RESEARCH ARTICLE

A low cost and input tailing method of quality control on multiple annealing, and looping-based amplification cycles-based whole-genome amplification products

Changyue Chen¹ | Jing Li¹ | JueFeng Wan^{2,3} | Yuan Lu¹ | Zhen Zhang^{2,3} | ZengHui Xu^{4,5}

¹Department of Medical Research, Shanghai MajorMed Diagnostics Company, Shanghai, China

²Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai, China

³Department of Oncology, Fudan University Shanghai Medical College, Shanghai, China

⁴Laboratory of Gene and Viral Therapy, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University of Chinese PLA, Shanghai, China

⁵ShangHai Cell Therapy Group Company, Shanghai, China

Correspondence

ChangYue Chen, Department of Medical Research, Shanghai MajorMed Diagnostics Company, Shanghai, China.

Email: changyue.chen@majorbio.com

Zhen Zhang, Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai, China.

Email: zhenzhang6@gmail.com and

ZengHui Xu, Laboratory of Gene and Viral Therapy, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University of Chinese PLA, Shanghai, China.

Email: zenghuixu@163.com

Funding information

The present study was funded by National Natural Science Foundation of China (grant no. 81372432).

Abstract

Background: Single-cell whole-genome sequencing provides novel insights into the nature of genetic heterogeneity in normal and diseased cells. However, amplification of formalin-fixed tissues with low cell numbers is still problematic and multiple annealing, and looping-based amplification cycles (MALBAC) is a commonly used whole-genome amplification (WGA) method with low cell numbers.

Methods: We developed a low-input tailing method to evaluate the MALBAC-based WGA from sub-nanogram or less quantities of input DNA. The tailing method uses 2100 BioAnalyzer to evaluate the size distribution of MALBAC products, and comparing the tailing with 10380 bp.

Results: Compared with a 22 loci qPCR panel, the tailing method provided a similar WGA evaluation efficiency in 13 samples on one set of study, with lower input, cheaper cost, shorter manual time, and a clear filtering cut off. Later, we demonstrated a strong correlation between tailing size and coverage breadth in another 29 samples on two sets of assays. As a result, the tailing method showed that it could predict whether a sequence breadth achieved 70% or not with 100% accuracy on these three sets of assays. Although further studies are needed, this tailing method is expected to be used as an excellent tool to select high-quality WGA products before library construction.

Conclusions: Our tailing method can provide a new WGA quality test to evaluate the WGA efficiency with 100% accuracy (42/42). Compared with qPCR panel, our tailing method needs lower input, cheaper cost, shorter manual time, a clear filtering cut off, and extendable high throughput as well as the same sensitivity.

1 | INTRODUCTION

Whole-genome DNA sequencing of single cells (scWGS) develops rapidly in recent years with the introduction of single-cell amplification methods.¹⁻⁶ There are three commonly used techniques: multiple displacement amplification (MDA), multiple annealing, and looping-based amplification cycles (MALBAC) and degenerate oligonucleotide primed PCR (DOP-PCR).⁷⁻⁹ However, amplification of

formalin-fixed tissues with low cell numbers is problematic. Among these methods, MDA technique is most suitable for fresh cells or tissues, however it demonstrates low efficiency when used with fragmented and/or cross-linked DNA.^{10,11} DOP-PCR can amplify ~6% of the genome and is suitable for the detection of copy number variation but not for the detection of single nucleotide variant.^{7,12-14} As comparison, MALBAC is superior in uniformity, coverage breadth, and reproducibility.¹⁵⁻¹⁷

Amplification from minute amounts of DNA from microdissected samples usually has a problem with low genome coverage breadth.¹⁸ The low coverage breadth not only limits tumor heterogeneity studies, but also increases the probability of high rates of sequencing artifacts.¹⁹ In these previous studies, the multiple loci qPCR panel has been applied for the whole genome amplification (WGA) quality control.^{8,20,21} In Navin's laboratory,⁸ a series of qPCRs were performed using 22 primer pairs independently.⁸ WGA reactions that showed 20/22 positive qPCR reactions were selected for subsequent library construction and next generation sequencing. However, qPCR panel cannot be widely used in the MALBAC-based WGA quality control due to its high input and the low yield. This prompted us to investigate

a new quality control (QC) method to evaluate WGA efficiency with low DNA input.

Herein, we report a low cost and input, easy-to-use method for MALBAC-based WGA QC method based on the size distribution of MALBAC products. The tailing size represents the maximal size of DNA fragment in capillary electrophoresis. When comparing tailing size with the peak at 10380 bp, tailings are divided into long tailing (>10380 bp peak) and short tailing (<10380 bp peak). Firstly, we compared tailing method to a 22 loci qPCR panel, we found that they gave us similar WGA evaluation efficiency in 13 samples. With promising findings from the first step, we further investigated another 29 samples, and in result we demonstrated a strong correlation between the cut-off length of tailing method

FIGURE 1 Overview of single-cell capture and sequencing and whole-genome amplification quality control. (A) Schematic overview of single cell capture and sequencing. Cells fixed on the slide were captured and collected onto the cover of tubes and multiple annealing, and looping-based amplification cycles (MALBAC) amplification was performed. The red arrows in the "Quality Control" represent tailing sizes. When comparing tailing size with the peak at 10380 bp, tailings are divided into long tailing (>10380 bp peak) and short tailing (<10380 bp peak). (B) Illustration of tailing method with electropherograms of 2100 Bioanalyzer. Samples with long tailing (black arrow) diagram correspond to >70% coverage breadth. Samples with short tailing (black arrow) diagram correspond to <70% coverage breadth. The Negative control (without temple) with short tailing represents the feature of continuous primer dimer peaks and no large amplicons over 10380 bp

(10380 bp) and 1× coverage breadth of 70% at 50× average depth (Figure 1B).

2 | RESULTS AND DISCUSSION

2.1 | Patients and tumor characteristics and CTC detection

Between September 2014 and March 2015, 31 patients were enrolled. We collected a 7.5 mL blood sample from each patient and sent for circulating tumor cell (CTC) detection and target sequencing. The number of CTCs varied from 1 to 42 in 31 patients. All CTCs from one patient were collected in one tube for 27 patients (Table 1, #Patient 1 to #Patient 27); for the other four patients, each CTC was collected in one tube. Sample 28,

29, 36, 37 are from #patient 28, Samples 30, 31, 33, 34, 42 are from #patient 29, Sample 32, 35, 41 are from #patient 30, Sample 38, 39, 40 are from #patient 31. Patients, corresponding samples, captured CTCs, tumor characteristics and target panels are listed in Table 1.

2.2 | Current WGA quality control using a 22 loci qPCR panel

As mentioned above, WGA quality control was necessary for MALBAC-based amplification. Current WGA quality control assay used several pairs of primers to target loci on different chromosomes for qPCR. WGA reactions that showed positive qPCR reactions were selected for subsequent library construction and next generation sequencing. Navin's lab consider looser filtering criteria

TABLE 1 Characterization of patients

Patient No	Sample No	Captured CTCs	Clinical stage	Targeted re-sequencing assays
#Patient 1	Sample 1	5	III	36 panel
#Patient 2	Sample 2	6	III	36 panel
#Patient 3	Sample 3	3	III	36 panel
#Patient 4	Sample 4	2	III	36 panel
#Patient 5	Sample 5	13	III	36 panel
#Patient 6	Sample 6	9	III	36 panel
#Patient 7	Sample 7	15	III	36 panel
#Patient 8	Sample 8	4	III	36 panel
#Patient 9	Sample 9	3	III	36 panel
#Patient 10	Sample 10	1	III	36 panel
#Patient 11	Sample 11	4	III	36 panel
#Patient 12	Sample 12	1	III	36 panel
#Patient 13	Sample 13	2	III	36 panel
#Patient 14	Sample 14	3	III	119 panel
#Patient 15	Sample 15	4	III	119 panel
#Patient 16	Sample 16	8	III	119 panel
#Patient 17	Sample 17	10	III	119 panel
#Patient 18	Sample 18	27	III	119 panel
#Patient 19	Sample 19	19	III	119 panel
#Patient 20	Sample 20	2	III	119 panel
#Patient 21	Sample 21	1	III	119 panel
#Patient 22	Sample 22	14	III	119 panel
#Patient 23	Sample 23	42	III	119 panel
#Patient 24	Sample 24	15	III	119 panel
#Patient 25	Sample 25	3	III	119 panel
#Patient 26	Sample 26	10	III	119 panel
#Patient 27	Sample 27	15	III	119 panel
#Patient 28	Sample 28/29/36/37	1	IV	WES
#Patient 29	Sample 30/31/33/34/42	1	IV	WES
#Patient 30	Sample 32/35/41	1	IV	WES
#Patient 31	Sample 38/39/40	1	III	WES

TABLE 2 qPCR results of 22 pairs of primers, tailing size, and corresponding coverage breadth for each sample

Criteria	Positive chro number	Name	50×, bases covered at ≥ 1× (%)	Tailing size (bp)	Accuracy
NTC	0/22				
Short tailing (<10380 bp)	8/22	Sample 1	19.41	5952	100% (42/42)
	9/22	Sample 2	23.50	5183	
	8/22	Sample 3	25.73	5760	
	7/22	Sample 4	28.79	2446	
	10/22	Sample 5	46.55	8701	
	19/22	Sample 6	66.32	8766	
	/	Sample 14	16.93	<10380	
	/	Sample 15	22.99	<10380	
	/	Sample 16	16.62	<10380	
	/	Sample 17	15.31	<10380	
	/	Sample 28	16.73	<10380	
	/	Sample 29	16.92	<10380	
	/	Sample 30	44.16	<10380	
	/	Sample 31	62.77	<10380	
	/	Sample 32	46.96	<10380	
	/	Sample 33	9.66	<10380	
	/	Sample 34	14.32	<10380	
	/	Sample 35	25.37	<10380	
Long tailing (>10380 bp)	22/22	Sample 7	75.45	20911	
	21/22	Sample 8	79.37	25103	
	22/22	Sample 9	81.11	14943	
	22/22	Sample 10	83.52	26862	
	21/22	Sample 11	87.37	18155	
	22/22	Sample 12	87.52	18217	
	22/22	Sample 13	88.66	12806	
	/	Sample 18	88.73	>10380	
	/	Sample 19	92.79	>10380	
	/	Sample 20	86.29	>10380	
	/	Sample 21	90.24	>10380	
	/	Sample 22	85.33	>10380	
	/	Sample 23	83.77	>10380	
	/	Sample 24	87.99	>10380	
	/	Sample 25	91.16	>10380	
	/	Sample 26	89.47	>10380	
	/	Sample 27	81.53	>10380	
	/	Sample 36	81.51	>10380	
	/	Sample 37	77.31	>10380	
	/	Sample 38	78.25	>10380	
	/	Sample 39	78.35	>10380	
	/	Sample 40	83.94	>10380	
	/	Sample 41	87.02	>10380	
	/	Sample 42	72.94	>10380	

NTC: negative control, using the same SYBR GREEN qPCR system without template.

"/" represent no data.

FIGURE 2 Electropherograms of multiple annealing, and looping-based amplification cycles (MALBAC) products. (A) Short tailing group Short tailing to (a) 5952 bp (black arrow, Sample 1), (b) 5183 bp (black arrow, Sample 2), (c) 8701 bp (black arrow, Sample 5), (d) 8766 bp (black red arrow, Sample 6), corresponding 1× coverage breadth of (a) 19.41%, (b) 23.50%, (c) 46.55%, (d) 66.32%. (B) Long tailing group Long tailing to (e) 20911 bp (black arrow, Sample 7), (f) 25103 bp (black arrow, Sample 8), (g) 14943 bp (black arrow, Sample 9), (h) 18217 bp (black arrow, Sample 12), corresponding 1× coverage breadth of (e) 75.45%, (b) 79.37%, (c) 81.11%, (d) 87.52%

(20/22 amplicons) for diminished signal due to hemizygous deletions during qPCR process.^{22,23}

Due to different coverage depth of each sample, all the sequencing data were normalized to a mean depth of 50×. Of 13 samples admitted to the MALBAC-based amplification, seven samples passed the qPCR filtering criteria and the minimum of 1× coverage breadth was 75.45% (Table 2). In contrast, the other six samples did not pass the qPCR filtering criteria (<20 amplicons) and the maximum of 1× coverage breadth was 66.32%. Our results showed that filtering criteria (20/22 amplicons) of qPCR quality panel directly correlated with 1× coverage breadth of 70%, which were concordant with Navin's report.⁸ However, we also noticed that once the number of positive amplicons

reached the filtering criteria (20/22 amplicons), the correlation between "the number of positive amplicons" and "Coverage breadth" was not linear. For example, while the 1× coverage breadth of sample 7 with 22/22 amplicons was 75.45%, the 1× coverage breadth of sample 11 with 21/22 amplicons was 87.37% (Table 2).

2.3 | WGA efficiency evaluated by tailing method correlates with qPCR panel and coverage breadth

In the beginning of our study, we tried to use 2100 BioAnalyzer to evaluate the size distribution of MALBAC products.²⁴ We analyzed 13 samples at first and found that there was a correlation between

the cut-off length of tailing method (10380 bp) and 1× coverage breadth of 70% (Table 2). Comparing the tailing sizes of MALBAC products with the 10380 bp marker, we divided these WGA products into two groups, short tailing group (<10380 bp) and long tailing group (≥10380 bp).

Figure 2A-D show the electropherograms of MALBAC products in the short tailing group. When the short tailing reached to approximately 8766 bp (Figure 2D), the 1× coverage breadth reached to 66.32%. The short tailings of MALBAC products of other three samples were separately reached to 5952 bp, 5183 bp, 8701 bp, while 1× coverage breadths of these samples were 19.41%, 23.50%, 46.55% correspondingly. 1× coverage breadths in the short tailing group were all below 70%.

Figure 2E-H show the electropherograms of MALBAC products in the long tailing group. Combined with the 1× coverage breadth data (Table 2), Figure 2E demonstrated that when the long tailing reached to 20911 bp, over the 10380 bp marker, the 1× coverage breadth reached to 75.45%. The long tailings of MALBAC products of the rest three samples reached to 25103 bp, 14943 bp, 18217 bp separately, while 1× coverage breadths of these samples were 79.37%, 81.11%, 87.52% correspondingly. 1× coverage breadths in the long tailing group were all above 70%.

Although the correlation between "Coverage breadths" and "Tailing size" is not linear, the criteria of tailing method (10380 bp) performed strong correlation with 1× coverage breadth of 70% at 50× mean depth. The accuracy of the tailing method for the prediction of 1× coverage breadth of 70% was 100% (13/13). In the first 36 gene study ($n = 13$), the median number of 1× coverage breadth was 83.52 in samples with long tailing vs 27.26 with short tailing (Mann-Whitney $P = 0.0012$; Figure 3A).

2.4 | Correlation between the cut-off length of tailing method and 1× coverage breadth

To further study the quality control criteria of tailing method, another 29 CTC samples were grouped into two, 14 of them were amplified

and sequenced with 119-gene panel (Table S2) and the rest 15 were done with whole-exome sequencing (WES). Followingly, all of the MALBAC products were purified and tested by tailing method. By looking into raw sequence data on average of 50× depth randomly, 10 of 14 samples with 119 gene panel and 7 of 15 with WES reached the WGA quality control criteria (tailings > 10380 bp, Table 2). The 1× coverage breadth of this long tailing group covered over 70% of genome. In contrast, the other four samples with 119 gene panel and eight with WES did not pass the standard (tailings <10380 bp, Table 2). Correspondingly, the 1× coverage breadth of this short tailing group covered <70%. The accuracy of the tailing method were 100% in 36-gene panel set (13/13), 119-gene panel set (14/14), and WES set (15/15).

When combining all data from assays using 36-gene panel, 119-gene panel and WES together, the median number of 1× depth coverage was 84.64% within long tailing group and 23.25% within short tailing group (Mann-Whitney $P < 0.0001$; Figure 3B). Also the minimum of 1× coverage breadth of the long tailing group was 72.94%, and maximum in the short tailing group was 66.32%. The overall accuracy of the tailing method was 100% (42/42) among all sets. From these results, it was shown that cut-off length of tailing method (10380 bp) in 50× mean depth was strong correlated with threshold value of 1× coverage breadth (70%).

Compared to the qPCR panel, the tailing method bears four main advantages. Firstly, the input amount of tailing method is ~1 ng, nearly 1% less of qPCR panel input requirement. It brings more convenience to quality control with larger scale of sample amount, especially with purified MALBAC products <100 ng. Secondly, there is a clear and universal cut-off length in tailing method (10380 bp) for better quality control. By contrast, the cutoff value of qPCR can be controlled by various factors, such as primer dimers when $C_t = 34-36$ in some chromosomes. Thirdly, tailing method takes shorter manual time. The tailing method needs a few minutes while qPCR panel needs more than 2 hours. Fourthly, the material cost of tailing method is less than qPCR as well. QC metric used in tailing method costs 6 dollars per sample while qPCR metric costs 48

FIGURE 3 Coverage breadth associated with size distribution of multiple annealing, and looping-based amplification cycles (MALBAC) products. (A) 1× coverage breadths of 36 gene panel assay of MALBAC products with long tailing ($n = 7$) or with short tailing ($n = 6$) (median 83.52 vs 27.26, Mann-Whitney $P = 0.0012$). (B) 1× coverage breadths of entire set of assays of MALBAC products with long tailing ($n = 24$) or with short tailing ($n = 18$) (median 84.64 vs 23.25, Mann-Whitney $P < 0.0001$)

dollars per sample. In addition, this tailing method may be further applied to high throughput screening instead of the current laborious QC test when conducted with other instruments, such as Agilent TapeStation 4000 and Advanced Analytical Fragment Analyzer. To sum up, tailing method can achieve the same WGA efficiency as qPCR panel does, but with lower input, cheaper cost, shorter manual time, a clear filtering cut off and extendable high throughput as well as the same sensitivity, which makes the tailing method a good WGA quality control criterion for MALBAC products.

3 | CONCLUSION

The tailing method can be used as a novel WGA quality control criterion to evaluate the WGA efficiency with rare cell samples. Compared to qPCR panel, the tailing method can reach the sample level of sensitivity with lower input, cheaper cost, shorter manual time, a clear filtering cut off and extendable high throughput. In our study, it was demonstrated that a tail of larger MALBAC amplicons (>10380 bp) is usually accompanied with better coverage although the relationship between is not linear. With further studies, the tailing method can be expected to be a great criteria to assess the WGA efficiency.

4 | EXPERIMENTAL PROCEDURE

4.1 | Patients, blood specimen collection and processing

A total of 31 patients with Locally Advanced Rectal Cancer (LARC, cT3-T4 and/or cN+) were enlisted from September 2014 to March 2015. They were treated with neoadjuvant CRT at Fudan University Shanghai Cancer Center. All patients signed the informed consent before chemoradiotherapy, and all clinical procedures were in accordance with the Helsinki declaration. The ethics committee of Fudan University Shanghai Cancer Center approved the study protocols and experimental procedures, including any relevant details. All experiments were performed in accordance with relevant guidelines and regulations. About 7.5 mL blood samples were obtained from all these 31 patients and loaded in evacuated ACD anticoagulant tubes for further CTC detection.

4.2 | Subtraction enrichment of CTCs and Identification of aneuploidy CTCs

The enrichment and identification of CTCs were performed according to CTCseq™ kit instruction (Majorbio, Shanghai, China). Samples were fixed on the slide, followed by counting and photographing. Each suspicious tumor cell coordinates were recorded to facilitate subsequent target cell identification. The identification principle of CTCs is that (a) they were negative for CD45 and (b) they were either positive in pan-cytokeratin staining or with heteroploid of chromosome 8. Afterwards, these slides were kept frozen at -20°C.

4.3 | Laser capture microdissection

The samples were loaded onto the stage of a Zeiss PALM MicroBeam (Zeiss, Jena, Germany) under a ×40 objective. After microdissection with a 355-nm laser beam, target cells were collected onto the cover of AdhesiveCap 200 opaque (Zeiss). A 5 µL lysis buffer (Malbac, Yikon Genomics, Nantong, China) was then put on the cover of AdhesiveCap 200 opaque. Samples were stored at -20°C.

4.4 | Single-cell WGA with MALBAC

Whole-genome amplification of lysed cells was performed using the MALBAC method⁹ following the standard protocol provided by the commercial MALBAC amplification kit (Yikon Genomics). In brief, these cells were centrifuged at 9230 g for 5 minutes and then collected and transferred to a new 200 µL PCR tube. These cells were incubated at 50°C for 50 minutes and then at 80°C for 10 minutes. Then 30 µL of freshly prepared preamplification mix was added to each tube and these cell mixes were incubated at 94°C for 3 minutes. After that, DNA was pre-amplified by a protocol of eight cycles of gradient temperature procedure (20°C for 40 seconds, 30°C for 40 seconds, 40°C for 30 seconds, 50°C for 30 seconds, 60°C for 30 seconds, 70°C for 4 minutes, 95°C for 20 seconds, and 58°C for 10 seconds) and kept on ice immediately. Afterward, 30 µL of the amplification reaction mix was added to each tube and samples were continuously amplified by a PCR procedure (94°C for 30 seconds, followed by 17 cycles of 94°C for 20 seconds and 58°C for 30 seconds and 72°C for 3 minutes).

4.5 | WGA quality control by qPCR panel

Whole-genome amplification quality control by qPCR panel was performed according to the protocol provided by Navin's laboratory⁸. About 22 pairs of primers (Sangon, Shanghai, China) were used to target 22 loci on different chromosomes. A total of 5 ng of each DNA sample was added to the AceQ qPCR SYBR GreenMaster Mix (Q111-02, Vazyme, Nanjing, China) and placed into 96-well plates. Then these samples were amplified in the ABI 7500 system (Thermo Fisher Scientific, Waltham, MA, USA) with an amplification procedure (95°C for 3 minutes, followed by 45 cycles of 95°C for 20 seconds and 60°C for 30 seconds). Afterward, the single-cell WGA product with positive qPCR results (>20 amplicons) was selected for subsequent library construction and next generation sequencing.

4.6 | WGA quality control on Agilent 2100

Multiple annealing, and looping-based amplification cycles products were firstly purified with 1.4× Ampure XP beads (Beckman, Brea, CA, USA) and then quantified with the QuantiFluor® dsDNA Dye and the Quantus™ Fluorometer (Promega, Madison, WI, USA). After that, each MALBAC product was diluted into 1 ng/µL and 1 µL of them were taken for subsequent analysis. A BioAnalyzer High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA) was used to visualize and identify the size range and the tailing size of MALBAC products.

More specifically, samples were injected into the separation channel and their components were electrophoretically separated. The Electropherogram tab of Agilent 2100 Bioanalyzer displayed a data plot of size over migration time vs fluorescence intensity. All the Figures were generated using Prism 5 (GraphPad Software, LaJolla, CA).

4.7 | Target gene sequencing and whole-exome sequencing

Sequencing libraries preparation with 36-gene multiplex PCR panel (MygeneSeq Technology, Shanghai, China, Table S1) was performed according to the manufacturer's instructions. The 36 genes, corresponding to 105 kb, were the high-frequency mutation in colorectal cancer patients according to the COSMIC database. For specific sequencing preparation, 100 ng of each MALBAC products were amplified by PCR and purified, and then used for constructing a barcoded DNA library (5144-08, Bioo Scientific, Austin, TX, USA). Sequencing libraries preparation with 119-gene panel (Integrated DNA Technologies, Skokie, IL, USA, Table S2) and SureSelect XT Human All Exon V5 (Agilent Technologies) were performed according to the manufacturer's instructions with 200 ng input MALBAC products. The size distribution of constructed libraries was measured by agilent 2100, and main peaks of those libraries ranged from 300 to 400 bp. Final concentrations were measured prior to sequencing by the Qubit 2.0 fluorometer (Life Technologies). The 36-genes panel libraries and the WES libraries were sequenced with paired-end flowcell lanes on the HiSeq4000 system (Illumina, San Diego, CA, USA) for 150 cycles. The 119-gene panel libraries were sequenced with paired-end flowcell lanes on the CN500 system (Illumina) for 75 cycles. All these sequence data were processed using CASAVA 1.8.1 pipeline (Illumina) converting BCL basecall files to fastq files.

4.8 | Software analysis

For the raw reads, cutadapt 1.12²⁵ was used for trimming the MALBAC primers and illumina adaptors. Then these trimmed sequencing reads were aligned to the human genome (hg19) by the Burrows-Wheeler Aligner (BWA- 0.7.12-r1039).²⁶ The mapping files in SAM format were converted to BAM format and sorted by SAMtools-1.2.²⁷ The Genome Analysis Toolkit (GATK v3.3-0-g37228af, <https://www.broadinstitute.org/gatk/>) was used for realigning the sorted BAM file.²⁸ Picard-tools-1.119 (<https://broadinstitute.github.io/picard/>) was used for eliminating PCR duplicates. After all these, data reads were re-calibrated and indel-realigned by GATK (<https://www.broadinstitute.org/gatk/>) before variation detection. Both base-level sequencing depth and coverage breadth were enumerated by the Depth Of Coverage module by bedtools. 1× Coverage breadth is defined as the percentage of nucleotide sites in the single-cell data with ≥1× coverage depth. The on target read ratios of all samples varies from 60% to 80%.

CONFLICT OF INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

CYC designed the experiments. CYC, JL, and JFW performed the experiments. YL analyzed the data. CYC, ZHX, and ZZ contributed to the manuscript writing.

ORCID

Changyue Chen <http://orcid.org/0000-0002-3534-6701>

REFERENCES

- Sherman MA, Barton AR, Lodato MA, et al. PaSD-qc: quality control for single cell whole-genome sequencing data using power spectral density estimation. *Nucleic Acids Res.* 2018;46(4):e20.
- Gawad C, Koh W, QuakeSR. Single-cell genome sequencing: current state of the science. *Nature Rev Genet.* 2016;17(3):175-188.
- Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-cell whole-genome amplification and sequencing: methodology and applications. *Ann Rev Genomics Hum Genet.* 2015;16:79-102.
- Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. *Nature.* 2011;472(7341):90-94.
- Ning L, Liu G, Li G, Hou Y, Tong Y, He J. Current challenges in the bioinformatics of single cell genomics. *Front Oncol.* 2014;4:7.
- Joesse SA, Gorges TM, Pantel K. Biology, detection, and clinical implications of circulating tumor cells. *EMBO Mol Med.* 2015;7(1):1-11.
- Hou Y, Wu K, Shi X, et al. Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *GigaScience.* 2015;4:37.
- Leung ML, Wang Y, Waters J, Navin NE. SNES: single nucleus exome sequencing. *Genome Biol.* 2015;16:55.
- Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science.* 2012;338(6114):1622-1626.
- Wang G, Maher E, Brennan C, et al. DNA amplification method tolerant to sample degradation. *Genome Res.* 2004;14(11):2357-2366.
- Babayian A, Alawi M, Gormley M, et al. Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells. *Oncotarget.* 2017;8(34):56066-56080.
- Arneson N, Hughes S, Houlston R, Done S. Whole-genome amplification by degenerate oligonucleotide primed PCR (DOP-PCR). *CSH Protoc.* 2008;2008:pdb.prot4919.
- Daigo Y, Chin SF, Gorringer KL, et al. Degenerate oligonucleotide primed-polymerase chain reaction-based array comparative genomic hybridization for extensive amplicon profiling of breast cancers: a new approach for the molecular analysis of paraffin-embedded cancer tissue. *Am J Pathol.* 2001;158(5):1623-1631.
- Cheung VG, Nelson SF. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc Natl Acad Sci USA.* 1996;93(25):14676-14679.
- Chen M, Song P, Zou D, et al. Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. *PLoS One.* 2014;9(12):e114520.
- Li N, Wang L, Wang H, et al. The performance of whole genome amplification methods and next-generation sequencing for pre-implantation genetic diagnosis of chromosomal abnormalities. *J Genet Genomics.* 2015;42(4):151-159.
- Ning L, Li Z, Wang G, et al. Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons. *Sci Rep.* 2015;5:11415.

18. Stoecklein NH, Erbersdobler A, Schmidt-Kittler O, et al. SCOMP is superior to degenerated oligonucleotide primed-polymerase chain reaction for global amplification of minute amounts of DNA from microdissected archival tissue samples. *Am J Pathol.* 2002;161(1):43-51.
19. Wong SQ, Li J, Tan AY, et al. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med Genomics.* 2014;7:23.
20. Wang Y, Waters J, Leung ML, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature.* 2014;512(7513):155-160.
21. Sho S, Court CM, Winograd P, et al. Precision oncology using a limited number of cells: optimization of whole genome amplification products for sequencing applications. *BMC Cancer.* 2017;17(1):457.
22. Beroukhi R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature.* 2010;463(7283):899-905.
23. Hicks J, Krasnitz A, Lakshmi B, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res.* 2006;16(12):1465-1479.
24. Panaro NJ, Yuen PK, Sakazume T, Fortina P, Kricka LJ, Wilding P. Evaluation of DNA fragment sizing and quantification by the agilent 2100 bioanalyzer. *Clin Chem.* 2000;46(11):1851-1853.
25. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* 2011;17(1):10.
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
27. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
28. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Chen C, Li J, Wan J, Lu Y, Zhang Z, Xu Z. A low cost and input tailing method of quality control on multiple annealing, and looping-based amplification cycles-based whole-genome amplification products. *J Clin Lab Anal.* 2019;33:e22697. <https://doi.org/10.1002/jcla.22697>