

An Alternative Approach to Synthesize cDNA Bypassing Traditional Reverse Transcription

Xiao-Xia Li · Fang Zheng · Yan-Li Jiao · Gang Guo · Bao-Li Wang · Zhi Yao

Published online: 29 January 2008
© Humana Press Inc. 2008

Abstract cDNAs of certain target genes are difficult to obtain by traditional reverse transcription. Herein we describe a novel method to synthesize cDNA based upon the use of the class IIS restriction enzymes. Briefly, the exons of a certain gene are separately PCR-amplified, each using the primers containing a recognition sequence of a certain class IIS restriction enzyme. All the fragments are restricted using the enzyme(s), resulting in the cohesive end of each exon that is complementary to the one in its adjacent exon. Then the fragments can be assembled together in their naturally occurring order. We successfully applied this method to acquire the coding sequence of *Hoxa7* gene. This approach is simple, highly efficient, less error prone and cost-effective, and can also be used to fuse different PCR-fragments from distinct genes to create a chimeric gene or to perform site-directed mutagenesis.

Keywords cDNA synthesis · DNA assembly · Class IIS restriction enzymes · Reverse transcription · Polymerase chain reaction

Xiao-Xia Li and Fang Zheng—These authors contributed equally to this work.

X.-X. Li · F. Zheng · Y.-L. Jiao · G. Guo · B.-L. Wang (✉)
Key Research Lab of Hormone and Development Affiliated to the Ministry of Health, Tianjin Medical University, Tianjin 300070, P.R. China
e-mail: blwang7212@yahoo.com.cn

X.-X. Li · Z. Yao (✉)
College of Basic Medical Science, Tianjin Medical University, Tianjin 300070, P.R. China
e-mail: yaozhi23@hotmail.com

Z. Yao
Tianjin Key Research Lab of Cellular and Molecular Immunology, Tianjin 300070, P.R. China

Introduction

Complementary DNA (cDNA) is often used in gene cloning, as gene probes or in the creation of a cDNA library, which is valuable for studying protein structure and function. The conventional method to acquire the cDNA fragment of a target gene is based on reverse transcriptase PCR (RT-PCR) amplification. Briefly, total RNA is isolated and reverse-transcribed into cDNA using a reverse transcriptase, then the resulting cDNA is used as a template for subsequent PCR amplification with primers specific for the target gene [1]. However, preparation of RNA is usually not a pleasant procedure as it requires intensive labor and high time consumption. Also, RNA is susceptible to endo- and exo-nuclease mediated degradation, rendering the initial stages of extraction and the downstream storage of the purified material more challenging than for DNA [2]. Although some strategies have been developed to prevent RNA degradation and increase yields such as thorough homogenization of tissue and cell samples, proper precipitation of RNA and optimized RNA storage conditions [3, 4], cDNAs of certain target genes are still difficult to obtain. These genes include the genes of rare transcripts, the genes of which the RNA templates are not always readily available such as those expressed in human cardiac tissue, and the long genes of which the RNA is not easy to be either reverse-transcribed in full-length using an ordinary reverse transcriptase or/and PCR-amplified using an ordinary polymerase. To solve the above-mentioned problems, several modified protocols combining chemical synthesis and PCR amplification of the target gene have recently been described for synthesis and assembly of such cDNA sequences. Representatives of these methods are the PCR-based thermodynamically balanced inside-out (TBIO) method [5], two-step total gene synthesis combining dual

asymmetrical PCR (DA-PCR) and overlap-extension PCR (OE-PCR) [6], and PCR-based two-step DNA synthesis (PTDS) [7]. However, since all these strategies are based upon chemical synthesis of a number of short nucleotide sequences approx. 50–60 nucleotides in length, major drawbacks are the high cost and high error rate of chemical synthesis. Moreover, because the assembly of long sequences is carried out by performing at least two steps of PCR amplifications, this further increases the error rate. Synthesis of full-length cDNA through artificial splicing of two or even more PCR-amplified exons using overlap-extension PCR (OE-PCR) seems to be a better method, thanks to its lower cost. Using this method, several sets of primers are synthesized, which correspond, respectively, to the exons comprising the full-length cDNA. The adjacent primers share several nucleotides complementary to each other. The exons are separately PCR-amplified, then a second-round overlap-extension PCR is carried out to splice the exons [8]. The advantage of this approach is the relative low cost because only a few short oligonucleotide primers are synthesized. However, this method is somewhat time-consuming because at least two rounds of PCRs are required to obtain the full-length target gene. Moreover, it cannot guarantee a high accuracy of the obtained product, due primarily to the required minimum of two rounds of PCR amplifications. Therefore, it is of great interest to develop a simple, less error prone, and more cost-effective method that guarantees the successful synthesis of the desired gene whose cDNA is difficult to obtain by traditional reverse transcription.

In the current study, we developed a novel method to synthesize cDNA based upon the use of class IIS restriction enzymes like *Eco31I*. In brief, the exons of a certain gene are separately PCR-amplified, each using a primer pair containing a recognition sequence of a certain class IIS restriction enzyme. All the fragments are restricted using the enzyme(s), resulting in cohesive ends of each exon being complementary to the one in its adjacent exon. Then the fragments can be assembled together in their naturally occurring order. This method is rapid, cost-effective, and accurate as compared to previously published protocols.

Materials and Methods

PCR Amplification

The method was exemplified by the synthesis of the coding sequence of the *Hoxa7* gene, a member of the *Hox* family. The *Hox* family of the clustered homeobox genes plays a fundamental role in the morphogenesis of the vertebrate embryo [9]. During organogenesis, the proteins encoded by

these genes act to trigger positional identity of embryonic cells. This determines the patterning and segment identity along the anterior–posterior axis of the skeleton and a variety of organ systems. The family members share the features of temporally and spatially restricted expression. It is likely that *Hoxa7* plays a role during skin development as the expression of human *Hoxa7* was weakly detected in the fetal dermis during the second trimester of fetal development, but could not be observed in the newborn or adult dermis [10]. To better understand the function and regulatory process of the *Hoxa7* gene, cDNA is often required for expression of the protein or generation of RNA probes. Our protocol provides an alternative method to synthesize a specific cDNA bypassing the traditional RT-PCR procedure in which RNA has to be extracted from fetal human tissues [10].

The coding sequence of *Hoxa7* is composed of only two exons, of which several starting and ending nucleotides of the translated region are outlined in Fig. 1. To synthesize cDNA of the *Hoxa7* gene (translated region), primers for amplification of the two exons of the gene were designed (Table 1). In the primers, we introduced the restriction sequence of *Eco31I* enzyme, which is a class IIS restriction endonuclease. Like other class IIS enzymes, *Eco31I* cuts outside its recognition site leaving 4-base 5'-overhangs that are predetermined by a downstream sequence, i.e.:



In the primers, the boxed bases in F1 and R2 are the recognition sequences of restriction enzyme *HindIII* and *EcoRI*, respectively, which are to be used to facilitate the cloning of the *Hoxa7* cDNA. The shaded regions in the primers R1 and F2 indicate the restriction sequence of *Eco31I*. Noteworthy is also that the bold bases 'TCAG' in F2 originate from the last four bases of the sense strand in the first exon of *Hoxa7*, such that the two expected 5'-overhangs (bold regions in the two primers) following *Eco31I* restriction are mutually complementary. The underlined regions in all the primers are homologous to the coding sequence of *Hoxa7*.

Two parallel PCR reactions were carried out to amplify the two exons separately in a volume of 50 μl , which included 10 ng of the human DNA extracted from blood cells, 0.4 μM of each primer, 200 μM dNTPs, 2.5 U *Pfu* DNA polymerase, 10 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 20 mM Tris-HCl (pH 8.8), 2 mM MgSO_4 , 0.1% Triton X-100, and 0.1 mg/ml nuclease-free bovine serum albumin (Fermentas). The reactions were performed in a Perkin Elmer 9600 thermocycler programed for an initial denaturation step of 3 min at 95°C, followed by 30 cycles of

Fig. 1 Schematic diagram of generation of *Hoxa7* cDNA. Note: The underlined bases indicate the restriction sites introduced by the primers. The starting and ending bases constituting the translated region of the exons are shaded gray

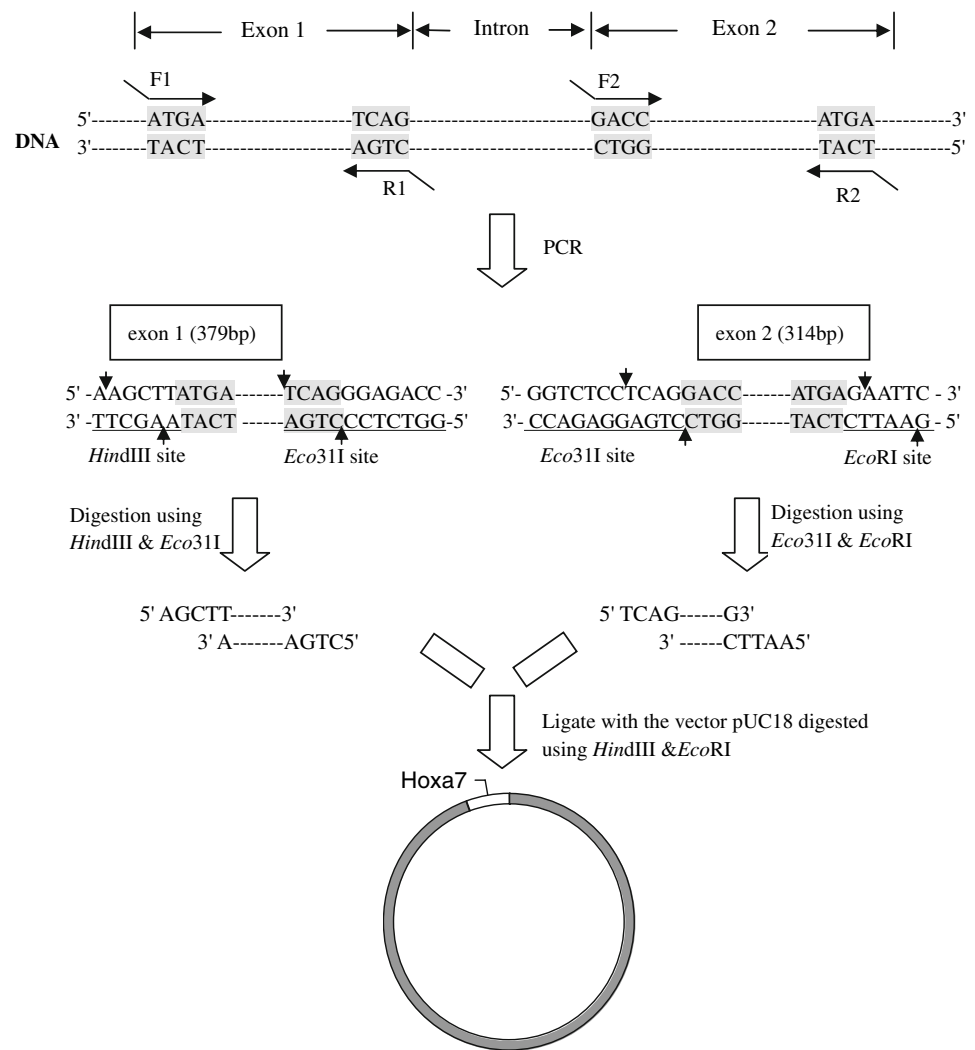


Table 1 The primers used for generation of *Hoxa7* cDNA

| Exons | Primers | Sequences for primers |
|-------|---------|---|
| 1 | F1 | 5'- GCG <u>AAGCTT</u> ATGAGTTCTTCGTATTATGTG -3' |
| | R1 | 5'- ATTGGTCTCC TCGA AGACCGCATCCAGG -3' |
| 2 | F2 | 5'- AATGGTCTCC TCAGG ACCTGACAGGAAGC -3' |
| | R2 | 5'- AGT <u>GAATTC</u> TCATTCTCTCTGCTTCTCTC -3' |

Note: The restriction sites for *HindIII* and *EcoRI* are in boxes. The restriction sites for *Eco31I* are shaded gray. The bases underlined are from the sequences of *Hoxa7* cDNA. The bold bases 'TCAG' in F2 are originated from the last four bases of the sense strand in the first exon of *Hoxa7* gene and complementary to the bold bases in R1

denaturation at 95°C for 30 s, annealing at 52°C for 30 s, and extension at 72°C for 1 min. The two PCR-amplified products were subjected to double digestion with *HindIII* or *EcoRI* (Fermentas) and *Eco31I* at 37°C for 2 h. Then the

digested fragments were purified from 2% agarose gel using the Bioflux gel extraction kit (Bioflux).

Cloning of *Hoxa7*

The plasmid pUC18 was digested using *HindIII* and *EcoRI* and afterwards gel-purified. The multi-component ligation was performed in a total volume of 10 µl containing 25 ng of each of the two enzyme-digested segments, 50 ng of the cleaved vector pUC18 and 3 U T4 DNA ligase (Promega) and incubated at 16°C for 16 h. Then the ligated product was transformed into *E. coli* JM109 and then the cells were plated onto an LB plate containing isopropyl β-D-thiogalactoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). White colonies were picked and plasmids were extracted and cleaved using *HindIII* and *EcoRI* to confirm the presence of the desired cDNA. Three of the recombinant clones were randomly selected and sequenced using M13 forward primer in Yingjun biotech company.

Results and Discussion

The principle of the *Eco31I* restriction-based approach is outlined in Fig. 1. The most essential feature of this method suggested in this article is as follows. The two exons coding for the *Hoxa7* gene (translated region) were amplified in two separate PCRs, with specially designed primers: two outer anchor primers, i.e., primers helping for the cloning, and two inner splicing primers, i.e., primers providing for the joining of the separately amplified segments. Each of the splicing primers consisted of a stretch identical to an edge of the corresponding DNA segment, and a pendant 5'-end which included a properly oriented recognition site of *Eco31I* endonuclease. The unique feature of *Eco31I* enzyme is that it recognizes six-base non-palindromic segments of a definite sequence and cuts DNA unilaterally with respect to the recognition site to yield a 5'-protruding 4-base overhang. This is quite different from usual class II endonucleases which recognize and cut inside palindromic sequences and result in self-complementary ends. The overhang generated by *Eco31I* is to be used for joining together the two DNA segments. For this goal, the adjacent splicing primers are designed in such a way that the two protruding ends at each prospective junction of the two segments are mutually complementary. Moreover, since this resulting overhang does not contain any part of *Eco31I* sequence, the joining will not introduce the *Eco31I* sequence, unlike the ligation following usual class II endonuclease treatment. Thus the joining of the segments leads to the naturally occurring sequence of the gene without introducing any undesired sequence.

In this work, the prospective four protruding bases "TCAG" in the pendent part of primer F2 originate completely from exon 1, while the protruding bases "CTGA" in the pendent part of primer R1 correspond only to the end of exon 1 itself. However, in the design of the adjacent splicing primers we had, to some extent, more flexibility. One splicing primer may have none to all of the protruding bases originating from its neighboring exon. And accordingly, the adjacent splicing primer should have all to none to ensure that the sum equals the number of the protruding bases produced by the restriction enzyme. For example, in the case of *Eco31I* restriction which produces a 4-base overhang, if one splicing primer contains one base from its neighboring exon, the adjacent splicing primer should have three bases from its neighboring exon such that the sum (i.e., four) equals the number of the protruding bases produced by *Eco31I*.

Thus, the paralleled amplifications of the two exons add the *Eco31I* site to the edge of the respective segment at the prospective junction. Following treatment using *Eco31I*, the two segments could be spliced to form the complete

coding sequence of the target gene, which could be easily cloned into a suitably treated vector.

In this work, we successfully employed the *Eco31I* restriction enzyme to synthesize cDNA of the *Hoxa7* gene. The two exons of *Hoxa7* were successfully amplified with the primer pairs, as shown in Fig. 2 (lanes 1 and 2). Multi-component ligation was carried out and the end product was transformed into JM109 at an efficiency of approximately 6×10^5 cfu/mg DNA. Over 80% of the transformants were white colonies on indicator plates containing IPTG and X-gal, and all ten white colonies selected for further examination were confirmed to be recombinant since the digestions of the plasmids they contained using *HindIII* and *EcoRI* endonucleases released a fragment of approx. 700 bp (Fig. 2, lane 3). This indicated that the two separately amplified fragments were successfully spliced and ligated into pUC18 plasmid in a directional manner. Three independent recombinant clones were selected, sequenced and blasted with Genbank database. The results revealed that all the three clones harbored the correct sequence of the target gene.

Although we take the assembly of *Hoxa7* coding sequence from two PCR-amplified exons as an example, the strategy described in this article is likely to work quite effectively if more exons are required to be spliced. In such cases, the target gene can be PCR-amplified into several segments determined by the number of exons it spans. Mutually complementary cohesive ends can be created among them using the strategy described herein. Then the segments can be ligated and cloned into the appropriately digested vector at one time. However, it should be noted that the protruding bases produced at different junctions should be different from one another, which is easy to

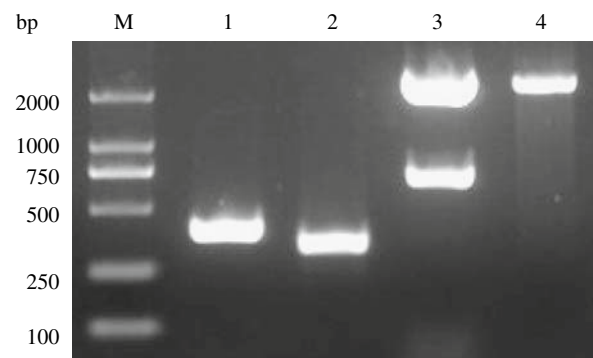


Fig. 2 Synthesis of cDNA of *Hoxa7* gene by *Eco31I* restriction-based strategy. Two exons of *Hoxa7* were successfully amplified with the primer pairs, which are 379 bp and 314 bp, respectively (lanes 1 and 2). Constructs were digested with *EcoRI* and *HindIII* endonucleases, and the DNA digests were separated on 2% agarose gel. The expected 693-bp fragment was released following cleavage of the construct (lane 3). As control, the pUC18 vector linearized using *EcoRI* and *HindIII* was run in lane 4

achieve by designing the primers in such a way that the exons can be spliced in the proper arrangement.

We have demonstrated the high efficiency of this method in synthesizing cDNAs of several silent genes.

Table 2 Alternative restriction sites of enzymes useful for the synthesis of cDNA

| Restriction enzyme | Target sequences |
|------------------------|---|
| <i>AlwI</i> | 5'-GGATCNNNN ▼ N-3' 3'-CCTAGNNNN N▲-5' |
| <i>Alw26I (BsmAI)</i> | 5'-GTCTCN▼NNNN-3' 3'-CAGAGN NNNN▲-5' |
| <i>BbsI (Bsc91I)</i> | 5'-GAAGACNN▼NNNN-3' 3'-CTTCTGNN NNNN▲-5' |
| <i>BbvI (Bst71I)</i> | 5'-GCAGCNNNNNNNN▼NNNN-3' 3'-CGTCGNNNNNNNN NNNN▲-5' |
| <i>BpmI</i> | 5'-CTGGAGNNNNNNNNNNNNNNNN NN▼-3' 3'-GACCTCNNNNNNNNNNNNNNNN▲NN-5' |
| <i>BsaI</i> | 5'-GGTCTC N▼NNNN-3' 3'-CCAGAGN NNNN▲-5' |
| <i>BsgI</i> | 5'-GTGCAGNNNNNNNNNNNNNN NN▼-3' 3'-CACGTCNNNNNNNNNNNNNN▲NN-5' |
| <i>BsmFI</i> | 5'-GGGACNNNNNNNNNN▼NNNN-3' 3'-CCCTGNNNNNNNNNN NNNN▲-5' |
| <i>BspMI</i> | 5'-ACCTGCNNNN▼NNNN-3' 3'-TGGACGNNNN NNNN▲-5' |
| <i>BsrDI (2/0)</i> | 5'-GCAATG NN▼-3' 3'-CGTTAC▲NN-5' |
| <i>EarI (Eam1104I)</i> | 5'-CTC TTCN▼NNN-3' 3'-GAGAAGN NNN▲-5' |
| <i>Eco57I</i> | 5'-CTGAAGNNNNNNNNNNNNNN NN▼-3' 3'-GACTTCNNNNNNNNNNNNNN▲NN-5' |
| <i>Esp3I (BsmBI)</i> | 5'-CGTCTCN▼NNNN-3' 3'-GCAGAGN NNNN▲-5' |
| <i>FokI</i> | 5'-GGATGNNNNNNNNNN▼NNNN-3' 3'-CCTACNNNNNNNNNN NNNN▲-5' |
| <i>HgaI</i> | 5'-GACGCNNNNNN▼NNNN-3' 3'-CTGCGNNNNNN NNNN▲-5' |
| <i>HphI</i> | 5'-GGTGANNNNNNNN N▼-3' 3'-CCACTNNNNNNNN▲N-5' |
| <i>MboII</i> | 5'-GAAGANNNNNNNN N▼-3' 3'-CTTCTNNNNNNNN▲N-5' |
| <i>PleI</i> | 5'-GAGTCNNNN▼N-3' 3'-CTCAGNNNN N▲-5' |
| <i>SapI</i> | 5'-GCTCTTCN▼NNN-3' 3'-CGAGAAGN NNN▲-5' |
| <i>SfaNI</i> | 5'-GCATCNNNNNN▼NNNN-3' 3'-CGTAGNNNNNN NNNN▲-5' |
| <i>Sth132I</i> | 5'-CCCGNNNN▼NNNN-3' 3'-GGGCNNNN NNNN▲-5' |

Sometimes we use this method to synthesize and assemble redundantly expressed genes. Although this kind of genes can be amplified by using traditional RT-PCR method, our strategy has shown its attractive advantage in terms of time span and effort consumption by omitting the laborious RNA preparation and reverse transcription procedure. Also, it might be more accurate since it omits the reverse transcription which may increase the probability of mismatched nucleotide incorporation because the reverse transcriptase does not possess 3'–5' exonuclease-dependent proofreading activity. The same approach can also be used to fuse different PCR fragments from distinct genes to create a chimeric gene or to perform site-directed mutagenesis.

As compared with overlap-extension PCR, which usually utilizes at least two rounds of PCRs for the synthesis of one fused DNA, one obvious advantage presented by our method is that only one round of PCR is required, which in turn minimizes the error rate of PCR. Accordingly, in the present study, the three randomly selected clones did not present mismatched nucleotides as confirmed by sequencing.

One concern about our method is that it may be restricted in use, due to the occasional presence of *Eco31I* recognition sites in target exons. However, this drawback can be easily overcome by employing other class IIS enzymes (Table 2). Moreover, different enzymes can be employed for different segments only if the enzymes used for the adjacent exons produce protruding bases of the same number.

Taken together, the cDNA assembly strategy described in this article provides an alternative to synthesize and amplify cDNA. It has at least the following advantages. First, it is simple, rapid, and far less effort-consuming as compared with traditional reverse transcription PCR, making it exceptionally appropriate for the synthesis of silent genes or long genes. Second, it employs only one round of PCR amplification, thus sharply limiting the mismatch of PCR.

Acknowledgments This work was supported by a grant from the National Natural Science Foundation of China (grant number 30670802) and key grants from the Tianjin Municipal Natural Science Foundation (grant no. 07JCZDJ07500, 06YFGPSH03300).

References

- Bertrand, E., Pictet, R., & Grange, T. (1994). Can hammerhead ribozymes be efficient tools to inactivate gene function? *Nucleic Acids Research*, 22, 293–300.
- Bustin, S. A., & Nolan, T. (2004). Pitfalls of quantitative reverse transcription polymerase chain reaction. *Journal of Biomolecular Techniques*, 15, 155–166.
- Eikmans, M., Baelde, H. J., de Heer, E., & Bruijn, J. A. (2000). Processing renal biopsies for diagnostic mRNA quantification:

- Improvement of RNA extraction and storage conditions. *Journal of the American Society of Nephrology*, 11, 868–873.
4. Puissant, C., & Houdebine, L. (1990). M. An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Biotechniques*, 8, 148–149.
 5. Gao, X., Yo, P., Keith, A., Ragan, T. J., & Harris, T. K. (2003). Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: A novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Research*, 31, e143.
 6. Young, L., & Dong, Q. (2004). Two-step total gene synthesis method. *Nucleic Acids Research*, 32, e59.
 7. Xiong, A. S., Yao, Q. H., Peng, R. H., Li, X., Fan, H. Q., Cheng, Z. M., & Li, Y. (2004). A simple, rapid, high fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Research*, 32, e98.
 8. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., & Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. *Gene*, 77, 61–68.
 9. Mark, M., Rijli, F. M., & Chambon, P. (1997). Homeobox genes in embryogenesis and pathogenesis. *Pediatric Research*, 42, 421–429.
 10. Stelnicki, E. J., Kömüves, L. G., Kwong, A. O., Holmes, D., Klein, P., Rozenfeld, S., Lawrence, H. J., Adzick, N. S., Harrison, M., & Largman, C. (1998). HOX homeobox genes exhibit spatial and temporal changes in expression during human skin development. *Journal of Investigative Dermatology*, 110, 110–115.