*Published online 4 July 2008*

*Nucleic Acids Research, 2008, Vol. 36, No. 15* ***e92***
*doi:10.1093/nar/gkn420*

# Chum-RNA allows preparation of a high-quality cDNA library from a single-cell quantity of mRNA without PCR amplification

**Takahiro Tougan, Daisuke Okuzaki and Hiroshi Nojima***

Department of Molecular Genetics, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received March 18, 2008; Revised June 7, 2008; Accepted June 18, 2008

## ABSTRACT

**Linear RNA amplification using T7 RNA polymerase is useful in genome-wide analysis of gene expression using DNA microarrays, but exponential amplification using polymerase chain reaction (PCR) is still required for cDNA library preparation from single-cell quantities of RNA. We have designed a small RNA molecule called chum-RNA that has enabled us to prepare a single-cell cDNA library after four rounds of T7-based linear amplification, without using PCR amplification. Chum-RNA drove cDNA synthesis from only 0.49 femtograms of mRNA (730 mRNA molecules) as a substrate, a quantity that corresponds to a minor population of mRNA molecules in a single mammalian cell. Analysis of the independent cDNA clone of this library ($6.6 \times 10^5$ cfu) suggests that 30-fold RNA amplification occurred in each round of the amplification process. The size distribution and representation of mRNAs in the resulting one-cell cDNA library retained its similarity to that of the million-cell cDNA library. The use of chum-RNA might also facilitate reactions involving other DNA/RNA modifying enzymes whose Michaelis constant ($K_m$) values are around 1 mM, allowing them to be activated in the presence of only small quantities of substrate.**

## INTRODUCTION

Comparison of gene-expression patterns between cells and/or tissues facilitates the identification of molecules activated by a particular physiological or pharmacological treatment. The use of gene-expression profiling is particularly important in neuroscience, clinical science, stem cell biology, and metagenomic analysis. In many cases, however, the amount of specimen tissue available is limited, allowing only small amounts of mRNA to be obtained. As such, amplification of the isolated RNA is obligatory to obtain the microgram amounts of RNA required for microarray analysis or cDNA library preparation. Without amplification, such amounts of RNA would be obtainable only from millions of cells. Although the polymerase chain reaction (PCR) is a powerful method for amplifying a single target DNA, the exponential amplification that can be achieved using multiple targets (from mixtures of DNA fragments or mRNA molecules) often produces a biased sample, since cDNAs of differing lengths and composition are amplified with differing efficiencies (1). The bias is especially conspicuous if amplified RNA is used for preparation of a cDNA library, as the library will provide an inaccurate impression of the abundance and diversity of various transcripts.

An alternative to PCR, for amplifying RNA, is the use of RNA polymerase, which is considered to generate non-biased RNA pools. The most commonly-used technique for RNA amplification is a linear amplification method first developed by Van Gelder, Eberwine and coworkers (2,3), in which small amount of RNA is primed with a synthetic oligonucleotide containing the T7 RNA polymerase promoter sequence located upstream of a polythymidylate region, and then T7 RNA polymerase is used to generate amplified antisense RNA (aRNA) after second-strand cDNA synthesis. This technique and the subsequent improved protocols, with or without combination of PCR (4–8), have allowed genome-wide microarray analysis of gene expression, using a single-cell amount of RNA as a starting material (9–12).

Notwithstanding the benefits of RNA polymerase, most cDNA library preparation from single-cell amounts of mRNA is performed using PCR amplification. Indeed, a protocol involving a nested round of cDNA synthesis and *in vitro* transcription in combination with PCR amplification (6,13,14) successfully reduces the required starting material without significantly reducing the overall sensitivity and fidelity. In contrast, use of RNA polymerase alone still requires one microgram total RNA after two-round amplification of complementary RNA (cRNA) (15).

*To whom correspondence should be addressed. Tel: +81 6 6875 3980; Fax: +81 6 6875 5192; Email: snj-0212@biken.osaka-u.ac.jp

This is mainly because the optimum concentration for most of the enzymes used in cDNA library preparation, such as reverse transcriptase (RTase), DNA polymerase I and DNA ligase, is more than 1 μM (16–18). That is, the amount of enzyme exceeds the single-cell amount of mRNA one million-fold. It should be remembered that an enzyme's rate of substrate conversion follows the Michaelis–Menten equation, where the Michaelis constant ($K_m$) is equivalent to the substrate concentration at which the rate of conversion is half of the maximum rate of conversion. Thus, if the $K_m$ value of an enzyme is higher than the practical concentration, the rate of substrate conversion becomes very slow.

The use of cDNA libraries retains some important advantages over microarray analysis, allowing the application of expression cloning techniques, identification of poly(A)-bearing non-coding RNA, expression analysis of infected unknown virus, and more. As such, we have investigated ways of circumventing the '$K_m$ problem'. We have designed a small dummy substrate termed chum-RNA, in which 'chum' refers to the 'friendly' action of this RNA in the reaction. Chum-RNA can be added to the reaction mixture to increase the effective quantity of substrate, thus increasing the substrate conversion rate of the enzyme. The chum-RNA can be easily removed following the completion of the reaction. In the present study, we show that the simple addition of chum-RNA allows cDNA synthesis to be initiated in the presence of less than 1 femtogram (fg) of mRNA, which equates to the concentration of a low copy number mRNA species in a single human cell. Using chum-RNA, we could for the first time synthesize sense RNA (sRNA) by RNA amplification, without the aid of PCR amplification. Use of chum-RNA aided linear RNA amplification, allowing us to successfully construct a high quality cDNA library that preserved the mRNA size distribution pattern seen in a cDNA library prepared from one million cells. The simplicity of the technique makes it a versatile strategy for reducing the quantity of starting materials required in a variety of methods, not only cDNA library preparation.

## MATERIALS AND METHODS

### Preparation of a cDNA library from a single-cell quantity of mRNA using chum-RNA

Chum-RNA was prepared by Gene Design Inc. (Osaka, Japan) by synthesizing and HPLC purifying the following oligonucleotide (Figure 1); 5′-AAU UCG UCU GGA CAC GAA AAA AAA AAA AAA AAA AAA AA-3′. The mRNA used for cDNA library construction was purified from about one million growing 293T cells using a Micro Poly(A) Purist kit (Applied Biosystems, Foster City, CA). The cDNA libraries were then constructed using a quantity of mRNA equivalent to that obtained from one cell or one million cells, using the linker-primer method with a pAP3neo vector, as described previously (19). Briefly, for construction of a single-cell cDNA library, the process of cDNA synthesis and mRNA amplification was repeated four times (Figure 1B, steps 1–5). For preparation of the one million cell cDNA library, the mRNA amplification process (Figure 1B, steps 3–7) was omitted. The amplification adaptor indicated in step 3, containing the T7 DNA-dependent RNA polymerase promoter (indicated by underlines), was prepared by annealing the following oligonucleotide pairs; sense T7 5′-CAC TAG TAC GCG TAA TAC GAC TCA CTA TAG GGA ATT CCC CGG G-3′; anti-sense T7 5′-pCCC GGG GAA TTC CCT ATA GTG AGT CGT ATT ACG CGT ACT AGT GAG CT-3′. The 5′ terminal of the anti-sense T7 sequence was phosphorylated to avoid unwanted ligation. Equal molar amounts of these oligonucleotides were mixed to a final concentration of 0.35 μg/μl in 10 mM Tris–HCl (pH 7.5) with 1 mM EDTA and 10 mM $MgCl_2$, warmed at 65°C for 2 min, then cooled at 37°C for 2 min and held at room temperature for 5 min. The resulting product can be stored at –20°C for several years. In step 5, an *in vitro* transcription is performed using a MEGAscript High Yield Transcription Kit (Applied Biosystems). A more detailed protocol for the preparation of a single-cell cDNA library using chum-RNA is described in the Supplementary Data Materials and Methods.

### RT–PCR

The efficiency of cDNA synthesis in the presence or absence of chum-RNA was confirmed by RT–PCR of *Homo sapiens* glyceraldehyde 3-phosphate dehydrogenase (HsGAPDH) mRNA using the following primer set which detects a band 902 base pairs (bp) in size; forward (HsGAPDH-F) 5′-CGA GAT CCC TCC AAA ATC AA-3′ and reverse (HsGAPDH-R) 5′-AGG GGT CTA CAT GGC AAC TG-3′. The annealing temperature for PCR was always at 50 or 55°C, while the number of amplification cycles was 30, 40 or 50.

HsGAPDH cDNA was amplified by PCR from 293T cDNA using the following primer set; forward (HsGAPDH-clone-F) 5′-ACA GTC AGC CGC ATC TTC TT-3′ and reverse (HsGAPDH-clone-R) 5′-TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT GGT TGA GCA CAG GGT ACT TTA TTG-3′. The amplified DNA fragment (~1.3 kb in length) was cloned into the pT7-Blue vector (Invitrogen, San Diego, CA) using the TA-cloning method, and its DNA sequence was confirmed. The plasmid was linearized by digestion with *Bam*HI, and then GAPDH mRNA was expressed in vitro using a MEGAscript High Yield Transcription Kit (Applied Biosystems) at 37°C for 4 h. The residual DNA fragment in the reaction mixture was digested by DNase treatment using TURBO DNA-free (Applied Biosystems) at 37°C for 1 h, and the successful removal of DNA was confirmed by PCR.

### Quality check of the single-cell cDNA library

To examine the quality of the single-cell cDNA library, we prepared plasmid DNA from 60 randomly selected clones, and then digested these using *Bam*HI to examine the size distribution of the inserted cDNA. DNA sequences were also determined and inserts were identified using a homology search with the BLASTN algorithm (http://www.ncbi.nlm.nih.gov).

### Microarray analysis

The plasmids carrying the single-cell and million-cell cDNA libraries were linearized by digestion with *Not*I, and their contents expressed *in vitro* using a MEGAscript High Yield Transcription Kit (Applied Biosystems) at 37°C for 4 h. The resulting mRNAs were independently reverse-transcribed using oligo-dT primers containing the T7 RNA polymerase promoter sequence to generate cDNAs, which were then subjected to *in vitro* transcription using T7 RNA polymerase to label the complementary RNAs (cRNAs) with Cy3-CTP or Cy5-CTP (Amersham Pharmacia Biotech, Piscataway, NJ). The Cy-labeled cRNAs from the single-cell cDNA library (1 μg) were then mixed with an equal amount of reverse color Cy-labeled cRNAs derived from the million-cell cDNA library. Hybridizations, rinsing, scanning and gene analysis on the Agilent's all human cDNA microarray (Hu44K) were conducted according to the manufacturer's protocol (G2940BA; Agilent Technologies Inc., Palo Alto, CA). Fluorophore reversal (dye swap) duplicates were used in two-color DNA microarray experiments.

## RESULTS

### Principle and design of Chum-RNA

Since the $K_m$ values for most of the enzymes used in the synthesis of cDNA exceed 1 μM, it is almost impossible to synthesize cDNA from very small quantities of mRNA, e.g. from a single mammalian cell, without the aid of PCR amplification. To circumvent this '$K_m$ problem', we conceived the idea of using mRNA-like small RNA molecules as dummy substrate, called 'chum-RNA' hereafter (Figure 1A). It is expected that the enzymes would 'bite' not only the large number of added chum-RNA but also the very small number of *bona fide* target mRNA molecules that are already included in the reaction mixture, and would initiate cDNA synthesis more rapidly than if the chum-RNA was not added (Figure 1B, step 1).

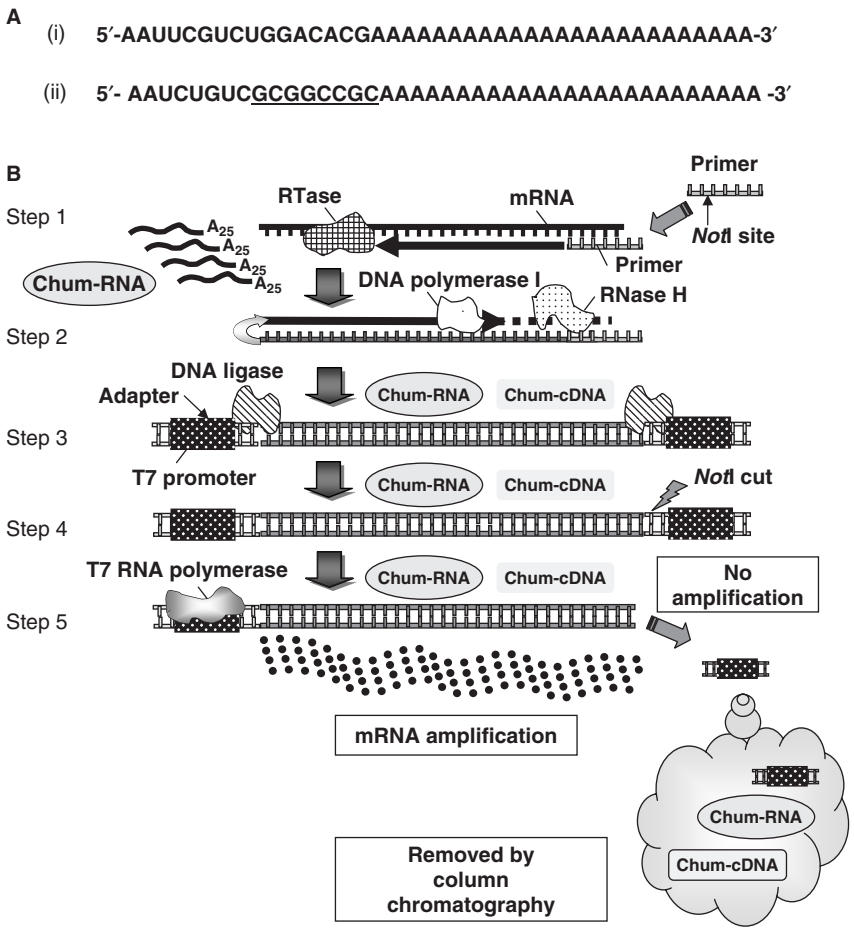


**Figure 1.** Structure of chum-RNA, and its usage for sense strand mRNA amplification and a single-cell cDNA library preparation. (**A**) Structure of typical chum-RNA molecules. (i) Sequence of the chum-RNA molecule used in this experiment. The 41-nucleotide RNA fragment consists of 16 randomly selected nucleotides followed by a 25-nucleotide poly-A tail, allowing an oligo-dT primer to associate and assemble with RTase to initiate cDNA synthesis. (ii) An alternate version of chum-RNA, carrying a *Not*I restriction site (underlined). (**B** and **C**) Schematic depiction of the procedure used for sense strand mRNA amplification (steps 1–5) and single-cell cDNA library preparation (steps 1–9). Chum-RNA is present in the reaction mixture throughout, and some of the chum-RNA may be turned into chum-cDNA during the procedure. Subsequently, the chum-RNA, chum-cDNA and the adapter that was cut off by *Not*I are removed by column chromatography at the end of steps 5 and 9. DNA is depicted by a grey line, while mRNA is depicted by a black line. See text for details.

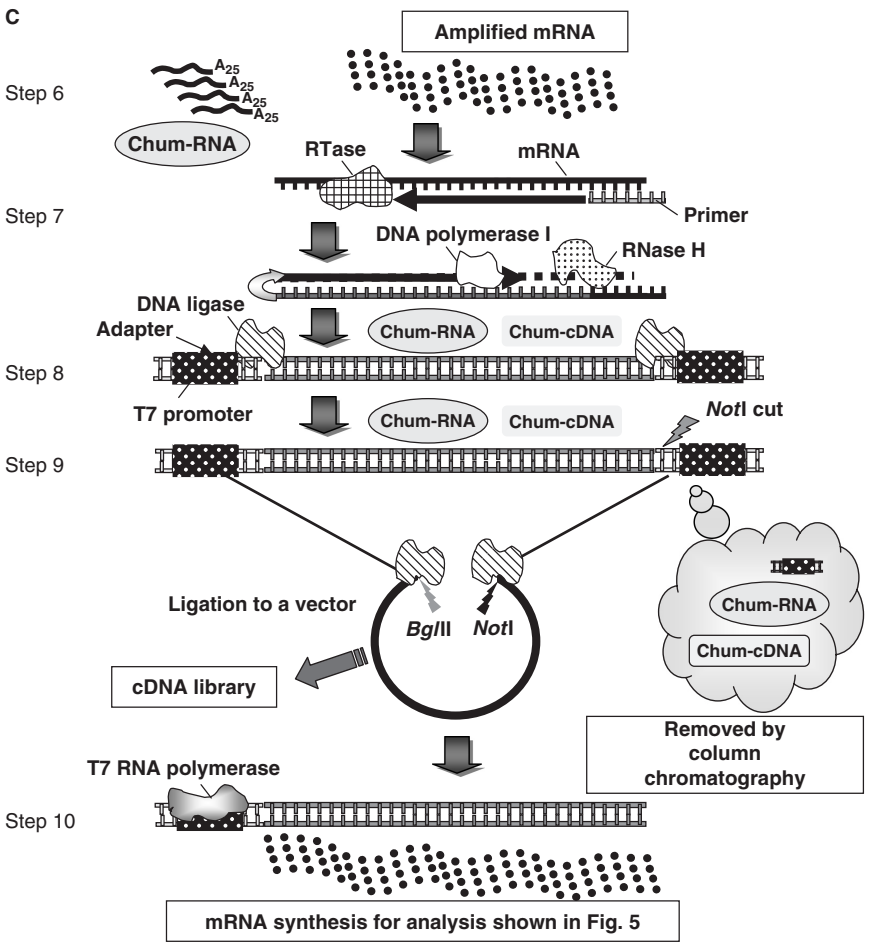


**Figure 1.** Continued.

The chum-RNA molecule consisted of 16 nucleotides that are designed not to form a hairpin, followed by a poly(A) tail 25 nucleotides in length that is required for hybridization with oligo-dT primer. Analysis of the sequence indicated that the chum-RNA was unlikely to form a hairpin structure. It is essential to add a poly(A) tail to chum-RNA because it allows the oligo-dT primer to easily associate with the chum-RNA molecule, activating the RTase and beginning cDNA synthesis (Figure 1A-i). In some experiments, we used a 41-nucleotide chum-RNA carrying a *Not*I restriction site (Figure 1A-ii; data not shown). We believe that the size and sequence of the chum-RNA could be varied, and modification with biotin might also be convenient, facilitating removal of the chum-RNA with avidin-conjugated magnetic beads following completion of the reaction.

It is expected that chum-RNA is turned into chum-cDNA because it can be recognized as a small mRNA by the enzymes (Figure 1B). However, it may not be amplified because chum-cDNA is removed by column chromatography at the end of steps 5 and 9 of Figure 1. Even if a small amount of chum-cDNA would escape this removal process, it may not be amplified afterwards because the MEGAscript Kit we used for mRNA amplification is designed to function best with transcription templates larger than 500 bp; the size of chum-cDNA is only 41 bp according to the manufacturer's protocol.

### Chum-RNA allows cDNA synthesis from a single-cell quantity of mRNA

To explore the utility of the chemically-synthesized chum-RNA, we first determined the optimal amount required to activate RTase, by varying the quantity of chum-RNA added to the reaction mixture. We performed cDNA synthesis (see Supplementary Protocol, step 2.1–2.16) using a 100-cell quantity of 293T mRNA (~1 ng) either in the presence of the denoted quantity of chum-RNA. We found that 3 μM (1 μg per 25 μl of reaction mixture) of chum-RNA produces the optimal result (Figure 2A). Larger amounts of chum-RNA appear to inhibit cDNA synthesis.

Using the optimal amount of chum-RNA, we next investigated whether chum-RNA exerts any dummy effects on RTase. mRNA was extracted from ~1 million 293T cells using a Micro Poly(A) Purist kit (Applied Biosystems). The single cell quantity of mRNA (10 pg) used as the substrate for cDNA synthesis was obtained by diluting the million-cell stock. We performed cDNA synthesis in the presence (+) or absence (–) of chum-RNA, and then the reaction product was probed for GAPDH cDNA by PCR. Using agarose gel electrophoresis (AGE), a band at 902 bp was

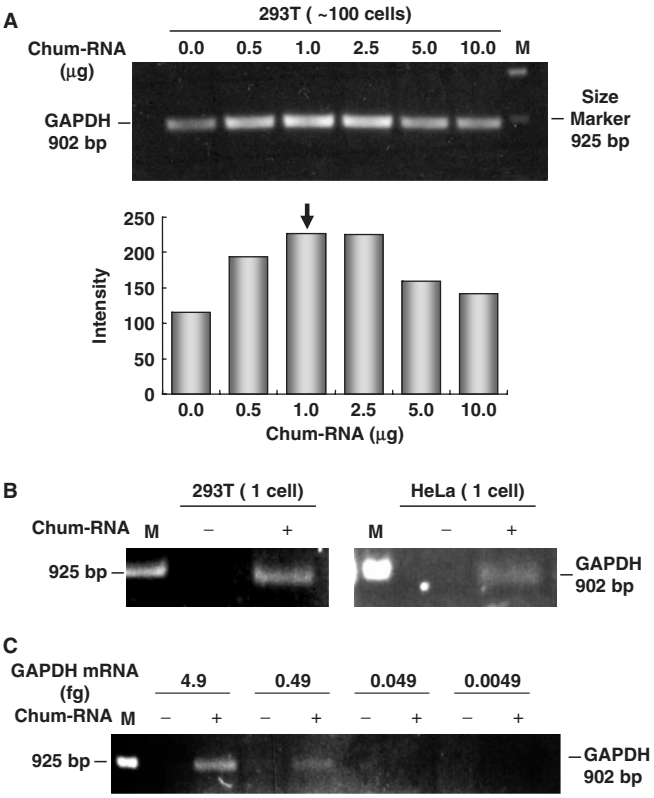


**Figure 2.** Chum-RNA facilitates sense strand mRNA amplification from a single-cell amount of mRNA. M denotes a molecular size marker (925 bp). (**A**) Determination of the optimum quantity of chum-RNA required to drive RTase to initiate GAPDH cDNA synthesis. Using a 100-cell quantity of 293T mRNA (~1 ng), cDNA synthesis was initiated in the presence of the denoted quantity of chum-RNA (Figure 1B, step 2). Successful cDNA synthesis was confirmed by PCR (30 cycles at 50°C) using the reaction product of Figure 1B, step 2 together with a primer-set that was designed to detect a band for GAPDH cDNA at 902 bp on AGE (upper panel). The bar graph, which was drawn by using Scion Image (Scion Corporation, Frederick, MA), compares the intensity of each band (lower panel). The arrow indicates the optimum quantity of chum-RNA (3 μM). (**B**) Synthesis of cDNA using a single-cell amount of mRNA from 293T (left, 10.5 pg) and HeLa (right, 10.1 pg) cells in the presence (+) or absence (–) of chum-RNA (3 μM). Successful cDNA synthesis was confirmed by the presence of GAPDH cDNA, which was detected by PCR (50 cycles at 50°C) using the reaction product of Figure 1B, step 2 and observed after AGE as a band at 902 bp. (**C**) Determination of the minimum amount of mRNA required for chum-RNA to drive RTase. The indicated amounts of GAPDH mRNA were reacted in the presence (+) or absence (–) of 3 μM chum-RNA (Figure 1B, step 2). We first cloned the amplified HsGAPDH cDNA into a plasmid and then its mRNA was prepared by *in vitro* transcription using a MEGAscript High Yield Transcription Kit (Applied Biosystems), which yielded 1.3 kb GAPDH mRNA. The original stock of GAPDH mRNA (4.9 μg/ml) was sequentially diluted ten-fold. Successful cDNA synthesis was confirmed by detection of the band for GAPDH cDNA (902 bp) that was generated by PCR.

detected, but only in the presence of chum-RNA (Figure 2B, left panel). The experiment was repeated using a single-cell amount of HeLa mRNA, which is expected to possess a distinct amount of GAPDH mRNA, and found a similar effect of the chum-RNA (Figure 2B, right panel). These observations indicate that the addition of chum-RNA facilitated cDNA synthesis when a single-cell amount of mRNA was used as a substrate.

To determine the minimum amount of mRNA required for initiation of cDNA synthesis by RTase in the presence of chum-RNA, GAPDH mRNA synthesized *in vitro* was used as a substrate. Only 0.49 fg, which corresponds to approximately 730 mRNA molecules of GAPDH, is sufficient for cDNA synthesis to begin, if chum-RNA is added in the reaction mixture (Figure 2C). This result suggests that chum-RNA can facilitate cDNA synthesis using mRNA from a single human cell (see below). Notably, no cDNA synthesis occurred even with a 10-fold higher amount of GAPDH mRNA (4.9 fg) if the chum-RNA was absent from the reaction mixture. We could also confirm that the 902 bp band is independent of Chum-RNA because no such band was observed in the rightmost lane, which shows the PCR reaction with a very low amount of GAPDH mRNA (Figure 2C, 0.0049 fg).

### Linear amplification of mRNA by RNA polymerase in the presence of chum-RNA

We next examined whether chum-RNA is useful during the process of double strand DNA synthesis and subsequent sense mRNA amplification using T7 RNA polymerase. First, cDNA was synthesized using RTase and a single-cell amount of 293T mRNA, in the presence or absence of chum-RNA (Figure 1B, step 1). The synthesized cDNA was converted into double strand DNA through degradation of the poly(A)$^+$ RNA strand with RNase H, followed by second strand synthesis with *Escherichia coli* DNA polymerase I (Figure 1B, step 2). Successful cDNA synthesis was confirmed by probing for GAPDH (Figure 3A) as described above (Figure 2). We next added an adaptor containing the promoter sequence for T7 RNA polymerase using DNA ligase (Figure 1B, step 3), cut with *Not*I and removed the oligo-dT primer side of the cDNA by column chromatography (Figure 1B, step 4); the latter process is essential for preventing the *in vitro* transcription of anti-sense mRNA in the next step. Then we conducted linear amplification of whole mRNA by *in vitro* transcription in the presence or absence of chum-RNA (Figure 1B, step 5) using T7 RNA polymerase and a MEGAscript High Yield Transcription Kit (Applied Biosystems) at 37°C for 4 h. Using this amplified mRNA, we repeated the cDNA synthesis (Figure 1B, step 1) and double strand DNA conversion (Figure 1B, step 2) to detect GAPDH mRNA. The entire GAPDH mRNA was successfully amplified in the presence of chum-RNA, as assessed by comparing the intensity of the GAPDH band (Figure 3B, plus lane) to the trace amount obtained in the absence of chum-RNA (Figure 3B, minus lane). This result reflects the powerful ability of chum-RNA to promote sense mRNA amplification using a single-cell quantity of substrate.

We next examined whether chum-RNA could facilitate the non-biased amplification of the entire mRNA content of a sample of 293T cells, using the Agilent 2100 Bioanalyzer (G2940BA; Agilent Technologies, Inc., Palo Alto, CA). Since the sensitivity of the Agilent apparatus was not sufficiently high to detect the amplified product from a single-cell amount of 293T mRNA (data not shown), we repeated the amplification process using

a 10-cell amount of 293T mRNA. Next, we labeled the amplified mRNA *in vitro* with fluorescent dye (Cy3-CTP) using T7 RNA polymerase and the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies). As shown in Figure 3C, the maximum intensity of the Cy3 signal observed at 1 kb was more than 30-fold stronger in the chum-RNA 'plus' sample (red line) than in the chum-RNA 'minus' sample (blue line). These results indicate that the successful linear amplification conducted here did not introduce significant bias in the vast majority of the mRNA population.

### Preparation of a highly representative single-cell cDNA library

We next examined whether chum-RNA would allow efficient construction of a cDNA library from a single-cell amount of mRNA. Hereafter, we will call this a single-cell cDNA library, even though we obtained the single-cell mRNA quantity by dilution of the original stock obtained from a million-cell quantity of mRNA. According to the linker-primer method (19), cDNA was synthesized from a single-cell amount of 293T mRNA, converted into double strand DNA, and an adaptor added. The DNA was digested with *Not*I, the unnecessary portion removed by column chromatography and mRNA was amplified using T7 RNA polymerase in the presence of chum-RNA, as described above (Figure 1B). Since this single round of mRNA amplification did not produce an amount of mRNA sufficient to make a good cDNA library harboring a high number of independent clones (complexity), the amplification procedure was repeated several times. Four rounds of mRNA amplification in the presence of chum-RNA produced a high-quality cDNA library containing independent cDNA clones of $6.6 \times 10^5$ colony forming unit (cfu), from only a single-cell amount of 293T mRNA. In contrast, omission of chum-RNA resulted in a very poor-quality cDNA library with only background quantities of independent cDNA clones ($5.3 \times 10^4$ cfu) and no evidence that any plasmid clones carried human cDNA inserts (data not shown).

To examine the quality of the single-cell cDNA library, 60 independent colonies were randomly picked and the length of their cDNA inserts was determined through digestion with appropriate restriction enzymes. It was found that 24 of 60 (40%) colonies possessed cDNA inserts longer than 0.1 kb, yielding an average cDNA insert size of 0.75 kb (Figure 4). Sequence analysis indicated that 15 colonies contained distinct human cDNAs

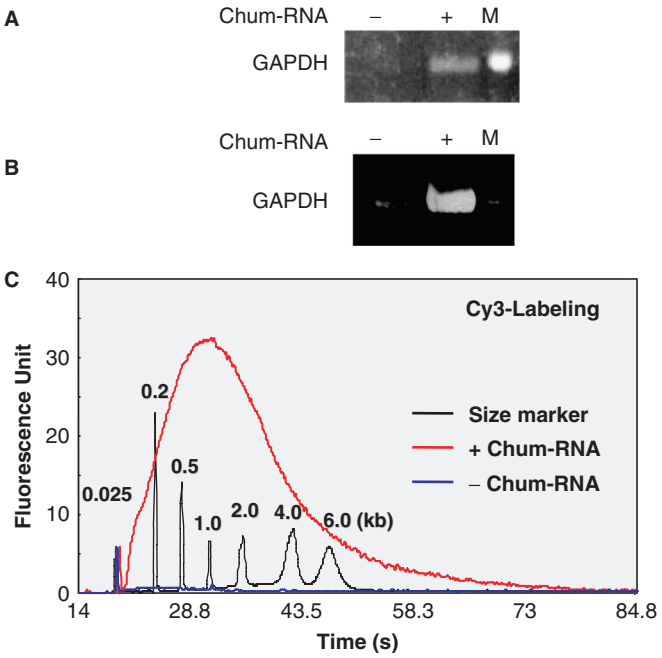


**Figure 3.** Chum-RNA allows non-biased linear amplification of mRNA using T7 RNA polymerase. M denotes the molecular size marker (925 bp). (**A**) Successful synthesis of cDNA using a single-cell amount of 293T mRNA in the presence (+) or absence (–) of chum-RNA (step 2 of Figure 1B), confirmed by detection of the band for GAPDH cDNA generated by RT-PCR (40 cycles at 55°C) on AGE. (**B**) Successful amplification of a single-cell amount of 293T mRNA in the presence (+) or absence (–) of 3 μM chum-RNA (step 7 of Figure 1C), confirmed by detection of GAPDH cDNA generated by RT-PCR (40 cycles at 55°C). (**C**) Chum-RNA allows non-biased amplification of the entire mRNA population from 293T cells. The amplified cDNA product (step 5 of Figure 1B) from a 10-cell amount of 293T mRNA in the presence (+) or absence (–) of 3 μM chum-RNA was labeled *in vitro* with fluorescent dye (Cy3-CTP) using T7 RNA polymerase and the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies). The resulting reaction product was analyzed using an Agilent 2100 Bioanalyzer (G2940BA; Agilent Technologies Inc., Palo Alto, CA). Numbers at each peak of the marker curve denote the molecular size in kilobases (kb). The ordinate and the abscissa represent the fluorescent intensity and the retention time of the bioanalyzer, respectively.

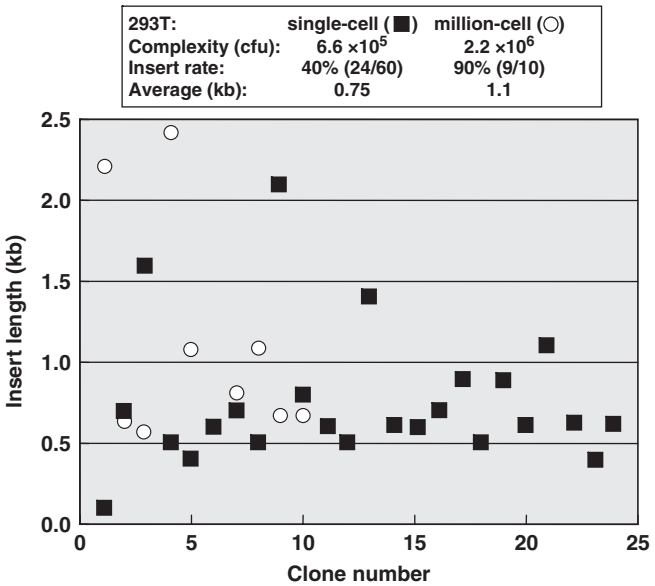


**Figure 4.** Size distribution of the insert cDNAs from the amplified versus unamplified cDNA libraries. Plasmid DNAs from 60 clones randomly selected from the 293T single-cell cDNA library amplified using chum-RNA (shown by filled squares), or from 10 clones randomly selected from the 293T million-cell cDNA library (unamplified, shown by open circles) were prepared, digested by *Bam*HI and the size of the insert cDNA analyzed by using AGE. DNA sequences were determined and subjected to a homology search with the BLASTN algorithm (http://www.ncbi.nlm.nih.gov; Table. S1). Only those clones harboring human cDNAs longer than 100 bp are presented. The independent cDNA clones (complexity), insert late (%) and average size of cDNA inserts (kb) of these cDNA libraries are listed above the panel.

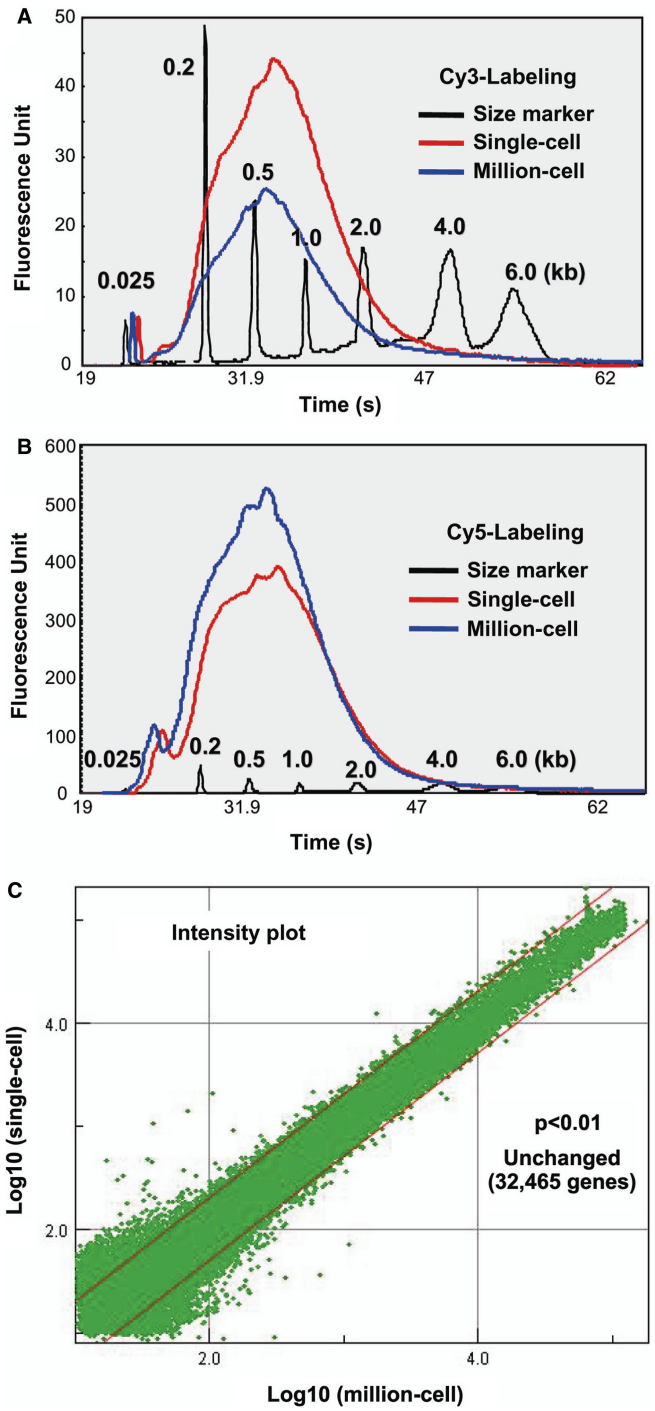


**Figure 5.** Comparison of the quality of amplified (single-293T cell) versus unamplified (1 million 293T cells) cDNA libraries. (**A** and **B**) Size distribution of mRNA synthesized from single-cell (red lines) versus million-cell (blue lines) cDNA libraries that shows the non-biased transcript amplification. Each plasmid form of the cDNA library (step 10 of Figure 1C) was labeled with Cy3-CTP (A) or Cy5-CTP (B) by using T7 RNA polymerase and the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies). The resulting reaction product was analyzed by using an Agilent 2100 Bioanalyzer (G2940BA). The numbers at each peak of the marker curve denote the molecular size in kb. The ordinate and the abscissa indicate the fluorescent intensity and the retention time of the bioanalyzer, respectively. (**C**) A scatterplot matrix comparing the expression profiles obtained from single-cell (*Y*-axis) versus million-cell (*X*-axis) cDNA libraries. Cy3-labeled (single-cell) or Cy5-labeled (million-cell) samples were hybridized to Agilent's all-human cDNA microarray (Hu44K), and then the color-swap experiment was also performed. Fluorescence intensities (signal intensity minus background) were normalized to median array densities to obtain normalized measures for each gene across all of the samples. The *X*-axis (million-cell) and *Y*-axis (single-cell) show virtual gene expression levels. The correlation coefficient was $R = 0.98301$. The red lines denote twofold changes. These results indicate that >90% of the loci in the amplified product (single-cell) are within 2-fold of the copy number of the unamplified product (million-cell).

(two clones were equal, Table S1), suggesting that the amplification process was non-biased, and able to generate a library of a similar quality to that of the 293T million-cell cDNA library (Table S2). The remaining nine colonies harbored DNA fragments from *E. coli*, bovine genome or plasmid vector. The contamination of *E. coli* DNA fragment may have occurred during the manufacturer's purifying process of the recombinant enzymes from the *E. coli* extract. Bovine DNA contamination is probably derived from the bovine serum albumin (BSA) included in some of the reaction mixtures of the enzyme.

### Microarray analysis of a single-cell cDNA library

To further assess the non-biased amplification of this cDNA library, we first examined the size distribution of the mRNAs transcribed from the plasmids carrying this single-cell cDNA library, using T7 RNA polymerase and an Agilent 2100 Bioanalyzer. As a control, we examined in parallel the million-cell 293T cDNA library prepared without amplification (original mRNA stock). These mRNAs were labeled with fluorescent dye (Cy3-CTP or Cy5-CTP) using the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies). The size distribution pattern of these Cy-labeled complementary RNA (cRNA) or mRNA from the single-cell and million-cell cDNA libraries showed similar distribution patterns throughout a wide size range, irrespective of which fluorescent dye was used, Cy3 or Cy5 (Figure 5A), providing a further indication that the linear amplification process did not introduce significant bias in the vast majority of mRNA population.

To more accurately determine the extent of variation caused by the amplification process, we next scanned the Agilent's all-human cDNA microarray (Hu44K) using the above-described Cy3 or Cy5-labeled cRNAs (Figure 5A and 5B). If any significant bias was introduced during the amplification procedure, a change in the expression profile would have been observed when the scanning data from the Cy3 or Cy5-labeled cRNAs was compared. Figure 5C shows a scatter plot matrix comparing the expression profiles obtained from amplified (single 293T cell) versus unamplified (1 million 293T cells) RNA. Over 90% of the human genes (37 465 spots yielded signals in our experiment) in the amplified product are within 2-fold (between the red lines in Figure 5C) of the copy number of the unamplified sample. Scatter plot analysis shows that the differential expression ratios derived from amplified and unamplified RNA were highly comparable ($R^2 = 0.9663$). These results further confirm the non-biased nature of amplification achieved during the preparation of this single-cell cDNA library.

## DISCUSSION

In the present study, we showed that a small mRNA molecule called chum-RNA is useful not only to facilitate cDNA synthesis but also for cDNA library preparation from a single-cell amount of mRNA, using T7 RNA polymerase to amplify RNA (Figure 1A). Chum-RNA can drive RTase to begin cDNA synthesis from quantities of substrate as low as those obtained from a single cell (10.5 pg of total RNA, Figure 2B). When GAPDH mRNA molecules synthesized *in vitro* were used as the substrate, chum-RNA was able to drive cDNA synthesis from only 0.49 fg of mRNA (730 mRNA molecules) as a substrate (Figure 2C), a quantity corresponding to that of a minor population of mRNA molecules in a single mammalian cell. The chum-RNA protocol presented here achieves a ~30-fold mRNA amplification (Figure 3C). Unlike PCR-based techniques, this protocol preserves the original size and species distribution of the mRNA pool. Indeed, *in vitro* labeling of the amplified cDNA with fluorescent dye revealed an mRNA distribution pattern similar to that observed in non-amplified cDNA synthesized from 1 million cells (Figure 3C), indicating that the amplification of the mRNA population was unbiased.

Notably, we were also successful in preparing a cDNA library from a single-cell amount of mRNA, after four rounds of amplification. This cDNA library preserved the mRNA size and species distribution (Figure 4; Table S2) observed in the million-cell cDNA library, confirming the non-biased nature of the amplification achieved using our chum-RNA protocol. The independent cDNA clone of this library ($6.6 \times 10^5$ cfu) also confirms the ~30-fold RNA amplification per one round of the amplification process (Figure 3C) because the number ($6.6 \times 10^5$) almost matches to the number ($28.5^4$) if 95% recovery rate was considered ($28.5 = 30 \times 0.95$). Fluorescent labeling of the mRNA transcribed from the single-cell library indicated that linear amplification using chum-RNA did not introduce any significant amplification bias (Figures 5A and B), and this result was confirmed using microarray analysis. To our knowledge, this is the first report of a technique for sense mRNA amplification and subsequent cDNA library preparation that does not rely on PCR. This technique could also be used for preparation of anti-sense RNA.

We initiated this study with the expectation that the addition of small RNA (chum-RNA) to the reaction mixture would accelerate the enzyme's rate of substrate conversion according to the Michaelis–Menten equation. Indeed, the first enzyme we tested, RTase, recognized not only the *bona fide* 293T mRNA (less than 1 pM) but also the chum-RNA (3 μM) as its substrate (step 1 of Figure 1B) and the reaction product of cDNA synthesis was observed after only 75 min of incubation (Figure 2 and Supplementary Protocol step 2.6). Without chum-RNA, the substrate conversion rate would have been too slow to detect the reaction product of cDNA synthesis over such an incubation. The other enzymes used in the subsequent reactions (DNA polymerase, RNase H and DNA ligase) also recognized both the *bona fide* 293T mRNA and chum-RNA as substrates (steps 2–4 of Figure 1B) and efficiently catalyzed the reactions to produce detectable amounts of cDNA derived from both mRNAs (Figure 2). Moreover, T7 RNA polymerase also helped to amplify both 293T mRNA and chum-RNA which was then removed by column chromatography (step 5 of Figure 1B). However, we cannot be entirely sure at present about how chum-RNA actually works; as an alternative mechanism it could activate replication enzymes and thereby promote the capture of rare mRNAs. Future research will be conducted to characterize the molecular mechanism behind the chum-RNA phenomenon.

Until now, methods for amplification of single-cell amounts of RNA have consisted primarily of exponential amplification using PCR, in combination with linear amplification using multiple T7 RNA polymerase reactions. One example is Ribo-SPIA, which generates micrograms of labeled cDNA from 5 ng of total RNA in 1 day for analysis on arrays or by PCR quantification (4). A modified amplification protocol generates microgram quantities of message-derived material from 100 ng of total RNA (13). Another protocol reduces by 1 million-fold the amount of input RNA needed for microarray analysis to as little as 10 ng, which corresponds to the amount obtainable from a few thousands cells (14). Such techniques are useful for genome-wide expression analysis, but the exponential amplification produced by PCR still introduces bias into the size and species distribution of the amplified product. Our technique is compatible with previously developed techniques, that is, chum-RNA can be added to any reaction mixture without compromising its results. For example, nearly 1 μg of total RNA is required for the preparation of a cDNA library, when RNA polymerase alone is utilized (14). This requirement is likely due to the difficulty with which DNA ligase initiates ligation of the adaptor, and the addition of chum-RNA might help to decrease the quantity of starting material required for this technique. As such, chum-RNA might constitute a versatile tool for boosting the activity of other DNA- or RNA-modifying enzymes with $K_m$ values around 1 mM, allowing these reactions to progress using only small quantities of substrate.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We are obliged to Dr Patrick Hughes for critically reading the manuscript. This work was supported in part by a grant-in-aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science and Technology of Japan and from the Japan Science, and Technology Agency (JST) to H.N. Funding to pay the Open Access publication charges for this article was provided by the Regional Research and Development Resources Utilization Program of JST.

*Conflict of interest statement*. None declared.

## REFERENCES


1. Thompson,J.R., Marcelino,L.A. and Polz,M.F. (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Res.*, **30**, 2083–2088.
2. Van-Gelder,R.N., von-Zastrow,M.E., Yool,A., Dement,W.C., Barchas,J.D. and Eberwine,J.H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl Acad. Sci. USA*, **87**, 1663–1667.
3. Phillips,J. and Eberwine,J.H. (1996) Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. *Methods*, **10**, 283–288.
4. Dafforn,A., Chen,P., Deng,G., Herrler,M., Iglehart,D., Koritala,S., Lato,S., Pillarisetty,S., Purohit,R., Wang,M. *et al.* (2004) Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. *Biotechniques*, **37**, 854–857.
5. Moll,P.R., Duschl,J. and Richter,K. (2004) Optimized RNA amplification using T7-RNA-polymerase based in vitro transcription. *Anal. Biochem.*, **334**, 164–174.
6. Kurimoto,K., Yabuta,Y., Ohinata,Y., Ono,Y., Uno,K.D., Yamada,R.G., Ueda,H.R. and Saitou,M. (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.*, **34**, e42.
7. Kurimoto,K., Yabuta,Y., Ohinata,Y. and Saitou,M. (2007) Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat. Protoc.*, **2**, 739–752.
8. Saitou,M., Yabuta,Y. and Kurimoto,K. (2008) Single-cell cDNA high-density oligonucleotide microarray analysis: detection of individual cell types and properties in complex biological processes. *Reprod. Biomed. Online*, **16**, 26–40.
9. Eberwine,J., Yeh,H., Miyashiro,K., Cao,Y., Nair,S., Finnell,R., Zettel,M., and Coleman,P. (1992) Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA*, **89**, 3010–3014.
10. Tietjen,I., Rihel,J.M., Cao,Y., Koentges,G., Zakhary,L. and Dulac,C. (2003) Single-cell transcriptional analysis of neuronal progenitors. *Neuron*, **38**, 161–175.
11. Osawa,M., Egawa,G., Mak,S.S., Moriyama,M., Freter,R., Yonetani,S., Beermann,F. and Nishikawa,S. (2005) Molecular characterization of melanocyte stem cells in their niche. *Development*, **132**, 5589–5599.
12. Jensen,K.B. and Watt,F.M. (2006) Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. *Proc. Natl Acad. Sci. USA*, **103**, 11958–11963.
13. Baugh,L.R., Hill,A.A., Brown,E.L. and Hunter,C.P. (2001) Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.*, **29**, E29.
14. Iscove,N.N., Barbara,M., Gu,M., Gibson,M., Modi,C. and Winegarden,N. (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat. Biotechnol.*, **20**, 940–943.
15. Ohara,R., Kikuno,R.F., Kitamura,H. and Ohara,O. (2005) cDNA library construction from a small amount of RNA: adaptor-ligation approach for two-round cRNA amplification using T7 and SP6 RNA polymerases. *Biotechniques*, **38**, 451–458.
16. Marcus,S.L. and Modak,M.J. (1976) Observations on template-specific conditions for DNA synthesis by avian myeloblastosis virus DNA polymerase. *Nucleic Acids Res.*, **3**, 1473–1486.
17. McClure,W.R. and Jovin,T.M. (1975) The steady state kinetic parameters and non-processivity of Escherichia coli deoxyribonucleic acid polymerase I. *J. Biol. Chem.*, **250**, 4073–4080.
18. Lehman,I.R. (1974) DNA ligase: structure, mechanism, and function. *Science*, **186**, 790–797.
19. Kobori,M., Ikeda,Y., Nara,H., Kato,M., Kumegawa,M., Nojima,H. and Kawashima,H. (1998) Large scale isolation of osteoclast-specific genes by an improved method involving the preparation of a subtracted cDNA library. *Genes Cells*, **3**, 459–475.