

Exonuclease combinations reduce noises in 3D genomics technologies

Siyuan Kong^{1,†}, Qing Li^{1,†}, Gaolin Zhang^{1,†}, Qiuja Li¹, Qitong Huang¹, Lei Huang¹, Hui Zhang², Yinghua Huang², Yanling Peng¹, Baoming Qin² and Yubo Zhang^{1,*}

¹Animal Functional Genomics Group, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China and ²CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China

Received January 04, 2019; Revised February 04, 2020; Editorial Decision February 07, 2020; Accepted February 19, 2020

ABSTRACT

Chromosome conformation-capture technologies are widely used in 3D genomics; however, experimentally, such methods have high-noise limitations and, therefore, require significant bioinformatics efforts to extract reliable distal interactions. Miscellaneous undesired linear DNAs, present during proximity-ligation, represent a main noise source, which needs to be minimized or eliminated. In this study, different exonuclease combinations were tested to remove linear DNA fragments from a circularized DNA preparation. This method efficiently removed linear DNAs, raised the proportion of annulation and increased the valid-pairs ratio from ~40% to ~80% for enhanced interaction detection in standard Hi-C. This strategy is applicable for development of various 3D genomics technologies, or optimization of Hi-C sequencing efficiency.

INTRODUCTION

Chromosome conformation capture-based technologies (C-technologies) are pillar methods in three dimensional (3D) genomics, which include chromosome conformation capture (3C) and 3C-derived technologies, such as circular chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C) and high-throughput chromatin conformation capture (Hi-C) (1,2). They are widely used to elucidate genome folding and to reconstruct 3D topological structure, both of which are crucial in transcriptional regulation and cellular fate determination (2). For example, 3C demonstrated that H19–Igf2 interaction affects the progression of myogenic differentiation (3); Hi-C allowed scientists to elucidate topologically

associated domains (TADs) that were mainly anchored by CTCF/Cohesin proteins, and which were associated with cell-specific gene expression (4). By employing these advanced methods, researchers can reconstruct 3D genome architecture, in eukaryotes, and probe unique insights into the topological mechanism of transcription and variations across the genome (5).

These methods are based on the principle that if the two DNA fragments interact with the same regulatory complex, both will be spatially contiguous (6). After proximity-dependent ligation between two DNA fragment-ends, quantification of ligation junctions is derived by PCR or sequencing, which is then used to measure DNA contact frequencies (7). Therefore, proximity-ligation offers a pivotal step for determining distal interaction identification (8). Proximity ligation is the basis of the Hi-C method. At the same time, noise could also be generated during the process (9). The desired circularized templates are present in a mixture with miscellaneous linear-type DNA, including dangling ends, internal fragments, contiguous ligation DNA, re-ligation DNA, dumped fragments, and so forth (Figure 1A) (10–12). Furthermore, PCR amplification also enhances artificial noise, due to the admixture of different DNA forms.

With a mixture of templates, both the amplification products and efficiency are dramatically affected by artifacts (heteroduplexes and chimeras), supercoiling and biases, which might significantly affect the proportion of valid Hi-C data or other C-technology data (Figure 1B) (2,13–16). Consequently, C-technologies remain as high-noise genomic methodologies (Figure 1) and, thus, require further optimization or development (11). Therefore, increasing the circularized ligation ratio and lowering the linear DNA quantity would be an important measure to optimize C-technologies. If the linear DNA is efficiently eliminated, the remaining circularized DNA will be directly read out

*To whom correspondence should be addressed. Tel: +86 75528434229; Fax: +86 75523251430; Email: ribon.001@163.com

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

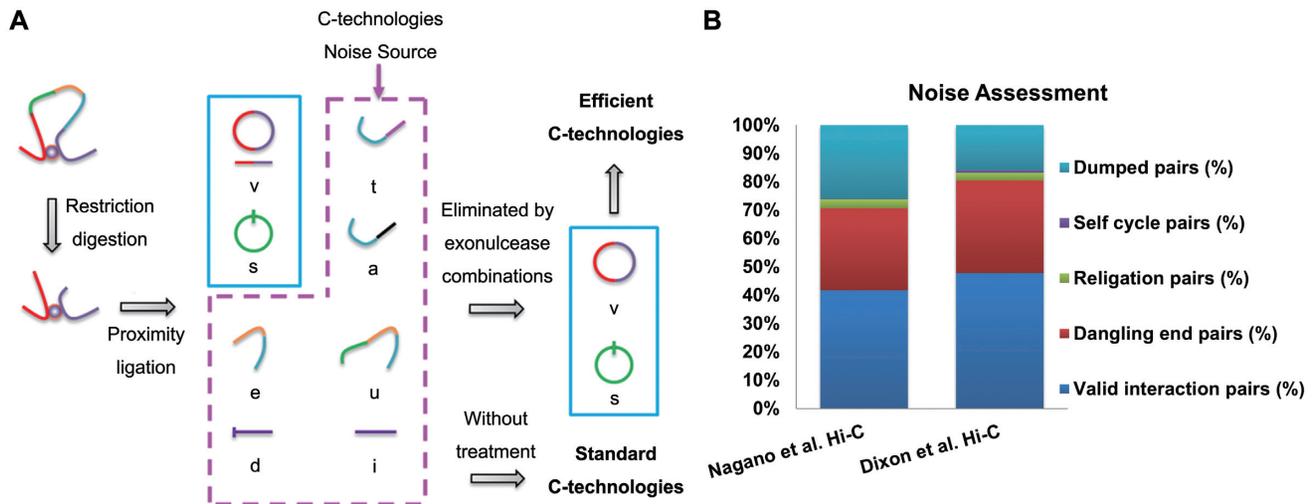


Figure 1. Experimental overview. (A) Noise sources generated by C-technology, aggregated from 3C, 4C, 5C and Hi-C protocols. v, valid ligation; s, self-ligation; e, re-ligation; d, same fragment dangling ends; t, intermolecular ligation; a, random ligation; u, contiguous ligation; i, same fragment internal. (B) HiC-Pro filtering result of Nagano *et al.* standard Hi-C (Standard Hi-C 2) and Dixon *et al.* standard Hi-C (Standard Hi-C 3).

(amplification, quantitation or sequencing), the principle of which is illustrated in Figure 1A.

Exonuclease is a powerful tool for DNA manipulation by cleaving nucleotides, one at a time, from the end (exo) of a polynucleotide chain (17–21). Various combinations of exonucleases have been applied to explore gene editing, DNA manipulation and technology development. For example, a combination of Lambda and RecJF exonuclease was used to develop a ChIP-exonuclease method (ChIP-exo) to define precise binding sites along the genome (22).

In this study, to increase the proportion of desired circularized templates, during C-technologies, we first tested different exonuclease combinations, in order to treat different DNA forms, derived from plasmids, to measure their efficiency. We here demonstrate that the exonuclease combinations of ‘Lambda and RecJF’, ‘Lambda and Exonuclease I’ or ‘Exonuclease I and Exonuclease III’ can be used to efficiently eliminate linearized DNAs (Figure 2A and Supplementary Figure S1A). Finally, we show that ‘Lambda and Exonuclease I’, ‘Lambda and RecJF’ or ‘Exonuclease I and Exonuclease III’ can also be employed for Hi-C development under the right conditions.

MATERIALS AND METHODS

Cell culture

Mouse C2C12 cells (ATCC[®], CRL-1772[™]) were cultured under the following conditions: DMEM (GIBCO, 11965092), 20% fetal bovine serum (FBS) (HyClone, 30071.03) and 1% penicillin/streptomycin (GIBCO, 15140), and maintained at 37°C with 5% CO₂ in Corning[®] 100 mm culture dishes. The cells were divided from one to four dishes, strictly based on when the confluence reaching 50–60%. Cells were then passaged using 2 ml of 0.25% trypsin (GIBCO, 25200056) at 37°C for 5 min, followed by blocking with 6 ml complete medium. Cells were then mixed, by pipetting, and then transferred into a 15 ml cen-

trifuge tube and centrifuged at 200 × g for 5 min. The supernatant was then discarded and cells were re-suspended, with 1 ml complete medium, followed by transfer of appropriate aliquots (200 μl) to new culture dishes. For C-technology tests, cells were harvested and cross-linked until they reached 6–8 × 10⁶ in number (100 mm dish, 80–90% cell confluence). Formaldehyde (HCHO) was added to the dish (280 μl of 37% HCHO to 10 ml medium, 1% final concentration) and cells were incubated at 37°C for 10 min. Crosslinking was stopped by adding 890 μl of 2.5 M glycine to the medium (0.2 M final concentration) at 25°C for 5 min, and the medium was then aspirated. Cells were washed, twice, with 10 ml of 1× PBS (GIBCO, 14190250) and then harvested from the dish by scraping with a plastic spatula. Remaining cells were washed into the centrifuge tube, using 0.5% BSA/PBS, followed by centrifugation at 500 × g, at 4°C for 5 min. The cells were either stored at –80°C or directly used for C-technology cell lysis assays.

Mouse E14 embryonic stem cells (mESCs) (ATCC[®], CRL-1821[™]) were cultured under specific conditions (KO-DMEM complete medium) in KO-DMEM (Invitrogen Life Technologies, 10829-018) supplemented with 15% FBS (Invitrogen Life Technologies, 10099-141), 1% penicillin/streptomycin (GIBCO, 15140), 1% MEM nonessential amino acids (NEAA) (Invitrogen Life Technologies, 11140-122), 1% 2-Mercaptoethanol (BME) (Sigma, M-7522), 1% GlutaMAX[™] (GIBCO, 35050061), 1% Vitamin C (Sigma, A8960), 1% LIF (Chemicon, ESG-1107), 1% PD0325901 (Selleck, S1036) and 3% CHIR-99021 (Selleck, S1263). Before cells were thawed, or passaged, dish bottoms were covered with warmed 0.1% gelatin (Sigma, E1270) (37°C). Next, the gelatin solution was removed and 0.05 million cells per 60 mm dish were seeded into KO-DMEM complete medium. In subculture, cell colony morphology, density and size were initially observed. In a 60 mm dish, efficient extraction of mESCs was achieved with an average colony size of 200–400 μm (in diameter) and a spacing of ~400 μm between

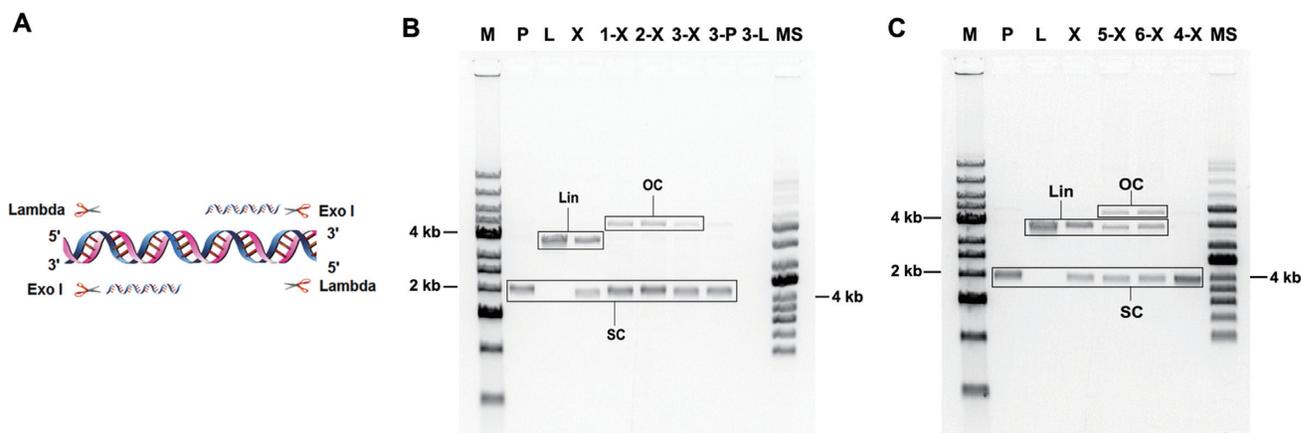


Figure 2. Experimental results for exonuclease combinations treatment. (A) The cleavage mechanism of Lambda and Exonuclease I combinations. Other exonuclease combinations (Lambda and RecJF; Exonuclease I and Exonuclease III) are shown in Supplementary Figure S1A and Supplementary Table S1. (B) Three exonuclease combinations (LRL; LRC; LIC) removed linear DNA from a paradigm mixture. M, 1 kb DNA ladder; P, pGL4.23 plasmid; L, linearized plasmid; X, mixture (plasmid and linear DNA 1:1); 1-X, LRL-Mixture, LRL to cut mixture; 2-X, LRC-Mixture, LRC to cut mixture; 3-X, LIC-Mixture, LIC to cut mixture; 3-P, LIC to cut plasmid; 3-L, LIC-Lin, LIC to cut linearized plasmid; MS, supercoiled ladder. (C) I+III combination and Exonuclease VIII, truncated elimination tests. 5-X, VIII4-Mixture, Exonuclease VIII, truncated within Buffer 4 to remove mixture; 6-X, VIIC-Mixture, Exonuclease VIII, truncated within CutSmart buffer to remove mixture; 4-X: I+III-Mixture, I+III to remove mixture. Loading samples for agarose gel electrophoresis were purified by phenol-chloroform.

colonies. After the cell clones were washed, with $1 \times$ DPBS (Invitrogen Life Technologies, 14190-144), the digestive enzyme accutase (Invitrogen Life Technologies, A1110501) was used instead of trypsin. Other steps were similar as for C2C12 cells (23).

Agarose gel electrophoresis and real-time qPCR

For agarose gel electrophoresis, loading samples (40 ng) were used for electrophoresis (0.8% agarose gel), performed at 100 V for 40 min. Gels were digitally imaged using a system adapted for photography (Bio-Rad Laboratories, CA, USA). After LRL (Lambda and RecJF within Lambda buffer) (NEB), LRC (Lambda and RecJF within Cutsmart buffer) (NEB), LIC (Lambda and Exonuclease I within Cutsmart buffer) (NEB) and I+III (I+III or IIIIC) (Exonuclease I and Exonuclease III within Cutsmart buffer) (NEB) elimination, qPCR was performed. Primer BH2 for qPCR was designed across both ends of the BamHI restriction site. Primer B and C were randomly designed away from the cutting site (<http://bioinfo.ut.ee/primer3/>) (Supplementary Table S15 and Supplementary Figure S1C). The experiments in triplicate were combined to enlarge the quantity and volume before purification. For same quantity (SQ) qPCR, after phenol/chloroform purification, each sample was diluted to 1 ng/ μ l and then subjected to a further 1000-fold dilution (1 μ l sample + 999 μ l H₂O) for quantification, with three replicates per sample. For same volume (SV) qPCR, each sample of the same volume was diluted 100-fold (1 μ l was diluted to 100 μ l by H₂O). The 20 μ l qPCR reaction mixtures were prepared comprised of 10 μ l SYBR qPCR Master Mix (Invitrogen Life Technologies, 4368702), 4 μ l H₂O, 2.5 μ l Primer F (1 μ M), 2.5 μ l Primer R (1 μ M) and 1 μ l Template.

All qPCR amplification reactions were performed using the following program: 95°C for 10 min for initial denaturation; 95°C for 15 s and 60°C for 2 min for 40 reaction

cycles; and 95°C for 15 s, 60°C for 2 min and 95°C for 15 s to develop a melt curve. Real-time fluorescence measurements were performed on an Applied Biosystems StepOne-Plus Real-Time qPCR System (Thermo Fisher Scientific, Shanghai, China). Relative DNA enrichment was calculated by $2^{-\Delta C_t}$. The balanced mix of plasmid and linearized DNA, without elimination, was regarded as standard 1. Data are displayed as means \pm the standard error (SEM) of triplicate experiments. Real-time qPCR was conducted following the MIQE guidelines (24).

Exonuclease combination pre-tests for linearized DNA elimination

The pGL4.23 (4283 bp, P1) (Supplementary Figure S1C) and pCDNA3.1-NLS-NgAgo (8246 bp, P2) plasmids were extracted, using Endo-free Plasmid Mini Kit II (E.Z.N.A.[®], D6950). Linearized plasmid DNA was obtained in the following 200 μ l solution/per reaction, including 23.4 μ l plasmid, 5 μ l BamHI-HF (HindIII for P2) (NEB), 20 μ l 10 \times Cutsmart Buffer (NEBuffer 2 for P2) and 151.6 μ l H₂O. Three reactions were performed at 37°C for 1 h. To increase the yield of linearized plasmid DNA, three digested solutions were combined, followed by addition of 1.2-fold phenol/chloroform/isoamyl alcohol for DNA extraction (Supplementary Figure S2). Linearized plasmid DNA elimination studies were then performed, in six groups, to test different exonuclease treatment combinations: LRL-Mixture, LRC-Mixture and LIC-Mixture (Supplementary Table S2); VIII4-Mixture, VIIC-Mixture and I+III-Mixture (Supplementary Table S3). In the treatment reactions, 250 ng plasmid and 250 ng linearized DNA were mixed as templates (Mix or Mixture). The other two groups were employed: 500 ng plasmid (LIC-plasmid) or linearized DNA (LIC-Lin) was incubated under the LIC treatment. Irrespective of whether LRL, LRC or LIC was employed, the quantity of exonucleases used was sufficient

to erase 500 ng linearized DNA. Each treatment was performed in triplicate. All eight treatments were incubated at 37°C for 1 h. After digestion, Qubit 2.0 Fluorimeter (Invitrogen, USA) was used to measure the residual DNA concentration. Alternatively, different time points of 1, 4 and 16 h were tested for DNA digestion, under I+IIIIC-Mix treatment conditions (Supplementary Table S5), with each incubation period having three replicates. A 1.2-fold volume of phenol/chloroform/isoamyl alcohol was used for DNA purification and the resultant DNA was analyzed by agarose gel electrophoresis. Agencourt Ampure XP bead (Beckman Coulter) purification was carried out as an alternative purification method (Supplementary Figures S4 and S5). Data are presented as mean \pm SEM.

Linear noises eliminating tests after proximity-ligation

In the earlier steps for chromosome conformation capture technologies, these samples were assayed by cross-linking, lysis, digestion and marking, followed by ligation and DNA purification (Supplemental S.2). For linear DNA eliminating tests, aliquots of 572 ng library DNA, per reaction (10 reactions), and 40 ng library DNA, per reaction (15 reactions), derived from C2C12, were used. Here, the DNA concentrations were measured by Qubit (Supplemental S.2.1 and Supplementary Table S6, Supplemental S.2.2 and Supplementary Table S9) and the solutions then combined and mixed thoroughly. For 572 ng library test, the resultant solution was divided, equally, into ten 1.5 μ l Eppendorf (EP) tubes (Supplementary Table S7–S8). But for 40 ng library tests, the combined solution was divided into 15 tubes (Supplemental S.2, Supplementary Table S10–S13). The content of each tube was then treated with either LRL, LRC, LIC, I+IIIIC, LRLP1 (P1 added as chaperon carrier), LRCP1, LICP1, or I+IIIICP1; as controls, we used Co and CoP1 (Supplementary Table S7–S8). Additionally, P2 plasmid was added as a replicate for the P1 chaperon carrier test for the 40 ng library test (Supplementary Table S11–S13). A 50 μ l reaction system was used for these tests, and was incubated at 37°C for 1 h. After digestion, 1.2-fold (60 μ l) Ampure XP bead purification was performed. Qubit was used to measure the linear DNA elimination efficiency. Agarose gel electrophoresis was used to detect different DNA forms (Supplemental S.2).

Hi-C improvement and bioinformatics analysis

The exo-Hi-C was basing on standard Hi-C (10). The linear DNA eliminating step was specifically introduced after proximity-ligation. LRC or LIC were employed to assess their effectiveness for eliminating linear interaction noise. For *in situ* system, *in situ* exo-Hi-C protocol was performed based on Rao *et al.* (Supplementary Figure S21 and Supplement Protocol) (25). LIC or IIIIC were employed to assess their effectiveness for eliminating linear interaction noise. These Hi-C libraries were sequenced (PE 150) on Illumina NovaSeq or Hiseq X-ten and cleaned FASTQ data were submitted to HiC-Pro 2.10.0 (<https://github.com/nservant/HiC-Pro>) (12). After mapping to the reference genome and filtering experimental artefacts, detailed statistics were generated. The Standard Hi-C 1 raw data were downloaded

from GEO-GSM862721 (26). The standard Hi-C 2 raw data were downloaded from GEO-GSM1718027 (27). The standard Hi-C 3 raw data were downloaded from GEO-GSM862720 (26). HiC-Pro terms are described in Supplementary Figure S10. They were side by side submitted for HiC-Pro analysis. For Exo-Hi-C, after HiC-Pro filtering, the correlation score was calculated by GenomeDISCO, according to raw matrices, at 50 kb resolution (28). Paired interactions, separated by different genomic distance, were analyzed by Empirical Cumulative Distribution Function (25). The first principal component (PC1) value for each chromosome was used to identify regions of the genome as belonging to either the A or B compartment. PCA of Hi-C data was performed by HOMER software with parameters (-res 50000 -window 100000 -genome mm9) (29). The A compartment (PC1 > 0) and B compartment (PC1 < 0) were obtained at 50 kb bin resolution (10,30). A series of Hi-C interaction frequency heatmaps were drawn by Juicebox (31,32). The Knight-Ruiz (KR) normalization was conducted. The regions at different resolutions can be reviewed using the Juicebox Browser (31,32). Aggregate Peak Analysis (APA) was performed by juicer with default parameters (33). The contact probability along with the genomic distance was counted by FitHiC with parameters -r 10000 -U 100 000 000 -L 10 000. Then the resulting contact probability was plotted against distance in a log-log plot (34). For downsampling, same amount of sequencing read pairs were sampled by Seqtk (<https://github.com/lh3/seqtk>). Loop detection was performed by HiCCUPS software (Default parameter: -r 5000, 10 000, 25 000) (25). The QuASAR-QC quality assessment was calculated by QuASAR software (35,36).

RESULTS AND DISCUSSION

Evaluation and qualification tests of different exonuclease combinations

In the present study, we employed different exonuclease combinations and designed a series of tests to remove different linear DNAs. This combination included LRL, LRC, LIC or I+IIIIC. To enhance the procedure, plasmid DNA was used for these tests, as it exists in three forms, circular DNA (SC), nicked circular DNA (OC) and linearized DNA (Lin) (Supplementary Figure S1B and S2A) (37). First, linearized plasmid DNA (pGL4.23, 4283 bp) was obtained by BamHI-HF digestion (Figure 2B and Supplementary Figure S2B). The band in the lane named 'Plasmid Lin' shows that the plasmid was completely linearized (Figure 2B, C and Supplementary Figure S2). For better simulation of a ligation solution containing circularized and linear DNAs, a circular plasmid was mixed with an equal quantity of linearized plasmid (called Mixture) (Figure 2B and C). Six treatments with various exonuclease combinations, or singular exonuclease, were assayed for linearized DNA elimination (Figures 2B-C, 3, Supplementary Tables S1 and S4, see Materials and Methods). The results showed that, although the mixture contained circular DNA, nicked circular DNA and linearized DNA, the LRL, LRC or LIC treatments could successfully remove linearized DNA (Figure 2B). LRL was consistent with a previously report (37).

In order to quantitatively assess the efficiency of elimination, DNA concentrations of digestion products were measured by Qubit (Figure 3A and Supplementary Table S4). The removal ratios of the mixtures were 61.4% (1-X: LRL-Mix), 66.6% (2-X: LRC-Mix) and 65.0% (3-X: LIC-Mix) (Supplementary Table S4). Qubit results were consistent with qPCR experiments (SV) (Figure 3B SV and Supplementary Figure S5, see Materials and Methods). To assess the subtle differences of the linearized elimination products, qPCR amplification was conducted with equal quantity (SQ) of templates, using primers either located across (BH2) or located away (B or C) from the BamHI restriction site (Figure 3B SQ and Supplementary Figure S3).

The concentrations determined using Qubit indicated that the removal ratio for LRC was higher (66.6%) than that obtained for LRL or LIC. Likewise, the BH2 primer-qPCR result (SQ) showed that the circular DNA ratio of LRC (2-X: 2.7-fold) was higher than that of LRL (1-X: 1.7-fold) and LIC (3-X: 1.7-fold). Our qPCR results with the B or C primers also showed this consistent tendency (Supplemental S.1.2 and Supplementary Figure S3). When pure linearized DNA was used as the substrate, after 1 h digestion, the elimination efficiency (Qubit result) was 99.0% (Figure 3A and Supplementary Table S4). In the same way, the pure linearized DNA elimination treatment also showed very low RQ (<0.001, Figure 3B) in qPCR. For cost efficiency, Exonuclease I could be employed as the CutSmart would serve as a universal buffer.

Exonuclease III acts at nicks, within dsDNA, to produce single-strand gaps, and so Exonuclease I+III can remove nicked circular and linearized DNA (Supplementary Table S1). Thus, we used this I+III combination for a 1 h incubation period, and purified circular dsDNA was then obtained (Figure 3C SQ) (Figures 2C: 4-X, 3C: 4-X and Supplementary Figure S6). But during these tests, we observed that Exonuclease I and III could also remove circular DNA, if they were over-digested for longer periods (Supplementary Figure S6). Hence, to better measure the efficiency of this process, we performed I+III-Mixture tests for different digestion times (Supplementary Figure S6). The electrophoresis results showed that 1 h was sufficient to remove nicked circular and linearized DNA. The left DNA ratio was 35.0% (1 h), 11.7% (4 h) and 3.2% (16 h) (Supplementary Table S5), which indicated a decline tendency along with prolonged time. As observed, a 16 h digestion treatment could remove nearly all the different DNA forms, with only a few DNAs being retained, which could be measured by Qubit (Supplementary Figure S6). Therefore, a I+III treatment could be regarded as one means to purify circular DNA from a mixture (in this test, ~1 h). However, performance of a pre-test to establish an appropriate digestion time and conditions would avoid the risk of losing informative circularized DNA for C-technology assay.

In addition, the elimination tests (Figure 2C) indicated that the efficiency of an exonuclease combination was better than a singular exonuclease treatment (Supplementary Table S1, Supplemental S.1.6). When 1.5-fold Lambda was added into LRC or LIC, Lambda could only remove linearized dsDNA but not ssDNA (Supplementary Figure S7). Therefore, combinatorial exonuclease treatments have po-

tential to optimize the C-technologies (Supplementary Table S1).

Eliminating linear noises of C-technologies

To assess the effect of combinatorial exonuclease treatments in C-technologies, high input (572 ng) and low input (40 ng) DNA libraries were constructed from C2C12 cells (Supplemental S.2). Here, exonuclease combinations LRL, LRC, LIC and I+III were used to remove linear noise. Agarose gel electrophoresis results (Figure 4A) showed that LRL (1), LRC (2) and LIC (3) groups successfully reduced linear DNA, after 1 h digestion, whereas the I+III (4) group over digested the library. For the high input library, the yield of remaining DNA from these four linear DNA elimination treatments were measured (Supplementary Table S7): the ratios were 34.2% (2: LRC), 23.0% (3: LIC), 18.2% (1: LRL) and 16.8% (4: I + III) (Figure 4B). These results were similar to those obtained with the low input library (Supplementary Table S11): here, the ratios were 9.3% (LRC), 8.1% (LIC), 6.0% (LRL) and 6.5% (I+III).

In these assays, after linear elimination it was difficult to detect the purified DNA, by agarose gel electrophoresis (Supplementary Figure S9B and C). Thus, we introduced circular plasmid into the library, as an internal reference (Figure 4C, Supplementary Figures S8B and S9D). An equal quantity of plasmid (see Materials and Methods) was added into these two replicates (572 and 40 ng of proximity-ligation library). LRLP1(1'), LRCP1(2'), LICP1(3') and I+IIICP1(4') groups were set as elimination treatments and CoP1(C') was set as control for the 40 ng library (Figure 4C). The electrophoresis results showed that, for the 572 ng library tests, LRLP1, LRCP1 and LICP1 had a consistent effect, whereas the library treated with I+IIICP1 was over-digested (Figure 4A and Supplementary Figure S8B). The ratios for these four treatments were 53.6% (LRLP1), 44.5% (LRCP1), 40.5% (LICP1) and 30.2% (I + III) (Supplementary Table S8 and Supplementary Figure S8B). These results were similar with the low input experiments [87.5% (1': LRLP1), 52.5% (2': LRCP1), 55.4% (3': LICP1) and 34.9% (4': I+IIICP1)] (Figure 4C, D and Supplementary Table S12). The remaining ratios, with plasmid P2, were 82.4% (LRLP2), 41.9% (LRCP2), 36.7% (LICP2) and 21.5% (I+IIICP2) (Supplementary Table S13, Supplementary Figures S9D and E). These two electrophoresis-gels clearly showed circular bands of internal reference, indicating that P2 was consistent with P1 as chaperon carrier (Figure 4C and Supplementary Figure S9D).

In conclusion, LRL, LRC and LIC combinations could reliably identify interactions and displayed a high consistency between the high and low input libraries. Thus, such treatments could be used to remove linear noise and purify proximal-ligation circularized DNA in C-technologies.

Exonuclease combinations reduce noise-source of Hi-C

In standard Hi-C, valid interaction pairs ratio are low (~30% to ~40%, Figures 1B and 5A) (26,27). To evaluate noise-elimination efficiency, 'Lambda and Exonuclease I' was tested to remove linear DNAs and to further con-

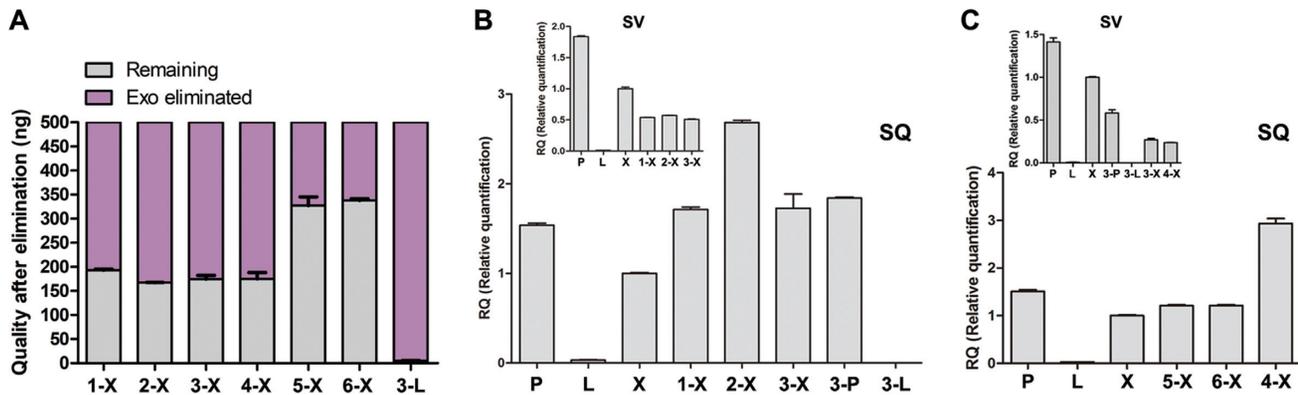


Figure 3. Quantification effects for different combination treatments. (A) Results of every linear elimination treatment, measured by Qubit. Plasmid (250 ng) and linear DNA (250 ng) were mixed, as templates (X, Mixture). Linear DNA (500 ng) (3-L, LIC-Lin) was set as control treatments. Every treatment had three replicates. 1-X, LRL-Mixture; 2-X, LRC-Mixture; 3-X, LIC-Mixture; 4-X, I+III-Mixture; 5-X, VIII4-Mixture; 6-X, VIIC-Mixture. (B) The qPCR results from three exonuclease combination treatments (in accordance with Figure 2B). The primer BH2 was used across the plasmid BamHI site. Same quality (SQ): The Qubit amounts of qPCR input for all treatments were consistent (Supplemental S.1.2). Same volume (SV): The qPCR input volumes were consistent (Supplemental S.1.4 and Supplementary Figure S5). P, pGL4.23 plasmid; L, linearized plasmid; X, Mixture; 3-P, LIC to cut plasmid; 3-L, LIC to cut linearized plasmid. Three experiments for each treatment were combined for enlarging the volume before purification. (C) The qPCR results correspond with Figure 2C. The primer was BH2. Data are presented as mean ± SEM.

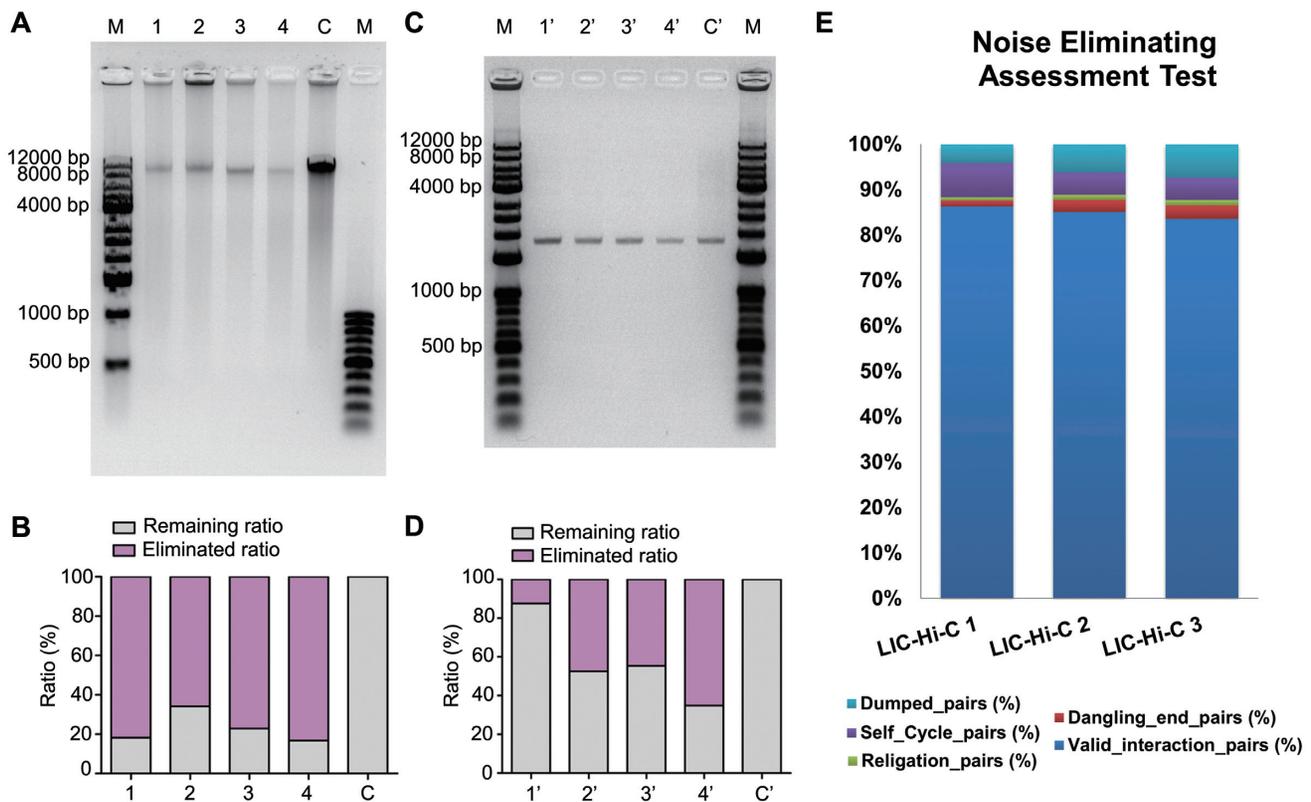


Figure 4. C-technology application and preliminary assessment in Hi-C. (A) DNA library (572 ng) linear noise elimination for C-technologies, after proximal-ligation. M (left), 1 kb DNA ladder; 1: LRL, LRL used for linear noise elimination; 2: LRC, LRC used for linear noise elimination; 3: LIC, LIC used for linear noise elimination; 4: I+III, I+III used for linear noise elimination; C: Co, without exonucleases digestion; M (right): 100 bp DNA ladder. (B) The yield ratio of remaining DNA after four linear elimination treatments, which was consistent to (A). The remaining ratio (gray) and eliminated ratio (purple) are shown together in the bar chart. (C) A 40 ng library was given four exonuclease treatment combinations, along with addition of plasmid chaperon carrier, as indicated. M: 1 kb + 100 bp DNA ladder. 1': LRLP1, LRL with P1 chaperon used for linear noise elimination; 2': LRC with P1 chaperon used for linear noise elimination; 3': LICP1, LIC with P1 chaperon used for linear noise elimination; 4': I+III, I+III with P1 chaperon used for linear noise elimination; C': P1 chaperon added, without exonucleases digestion. (D) The yield ratio of remaining DNA, after four linear DNA elimination treatments, which was consistent to (C). (E) HiC-Pro filtering results of Hi-C with exonuclease elimination LIC, compared with standard Hi-C in Figure 1B. LIC-Hi-C had three experimental replicates (LIC-Hi-C 1; LIC-Hi-C 2; and LIC-Hi-C 3). HiC-Pro terms are as indicated in Supplemental S.3.

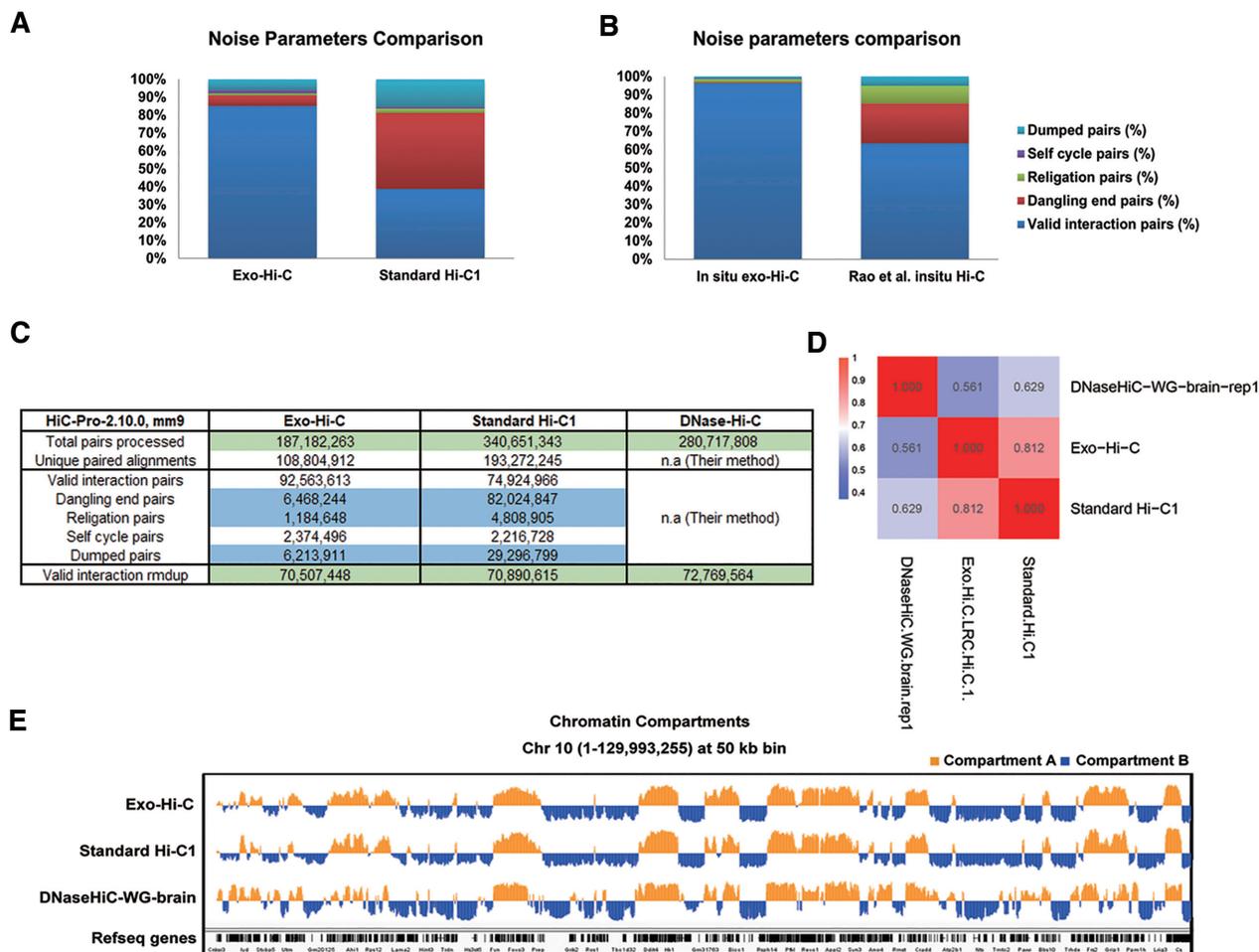


Figure 5. Exonuclease combination eliminates high-noise of standard Hi-C, meanwhile maintains a high degree of correlation. (A) HiC-Pro filtering results of Hi-C with exonuclease elimination LRC (termed as Exo-Hi-C), compared with Dixon *et al.* Standard Hi-C 1 (standard Hi-C data in mouse ES cells) (26). Valid interaction pairs of Exo-Hi-C were twice as much as that with Standard Hi-C 1. (B) HiC-Pro filtering results of *in situ* exo-Hi-C (IIIC combination) and Rao *et al.* *in situ* Hi-C (25). (C) Specific data comparison between Exo-Hi-C, standard Hi-C 1 and DNase-Hi-C. To obtain valid interaction data of near equivalent size (~71 million), Standard Hi-C 1 required ~341 million clean data. DNase-Hi-C required ~281 million clean data. Exo-Hi-C was sequenced ~187 million clean data. In detail, the noise associated with dangling end pairs, dumped pairs and religation pairs were greatly reduced. Valid interaction pairs mean valid pairs and unique paired alignments mean total pairs, which were consistent with Supplementary Table S14. (D) The correlation between Exo-Hi-C, Standard Hi-C 1 and DNase Hi-C, according to interaction frequency at 50 kb resolution. The correlation was related with cell types and exonuclease treatment did not alter the cell feature. Exo-Hi-C and Standard Hi-C 1 were from mESC cells. DNase Hi-C were from mouse brain cells. (E) Exo-Hi-C and Standard Hi-C 1 were very consistent in A/B compartment level. The slide window of chromosome 10 (chr 10) is shown. The positive PC1 value reflected A compartments (marked as orange peak). The negative PC1 regions were B compartments (marked as navy blue peak). The Refseq Genes were shown correspondingly.

struct Hi-C libraries (Figure 1A). By this treatment, the noise ratio was significantly decreased. As shown in Figure 4E, LIC efficiently reduced the ratio of dangling end pairs to 2.3%, compared with a T4 DNA polymerase treatment (30.8%) (Figure 1B). The wrong size (dumped pairs) fragment ratio was also reduced (from 20.9% to 5.9%). At the same time, a small increase was observed in self-ligation reads, which could be removed by bioinformatics analysis (11). Thus, LIC could efficiently reduce various linear noise in Hi-C and increased the ratio of circularized interaction DNA (from ~40% to ~80%) (Figure 4E).

Potential application to C-technologies

To further evaluate the elimination effect, we next prepared

mESCs Hi-C libraries (38) with ‘Lambda and RecJF’ combination, herein referred to as Exo-Hi-C (Figure 5A and Supplementary Table S14). Standard Hi-C 1 was downloaded from previous work by Dixon *et al.* (26). Mapping and filtering processes were performed by HiC-Pro pipeline (12). The ratios of valid interaction pairs of Exo-Hi-C (85.0%) was increased compared to Standard Hi-C 1 (39.0%) (Figure 5A). The noise for dangling end pairs (decreased from 82.02 to 6.47 million), dumped pairs (decreased from 29.30 to 6.21 million) and religation pairs (decreased from 4.81 to 1.18 million) were dramatically reduced (Figure 5C). In addition, exonuclease elimination did not significantly alter the ratios of cis and trans interaction in Hi-C (Supplementary Figure S11). Through Empirical Cumulative Distribution analysis and correlation analysis,

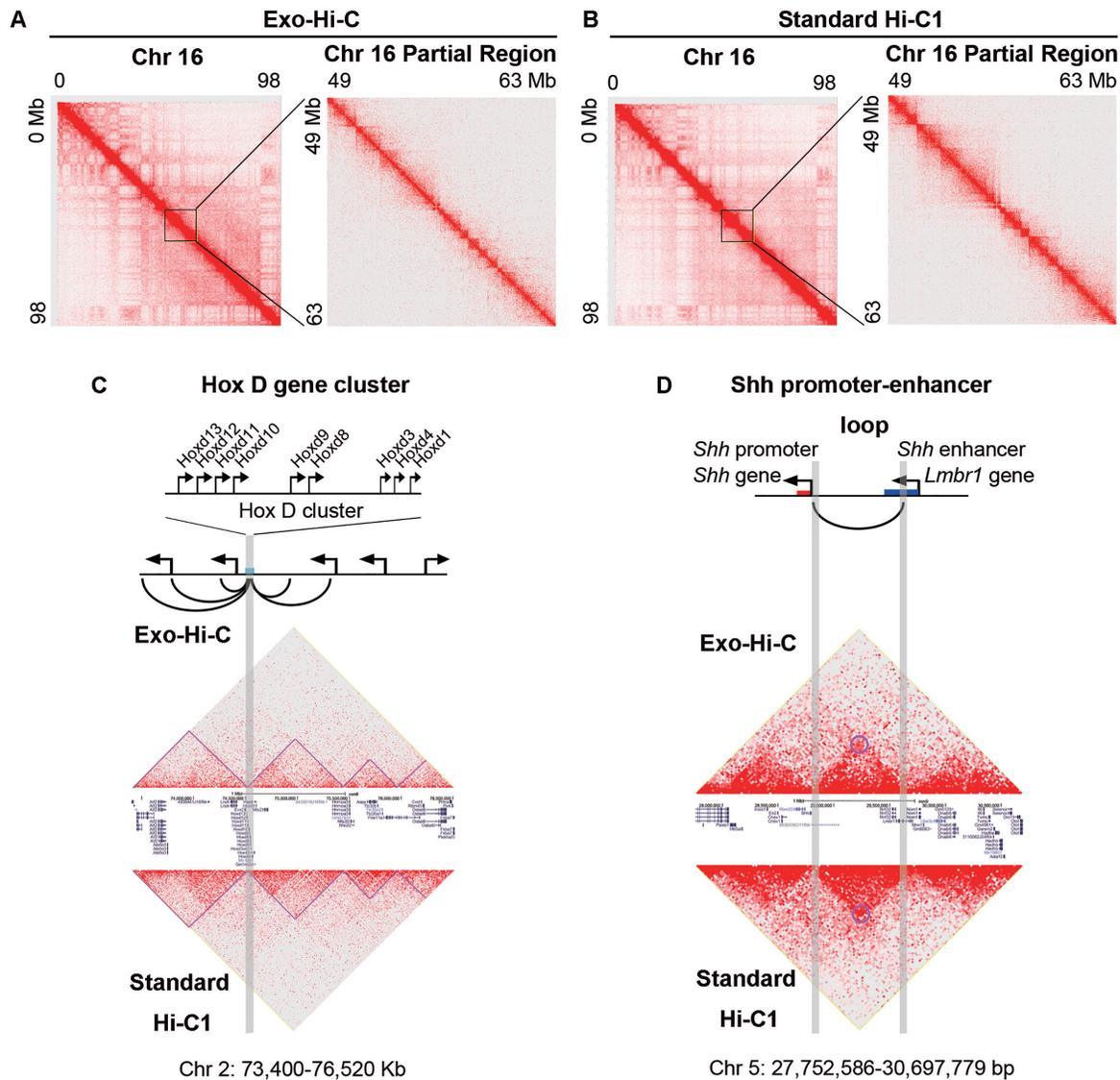


Figure 6. Hierarchical structures were compared between Exo-Hi-C and standard Hi-C, which were robust in Exo-Hi-C after the noise elimination. (A) The Knight-Ruiz (KR) normalized Hi-C interaction frequency heatmaps of chromosome 16 (250 kb bin) and its partial region 49–63 M (25 kb) in mESCs from Exo-Hi-C. (B) The KR normalized Hi-C interaction frequency heatmaps of chromosome 16 (250 kb bin) and its partial region 49–63 M (25 kb) in mESCs from Standard Hi-C 1. (C) *Hox D* gene cluster locus was zoomed-in at chromosome 2; 73.4–76.5 M (10 kb bin) on the normalized Hi-C interaction frequency heatmaps. The sub-TADs and TADs surrounding the *Hox D* were distinctly displayed (purple line). The upside was Exo-Hi-C and the downside was Standard Hi-C 1. The middle Track was RefSeq Genes. The French grey lane highlighted the *Hox D* gene cluster. The loop symbols conceptualized the interactions between *Hox D* genes and their regulatory elements in the flanking domains. (D) The enhancer-promoter loop of *Shh* gene interaction (chromosome 5; 27.75–30.70 M, at 25 kb bin) was zoomed-in. Exo-Hi-C displayed the robust loop structure as standard Hi-C. The French grey lanes highlighted the *Shh* gene promoter (left) and enhancer (right, within an intron of the *Lmbr1* gene) regions.

we could infer that combinatorial exonuclease treatments primarily removed noise, and thus, would not likely influence valid interactions and its related characteristics (Figure 5D, Supplementary Figure S12). We then performed QuASAR at 40 kb and 1 Mb resolution (35,36). The results showed that Exo-Hi-C and Standard Hi-C were comparable in quantity (QuASAR-QC scores) at both resolutions (Supplementary Figure S13).

We further compared global chromatin interaction patterns and a series of defined structures at different resolutions. The A/B compartments were representatively showed on chromosome 10. Exo-Hi-C and Standard Hi-C 1 showed

consistency at global regions (Figure 5E). To further probe different levels of chromatin conformation, we generated Hi-C interaction frequency heatmaps, at various chromatin and their regions. With 250, 100 and 25 kb resolution, both Exo-Hi-C and Standard Hi-C 1 showed similar interaction grids (Figure 6A, B and Supplementary Figure S14–S16) (39). We also tested several reported gene clusters, or DNA-loops, for validation. For example, *Hox A* (a1–a13) and *Hox D* (d1–d13) gene clusters are similar in gene regulation transform patterns and are regulated by long-range enhancers in gene-poor regions or in sub-TADs, or cross the nearby TADs (40–45). At 10 kb resolution, interaction

heatmaps of Exo-Hi-C successfully demonstrated *Hox D* (Figure 6C) or *Hox A* gene clusters (Supplementary Figure S17), accompanying their surrounding domains. Compared to DNase Hi-C, the Exo-Hi-C provided more distinct and informative interaction information (Supplementary Figure S18). We also explored the *Shh* gene-related region (46). It was reported that disruption of an enhancer, in an intron of the *Lmbr1* gene, led to *Shh* gene misexpression and a polydactyl phenotype (46,47). This enhancer is 1 Mb away from the *Shh* gene (47), and the Exo-Hi-C could detect this enhancer-promoter loop and showed consistent result to that obtained using Standard Hi-C 1 (Figure 6D and Supplementary Figure S18).

We generated contact probability (log)-Distance (Mb) plots for Exo-Hi-C, standard Hi-C and DNase Hi-C (Supplementary Figure S19) (48). As shown, $P(s) \sim s^{-1}$ is a black solid line, which is following a fractal globule model function and its contact probability own decaying characteristic of prediction(10). $P(s) \sim s^{-0.5}$ is a black dashed line, whose trends reflect predicted mitotic states (30). Relative contact probability plots (500 kb–7 Mb) demonstrated that the features of Hi-C were well-preserved during noise elimination. We performed aggregate peak analysis (APA) (Supplementary Figure S20), which displays the aggregate enrichment from an entire set of putative DNA-loops in a contact matrix and can be used to evaluate signal aggregate degree. Exo-Hi-C showed marked visual enrichment at the center of these plots. Compared with the previous methods, Exo-Hi-C achieved desirable signal aggregation, which met the characteristic of Hi-C aggregation peaks.

A similar approach, this elimination principle may also be used to improve *in situ* Hi-C after *in situ* ligation (25). For *in situ*, *in situ* exo-Hi-C was specifically introduced by linear DNA eliminating step after proximity-ligation basing on Rao *et al.* *in situ* Hi-C (Supplementary Figure S21). Here, three replicates for 2 types (IIIIC and LIC) of exonuclease combinations were used to test the effect of elimination. The *in situ* Hi-C was the control group prepared by us and Rao *et al.* *in situ* Hi-C was downloaded (25). Compared with the Rao *et al.* datasets, *in situ* exo-Hi-C yielded much lower noise levels and higher valid interaction pairs ratio (Figure 5B and Supplementary Figure S22). Furthermore, both combinations showed higher consistency and more validated pairs than those of *in situ* Hi-C datasets (Supplemental S.4) (25). The related assessment was also shown in the Supplementary material (Supplementary Figure S23–S25) (49). After sampled same amount of sequencing read pairs, HiCCUPS, QuASAR-QC and APA were also applied to analysis. For *in situ*, the loop numbers of different groups was showed as IIIIC-*in situ* Hi-C > LIC-*in situ* Hi-C > *in situ* Hi-C (control) > Rao *et al.* *in situ* Hi-C. The difference between *in situ* exo-Hi-C and *in situ* Hi-C reached 3–6-fold (Supplementary Figure S26). These should be compelling. QuASAR-QC scores of *in situ* exo-Hi-C were also higher than control groups. Concretely, the situation was IIIIC-*in situ* Hi-C > LIC-*in situ* Hi-C > *in situ* Hi-C (control) > Rao *et al.* *in situ* Hi-C (Supplementary Figure S27), which provided evidence that *in situ* exo-Hi-C was less noisy. Furthermore, we also used APA to evaluate. *In situ* exo-Hi-C (IIIIC-*in situ* Hi-C and LIC-*in situ* Hi-C) reflected the good signal aggregate degree similar with control and previous

Rao *et al.*'s groups (Supplementary Figure S28). Therefore, our strategy is effective in decreasing the noise level.

In conclusion, C-technologies are frequently employed technologies for 3D genomics. For these high-noise methodologies, removing undesired linear DNAs, generated by cyclization steps, are critical to reduce their noise sources, as this will be beneficial for downstream 3D interaction identification. In this study, we developed four strategies (LRL, LRC, LIC and IIIIC) to eliminate linear DNA noise, which could increase the yield of valid-pairs (valid templates) for C-technologies. By using these three exonuclease combinations, we successfully optimized Hi-C technology, especially in dilution Hi-C from ~40% to ~80%. Our technique has high potential to be applied to a wide range of C-technologies, such as 3C, 4C (1), 5C, and ChIP combining 3C (ChIP-Loop), because 3C, 4C, 5C, Capture-Hi-C, etc., are dilution based technologies and these technologies are high-noise ones (Supplementary Figure S29) (2,4,8,50). Therefore, adjustment of a pivotal step of C-technologies or the data improvement, even a minor one, would have long term effects on related researches. The high-efficiency, convenience and low cost of our methodology make it very promising application in the field of 3D genomics.

DATA AVAILABILITY

The datasets (Standard Hi-C 1, 2 and 3) analyzed during the current study are available in the GEO repository (GSM862721, GSM1718027 and GSM862720). The datasets (DNase-Hi-C and *in situ* DNase-Hi-C) analyzed during the current study are available in the GEO repository (GSM1370433, GSM1370434, GSM1689794, GSM1689795, GSM1689796 and GSM1689797). The datasets (Rao *et al.* *in situ* Hi-C) analyzed during the current study are available in the GEO repository (SRR1658716, SRR1658722 and SRR1658724). The datasets generated and analyzed during the current study are accessible through GEO Series accession number GSE125656. HiC-Pro filtering results are listed in Supplement Tables S1 and S2.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Author contributions: Y.Z. conducted and designed the experiments. S.K., G.Z. and Q.L. performed the experiments. Q.L. and Q.H. conducted the bioinformatics work. L.H. and Y.P. participated in discussions and gave good advice. H.Z., Y.H. and B.Q. provided some important experiment materials. The manuscript was written by S.K. and Y.Z. All authors read and approved the final manuscript. We got the help from Prof. William J. Lucas, UC Davis to revise our manuscript.

FUNDING

Thousand Talents Plan for Young Professionals [to Y.Z.]; National Natural Science Foundation of China

[31970592]; National Key Research and Development Program of China [2018YFA0903201]; Agricultural Science and Technology Innovation Program; Fundamental Research Funds for Central Non-profit Scientific Institution [Y2017CG26]; Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission [CAAS-XTX2016001-3]; Elite Young Scientists Program of Chinese Academy of Agricultural Sciences [CAASQNYC-KYYJ-41]; Special fund for industrial development in Dapeng New District [KY20180116]. Funding for open access charge: Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences.

Conflict of interest statement. None declared.

REFERENCES

- Schmitt, A.D., Hu, M. and Ren, B. (2016) Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.*, **17**, 743–755.
- Denker, A. and De, L.W. (2016) The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.*, **30**, 1357–1382.
- Borensztein, M., Monnier, P., Court, F., Louault, Y., Ripoché, M.A., Turet, L., Yao, Z., Tapscott, S.J., Forne, T., Montarras, D. *et al.* (2013) Myod and H19–Igf2 locus interactions are required for diaphragm formation in the mouse. *Development*, **140**, 1231–1239.
- Yu, M. and Ren, B. (2017) The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.*, **33**, 265–289.
- Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Włodarczyk, J., Ruszczycycki, B. *et al.* (2015) CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
- Gavrilov, A.A., Gushchanskaya, E.S., Strelkova, O., Zhironkina, O., Kireev, I.I., Iarovaia, O.V. and Razin, S.V. (2013) Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res.*, **41**, 3563–3575.
- de Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev.*, **26**, 11–24.
- Dekker, J. (2006) The three ‘C’ s of chromosome conformation capture: controls, controls, controls. *Nat. Methods*, **3**, 17–21.
- Kong, S. and Zhang, Y. (2019) Deciphering Hi-C: from 3D genome to function. *Cell Biol. Toxicol.*, **35**, 15–32.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
- Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P. and Andrews, S. (2015) HiCUP: pipeline for mapping and processing Hi-C data [version 1; peer review: 2 approved, 1 approved with reservations]. *F1000research*, **4**, 1310–1310.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C., Vert, J., Heard, E., Dekker, J. and Barillot, E. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.*, **16**, 259–259.
- Grob, S. and Cavalli, G. (2018) Technical review: A hitchhiker’s guide to chromosome conformation capture. *Methods Mol. Biol.*, **1675**, 233–246.
- Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., van den Heuvel, A., Kockx, C., Palstra, R.J., Wendt, K.S., Grosveld, F., van Ijcken, W. *et al.* (2013) Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat. Protoc.*, **8**, 509–524.
- Oudelaar, A.M., Davies, J.O.J., Downes, D.J., Higgs, D.R. and Hughes, J.R. (2017) Robust detection of chromosomal interactions from small numbers of cells using low-input Capture-C. *Nucleic Acids Res.*, **45**, e184.
- Niu, L., Shen, W., Huang, Y., He, N., Zhang, Y., Sun, J., Wan, J., Jiang, D., Yang, M., Tse, Y.C. *et al.* (2019) Amplification-free library preparation with SAFE Hi-C uses ligation products for deep sequencing to improve traditional Hi-C analysis. *Commun Biol.*, **2**, 267.
- Subramanian, K., Rutvisuttinunt, W., Scott, W. and Myers, R.S. (2003) The enzymatic basis of processivity in lambda exonuclease. *Nucleic Acids Res.*, **31**, 1585–1596.
- Lovett, S.T. and Kolodner, R.D. (1989) Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of *Escherichia coli*. *PNAS*, **86**, 2627–2631.
- Okamoto, N., Lee, A., Kano, T. and Lee, T.D. (1994) Homogeneous fluorescence detection method for human leukocyte antigen-DR typing following polymerase chain reaction amplification with sequence-specific primer. *Anal. Biochem.*, **221**, 340–347.
- Chen, X., Lin, C., Chen, Y., Luo, F., Wang, Y. and Chen, X. (2016) Terminal protection of a small molecule-linked loop DNA probe for turn-on label-free fluorescence detection of proteins. *Biosens. Bioelectron.*, **83**, 97–101.
- Huang, H., Shi, S., Zheng, X. and Yao, T. (2015) Sensitive detection for coralyne and mercury ions based on homo-A/T DNA by exonuclease signal amplification. *Biosens. Bioelectron.*, **71**, 439–444.
- Serandour, A.A., Brown, G.D., Cohen, J.D. and Carroll, J.S. (2013) Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome Biol.*, **14**, 1–9.
- Ying, Q., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B.W., Woodgett, J.R., Cohen, P. and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J.F., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W. and Shipley, G.L. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D. and Lander, E.S. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
- Nagano, T., Várnai, C., Schoenfelder, S., Javierre, B.M., Wingett, S.W. and Fraser, P. (2015) Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.*, **16**, 175–175.
- Ursu, O., Boley, N., Taranova, M., Wang, Y.X.R., Yardimci, G.G., Noble, W.S. and Kundaje, A. (2018) GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics*, **34**, 2701–2707.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
- Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X., He, J., Xiang, Y., Wang, Q., Li, Y. *et al.* (2017) Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature*, **547**, 232–235.
- Durand, N.C., Robinson, J., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S. and Aiden, E.L. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.*, **3**, 99–101.
- Robinson, J., Turner, D., Durand, N.C., Thorvaldsdottir, H., Mesirov, J.P. and Aiden, E.L. (2018) Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.*, **6**, 205740.
- Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S. and Aiden, E.L. (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.*, **3**, 95–98.
- Ay, F., Bailey, T.L. and Noble, W.S. (2014) Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.*, **24**, 999–1011.
- Yardimci, G.G., Ozadam, H., Sauria, M.E.G., Ursu, O., Yan, K.-K., Yang, T., Chakraborty, A., Kaul, A., Lajoie, B.R., Song, F. *et al.* (2019) Measuring the reproducibility and quality of Hi-C data. *Genome Biol.*, **20**, 57–57.
- Sauria, M.E.G. and Taylor, J. (2017) QuASAR: quality assessment of spatial arrangement reproducibility in Hi-C data. bioRxiv doi:

- <https://doi.org/10.1101/204438>, 14 November 2017, preprint: not peer reviewed.
37. Balagurumoorthy,P., Adelstein,S.J. and Kassis,A.I. (2008) Method to eliminate linear DNA from mixture containing nicked circular, supercoiled, and linear plasmid DNA. *Anal. Biochem.*, **381**, 172–174.
 38. Dekker,J., Belmont,A.S., Guttman,M., Leshyk,V.O., Lis,J.T., Lomvardas,S., Mirny,L.A., O’Shea,C.C., Park,P.J., Ren,B. *et al.* (2017) The 4D nucleome project. *Nature*, **549**, 219–226.
 39. Ay,F. and Noble,W.S. (2015) Analysis methods for studying the 3D architecture of the genome. *Genome Biol.*, **16**, 183–183.
 40. Phillipscremins,J.E. and Corces,V.G. (2013) Chromatin insulators: linking genome organization to cellular function. *Mol. Cell*, **50**, 461–474.
 41. Woltering,J.M., Noordermeer,D., Leleu,M. and Duboule,D. (2014) Conservation and divergence of regulatory strategies at Hox loci and the origin of tetrapod digits. *PLoS Biol.*, **12**, e1001773.
 42. Lonfat,N., Montavon,T., Darbellay,F., Gitto,S. and Duboule,D. (2014) Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science*, **346**, 1004–1006.
 43. Williamson,I., Berlivet,S., Eskeland,R., Boyle,S., Illingworth,R.S., Paquette,D., Dostie,J.E. and Bickmore,W.A. (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev.*, **28**, 2778–2791.
 44. Andrey,G., Montavon,T., Mascrez,B., Gonzalez,F., Noordermeer,D., Leleu,M., Trono,D., Spitz,F. and Duboule,D. (2013) A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science*, **340**, 1234167–1234167.
 45. Berlivet,S., Paquette,D., Dumouchel,A., Langlais,D., Dostie,J. and Kmita,M. (2013) Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS Genet.*, **9**, e1004018.
 46. Lettice,L.A., Horikoshi,T., Heaney,S.J.H., Van Baren,M.J., Der Linde,H.C.V., Breedveld,G.J., Joosse,M., Akarsu,N., Oostra,B.A. and Endo,N. (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *PNAS*, **99**, 7548–7553.
 47. Smallwood,A. and Ren,B. (2013) Genome organization and long-range regulation of gene expression by enhancers. *Curr. Opin. Cell Biol.*, **25**, 387–394.
 48. Ma,W., Ay,F., Lee,C., Gulsoy,G., Deng,X., Cook,S., Hesson,J., Cavanaugh,C., Ware,C.B. and Krumm,A. (2015) Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat. Methods*, **12**, 71–78.
 49. Deng,X., Ma,W., Ramani,V., Hill,A., Yang,F., Ay,F., Berletch,J.B., Blau,C.A., Shendure,J., Duan,Z. *et al.* (2015) Bipartite structure of the inactive mouse X chromosome. *Genome Biol.*, **16**, 152–152.
 50. Gavrilov,A.A., Golov,A.K. and Razin,S.V. (2013) Actual ligation frequencies in the chromosome conformation capture procedure. *PLoS One*, **8**, e60403.