Next-generation whole genome amplification methods for CNV and SNV detection from single cells

Luke Sherlin*, Marta Gonzalez, Greyson King, Datta Mellacheruvu, Fang Sun, Julie Laliberte, Emmanuel Kamberov & Andrew Farmer

Takara Bio USA, Inc., Mountain View, CA 94043, USA *Corresponding author: luke_sherlin@takarabio.com

Abstract

Preparation of amplified genomic material from small amounts of DNA or single cells is extremely important in assisting research involving genetic analyses of clinical samples aimed at identifying the best treatment regimen and molecular diagnoses of diseases such as cancer. Technologies that allow for accurate and reproducible detection of single nucleotide variations (SNVs) and copy number variations (CNVs) in genomic material from limited samples need to do so with high fidelity and high genome coverage. Additionally, these technologies should be flexible enough to be used in a variety of analytical platforms. To address these needs, we have developed the PicoPLEX® WGA V2 System (PicoPLEX WGAv2), a platform-agnostic whole genome amplification system, and the PicoPLEX Gold Single Cell DNA-Seq Kit (PicoPLEX Gold), a complete cells-to-library solution for Illumina® sequencers. These systems use optimized enzymes, primers, and protocols for exceptional sequencing coverage, uniformity, and accuracy in detecting SNVs, all while increasing the resolution for CNV detection relative to previous versions. Both systems maintain the technology's simple workflow and unmatched cell-to-cell reproducibility that is a hallmark of the PicoPLEX technology.

In this study, we demonstrate CNV detection to 5.5-Mb resolution at a depth of 1 million read pairs in a single cell with validated copy number gains and losses. SNV detection and reproducibility are shown to be superior to competitive technologies.

1 PicoPLEX Gold technology—principle and workflow

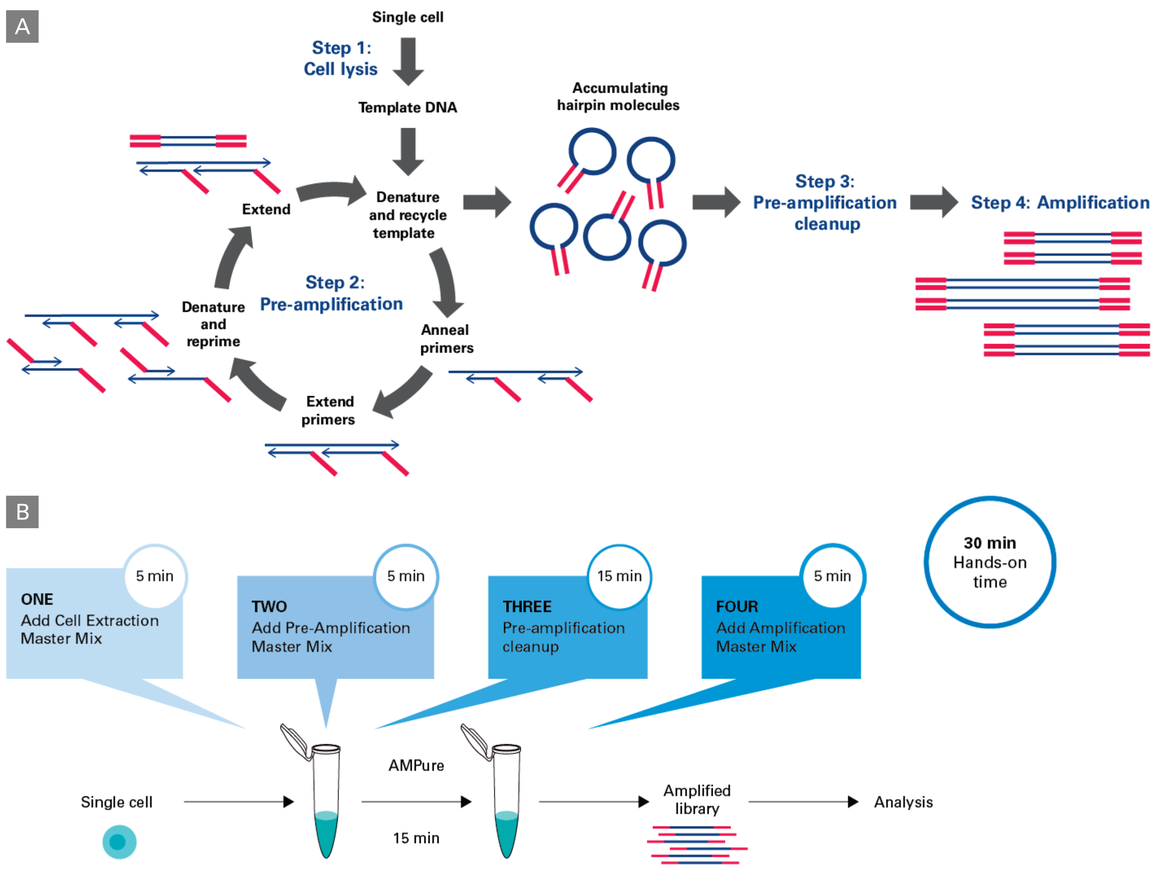

Figure 1. An overview of the principle of PicoPLEX technology and workflow schematic. Panel A. Step 1: A single cell is lysed, and the DNA is released free of proteins. Step 2: Multiple rounds of quasi-linear amplification of the single-cell gDNA is performed. The formation of hairpins prevents subsequent amplification of the products and promotes re-utilization of the original template. Step 3: Sample is cleaned up to remove extra primers. Step 4: gDNA is amplified and sample-barcoded adapters compatible with NGS technologies are added. **Panel B.** Schematic of the PicoPLEX Gold protocol showing a streamlined workflow with minimum hands-on time.

2 High-fidelity detection of single nucleotide variants

The PicoPLEX WGAv2 system is more accurate at detecting SNVs when compared to DOPlify and REPLI-g technologies

Depth of SNV position ≥10 Allele frequency ≥20%		PicoPLEX WGAv2 1 cell		PicoPLEX WGAv2 5 cells		DOPlify 1 cell		DOPlify 5 cells		REPLI-g 1 cell		REPLI-g 5 cells	
	Bulk	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2
Number of SNVs called	74	57	67	69	67	34	57	62	67	Failed	Failed	40	Failed
Number of false positives		3	1	0	1	5	1	0	7	Failed	Failed	0	Failed
Average false positives		0.02%		0.005%		0.03%		0.035%		Failed		0	Failed
Call rate		78%	92%	95%	92%	47%	78%	85%	92%	Failed	Failed	55%	Failed
Average call rate		85%		93%		62%		88%		Failed		55%	Failed
Missed		17	7	5	7	40	17	12	7	Failed	Failed	34	Failed
Average locus dropouts		16.2%		8.1%		38.5%		12.8%		Failed		45.9%	Failed
Number of heterozygous SNVs called	45	45	38	45	45	36	31	38	41	Failed	Failed	32	Failed
Average allele dropouts		7.8%		0.0%		25.6%		12.2%		Failed		71.1%	Failed

Figure 2. Accurate detection of targeted SNVs using the PicoPLEX WGAv2 system. Whole genome amplification products from single- or five-cell samples of a GM12878 cell line (Coriell Institute) were prepared in replicates, using a prototype of the PicoPLEX WGAv2 kit, the PerkinElmer DOPlify kit, and the QIAGEN REPLI-g (MDA) kit. 10 ng of each WGA sample served as input for the Accel-Amplicon Sample-ID Panel (Swift Biosciences). 10 ng of non-WGA gDNA from a bulk NA12878 cell line was used as control in the same amplicon panel library preparation kit. Although sufficient yield was produced by the REPLI-g system, only one five-cell sample produced sufficient amplicon material to sequence, and therefore no data is shown for the other samples. Sequencing was performed on an Illumina MiSeq® platform using V2 chemistry and a read length of 2 x 75 bp. The false-positive rate is calculated using the cumulative coverage of all amplicons in this panel (e.g., 2 false positives times 100, divided by ~10,000 bp results in a 0.02% false-positive rate). An intersection of Genome In a Bottle (GIB) variants to hg19 (human genome assembly GRCh37, Ensembl) showed a total of 78 variants. Due to the amplicon design, paired read lengths of 75 bp were too short to capture in 4 out of the 78 SNVs; therefore, the total number of capturable SNVs was cut to 74. VarDict was used to interpret SNVs from BAM files using the following criteria: depth of SNV position ≥10 reads (10X coverage), allele frequency ≥20%. SNVs detected by all WGA chemistries were compared to those observed in the bulk gDNA control and reported as numbers and percentages.

3 Greatly improved genome coverage and superior reproducibility between two single cells using the PicoPLEX Gold kit

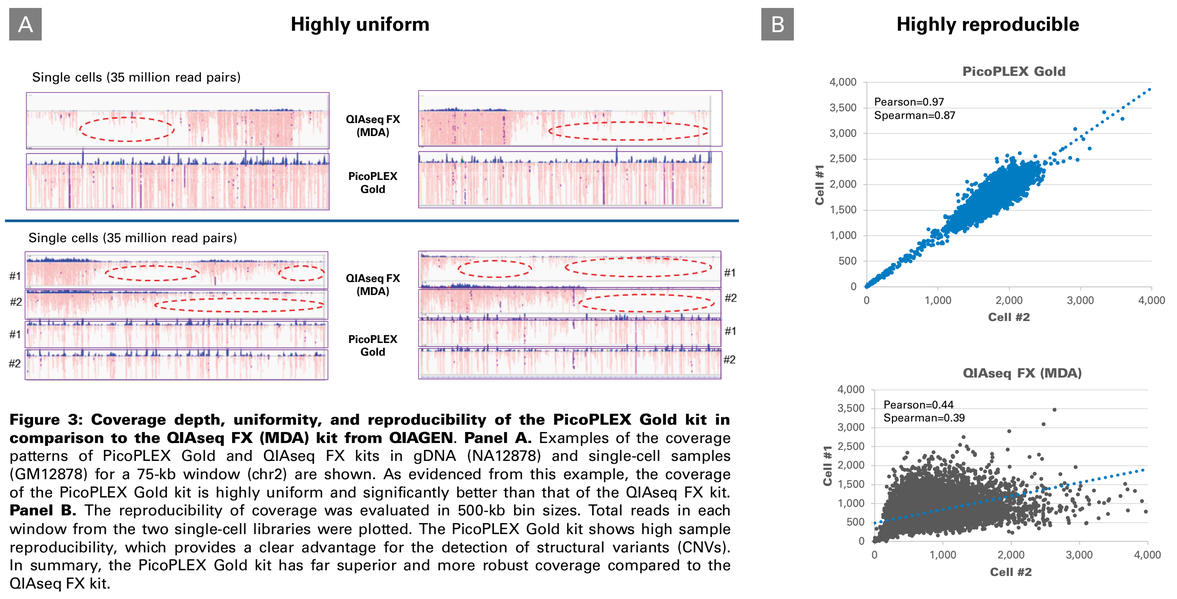

Figure 3: Coverage depth, uniformity, and reproducibility of the PicoPLEX Gold kit in comparison to the QIAseq FX (MDA) kit from QIAGEN. Panel A. Examples of the coverage patterns of PicoPLEX Gold and QIAseq FX kits in gDNA (NA12878) and single-cell samples (GM12878) for a 75-kb window (chr2) are shown. As evidenced from this example, the coverage of the PicoPLEX Gold kit is highly uniform and significantly better than that of the QIAseq FX kit. **Panel B.** The reproducibility of coverage was evaluated in 500-kb bin sizes. Total reads in each window from the two single-cell libraries were plotted. The PicoPLEX Gold kit shows high sample reproducibility, which provides a clear advantage for the detection of structural variants (CNVs). In summary, the PicoPLEX Gold kit has far superior and more robust coverage compared to the QIAseq FX kit.

4 Accurate detection of segmental aneuploidies with low-pass sequencing

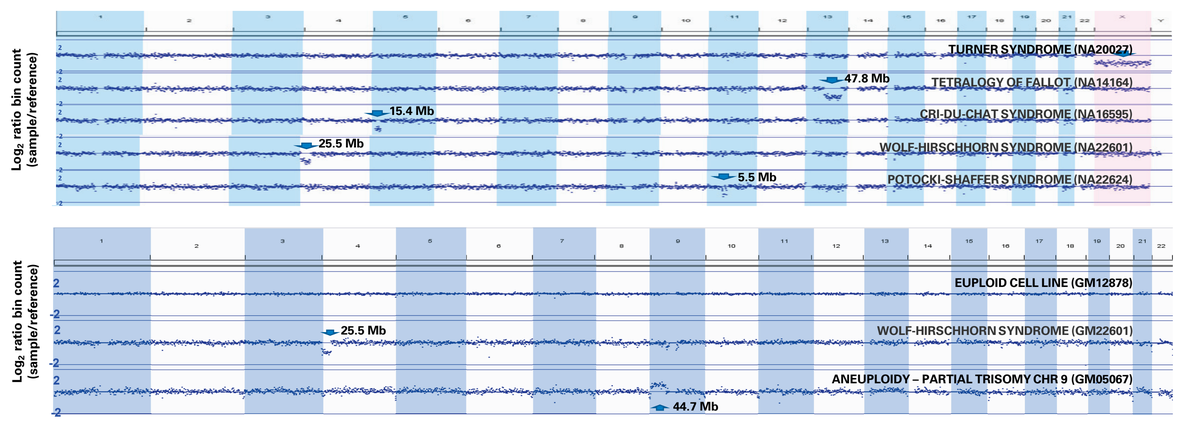

Figure 4. Characterized CNVs detected using the PicoPLEX Gold and PicoPLEX WGAv2 kits. Top Panel. Characterized CNVs detected using the PicoPLEX Gold kit. Genomic DNA from the CNV Panel (Coriell Institute, Cat. # CNVPANEL01) was diluted to 15 pg and processed with the PicoPLEX Gold kit. The resulting libraries were sequenced on an Illumina MiSeq platform using V3 chemistry at a read length of 2 x 75 bp. FASTQ files were trimmed to remove adapters and the first 14 bases and then aligned to the human genome assembly GRCh37. The reference consisted of 24 replicates of 5 sorted peripheral blood mononuclear cells (PBMC) from female and male individuals equally processed using the PicoPLEX Gold kit. **Bottom Panel.** Single cells from various cell lines (GM22601, GM05067, and GM12878) were amplified using a prototype of the PicoPLEX WGAv2 kit. 1 ng of the purified product was used as input for a Nextera XT kit and sequenced on an Illumina MiSeq platform using V3 chemistry at a read length of 2 x 75 bp. FASTQ files were trimmed to remove adapters and then aligned to the human genome assembly GRCh37. Only autosomes are reported. For both panels, alignment was normalized to 1 million reads, and the number of reads per bin of 1 Mb was calculated using bedtools 2.25.0. The $\log_2$ ratio of the bin counts (sample/reference) was plotted using the Integrative Genomics Viewer.

5 Best-in-class reproducibility

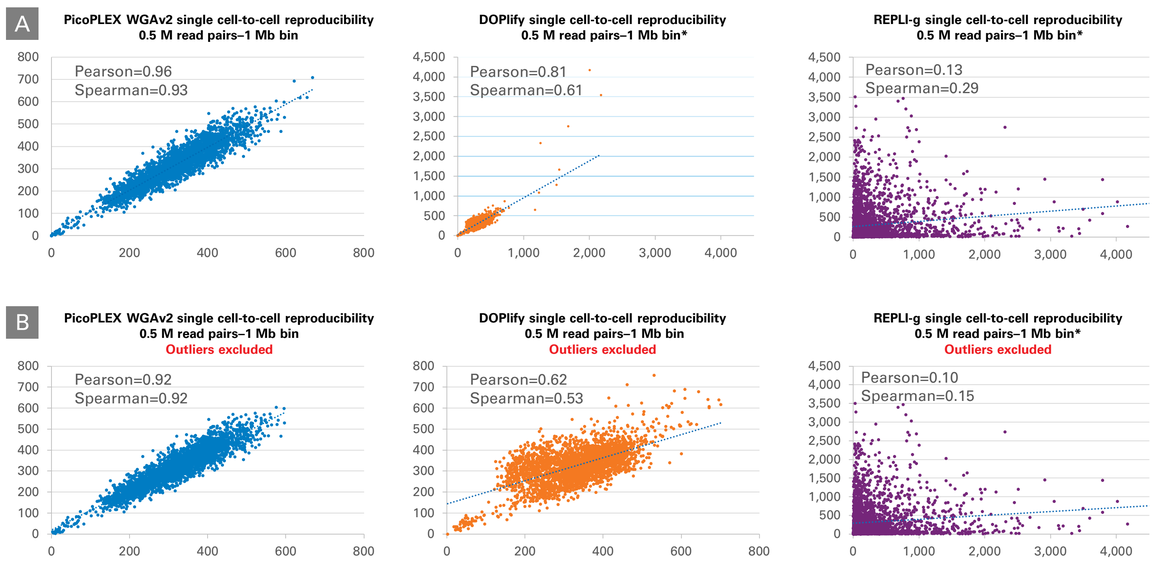

Figure 5. Whole genome coverage reproducibility of the prototype PicoPLEX WGAv2 kit in comparison to PerkinElmer DOPlify and QIAGEN REPLI-g (MDA) kits. Whole genome amplification (WGA) products were prepared from single-cell samples of GM12878 in replicates, using a prototype of PicoPLEX WGAv2, DOPlify, and REPLI-g (MDA) kits. One ng of amplified product was used as input for a Nextera XT kit and the resulting libraries were sequenced on an Illumina MiSeq platform using V3 chemistry and a read length of 2 x 75 bp. After read alignment to human genome assembly GRCh37 and normalization to 1 million reads (0.5 million read pairs), the number of reads per bin of 1 Mb was calculated using bedtools 2.25.0. Total reads in each window from two single-cell libraries were plotted, with Pearson and Spearman correlations calculated and indicated on each graph. **Panel A.** Samples containing outliers. *Due to the different biases of the technologies compared, the graphs have different scales since no correlation points have been excluded. **Panel B.** Samples with outliers excluded from calculations. Results from the PicoPLEX WGAv2 kit show high reproducibility and robust coverage compared to the DOPlify and REPLI-g kits, which provides a clear advantage for the detection of structural variants (CNVs).

Conclusions

- The newly launched PicoPLEX Gold and PicoPLEX WGAv2 kits show improved performance for the detection of CNVs and SNVs
- Comparison of the PicoPLEX Gold and QIAGEN REPLI-g library-prep systems shows better reproducibility and superior CNV and SNV detection with the PicoPLEX Gold kit
- Comparison of the PicoPLEX WGAv2 kit to QIAGEN REPLI-g and PerkinElmer DOPlify systems shows superior reproducibility and accuracy for SNV detection with the PicoPLEX WGAv2 kit
- Both PicoPLEX Gold and PicoPLEX WGAv2 kits demonstrate the capability to detect segmental aneuploidies to a resolution of 25 Mb with as few as 1 million reads
- Both PicoPLEX Gold and PicoPLEX WGAv2 kits employ the gold-standard PicoPLEX technology that enables fast, accurate, and reproducible readouts

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis