Highly reproducible and accurate single cell whole-genome amplification using next-generation SMARTer PicoPLEX technology

Accurate, reproducible detection of genomic variants such as single nucleotide variants (SNVs) and copy number variants (CNVs) from small amounts of DNA, single cells, or fixed tissue is critical for genetic analysis of clinical samples, with the broader goal of assisting molecular diagnosis of diseases such as cancer (Wang and Navin 2015). To this end, we optimized our PicoPLEX technology to develop the SMARTer PicoPLEX Gold Single Cell DNA-Seq Kit (PicoPLEX Gold) which enables highly reproducible CNV and SNV detection from 1 to 5 cells or small amounts of purified genomic DNA (gDNA), and is an excellent tool for research in cancer genomics, reproductive health, developmental biology, and other related fields.

The first generation of PicoPLEX technology, namely the original SMARTer PicoPLEX WGA Kit (PicoPLEX WGA), was optimized for the reproducible detection of aneuploidies and CNVs in single cells (Zhang et al. 2017; Deleye et al. 2017; Babayan et al. 2017; Biezuner et al. 2017). To enable accurate detection of SNVs, we revamped the PicoPLEX chemistry using optimized enzymes, primers, and protocols that improve sequencing coverage, uniformity, and the accuracy of genomic variant detection. The second generation, PicoPLEX Gold kit has significantly improved genome coverage and fidelity, which expands the utility and scope of applications of this technology, such as the analysis of SNVs, indels, and other small structural variants from single cells. These technological enhancements also improve the sample-to-sample reproducibility and hence the resolution of CNV detection.

Introduction

Principle and workflow

The PicoPLEX Gold kit features a streamlined workflow which involves four steps, from input to amplified NGS libraries, and can be completed in less than three hours. The kit is based on our patented PicoPLEX technology for single cell whole-genome amplification and consists of high-fidelity DNA polymerases and optimized primers. PicoPLEX Gold is compatible with single cells (unfixed or fixed) and purified human genomic DNA.
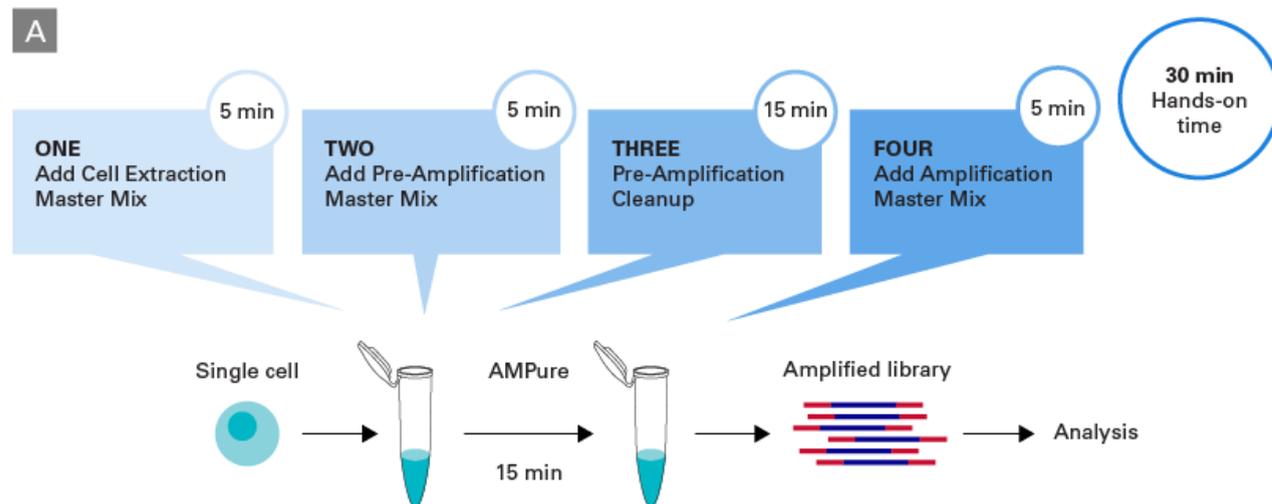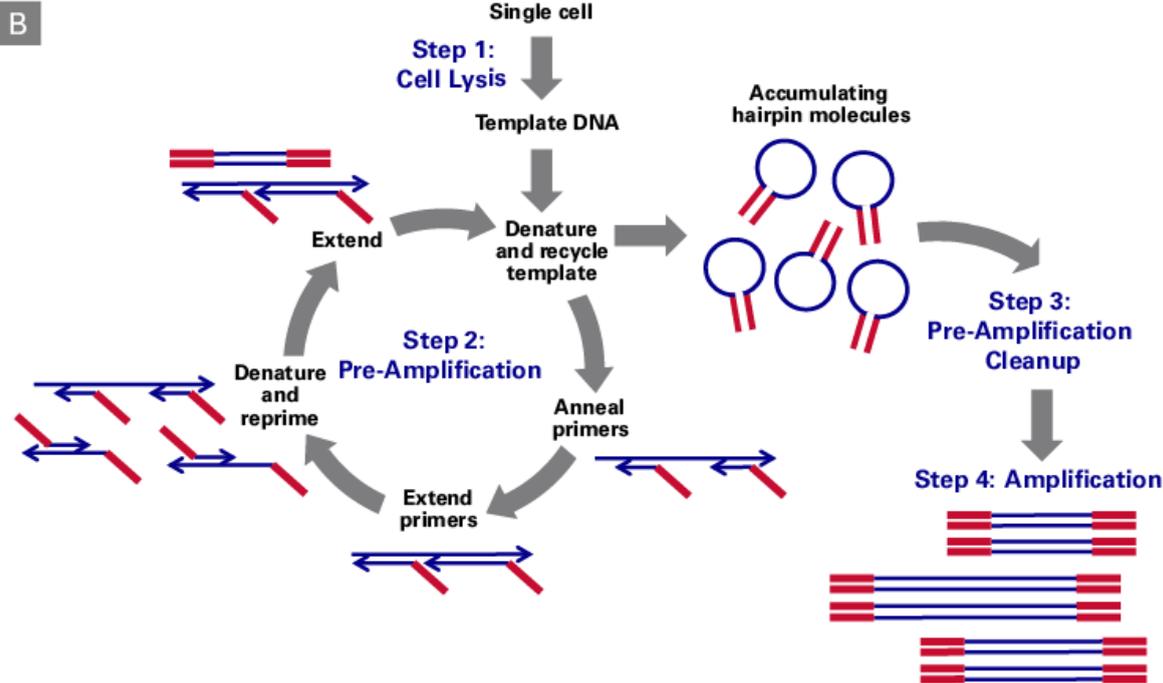
Figure 1. SMARTer PicoPLEX Gold Single Cell DNA-Seq technology. Panel A. Schematic depicting the simple, four-step PicoPLEX Gold workflow with minimum hands-on time. **Panel B.** Schematic illustrating the PicoPLEX Gold chemistry. Cellular gDNA extracted in Step 1 is used as the template for multiple cycles of quasi-random priming and linear amplification followed by exponential library amplification.

Results

Library characteristics

PicoPLEX Gold uses multiple rounds of template re-priming to generate micrograms of amplified DNA. The success of amplification from single cells is effectively 100%, reducing the failure rate to a minimum. The consistency of amplification is evident from Figure 2 where amplification curves for triplicates of five cells and single cells are depicted in purple and blue respectively. The lack of amplification of no template controls (NTC) indicates a high level of purity of the reagents and the sensitivity of the chemistry.
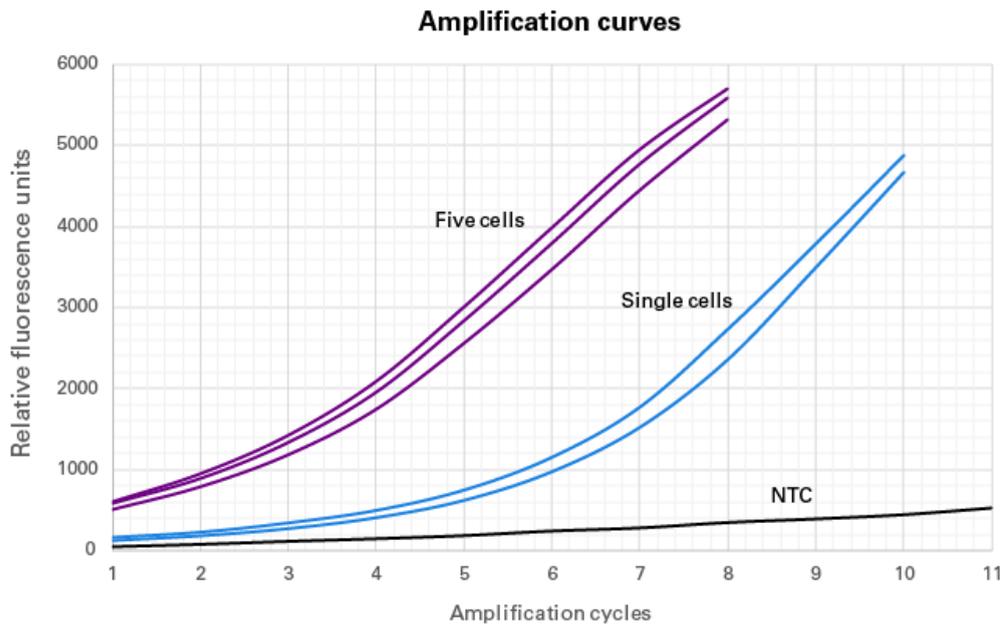
Figure 2. Real-time analysis of library amplification using the PicoPLEX Gold kit. A typical real-time amplification analysis of libraries prepared with PicoPLEX Gold kit using triplicates of single (blue) and five (purple) GM12878 cells, relative to the NTC (grey). Results were obtained using the CFX96 Touch Real-Time PCR Detection System with EvaGreen and fluorescein as the dyes.

Improvements over the first generation of PicoPLEX technology

The performance of the two generations of PicoPLEX products was evaluated side by side. Libraries were prepared from 15 pg of gDNA (NA12878) using the PicoPLEX Gold Kit or SMARTer PicoPLEX WGA Kit (PicoPLEX WGA) and sequenced on an Illumina® NextSeq® platform to a depth of ~35 million read pairs (PE 2 x 150 bp). As depicted in Table I, the genome coverage was improved when using PicoPLEX Gold (50%) compared to PicoPLEX WGA (33%). Additionally, the duplication rate was significantly reduced in PicoPLEX Gold data (9%) versus PicoPLEX WGA (21%). The use of high fidelity polymerases in the PicoPLEX Gold Kit reduced the allele drop-in to a minimum, allowing higher confidence in SNV detection (Figure 4D, below).

	SMARTer PicoPLEX WGA Kit	SMARTer PicoPLEX Gold Single-cell DNA-seq Kit
Genome Coverage (>1X)	~33%	~50%
Duplication rate	~21%	~9%
Fidelity	High	Very high
Reproducibility	Very high	Very high
Workflow	Single-tube, 3-step 2.5 hr workflow. Needs Nextera® library prep.	4-step 3 hr workflow. Direct NGS library.

Table I. Improvements in PicoPLEX Gold in comparison to the first generation SMARTer PicoPLEX WGA kit. Libraries were prepared using 15 pg of (NA12878) gDNA and sequenced to a depth of ~35 million read pairs (PE 2 x 150 bp). PicoPLEX Gold has an improved coverage, lower duplication rate, higher fidelity, and best-in-class sample-to-sample reproducibility.

Genome coverage, uniformity, and reproducibility

We measured coverage depth, uniformity, and reproducibility of PicoPLEX Gold in comparison to QIAseq FX Multiple Displacement Amplification (MDA) technology for two individual single cells. The coverage of PicoPLEX Gold was similar to QIAseq FX (MDA) at lower depths and greater at higher depths (Figure 3A). Notably, PicoPLEX Gold has a highly uniform coverage pattern that is considerably better than that of QIAseq FX (Figure 3B). The reproducibility of coverage between two single cells was significantly higher for PicoPLEX Gold, which provides a clear advantage for the detection of structural variants (Figure 3C).
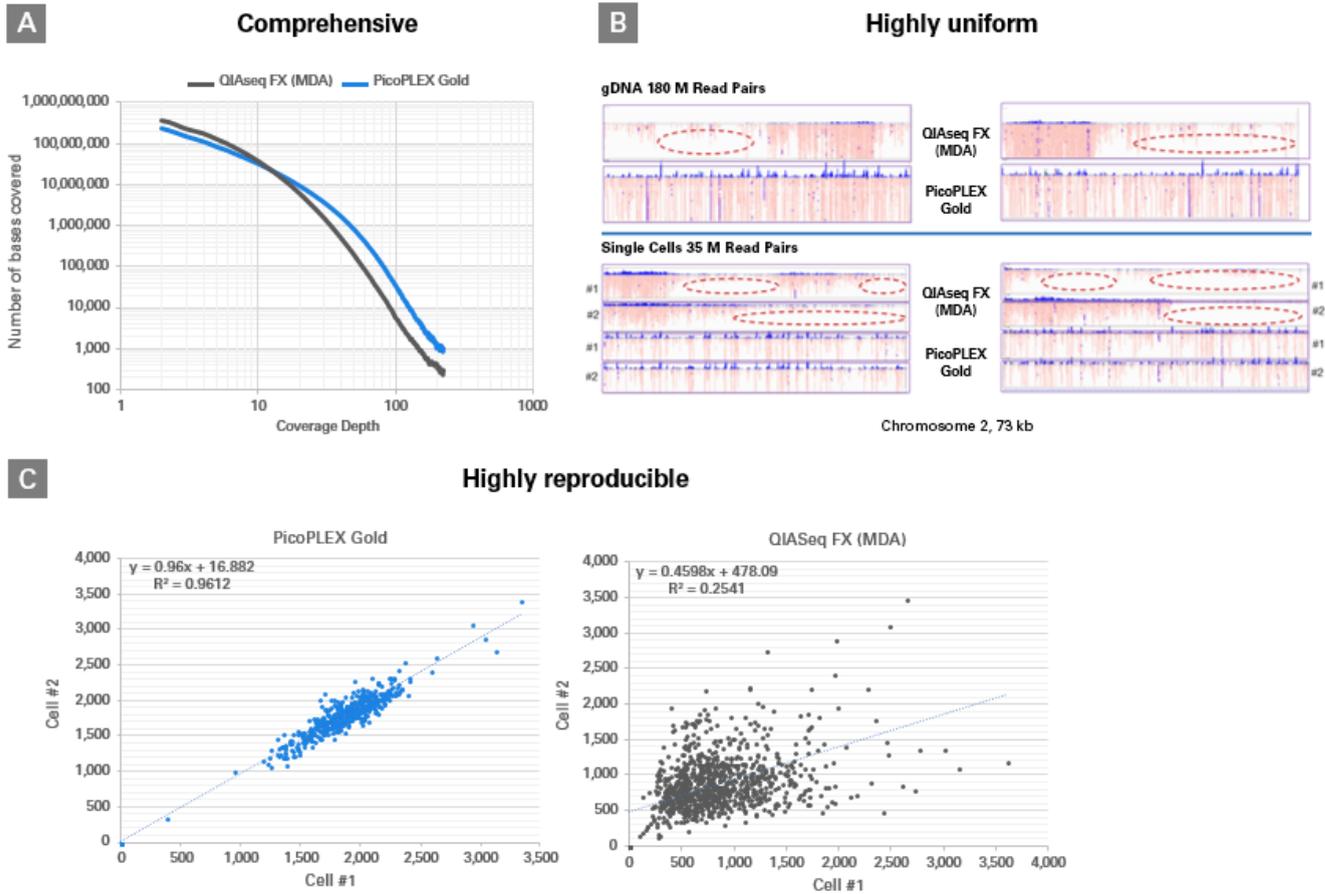
Figure 3. Key features of PicoPLEX Gold: comprehensive coverage, high uniformity, and reproducibility (in comparison to QIAseq FX). Panel A. A log-log plot showing the number of bases covered at various depths of sequencing (~35M read pairs, PE 2 x 150 bp). The coverage of PicoPLEX Gold was similar to QIAseq FX (MDA) at lower depths and greater at higher depths. **Panel B.** Examples of the coverage patterns of PicoPLEX Gold and QIAseq FX in gDNA (NA12878) and single-cell samples (GM12878) for a 75-kb window (chr2). MDA has a higher propensity to leave large gaps in the genome, whereas PicoPLEX Gold has a more uniform coverage. **Panel C.** The reproducibility of coverage evaluated by comparing total reads in 100-kb bins. The consistency of the total reads in each window from the two single-cell libraries is significantly higher for PicoPLEX Gold (left), both in comparison to QIAseq FX (right) and to other technologies in the market (data not shown).

Improved recovery and accuracy of SNV detection

The two key artifacts of single-cell WGA are allele drop-out (ADO; false negatives) and allele drop-in (ADI; false positives). In addition to the high recovery rate of SNVs, PicoPLEX Gold has superior performance in both ADO and ADI rates. We evaluated the performance of PicoPLEX Gold using GM12878 single and five cells sequenced to a depth of ~35 million read pairs (PE 2 x 150 bp; deduplicated) and benchmarked the error rates using the NIST Genome in a Bottle data set. SNVs were generated using the standard GATK pipeline and filtered stringently (10X minimum depth and GATK quality score of 75 or higher). The total number of SNVs detected by PicoPLEX Gold was 3.5X higher for single cells and 9X for five cells compared to QIAseq FX (MDA) (Figure 4A). The improved allele balance (Figure 4B) leads to a significantly reduced allele drop out, which is up to 5-fold lower compared to QIAseq FX (MDA) (Figure 4C). The high-fidelity polymerases used in PicoPLEX Gold resulted in significantly lower false-positive rates (Figure 4D).
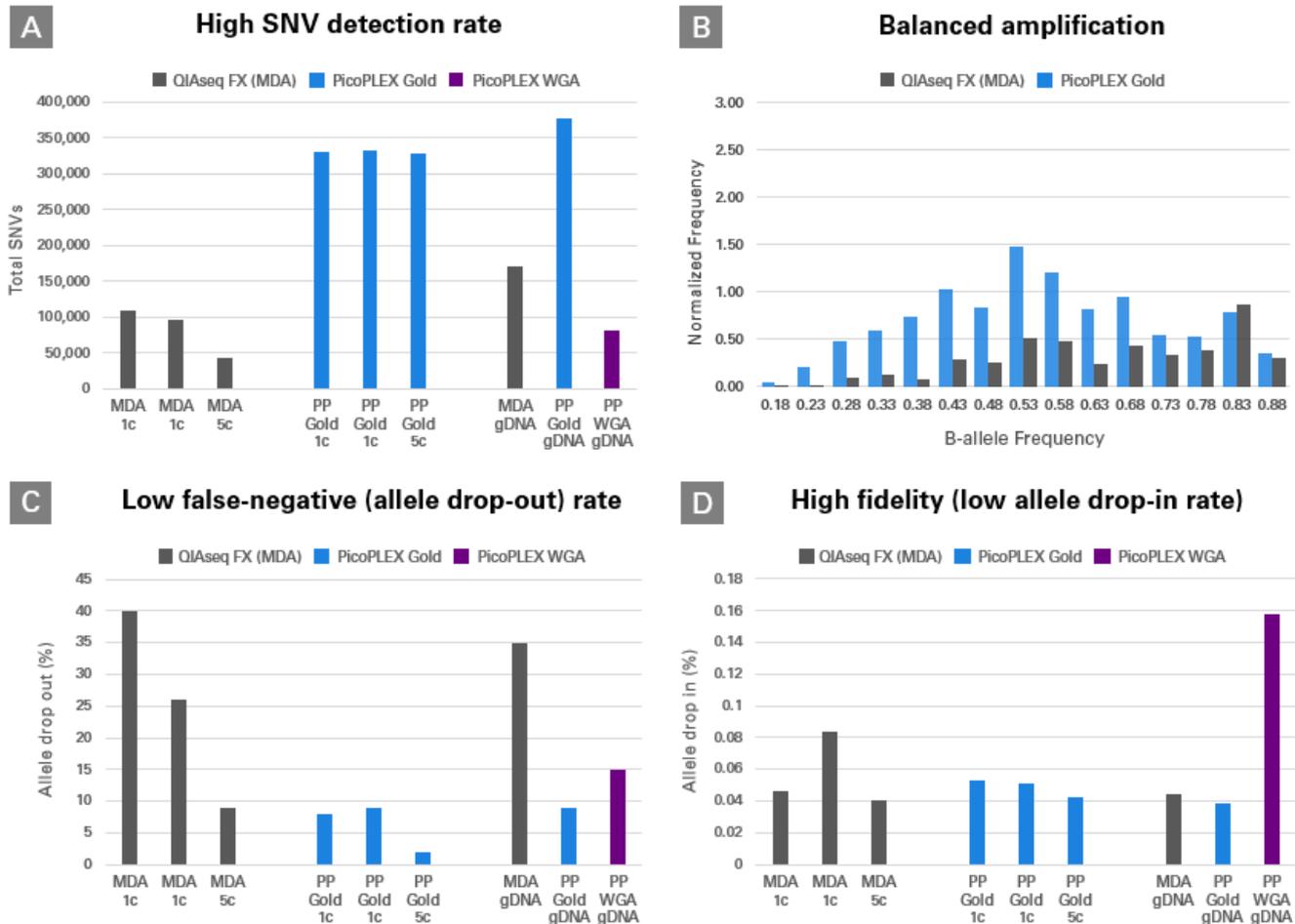
Figure 4. High-quality single nucleotide variant (SNV) detection with PicoPLEX Gold. Panel A. Comparison of SNV-detection rate between QIAseq FX (MDA), PicoPLEX Gold (PP Gold), and SMARTer PicoPLEX WGA (PP WGA) kits from one cell (1c), five cells (5c) or 15 pg of NA12878 gDNA inputs. Single and five cells were sequenced to a depth of ~35M read pairs, and gDNA samples to a depth of ~40M read pairs. The high fidelity and robust coverage of PicoPLEX Gold (blue bars) provide a clear advantage in detecting a greater (~2–9 fold) number of high-quality SNVs compared to QIAseq FX (gray bars) and PicoPLEX WGA (purple bar). **Panel B.** The symmetric distribution of the B-allele frequencies for PicoPLEX Gold (blue bars), centered around 0.5, indicating a balanced recovery of both alleles. PicoPLEX has better allele balance compared to QIAseq FX (MDA) (gray bars). **Panel C.** Unbiased amplification of PicoPLEX Gold results in the lowest allele drop-out (false-negative) rates among all single cell library-preparation technologies tested. **Panel D.** High fidelity of the polymerases used in PicoPLEX Gold kit (blue bars) leads to minimal allele drop-in rates that are comparable to QIAseq FX (gray bars) and significantly lower than PicoPLEX WGA (purple bar).

CNV detection from a single cell with low-pass sequencing

SMARTer PicoPLEX technology has been the gold standard for detecting aneuploidies in single cells, largely due to its unparalleled sample-to-sample reproducibility. PicoPLEX Gold builds upon the advantages of legacy products by improving the resolution of CNV detection and enables the identification of small structural variants. The superior CNV detection of PicoPLEX Gold was demonstrated using two single NCI-H929 cells, which are known to have stable aneuploidies of various sizes. We evaluated the ability of the kit to consistently detect CNVs at various depths of sequencing in comparison to a bulk sample sequenced at a higher depth. The Log2 ratios of the total number of reads in 50 kb bins from NCI-H929 cells (sample) to that of GM12878 cells (euploid reference) were plotted (Figure 5). PicoPLEX Gold detected the same small aneuploidies (100–500 kb in size) at different read depths: 17.5 million, 8.5 million, and even 2.5 million read pairs. The consistent detection of these structural variants in two biological replicates (Figure 5A and 5B) demonstrates a high degree of sample-to-sample reproducibility.
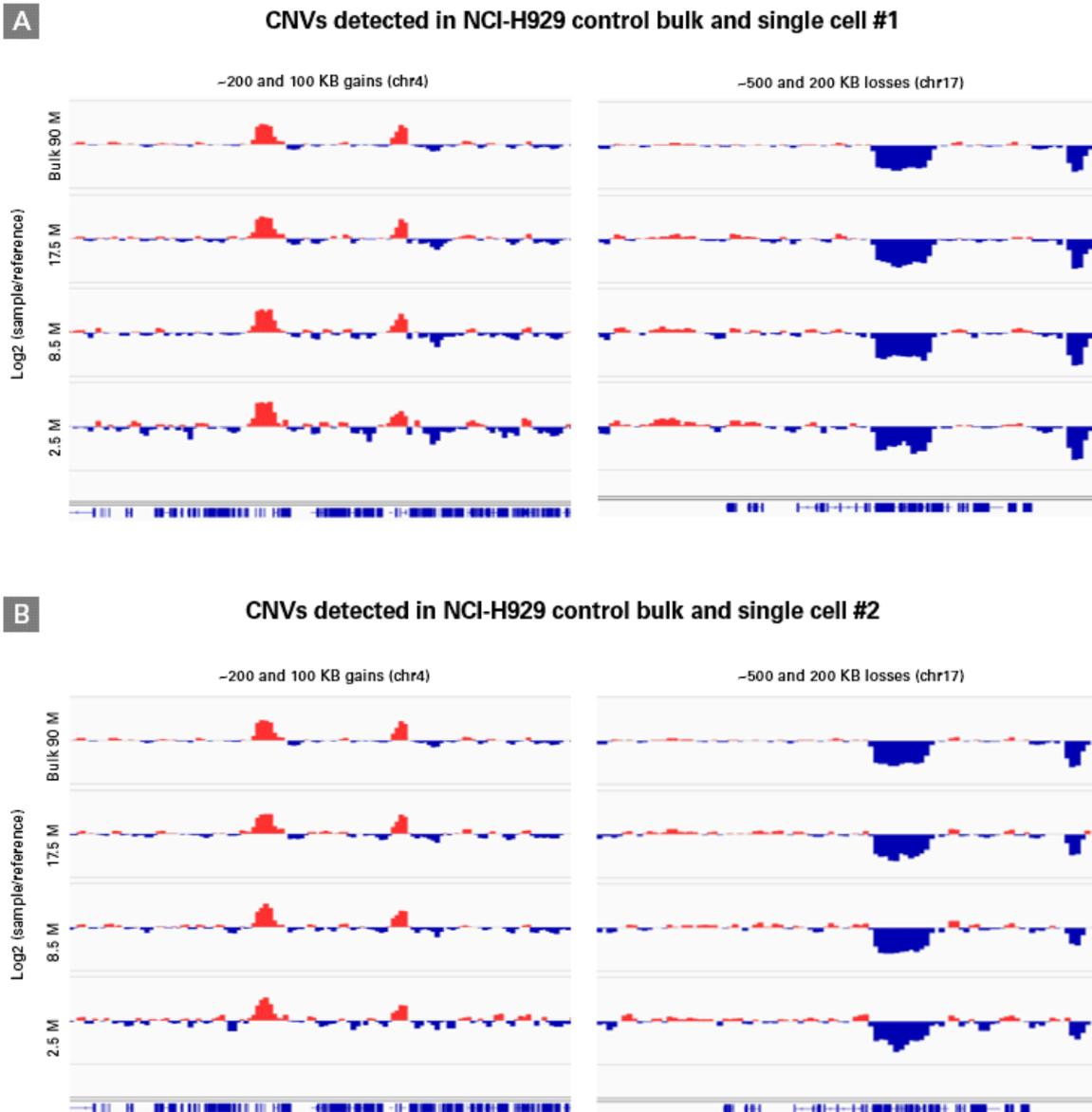
Figure 5. CNVs detected in two individual cells using the SMARTer PicoPLEX Gold Single Cell DNA-seq Kit. Log_2 ratio of the total number of reads in 50-kb bins from single NCI-H929 cells, shown as one cell in Panel A and a second cell in Panel B. Red bars represent copy-number gains while blue bars represent losses. The top row of the graphs in each panel depicts the control bulk sample sequenced to a depth of 90 million read pairs. The highly reproducible coverage of the SMARTer PicoPLEX Gold kit enables the accurate detection of structural variants as small as 100 kb, even at shallow sequencing depths (2.5–8.5 million read pairs).

Conclusions

In summary, PicoPLEX Gold has best-in-class sample-to-sample reproducibility, significantly improved breadth of coverage, and very high fidelity of amplification. Single-cell libraries generated using PicoPLEX gold are ideal for high-resolution CNV detection and accurate SNV detection. These enhancements make PicoPLEX Gold the preferred technology for several single-cell genomics research applications, such as detecting aneuploidies in embryo biopsies, characterizing the heterogeneity and tumor evolution of cancer tissues, and profiling circulating tumor and immune cells.

Methods

Sample preparation

GM12878 cells were sourced from the Coriell Institute, stained with CD81-FITC antibody, and flow sorted using a BD FACSJazz instrument. NCI-H929 cells were obtained from ATCC and processed similarly. Libraries were prepared according to the SMARTer PicoPLEX Gold Single Cell DNA-Seq Kit user manual and sequenced on an Illumina NextSeq (PE 2 x 150 bp).

Bioinformatic analysis

FASTQ reads were trimmed to remove the primer sequence from the 5' end of the read. Trimmed reads were aligned using BWA (default parameters). Single nucleotide variants were generated using GATK (according to its best-practices guidelines, found at https://software.broadinstitute.org/gatk/best-practices/) and filtered at a minimum depth of 10X, with a minimum quality score of 75. Allele drop-out rates were calculated as described in Leung et al. 2015. CNVs were generated using CNV-seq (Xie and Tammi, 2009). Briefly, normalized counts in 50 kb bins from H929 cells were compared to GM12878 cells (euploid reference) to detect CNVs.

References/citations

Babayan, A. *et al.* Comparative study of whole genome amplification and next-generation sequencing performance of single cancer cells. *Oncotarget* **8,** 56066–56080 (2017).

Biezuner, T. *et al.* Comparison of seven single cell Whole Genome Amplification commercial kits using targeted sequencing. *BioRxiv* **186940,** (2017). doi:10.1101/186940.

Deleye, L. *et al.* Performance of four modern whole genome amplification methods for copy number variant detection in single cells. *Sci. Rep.* **7,** 3422 (2017).

Leung, M.L, *et al.* SNES: single nucleus exome sequencing. *Gen. Biol.* **16,** 55 (2015).

Wang, Y. & Navin, N. E. Advances and applications of single-cell sequencing technologies. *Mol. Cell* **58,** 598–609 (2015).

Xie, C. & Tammi, M. T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* **10,** 80 (2009).

Zhang, X. *et al.* The comparison of the performance of four whole genome amplification kits on ion proton platform in copy number variation detection. *Biosci. Rep.* **37,** BSR20170252 (2017).

Related Products

Cat. #	Product	Size	License	Quantity	Details
R300669	PicoPLEX® Gold Single Cell DNA-seq Kit	24 Rxns		*	

The SMARTer PicoPLEX Gold Single Cell DNA-Seq Kit is designed to generate high-quality DNA libraries from single cells for sequencing on Illumina platforms. The kit is based on innovative PicoPLEX technology and yields libraries with superior reproducibility from 1–5 mammalian cells or genome equivalents of gDNA as input, providing reliable results from challenging samples such as formalin-fixed single cells. Single index, dual index, and unique dual index kits are available and must be purchased separately. This product contains reagents for 24 reactions.

The kit features a streamlined workflow consisting of four steps—from input to amplified NGS libraries—that can be completed in less than four hours. The kit uses high-fidelity enzymes and optimized primers and is suitable for accurate and reproducible profiling of copy number variants (CNVs), single-nucleotide variants (SNVs), indels, and other small structural variants. The kit is compatible with single cells and purified DNA from humans, animals, and bacteria, as well as single flow-sorted chromosomes.

Documents	Components	Image Data

Cat. #	Product	Size	License	Quantity	Details
R300670	PicoPLEX® Gold Single Cell DNA-seq Kit	96 Rxns		*	
R300698	PicoPLEX® Gold Single Cell DNA-Seq Kit	384 Rxns		*	

Add to Cart

Takara Bio USA, Inc.