

Research Article

The comparison of the performance of four whole genome amplification kits on ion proton platform in copy number variation detection

Xinyi Zhang^{1,*}, Bo Liang^{2,3,*}, Xiaoyan Xu¹, Feifei Zhou¹, Lingyin Kong⁴, Jingjing Shen⁴, Yingying Xia², Liming Xuan⁴, Yan Mao², Yongfeng Xue², Caixia Liu^{5,6} and Jichun Tan¹

¹ Reproductive Medical Center of Gynecology and Obstetrics Department, Shengjing Hospital Affiliated to China Medical University, Shenyang, Liaoning 110004, China; ² Basecare Medical Device Co., Ltd., Suzhou, Jiangsu 215000, China; ³ State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.; ⁴ Cyagen Health Inc., Suzhou, Jiangsu 215400, China; ⁵ Key Laboratory of Maternal-Fetal Medicine of Liaoning Province, Shengjing Hospital Affiliated to China Medical University, Shenyang, Liaoning 110004, China.; ⁶ Key Laboratory of Obstetrics and Gynecology of Higher Education of Liaoning Province, Shengjing Hospital Affiliated to China Medical University, Shenyang, Liaoning 110004, China.

Correspondence: Jichun Tan (tjczjh@163.com) and Caixia Liu (liucx1716@163.com)



With the development and clinical application of genomics, more and more concern is focused on single-cell sequencing. In the process of single-cell sequencing, whole genome amplification is a key step to enrich sample DNA. Previous studies have compared the performance of different whole genome amplification (WGA) strategies on Illumina sequencing platforms, but there is no related research aimed at Ion Proton platform, which is also a popular next-generation sequencing platform. Here by amplifying cells from six cell lines with different karyotypes, we estimated the data features of four common commercial WGA kits (PicoPLEX WGA Kit, GenomePlex Single Cell Whole Genome Amplification Kit, MALBAC Single Cell Whole Genome Amplification Kit, and REPLI-g Single Cell Kit), including median absolute pairwise difference, uniformity, reproducibility, and fidelity, and examined their performance of copy number variation detection. The results showed that both MALBAC and PicoPLEX could yield high-quality data and had high reproducibility and fidelity; and as for uniformity, PicoPLEX was slightly superior to MALBAC.

Introduction

With the decreasing cost and increasing clinical application of next-generation sequencing (NGS), more and more concerns are focused on NGS. Ion Proton is a semiconductor-based NGS platform [1]. The reads yielded by Ion Proton platform are length-variable single-end reads [2], which is quite different from those of Illumina sequencing platforms, e.g. Hiseq series [3,4]. Ion Proton is one of the currently main NGS platforms and is widely applied to diverse fields, including exome sequencing, transcriptome sequencing, single-cell sequencing etc. [5-9], and especially it often occurs in clinical examination, such as noninvasive prenatal diagnosis [10] and preimplantation genetic screening [11].

Single-cell sequencing offers cellular-level genetic information, so it is an important approach to study early embryo development, neuron genomic heterogeneity, carcinogenesis, and uncultivable bacteria etc. [12,13]. Along the process of single-cell sequencing, whole genome amplification (WGA) is a key step [14], because of the low quantity of available materials. There are plenty of commercial WGA kits and different kits are based on different strategies: multiple displacement amplification (MDA), degenerate-oligonucleotide-primed PCR (DOP-PCR), and multiple annealing and looping-based amplification cycles (MALBAC, MALBAC-like) are three major WGA strategies. Different strategy may cause different WGA performance and some studies have paid efforts in the comparison of WGA methods

*These authors contributed equally to this work.

Received: 15 March 2017

Revised: 16 May 2017

Accepted: 01 June 2017

Accepted Manuscript Online: 01 June 2017

Version of Record published: 7 July 2017

[15–19]. However, current related researches are conducted on Illumina sequencing platforms, the results have not validated on Ion Proton platform.

To address this issue, we tested four common commercial WGA kits—PicoPLEX WGA Kit (which we called PicoPLEX, Rubicon Genomics, U.S.A.), GenomePlex Single Cell Whole Genome Amplification Kit (GenomePlex, Sigma–Aldrich, U.S.A.), MALBAC Single Cell WGA Kit (MALBAC, Yikon Genomics, China), and REPLI-g Single Cell Kit (MDA, Qiagen, Germany). The first two kits were based on DOP-PCR method, and the two following were based on MALBAC and MDA respectively. The amplified products were sequenced on Ion Proton platform, and then we evaluated the data features, compared the difference among different sample types, and examined their performance in copy number variation (CNV) detection. In this way, we found that MALBAC and PicoPLEX could yield high-quality data and exhibited high reproducibility, and in terms of uniformity, PicoPLEX was slightly better than MALBAC. Our results provided system comparison among the four common WGA kits and guidance for future studies related to single-cell sequencing on Ion Proton platform.

Materials and methods

Samples

Six cell lines in the study were bought from the Coriell Institute (Coriell Biorepositories, U.S.A.), and their karyotypes covered common CNV types, including four aneuploids (+13, +18, +21, and XO), one microdeletion ($\text{del}(16)(\text{pter}\rightarrow\text{p12}::\text{p11.2}\rightarrow\text{qter})$), and one unbalanced translocation ($\text{der}(4)\text{t}(4;12)(\text{p16.1};\text{p12})\text{pat}$). Single-cell, multi-cell, and genomic DNA (gDNA) samples were separately prepared for the six cell lines.

Single cell and multi cells (3–5 cells) were isolated by micromanipulating under a dissection microscope (Olympus SZ61, Japan) using a finely pulled glass Pasteur pipette. Then, the isolated cells were washed twice with phosphate-buffered saline (PBS) and transferred into 200 μl PCR tubes containing 2.5 μl of PBS.

The gDNAs of the cell lines were extracted by Qiagen DNeasy Blood and Tissue Kit (Qiagen, Germany) according to the manufacturer's instructions. The concentration of purified gDNA was determined using Quant-iT Assay Kit with Qubit fluorometer (Invitrogen, U.S.A.) and purity analysis was conducted by Nanodrop 2000 (Thermo Scientific, Germany). Then, 6 ng of gDNA was extracted and serially diluted 1000-fold by Nuclease-Free Water (Qiagen, Germany) to adjust the concentration to 6 $\text{pg}/\mu\text{l}$. We aliquoted 2.5 μl of the diluted solution so that each aliquot contained 15 pg of gDNA, which was equivalent to the amount of the DNA in a single cell (~ 6.6 pg of DNA), taking into account the random loss of DNA due to limiting dilution.

These samples were stored in -80°C before WGA. All procedures were performed in a dedicated laboratory with positive pressure HEPA-filtered air to avoid contamination.

Whole genome amplification

According to the recommended protocols, Single-cell, multi-cell, and gDNA samples were amplified by four common commercial WGA kits: PicoPLEX WGA Kit (Rubicon Genomics, U.S.A., version: R30050), GenomePlex Single Cell Whole Genome Amplification Kit (Sigma–Aldrich, U.S.A., version: WGA4-50RXN), MALBAC Single Cell WGA Kit (Yikon Genomics, China, version: YK001B), and REPLI-g Single Cell Kit (Qiagen, Germany, version: 150343). After WGA, the products were purified by Agencourt Ampure XP beads (Beckman Coulter, U.S.A.). The purified DNA was quantified using the Quant-iT Assay Kit with Qubit fluorometer (Invitrogen, U.S.A.), and the purity was determined by Nanodrop 2000 (Thermo Scientific, U.S.A.). Then the amplified products were separated by 2% agarose gels electrophoresis in TBE buffer (90 mmol/l Tris-borate, pH 8.0, 2 mmol/l EDTA), stained by SYBR Green (Invitrogen, U.S.A.), and visualized by Image Master software (Bio-Rad, U.S.A.).

Library construction and whole-genome DNA sequencing

The WGA products were fragmented to an average size of 170 bp and end ligated with sequencing adaptors carrying identifiable barcode for Ion Proton platform. Tagged fragments were then amplified by PCR with adaptor primers to generate 250 bp (80 bp adaptors/barcode and 170 bp insert DNA) sequencing libraries. Finally, the libraries with different barcode were pooled together and sequenced on Ion Proton platform with PI chip.

Data analysis

All sequencing data were aligned to the human genome reference sequences (version: NCBI Build37/hg19) by TMAP software. Then, PCR duplications were removed by Picard software, and uniquely mapped reads, reads with mapping quality score (MAPQ) ≥ 10 and length ≥ 35 bp, were kept for the following analysis.

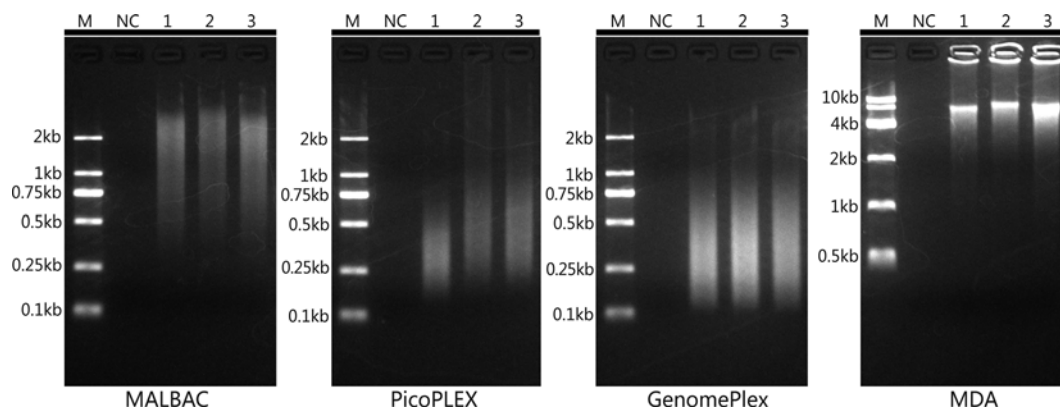


Figure 1. Gel electrophoresis of WGA products

M, DNA molecular weight markers; NC, negative control; Lane 1, amplified gDNA sample; Lane 2, single-cell sample; Lane 3, multi-cell sample.

Table 1 Basic statistics of single cells amplified by different kits

Reagent	Raw reads	Depth	Coverage (%)	Mapping rate (%)	Duplication rate (%)	Unique reads ratio (%)	GC (%)	MAPD
MALBAC	5,244,767	0.26	15.82	98.05	6.98	58.11	40.91	0.31
MDA	4,699,604	0.31	8.06	99.21	10.26	55.46	40.46	2.48
GenomePlex	4,768,010	0.23	13.72	97.63	10.06	40.95	40.90	0.42
PicoPLEX	5,791,391	0.30	15.37	97.20	11.33	39.19	40.92	0.24
Bulk	4,644,129	0.25	18.71	98.79	9.61	57.34	40.91	0.20

For each sample, all chromosomes were divided into segments of the same size (20 kbp), called windows. The number of uniquely mapped reads and GC content in each window was determined. Then, we filtered windows containing no reads and windows with 'N's in the reference sequences. The number of uniquely mapped reads in the remaining windows was corrected and normalized based on their GC content by LOESS regression. Finally, taking corresponding samples from normal humans' oral epithelium as control, we applied the circular binary segmentation (CBS) algorithm to segment the genome into regions with equal copy number, and screened out CNVs whose mean of the log of normalized read count were less than -0.51 or more than 0.37 and size were more than 4 Mbp.

Results

The yield and size of WGA products

Single-cell, multi-cell, and gDNA samples were amplified by MALBAC, PicoPLEX, GenomePlex, and MDA respectively. Gel analysis was used to determine the yield and size range of the WGA products (Figure 1). Among different sample types, the yield and size range were similar in the same WGA kit. And the yield was $1.0 \pm 0.32 \mu\text{g}$ in MALBAC, $2.0 \pm 0.19 \mu\text{g}$ in PicoPLEX, $4.4 \pm 2.3 \mu\text{g}$ in GenomePlex, and $13.6 \pm 6.7 \mu\text{g}$ in MDA; the size of products distributed approximately 0.5–2.0 kb in MALBAC, 0.25–1.0 kb in PicoPLEX, 0.1–1.0 kb in GenomePlex and 4.0–10.0 kb in MDA. Above results were consistent with the results of previous studies [16,20].

Quality assessment of sequencing data

In order to estimate the quality of data generated from the four kits on Ion Proton platform, we evaluated the quantity of raw reads, depth, coverage, mapping rate, duplication rate, the ratio of unique reads, genomic GC percent, and median absolute pairwise difference (MAPD) (Table 1 and Figure 2). The results showed that although the means of GC percents of the four kits were almost the same, the peaks of GC percents of MALBAC and PicoPLEX slightly shifted to the right comparing with the bulk data, while those of MDA and GenomePlex were highly consistent with the bulk data (Figure 2b). With similar number of raw reads to the other three kits, the coverage of MDA was the lowest among the four kits and the coverage on different depths suggested that the uniformity of depth in MDA was much worse than the other three kits and deviated severely from the bulk data. According to Lander et al. [21] and our practical experience, we regarded sequencing data whose cumulative coverage on 20% of the average depth was less

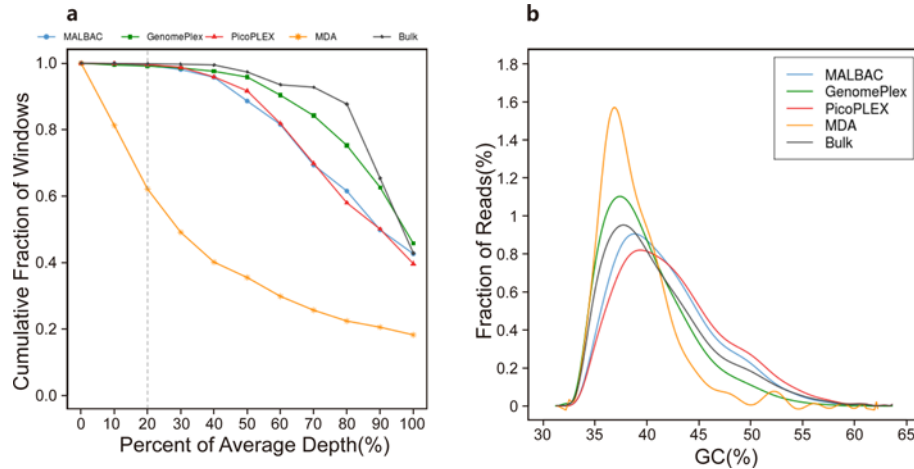


Figure 2. Uniformity of coverage and GC bias analysis

(a) The cumulative distribution of read depth. The cumulative fraction of windows on the depth of 20% of average depth (dash line) is a metric for data quality assessment. (b) The reads distribution under different GC content levels.

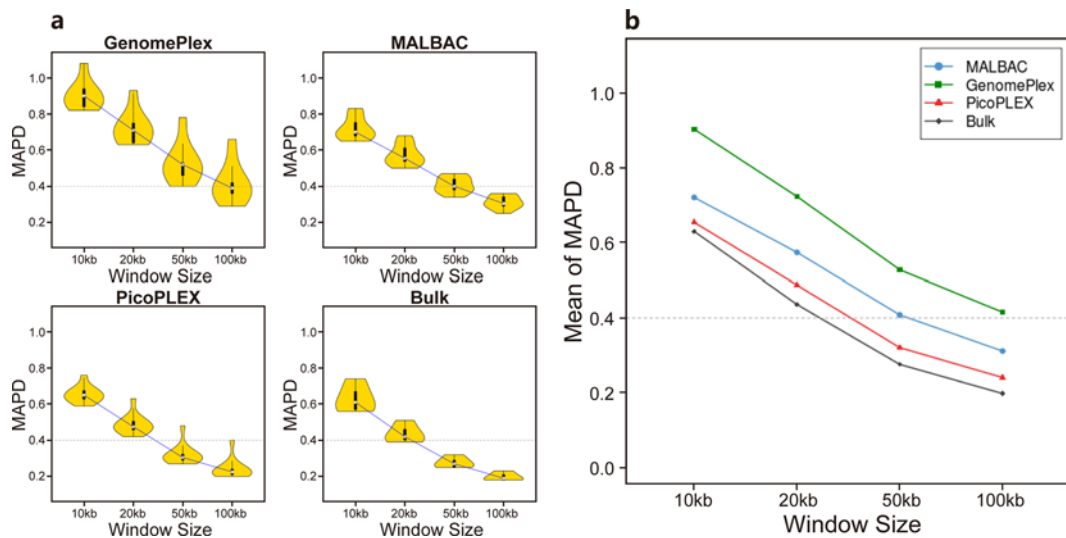


Figure 3. MAPDs of WGA kits and bulk data in different window sizes

(a) MAPDs of read count yielded by GenomePlex, MALBAC, PicoPLEX, and the bulk-cell sample respectively, are compared in different window sizes. The threshold of MAPD is 0.4 (dash line). (b) The average MAPDs among all samples of GenomePlex, MALBAC, PicoPLEX, and the bulk data respectively, are plotted against window sizes.

than 90% as not quantitated data for further analysis. In our results, the cumulative coverage on 20% of the average depth of MDA was only 62%, while the other three kits were ~99%.

To assess copy number (CN) noise across the whole genome, we employed MAPD algorithm [22], which was widely applied in quality assessment for microarray data and also was adapted on Hiseq platform [23,24]. MAPD algorithm was the median of the absolute values of deviation between the log₂ CN ratios of every pair of adjacent windows. Theoretically, higher MAPD scores reflected greater genomic noise and we eliminated data whose MAPD score was higher than 0.4.

We tested the performance of data in different window sizes (Figure 3) and found that larger window size contained less genomic noise, at the cost of CNV resolution. MAPDs in 100kb windows of both PicoPLEX and MALBAC passed the threshold (0.23 ± 0.05 and 0.31 ± 0.04 respectively), and the mean of MAPD of PicoPLEX was lower and closest to that of the bulk data in the four kits. MAPD of MDA was as high as 2.48 even in 100 kb windows, consistent with the bad uniformity of coverage with depth in MDA.

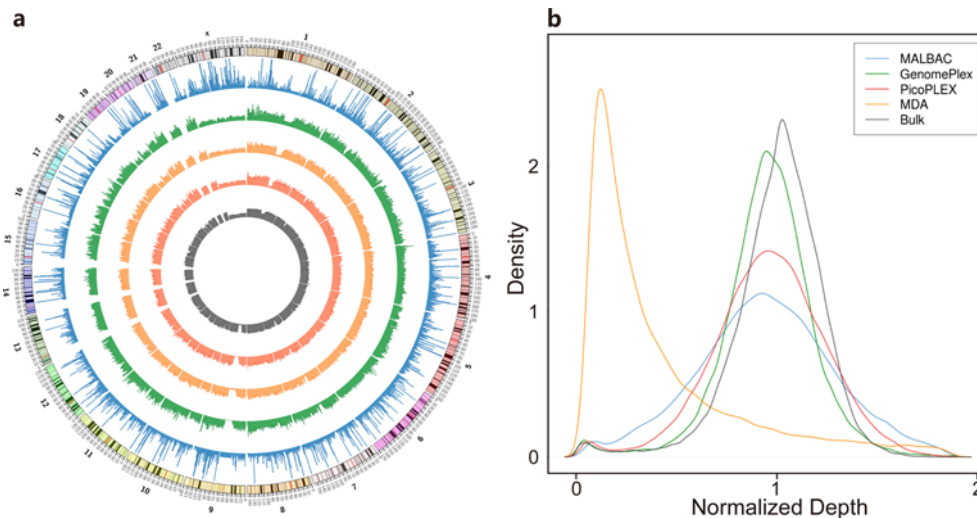


Figure 4. Uniformity of read depth across the whole genome

(a) Read depth distribution across the whole genome. The five circles from inside to outside are normalized read depth in 100 kb windows of the bulk data (black), GenomePlex (light red), PicoPLEX (orange), MALBAC (green), and MDA (blue). The circle outside them is the chromosomal coordinate in a pter-qter orientation, which is zoomed out by 1,000,000 times, i.e. the actual position is the number on the circle multiplying 1 Mbp. (b) The Kernel density plots of normalized read depth in 100 kb windows of the four WGA kits and the bulk data.

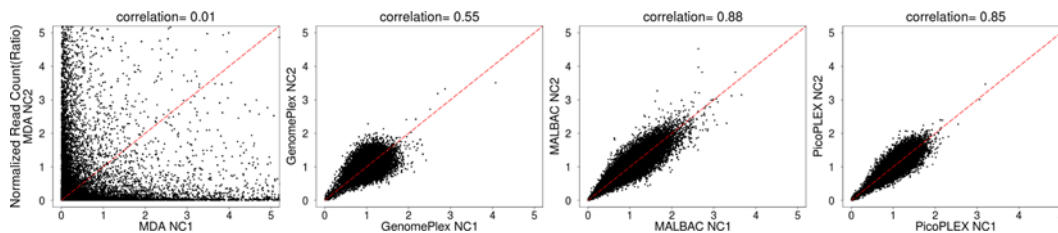


Figure 5. Amplification reproducibility of WGA kits

The normalized read count in 100 kb windows of WGA products is plotted against that of another repeat of the same WGA kit, and the Pearson correlation coefficient between the two repeats is showed on the top. From left to right are plots of MDA, GenomePlex, MALBAC, and PicoPLEX.

Uniformity and reproducibility analysis

A uniform distribution of read depth was the basis of CNV analysis. Among the four kits, the narrow distribution of read depth approximately 1 in GenomePlex and PicoPLEX suggested their high uniformity, while the distribution of MALBAC was wider and had fat tails, indicating relatively lower uniformity in MALBAC. The distribution of MDA severely deviated from 1 and the distributions of the other three WGA kits and the bulk data, supporting the conclusion mentioned above (Figure 4).

To estimate the reproducibility of different kits, we compared the normalized read ratios between two repeats of the four kits and calculated their Pearson correlation coefficient (Figure 5). The correlation coefficients between two repeats of MALBAC and PicoPLEX were 0.88 and 0.85 respectively, while that of GenomePlex was only 0.55. MDA showed the lowest correlation coefficients, i.e. 0.01, indicating that there was nearly no linear correlation between two repeats of MDA.

In summary, GenomePlex products had high uniformity but low reproducibility and MALBAC had high reproducibility but lower uniformity than GenomePlex and PicoPLEX, while MDA had bad uniformity and bad reproducibility. PicoPLEX was superior to the other three kits in both uniformity and reproducibility, and was recommended for CNV analysis on Ion Proton platform.

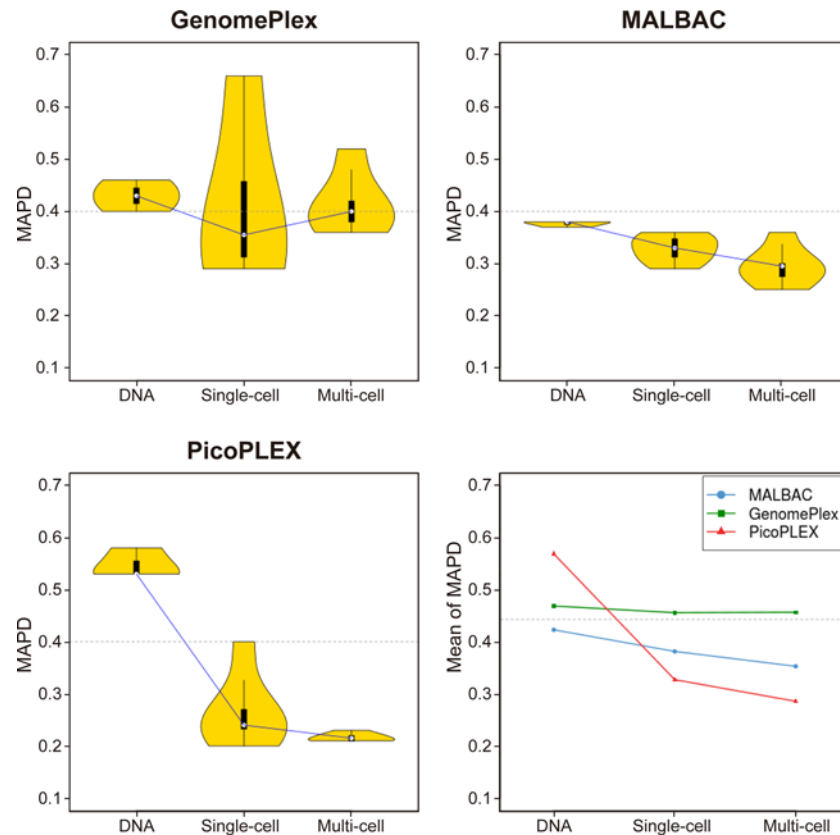


Figure 6. MAPDs of WGA kits in different types of samples

MAPDs of read count yielded by GenomePlex, MALBAC, and PicoPLEX respectively, are compared in different sample types.

Comparison among different types of samples

To examine the performance of different kits in different types of samples, we extracted gDNA, single cell, and multi cells from normal humans' oral epithelium and conducted whole genome amplification on all of them by GenomePlex, MALBAC, and PicoPLEX (MDA was excluded because its performance was not suitable for CNV analysis as mentioned above), and then sequenced them on Ion Proton platform. Except for GenomePlex whose MAPDs fluctuated around 0.40, the MAPD of MALBAC and PicoPLEX decreased from gDNA to multi-cell samples (Figure 6), indicating a lower genomic noise level in multi-cell samples.

In order to evaluate the amplification bias of different WGA kits in different sample types, we plotted the normalized read count of samples amplified by different WGA kits against the unamplified one (Figure 7). The correlation coefficients between samples amplified by MALBAC and the bulk-cell sample fluctuated around 0.62 among different sample types, and the correlation coefficients of GenomePlex and PicoPLEX demonstrated higher consistency with unamplified data in multi-cell samples than in gDNA or single-cell samples, indicating smaller amplification bias in multi-cell samples. In total, PicoPLEX had the greatest correlation coefficients except for gDNA samples, suggesting it contained less amplification bias than others, i.e. PicoPLEX showed the best fidelity.

CNVs detection

To estimate the effect of the four kits on variation detection, we selected cell lines of six different CNVs with confirmed karyotypes, including four aneuploids, one microdeletion, and one unbalanced translocation, and detected CNVs in their multi-cell samples of the four kits. Using the same software and parameters, the four WGA kits showed different performance in CNV detection (Table 2 and Figure 8). The CNVs detected in PicoPLEX and MALBAC products were completely consistent with the confirmed CNVs of each cell line, while there were two false positives results in GenomePlex. The detected CNV results were totally wrong in MDA under the same condition, associating with its bad genomic uniformity of depth.

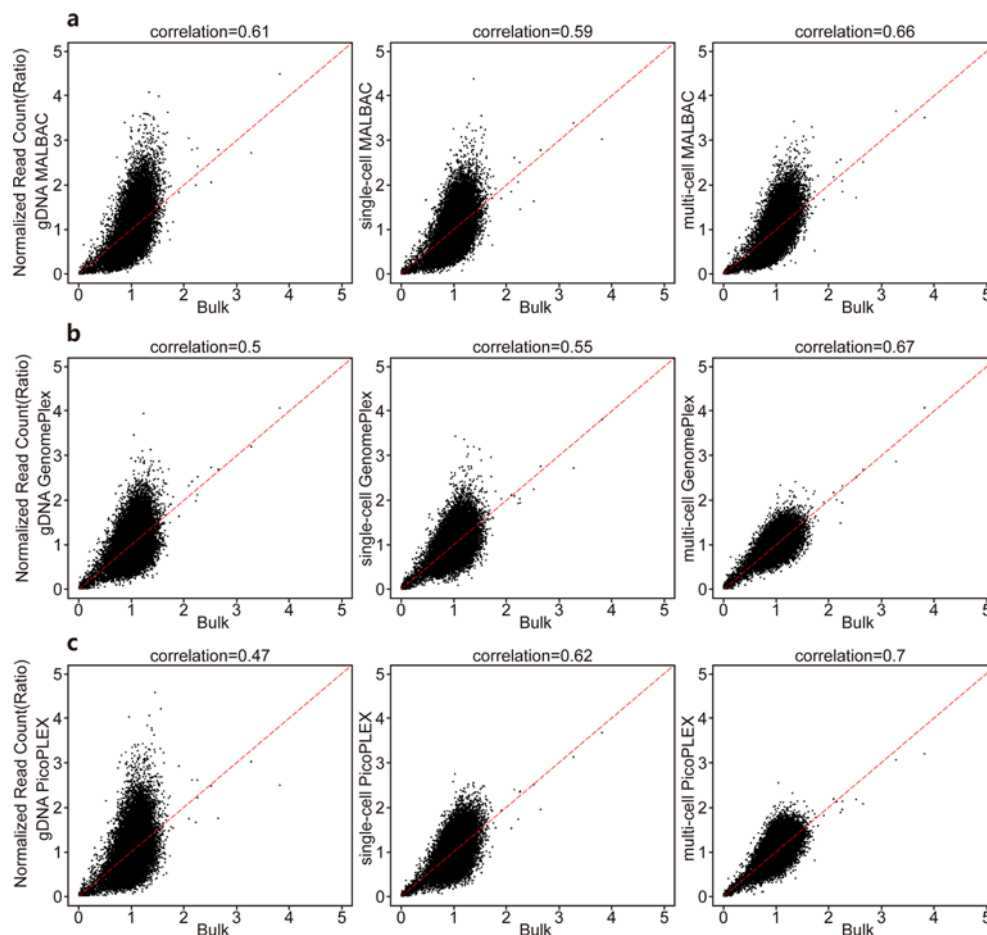


Figure 7. Amplification bias of WGA kits in different types of samples

The normalized read count of WGA products of (a) MALBAC, (b) GenomePlex, and (c) PicoPLEX in different sample types is plotted against that of the unamplified sample. From left to right are plots of gDNA samples, single-cell samples, and multi-cell samples respectively.

Table 2 CNVs detection results of six confirmed cell lines

Cell Line	Confirmed CNV	MALBAC	PicoPLEX	GenomePlex	MDA
GM00857	45,XO	45,XO	45,XO	45,XO	+21, -22
GM05875	46,XX, del(16)(pter ->p12::p11.2->qter)	46,XX, del(16p12.1-p11.2)	46,XX, del(16p12.1-p11.2)	46,XX, del(16p12.1-p11.2)	NA
GM04592	47,XX,+21	47,XX,+21	47,XX,+21	46,XX,+21, del(1p31.1-p22.3)	-22
GM01359	47,XY,+18	47,XY,+18	47,XY,+18	47,XY,+18, dup(6q15-q16.1)	NA
GM03330	47,XY,+13	47,XY,+13	47,XY,+13	47,XY,+13	NA
GM01183	46,XY, der(4)t(4;12) (p16.1;p12)pat	46,XY, del(4p16.3-p16.2), dup(12p13.33-p12.1)	46,XY, del(4p16.3-p16.2), dup(12p13.33-p12.1)	46,XY, del(4p16.3-p16.2), dup(12p13.33-p12.1)	NA
Bulk	46,XX	46,XX	46,XX	46,XX	46,XX

Results consistent with the confirmed karyotypes were bolded.

Discussion

In the present study, we examined the performance of four common WGA kits on Ion Proton sequencing platform. We extracted gDNA, single-cell, and multi-cell samples from six positive cell lines and humans' oral epithelium, conducted whole genome amplification on them by MDA, GenomePlex, MALBAC, and PicoPLEX respectively, and

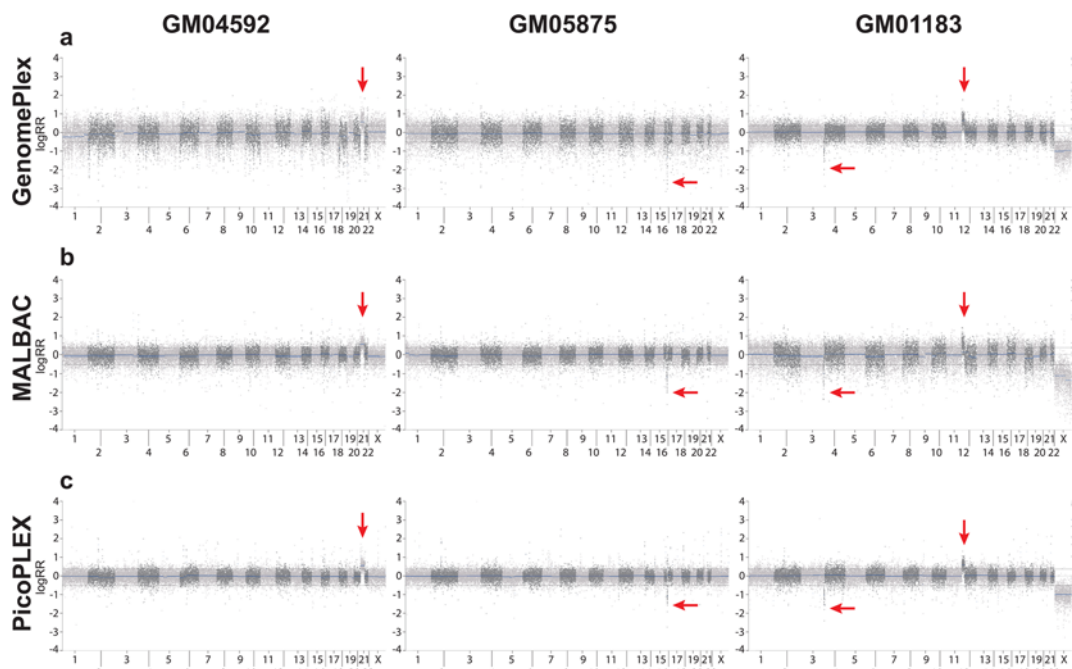


Figure 8. Plots of copy number of samples from different cell lines amplified by different WGA kits

The log of normalized read count (logRR) of WGA products of (a) GenomePlex, (b) MALBAC, and (c) PicoPLEX is plotted. From left to right are plots of samples from cell lines of aneuploids (GM04592), microdeletion (GM05875), and unbalanced translocation (GM01183) respectively. The x-axis of each plot stands for 23 chromosomes of human without chromosome Y, and points of adjacent chromosomes are drawn in different colors. Segments marked by red arrows are confirmed CNVs in the cell line. Because the control consists of female samples while the gender of GM01183 is male, the copy number of chromosome X in GM01183 appears less than normal.

sequenced them on Ion Proton platform. Then we estimated the quantity and size of WGA products, the quality, uniformity, reproducibility, and fidelity of sequencing data, and CNV detection performance of the four WGA kits. The results indicated that both MALBAC and PicoPLEX could yield high-quality data and showed high reproducibility and fidelity in reads distribution, but as for uniformity and genomic noise, according to MAPDs, PicoPLEX performed slightly better than MALBAC and closer to the distribution of the bulk-cell sample. Although GenomePlex had high uniformity, it contained excessive genomic noise, which had increased false positive errors in CNV detection. MDA showed the worst uniformity, reproducibility, and CNV detection performance among the four WGA kits, indicating that it was not suitable for CNV analysis.

In practical manufacture and application of WGA kits, there ought to be a reference standard in the kit. In theory, the reference standard could be a lysed single cell isolated from positive cell lines or corresponding DNA. Since DNA had an advantage in preservation and was more convenient in operation, we evaluated the property of the WGA products from gDNA sample, attempting to replace lysed cells with gDNA as reference standards. The MAPDs of the amplified gDNA samples were more than those of single-cell or multi-cell samples amplified by the same WGA kits, suggesting that there was more genomic noise in the amplified gDNA samples. The correlation coefficients between amplified gDNA samples and the bulk data was the lowest among the three sample types, which reflected the low fidelity in amplified gDNA samples. The bad performance of gDNA samples amplified by WGA kits might result from the nonrandom distributed DNA fragments in gDNA diluents.

Considering the relatively low throughput on Ion Proton, low-coverage whole genome sequencing was common on this platform but it was also a challenge for variant calling. To detect different variants, there were different choices of WGA kits. For small CNV detection, PicoPLEX and MALBAC were both good choices; for large CNV detection, PicoPLEX, MALBAC, and GenomePlex were competent; and for SNP calling, MDA and MALBAC were appropriate. Different projects focused on different variants. For example, preimplantation genetic screen currently focused on CNV detection [25-27], inheritable diseases studies were mainly interested in SNP calling [28-31], and cancer-related studies might concern CNVs, SNPs, and other variants [32,33]. There was no best method for all projects, but there was always a more appropriate one for a certain project.

In conclusion, we quantitatively examined the performance of four common commercial WGA kits on Ion Proton platform. Based on samples extracted from six positive cell lines and normal humans' oral epithelium, we estimated the data yield, data quality, uniformity, reproducibility, fidelity, and CNV detection performance of PicoPLEX, GenomePlex, MALBAC, and MDA. At a similar yield and laboratorial situation, PicoPLEX showed the best performance in the quality of sequencing data, uniformity of read depth, amplification reproducibility and fidelity, so it was recommended for CNV calling on Ion Proton platform. Results in the present study would guide the selection of WGA kits on Ion Proton platform and assist the application of single-cell sequencing on this platform.

Funding

This work was supported by National High-Tech R&D Program (863)[grant number 2015AA020407]; Clinical Research Center for Obstetrical Diseases and Collaborative Network Construction of Liaoning Province [grant number 2016007010]; and Science and Technology Support Project of Jiangsu Province [grant number BE2013654]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We sincerely thank our colleagues at Basecare Medical Device Co., Ltd., for the work.

Author Contribution

X.Z., B.L., and J.T. conceived the study, X.X. and F.Z. prepared the samples, X.Z., J.S., Y.M., and Y.F.X. conducted the experiments, B.L., L.K., Y.Y.X., and L.X. analyzed the data. X.Z. and B.L. wrote the draft of the manuscript, C.L. and J.T. supervised the work, and all authors reviewed and approved the manuscript.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Abbreviations

CNV, copy number variation; EDTA, ethylenediaminetetraacetic acid; gDNA, genomic DNA; LOESS, local regression, a non-parametric regression method based on k-nearest neighbors algorithm (k-NN); MAPD, median absolute pairwise difference; MDA, multiple displacement amplification; NGS, next-generation sequencing; PBS, phosphate-buffered saline; SNP, single-nucleotide polymorphism; TBE, Tris-Borate-EDTA buffer; WGA, whole genome amplification.

References

- 1 Rothberg, J.M. et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352, doi:10.1038/nature10242
- 2 Merriman, B., Ion Torrent, R., Team, D. and Rothberg, J.M. (2012) Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* **33**, 3397–3417, doi:10.1002/elps.201200424
- 3 Samura, O (2014) A commentary on comparison of the performance of Ion Torrent chips in noninvasive prenatal trisomy detection. *J. Hum. Genet.* **59**, 421–422, doi:10.1038/jhg.2014.56
- 4 Boland, J.F. et al. (2013) The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Hum. Genet.* **132**, 1153–1163, doi:10.1007/s00439-013-1321-4
- 5 Shih, A.H. et al. (2015) Mutational cooperativity linked to combinatorial epigenetic gain of function in acute myeloid leukemia. *Cancer Cell* **27**, 502–515, doi:10.1016/j.ccell.2015.03.009
- 6 Belloni, E. et al. (2015) Whole exome sequencing identifies driver mutations in asymptomatic computed tomography-detected lung cancers with normal karyotype. *Cancer Genet.* **208**, 152–155, doi:10.1016/j.cancergen.2015.02.004
- 7 Wen, Z. et al. (2014) Synaptic dysregulation in a human iPS cell model of mental disorders. *Nature* **515**, 414–418, doi:10.1038/nature13716
- 8 Rahman, H. et al. (2014) Transcriptome analysis of salinity responsiveness in contrasting genotypes of finger millet (*Eleusine coracana* L.) through RNA-sequencing. *Plant Mol. Biol.* **85**, 485–503, doi:10.1007/s11103-014-0199-4
- 9 Motoike, I.N. et al. (2014) Validation of multiple single nucleotide variation calls by additional exome analysis with a semiconductor sequencer to supplement data of whole-genome sequencing of a human population. *BMC Genomics* **15**, 673, doi:10.1186/1471-2164-15-673
- 10 Liao, C. et al. (2014) Noninvasive prenatal diagnosis of common aneuploidies by semiconductor sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7415–7420, doi:10.1073/pnas.1321997111
- 11 Deleye, L. et al. (2015) Shallow whole genome sequencing is well suited for the detection of chromosomal aberrations in human blastocysts. *Fertil. Steril.* **104**, 1276.e1271–1285.e1271, doi:10.1016/j.fertnstert.2015.07.1144
- 12 Liang, J., Cai, W. and Sun, Z (2014) Single-cell sequencing technologies: current and future. *J. Genet. Genomics* **41**, 513–528, doi:10.1016/j.jgg.2014.09.005
- 13 Wang, Y. and Navin, N.E. (2015) Advances and applications of single-cell sequencing technologies. *Mol. Cell* **58**, 598–609, doi:10.1016/j.molcel.2015.05.005
- 14 Gawad, C., Koh, W. and Quake, S.R. (2016) Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188, doi:10.1038/nrg.2015.16

- 15 Ning, L. et al. (2015) Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons. *Sci. Rep.* **5**, 11415, doi:10.1038/srep11415
- 16 Li, N. et al. (2015) The performance of whole genome amplification methods and next-generation sequencing for pre-implantation genetic diagnosis of chromosomal abnormalities. *J. Genet. Genomics* **42**, 151–159, doi:10.1016/j.jgg.2015.03.001
- 17 Huang, L., Ma, F., Chapman, A., Lu, S. and Xie, X.S. (2015) Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu. Rev. Genomics Hum. Genet.* **16**, 79–102, doi:10.1146/annurev-genom-090413-025352
- 18 Hou, Y. et al. (2015) Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *Gigascience* **4**, 37, doi:10.1186/s13742-015-0068-3
- 19 de Bourcy, C.F. et al. (2014) A quantitative comparison of single-cell whole genome amplification methods. *PLoS ONE* **9**, e105585, doi:10.1371/journal.pone.0105585
- 20 Treff, N.R., Su, J., Tao, X., Northrop, L.E. and Scott, Jr, R.T. (2011) Single-cell whole-genome amplification technique impacts the accuracy of SNP microarray-based genotyping and copy number analyses. *Mol. Hum. Reprod.* **17**, 335–343, doi:10.1093/molehr/gaq103
- 21 Lander, E.S. and Waterman, M.S. (1988) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239
- 22 Affymetrix. (2008) Median of the absolute values of all pairwise differences and quality control on Affymetrix genome-wide human SNP array 6.0, available at http://tools.thermofisher.com/content/sfs/brochures/mapd_snp6_whitepaper.pdf
- 23 Masson, A.L. et al. (2016) Copy number variants associated with 18p11.32, DCC and the promoter 1B region of APC in colorectal polyposis patients. *Meta Gene* **7**, 95–104, doi:10.1016/j.mgene.2015.12.005
- 24 Cai, X. et al. (2014) Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **8**, 1280–1289, doi:10.1016/j.celrep.2014.07.043
- 25 Magli, M.C. et al. (2016) Preimplantation genetic testing: polar bodies, blastomeres, trophoctoderm cells, or blastocoelic fluid? *Fertil. Steril.* **105**, 676e675–683e675, doi:10.1016/j.fertnstert.2015.11.018
- 26 Lukaszuk, K. et al. (2016) Next generation sequencing for preimplantation genetic testing of blastocysts aneuploidies in women of different ages. *Ann. Agric. Environ. Med.* **23**, 163–166, doi:10.5604/12321966.1196874
- 27 Chang, J., Boulet, S.L., Jeng, G., Flowers, L. and Kissin, D.M. (2016) Outcomes of in vitro fertilization with preimplantation genetic diagnosis: an analysis of the United States Assisted Reproductive Technology Surveillance Data, 2011–2012. *Fertil. Steril.* **105**, 394–400, doi:10.1016/j.fertnstert.2015.10.018
- 28 North, T.L. et al. (2016) A study of common Mendelian disease carriers across ageing British cohorts: meta-analyses reveal heterozygosity for alpha 1-antitrypsin deficiency increases respiratory capacity and height. *J. Med. Genet.* **53**, 280–288, doi:10.1136/jmedgenet-2015-103342
- 29 Ma, Y. et al. (2016) GJB2, SLC26A4, and mitochondrial DNA12S rRNA hot-spots in 156 subjects with non-syndromic hearing loss in Tengzhou, China. *Acta Otolaryngol.* 1–6, doi:10.3109/00016489.2016.1164893
- 30 Xu, Y. et al. (2015) Embryo genome profiling by single-cell sequencing for preimplantation genetic diagnosis in a beta-thalassemia family. *Clin. Chem.* **61**, 617–626, doi:10.1373/clinchem.2014.228569
- 31 Ben Charfeddine, I. et al. (2012) Steroid 21-hydroxylase gene mutational spectrum in 50 Tunisian patients: characterization of three novel polymorphisms. *Gene* **507**, 20–26, doi:10.1016/j.gene.2012.07.027
- 32 Pfarr, N. et al. (2016) Copy number changes of clinically actionable genes in melanoma, non-small cell lung cancer and colorectal cancer - a survey across 822 routine diagnostic cases. *Genes Chromosomes Cancer* **55**, 821–833, doi:10.1002/gcc.22378
- 33 Hu, L. et al. (2016) Clinical significance of long non-coding RNA CASC8 rs10505477 polymorphism in lung cancer susceptibility, platinum-based chemotherapy response, and toxicity. *Int. J. Environ. Res. Public Health* **13**, pii: E545, doi:10.3390/ijerph13060545