OXFORD

BIOLOGY Methods & Protocols

Biology Methods and Protocols, 2024, bpae037

https://doi.org/10.1093/biomethods/bpae037
Advance Access Publication Date: 25 May 2024
Methods Article

An optimized ligation-mediated PCR method for chromosome walking and fusion gene chromosomal breakpoints identification

Jrhau Lung [1,2,3,*], Ming-Szu Hung [2,4,5], Chao-Yu Chen [6,7,8], Tsung-Ming Yang [4], Chin-Kuo Lin [4], Yu-Hung Fang [4,5], Yuan-Yuan Jiang [4], Hui-Fen Liao [9] and Yu-Ching Lin [2,4,5]

[1]Department of Medical Research and Development, Chang Gung Memorial Hospital, Chiayi, 613, Taiwan
[2]Department of Medicine, College of Medicine, Chang Gung University, Taoyuan, 333, Taiwan
[3]Center for General Education, Chang Gung University of Science and Technology, Chiayi Campus, Chiayi, 613, Taiwan
[4]Department of Pulmonary and Critical Care Medicine, Chang Gung Memorial Hospital, Chiayi, 613, Taiwan
[5]Department of Respiratory Care, Chang Gung University of Science and Technology, Chiayi Campus, Chiayi, 613, Taiwan
[6]Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital, Chiayi, 613, Taiwan
[7]Department of Early Childhood Care and Education, Shu-Zen Junior College of Medicine and Management, Kaohsiung, 821, Taiwan
[8]Graduate Institute of Clinical Medical Sciences, College of Medicine, Chang Gung University, Taoyuan, 333, Taiwan
[9]Department of Biochemical Science and Technology, National Chiayi University, Chiayi, 600, Taiwan

*Correspondence address. Department of Medical Research and Development, Chang Gung Memorial Hospital, Chiayi, 613, Taiwan, Republic of China. Tel: 88653621000~2756; E-mail: jrhaulung@gmail.com

Abstract

Molecular techniques that recover unknown sequences next to a known sequence region have been widely applied in various molecular studies, such as chromosome walking, identification of the insertion site of transposon mutagenesis, fusion gene partner, and chromosomal breakpoints, as well as targeted sequencing library preparation. Although various techniques have been introduced for efficiency enhancement, searching for relevant single molecular event present in a large-sized genome remains challenging. Here, the optimized ligation-mediated polymerase chain reaction (PCR) method was developed and successfully identified chromosomal breakpoints far away from the exon of the new exon junction without the need for nested PCR. In addition to recovering unknown sequences next to a known sequence region, the high efficiency of the method could also improve the performance of targeted next-generation sequencing (NGS).

Keywords: ligation-mediated PCR; chromosome walking; chromosomal breakpoint; fusion gene; targeted sequencing

Introduction

Identification of unknown sequences adjacent to a known sequence region has great applications in various molecular studies including cDNA cloning, chromosome walking, transposon insertion site determination, as well as fusion gene partner and chromosomal breakpoint identification. Various methodologies have been developed and optimized for specific molecular targets, including inverse PCR [1–3], adaptor ligation-mediated PCR [4–6], and degenerate priming PCR [7–10]. Interested targets are frequently hidden behind the nonspecific amplification, and nested PCR is often introduced to eliminate nonspecific products. The sequential amplification of a sample would preferentially enrich short amplicons, leaving the expansion of interest sequence target behind. Even when the interested sequences are successfully amplified, only short sequence information less than 1 kb is obtained. It is still insufficient and inefficient for uncovering a molecular event far away from the known sequence side when the event could not be simply resolved by mapping the retrieved short sequence next to the known sequence region to the reference genome, such as chromosomal breakpoint.

Here, we report an efficient and optimized ligation-mediated PCR methodology, for chromosomal breakpoint identification (Fig. 1). We validate the protocol and smoothly recover amplicons containing chromosomal breakpoints in sizes of ~3 kb, ~4.5 kb, and 5.1 kb upstream of exon adjacent to the new exon junction of a *RET* fusion gene-positive cell line, LC-2/Ad, an *ALK* fusion gene-positive cell line, H3122, and an *ALK* fusion gene-positive lung cancer patient sample, respectively, without the need for nested PCR. The optimized method could be applied to improve the efficiency and performance of any application adopting ligation-mediated PCR.

Materials and methods

Genomic DNA extraction

Genomic DNA of lung cancer cell lines, LC-2/ad and H3122, and the patient sample were extracted using Allprep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instruction. Lung cancer patient samples used in the study were collected with written and signed informed consent and requested from the Research Specimen Processing Laboratory, Chang Gung Memorial Hospital, Chiayi, and Biobank, Chang Gung Memorial Hospital, Chiayi after approval from the

Received: 15 March 2024. **Revised:** 2 May 2024. **Editorial decision:** 12 May 2024. **Accepted:** 24 May 2024

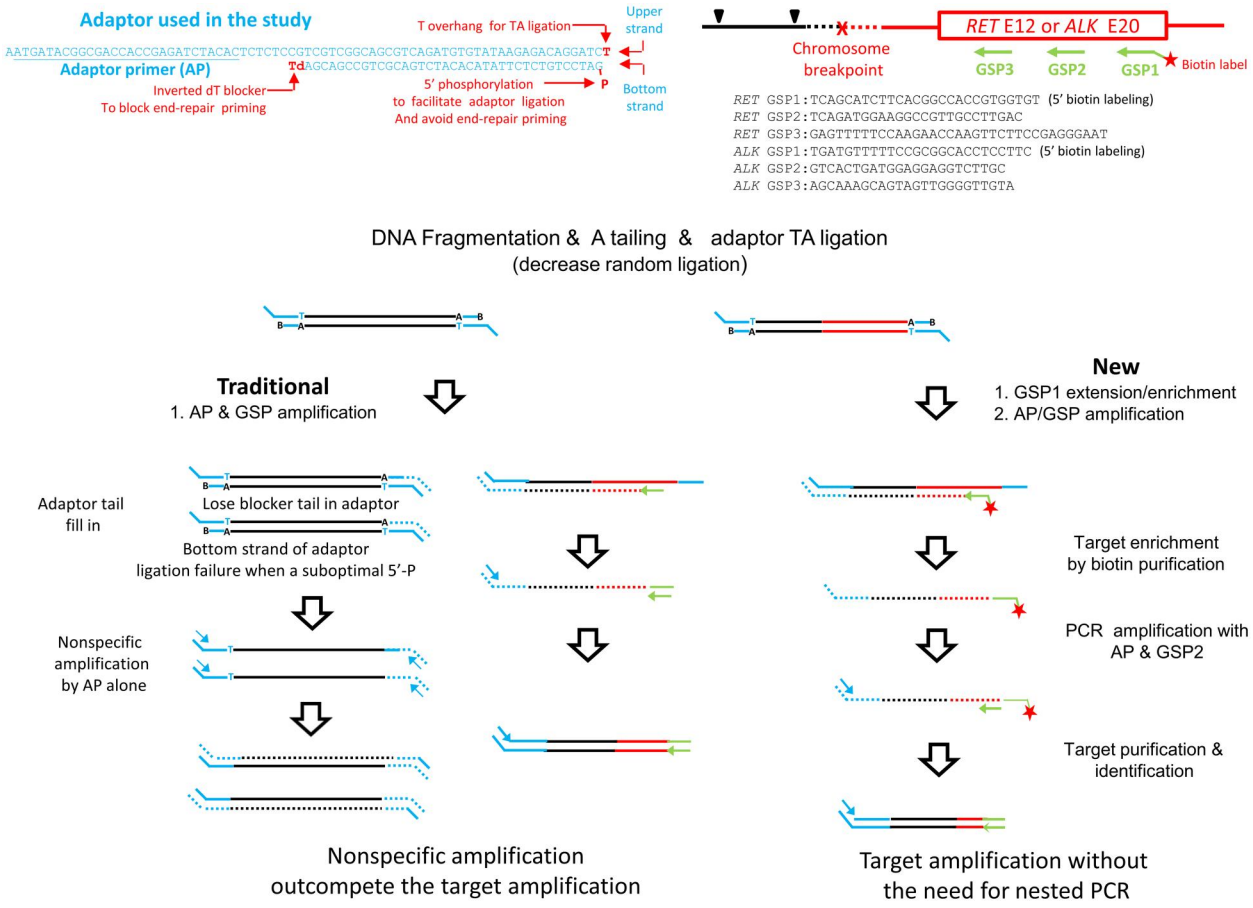

Figure 1. The principle of the optimized ligation-mediated PCR. The detailed primer and adaptor sequences and modifications are shown at the top of the figure. The traditional and the optimized methods all start with library fragmentation and adaptor ligation. In the optimized one, the attachment of the adaptor to library DNA is enhanced by TA ligation. The downstream target amplification could encounter nonspecific amplification by the adaptor primer alone due to adaptor end repairing during the early phase of PCR caused by the suboptimal 5' phosphorylation labeling or 3' extension blocker attachment to the bottom strand of the adaptor (on the bottom left of the figure). These complications are avoided by introducing the extension step using a biotin-labeled gene-specific primer and capture enrichment step (on the bottom right of the figure). These modifications make the optimized method capable of a highly specific and efficient target amplification without the need for additional nested PCR.

institutional review board (IRB No: 201600631B0, 201600631 B0C501 and 201600631B0C502).

Restriction enzyme cleavage and A-tailing

First, 100 ng genomic DNA was digested using restriction enzymes to create blunt ends (New England Biolabs, Beverly, MA, USA) in the condition following the manufacturer's protocol. The choice of restriction enzymes to determine the chromosomal breakpoint for a specific fusion gene using the ligation-mediated PCR relies on a prior analysis of the restriction enzyme map of the genomic sequences adjacent to the exon of the new exon junction. Restriction enzymes having no cutting site close to the exon adjacent to the new exon junction are the top priority for the library preparation, otherwise, the ligation-mediated PCR could have a very low chance of obtaining an amplicon containing the chromosomal breakpoint when adopting a gene-specific primer residing on the exon adjacent to the new exon junction in the ligation-mediated PCR. For this reason, DraI digestion was chosen for the identification of the chromosomal breakpoint of the *RET* fusion gene, and SmaI or EcoRV was chosen for the identification of the chromosomal breakpoint of the *ALK* fusion gene. After digestion, the DNA fragments were purified using phenol-chloroform extraction and ethanol precipitation procedure. The A-tailing to the blunt end DNA fragments was performed using Taq DNA polymerase (Sigma-Aldrich, St Louis, MO, USA) with 1.5 mM $MgCl_2$, and 330 μM deoxyadenosine triphosphate (dATP) at 72°C for 1 hr. After the reaction, the DNA fragments were again purified using phenol-chloroform extraction and ethanol precipitation procedure.

Adaptor ligation

The top strand of the adaptor used in this study was modified from the Illumina Index 2 (i5) Adapter sequence with an additional T in the 3' terminal, and the bottom strand of the adaptor was the complementary sequence of the read 1 sequence of Nextera Transposase Adapter with 5' phosphorylation and 3' inverted dT modification (Fig. 1). Both oligos were ordered from Integrated DNA Technologies (IDT, Singapore). To make the adaptor for ligation, 3 μM of each oligo was mixed and heated up to 95°C for 5 min and slowly cooled down to 25°C at a rate of 1°C/15 sec to form the final partial duplex adaptor. For the adaptor ligation, 100 ng of template DNA with 3' A-tailing and 25 pmol adaptor was incubated with the Blunt/TA ligase master mix (New England Biolabs) at room temperature for 2 hr. After ligation, adaptor-linked template DNA was purified using 0.85X volume of AMPureXP beads (Beckman Coulter, Brea, CA, USA)

Gene-specific primer extension and capture enrichment

For enrichment of the target sequence, the extension reaction was performed using a 20 ng adaptor-ligated template and 200 nM 5' Biotin-labeled specific primer with Emerald Taq Master Mix (Takara, Tokyo, Japan) in the following thermal cycling condition: 98°C, 3 min; 25 cycles of 98°C, 20 sec, 65°C, 30 sec, 72°C, 2 min; 72°C, 2.5 min. The extension product was then capture-enriched using Dynabeads™ MyOne™ Streptavidin C1 bead (Thermo Fisher, Waltham, MA, USA) after removing the unextended primer using 0.85X volume of Ampure XP beads.

PCR reaction for target amplification

Target amplification was performed using 1/10 volume of capture-enriched extension product, 200 nM of adaptor primer (AP) and gene-specific primers (GSP1, GSP2, and GSP3), and Emerald Taq Master Mix (Takara) with the following thermal cycling condition: 98°C, 3 min; 40 cycles of 95°C, 30 sec, 65°C, 30 sec, 72°C, 5 min; 72°C, 5 min.

Results

Optimizations in the advanced version of ligation-mediated PCR include choosing blunt-end restriction enzymes to avoid incomplete end repair, A-tailing to avoid random library fragment ligation, extension blocker in the tail of the bottom strand of adaptor to avoid adaptor end-repair priming in adaptor during the early cycle of amplification, and a biotin-labeled gene-specific primer extension and enrichment before final amplification to avoid amplification from adaptor primer alone (Fig. 1). A *RET* fusion gene-positive lung cancer cell line, LC-2/ad (*CCDC6* E1: *RET* E12), and an *ALK* fusion gene-positive lung cancer cell line, H3122 (*EML4* E13: *ALK* E20), were employed to demonstrate the performance of the method. In view of the *RET* fusion gene most frequently involved with exon 8 and exon 12 of *RET* gene and *ALK* fusion gene frequently involved with exon 20 of *ALK* gene, blunt-end restriction enzymes with no cutting site between intron 7 to intron 11 of *RET*, such as DraI, or with cutting site far away from the exon 20 of *ALK*, such as EcoRV and SmaI, was chosen to digest the genomic DNA of the two cell lines. After A-tailing, adaptor ligation, single primer extension, and streptavidin capture enrichment, 3 kb and 4.5 kb bands for LC-2/ad cells and H3122 cells were efficiently and successfully amplified in the first PCR reaction, respectively. The performance and reliability of the ligation-mediated PCR method for the identification of the chromosomal breakpoint of a fusion gene in clinical samples were also demonstrated using an *ALK* fusion gene-positive lung cancer sample identified in our earlier work [11] and a 5.1 kb amplicon was recovered (Fig. 2). The chromosomal breakpoints in these amplified fragments were then uncovered by Sanger sequencing with primer walking from either the adaptor primer or gene-specific primer side, and the chromosomal breakpoints were 1.7 kb, 0.3 kb and 1.1 kb upstream to the gene-specific primer side in LC-2/ad cells, H3122 cells and the *ALK* fusion gene-positive lung cancer sample, respectively (Fig. 2).

Discussion

Techniques that help identify or enrich unknown sequence fragments adjacent to known sequence regions have been widely applied in various molecular studies, such as the determination of the insertion site of transposon mutagenesis, identification of fusion partners, and sequencing library preparation. Each strategy is developed for specific study purposes and has its own merits and demerits. The inverse PCR is developed to help efficiently determine the insertion site of a known sequence in a genome using a primer pair extending outward from the insertion sequence to the flanked unknown sequence after restriction enzyme digestion and circularization after ligation [1]. Although the success of inverse PCR is dependent on the high efficiency of template circularization, linearization of the circulated templates by cutting in the middle of the insertion sequence is found to improve the amplification efficiency of inverse PCR because a circular template would pose restraints for template denature and primer binding [12,13]. However, the additional cutting step might also break the sequence linkage between the primer set and pose a great uncertainty to the method. Degenerate priming PCR can be easily performed without the need for prior DNA manipulation but will generally adopt a large degenerate primer set to increase the success rate and require a great deal of work in downstream screening. Owing to the nonspecific binding nature of degenerate primer, the method relies on multiple nested-PCR to discriminate nonspecific amplification using consecutive internal specific primers. Even so, substantial final amplification products could still come from the single degenerate primer and very often only short amplicons are returned [9, 10]. Reliable methods for retrieving long sequences would be valuable when target molecular events could not be resolved by simply mapping identified new sequences to the reference genome.

The method is an advanced version of the ligation-mediated PCR. The success of ligation-mediated PCR relies on an efficient and reliable attachment of adaptors to every terminal of the library fragment. Moreover, major obstacles for the success of ligation-mediated PCR, including adaptor self-ligation, template intermolecular random ligation, adaptor end-repair priming, and nonspecific amplification by adaptor primer, were tailored and optimized at once in this method. Efficient ligation of adaptors to the library fragments and suppression of intramolecular random ligation is achieved by conversion of the library fragment with 3' A-tailing using Taq-DNA polymerase. It should be noted that the non-template nucleotide addition by the polymerase is not a very explicit process. The efficiency of nucleotide addition is terminal sequence context-dependent and A-tailing efficiency is just slightly greater than the addition of other nucleotides [14–16]. To improve the A-tailing efficiency, a library fragment was generated using restriction enzymes to create blunt ends, and only dATP was supplemented in the Taq DNA polymerase-mediated non-template nucleotide addition reaction. The quality of the adaptor is another critical factor for the success of the ligation-mediated PCR. Although the bottom strand of the adaptor is synthesized with 5'-phosphorylation and 3'-inverted dT modification, the labeling efficiency is generally suboptimal. This could make adaptor end-repair filling occur in the early phase of the downstream PCR and let amplification from adaptor primer alone easily outcompete the target amplification. To avoid these complications, capture enrichment after extension using a biotin-labeled gene-specific primer is introduced to eliminate the nonspecific amplification by adaptor primer. Integration of all these modifications together makes an efficient and reliable target recovery possible without the need for additional nested PCR enrichment. To facilitate resolving the chromosomal breakpoint in the amplicon after the ligation-mediated PCR, the incorporation of the next-generation sequencing, especially those that generate long reads, such as Nanopore, can be envisaged by simply attaching the corresponding sequencing adaptor to the amplicon. One thing should be noted, however, that the current method

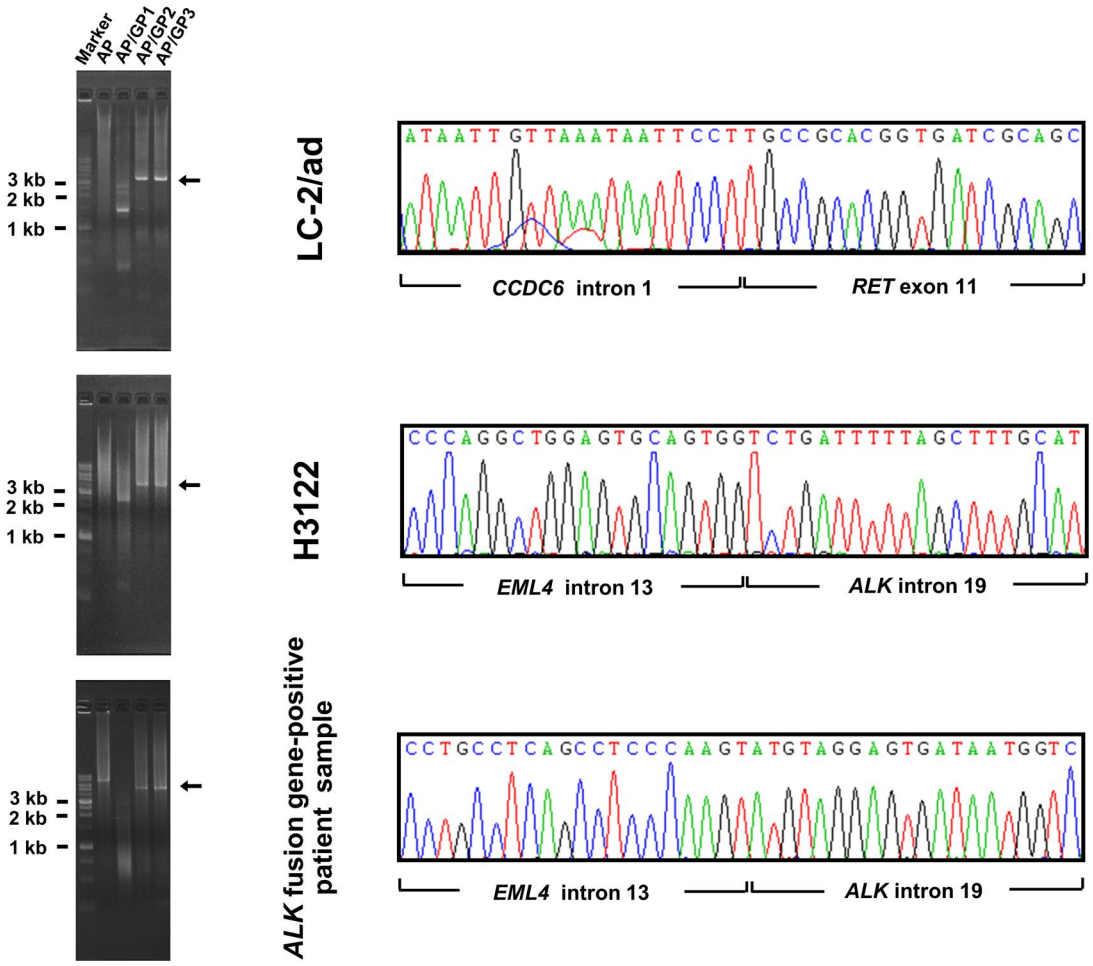

Figure 2. The application of the optimized ligation-mediated PCR method for identifying the chromosomal breakpoint of oncogenic fusion genes in LC-2/ad and H3122 cells, and an *ALK* fusion gene-positive lung cancer patient sample. The amplification products after the procedure were resolved using agarose gel electrophoresis. Target bands from the amplification by the gene-specific primer and adaptor primer are labeled with black arrows. The chromosomal breakpoint was confirmed by Sanger sequencing with primer walking.

relies on obtaining a single amplicon using a gene-specific primer on the exon adjacent to the new exon junction with the adaptor primer to retrieve the chromosomal breakpoint of a fusion gene. It would be challenging to identify a chromosomal breakpoint using the current method if the chromosomal breakpoint of a fusion gene is more than 20–30 kb away from the exon adjacent to the new exon junction beyond the current amplification limit of the long-range PCR, such as those involved in the minor *BCR-ABL* fusion gene formation. Besides, a restriction enzyme without any cutting site across an ultra-long sequence region between the chromosomal breakpoint and the exon adjacent to the new exon junction to make a single amplicon containing the chromosomal breakpoint in the ligation-mediated PCR possible is also unlikely. These would need to incorporate multiple gene-specific primers spanning the entire region, including those residing in the adjacent intron, in the ligation-mediated PCR to amplify and resolve the entire sequence piece by piece simultaneously using the next-generation sequencing to identify the chromosomal breakpoint of a fusion gene.

Potential applications of the optimized method in targeted next-generation sequencing

Targeted next-generation sequencing is an important tool for disease diagnosis and screening. Owing to high levels of multiplexing and amplification, off-target reads frequently occur. Current library preparation in targeted next-generation sequencing often adopts the long hybridization capture probe [17,18] or padlock probe [19, 20] for target region enrichment, which often requires a significant amount of input DNA and is not applicable when only a small quantity of sample is available. The procedure established in the method retains both the advantages of the hybridization capture-based method by harnessing the biotin-labeled primer for target enrichment and the amplicon-based method using an internal gene-specific primer for target amplification could permit a lower amount of DNA input, higher on-target rate, better uniformity, and lower cost. The advanced version of ligation-mediated PCR could further extend the application scope of targeted sequencing and benefit more patients.

Acknowledgements

The authors would like to express their gratitude to Chang Gung Memorial Hospital, Chiayi for their support to the study. The authors also would like to acknowledge the Research Specimen Processing Laboratory, Chang Gung Memorial Hospital, Chiayi, and Biobank, Chang Gung Memorial Hospital, Chiayi for providing specimens for the study.

Author contributions

Jrhau Lung (Conceptualization [equal], Data curation [equal], Funding acquisition [equal], Methodology [equal], Writing—original draft [equal], Writing—review & editing [equal]), Ming-Szu Hung (Data curation [equal]), Chao-Yu Chen (Investigation [equal]), Tsung-Ming Yang (Resources [equal]), Chin-Kuo Lin (Resources [equal]), Yu-Hung Fang (Resources [equal]), Yuan-Yuan Jiang (Formal analysis [equal]), Hui-Fen Liao (Resources [equal]), and Yu-Ching Lin (Resources [equal], Supervision [equal])

Funding

The study was supported by grants CORPG6K0041-3, CMRPG6N0151-3, and CORPG6N0311 from Chang Gung Memorial Hospital, Chiayi, Taiwan.

Conflict of interest statement

All authors declare that there is no competing interests for the study.

Data availability

All data supporting the study are disclosed in this article. Other data related to the study are available from the corresponding author upon request.

References

1. Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. *Genetics* 1988;**120**:621–3.
2. Peng Z, Zhao Z, Nath N *et al.* Generation of long insert pairs using a Cre-LoxP Inverse PCR approach. *PLoS One* 2012;**7**:e29437.
3. Uchiyama T, Watanabe K. Improved inverse PCR scheme for metagenome walking. *Biotechniques* 2006;**41**:183–8.
4. Espelund M, Jakobsen KS. Cloning and direct sequencing of plant promoters using primer-adapter mediated PCR on DNA coupled to a magnetic solid phase. *Biotechniques* 1992;**13**:74–81.
5. Riley J, Butler R, Ogilvie D *et al.* A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res* 1990;**18**:2887–90.
6. Tsuruta Y, Iwagami S, Furue S *et al.* Detection of human T cell receptor cDNAs (α, β, γ and δ) by ligation of a universal adaptor to variable region. *J Immunol Methods* 1993;**161**:7–21.
7. Lang M, Orgogozo V. Identification of homologous gene sequences by PCR with degenerate primers. *Methods Mol Biol* 2011;**772**:245–56.
8. Liu YG, Mitsukawa N, Oosumi T *et al.* Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J* 1995;**8**:457–63.
9. Singer T, Burke E. High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. *Methods Mol Biol* 2003;**236**:241–72.
10. Wu L, Di DW, Zhang D *et al.* Frequent problems and their resolutions by using thermal asymmetric interlaced PCR (TAIL-PCR) to clone genes in Arabidopsis T-DNA tagged mutants. *Biotechnol Biotechnol Equip* 2015;**29**:260–7.
11. Lung J, Lin YC, Hung MS *et al.* A sensitive and high throughput TaqMan-based reverse transcription quantitative polymerase chain reaction assay efficiently discriminates ALK rearrangement from overexpression for lung cancer FFPE specimens. *Lung Canc* 2016;**94**:114–20.
12. Laghi L, Randolph AE, Malesci A *et al.* Constraints imposed by supercoiling on in vitro amplification of polyomavirus DNA. *J Gen Virol* 2004;**85**:3383–8.
13. Ochman H, Ayala FJ, Hartl DL. Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. *Methods Enzymol* 1993;**218**:309–21.
14. Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer Modifications that facilitate genotyping. *BioTechniques* 1996;**20**:1004–10.
15. García PB, Robledo NL, Islas ÁL. Analysis of non-template-directed nucleotide addition and template switching by DNA polymerase. *Biochemistry* 2004;**43**:16515–24.
16. Zhou Y, Bo F, Tian T *et al.* Excessive addition split peak formed by the non-templated nucleotide addition property of Taq DNA polymerase after PCR amplification. *Front Bioeng Biotechnol* 2023;**11**:1180542.
17. Kozarewa I, Armisen J, Gardner AF *et al.* Overview of target enrichment strategies. *Curr Protoc Mol Biol* 2015;**112**:7.21. 21–7.21.23.
18. Syvänen AC, Bengtström M, Tenhunen J, Söderlund H. Quantification of polymerase chain reaction products by affinity-based hybrid collection. *Nucleic Acids Res* 1988;**16**:11327–38.
19. Li JB, Gao Y, Aach J *et al.* Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* 2009;**19**:1606–15.
20. Shen P, Wang W, Krishnakumar S *et al.* High-quality DNA sequence capture of 524 disease candidate genes. Proc Natl Acad Sci USA 2011;**108**:6549–54.

Biology Methods and Protocols, 2024, 00, 1–5
https://doi.org/10.1093/biomethods/bpae037
Methods Article