Published online 26 July 2017

Nucleic Acids Research, 2017, Vol. 45, No. 17 ***e153***
doi: 10.1093/nar/gkx648

Targeted chromatin ligation, a robust epigenetic profiling technique for small cell numbers

Mark A. Zarnegar, Felicia Reinitz, Aaron M. Newman and Michael F. Clarke*

Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA 94305, USA

Received February 15, 2017; Revised June 27, 2017; Editorial Decision July 13, 2017; Accepted July 19, 2017

ABSTRACT

The complexity and inefficiency of chromatin immunoprecipitation strategies restrict their sensitivity and application when examining rare cell populations. We developed a new technique that replaces immunoprecipitation with a simplified chromatin fragmentation and proximity ligation step that eliminates bead purification and washing steps. We present a simple single tube proximity ligation technique, targeted chromatin ligation, that captures histone modification patterns with only 200 cells. Our technique eliminates loss of material and sensitivity due to multiple inefficient steps, while simplifying the workflow to enhance sensitivity and create the potential for novel applications.

INTRODUCTION

Chromatin immunoprecipitation (ChIP) combined with genome-wide next generation sequencing (ChIP-seq) has become an established research tool for investigating broad areas of biology. Standard ChIP-seq typically requires large numbers of cells (1–10 million), limiting its utility in situations when only small numbers of cells can be obtained. For example, biopsy specimens of human tissues are often limited to fewer than 50 000 cells. Moreover, organs and tissues contain complex mixtures of cells containing rare subpopulations, such as in bone marrow, where 1/20 000 cells are hematopoietic stem cells. Thus, applying ChIP-seq to understand biological processes such as stemness and differentiation has been hindered by the need for a large number of cells.

A number of techniques for applying ChIP-seq with low cell numbers (<100 000 cells) have been previously described (1–9) (Supplementary Table S1) including methods optimized for fewer than 10 000 cells (5–8). While some of these methods can increase the recovery of enriched material and improve the efficiency of immunoprecipitation for low cell counts (5,9), they suffer from complicated or inefficient workflows that lead to loss of material at key steps (e.g. immunoprecipitation and washing). These losses, coupled with the small amounts of recovered material, further reduce ChIP-seq sensitivity (due in part to low efficiency conversion of enriched DNA to sequencing libraries). Moreover, methods for applying ChIP to <10 000 cells have been inconsistent or not demonstrated to work with some common histone marks (5–9). Attempts to overcome these shortcomings have produced prohibitively high methodological complexity, requiring an ever-increasing level of expertise for researchers to reproducibly execute protocols and obtain sufficient data quality with decreasing numbers of cells. For epigenetic investigations of rare cell populations to be routinely performed by researchers of variable skill levels, without expensive and complicated devices and procedures, we have developed a new technique for profiling epigenetic landscapes that enhances sensitivity and simplifies the workflow.

We present a simple, novel, bead-free approach for detecting genome-wide histone modification patterns using targeted chromatin ligation (TCL). Our strategy uses proximity ligation of antibody bound adapter, followed by selective amplification of ligated chromatin to enhance the signal relative to background. Our approach utilizes a simple chromatin fragmentation strategy, eliminates the need for bead-based immunoprecipitation and washing and purifies all DNA, allowing unligated nucleotides to provide a carrier effect instead of using additional material. The entire procedure has less processing and handling steps, and less hands-on time than conventional ChIP-seq (Supplemental Table S2), thus providing greatly reduced methodological complexity while generating improved sensitivity and ease of use.

MATERIALS AND METHODS

Targeted chromatin ligations

Reagents. Chromatin Digestion Buffer (CBD): 33 mM Tris-acetate, pH 7.9, 66 mM potassium acetate, 10 mM magnesium acetate, 0.25% Triton X-100, 1 mM EGTA, 10 mM sodium butyrate. Two-times TCL (and N-ChIP) dilution buffer (TDB): (220 mM KCl, 50 mM Tris-acetate, pH 7.9, 0.2% Sarkosyl (Teknova S3376), 0.2% sodium deoxycholate, 1.75% Triton X-100, 40 mM EDTA, 1 mM EGTA). The enzyme mix (EM) used to fragment chromatin

*To whom correspondence should be addressed. Tel: +1 650 736 9639; Fax: +1 650 736 2961; Email: mfclarke@stanford.edu

contains an equal volume of SaqAI (MseI), FspBI (BfaI), Csp6I, and NdeI from Thermo Fisher (FD2174, FD1764, FD0214, FD0583). A protease Inhibitor (PI) cocktail solution (Roche #4693159001 dissolved in phosphate buffered-saline (PBS) to produce a 20× stock) was added to chromatin digestions.

Antibodies used include. Anti-H3K4me3 (Abcam ab8580), anti-H3K27me3 (Active Motif #39155), anti-H3K36me3 (Abcam ab9050) and anti-H3K27ac (Active Motif #39133) were conjugated with Abcam streptavidin conjugation kit (ab102921). After conjugation, antibodies were concentrated with Pierce concentrator columns (100 MWCO 0.5 ml), then diluted to 1 μg/μl with PBS and final concentrations of 150 mM NaCl and 30% glycerol. To prepare working stocks of antibody–adapter complexes, 5 μg of antibody (∼33 pmol) were incubated in 25 μl 1× TCL buffer (equal volumes CBD + TDB) with 41.25 pmol TCL adapters (Supplemental Table S4, ordered from Integrated DNA Technologies) for 2+ h at 4 °C. Antibody–adapter stocks were then diluted to 25–50 ng/μl where appropriate, with 1× TCL buffer. We used T4 DNA ligase (EL0011) and Ligation Buffer (Fisher FERB69). Q5 High Fidelity 2× master mix was used for PCR amplification (New England Biolabs M0492). For transposition based library construction, NEXTERA DNA prep kit (Illumina FC-121–1031) was used. We also used Axygen beads for purifying/size selecting libraries after indexing (Fisher MAGPCRCL5).

Protocol. Chromatin fragmentation was performed by adding 10 μl of digestion mix (150 μl CDB + 8 μl PI + 4 μl EM) to the cell pellet (spun down at ∼1000 G for 10 min) in 1.7 ml tubes (Axygen MCT-175-C). Cells were resuspended by pipetting ∼10×. Samples were then placed in a water bath for 30 min at 37°C. Digestion was stopped by addition of an equal volume of TDB.

A total of 3–5 μl of antibody–adapter complex was added to each TCL sample, mixed by pipetting ∼10×, and then samples were placed at 4°C overnight in a rack without mixing. For MCF7 TCLs, the recommended amounts of antibody bound by adapter are: ∼200 ng anti-H3K27me3, ∼80 ng anti-H3K36me3, ∼40 ng anti-H3K4me3, or ∼100 ng anti-H3K27ac. For neurospheres or other normal mouse cells, the recommended amounts are: ∼80 ng anti-H3K27me3, ∼40 ng anti-H3K36me3 or ∼20 ng anti-H3K4me3). For other cell types/lines, it is recommended to test antibody-adapter amount, beginning with a quantity proportional to the DNA content/cell relative to MCF7 or normal mouse cells.

The next day, samples were placed on the work bench and allowed to reach room temperature (∼15 min). A total of 180 μl of ligation mix (1× ligation buffer + 1 unit ligase) was then added to each sample and mixed by pipetting 2× then samples were incubated for 10 min at RT. A total of 20 μl of 10% Sarkosyl solution was added to each sample, followed by 10 μl of proteinase K (10 mg/ml). Samples were incubated for 1+ h at 65°C to digest protein. DNA was column purified (ZYMO DNA clean and concentrator-5) and eluted in 15 μl EB.

The purified TCL DNA was next used in a 60 μl PCR amplification reaction with 2× Q5 polymerase mix (98°C for 10 s, 63°C for 30 s, 72°C for 2 min). For TCL reactions with two adapters, ∼15–18 cycles were used. For TCL reactions with a single adapter, ∼25–30 cycles of amplification were used. Single adapter/primer amplifications are ∼40% as efficient as standard PCR, as determined by qPCR and thus equivalent to ∼15–18 cycles of standard PCR. After amplification, samples were purified with ZYMO columns (30 μl EB) then quantified with a Qubit 3.0 and HS dsDNA assay kit. Amplifications typically yielded ∼100–700 ng of DNA for 2000 cell TCLs. All TCL samples used in this manuscript were produced using single adapter (A) TCL reactions.

Chromatin immunoprecipitations

Approximately 1 million MCF7 cells were resuspended in 0.25 ml CBD + PI + 10 μl of EM. Chromatin digestions were performed at 37°C for 30 min, followed by dilution with 0.25 ml TDB. Insoluble material was removed by centrifugation at 10 000G for 10 min followed by transferring the solubilized chromatin solution to a new tube. Chromatin was then pre-cleared with 50 μl magnetic Protein A-Dynabeads for 2 h (Invitrogen 10002D). Dynabeads were prepared by washing and resuspension with 1× TCL buffer prior to use. A total of 50 μl of chromatin solution was saved for Input. Another 50 μl Dynabeads, with either 1 μg of anti-H3K36me3, 2 μg of anti-H3K27me3 or 1 μg of anti-H3K27ac, was added to the chromatin solution and then they were incubated overnight at 4°C with rotation. Bead bound chromatin was washed twice with 1× TCL buffer, once with 1× TCL buffer containing 0.3 M NaCl and twice with TE. DNA was eluted by resuspending beads in 100 μl TE containing 1% sodium dodecylsulphate (SDS) and 10 μg Proteinase K (10 mg/ml), followed by incubation at 65°C for 2 h, with mixing every ∼15 min. Beads were removed by magnet and DNA was column purified using Zymo columns and 30 μl elution buffer. ChIP enriched DNA was quantified using a Qubit 3.0 and dsDNA HS assay. N-ChIPs yielded ∼200–300 ng (H3K36me3), ∼30–60 ng (H3K27me3) and ∼60–140 ng DNA (H3K27ac).

For low cell number ChIPs, 200 000 cells were digested as described above, in 0.2 ml digestion volume and processed identically to generate 0.4 ml of pre-cleared chromatin at 500 cell/μl. The 10 000, 2000, 400 or 200 cell equivalents were then aliquoted to PCR tubes and diluted to 200 μl with 1× TCL buffer. We used 125 ng anti-H3K36me3, 250 ng anti-H3K27me3 or 125 ng anti-H3K27ac for each ChIP, with 15 μl of beads. After overnight incubation, samples were washed as described above, then eluted in 50 μl TE for PK digestion. After column purification, samples were eluted in 10 μl.

QPCR analysis

Primers were selected for use after being verified to have similar amplification curves across a 10,000-fold range of input, with no amplification in no template controls, prior to conducting any qPCR analysis. Approximately 20–50 ng of amplified TCL DNA and ChIP DNA for H3K27me3 samples were analyzed by qPCR (10 μl reactions, performed in triplicate) using an AB 7900HT and SYBR Green PCR Master Mix (Applied Biosystems 4309155). SDS software version 2.4 was used to analyze qPCR data. Standard

40 cycle reaction conditions were used (95°C for 10 s, 60°C for 10 s, 72°C for 1 min). Primers (see Supplementary Table S3) were used at 250 nM. Data were reported as Normalized Fold Signal by first calculating the ratio of input, then normalizing all data points against a chosen negative control region.

Library construction

About 25–40 ng of amplified TCL DNA or high cell number N-ChIP enriched DNA were used for library construction using transposition based NEXTERA (followed manufacturer's protocol with ∼8 PCR cycles for indexing). Input samples were made into libraries using NEXTERA with ∼10 cycles of PCR for indexing. Libraries for low cell number ChIP samples (and input) generated from 10 000 to 200 cells were made using NEXTERA XT (followed manufacturer's protocol with 14 cycles of PCR for indexing). Indexed samples were quantified by Qubit 3.0, then pooled to produce ∼5 nM samples ready for submission to sequencing facilities. TCL and N-ChIP libraries were sequenced on a NextSeq500 to obtain 75-bp single end reads with read depths of ∼30–60 million reads.

Cell preparation

MCF7 (ATCC HTB-22) cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and penicillin-streptomycin-glutamine. Cells were trypsinized, pelleted, washed 2× with PBS, then resuspended and counted by hemocytometer. For TCLs, cells were then diluted to 100 cells/μl and 2000 cells were aliquot to 1.7 ml tubes containing 200 μl PBS. Cells were then pelleted at ∼1000G for 5–10 min for TCL. For TCLs with 10 000–200 cells, after counting cells, they were diluted to 500 cells/μl, then serially diluted to 100 cells/μl and 20 cells/μl, prior to making the 10 000–200 cell aliquots pelleted for TCL.

To produce neurospheres, mice (Black6 from a mixed C57Bl6 and B6C3 background) were euthanized by CO_2, decapitated and their brains immediately removed. The subventricular zone (SVZ) was micro-dissected and stored in ice-cold PBS for further processing. The tissue was digested using Liberase DH (Roche) and DNase I (250 U ml^{-1}) at 37 °C for 20 min followed by trituration. Digested tissue was washed in ice-cold HBSS without calcium and magnesium, filtered through a 40-μm filter and FACS-sorted as Lineage$^-$CD24$^-$ cells into neurosphere growth media that is, Neurobasal-A (Invitrogen) supplemented with Glutamax (Life Technologies), 2% B27-A (Invitrogen), mouse recombinant epidermal growth factor (EGF; 20 ng ml^{-1}) and basic fibroblast growth factor (bFGF; 20 ng ml^{-1}) (Shenandoah Biotechnology). Lineage cells were depleted using mouse CD45, CD31 and Ter119 (Biolegend).

After neurospheres formed, they were FACS-sorted into CD15$^+$ Egfr$^+$ and CD15$^-$ Egfr$^+$ or CD15$^-$ Egfr$^-$ cells. In total, ∼7000 cells were sorted for each population before individually processing ∼1000 sorted events/cells for each TCL, as described for MCF7 cells. For FACS analysis the cells were stained with anti-CD15–fluorescein isothiocyanate (FITC) (MMA; BD), and EGF complexed with Alexa647–streptavidin (Life Technologies).

ENCODE data

MCF7 aligned reads from GEO datasets GSM945854, GSM970218, GSM970217 and GSM945859 were downloaded from http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeSydhHistone.

Sequence analysis

Raw sequence reads were uploaded to Galaxy (usegalaxy.org) and aligned to the human genome (hg19) or mouse genome (mm9) using Bowtie (-1,15; -e, 40; -v, 2; -m, 1) and Bowtie2 (-very-fast-local), respectively. Only uniquely mapped reads were retained for further analysis. Alignment files were used to produce signal tracks with DeepTools (100 bp bins with 500 bp read extensions and RPKM normalization). We also filtered out ENCODE blacklisted regions (downloaded from https://sites.google.com/site/anshulkundaje/projects/blacklists). Resulting signal files were used for principal component analysis (PCA) and correlation analysis. Signal files were loaded into the Broad Intitute's IGV browser to visualize data. The neurosphere H3K4me3 data were further filtered to show only reads within promoters (defined as 5000 bp around transcriptional start sites based on RefSeq). To calculate fraction of reads in peaks (FRIP), we used MACS2 (-nomodel, $P = 0.01$, -broad,cuttoff 0.1, duplicates = auto, extension 200) to call peaks using ENCODE MCF7 ChIP data. The resulting Broad Peak BED files were used with Samtools to extract all reads located within peak regions and compared read counts to those of the unfiltered alignment files. Cross-correlation plots were generated for de-duplicated BAM files using Phantompeakqualtools (13), with strand shifts ranging from 0 to 1000 bp at a step size of 5 bp, and otherwise default parameters. We employed ngs.plot (14) with default parameters to generate aggregation plots across all TSS intervals in the hg19 reference genome.

Availability of sequence data

All raw data (FASTQ) files and signal files (BigWig) files used for the production of this manuscript have been deposited into the Sequence Read Archive (SRA) and are available through the Gene Expression Omnibus accession number GSE94804.

RESULTS

Development of targeted chromatin ligation

Development of the TCL procedure (Figure 1A) began with determining suitable chromatin digestion conditions to produce a broad DNA ladder. ChIP-seq techniques utilize one of two strategies to fragment chromatin, sonication or enzymatic digestion with micrococcal nuclease. Sonication is the preferred method when using fixed chromatin, but requires a larger working volume that increases loss of material through greater absorption and destroys some epitopes, contributing to loss of material and limited sensitivity of the assay. Micrococcal nuclease is the preferred method when working with native chromatin, but as with sonication, the

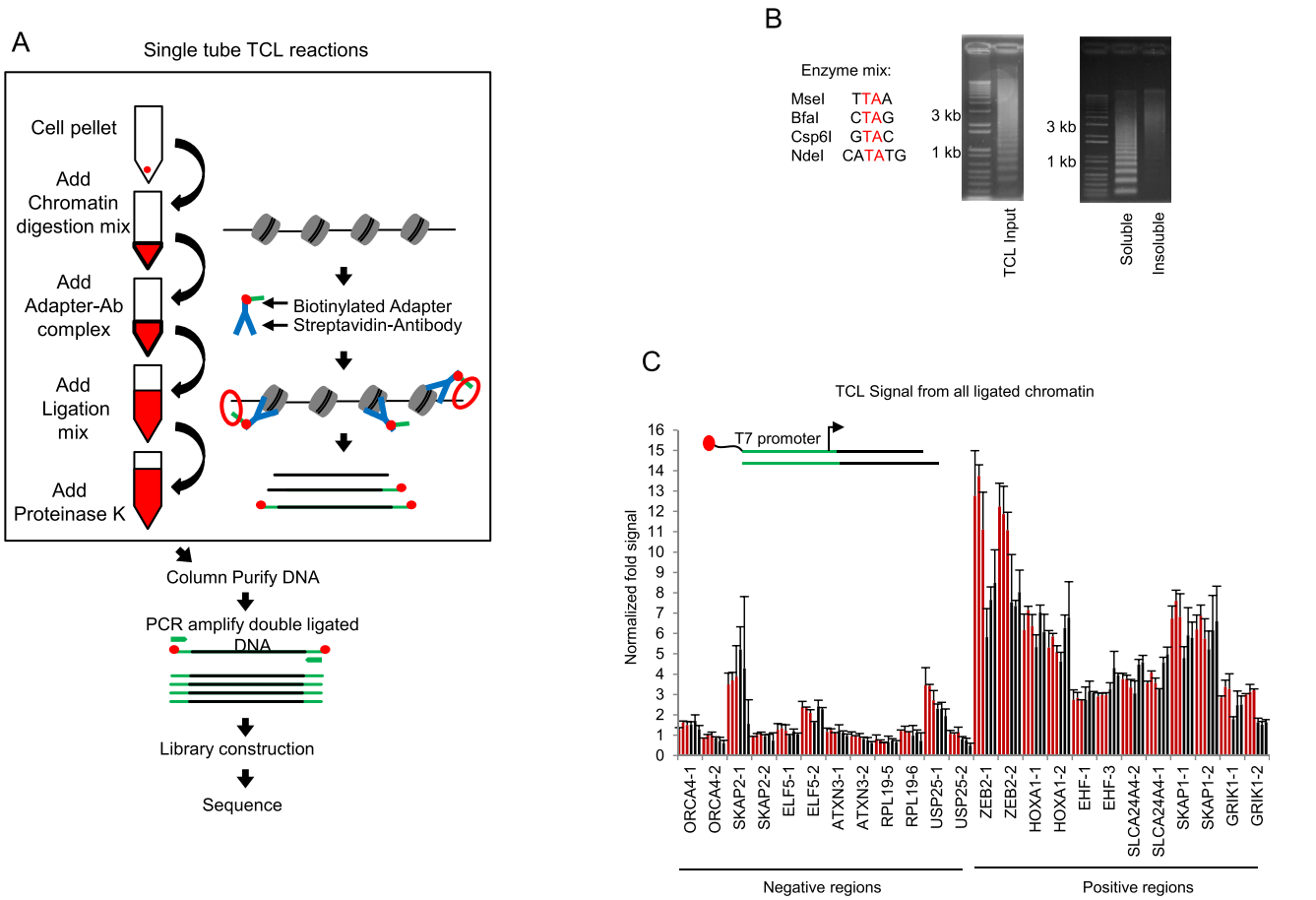

Figure 1. Targeted chromatin ligation (TCL) work flow and chromatin preparation. (**A**) The single tube TCL work flow is depicted (black box), which is followed by amplification and library construction. (**B**) Gel analysis of restriction enzyme fragmented chromatin. A mix of four restriction enzymes was used to digest chromatin, as indicated. A representative gel of input from an MCF7 cell digest with conditions used for all TCL-seq replicates (see 'Materials and Methods' section for details) and a representative gel of soluble MCF7 DNA (N-ChIP) input and the insoluble fraction are shown. (**C**) TCL-qPCR signal to noise ratio is not sensitive to reaction parameters when qPCR analyzing all ligated chromatin fragments after T7-based amplification. TCL-qPCR data were first normalized to input, then normalized against a negative region to generate normalized fold signal. Red bars represent signal from TCL reactions using three different molecular ratios of biotinylated adapter bound to streptavidin conjugated antibody (2:1, 3:1 and 4:1, respectively). Black bars represent signal from TCL reactions using three different quantities of antibody loaded with adapter at a 2:1 ratio (200, 400 and 600 ng, respectively). Data shown are the mean of three or more TCL replicates. Error bars represent S.D. Each bar represents a different genomic region predicted to be negative or positive for H3K27me3 modifications using ENCODE data tracks, and two primer sets for non-overlapping regions of the same gene were used to gauge consistent coverage.

ends of the chromatin are not uniform and require processing that is inefficient and laborious. Additionally, careful titration of micrococcal nuclease is required to prevent over digestion. We sought to simplify the chromatin preparation step so that the experimenter could avoid loss of material, eliminate the need for sonication equipment, and easily avoid excessive digestion. We therefore decided to use a cocktail of restriction enzymes (three 4-base cutters and one 6-base cutter) that produce identical dinucleotide overhangs. Such overhangs can be efficiently ligated relative to the blunt ends or single nucleotide overhangs produced by processing chromatin fragmented by sonication or micrococcal nuclease. We digested unfixed cells in a 10 μl volume of buffer that permeabilizes the cells, followed by a 2-fold dilution in buffer that terminates digestion with EDTA and lyses the cell. Since our TCL procedure does not use beads, we do not need to pellet debris and transfer material to new tubes, a step required by ChIP procedures that reduces background, but contributes to material loss. After testing digestion conditions we were able to generate consistent DNA ladders for TCL and ChIP (Figure 1B).

After determining suitable chromatin digestion conditions, we proceeded to test various reaction parameters of TCLs: antibody concentrations, adapter concentrations, salt concentrations and ligation conditions (volume, temperature, ligase amount). We used 2000 MCF7 cells for each TCL reaction during testing, and all initial testing used anti-H3K27me3 antibody conjugated to streptavidin. After chromatin fragmentation, 3–5 μl of streptavidin-conjugated antibody with bound adapter was added to the 20 μl fragmented chromatin solution, then incubated overnight (Figure 1A). As rotation or agitation of dilute small volume samples could contribute to significant loss of material through absorption to surfaces, we incubated TCL samples without mixing during the overnight incubation. The next day, 180 μl of a ligation mix was added to samples to allow

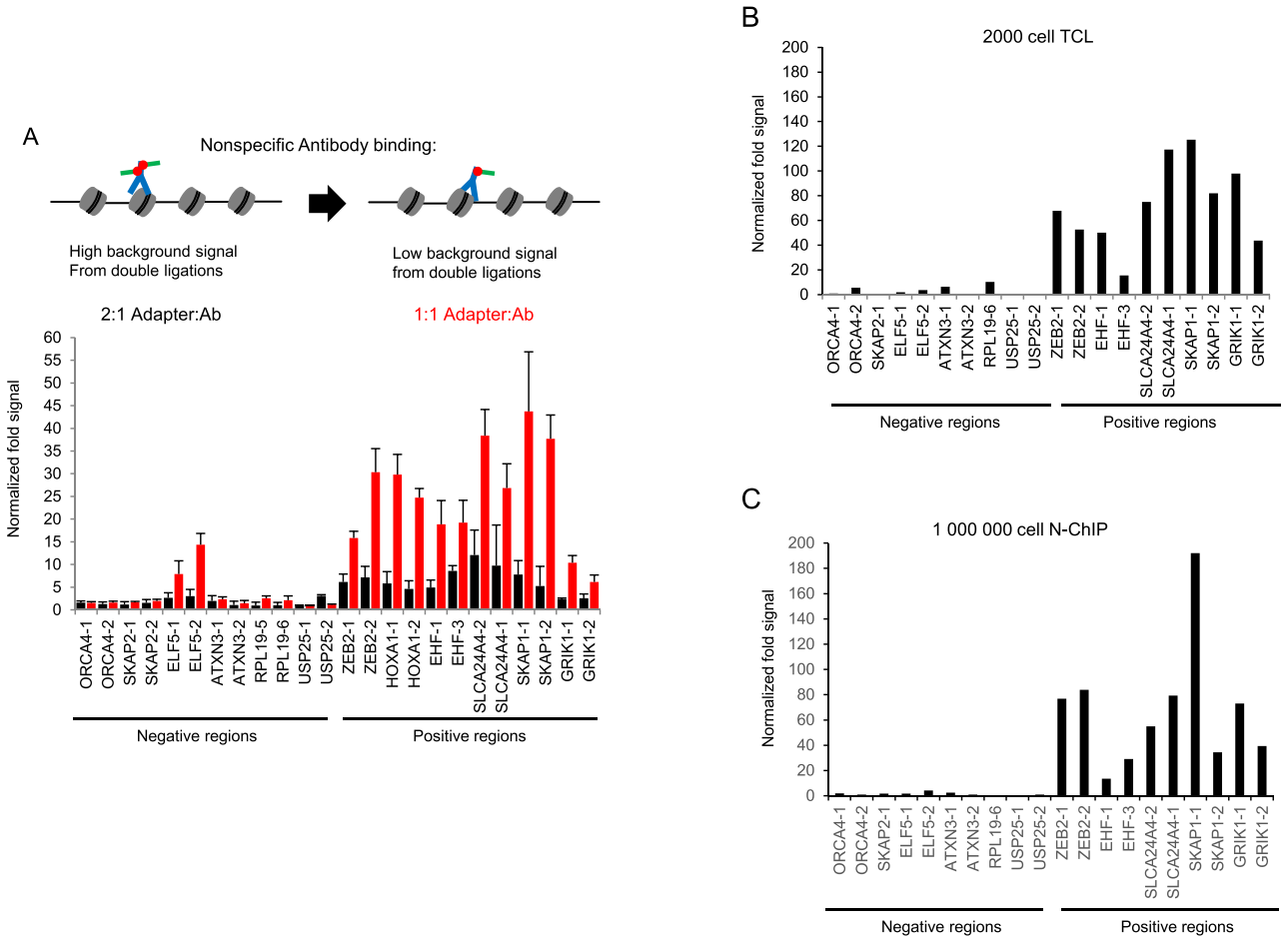

Figure 2. Analysis of a critical TCL reaction parameter. (**A**) A schematic and TCL-qPCR data from optimization of a critical parameter of TCL reactions is provided. The TCL-qPCR data are shown as fold signal described above. Black and red data bars were generated with different adapter to antibody molecular ratios (2:1 and 1:1, respectively). Data shown as mean + S.D. from three independent TCL replicates. (**B**) qPCR data (mean from technical replicates, no error bars provided) from a single representative 2000 cell TCL reaction for H3K27me3 is shown. Data presented as described above. (**C**) qPCR data (mean from technical replicates, no error bars provided) from a representative 1 million cell N-ChIP sample for H3K27me3 is shown. Data presented as described above.

adapters to ligate to chromatin ends. We found the most important parameters for successful TCL reactions to be a 1:1 adapter to antibody ratio and PCR selection of only double ligated chromatin.

Initially, we considered that most ligated chromatin fragments might be ligated at only one end, so a T7 promoter was included on the adapters. TCL reactions were then analyzed by T7-based RNA amplification, followed by reverse transcription and qPCR. When we tested various reaction conditions, we observed highly robust yet limited signal that was insensitive to antibody and adapter concentrations (Figure 1C), or other conditions tested (data not shown). The importance of using a low amount of adapter became clear when we examined only double ligated chromatin amplified by PCR (Figure 2A). Combining selection for double ligations and reducing the ratio of adapter to antibody to 1:1 increased signal to noise detected by TCL-qPCR and made signal sensitive to antibody titration. To evaluate these optimizations, we compared the performance of TCL-qPCR using only 2000 cells with native chromatin ChIPs (N-ChIP) performed with one million cells and identical reagents as TCL. Both methods yielded comparable signals (Figure 2B and C).

Adapter and library construction considerations

During the initial development of TCL, we performed reactions using antibodies loaded with two adapters, an 'A' adapter and a 'B' adapter. We were concerned that using two adapters would limit sensitivity when nearing the lower limit of input cell numbers for TCL reactions. Two adapters lead to formation of four species of double ligation, but PCR amplification would only efficiently amplify half those species of ligation products, potentially leading to drop out of signal during amplification (Supplementary Figure S1A and B). We therefore switched to using a single adapter for TCL reactions and found that one adapter is preferable, in part due to elimination of PCR generated primer dimer artefacts (Supplementary Figure S1C and D, red arrows). However, as head to tail annealing of denatured double ligated DNA having the same adapter on either end suppressed PCR amplification efficiency (~35–45% efficiency

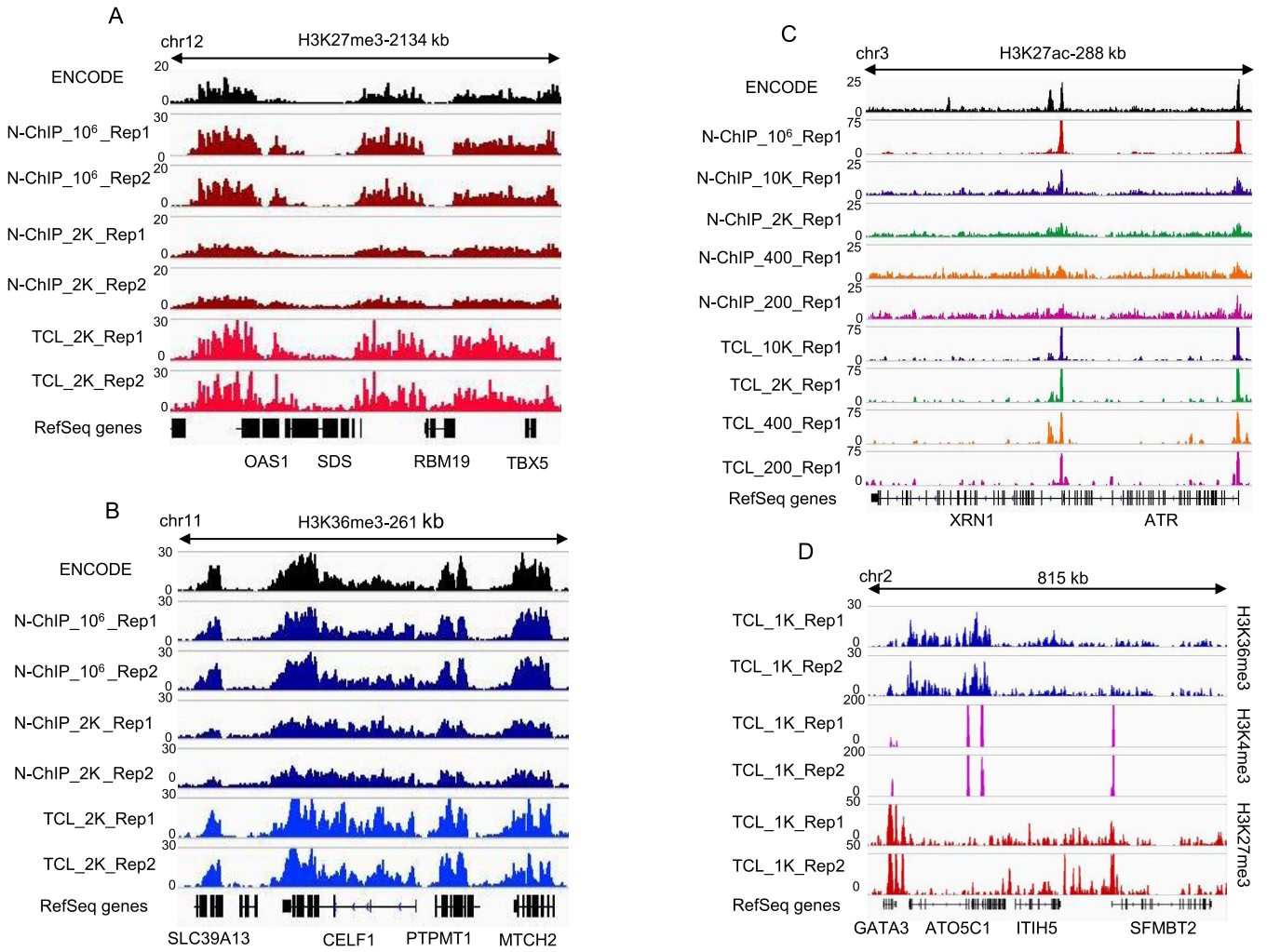

Figure 3. TCL-seq generates high quality data for multiple histone marks and only 200 cells. (**A**) Normalized 2000 cell TCL-seq data for H3K27me3 across a random genomic window is shown in comparison to ChIP-seq data. N-ChIP-seq data was generated using a million cells or 2000 cells. ENCODE data were generated using ∼1 million cells. (**B**) Normalized 2000 cell TCL-seq data for H3K36me3 is presented as described above. (**C**) Normalized TCL-seq data for H3K27ac are shown in comparison to ChIP-seq data as described above. TCL-seq and N-ChIP-seq samples were generated with decreasing cell numbers (10 000, 2000, 400 and 200 cells). (**D**) TCL-seq data produced with <1000 neurosphere cells (∼1000 sorted events) are shown for three different histone marks. H3K36me3 and H3K27me3 signal tracks were generated using all uniquely mapped reads. H3K4me3 signal tracks show only reads across promoters (5000 bp around transcriptional start sites).

based on qPCR, data not shown), we needed to increase the number of PCR cycles during amplification to compensate.

Having validated TCL reactions with either one or two adapters by qPCR, we proceeded with library construction to evaluate genome-wide histone profiles. PCR amplified TCL DNA (Supplementary Figure S1C and D) requires further fragmentation to produce next generation sequencing libraries. We used transposition based tagmentation that simultaneously fragments DNA while inserting sequencing adapters, to construct sequencing libraries.

Genome-wide analysis of TCL versus ChIP

After sequencing, we downloaded ENCODE data for MCF7 (10) and analyzed all libraries using Galaxy. We then performed both qualitative and quantitative assessments of TCL performance. First, we visually examined genome-wide data quality using normalized signal tracks. When comparing data generated from 2000 cell TCLs and 2000 cell ChIPs to ChIP samples produced with a million or more cells, H3K27me3 chromatin profiles generated with TCL were virtually indistinguishable across a range of genomic intervals (Figure 3A). Genomic windows comparing H3K36me3 marks between 2000 cell TCLs and ChIP were also highly concordant (Figure 3B). While 2000 cell ChIPs were able to produce signal reminiscent of one million cell ChIPs and 2000 cell TCLs, the signal was clearly reduced (Figure 3A and B). To further evaluate the sensitivity of TCL, we then compared H3K27ac chromatin landscapes produced with high cell number ChIPs to TCL and ChIP data generated with decreasing cell numbers (10 000–200 cells). Strikingly, the TCL signal was consistently retained even as the number of cells used was reduced to 200, while the ChIP signal dropped sharply even at 10 000 cells and continued to decrease with less cells (Figure 3C and Sup-

Table 1. Fraction of reads in peaks (FRIP) were calculated for MCF7 TCL-seq and ChIP-seq data for H3K36me3, H3K27me3 and H3K27ac histone marks

Sample	All unique reads	Unique reads in peaks	FRIP
TCL-2000-H3K36me3-Rep1	60930711	20533106	0.3369910783
TCL-2000-H3K36me3-Rep2	33857568	11840806	0.3497240558
N-ChIP-106-H3K36me3-Rep1	41255535	18713815	0.4536073766
N-ChIP-106-H3K36me3-Rep2	38408827	18624579	0.4849036134
N-ChIP-2K-H3K36me3-Rep1	72295596	22685546	0.313788768
N-ChIP-2K-H3K36me3-Rep2	57126237	16772871	0.2936106399
ENCODE MCF7-H3K36me3 Rep1	25318535	11194088	0.4421301627
ENCODE MCF7-H3K36me3 Rep2	28802418	11257299	0.3908456228
TCL-2000-H3K27me3-Rep1	78857866	17468201	0.2215150103
TCL-2000-H3K27me3-Rep2	29413373	7013869	0.2384585066
N-ChIP-106-H3K27me3-Rep1	40852490	13200650	0.3231296305
N-ChIP-106-H3K27me3-Rep2	27563539	8757953	0.3177368842
N-ChIP-2K-H3K27me3-Rep1	71475678	13297770	0.1860460841
N-ChIP-2K-H3K27me3-Rep2	68619254	12238198	0.1783493303
ENCODE MCF7-H3K27me3 Rep1	34594045	9585753	0.2770925747
ENCODE MCF7-H3K27me3 Rep2	31179986	7949403	0.2549521029
TCL-103-H3K27Ac-Rep1	24972549	6817962	0.273018265
TCL-103-H3K27Ac-Rep2	23781641	6347332	0.2669005053
N-ChIP-103-H3K27Ac-Rep1	29602838	7632053	0.2578149095
N-ChIP-103-H3K27Ac-Rep2	23317320	6187630	0.2653662599
TCL-2000-H3K27Ac-Rep1	25766600	7664231	0.2974482858
TCL-2000-H3K27Ac-Rep2	17713297	4775925	0.2696237183
N-ChIP-2000-H3K27Ac-Rep1	33092392	7621475	0.2303089786
N-ChIP-2000-H3K27Ac–Rep2	29981764	6524489	0.2176152477
TCL-400-H3K27Ac-Rep1	26458816	8055151	0.3044410982
TCL-400-H3K27Ac-Rep2	35314640	10827426	0.306598793
N-ChIP-400-H3K27Ac-Rep1	43562965	9279297	0.2130088482
N-ChIP-400-H3K27Ac-Rep2	31704721	6906183	0.2178282219
TCL-200-H3K27Ac-Rep1	26839135	6982965	0.260178467
TCL-200-H3K27Ac-Rep2	23096721	5580259	0.2416039489
N-ChIP-200-H3K27Ac-Rep1	41129913	8753920	0.2128358502
N-ChIP-200-H3K27Ac-Rep2	13217284	2990945	0.2262904391
N-ChIP-106-H3K27Ac-Rep1	35190909	16858415	0.4790559687
N-ChIP-106-H3K27Ac-Rep2	29073330	11591090	0.3986846364
ENCODE MCF7-H3K27Ac Rep1	25799447	7325880	0.2839549235
ENCODE MCF7-H3K27Ac Rep2	23868427	6138221	0.2571690627

Peaks were called using MACS2 peak calling parameters suitable for broad peaks, as recommended by ENCODE. Identical peak calling parameters were used for all samples.

plementary Figure S2). While we did not perform N-ChIPs for H3K4me3 and there is no equivalent ENCODE data for that histone mark to make a comparison, we did generate highly reproducible 2000 MCF7 cell TCLs for H3K4me3; visual examination relative to other histone marks revealed high feature specificity with clearly reproducible peak patterns (Supplementary Figure S3). Since MCF7 is a human cancer cell line with an abnormal karyotype and high DNA content (∼10 pg/cell), we sought to test TCL on a normal low DNA content (∼5 pg/cell) cell type. We also sought to ensure that TCL works in the context of sorting limited cell numbers. To accomplish this, we sorted mouse brain derived neurosphere cells and performed TCLs on ∼1000 sorted events (<1000 actual cells). The resulting epigenetic profiles were highly concordant (Figure 3D).

To quantitatively evaluate the robustness of genome wide TCL data, both across biological replicates and in comparison, to ChIP data, we used several approaches, including: examination of signal to noise represented by the FRIP (11,12), correlation analysis and PCA. First, we calculated the FRIP for all MCF7 data profiled by both TCL and ChIP, and found that low cell number TCL samples had superior FRIPs when compared to low cell number ChIPs, and nearly equivalent FRIPs compared to ENCODE ChIPs (Table 1). Notably, the FRIPs generated with 200 and 400 cell H3K27ac TCLs averaged 25.1 and 30.5%, respectively, compared to ENCODE ChIPs that averaged 27.1% (Table 1). In contrast, ChIP samples produced with 2000, 400 or 200 cells had FRIP scores 3–9% less than comparable TCLs, consistent with the visually inferior signal to noise observed (Figure 3A–C). We also evaluated signal to noise ratios by strand cross-correlation analysis (11,13) and found the results to be consistent with FRIP, as low cell TCLs had higher normalized strand coefficients and higher relative strand correlations compared to equivalent cell number ChIPs (Supplementary Figure S4A). Aggregation plots (14) of H3K27ac around TSS intervals also suggest that TCLs are more similar to one million cell N-ChIPs than low cell N-ChIPs (Supplementary Figure S4B). Next, we produced genome-wide Pearson correlation plots using 2 kb genomic windows. Strong correlations between TCL and ChIP data were observed (Figure 4A). Genome-wide correlations for TCL versus high cell number N-ChIP averaged $r = 0.69$ for H3K27me3, $r = 0.83$ for H3K36me3 and $r = 0.6$ for H3K27ac. Correlations between biological replicates of TCLs were high and comparable to ChIPs, with nearly all replicates having $r > 0.8$ using 2 kb genomic bins (Figure 4A and Supplementary Figure S5). Biological replicates of 2000 cell H3K4me3 TCLs were also highly correlated with $r = 0.88$. Additionally, analysis of the 10 000–200 cell H3K27ac TCLs demonstrated high correlations between biological replicates and comparing 10 000 cell TCLs to 400 cell TCLs revealed high correlations (Supplementary Figure S5). Genome wide correlations for the

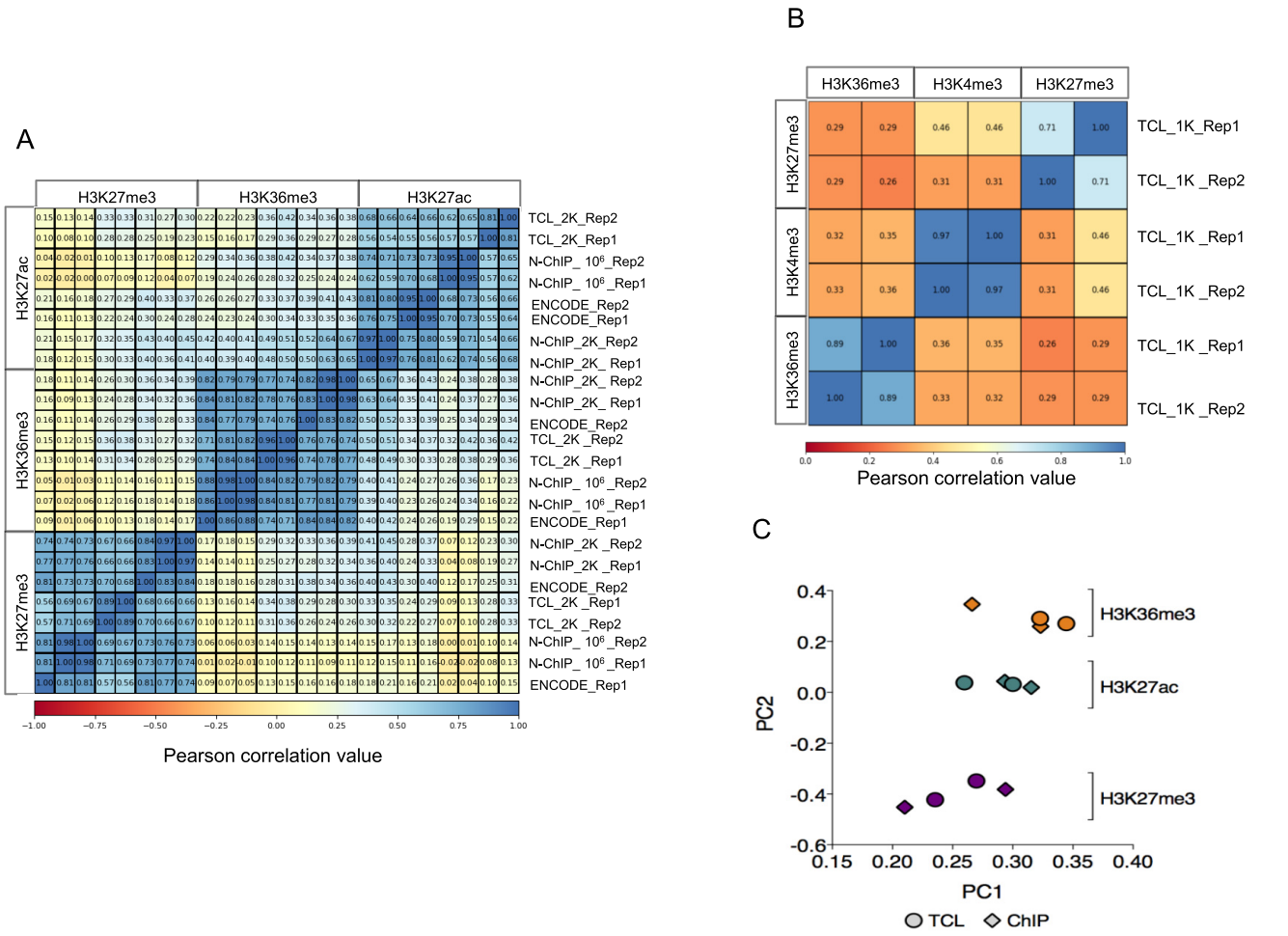

Figure 4. Correlation and principal component analysis (PCA) of genome wide data. (**A**) A heat map of Pearson correlations for TCLs (2000 cell samples), N-ChIP (2000 cell and 1 000 000 cell samples) and ENCODE ChIP data are shown. Deeptools in Galaxy was used to generate the heat map image using 2 kb bins. (**B**) A heat map showing Pearson correlations for neurosphere TCL data are shown and was generated as described above. (**C**) PCA of 2000 cell TCL data and ENCODE ChIP data is shown. Data were generated using Deeptools and a 500-bp bin size.

<1000 cell neurosphere TCLs were also very high with $r = 0.71–0.97$ (Figure 4B). Finally, we performed PCA analysis on the TCL and ChIP data from MCF7 samples using various bin sizes, from 500 bp to 20 kb, and found that the TCL data and ChIP data clustered together based on histone marks. Increasing the bin size had a negligible effect as data were already well clustered using small 500-bp windows (Figure 4C). Notably, the TCL samples appeared to cluster more tightly than the ChIP samples, indicating strong reproducibility (Figure 4C).

DISCUSSION

The simple workflow of the TCL technique, outlined in Figure 1, begins with resuspension of unfixed cells in digestion buffer containing enzymes that generate native chromatin fragments with identical dinucleotide overhangs. While ChIP-seq methods seek to fragment chromatin into small fragments (250–500 bp), which reduces background chromatin binding to beads, facilitates library construction and maximizes data resolution, our bead-free strategy and library construction method obviate the need for small fragments and allows the strategic use of larger chromatin fragments for greater sensitivity.

The large chromatin fragments facilitate relatively symmetrically distributed background binding events, presumably to unmodified histones, that likely drives most ligation events detected in Figure 1C. Since our technique eliminates washing to improve stringency, and most background appears to be driven by background antibody binding and not reaction parameters, signal specificity was limited when examining all ligation events. We took advantage of the apparent background binding by hypothesizing that if the initial ligation of chromatin ends is driven by specific or non-specific antibody binding, secondary ligation events should be driven by higher locally concentrated adapters as a function of antibody specificity and concentration. Therefore, using enough antibody to have at least one non-specific binding event per chromatin fragment, but not two, should facilitate capturing more frequent specificity driven double ligations when limiting the amount of adapter bound to the antibody, and producing increased signal to noise while also maintaining sufficient depth of coverage for regions with lower abundancy of modified histone targets across multin-

ucleosome fragments that might fail to produce double ligations without nonspecific binding events. Since double ligations should occur more frequently with bigger fragments that can support both specific and non-specific antibody binding, amplification of double ligated chromatin should select for larger fragments. The data presented in Supplementary Figure 1C supports this interpretation and shows that ligations in the presence of IgG or no antibody selects smaller fragments where ligations are driven by diffusion and intermolecular ligation, not intramolecular proximity ligation.

As cell numbers used for epigenetic profiling decrease, it may be expected that methods will become increasingly sensitive to antibody quality. For example, we originally tested three ChIP-seq validated antibodies against H3K27me3, including Millipore #07–499, used by ENCODE and found all three worked well for high cell number ChIP. Indeed, all three produced similar results for TCL-qPCR, but one produced sequence data that better matched ENCODE H3K27me3 signal tracks (data not shown). Thus, researchers should not assume all ChIP validated antibodies are compatible with TCL or any other low cell genome wide profiling technique.

While amplification of ligated material prior to transposase based library construction masks the duplication rate, and single end reads do not support accurate estimation of duplication rates, we did assess duplication rates and found that ∼17–27% of unique reads map to identical 5′ sequences. That likely overestimates the real duplication rate, suggesting the library complexity produced by TCL remains high, and is superior to other low cell epigenetic profiling techniques. For example, the single end data from ChIP-seq with 10 000–100 cells, produced by a microfluidic ChIP-seq device (7), had duplication rates calculated to be in the range of 55–80%. The apparent low duplication, along with the robustness demonstrated by our PCA and correlation analysis clearly indicate that TCL produces high quality robust data.

We have demonstrated a greatly simplified approach for producing high quality histone modification profiles that is unique and distinct from ChIP. Key advantages of TCL include greatly reduced handling through elimination of inefficient immunoprecipitation, washing and the subsequent need for inefficient enzymatic end repair and single nucleotide or blunt-end ligation steps with picogram quantities of starting material. These qualities should make TCL more amenable to microfluidic adaptation, automation and further optimization with even less than the 200 cells tested here. While the current iteration of TCL was designed and tested only for mapping histone modifications, we are currently working to adapt TCL for use with transcription factors. We also believe TCL offers the opportunity for studies beyond what is possible for ChIP, such as multiplexing of histone modifications or transcription factor co-occupancy without re-ChIP. For example, it may be possible to preload antibodies against H3K27me3 and H3K4me3 with different barcoded adapters so that their simultaneous use can allow direct amplification and detection of true bivalent chromatin. TCL thus provides robust epigenetic profiles from low cell numbers in an easy to execute approach with the potential for novel applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

California Institute for Regenerative Medicine [GC1R-06673-A to Michael Snyder, M.F.C]; Breast Cancer Research Foundation [SPO 36622 to M.F.C]; Ludwig Foundation [1149727–103-KHABQ to M.F.C.]. Funding for open access charge: Ludwig Foundation [1149727–103-KHABQ].

Conflict of interest statement. Mark A. Zarnegar and Michael F. Clarke are co-inventors of Targeted Chromatin Ligation, for which Stanford University is pursuing a patent.

REFERENCES

1. Gilfillan,G.D., Hughes,T., Sheng,Y., Hjorthaug,H.S., Straub,T., Gervin,K., Harris,J.R., Undlien,D.E. and Lyle,R. (2012) Limitations and possibilities of low cell number ChIP-seq. *BMC Genomics*, **13**, 645.
2. Adli,M., Zhu,J. and Bernstein,B.E. (2010) Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat. Methods*, **8**, 615–618.
3. 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*, **7**, 565–567.
4. Schmidl,C., Rendeiro,A.F., Sheffield,N.C. and Bock,C. (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods*, **10**, 963–965.
5. Zheng,X., Yue,S., Chen,H., Weber,B., Jia,J. and Zheng,Y. (2015) Low-cell-number epigenome profiling aids the study of lens aging and hematopoiesis. *Cell Rep.*, **7**, 1505–1518.
6. Lara-Astiaso,D., Weiner,A., Lorenzo-Vivas,E., Zaretsky,I., Jaitin,D.A., David,E., Keren-Shaul,H., Mildner,A., Winter,D., Jung,S., Friedman,N. and Amit,I. (2014) Immunogenetics. Chromatin state dynamics during blood formation. *Science*, **345**, 943–949.
7. Brind'Amour,J., Liu,S., Hudson,M., Chen,C., Karimi,M.M. and Lorincz,M.C. (2015) An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6**, 6033.
8. van Galen,P., Viny,A.D., Ram,O., Ryan,R.J., Cotton,M.J., Donohue,L., Sievers,C., Drier,Y., Liau,B.B., Gillespie,S.M. *et al.* (2016) A multiplexed system for quantitative comparisons of chromatin landscapes. *Mol. Cell*, **1**, 170–180.
9. Cao,Z., Chen,C., He,B., Tan,K. and Lu,C. (2015) A microfluidic device for epigenomic profiling using 100 cells. *Nat. Methods*, **10**, 959–962.
10. Dunham,I., Kundaje,A., Aldred,S.F., Collins,P.J., Davis,C.A., Doyle,F., Epstein,C.B., Frietze,S., Harrow,J., Kaul,R. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
11. Landt,S.G., Marinov,G.K., Kundaje,A., Kheradpour,P., Pauli,F., Batzoglou,S., Bernstein,B.E., Bickel,P., Brown,J.B., Cayting,P. *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.*, **9**, 1813–1831.
12. Kellis,M., Wold,B., Snyder,M.P., Bernstein,B.E., Kundaje,A., Marinov,G.K., Ward,L.D., Birney,E., Crawford,GE, Dekker,J. *et al.* (2014) Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 6131–6138.
13. Marinov,G.K., Kundaje,A., Park,P.J. and Wold,B.J. (2014) Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda).*, **4**, 209–223.
14. Shen,L., Shao,N., Liu,X and Nestler,E. (2014) ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics.*, **15**, 284.