Published online 15 September 2014 *Nucleic Acids Research, 2014, Vol. 43, No. 6* ***e35***
doi: 10.1093/nar/gku818

TELP, a sensitive and versatile library construction method for next-generation sequencing

Xu Peng[1,†], Jingyi Wu[2,†], Reinhard Brunmeir[1], Sun-Yee Kim[1], Qiongyi Zhang[1], Chunming Ding[1], Weiping Han[3,4], Wei Xie[2] and Feng Xu[1,4,*]

[1]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), Singapore 117609, Singapore, [2]Tsinghua University-Peking University Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China, [3]Laboratory of Metabolic Medicine, Singapore Bioimaging Consortium, A*STAR, Singapore 138667, Singapore and [4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore

Received January 23, 2014; Revised July 07, 2014; Accepted August 29, 2014

ABSTRACT

Next-generation sequencing has been widely used for the genome-wide profiling of histone modifications, transcription factor binding and gene expression through chromatin immunoprecipitated DNA sequencing (ChIP-seq) and cDNA sequencing (RNA-seq). Here, we describe a versatile library construction method that can be applied to both ChIP-seq and RNA-seq on the widely used Illumina platforms. Standard methods for ChIP-seq library construction require nanograms of starting DNA, substantially limiting its application to rare cell types or limited clinical samples. By minimizing the DNA purification steps that cause major sample loss, our method achieved a high sensitivity in ChIP-seq library preparation. Using this method, we achieved the following: (i) generated high-quality epigenomic and transcription factor-binding maps using ChIP-seq for murine adipocytes; (ii) successfully prepared a ChIP-seq library from as little as 25 pg of starting DNA; (iii) achieved paired-end sequencing of the ChIP-seq libraries; (iv) systematically profiled gene expression dynamics during murine adipogenesis using RNA-seq and (v) preserved the strand specificity of the transcripts in RNA-seq. Given its sensitivity and versatility in both double-stranded and single-stranded DNA library construction, this method has wide applications in genomic, epigenomic, transcriptomic and interactomic studies.

INTRODUCTION

Next-generation sequencing (NGS) technology enables the comprehensive analysis of genomes, epigenomes, transcriptomes and interactomes. Chromatin immunoprecipitation (ChIP) coupled with NGS (ChIP-seq) and cDNA sequencing (RNA-seq) are becoming standard tools for genome-wide profiling of histone modifications, transcription factor binding and gene expression. One crucial step of NGS is sequencing library construction, in which minimal amounts of DNA are faithfully amplified and offered proper structure for sequencing. Although all NGS platforms provide standard protocols for sequencing library construction, these protocols are not without limitations. For example, the standard ChIP-seq methods require nanogram quantities of immunoprecipitated DNA (1), which is challenging for many ChIP studies on transcription factor binding, especially when cell numbers are limited (e.g. stem cells isolated *in vivo* and many clinical samples). Thus, there is a need for highly sensitive library construction methods to apply ChIP-seq to rare cell types.

The reads obtained from NGS are shorter than those obtained from conventional capillary electrophoresis-based Sanger sequencing. This fact poses challenges for accurate sequence alignment to the reference genome, especially to regions with a high abundance of repetitive sequences. A solution to this limitation is paired-end sequencing, which provides sequence information on both ends of DNA fragments and thus enables precise mapping of the sequencing reads to the reference genome or, more significantly, to an unknown genome (*de novo* sequencing).

RNA-seq is a powerful tool for novel transcript identification and known transcript quantification. Compared with traditional microarray-based methods, RNA-seq offers an unlimited dynamic range and superior sensitivity for quantitative applications. Moreover, RNA-seq makes it possible to annotate the 5′ and 3′ ends and splice junctions of all transcripts at the genome-wide level. Another advantage of RNA-seq is that information regarding the originally transcribed strand can be determined through

*To whom correspondence should be addressed. Tel: +65 64070427; Fax: +65 67766840; Email: xu_feng@sics.a-star.edu.sg
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

strand-specific RNA sequencing. This valuable information helps to identify antisense transcripts, determine the boundaries of adjacent genes transcribed on opposite strands and quantify the expression levels of overlapping genes (2).

Here, we report a novel and versatile library construction method that can be applied to both ChIP-seq and RNA-seq on the widely used Illumina platform. We named our sequencing library preparation method TELP, which stands for tailing (T), extension (E), ligation (L) and PCR (P), the four main steps in the protocol. Using this method, we first generated high-quality epigenomic maps through ChIP-seq for adipocytes, a cell type of interest to us. Adipocytes are the main storage site for excess energy in mammals. These cells are originally derived from multipotent mesenchymal stem cells (MSCs). The adipocyte maturation process is largely controlled by various adipogenic factors, among which, peroxisome proliferator-activated receptor γ (PPARγ) plays a central role (3). PPARγ is the only factor that is both necessary and sufficient to promote adipogenesis, and thus, it is considered the master regulator of this process (4,5).

To validate our library preparation method in ChIP-seq analysis of transcription factor binding, we profiled PPARγ binding by TELP in mature adipocytes. Significantly, we further showed that TELP is a highly sensitive ChIP-seq library construction method that can be successfully applied to as little as 25 pg of starting ChIP DNA. Another feature of TELP is that it allows paired-end sequencing. Given that TELP is compatible with both double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) library construction, we applied it to cDNA sequencing to systematically study gene expression dynamics during adipogenesis. Moreover, TELP preserves the strand specificity of the transcripts in RNA-seq, supporting its broad applications in the identification of antisense transcripts and quantification of overlapping transcripts.

MATERIALS AND METHODS

Cell culture, differentiation and staining

The mouse adipocyte progenitor cell line 3T3-L1 and the multi-potent MSC line C3H 10T1/2 were obtained from the American Type Culture Collection. These cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% calf bovine serum (3T3-L1) or fetal bovine serum (FBS) (10T1/2). Adipogenic induction was performed by treating confluent cells with DMEM containing 10% FBS, 1 μM dexamethasone, 0.5 mM 3-isobutyl-1-methylxanthine and 10 μg/ml insulin for 2 days, followed by 10 μg/ml insulin alone for 5 days. The PPARγ agonist rosiglitazone (1 μM) was added for 10T1/2 cell differentiation (6). Oil-Red-O staining was performed as previously described (7).

ChIP and high-sensitivity ChIP-seq library preparation

ChIP assays were performed essentially as described (6). Briefly, mature adipocytes were crosslinked with 1% formaldehyde for 10 min at 37°C; then, crude nuclei were purified (8). Chromatin was fragmented by sonication with a Bioruptor UCD-300 (Diagenode) to obtain fragments ∼200–800 bp in length (confirmed by electrophoresis on a 1% agarose gel). For each ChIP assay, 2–5 μg of antibodies were added and incubated at 4°C overnight. The antibodies used were the following: anti-H3 K4 trimethylation (me3; Abcam, ab8580, Lot-GR73721-1); anti-H3 K9 acetylation (ac; Abcam, ab4441, Lot-GR78058-1); anti-H3K27ac (Upstate Biotechnology, 07–595, Lot-DAM1713305); anti-H3K27me3 (Upstate Biotechnology, 07–449, Lot-1999681) and anti-PPARγ (Cell Signaling, 81B8). Rabbit IgG (Sigma, I-5006) was used as a negative control.

ChIP DNA was quantified with a Qubit fluorometer (Invitrogen, Q32857) using the Quant-iT dsDNA HS assay kit (Invitrogen, Q32851). The Illumina standard ChIP-seq library was prepared with a ChIP-seq Sample Prep Kit (Illumina, IP-102–1001), using 10 ng of ChIP DNA as starting material. Other sequencing libraries were constructed from 25 pg to 1 ng of ChIP DNA using TELP as described below.

First, ChIP DNA fragments were end-repaired by mixing 32.6 μl of DNA sample, 4 μl of 10× T4 ligase buffer (NEB, B0202S), 1.6 μl of 10 mM dNTP mix (NEB, N0447S), 0.8 μl of T4 PNK (NEB, M0201S), 0.8 μl of T4 DNA polymerase (NEB, M0203S) and 0.16 μl of Klenow fragment (NEB, M0210S) and incubating at 20°C for 30 min. This step is only required for library construction from small amounts of DNA fragmented by mechanical shearing (e.g. sonication). End repair products were purified with the MinElute PCR purification kit (Qiagen, 28006).

Next, poly-C tailing was initiated by mixing 28 μl of end-repaired DNA, 1 μl of 10× EX buffer (Takara, supplied with RR006A) and 1 μl of 1 mM dCTP (NEB, N0446S) and then denaturing the DNA. Subsequently, 1 μl of terminal deoxynucleotidyl transferase (TDT; NEB, M0315S) was added, and the reaction mixture was incubated at 37°C for 35 min. After poly-C tailing, TDT was inactivated by heating to 75°C for 20 min.

An extension step was performed by adding the following extension mix to the above-mentioned TDT reaction: 6.2 μl of H_2O; 0.8 μl of KAPA2G Robust HS (KAPA, KK5515); 12 μl of 5× KAPA buffer A (KAPA, supplied with enzyme); 4.8 μl of 2.5 mM dNTP (Takara, supplied with RR006A) and 6 μl of 2 μM biotin-labeled anchor primer. The anchor primer was designed to have nine consecutive Gs plus an H (which represents the nucleotides A, T or C, but not G) at the 3′-end to promote proper annealing at the beginning of the poly-C tail (as illustrated in Figure 1). The extension program was as follows: (i) 95°C for 3 min; (ii) 47°C for 1 min, 68°C for 2 min, 16 cycles and (iii) 72°C for 10 min. After the reaction, excess anchor primers were removed by exonuclease I (Exo I) (NEB, M0293S) digestion for 1 h at 37°C.

After Exo I treatment, the extension products were captured by pre-washed magnetic streptavidin C1 beads (Invitrogen, 650.01) in 1× Binding & Wash (B&W) buffer (10 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl) in a 1.5-ml Eppendorf Lobind tube (Eppendorf, 0030 108.051). DNA binding was performed in a thermomixer (Eppendorf, 5355 000.011) for 30 min by mixing at 1400 rpm (10 s on, 10 s off) at 23°C. Then, the beads were washed once with 100 μl of 1× B&W buffer, three times with 150 μl of EBT buffer

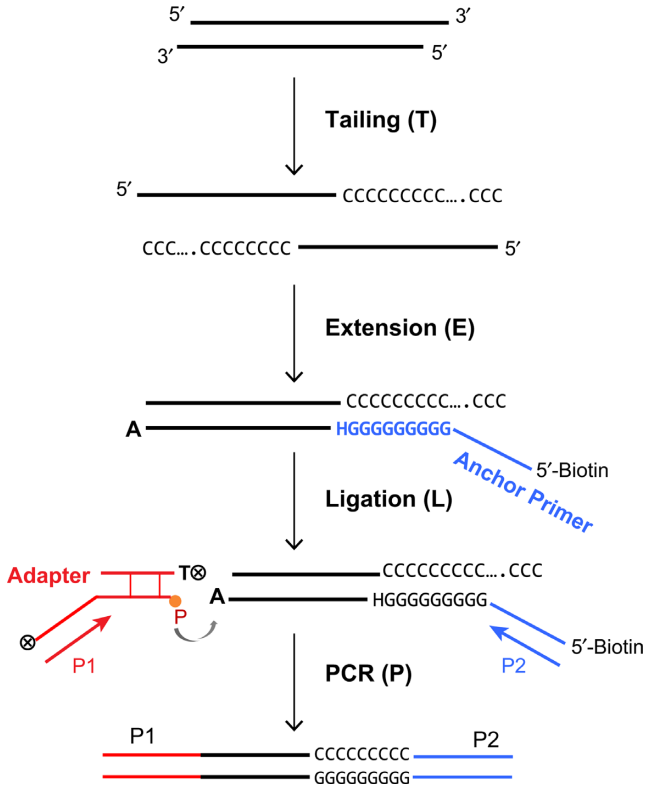

Figure 1. A schematic of the TELP sequencing library construction method. First, the tailing (T) step was performed using TDT to add a poly-C tail to the 3′-end of dsDNA or ssDNA fragments. This step was followed by the extension (E) step initiated from a biotin-labeled anchor primer annealed to the poly-C tail. Then, an adapter was added in the ligation (L) reaction to yield the templates for PCR (P) amplification. After these four simple steps, the library is ready for NGS on an Illumina platform. Details of the TELP method are described in the Materials and Methods section.

(10 mM Tris-HCl pH 8.0, 0.02% Triton X-100) and resuspended in 8.4 μl of elution buffer (EB; 10 mM Tris-HCl pH 8.0) in preparation for adapter ligation.

Ligation adapters were prepared by annealing oligo Adp_A and Adp_B together. The 3′-end of both oligos was blocked by a phosphate group to prevent self-ligation. Adapter ligation was performed on beads in 20 μl of reaction containing 1 μl of Quick ligase (NEB, M2200L), 10 μl of 2× Quick ligation buffer, 8.4 μl of the suspension of beads and extension products and 0.6 μl of 10 mM adapter. The ligation products were then washed once with 100 μl of 1× B&W buffer and three times with 150 μl of EBT buffer and recovered by elution per the manufacturer's instructions.

Final polymerase chain reaction (PCR) amplification was performed on ligation products to yield libraries ready for sequencing on the Illumina platforms. First-round PCR was performed in a 40-μl reaction containing 30 μl of ligation products, 0.4 μl of EX Taq HS (Takara, RR006A), 4 μl of 10x EX buffer, 3.2 μl of 2.5 mM dNTP, 1.2 μl of 20 μM P1_L primer and 1.2 μl of 20 μM P2M_G5 primer. The PCR program was as follows: (i) 95°C for 3 min; (ii) 95°C for 30 s, 60°C for 30 s, 72°C for 2 min, 14–19 cycles and (iii) 72°C for 7 min. The PCR products were purified using the MinElute PCR purification kit and eluted in 25 μl of EB.

For multiplex sequencing, index sequences were added by the second round PCR in a 20 μl reaction containing 4 μl of the first round PCR product, 10.2 μl of H_2O, 0.2 μl of EX Taq HS, 2 μl of 10× EX buffer, 1.6 μl of 2.5 mM dNTP, 1 μl of 20 μM P1 and 1 μl of 20 μM index primers. The PCR program was as follows: (i) 95°C for 3 min; (ii) 95°C for 30 s, 55°C for 30 s, 72°C for 2 min, 5–7 cycles and (iii) 72°C for 7 min. After PCR amplification, the fragments were subjected to size selection on a 2% ultra-pure agarose gel (Invitrogen, 16500-500) to recover DNA fragments 200–500 bp in length using a Gel Extraction kit (Qiagen, 28706). Purified DNA libraries were eluted in 30 μl of EB and quantified by Qubit in preparation for NGS on an Illumina GAII or HiSeq2000 platform.

Construction of mimic sequencing library from dsDNA and ssDNA

Phi X174 DNA was used as a template to generate a 294-bp PCR fragment using primers P_F and P_R. A phosphate group was added to the 5′ end of the P_R primer to facilitate Lambda exonuclease digestion and thus produce ssDNA (+ strand). One nanogram of purified ssDNA or the original double-stranded PCR product was amplified using the TELP protocol. The mimic sequencing libraries prepared from ssDNA and dsDNA were then cloned into the plasmid pTZ57R/T using an InsTAclone PCR cloning kit (Thermo, K1214). Subsequently, 13 clones from the ssDNA mimic library and 9 clones from the dsDNA mimic library were sequenced using Sanger sequencing.

RNA extraction, RT-qPCR and RNA-seq library preparation

Total RNA was extracted using TRIzol reagent (Ambion, 15596-026) from both preadipocytes and mature adipocytes during 10T1/2 cell differentiation. Expression of adipocyte marker genes *Pparg2, Fabp4* and *Cebpa* was examined using RT-qPCR as described (6). mRNA purification was performed on 3 μg of total RNA with microPoly(A) Purist Kit (Ambion, AM1919). Purified mRNA was sonicated with Bioruptor, treated with amplification-grade DNase I (Invitrogen, 18068-015) and reverse transcribed with random 9-mers (N9) and SuperScript II Reverse Transcriptase (Invitrogen, 18064-022) in the presence (Strand-specific) or absence of actinomycin D (Sigma-Aldrich, A1410) (2). Alternatively, mRNA was treated with DNase I, reverse transcribed with poly-T primer (T18) and then sonicated to produce cDNA of the proper size for sequencing library construction. Excess RT primers were removed by exonuclease I digestion. Following removal of mRNA templates by RNase A treatment (Fermentas, EN0531), cDNA was purified with the MinElute PCR purification kit and applied to sequencing library construction using the TELP protocol as described above. For comparison, a library was prepared from 5 μg of the same total RNA samples using standard Illumina RNA-seq library preparation, and the library was sequenced on a HiSeq2000 platform.

Sequencing library validation

ChIP-seq libraries were validated by comparing the original H3K4me3 ChIP signal to the signals from generated

libraries including the Illumina standard library (10 ng), the TELP 1 ng library, the TELP 100 pg library and the TELP 25 pg library. ChIP DNA and sequencing libraries were quantified through real-time quantitative PCR analysis using SYBR Green and a 7900HT Fast Real-Time PCR System (Applied Biosystems). ChIP-qPCR primers targeting genomic loci that were H3K4me3 positive (positive control primers PC_1–4) and negative (negative control primers NC_1–5) were designed to have melting temperatures (T_ms) near 60°C. The primer sequences for ChIP-qPCR are listed in Supplementary Table S1. The ChIP signal at the *Cebpa* promoter region (PC_1) was arbitrarily given the value of 100%, and the signals at all other regions are reported as a percentage of PC_1 to show relative enrichments.

RNA-seq libraries were validated by comparing the original cDNAs (preadipocytes and mature adipocytes of 10T1/2 differentiation) to the corresponding TELP RNA-seq libraries. The expression of the key adipogenic regulatory genes *Pparg2, Fabp4, Cebpa, CD36, Retn* and *Dlk1* was determined through real-time qPCR. The primer sequences for RT-qPCR are listed in Supplementary Table S1. All gene expression data in this study were normalized to the riboprotein gene *36B4* (9).

NGS

ChIP-seq and RNA-seq libraries prepared by TELP and Illumina standard protocol were sequenced on the Illumina GAII or HiSeq2000 platform per the manufacturers' instructions.

ChIP-seq data analysis

All ChIP-seq data sets were aligned to the mouse genome (mm9) using the Bowtie program (version 2.0.4). Alignments were performed with the following parameters in addition to the default settings: -t -q -p 8 -N 1 -L 25. We only kept the unique alignments and duplicates were removed using Picard for all the samples. Reads statistics and performance measures for all the sequencing libraries in our study were summarized in Supplementary Table S2. For the downstream analysis, we normalized the read counts for the ChIP samples by computing the numbers of reads per kilobase of bin per million reads sequenced (RPKM). To minimize the batch and cell-type variations, the RPKM values were further normalized through Z-score transformation. MACS (10) was used to identify H3K4me3 and PPARγ binding peaks. To visualize the ChIP-seq signals for each histone modification and PPARγ using the UCSC genome browser tracks, we extended each read to 300 bp and counted the read coverage for each base.

RNA-seq data analysis

All RNA-seq data sets were aligned to the mouse genome (mm9) using TopHat (version 2.0.6). Alignments were performed with the following parameters in addition to the default settings: -p 2 -solexa1.3-quals. The mapped reads were further analyzed using the Cufflinks program (11), and the expression levels for each transcript were quantified as fragments per kilobase of transcript per million mapped reads (FPKM) based on the RefSeq database.

Comparison of gene expression levels and various histone modifications in adipocytes

To show the correlation between the levels of gene expression and histone modifications at promoters, we first ranked 22,037 genes based on their expression values (FPKM). The averaged enrichment level (Z-score normalized RPKM) of each histone modification (within TSS +/− 2.5 kb) for the same gene was shown accordingly.

Comparison of the ChIP-seq signals between libraries prepared by the standard Illumina method and TELP

Scatter plots and Venn diagrams were used to show the correlation of the H3K4me3 ChIP signal between libraries prepared using the standard Illumina protocol (using 10 ng ChIP DNA) or TELP (using 1 ng, 100 pg or 25 pg ChIP DNA). The ChIP-seq enrichment (RPKM) from various samples was compared at the H3K4me3 peaks ($n = 10{,}442$) identified by MACS using the sample prepared by the standard Illumina method. The Pearson correlation coefficients were calculated and are shown. The overlap of H3K4me3 peaks between samples prepared by the Illumina and TELP methods are illustrated in Venn diagrams. The H3K4me3 peaks of 1 ng, 100 pg and 25 pg samples were also identified by MACS. As the 25 pg sample yielded significantly more peaks ($n = 28{,}314$) than other samples, we set an additional selection criteria (Z-score-normalized RPKM > 0) to filter out the weak peaks. The remaining strong H3K4me3 peaks ($n = 10{,}733$) were used for further comparison.

Comparison of the RNA-seq signals in the Illumina and TELP libraries

Scatter plots and a Venn diagram were used to show the correlation between the RNA-seq signal of preadipocytes and mature adipocytes prepared by the Illumina and the TELP methods. Correlation of FPKM between different samples was calculated, and the Pearson correlation coefficients are shown. Activated genes during adipogenesis were defined as genes that showed high levels of expression (FPKM > 5) in mature adipocyte and at least a 2-fold upregulation in mature adipocytes compared with preadipocytes. The overlap of activated genes identified from samples prepared by the Illumina and the TELP methods is indicated in the Venn diagram.

Comparison of PPARγ binding profiles between primary and 10T1/2 adipocytes

PPARγ binding peaks ($n = 10{,}940$) in 10T1/2-derived mature adipocytes were identified by MACS with default settings over the background signal (IgG ChIP-seq). The corresponding PPARγ binding profiles in eWAT- and iWAT-derived adipocytes have been described previously (12). Venn diagrams were used to show the overlaps between PPARγ binding peaks in 10T1/2 adipocytes and eWAT or iWAT adipocytes.

RESULTS

Development of the TELP sequencing library construction method

ChIP-seq has been widely used to profile genome-wide chromatin modification dynamics and transcription factor binding. RNA-seq is also becoming a standard tool for comprehensive transcriptome analysis. We have developed a versatile sequencing-library construction method that is compatible with both ssDNA (e.g. cDNA) and dsDNA (e.g. ChIP DNA), which we have termed TELP. Libraries prepared by TELP can be directly sequenced on the short-read sequencing platforms provided by Illumina.

In the initial step of TELP, a poly-C tail was added to the 3′-end of ssDNA or dsDNA using TDT (Figure 1). We first examined the effects of buffer, ion conditions and reaction times on the tailing efficiency (Supplementary Figure S1A). We found that the tailing reaction was optimal when TDT stock buffer was used with a Co^{++} supplement (Lane 4 in Panel 1, Supplementary Figure S1B). Then, we evaluated the influence of the length of the poly-C tail on the final yield of the extension product. Surprisingly, we found that an excess length of poly-C tail had an adverse effect on the next extension reaction as indicated by the weaker PCR band from the extension product (Lane 4 in Panel 2, Supplementary Figure S1B). In contrast, shorter poly-C tails promote the extension reaction (Lanes 1 and 3 in Panel 2, Supplementary Figure S1B). Moreover, when using a minute amount of starting material (300 or 30 pg DNA), TDT tailing using Ex-Taq buffer provided the highest efficiency in the extension reaction (Lane 1, Supplementary Figure S1C). We thus decided to use TDT plus Ex-Taq buffer for poly-C tailing. To estimate the length of the poly-C tail added under these conditions, a shorter PCR fragment (175 bp) was used in the tailing reaction. As shown in Supplementary Figure S1D, ∼15–20 cytosines were added to the 3′-end of this PCR fragment.

Next, we compared the extension efficiencies of several commercially available DNA polymerases. As shown in Supplementary Figure S1E, KAPA-2G had the highest efficiency in extension reactions as judged by qPCR of extension products. Therefore, KAPA-2G was chosen for the extension step. In this step, we used a biotin-labeled anchor primer (Figure 1) to facilitate downstream purification of extension products. Excess anchor primer was removed by exonuclease I (Exo I) digestion after extension to prevent it from ligating to the adapter added later in the ligation step. This step is essential for the proper amplification of the sequencing library (Supplementary Figure S1F, right panel). When Exo I treatment was omitted, the anchor primer-adapter dimer dominated the PCR reaction, leading to the failure of library preparation (Supplementary Figure S1F, left panel).

After Exo I digestion, extension products were captured by magnetic streptavidin beads, and adapter ligation was performed on the beads. To prevent adapter self-ligation, the 3′ ends of both oligo Adp_A and Adp_B were blocked by a phosphate group. The ligation reaction was performed at 4°C overnight before the products were eluted to yield templates for final PCR amplification (Figure 1). For multiplex sequencing, index sequences were added in the PCR step to generate libraries ready for sequencing on the Illumina platforms (Supplementary Figure S1G, left panel). It should be noted that when the starting material was minimal (e.g. 25 pg), we observed significant byproduct formation after PCR amplification (Supplementary Figure S1G, right panel); however, the sequencing library was not significantly affected after size selection on an agarose gel, as judged by the sequencing results (detailed below).

High-quality ChIP-seq data sets generated by TELP in murine adipocytes

After establishing the TELP protocol, we sought to validate it in murine adipocytes. To obtain mature adipocytes, murine preadipocytes 3T3-L1 and MSCs 10T1/2 were differentiated *in vitro*. After 7 days of adipogenic induction, these cells become lipid-laden fat cells as judged by Oil-Red-O staining (Supplementary Figure S2A). In addition, significant adipocyte marker genes such as *Pparg2*, *Cebpa* and *Fabp4* were dramatically upregulated (Supplementary Figure S2B), indicating high efficiency adipogenesis.

Histone modifications are fundamentally involved in gene regulation (13–15). Depending on the modification site and type, these chromatin markers can be signatures of active promoters (H3K4me3), gene bodies (H3K36me3) or enhancer elements (H3K27ac) (16). We performed ChIP assays using antibodies specific to histone H3K4me3, K9ac, K27ac and K27me3 in mature adipocytes derived from 10T1/2 cells. The quantity of ChIP DNA obtained from these assays was at the nanogram level as determined using a Qubit fluorometer. Using TELP, we prepared sequencing libraries from these ChIP DNAs and sequenced them on the Illumina HiSeq2000 platform. In parallel, we performed RNA-seq using the standard Illumina protocol to profile genome-wide gene expression levels in the same cells. Figure 2A shows the chromatin state as well as the mRNA expression profile near the murine *Pparg* gene, which is highly expressed in mature adipocytes.

Consistent with the known functions for the profiled four chromatin modifications, we found the following: (i) H3K9ac and H3K4me3 were preferentially enriched at known promoters, consistent with previous findings that they are involved in gene activation as well as transcriptional initiation (13,17); (ii) H3K27ac was primarily localized to promoters, intronic and intergenic regions, in agreement with their roles as markers of *cis*-regulatory elements (18–20) and (iii) H3K27me3, a repressive chromatin marker (14), was largely absent from the actively transcribed *Pparg* gene and distributed broadly across the inactive flanking regions. Moreover, our epigenomic profiles generated by TELP were highly consistent with the profiles generated by the Illumina standard protocol in 3T3-L1 cells at the *Pparg* locus (7) (Supplementary Figure S3), demonstrating that these two techniques performed similarly in ChIP-seq analysis. As a negative control, sequencing reads from input DNA were evenly distributed across the genome (Figure 2A). After obtaining the epigenomic maps of these four histone modifications in mature adipocytes, we compared our ChIP-seq data sets to the RNA-seq data set at the genome-wide level. In line with previous knowledge,

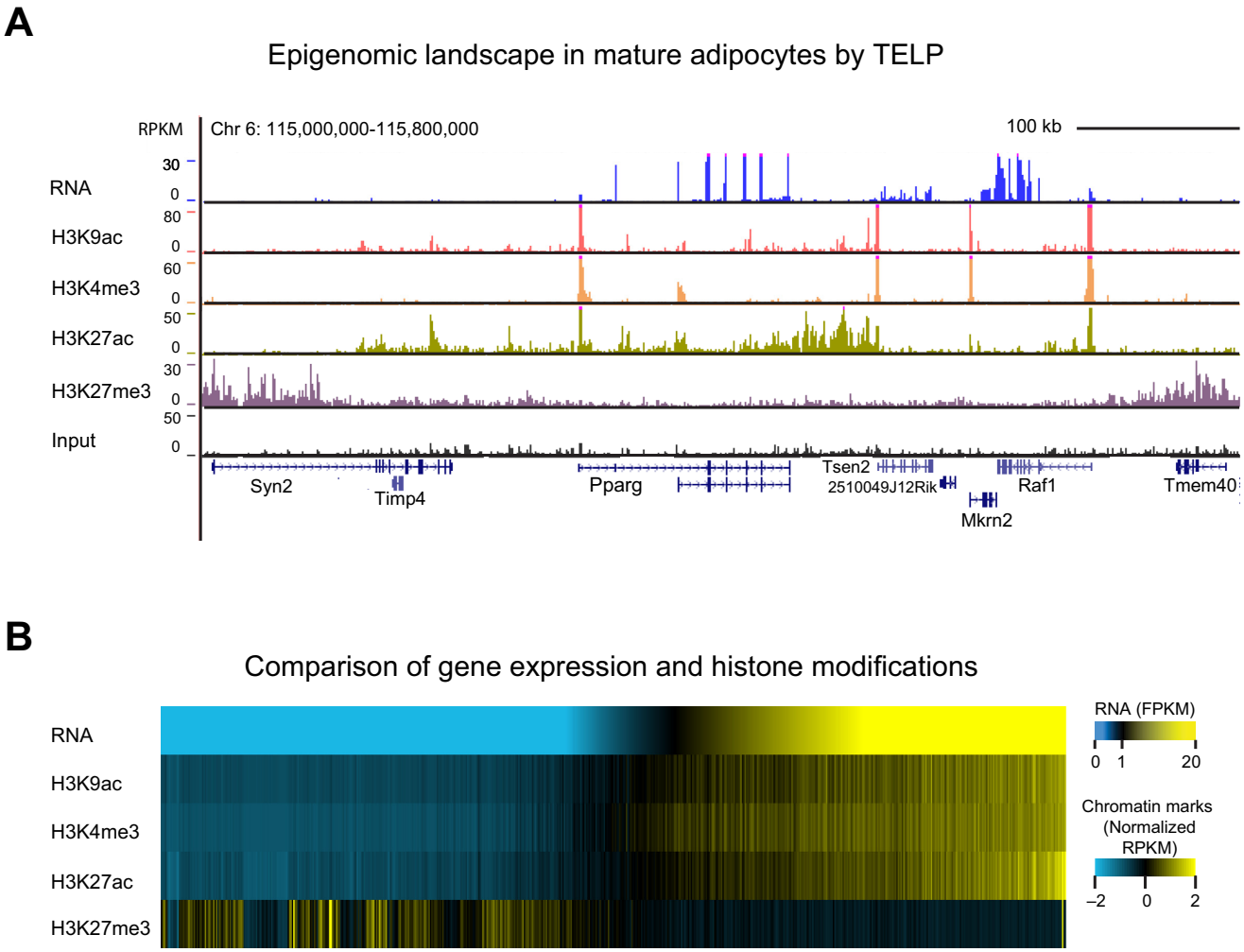

Figure 2. High-quality epigenomic maps generated by TELP in mature adipocytes. (**A**) Epigenomic maps of histone H3 K9 acetylation (ac), K4 trimethylation (me3), K27ac and K27me3 in mature adipocytes derived from 10T1/2 cells. The RNA-seq data generated by the Illumina standard protocol is shown in parallel to illustrate the gene expression levels. The data from input DNA was included as a negative control. (**B**) A heat map is shown for the expression levels of genes in adipocytes ranked from the lowest expression (left) to the highest expression (right). The levels of various histone markers at promoters (averaged Z-score normalized RPKM values within TSS +/− 2.5 kb) for the same genes are shown in parallel.

H3K9ac, H3K27ac and H3K4me3 show strong positive correlations with the transcriptional status of corresponding genes, while H3K27me3 shows a negative correlation with gene activity (Figure 2B).

In addition to chromatin modification profiling, TELP was used to generate the genome-wide binding map of the adipogenic master regulator PPARγ in mature adipocytes. As shown in Supplementary Figure S4A, PPARγ binds to the same upstream and promoter regions of the *LPL* gene in mature adipocytes derived from 10T1/2 cells as it does in progenitor cells isolated from epididymal and inguinal white adipose tissue (eWAT and iWAT) (12). At the genome-wide level, our quantitative analyses revealed that the majority of PPARγ binding peaks in 10T1/2 adipocytes were present in eWAT- (8270 out of 10,940 binding peaks, ∼76%) and iWAT- (6847 out of 10,940 binding peaks, ∼63%) derived adipocytes (Supplementary Figure S4B). However, large portions of the PPARγ binding sites in eWAT- (27,913 out of 36,183 binding sites, ∼77%) and iWAT- (8336 out of 15,183 binding sites, ∼55%) derived adipocytes were not found in 10T1/2 adipocytes. Given that similar observations have been reported in the comparison between primary and 3T3-L1 adipocytes (12), we conclude that the PPARγ binding differences between 10T1/2 adipocytes and eWAT or iWAT adipocytes mainly reflect functional differences between cell types but not differences in ChIP-seq methodologies. Together, these data demonstrate that the epigenomic maps generated in adipocytes using TELP are of high quality and are consistent with previously published results.

Highly sensitive sequencing library preparation from 25 pg of ChIP DNA using TELP

ChIP-seq is a powerful tool for the genome-wide mapping of chromatin modifications and transcription factor binding. However, the application of this technique has been hindered by the requirement of a substantial quantity of starting materials (usually at the nanogram level) for the

preparation of sequencing libraries (1). In the standard library construction method, a significant amount of DNA is lost due to multiple steps of DNA purification (column and gel purification) that occur before PCR amplification. In TELP, we minimized the number of DNA purification steps to avoid such sample loss, which should significantly increase its sensitivity with low amounts of ChIP DNA. To test this hypothesis, we used H3K4me3 ChIP DNA from mature 3T3-L1 adipocytes and quantified it with the Qubit fluorometer (Invitrogen). We then serially diluted the DNA to obtain starting DNA amounts of 1 ng, 100 pg and 25 pg for TELP library preparation. In parallel, 10 ng of the same ChIP DNA was used as starting material for the standard Illumina protocol for comparison.

We first examined the fidelity of the sequencing libraries by qPCR. To this end, we selected nine genomic regions including both H3K4me3-positive and -negative loci based on a previously published ChIP-seq data set (7). Then, we determined the relative enrichments of H3K4me3 at these genomic sites in all four sequencing libraries and compared the results to the original ChIP data. As shown in Figure 3A, H3K4me3 enrichments were faithfully represented in the standard Illumina library (Illumina 10 ng) and the TELP library from 1 ng DNA (TELP 1 ng) compared to the original ChIP qPCR results, while moderate variations were observed in TELP libraries prepared from low amounts of ChIP DNA (TELP 100 and 25 pg).

Next, we sequenced these libraries on the Illumina sequencer and aligned the sequencing reads to the mouse reference genome. Visual analysis of the ChIP-seq results from the UCSC browser suggested high concordance among these H3K4me3 maps (Figure 3B), indicating successful preparation of sequencing libraries from as little as 25 pg ChIP DNA by TELP. In a more quantitative analysis, we determined the correlations of H3K4me3 signal intensities from the standard Illumina library and from those generated by TELP using various amounts of ChIP DNA. High Pearson correlation coefficients of 0.95, 0.89 and 0.88 were observed in the comparison between Illumina 10 ng library and TELP 1 ng, 100 pg and 25 pg libraries, respectively (Figure 3C). Moreover, an additional quantitative comparison revealed that at the genome-wide level, there were 96.3, 96.6 and 83.3% overlaps between H3K4me3-marked promoters identified in the Illumina 10 ng library and TELP 1 ng, 100 pg and 25 pg libraries, respectively (Figure 3D). In an attempt to address the reproducibility issue, we performed TELP library construction and sequencing in triplicate using minute amounts of ChIP DNA at 100 pg and 25 pg levels. As shown in Supplementary Figure S5A, all TELP ChIP-seq experiments produced highly consistent H3K4me3 profiles across a 150-kb region on chromosome 6. At the genome-wide level, high correlation of H3K4me3 signal intensities among the three TELP 100 pg libraries was observed ($R = 0.87$), and the overlaps of H3K4me3 peaks were between 89.9% and 95.4% (Supplementary Figure S5B and SC). Similarly, high Pearson correlation coefficients of 0.80, 0.79 and 0.76 were observed in the pair-wise comparison among the three TELP 25 pg libraries, and the overlaps of H3K4me3 peaks were between 85.2% and 91.9% (Supplementary Figure S5D and SE). In summary, these data demonstrated that TELP is highly sensitive for preparing libraries from low amounts of DNA.

TELP library preparation is compatible with paired-end sequencing

Paired-end sequencing is advantageous over single-end sequencing because it reduces ambiguity in the generated reads and thus accurately identifies transcription factor binding or chromatin modification enrichment sites. As shown in the structure of the TELP multiplex sequencing library (Figure 4A), the first-end sequencing was performed using PE1; then, the index sequence was determined using primer P_code. The second-end sequencing was achieved using primer G9_PE2. We sequenced two TELP H3K4me3 ChIP-seq libraries (1 ng and 100 pg) from both ends and found virtually identical patterns of this modification from the data based on the first- and second-end sequencing (Figure 4B).

High fidelity RNA-seq library preparation using TELP

After successfully utilizing TELP to prepare ChIP-seq libraries, we extended its application to cDNA sequencing (RNA-seq). We first prepared cDNA through reverse transcription of the mRNA purified from both preadipocytes and mature adipocytes derived from 10T1/2 cells. Subsequently, we constructed RNA-seq libraries from these cDNAs by TELP. To validate the TELP RNA-seq libraries, we selected six regulatory genes of adipogenesis and examined their expression levels through qPCR. As shown in Supplementary Figure S6, for original cDNAs and RNA-seq libraries prepared by TELP, we observed the same dynamic trends for the examined genes between preadipocytes and mature adipocytes. Quantitatively, there were moderate variations between RNA-seq libraries and the original cDNAs, which did not impose a significant impact on the quality of RNA sequencing results (see below).

We then sequenced the TELP RNA-seq libraries and also performed Illumina standard RNA-seq in parallel for comparison. Figure 5A shows the gene expression profiles generated by the standard Illumina protocol or TELP in preadipocytes and mature adipocytes for the following three genes: (i) a constantly expressed house-keeping gene, *36B4*; (ii) an upregulated gene, *RETN* and (iii) a downregulated gene, *DLK1*. The expression patterns for these three genes were very similar between standard Illumina and TELP RNA-seq, suggesting the high quality of TELP libraries. At the genome-wide level, quantitative analysis revealed that the correlations of signal intensities from standard Illumina RNA-seq and TELP RNA-seq were 0.94 and 0.93 (Pearson correlation coefficients) for preadipocytes and mature adipocytes, respectively (Figure 5B).

In an effort to further demonstrate the consistency of standard Illumina RNA-seq and TELP RNA-seq, we calculated the genes upregulated by at least 2-fold in mature adipocytes compared to preadipocytes based on the RNA-seq data. In total, we identified 886 genes that meet the criteria from the standard Illumina RNA-seq data and 869 genes from the TELP RNA-seq data. Among them, 716 genes (80.8%) were common to both data sets, indicating

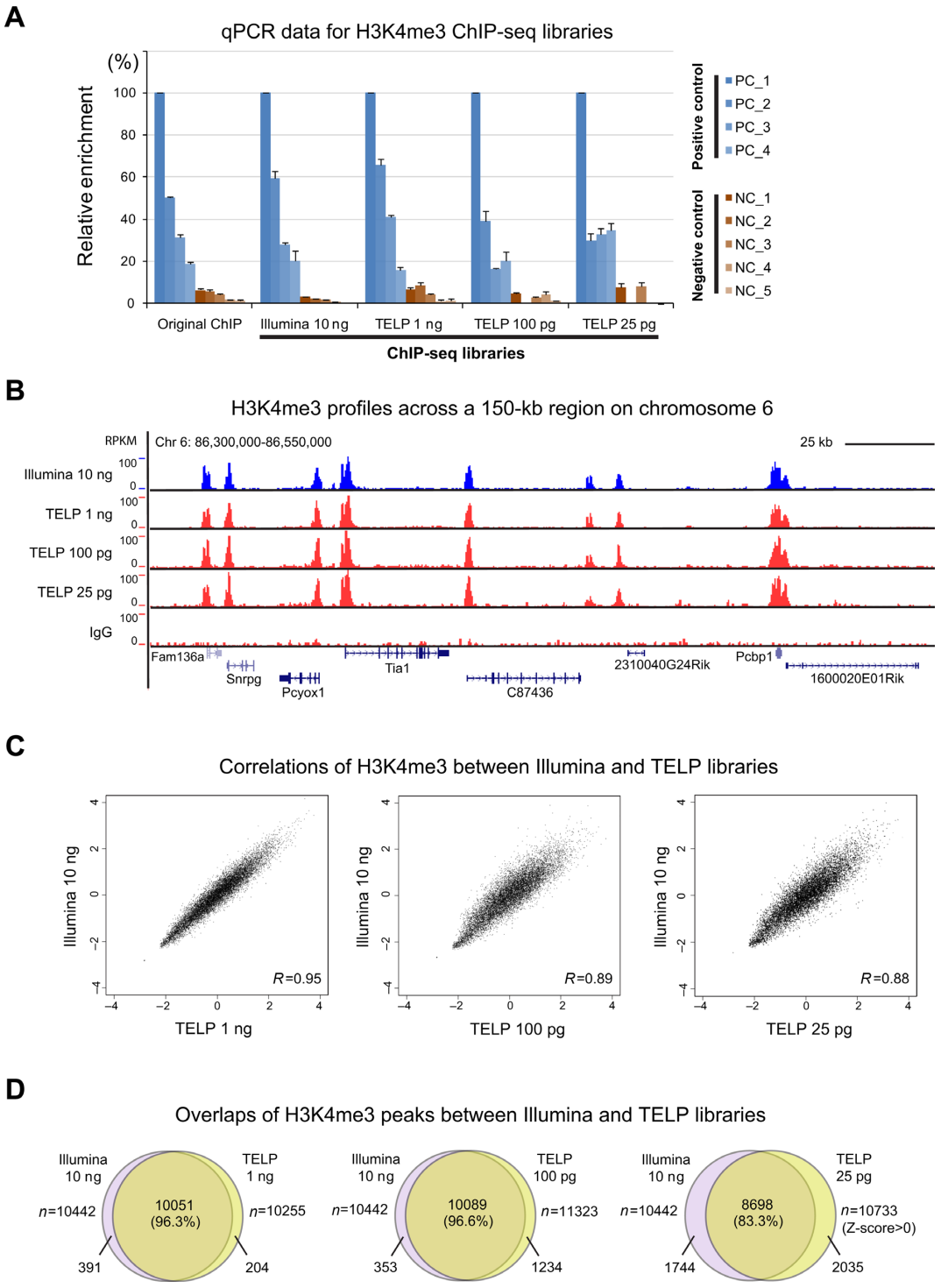

Figure 3. Highly sensitive preparation of a sequencing library from 25 pg of ChIP DNA using TELP. (**A**) Validation of the sequencing libraries by qPCR. Histone H3K4me3 levels at nine genomic regions were determined through qPCR for sequencing libraries prepared using either the standard Illumina protocol with 10 ng ChIP DNA (Illumina 10 ng) or using TELP with the indicated amounts of ChIP DNA. The data were normalized to the promoter of the *Cebpa* gene (positive control, PC_1). The details of the qPCR primers are given in the Materials and Methods section. These results are the averages of three independent qPCR experiments, and the error bars indicate standard deviations. (**B**) A comparison of the H3K4me3 ChIP-seq signals from 10 ng ChIP DNA using the standard Illumina library construction method with TELP ChIP-seq analyses from 1 ng, 100 pg and 25 pg ChIP DNA. The data from a similar ChIP experiment using rabbit IgG were included as a negative control. (**C**) Scatter plots showing the comparison of the H3K4me3 levels (averaged RPKM signal intensities at the called H3K4me3 peaks, based on Illumina 10 ng) between those generated by the standard Illumina protocol and those generated by TELP using various amount of ChIP DNA. Pearson correlation coefficients, *R*, are indicated. (**D**) Venn diagrams showing overlaps between H3K4me3-marked regions identified by Illumina ChIP-seq (10 ng DNA) and TELP ChIP-seq (1 ng, 100 pg and 25 pg DNA). As significantly more peaks were called from the 25 pg DNA library due to elevated background noise ($n = 28{,}314$), we selected those strong H3K4me3 peaks ($n = 10{,}733$) by setting an additional selection criteria (Z-score normalized RPKM > 0) for this analysis.

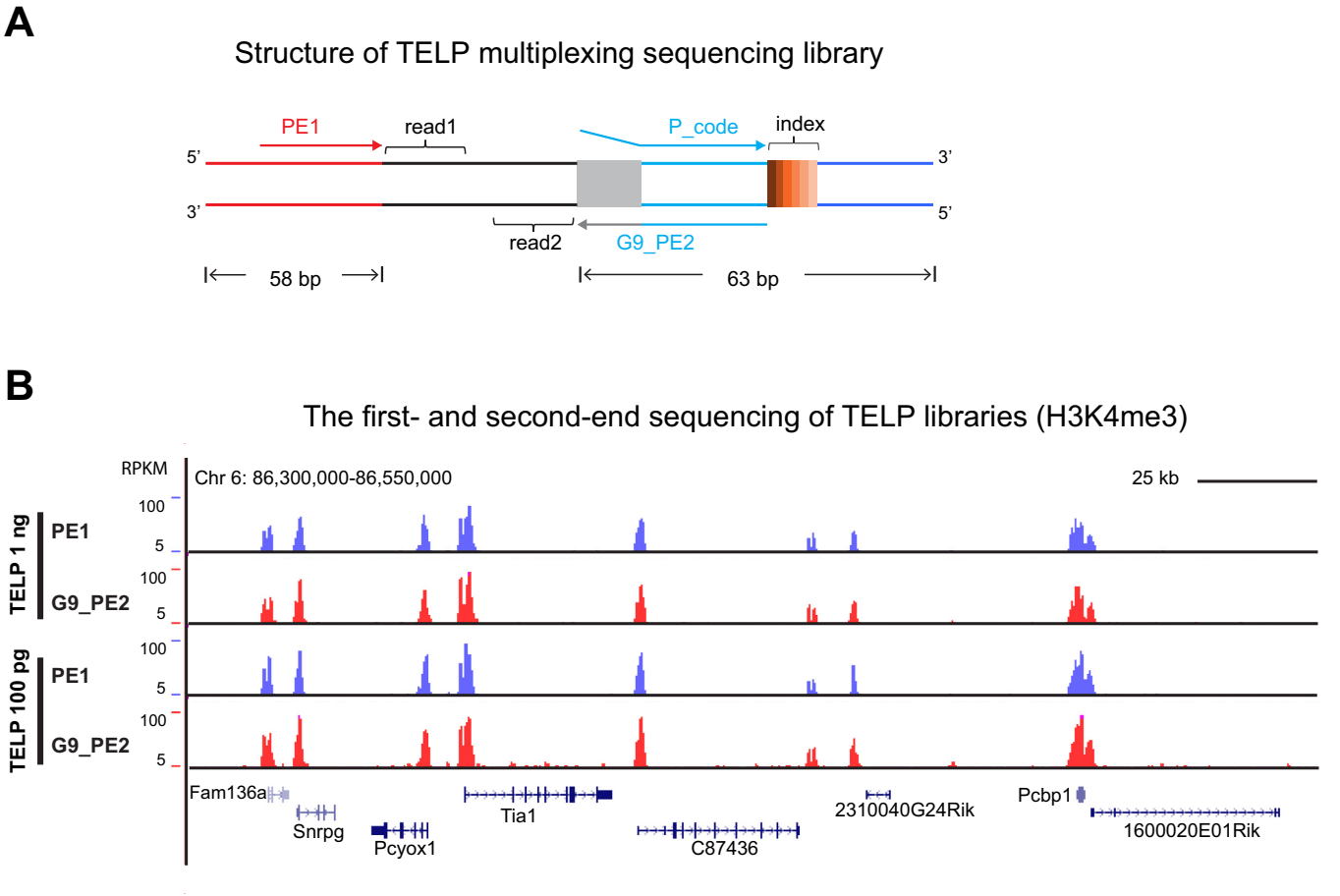

Figure 4. TELP library preparation is compatible with paired-end sequencing. (**A**) The structure of the multiplex sequencing library generated by TELP and the directions of the first- and second-end sequencing. The gray box represents consecutive G:C pairs, and the P_code primer was used to sequence the index. (**B**) Highly consistent H3K4me3 ChIP-seq signals from the first- and second-end sequencing of TELP libraries.

the majority of upregulated genes can be identified by either method (Figure 5C).

Strand-specific RNA sequencing achieved by TELP

Strand-specific RNA sequencing can add substantial value to the RNA-seq data sets, especially in applications such as the identification of regulatory antisense transcripts (21). However, many RNA-seq library construction methods do not preserve information regarding which strand was transcribed. This is mainly due to the synthesis of double-stranded cDNA using random primers. Existing strand-specific RNA-seq protocols involve extra steps such as selective labeling of the cDNA strand with dUTP (22) or bisulfite conversion of cytosines to uracils in RNA (23). As TELP is compatible with library preparation from both ssDNA and dsDNA, we tested whether it could maintain the strand specificity in RNA-seq. As an initial test, we constructed mimic sequencing libraries from both dsDNA and ssDNA using TELP (Figure 6A). After addition of the anchor primer and adapter sequences, the resulting mimic libraries were 92 bp longer than the original DNA fragments (Figure 6B). Both ssDNA and dsDNA mimic libraries were cloned and subsequently sequenced by Sanger sequencing. The sequencing results showed that all 13 clones from the ssDNA mimic library contain the same positive strand sequence. In contrast, nine clones from the dsDNA mimic library contain either the positive or the negative strand sequence at a random rate (four versus five). We conclude that the strand-specific information was faithfully maintained by TELP in these mimic libraries (Figure 6C).

Next, we tested TELP in constructing real strand-specific RNA-seq libraries. Poly-A RNA was reverse transcribed using a poly-T primer (T18) to generate the first-strand cDNA. Then, the first-strand cDNA was sonicated with a Bioruptor to obtain ssDNA fragments of the proper length. Alternatively, the first-strand cDNA was generated using a random N9 primer in the presence of actinomycin D (2). The sequencing library was prepared from these cDNA fragments by TELP. Finally, the cDNA library was sequenced on an Illumina platform (Supplementary Figure S7A). Sequencing results show that strand specificity was preserved by TELP for individual genes (Supplementary Figure S7B) and at the genome-wide level (Figure 6D). Together, these data clearly demonstrate that TELP preserves strand information for both DNA and RNA samples.

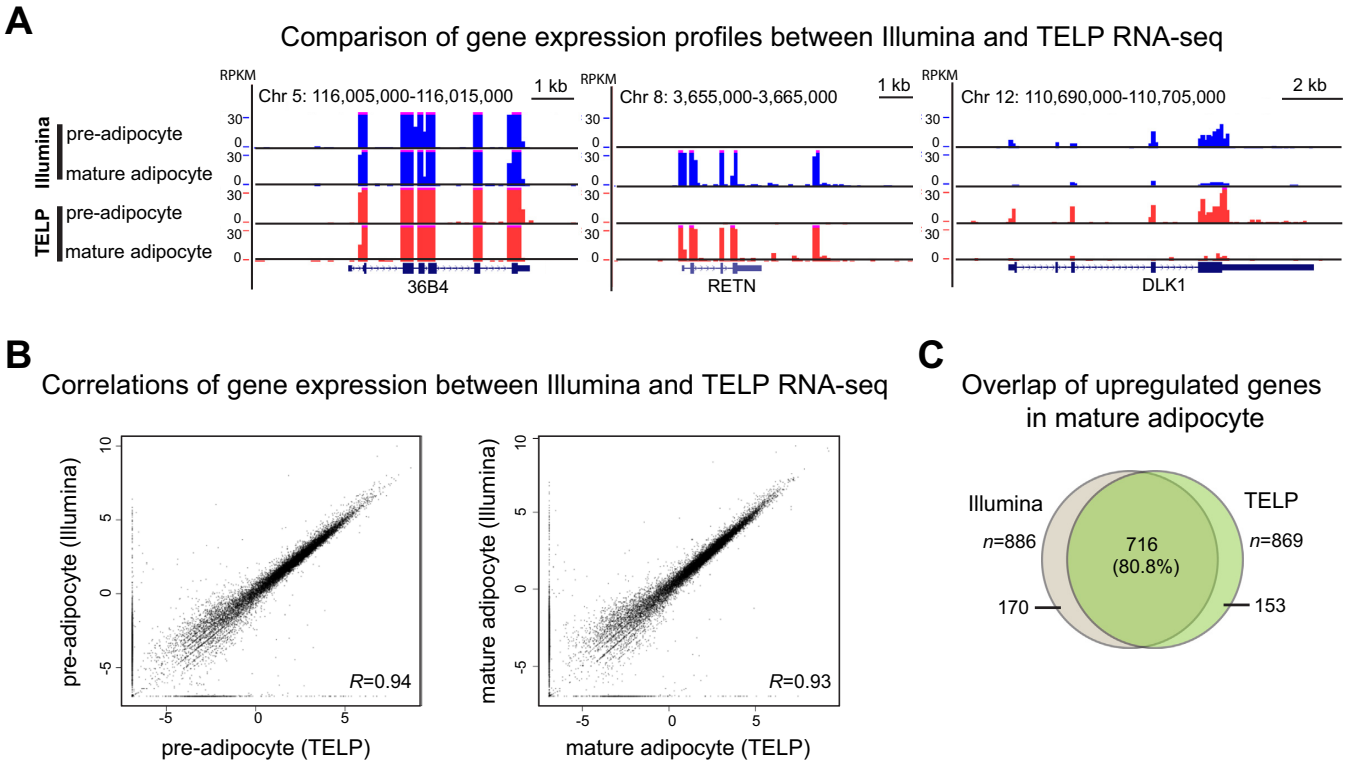

Figure 5. High fidelity RNA-seq library preparation by TELP. (**A**) A comparison of gene expression profiles generated by the standard Illumina protocol or TELP in preadipocytes and mature adipocytes. (**B**) The quantitative comparison of signal intensities of the RNA-seq data generated by the standard Illumina protocol or TELP. Pearson correlation coefficients, *R*, are indicated. (**C**) Overlap of upregulated genes in mature adipocytes determined by the standard Illumina protocol or TELP. This result is based on a single experiment.

DISCUSSION

NGS has a broad spectrum of applications in virtually all branches of biological research. Therefore, novel methodologies pushing the boundary of its usage are highly desired. Here, we report the development of a novel method of sequencing library construction that we have termed TELP. TELP is versatile, and our data show that it is compatible to both ChIP-seq and RNA-seq. Theoretically, TELP can be applied to any dsDNA or ssDNA library preparation. In addition, we showed that TELP is highly sensitive in generating epigenomic maps of H3K4me3 from minute amounts of ChIP DNA (25 pg). Given its versatility and sensitivity, we anticipate that TELP will be useful in many areas of genomic and epigenomic research.

During the course of method development of TELP, we systematically examined the effects of buffers, enzymes and reaction times on each step and also on the final outcome. Surprisingly, optimal conditions for one step may not necessarily benefit the whole protocol. For example, poly-C tailing was most efficient when TDT stock buffer was used with the supplement of Co^{++} (Supplementary Figure S1B). However, the excess length of the poly-C tail had an adverse effect on the next extension reaction. This result is likely due to the increased failure of proper annealing of the anchor primer to the beginning of the poly-C tail, which resulted in the mismatch of the last nucleotide (H) on the anchor primer and inhibition of the extension reaction. Further analysis showed that the extension reaction is most efficient when the poly-C tail is ∼15–20 nucleotides in length (Supplementary Figure S1B–D).

After extension, we used magnetic streptavidin beads to capture the products and performed adapter ligation on the beads. This reduced the DNA purification steps, which cause major DNA loss. Similar approaches have been adopted in other sequencing protocols for the same purpose (24,25). In the final PCR amplification step, we used a two-step protocol in which a pre-library was prepared in the first step, and the index sequences were added for multiplex sequencing in the second step. This process can be easily substituted with a one-step PCR protocol using a longer primer containing the index sequence (data not shown).

After establishing the TELP protocol, we first applied it to ChIP-seq library construction and generated high-quality epigenomic maps of H3K4me3, H3K9ac, H3K27ac and H3K27me3 in murine adipocytes. As expected, H3K9ac and H3K4me3 were found at active promoters, and H3K27ac was found at intronic and intergenic regions, while H3K27me3 was mainly found at inactive genomic regions (Figure 2A). Moreover, our ChIP-seq data set showed good consistency with a published data set obtained by standard Illumina sequencing in another model of murine adipocytes, 3T3-L1 cells (7) (Supplementary Figure S3). In parallel, we profiled the genome-wide binding of the master regulator of adipogenesis, PPARγ. Our results revealed PPARγ binding at most of its target genes (Supplementary Figure S4), again indicating the fidelity of the

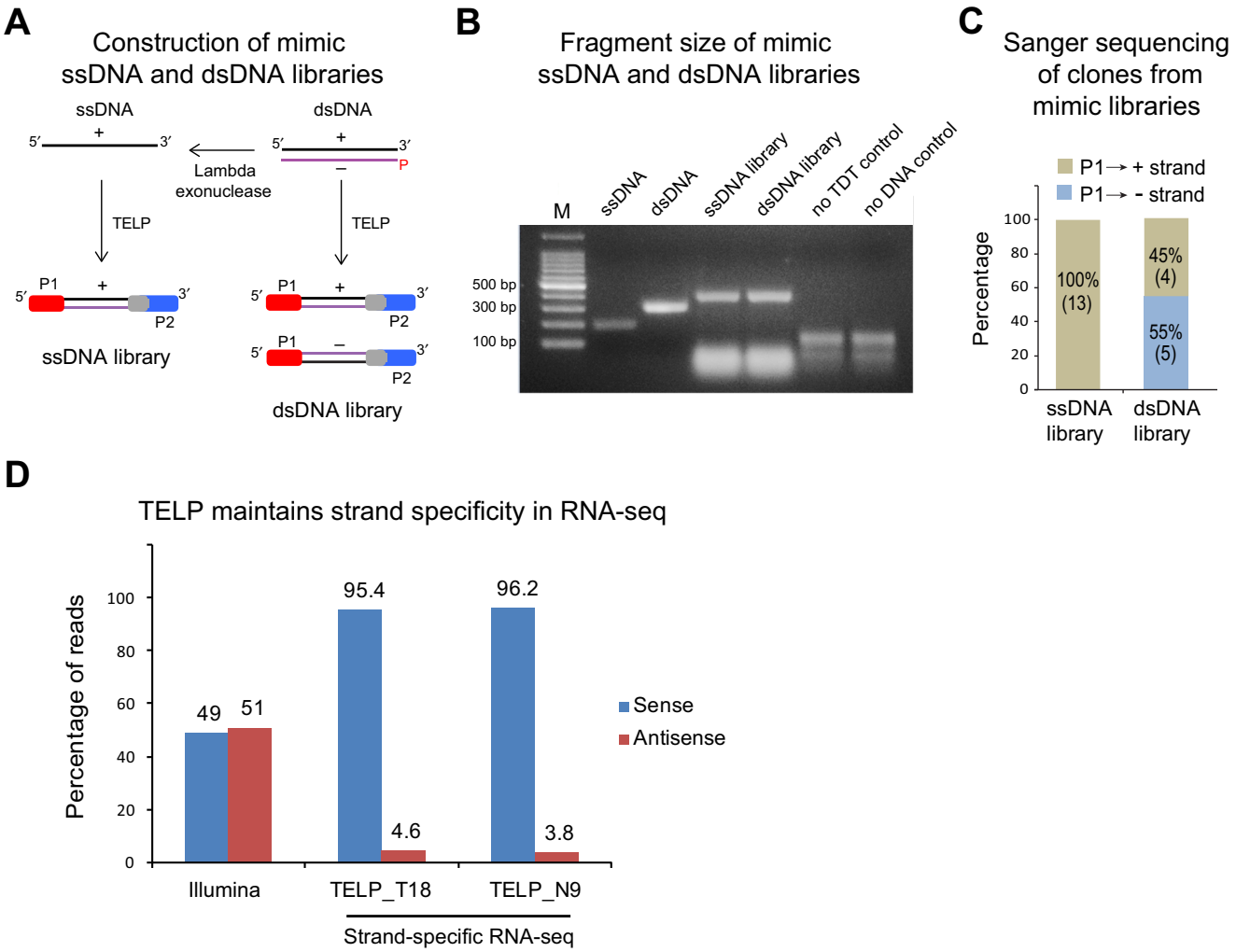

Figure 6. Strand-specific RNA sequencing achieved by TELP. TELP is compatible with both ssDNA and dsDNA amplifications. (**A**) A 294-bp PCR fragment was used as an example of dsDNA. This fragment was digested by Lambda exonuclease to generate an ssDNA fragment (+ strand). Then, mimic sequencing libraries were prepared from ssDNA and dsDNA using TELP. (**B**) ssDNA, dsDNA and the mimic libraries prepared from them were analyzed on an agarose gel. ssDNA migrates faster than dsDNA of the same length. Mimic libraries are 92 bp longer than original DNA fragments after the addition of the anchor primer and adapter sequences. The lower bands in these two lanes are excess PCR primers. Reactions without TDT or DNA template were included as negative controls; the bands in these two lanes are byproducts from adapter-anchor primer ligation. (**C**) Mimic sequencing libraries prepared from ssDNA and dsDNA were cloned and subsequently sequenced through Sanger sequencing. A total of 13 clones from the ssDNA mimic library all contained the positive strand sequence, while nine clones from the dsDNA mimic library contained either the positive or the negative strand sequence at a random rate (four versus five), showing that strand specificity was preserved in TELP library construction. (**D**) Genome-wide fidelity of strand specificity for RNA sequencing using the TELP protocol in mature adipocytes.

TELP method. In an attempt to examine TELP's sensitivity, we successfully prepared a sequencing library from as little as 25 pg ChIP DNA (Figure 3). This is, to our knowledge, one of the highest sensitivities that have been achieved to date (26–28) in ChIP-seq. Paired-end sequencing is another desirable feature of NGS, and we clearly showed that TELP libraries are compatible with this technique (Figure 4).

In addition to sensitive and faithful ChIP-seq library preparation, we also generated transcriptomic profiles for both preadipocytes and mature adipocytes using TELP. Our data sets showed very high consistency with the RNA-seq data sets obtained using the standard Illumina protocol (Figure 5). Moreover, TELP preserved strand specificity in RNA-seq, a feature that is useful in the identification of regulatory antisense transcripts, determining the boundaries of adjacent genes transcribed from opposite strands and quantification of overlapping transcripts (2).

Preserving end information of original DNA is another feature of TELP (Figure 1). Because end repair is only required for library construction from small amounts of DNA fragmented by mechanical shearing, in most cases, the original DNA ends will be intact during library construction. This feature is especially useful in the identification of DNA damage hotspots and DNA replication stalling sites *in vivo* (29).

As detailed in the Materials and Methods section, TELP only involves one optional column purification (after end repair for small amounts of DNA fragmented by mechanical shearing) and one binding step with magnetic beads before the final PCR amplification. This is in contrast to

the three column-purification steps and one agarose gel-purification step in the standard Illumina protocol. The minimized number of DNA purification steps not only significantly decreases sample loss but also eliminates the risk of DNA cross contamination during DNA gel purification. As TELP libraries are already indexed when applied to agarose gel electrophoresis for size selection, we can even pool multiple libraries together and purify them simultaneously. Considering the dramatically increasing capabilities of Illumina sequencers, multiplex sequencing will become prevalent in the future. Toward this end, TELP will be more efficient than standard methods in handling large numbers of samples.

In summary, we have shown that TELP is a sensitive and versatile library construction method for NGS. Given that the cost of sequencing continues to decline, establishing such a protocol in-house could substantially lower the cost for both ChIP-seq and RNA-seq. To apply this method to rare cell types and limited clinical samples, future studies on combining TELP with a robust protocol for small cell-number ChIP assays (26–28,30–33) are required. In principle, this method could also be useful in transcriptomic studies from limited amounts of starting materials.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to the members of the Feng Xu and Wei Xie laboratories for their critical comments and discussion throughout this work. We also thank Dr Judith L. Swain for her helpful advice and critical reading of the manuscript.

FUNDING

Intramural funding from the Agency for Science, Technology and Research (A*STAR) of Singapore (to F.X.). Funding for open access charge: A*STAR of Singapore.
Conflict of interest statement. None declared.

REFERENCES

1. Park,P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **10**, 669–680.
2. Levin,J.Z., Yassour,M., Adiconis,X., Nusbaum,C., Thompson,D.A., Friedman,N., Gnirke,A. and Regev,A. (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods*, **7**, 709–715.
3. Farmer,S.R. (2006) Transcriptional control of adipocyte formation. *Cell Metab.*, **4**, 263–273.
4. Tontonoz,P., Hu,E. and Spiegelman,B.M. (1994) Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. *Cell*, **79**, 1147–1156.
5. Tontonoz,P. and Spiegelman,B.M. (2008) Fat and beyond: the diverse biology of PPARgamma. *Annu. Rev. Biochem.*, **77**, 289–312.
6. Zhang,Q., Ramlee,M.K., Brunmeir,R., Villanueva,C.J., Halperin,D. and Xu,F. (2012) Dynamic and distinct histone modifications modulate the expression of key adipogenesis regulatory genes. *Cell Cycle*, **11**, 4310–4322.
7. Mikkelsen,T.S., Xu,Z., Zhang,X., Wang,L., Gimble,J.M., Lander,E.S. and Rosen,E.D. (2011) Comparative epigenomic analysis of murine and human adipogenesis. *Cell*, **143**, 156–169.
8. Shechter,D., Dormann,H.L., Allis,C.D. and Hake,S.B. (2007) Extraction, purification and analysis of histones. *Nat. Protoc.*, **2**, 1445–1457.
9. Laborda,J. (1991) 36B4 cDNA used as an estradiol-independent mRNA control is the cDNA for human acidic ribosomal phosphoprotein PO. *Nucleic Acids Res.*, **19**, 3998.
10. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
11. Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J. and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.
12. Siersbaek,M.S., Loft,A., Aagaard,M.M., Nielsen,R., Schmidt,S.F., Petrovic,N., Nedergaard,J. and Mandrup,S. (2012) Genome-wide profiling of peroxisome proliferator-activated receptor gamma in primary epididymal, inguinal, and brown adipocytes reveals depot-selective binding correlated with gene expression. *Mol. Cell Biol.*, **32**, 3452–3463.
13. Santos-Rosa,H., Schneider,R., Bannister,A.J., Sherriff,J., Bernstein,B.E., Emre,N.C., Schreiber,S.L., Mellor,J. and Kouzarides,T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature*, **419**, 407–411.
14. Cao,R., Wang,L., Wang,H., Xia,L., Erdjument-Bromage,H., Tempst,P., Jones,R.S. and Zhang,Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039–1043.
15. Xu,F., Zhang,K. and Grunstein,M. (2005) Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell*, **121**, 375–385.
16. Li,B., Carey,M. and Workman,J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707–719.
17. Agalioti,T., Chen,G. and Thanos,D. (2002) Deciphering the transcriptional histone acetylation code for a human gene. *Cell*, **111**, 381–392.
18. Heintzman,N.D., Stuart,R.K., Hon,G., Fu,Y., Ching,C.W., Hawkins,R.D., Barrera,L.O., Van Calcar,S., Qu,C., Ching,K.A. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, **39**, 311–318.
19. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 21931–21936.
20. Rada-Iglesias,A., Bajpai,R., Swigut,T., Brugmann,S.A., Flynn,R.A. and Wysocka,J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279–283.
21. Core,L.J., Waterfall,J.J. and Lis,J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
22. Parkhomchuk,D., Borodina,T., Amstislavskiy,V., Banaru,M., Hallen,L., Krobitsch,S., Lehrach,H. and Soldatov,A. (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.*, **37**, e123.
23. He,Y., Vogelstein,B., Velculescu,V.E., Papadopoulos,N. and Kinzler,K.W. (2008) The antisense transcriptomes of human cells. *Science*, **322**, 1855–1857.
24. Gansauge,M.T. and Meyer,M. (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.*, **8**, 737–748.
25. Meyer,M., Kircher,M., Gansauge,M.T., Li,H., Racimo,F., Mallick,S., Schraiber,J.G., Jay,F., Prufer,K., de Filippo,C. *et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science*, **338**, 222–226.
26. Shankaranarayanan,P., Mendoza-Parra,M.A., Walia,M., Wang,L., Li,N., Trindade,L.M. and Gronemeyer,H. (2011) Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat. Methods*, **8**, 565–567.
27. Adli,M., Zhu,J. and Bernstein,B.E. (2010) Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat. Methods*, **7**, 615–618.

28. Ng,J.H., Kumar,V., Muratani,M., Kraus,P., Yeo,J.C., Yaw,L.P., Xue,K., Lufkin,T., Prabhakar,S. and Ng,H.H. (2013) In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell*, **24**, 324–333.
29. Zhou,Z.X., Zhang,M.J., Peng,X., Takayama,Y., Xu,X.Y., Huang,L.Z. and Du,L.L. (2013) Mapping genomic hotspots of DNA damage by a single-strand-DNA-compatible and strand-specific ChIP-seq method. *Genome Res.*, **23**, 705–715.
30. Acevedo,L.G., Iniguez,A.L., Holster,H.L., Zhang,X., Green,R. and Farnham,P.J. (2007) Genome-scale ChIP-chip analysis using 10,000 human cells. *BioTechniques*, **43**, 791–797.
31. Dahl,J.A. and Collas,P. (2008) MicroChIP–a rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies. *Nucleic Acids Res.*, **36**, e15.
32. Wu,A.R., Hiatt,J.B., Lu,R., Attema,J.L., Lobo,N.A., Weissman,I.L., Clarke,M.F. and Quake,S.R. (2009) Automated microfluidic chromatin immunoprecipitation from 2,000 cells. *Lab Chip*, **9**, 1365–1370.
33. O'Neill,L.P., VerMilyea,M.D. and Turner,B.M. (2006) Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat. Genet.*, **38**, 835–841.