Clinica Chimica Acta 495 (2019) 562–569

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Evaluation of analytical factors associated with targeted *MEFV* gene sequencing using long-range PCR/massively parallel sequencing of whole blood DNA for molecular diagnosis of Familial Mediterranean fever

Takayuki Ishige[a,*], Sakae Itoga[a], Kenji Kawasaki[a], Emi Utsuno[b], Minako Beppu[a,b], Setsu Sawai[a,b], Motoi Nishimura[a,b], Tomohiko Ichikawa[b], Fumio Nomura[b], Kazuyuki Matsushita[a,b]

[a] Division of Laboratory Medicine, Chiba University Hospital, Chiba, Japan
[b] Division of Clinical Genetics, Chiba University Hospital, Chiba, Japan

ARTICLE INFO

Keywords:
Long-range PCR
Post-PCR cleanup
Library preparation
Massively parallel sequencing

ABSTRACT

Background: Long-range PCR (LR-PCR) is used to enrich the target regions of the genome. This study aimed to establish the pipeline of targeted gene sequencing using LR-PCR and massively parallel sequencing (MPS).
Methods: The 14-kb-long *MEFV* gene, including the entire coding exons, was selected as a target gene and amplified using LR-PCR. The evaluated analytical factors were as follows: LR-PCR conditions, three types of post-PCR cleanup methods, and two types of MPS library preparation methods.
Results: With regard to LR-PCR conditions, Tks Gflex DNA polymerase at 7-min (30-s/kb) annealing/extension with 100-ng genomic DNA input had the highest yield. Regarding post-PCR purification methods, the magnetic beads-based method had high recovery and purity. In the MPS library preparation methods, the ligation-based method had a higher base coverage in the target (94.58%), uniformity of base coverage (99.95%), and target bases with no strand bias (97.40%). The exonic variants determined by Sanger sequencing were detected by both ligation- and transposon-based methods.
Conclusions: Various analytical factors were evaluated, and the pipeline of targeted gene sequencing using LR-PCR and MPS was established. These data can enable the optimization of targeted gene sequencing using LR-PCR and MPS in the clinical laboratory.

1. Introduction

In recent laboratory medicine, polymerase chain reaction (PCR) and sequencing are essential technologies for molecular diagnosis [1]. Various types of PCR such as real-time PCR, multiplex PCR, and long-range PCR (LR-PCR) are used to analyze specific targets in the genome [1]. LR-PCR for > 10 kb is challenging because it requires detailed examination of DNA polymerases and cycling conditions; however, it is a powerful tool for the amplification of multiple target regions (e.g., entire coding exons of a gene) [2]. For sequencing, two types of technologies, Sanger sequencing and massively parallel sequencing (MPS), are applied in the clinical laboratory [1]. Sanger sequencing is the gold standard technique for DNA sequencing [1]. The applications of MPS are rapidly increasing because of its cost-effectiveness and scalability.

Three methods, including LR-PCR, multiplex PCR, and hybridization capture, were mainly used for target enrichment of MPS [3]. In these methods, LR-PCR is cost-effective and widely used for targeted gene sequencing (e.g., *BRCA1*, *BRCA2*, *PKD1*, *PKD2*, *PMS2*, and *HLA* genes) [4–9]. Fig. 1 shows the general workflow of LR-PCR and MPS. Initial LR-PCR is required for the amplification of target regions, purification of PCR products for downstream applications, and library construction for MPS. These analytical factors affect the quality of DNA sequencing results. Therefore, analytical factors associated with LR-PCR and MPS must be evaluated and optimized for accurate and effective molecular diagnosis in the clinical laboratory.

Familial Mediterranean fever (FMF) is an autoinflammatory disorder characterized by recurrent attacks of fever and inflammation in the peritoneum, synovium, or pleura, accompanied by pain [10]. The international FMF consortium identified the *MEFV* gene as the causative gene for FMF in 1997 [11,12]. *MEFV* (accession number: NG_007871) is located on chromosome 16p13.3 and comprises 14,600 bases, 10 coding exons, and 781 amino acids in the translated protein (Pyrin) [13]. *MEFV* is sequenced in clinical laboratories to diagnose FMF.

This study aimed to establish the analytical pipeline of targeted

* Corresponding author at: Division of Laboratory Medicine, Chiba University Hospital, 1-8-1 Inohana, Chuo-ward, Chiba-City, Chiba 260-8677, Japan.
E-mail address: ishige-t@chiba-u.jp (T. Ishige).

https://doi.org/10.1016/j.cca.2019.06.001
Received 18 March 2019; Received in revised form 27 May 2019; Accepted 3 June 2019
Available online 04 June 2019
0009-8981/

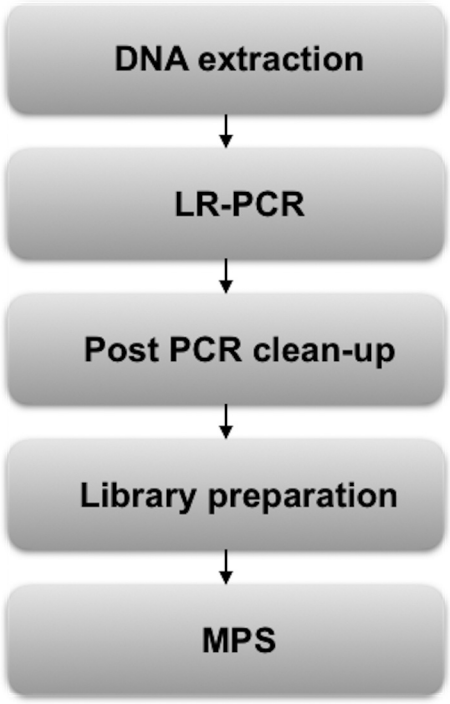

Fig. 1. Workflow of LR-PCR and MPS. In this study, LR-PCR, post-PCR clean-up, and MPS library preparation methods were evaluated.

gene sequencing by evaluating various analytical factors associated with it using LR-PCR and MPS. *MEFV* was selected as the target gene for two reasons; we have been conducting molecular diagnosis of FMF during the past 20 years [14,15]. Also, *MEFV* is a suitable target for LR-PCR and MPS because of its gene size and number of exons. The evaluated analytical factors included LR-PCR conditions (four types of DNA polymerases with proofreading activity and cycle protocols), three types of post-PCR clean-up methods (spin column, magnetic beads, and enzyme-based methods), and two types of MPS library preparation methods (ligation- and transposon-based methods).

2. Materials and methods

2.1. Human genomic DNA

The human genomic DNA, which was extracted from the pooled whole blood of multiple healthy volunteers by isolating high molecular weight DNA according to the method described in a study by Blin and Stafford [16], was purchased from Takara (Tokyo, Japan).

2.2. Primer design

The primer pairs for LR-PCR were designed using the Primer3 program [17]. For LR-PCR by two-step amplification, the melting temperature (Tm) of the primer was 68 °C. The modified general primer selection conditions were as follows: primer size, minimum 25 bases, optimum 30 bases, and maximum 35 bases; primer Tm, minimum 65 °C, optimum 68 °C, and maximum 71 °C. The primers for LR-PCR used in this study were 5′-TCC TCT GAA CCT GTA AGA AGA GAA CAC AGC-3′ (forward, Tm 67.9 °C, 30 bases, position hg19 chr16 [NC_000016.9]: 3306, 756–3306,785) and 5′-ATC CAT GGT GTG TCA TCA GTA CAT GTC TTC-3′ (reverse, Tm 68.0 °C, 30 bases, position hg19 chr16 [NC_000016.9]:3,292,920-3,292,949). The theoretical product size of LR-PCR was 13,866 bp, and the GC content was 51.5%.

2.3. DNA quantification and agarose gel electrophoresis

DNA quantification was performed using QuantiFluor ONE dsDNA System (Promega) on Quantus fluorometer (Promega) according to the manufacturer's protocol. Agarose gel electrophoresis was performed using E-Gel iBase Power System and E-Gel Safe Imager (Thermo Fisher Scientific). The E-Gel 0.8% Agarose (Thermo Fisher Scientific) and E-Gel SizeSelect 2% (Thermo Fisher Scientific) were used for LR-PCR and the MPS library size selection, respectively.

2.4. LR-PCR

2.4.1. Reaction mixture

Ex Taq DNA polymerase (Ex Taq; hot start version, Takara): The 50-μL reaction mixture contained 1 × Ex Taq buffer, 0.025 U/μL of Ex Taq HS DNA polymerase, 0.2 mM of dNTP, 0.2 μM of primer pair, and 100 ng of genomic DNA.

LA Taq DNA polymerase (LA Taq; hot start version, Takara): The 50-μL reaction mixture contained 1 × LA PCR buffer II, 0.05 U/μL of LA Taq HS DNA polymerase, 0.4 mM of dNTP, 0.2 μM of primer pair, and 100 ng of genomic DNA.

PrimeSTAR GXL DNA polymerase (PS-GXL; Takara): The 50-μL reaction mixture contained 1 × PrimeSTAR GXL buffer, 0.025 U/μL of PrimeSTAR GXL DNA polymerase, 0.2 mM of dNTP, 0.2 μM of primer pair, and 100 ng of genomic DNA.

Tks Gflex DNA polymerase (Tks-Gflex; Takara): The 50-μL reaction mixture contained 1 × Tks Gflex buffer, 0.025 U/μL of Tks Gflex DNA polymerase, 0.2 μM of primer pair, and 100 ng of genomic DNA.

2.4.2. Cycling protocol

Two-step amplification was used for LR-PCR. Different annealing/extension conditions, 3.5 min (15 s/kb), 7 min (30 s/kb), 10.5 min (30 s/kb), and 14 min (60 s/kb), were evaluated. The cycling conditions were as follows: 94 °C for 2 min; 30 cycles at 98 °C for 10 s, and 68 °C for 3.5 min, 7 min, 10.5 min, or 14 min.

2.5. Post-PCR clean-up

2.5.1. Spin column-based method

The NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany) were used for the spin column-based method, according to the manufacturer's protocol.

2.5.2. Magnetic bead-based method

The Agencourt AMPure XP (Beckman Coulter) was used for the magnetic bead-based method, according to the manufacturer's protocol.

2.5.3. Enzyme-based method

Enzymatic purification was performed using exonuclease I (New England Biolabs, MA, USA) and shrimp alkaline phosphatase (New England Biolabs). Subsequently, 20 U of exonuclease I and 2 U of shrimp alkaline phosphatase were added in 10 μL of PCR products and incubated for 15 min at 37 °C. The mixture was incubated for 15 min at 80 °C to inactivate both enzymes.

2.6. Sanger sequencing

Sanger sequencing of eight regions of *MEFV* was performed using the 3130 Genetic Analyzer (Thermo Fisher Scientific, MA, USA). The BigDye Terminator v3.1 cycle sequencing kit (Thermo Fisher Scientific) was used for the cycle sequencing reaction. Each 10-μL reaction mixture contained 1 μL of BigDye Terminator ready reaction mix, 2 μL of 5 × sequencing buffer, 0.4 μM of each sequencing primer (Supplementary Table 1), and purified PCR products; the cycling conditions were as follows: 96 °C for 1 min, followed by 25 cycles at 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 1 min. The cycle sequencing products were purified

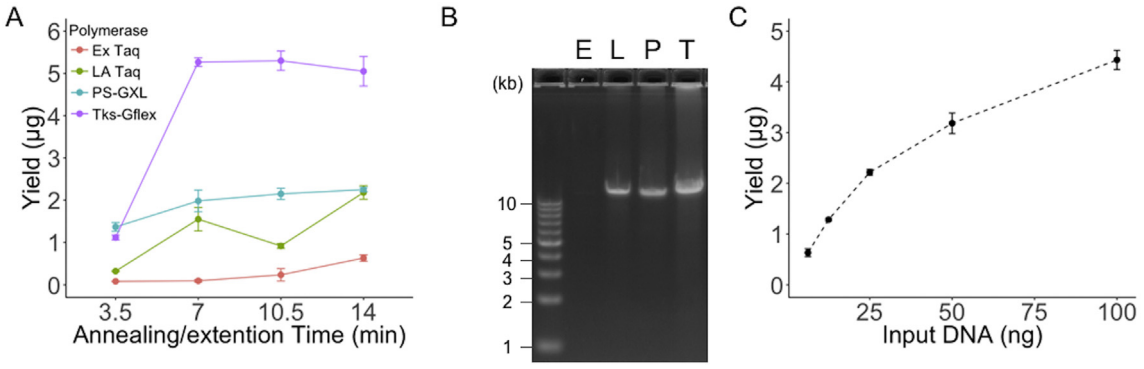

Fig. 2. Evaluation of various LR-PCR conditions. (A) Comparison of PCR product yields amplified under the conditions of different DNA polymerases and annealing/extension times. Error bars represent standard deviations of three samples. (B) Agarose gel electrophoresis of 14-kb amplicon at 30-s/kb annealing/extension. Abbreviations: E, Ex Taq DNA polymerase; L, LA Taq DNA polymerase; P, PS-GXL DNA polymerase; T, Tks Gflex DNA polymerase. (C) Comparison of PCR product yields amplified using different initial input DNA templates. Error bars represent standard deviations of three samples.

using the BigDye XTerminator purification kit (Thermo Fisher Scientific), according to the manufacturer's protocol. Allelic frequency was determined using a previously described method [18].

2.7. Library preparation

2.7.1. Ligation-based method

The Ion Xpress Plus Fragment Library Kit (Thermo Fisher Scientific) was used as the ligation-based MPS library preparation method according to the manufacture's instruction. Briefly, the Ion Shear Plus Enzyme was used to fragment 1000 ng of PCR products for 50 min at 37 °C. After the fragmented PCR products were purified, adaptor ligation and nick-repair reactions were performed.

2.7.2. Transposon-based method

The MuSeek Library Preparation Kit for Ion Torrent (Thermo Fisher Scientific) was used as the transposon-based MPS library preparation method according to the manufacture's instruction. Briefly, 100 ng of PCR products was used for the transposon-mediated fragmentation reaction for 3 min at 30 °C. After the fragmented PCR products were purified, adaptor addition PCR was performed.

2.8. Massively parallel sequencing

The sequencing template of 200 base-read libraries was prepared using the Ion PGM Template OT2 200 kit (Thermo Fisher Scientific) with the Ion OneTouch system (Thermo Fisher Scientific). MPS was performed using the Ion PGM Sequencing 200 kit v2 (Thermo Fisher Scientific) and the Ion 314-chip kit v2 (Thermo Fisher Scientific) with the Ion Torrent PGM semiconductor sequencer (Thermo Fisher Scientific).

2.9. Real-time PCR

Real-time PCR was performed to assess the PCR amplification efficiency of each *MEFV* region (total 20 regions). Genomic DNA with serial dilution (100 ng/µL, 50 ng/µL, 25 ng/µL, and 12.5 ng/µL) was used. Real-time PCR was performed using a LightCycler 480 instrument (Roche). The 10-µL reaction mixture consisted of 1 × LightCycler 480 SYBR Green I Master, 0.5 µM of primer (Supplementary Table 2), and 1 µL of genomic DNA, and the cycling conditions were as follows: initial denature, 95 °C for 10 min; real-time PCR, 45 cycles at 99 °C for 10 s, 60 °C for 15 s, and 72 °C for 15 s (signal acquisition); melting, 99 °C for 5 s, 65 °C for 1 min, with a continuous increase in temperature from 65 °C to 99 °C at a rate of 0.11 °C/s with 5 signal acquisitions per degree; and cooling, 40 °C for 30 s.

2.10. Bioinformatics and statistical analysis

Sequencing data were mapped on the *Homo sapiens* genome (hg19) on the Torrent server. The MPS results were visualized on Integrative Genomics Viewer 2.5.0 [19]. The Ion Reporter software (version 5.2, Thermo Fisher Scientific) was used to detect sequence variants. The coverage analysis of each base of the target region and subsampling was performed using Torrent Suite Software Coverage Analysis Plugin (version 5.4.0.5, Thermo Fisher Scientific) and SAMtools [20], respectively. Statistical analysis and data visualization were performed using R version 3.5.1 statistical software [21]. Sliding window analysis was performed using the "evobiR" package. The window size was set at 100. Kruskal-Wallis rank sum test, paired *t*-test and Spearman rank-order correlation test were performed using the "stats" package. Data visualization and locally estimated scatterplot smoothing (LOESS) regression were performed using the "ggplot2" package. A p-value of < 0.05 was considered to be statistically significant.

3. Results

3.1. Comparison of DNA polymerases for LR-PCR

First, we compared four types of DNA polymerases with proofreading activity for LR-PCR. The PCR product yield was positively correlated with annealing/extension time and input DNA volume (Fig. 2A). In Ex Taq and LA Taq DNA polymerases, the PCR product yield increased with an increase in annealing/extension time. In contrast, the yield plateaued at 7-min (30-s/kb) annealing/extension with regard to PS-GXL and Tks Gflex DNA polymerases. The highest yield was obtained using Tks Gflex DNA polymerase at 7-min (30-s/kb) annealing/extension (Fig. 2B; mean, 5.3 µg). In addition, the PCR product yield was correlated with the input genomic DNA volume (Fig. 2C). Furthermore, LR-PCR using Tks Gflex DNA polymerase at 7-min (30-s/kb) annealing/extension with 100-ng genomic DNA extracted using other methods (magnetic bead-based automation and silica column-based methods) was successful (data not shown).

3.2. Evaluation of post-PCR clean-up methods

The recovery and purity of post-PCR purification methods were evaluated. Purity was determined from the quality values (QV) of Sanger sequencing. We compared three types of purification methods: spin columns, magnetic beads, and enzymes. Of these, enzymatic purification had the highest recovery, followed by magnetic bead and spin column purifications (Fig. 3A). The mean recoveries (%) and standard deviations of each method were as follows: enzymes, 96.4%, 0.9; magnetic beads, 73.3%, 2.9; and spin columns, 29.9%, 12.4,

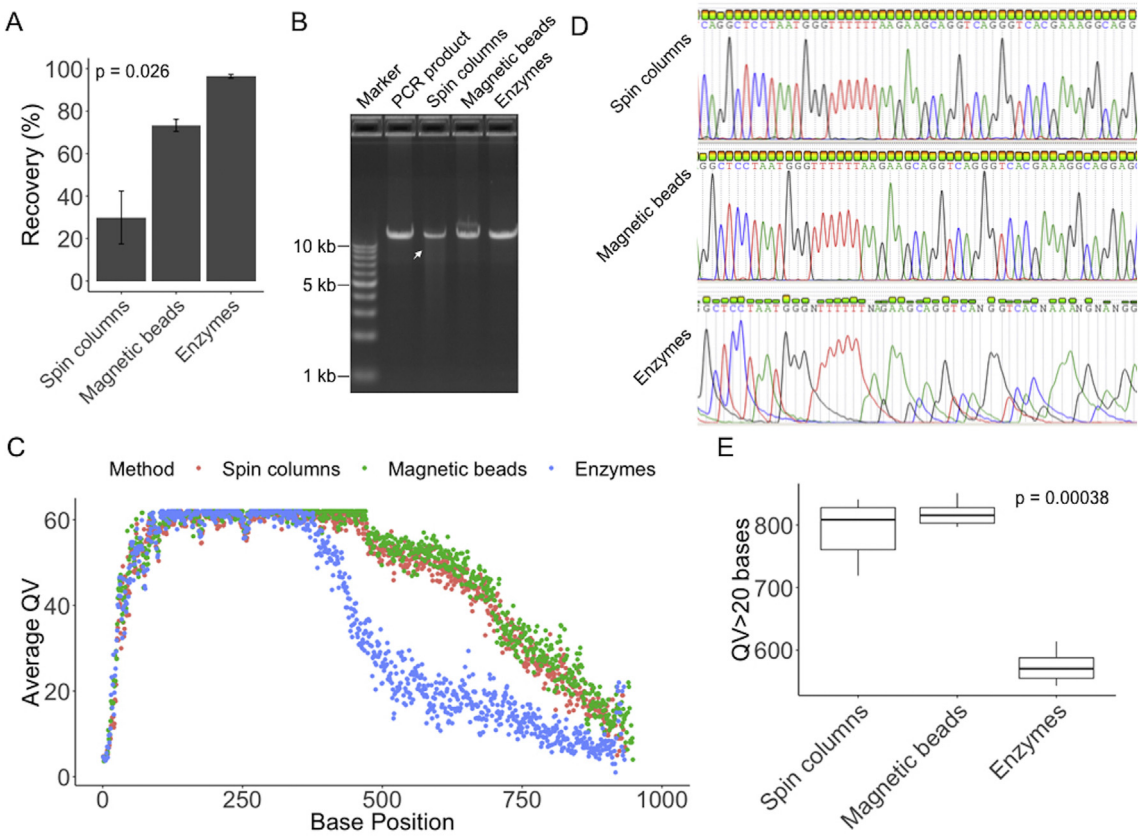

Fig. 3. Comparison of post-PCR clean-up methods. (A) Extraction recovery of three purification methods. Error bars represent standard deviations of three samples. (B) Electrophoresis of PCR products with or without purification. The marker was a 1-kb ladder marker. (C) Mean QV distribution of each nucleotide position in Sanger sequencing. Mean QV data were obtained from eight sequenced region of *MEFV*. (D) Typical electropherograms of Sanger sequencing at 530–580 base positions. (E) The number of QV > 20 bases using three purification methods. The *p*-values were determined by Kruskal-Wallis rank sum test.

respectively. Smearing was observed in agarose gel ELP of PCR products purified by spin column purification (Fig. 3B). In Sanger sequencing, the QV distributions were similar in the three methods up to the 400 base position (Fig. 3C). However, the QV of enzymatic purification decreased more steeply than that of silica column and magnetic bead purifications (Fig. 3C–D). Therefore, QV > 20 bases in enzymatic purification were significantly fewer than those of silica column and magnet bead purifications (Fig. 3E). The mean QV > 20 bases and standard deviations of each method were as follows: enzymes, 573, 24; magnetic beads, 819, 21; and spin columns, 794, 46, respectively.

3.3. Comparison of MPS library preparation methods

Two MPS library preparation methods, ligation- and transposon-based methods, were compared. In contrast to the ligation-based method, library amplification PCR was required for adaptor addition in the transposon-based method. Table 1 shows the result of coverage analysis. Compared with the transposon-based method, the ligation-based method had greater base coverage in the target gene (Table 1 and Fig. 4A). However, mapped sequence reads on chromosomes, except for chromosome 16, using transposons were significantly higher than those using ligation (Fig. 4B). These ratios were well correlated with chromosome length (Fig. 4C). Furthermore, the uniformity of base coverage and target bases with no strand bias were higher in the ligation-based method than in the transposon-based method. Fig. 5A shows normalized base coverage at the *MEFV* region on chromosome 16. The base coverage of exon 2 using the transposon-based method was extremely lower than that using the ligation-based method. Sliding window analysis of the *MEFV* region revealed a negative correlation between

Table 1
Base coverage statistics of MPS.

Category	Ligation	Transposon
Base coverage on target[a]	94.58%	87.14%
Uniformity of base coverage[b]	99.95%	96.98%
Target bases with no strand bias[c]	97.40%	95.41%

[a] The percentage of all bases covered by reads aligned to the reference that covered bases in target regions.

[b] The percentage of bases in all targeted regions covered by at least 0.2× the mean base coverage depth.

[c] The percentage of all targets that did not show bias towards forward or reverse strand read alignments. An individual target is considered to have read bias if it has at least 10 reads and the fraction of forward or reverse reads to total reads is > 70%.

normalized base coverage and local GC content (Fig. 5B). Real-time PCR experiments of each exon were performed to confirm the effect of GC content on PCR efficiency. PCR efficiency (%) of the middle exon 2 region was < 50 (Fig. 5C). The Tm of PCR products of this region was very high because of its high GC content (Tm, 95.5 °C; GC content, 75%). The Tm values and GC contents of all regions were well correlated, and these were negatively correlated with PCR efficiency (Fig. 6).

Moreover, the exonic variants determined by Sanger sequencing were detected using MPS (both ligation- and transposon-based methods). The variant frequencies detected using Sanger sequencing and MPS were well correlated (Table 2). The coefficient of determination (RSQ) using Sanger sequencing in the ligation-based method (RSQ = 0.98) was higher than that in the transposon-based method (RSQ = 0.95). The MPS result of ligation-based method and detected

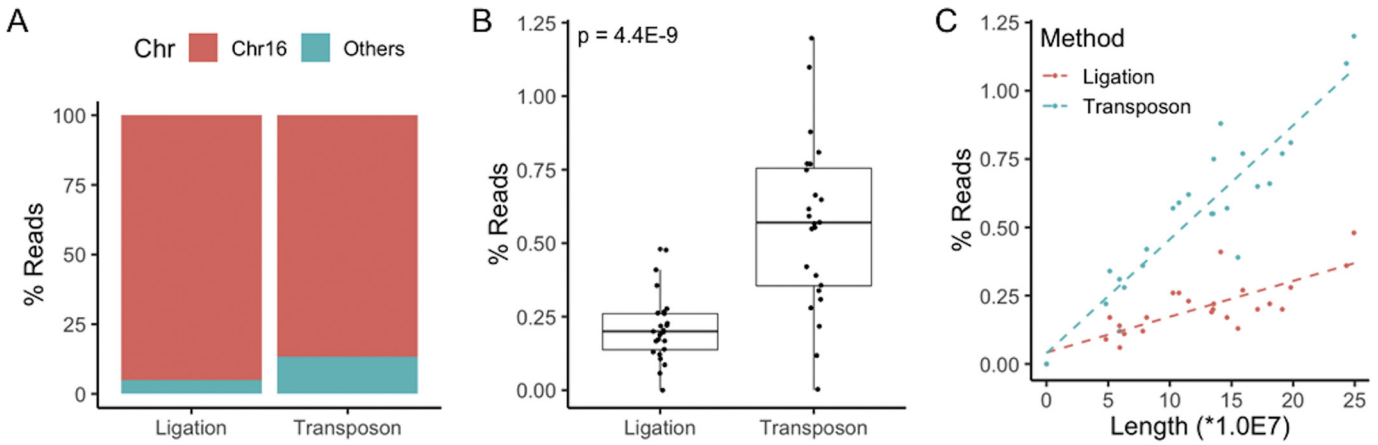

Fig. 4. Comparison of the sequence reads mapped on hg19 obtained by two MPS library preparation methods. (A) Percentage of mapped sequence reads on chromosome 16 and others. (B) Percentage of mapped reads on chromosomes, except for chromosome 16. Paired *t*-test determined the p-value. (C) Comparison of chromosome length and percentage mapped reads on chromosomes, except for chromosome 16. Spearman rho and p-value were as follows: ligation, rho = 0.73, $p = 5.7E-5$; and transposon, rho = 0.89, $p = 7.7E-9$.

variants in *MEFV* gene were shown in Fig. 7. Thirty-eight variants were detected by Ion Reporter software, and almost were in non-coding regions.

We performed the coverage analysis of subsampling MPS data to evaluate the association between minimum-base coverage and total sequence reads (Fig. 8). The slope and intercept were as follows: ligation, slope = 0.0016, intercept = −5.5; transposon, slope = 0.00016, intercept = 0.69. The required total sequence reads in the ligation- and transposon-based methods were 12,526 and 197,748, respectively, to obtain a minimum base coverage of 30.

4. Discussion

In this study, we evaluated various analytical factors to establish the pipeline of targeted gene sequencing using LR-PCR and MPS. The 14-kb-long *MEFV* gene was selected as the target gene for LR-PCR. LR-PCR conditions, post-PCR purification methods, and MPS library preparation methods were evaluated. First, LR-PCR using Tks Gflex DNA

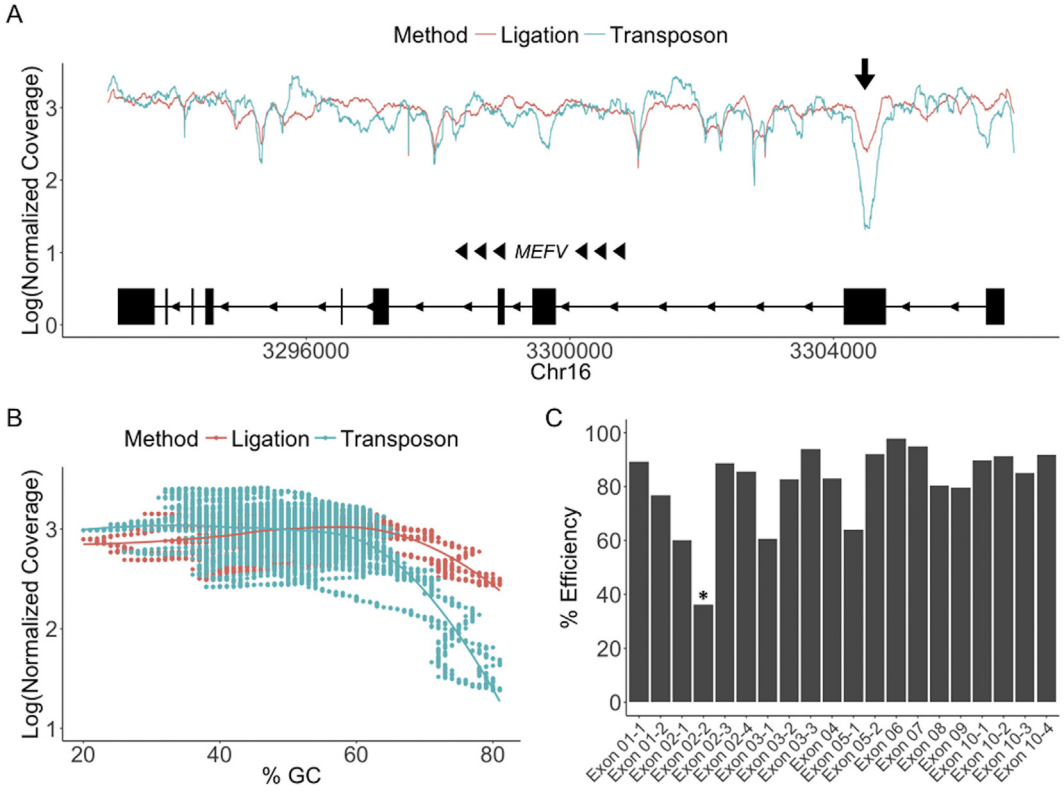

Fig. 5. Comparison of the normalized base coverage in the *MEFV* gene obtained by two MPS library preparation methods. (A) The normalized coverage of each nucleotide position in *MEFV*. The gene structure of *MEFV* is shown at the bottom of the plot. Filled boxes indicate coding exons. *MEFV* is located in the reverse strand of hg19 reference genome. The arrow indicates the exon 2 region of *MEFV*. (B) Association between normalized coverage and local GC content determined by sliding window analysis. The lines indicate LOESS regression lines. (C) Real-time PCR efficiencies of each region. The asterisk indicates the middle exon 2 of *MEFV* with lower PCR efficiency (< 50%).

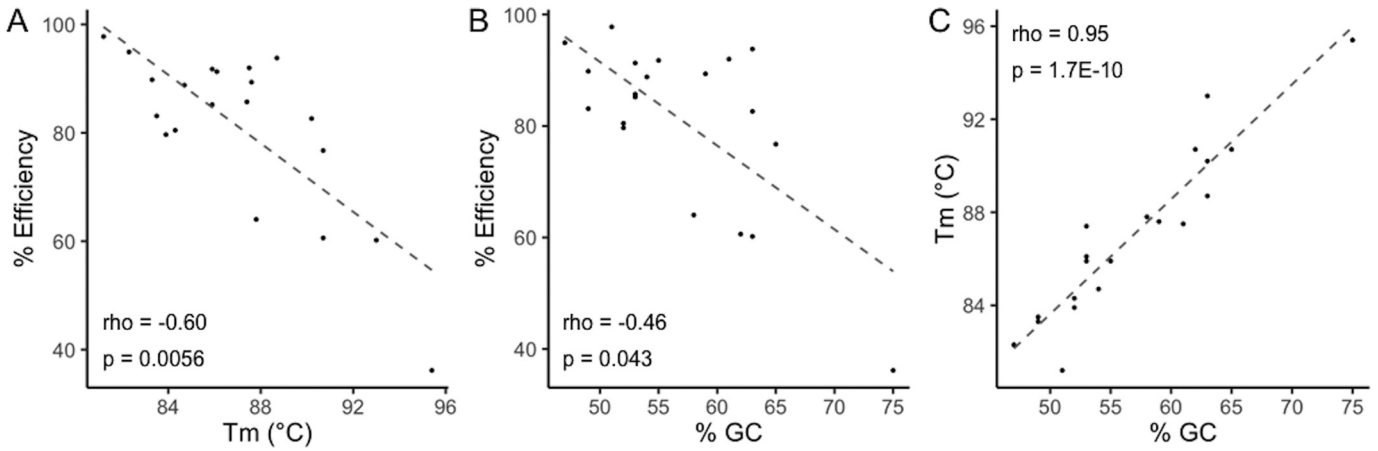

Fig. 6. Comparison of PCR efficiency (%), Tm (°C), and GC content (%). Real-time PCR of 20 regions of *MEFV* were performed. PCR efficiency and Tm were determined by real-time PCR experiments of each *MEFV* region. (A) Efficiency versus Tm. (B) Efficiency versus GC content. (C) Tm versus GC content. The p-values were determined by Spearman rank-order correlation test.

Table 2
Comparison of exonic variants in the *MEFV* gene detected by Sanger sequencing and MPS.

Chr16	Exon	Variant	Frequency		
			Sanger	MPS-L	MPS-T
3293888	9	c.1764G > A (p.P588=)	0.21	0.25	0.20
3297073	5	c.1530T > C (p.D510=)	0.78	0.77	0.74
3297175	5	c.1428A > G (p.Q476=)	0.74	0.77	0.68
3297181	5	c.1422G > A (p.E474=)	0.79	0.79	0.75
3299749	3	c.942C > T (p.R314=)	0.80	0.85	0.80
3304463	2	c.605G > A (p.R202Q)	0.18	0.21	0.23
3304573	2	c.495C > A (p.A165=)	0.52	0.62	0.62
3304654	2	c.414A > G (p.G138=)	0.57	0.66	0.53
3304762	2	c.306T > C (p.D102=)	0.60	0.66	0.55

Note: RefSeq accession numbers of *MEFV* gene and Pyrin protein are NM_000243 and NP_000234, respectively. The frequencies of variants were variable because this genomic DNA was extracted from the pooled whole blood of multiple healthy volunteers.
Abbreviations: MPS-L, massively parallel sequencing with ligation-based library; MPS-T, massively parallel sequencing with transposon-based library.

polymerase at 7-min (30-s/kb) annealing/extension with 100-ng genomic DNA input had the highest yield and was suitable for the amplification of 14-kb-long PCR products. Second, with regard to post-PCR purification methods, the magnetic bead-based method had high recovery and purity. Subsequently, regarding MPS library preparation methods, the ligation-based method had high base reads on the target, uniformity of coverage, and target bases with no strand bias. Furthermore, the RSQ of variant frequency was well correlated between Sanger sequencing and MPS using a ligation-based library. Therefore, the ligation-based method without library amplification PCR is preferred for MPS library preparation.

The LR-PCR performed in this study was similar condition to the previous study, regarding to the use of proofreading DNA polymerases, two-step cycling, and short denaturation time [2]. In the single-gene analysis, LR-PCR is considered to be an easy-to-use and cost-effective target-enrichment method. Compared with other methods, the advantages of LR-PCR are the following: 1) Sanger sequencing for the variant confirmation is applicable using LR-PCR products, 2) the amplification bias of each region is minimized, and 3) well-designed primers can specifically amplify the target gene despite the existence of highly homologous pseudogenes such as *PKD1* and *PMS2* genes [6–8]. Thus, LR-PCR is useful for the analyses of genes with high GC-rich regions and pseudogenes. In addition, the median gene length is 21,209 bases according to the consensus coding sequence (CCDS) database [22]. Thus, several genes can be amplified using LR-PCR with one primer pair. In addition, LR-PCR can analyze the entire regions of gene including exons and introns [23].

Among the post-PCR cleanup methods evaluated in this study, enzymatic clean-up has the highest recovery and shortest handling time (approximately 5 min). However, the lack of purity affected the quality of Sanger sequencing. Although the silica column-based method had high purity, the recovery was the lowest. Therefore, the magnetic bead-based method was considered to be a suitable post-PCR cleanup method of LR-PCR because of its recovery and purity. MPS has been reported to provide false-positive results because of chemical, platform, and bioinformatic biases [24]. Confirmatory tests will be required for determining the clinically relevant and/or unknown variants. LR-PCR products purified using the magnetic bead-based method can be used for both MPS and Sanger sequencing, which are used to confirm the variants detected using MPS.

In the library preparation for MPS, adaptor addition PCR was required in the transposon-based method. Although the input DNA volume in the transposon-based method was lesser than that in the ligation-based method, the base coverage was affected by the adaptor addition PCR amplification bias. The transposon-based method significantly increased the percentage of sequence reads on chromosomes, except for chromosome 16. Other genomic regions were amplified using adaptor addition PCR. Furthermore, the PCR amplification bias causes strand and variant frequency biases. In general, a minimum of 30× base coverage with balanced reads may be sufficient for germline testing if only detection of heterozygous or homozygous variants of two alleles are considered [25,26]. Approximately 200,000 sequence reads were required to obtain a minimum of 30× base coverage in the transposon-based method, which is 15-fold higher than the ligation-based method. The uniformity of base coverage is an important factor that influences the efficiency of targeted gene sequencing using MPS. Thus, the ligation-based method is preferable for obtaining uniform sequence coverage with no strand bias.

The major limitation of LR-PCR is the requirement of high-quality genomic DNA. Thus, degraded DNA samples, such as formalin-fixed paraffin-embedded DNA and cell-free DNA, cannot be analyzed [27]. In addition, LR-PCR may not be suitable for enriching the target regions if the protein-coding regions of the gene with large intronic regions are analyzed. Although the single nucleotide variants and small indels can be detectable by LR-PCR, the detection of large indels such as copy number variants (CNVs) may be difficult. However, it has been reported the CNVs were very rare in the FMF patients [28]. In this study, we used the Ion Torrent PGM semiconductor sequencer and its compatible

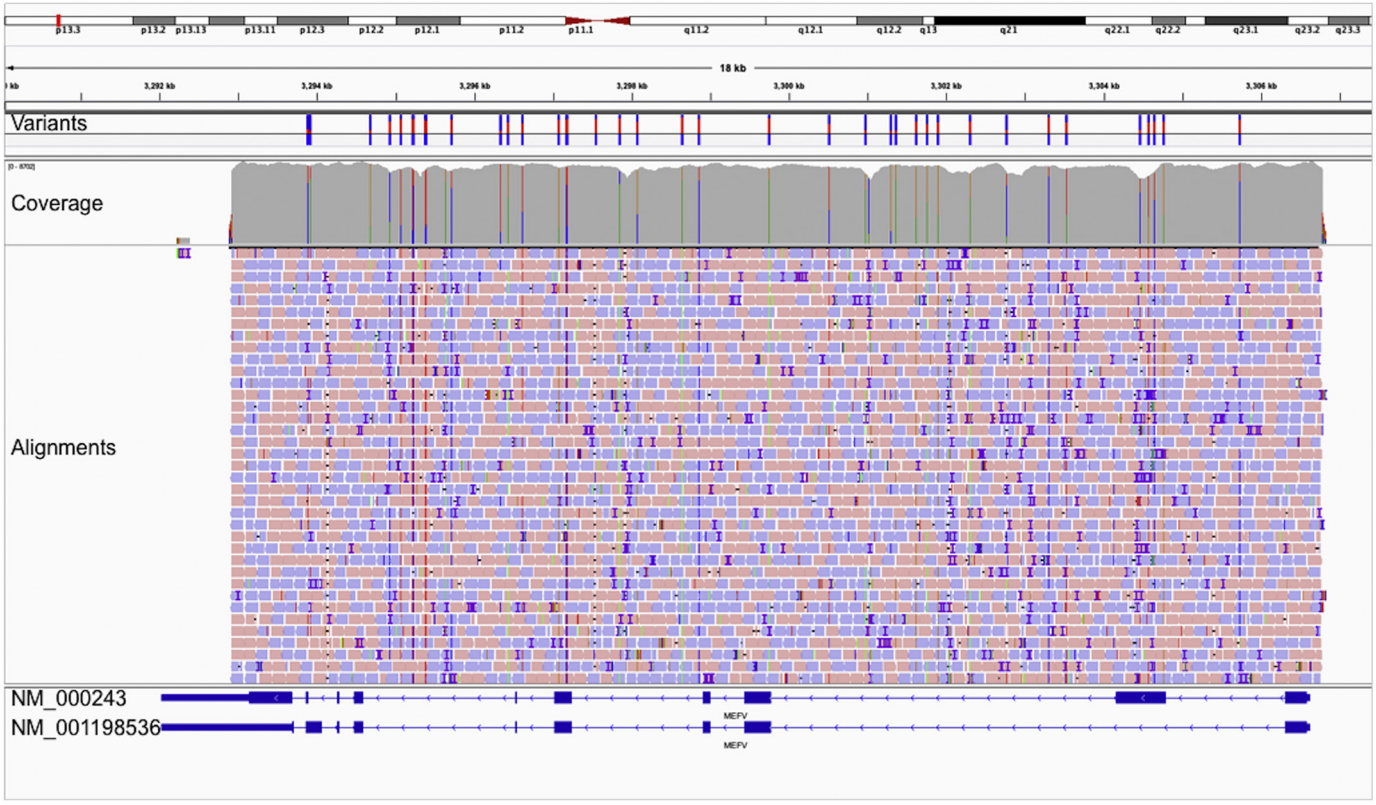

Fig. 7. The MPS result of ligation-based method and detected variants in *MEFV* gene.

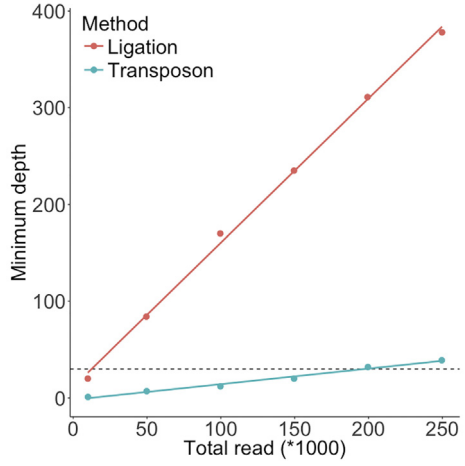

Fig. 8. Estimation of minimum base coverage in various total reads. The dotted line indicates minimum base coverage of 30.

library preparation kits. Further studies with other sequencing technologies and library preparation reagents would be required.

In conclusion, we established the analytical pipeline of targeted gene sequencing using LR-PCR and MPS. These data can enable the optimization of target gene sequencing using LR-PCR and MPS in the clinical laboratory.

Acknowledgments

This study was supported by Kurozumi Medical Foundation, Japan (Project Number J18KF00088).

Disclosure

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2019.06.001.

References

[1] M.L. Landsverk, L.-J.C. Wong, Clinical molecular diagnostic techniques: a brief review, in: L.-J.C. Wong (Ed.), Next Generation Sequencing: Translation to Clinical Diagnostics, Springer, New York, 2013, pp. 19–36.
[2] S. Cheng, C. Fockler, W.M. Barnes, R. Higuchi, Effective amplification of long targets from cloned inserts and human genomic DNA, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 5695–5699.
[3] H. Cui, Methods of gene enrichment and massively parallel sequencing technologies, in: L.-J.C. Wong (Ed.), Next Generation Sequencing: Translation to Clinical Diagnostics, Springer, New York, 2013, pp. 39–58.
[4] I. Hernan, E. Borràs, M. de Sousa Dias, M.J. Gamundi, B. Mañé, G. Llort, J.A. Agúndez, M. Blanca, M. Carballo, Detection of genomic variations in BRCA1 and BRCA2 genes by long-range PCR and next-generation sequencing, J. Mol. Diagn. 14 (2012) 286–293.
[5] H. Ozcelik, X. Shi, M.C. Chang, E. Tram, M. Vlasschaert, N. Di Nicola, A. Kiselova, D. Yee, A. Goldman, M. Dowar, B. Sukhu, R. Kandel, K. Siminovitch, Long-range PCR and next-generation sequencing of BRCA1 and BRCA2 in breast cancer, J. Mol. Diagn. 14 (2012) 467–475.
[6] A.Y. Tan, A. Michaeel, G. Liu, O. Elemento, J. Blumenfeld, S. Donahue, T. Parker, D. Levine, H. Rennert, Molecular diagnosis of autosomal dominant polycystic kidney disease using next-generation sequencing, J. Mol. Diagn. 16 (2014) 216–228.
[7] Y.C. Tan, A. Michaeel, J. Blumenfeld, S. Donahue, T. Parker, D. Levine, H. Rennert, A novel long-range PCR sequencing method for genetic analysis of the entire PKD1 gene, J. Mol. Diagn. 14 (2012) 305–313.
[8] M. Clendenning, H. Hampel, J. LaJeunesse, A. Lindblom, J. Lockman, M. Nilbert, L. Senter, K. Sotamaa, A. de la Chapelle, Long-range PCR facilitates the identification of PMS2-specific mutations, Hum. Mutat. 27 (2006) 490–495.
[9] Y. Ozaki, S. Suzuki, A. Shigenari, Y. Okudaira, E. Kikkawa, A. Oka, M. Ota, S. Mitsunaga, J.K. Kulski, H. Inoko, T. Shiina, HLA-DRB1, -DRB3, -DRB4 and -DRB5 genotyping at a super-high resolution level by long range PCR and high-throughput

sequencing, Tissue Antigens 83 (2014) 10–16.
[10] M. Alghamdi, Familial Mediterranean fever, review of the literature, Clin. Rheumatol. 36 (2017) 1707–1713.
[11] French FMF Consortium, A candidate gene for familial Mediterranean fever, Nat. Genet. 17 (1997) 25–31.
[12] The International FMF Consortium, Ancient missense mutations in a new member of the RoRet gene family are likely to cause familial Mediterranean fever, Cell. 90 (1997) 797–807.
[13] N.A. O'Leary, M.W. Wright, J.R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C.M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V.S. Joardar, V.K. Kodali, W. Li, D. Maglott, P. Masterson, K.M. McGarvey, M.R. Murphy, K. O'Neill, S. Pujar, S.H. Rangwala, D. Rausch, L.D. Riddick, C. Schoch, A. Shkeda, S.S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R.E. Tully, A.R. Vatsan, C. Wallin, D. Webb, W. Wu, M.J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T.D. Murphy, K.D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation, Nucleic Acids Res. 44 (2016) D733–D745.
[14] S. Kim, M. Ikusaka, G. Mikasa, A. Basugi, Y. Ohira, S. Nishizawa, S. Itoga, M. Sunaga, F. Nomura, Clinical study of 7 cases of familial Mediterranean fever with MEFV gene mutation, Intern. Med. 46 (2007) 221–225.
[15] D. Kishida, M. Yazaki, A. Nakamura, F. Nomura, T. Kondo, T. Uehara, M. Ikusaka, A. Ohya, N. Watanabe, R. Endo, S. Kawaai, Y. Shimojima, Y. Sekijima, One novel and two uncommon MEFV mutations in Japanese patients with familial Mediterranean fever: a clinicogenetic study, Rheumatol. Int. 38 (2018) 105–110.
[16] N. Blin, D.W. Stafford, A general method for isolation of high molecular weight DNA from eukaryotes, Nucleic Acids Res. 3 (1976) 2303–2308.
[17] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3–new capabilities and interfaces, Nucleic Acids Res. 40 (2012) e115.
[18] T. Ishige, S. Itoga, K. Matsushita, F. Nomura, Locked nucleic acid probe enhances Sanger sequencing sensitivity and improves diagnostic accuracy of high-resolution melting-based KRAS mutational analysis, Clin. Chim. Acta 457 (2016) 75–80.
[19] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.
[20] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics. 25 (2009) 2078–2079.
[21] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018 URL https://www.R-project.org/.
[22] C.M. Farrell, N.A. O'Leary, R.A. Harte, J.E. Loveland, L.G. Wilming, C. Wallin, M. Diekhans, D. Barrell, S.M. Searle, B. Aken, S.M. Hiatt, A. Frankish, M.M. Suner, B. Rajput, C.A. Steward, G.R. Brown, R. Bennett, M. Murphy, W. Wu, M.P. Kay, J. Hart, J. Rajan, J. Weber, C. Snow, L.D. Riddick, T. Hunt, D. Webb, M. Thomas, P. Tamez, S.H. Rangwala, K.M. McGarvey, S. Pujar, A. Shkeda, J.M. Mudge, J.M. Gonzalez, J.G. Gilbert, S.J. Trevanion, R. Baertsch, J.L. Harrow, T. Hubbard, J.M. Ostell, D. Haussler, K.D. Pruitt, Current status and new features of the Consensus Coding Sequence database, Nucleic Acids Res. 42 (2014) D865–D872.
[23] H. Jia, Y. Guo, W. Zhao, K. Wang, Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer, Sci. Rep. 4 (2014) 5737.
[24] M.A. Quail, M. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow, Y. Gu, A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers, BMC Genomics 13 (2012) 341.
[25] J.S. Parla, I. Iossifov, I. Grabill, M.S. Spector, M. Kramer, W.R. McCombie, A comparative analysis of exome capture, Genome Biol. 12 (2011) R97.
[26] Y. Xu Asan, H. Jiang, C. Tyler-Smith, Y. Xue, T. Jiang, J. Wang, M. Wu, X. Liu, G. Tian, J. Wang, J. Wang, H. Yang, X. Zhang, Comprehensive comparison of three commercial human whole-exome capture platforms, Genome Biol. 12 (2011) R95.
[27] E. Zonta, P. Nizard, V. Taly, Assessment of DNA integrity, applications for cancer research, Adv. Clin. Chem. 70 (2015) 197–246.
[28] M.E. van Gijn, S. Soler, C. de la Chapelle, M. Mulder, C. Ritorre, M. Kriek, L. Philibert, M. van der Wielen, J. Frenkel, S. Grandemange, E. Bakker, J.K. Ploos van Amstel, I. Touitou, Search for copy number alterations in the MEFV gene using multiplex ligation probe amplification, experience from three diagnostic centres, Eur. J. Hum. Genet. 16 (2008) 1404–1406.