivity was diminished when low number of cells was used as starting material (Fig. 4c). With low numbers of cells, only most of open chromatin regions with high fold enrichment were identified (Fig. 4d). It means that SALP-seq can identify those highly open chromatin regions even with as few as 500 cells. Moreover, comparison of peaks of $10^5$ and $5 \times 10^4$ HepG2 cells revealed that SALP-seq has high reproducibility in identifying open chromatin regions (Fig. 4e), especially those with high openness (Fig. 4f).

## Construction of NGS library with enzymatically digested or sonicated gDNA by SALP-seq

To show that SALP could be used to construct NGS library with any other DNA fragments, we finally constructed NGS libraries with Hind III-digested and sonicated HepG2 gDNA by using an adapted SALP procedure (Fig. 5a). In this procedure, after the SSA was ligated and elongated, a T adaptor with one T base overhang was ligated to the other end of DNA fragments. Then DNA was amplified with two primers annealing to SSA and T adaptor for generating the Illumina compatible sequencing libraries. The results revealed that the NGS libraries could be successfully constructed with the enzymatically and physically sheared gDNAs (Additional file 9: Figure S5), which were also sequenced by using Illumina Hiseq platform together with tagmented DNA (Additional file 7: Table S3). For checking the coverage of the constructed libraries, the genome-wide reads density of two NGS libraries was calculated. The

Wu *et al. BMC Genomics* (2018) 19:143

Page 8 of 12



**Fig. 4** Chromatin openness state captured by SALP-seq with different numbers of HepG2 cells. **a** Reads density of different numbers of cells in whole genome scale. **b** UCSC browser track showing some typical open chromatin regions identified with different numbers of cells by SALP-seq. The chromatin openness state identified by ENCODE was also shown to compare. **c** Peaks obtained with different numbers of cells. The overlapped peaks were shown. **d** Fold enrichment of peak obtained with different numbers of cells. **e** Comparison of overlapped peaks enriched by the SALP-seq reads of $10^5$ and $5 \times 10^4$ HepG2 cells. The peaks were called with the same numbers of randomly sampled reads ($10^7$) of two cell samples. **f** Comparison of fold enrichment of different SALP-seq peaks of HepG2 cells. Overlap: peaks overlapped between $10^5$ and $5 \times 10^4$ HepG2 cells; Not overlap: peaks not overlapped between $10^5$ and $5 \times 10^4$ HepG2 cells

results indicate that the reads were uniformly distributed on chromosomes (Fig. 5b and c; Additional file 10: Figure S6 and Additional file 11: Figure S7), indicating that the NGS library could be successfully constructed with the enzymatically and physically sheared gDNAs by using SALP. Additionally, the genome-wide distributions of the predicted Hind III digestion sites basically matched with the reads density of the NGS library constructed with Hind III-digested gDNA (Fig. 5b and Additional file 10: Figure S6). However, some significant differences were also found in several chromosomes such as Chromosome 5, 9 and 11 (Additional file 10: Figure S6), which may be caused by the difference in sequences between the reference genome (hg19) and the HepG2 cell genome.

## Discussion

In the current NGS era, as development of the next-generation sequencing, NGS has great advantage over conventional methods in production of large volumes of sequence data [2]. With this advantage, NGS has been

used in virtually all branches of biological and medical researches. As the fundamental step of NGS, the library construction methodology has been attracting increasing attention. Here, we report the development of a novel method, named as SALP, for constructing the NGS library. SALP can be used to construct high-quality NGS libraries with enzymatically or physically sheared gDNA. Importantly, SALP can also be used to construct high-quality NGS libraries based on Tn5 tagmentation in an unbiased high-throughput format, which can be widely used in the current increasing research of chromatin state.

With the feature of "cut and paste" of the transposase, a Tn5-based library construction method, named ATAC-seq, has significant advantages over conventional library construction methods, such as a few steps and low input requirements. However, ATAC-seq is limited by an inherent drawback that as many as 50% of tagmented fragments could not be sequenced due to using two different Tn5 adaptors. The Tn5-based SALP overcomes this limitation. In the Tn5-based SALP procedure, only one kind of Tn5
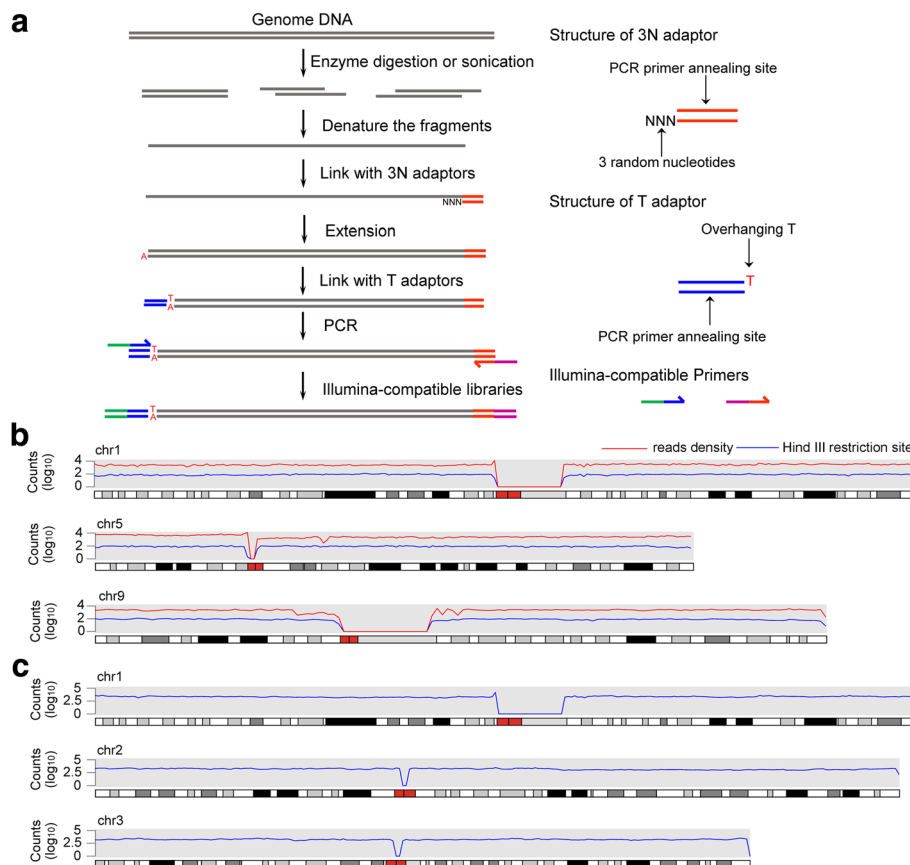
Wu *et al. BMC Genomics* (2018) 19:143

Page 9 of 12



**Fig. 5** NGS library construction with alternatively sheared gDNA by using SALP method. **a** Schematic of library construction with the enzymatically digested and sonicated DNA by using SALP method. **b** Distribution of reads density of NGS library constructed with the Hind III-digested HepG2 gDNA and the density of predicted Hind III restriction sites in whole genome. **c** Distribution of reads density of NGS library constructed with the sonicated HepG2 gDNA. The information of other chromosomes was shown as Additional file 10: Figure S6 and Additional file 11: Figure S7

transposome adaptor was used to assemble with Tn5. By using SSA, all fragments generated by Tn5 tagmentation can be sequenced. We also optimized the other key adaptor, SSA, for SALP. The SSAs with 3′ end overhang ranging from 1 to 4 random base were tested, which demonstrated that the SSA with overhang of 3 random bases showed the highest ligation efficiency in the key step of SALP. Nevertheless, we found that the SSAs with overhang of 1–4 random bases all successfully produced library. The SSAs with overhang of 1–2 random bases showed limited ligation efficiency, which may be related to the reason that the overhangs of these SSAs might be too short to form the stable structure under the ligase reaction condition and few SSAs were linked to DNA fragments. The SSA with overhang of 3 random bases could anneal with the end of the single-stranded fragments efficiently and tightly, which allows ligase to ligate SSA efficiently. When the SSA with overhang of 4 random bases was used, stable hybrids could be formed by 3 of 4 random bases; however, a one-base gap or mismatch can be easily generated in the annealed hybrids, which can reduce the efficiency of ligation and elongation.

Chromatin organization with cell type-specificity can affect accessibility and activity of regulatory elements and the manifestation of unique cellular phenotypes [37]. Chromatin openness state as an important feature of chromatin structure exerts significant influence on gene transcription. SALP-seq can be used to identify the open chromatin region easily. Comparing the open chromatin regions identified by SALP-seq and ATAC-seq of GM12878 cell line revealed that the reads density of SALP-seq were higher than that of ATAC-seq in co-located peaks. This phenomenon is caused by the principle of SALP method. In SALP, a double-strand DNA (dsDNA) fragment was firstly denatured into two single-strand DNAs (ssDNAs), which were then linked with SSA and elongated into two dsDNAs like DNA replication. This process allows one input dsDNA fragment to be sequenced twice in SALP-seq, which can reduce the sequencing error and mutation robustly. This process allows SALP-seq to detect the low- or ultra-low-frequency mutations with strand specificity, which are often related to somatic mutations [38]. Compared with

Wu *et al. BMC Genomics* (2018) 19:143

Page 10 of 12

ATAC-seq, SALP-seq can identify more peaks due to its capability to find peaks with low fold enrichment. These results are caused by the fact that all Tn5 tagmented fragments could be sequenced in SALP-seq and many regions with lower openness state can thus be captured by SALP-seq but may be missed by ATAC-seq.

The Tn5 adaptor used in SALP-seq are specially designed, which has a unique barcode sequence between the ME and PCR primer annealing sites. With this design, chromatins of different cell samples tagmented by different BTAs can be pooled together. The chromatin cocktail can then be treated as one DNA sample, which greatly simplifies the library construction of multiple cell or DNA samples. The chromatin openness state of various cells can be easily performed with minimal bias caused by library construction process, which is beneficial to the comparative characterization of chromatin state. This study compared the chromatin openness state of four difference cell lines including GM12878, HepG2, HeLa, and 293T with SALP-seq in this way, which clearly revealed the common and cell-type-specific open chromatin regions among 4 cell lines. In our knowledge, this is the first time to explore the chromatin state of different cells like this, which is currently realized by using multiple PCR amplifications of different cell samples with different Illumina index primers in ATAC-seq. The tagmentation-based SALP-seq should be useful for comparative characterization of chromatin state of clinical tissues [22].

It should be pointed out that the NGS library construction strategy described in Fig. 5a can also be used to construct DNA library in high throughput by using the barcoded SSA. In this case, a double-stranded barcode sequence can be placed between the 3′ overhang and primer-annealing site. By using such barcoded SSAs, the SSA-ligated ssDNAs from different DNA samples can be pooled together. The pooled DNAs can be treated as one DNA sample to undergo the followed library construction steps, including elongation, T adaptor ligation and PCR amplification. As using the barcoded Tn5 adaptors, this can simplify the experimental operation, reducing artificial bias, and saving the labor, time and cost.

Finally, SALP should be useful for the NGS library construction of highly degraded DNA and free blood DNA, including circulating free DNA (cfDNA), circulating tumor DNA (ctDNA), and cell-free fetal DNA (cffDNA). Due to the high SSA ligation efficiency, SALP should be useful for the amplification, detection, and NGS library construction of free blood DNA and highly degraded DNA. SALP should find its wide applications in liquid biopsy, in vitro diagnostic products (IVD), non-invasive prenatal testing (NIPT), forensic medicine, and even paleontology and archaeology (treating ancient DNA) in the future. Although several single-stranded DNA library preparation methods have already been used in these fields [5–8], the wide applications of these methods are still limited by their intrinsic drawbacks. In our opinion, SALP overcomes these shortcomings due to its simplicity and high efficiency.

## Conclusions

In conclusion, we have developed a new method, named SALP, for constructing NGS library, which is based a novel single strand adaptor linking technique. This method can be used to construct NGS library with the DNA fragments sheared by various methods including tagmentation, sonication and enzymatic digestion. This method constructs NGS library in several simple steps just dependent on two cheap routine enzymes, T4 DNA ligase and Taq polymerase. Combining with Tn5 tagmentation technique, this method can simply the NGS library construction procedure in five steps, including tagmentation, SSA ligation, elongation, T adaptor ligation, and index PCR amplification. The Tn5-based SALP-seq overcomes the key shortcoming of the current ATAC-seq technique and develops a high-throughput procedure for constructing NGS libraries with multiple cell samples. The Tn5-based SALP-seq can thus be widely used to characterize the comparative chromatin openness states or genome sequences of various tissues or cells, such as cancerous tissues from various patients, which may be contribute to the current increasing personalized and precision medicine. SALP can construct NGS library with low number of cells in a high-throughput format, which is also useful for analyzing clinical tissue samples.

## Additional file

**Additional file 1: Table S1.** Oligonucleotides used as adaptors and PCR primers. (DOCX 444 kb)

**Additional file 2: Table S2.** Barcodes on Barcoded Tn5 adaptors for labeling different cell samples. (DOCX 395 kb)

**Additional file 3: Figure S1.** Validation of SALP method. (DOCX 15 kb)

**Additional file 4:** Cloning sequencing. (DOCX 14 kb)

**Additional file 5: Figure S2.** The structure of SALP library. (DOCX 309 kb)

**Additional file 6: Figure S3.** Verification of the ligation efficiency of SSA adaptors. (DOCX 21 kb)

**Additional file 7: Table S3.** Reads from a lane of Illumina Hiseq X Ten sequencing. (DOCX 68 kb)

**Additional file 8: Figure S4.** Comparison of fold enrichment of two types of GM12878 SALP-seq peaks. (DOCX 187 kb)

**Additional file 9: Figure S5.** Construction of NGS library of gDNAs sheared by sonication and restriction endonuclease digestion with SALP method. (DOCX 15 kb)

**Additional file 10: Figure S6.** Comparison of the distribution of Hind III digestion library reads density and Hind III restriction sites through the whole genome. (DOCX 45 kb)

**Additional file 11: Figure S7.** Reads distribution of sonication library. (DOCX 295 kb)

Wu *et al. BMC Genomics* (2018) 19:143

Page 11 of 12

## Abbreviations

## Acknowledgements

## Funding

## Availability of data and materials

The raw reads data from SALP-seq are available at NCBI GEO with the accession number: GSE 104162. Supplementary Data are available at Online.

## Authors' contributions

## Ethics approval and consent to participate

## Consent for publication

## Competing interests

## References

1. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418–26.
2. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet. 2010;11(1):31–46.
3. Mardis ER. DNA sequencing technologies: 2006-2016. Nat Protoc. 2017;12(2):213–8.
4. Adey A, Morrison HG, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. 2010;11(12):R119.
5. Lazinski DW, Camilli A. Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. BioTechniques. 2013;54(1):25–34.
6. Gansauge MT, Gerber T, Glocke I, Korlevic P, Lippik L, Nagel S, Riehl LM, Schmidt A, Meyer M. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. 2017;45(10):e79.
7. Gansauge MT, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat Protoc. 2013;8(4):737–48.
8. Peng X, Wu J, Brunmeir R, Kim SY, Zhang Q, Ding C, Han W, Xie W, Xu F. TELP, a sensitive and versatile library construction method for next-generation sequencing. Nucleic Acids Res. 2015;43(6):e35.
9. Raine A, Manlig E, Wahlberg P, Syvanen AC, Nordlund J. SPlinted ligation adapter tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. Nucleic Acids Res. 2017;45(6):e36.
10. Syed F, Grunenwald H, Caruccio N. Next-generation sequencing library preparation: simultaneous fragmentation and tagging using in vitro transposition. Nat Meth. 2009;6(11). https://doi.org/10.1038/nmeth.f.1272.
11. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet. 2014;15(2):69–81.
12. Johnson DG, Dent SY. Chromatin: receiver and quarterback for cellular signals. Cell. 2013;152(4):685–9.
13. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9):613.
14. Denny SK, Yang D, Chuang C-H, Brady JJ, Lim JS, Grüner BM, Chiou S-H, Schep AN, Baral J, Hamard C. Nfib promotes metastasis through a widespread increase in chromatin accessibility. Cell. 2016;166(2):328–42.
15. Guo H, Hu B, Yan L, Yong J, Wu Y, Gao Y, Guo F, Hou Y, Fan X, Dong J. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. Cell Res. 2017;27(2):165.
16. Wu C, Chen Z, Dardalhon V, Xiao S, Thalhamer T, Liao M, Madi A, Franca RF, Han T, Oukka M. The transcription factor musculin promotes the unidirectional development of peripheral Treg cells by suppressing the TH2 transcriptional program. Nat Immunol. 2017;18(3):344–53.
17. Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, Zhang B, Liu B, Wang Q, Xia W. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature. 2016;534(7609):652–71.
18. Moskowitz DM, Zhang DW, Hu B, Le Saux S, Yanes RE, Ye Z, Buenrostro JD, Weyand CM, Greenleaf WJ, Goronzy JJ. Epigenomics of human CD8 T cell differentiation and aging. Sci Immunol. 2017;2(8):eaag0192.
19. Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. Nat Meth. 2016;13(12):1013–20.
20. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol Metab. 2016;5(3):233–44.
21. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016;48(10):1193–203.
22. Britton E, Rogerson C, Mehta S, Li Y, Li X, Fitzgerald RC, Ang YS, Sharrocks AD. Open chromatin profiling identifies AP1 as a transcriptional regulator in oesophageal adenocarcinoma. PLoS Genet. 2017;13(8):e1006879.
23. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harb Protoc. 2006;2006(1):pdb. prot4455.
24. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
25. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
26. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–89.
27. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from gene ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45(D1):D183–9.
28. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(suppl_2):W202–8.
29. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. 2016;44(D1):D116–25.
30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
31. Goryshin IY, Reznikoff WS. Tn5 in vitro transposition. J Biol Chem. 1998;273(13):7367–74.
32. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Meth. 2013;10(12):1213–8.
33. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75–82.

Wu *et al. BMC Genomics* (2018) 19:143

Page 12 of 12

34. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311–22.

35. Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (formaldehyde assisted isolation of regulatory elements). Methods. 2009;48(3):233–9.

36. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010;2010(2):pdb. prot5384.

37. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet. 2011;12(1):7.

38. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013;23(5):843–54.