SMARTer PicoPLEX Gold: A New Generation of Single Cell NGS Library with High Reproducibility, and Greatly Improved Coverage and Fidelity for Precision Medicine

Fang Sun*, Datta Mellacheruvu, Bayu Sisay, John Langmore, Graeme McLean, Andrew Farmer, Emmanuel Kamberov

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

Introduction

Accurate, reproducible detection of mutations and copy number variations (CNV) from small amounts of DNA, including single cells and fixed tissue, is key for genetic analysis of clinical samples to assist in identifying the best treatment regimen and molecular diagnoses of diseases such as cancer. Increasing demand for genetic analysis from limited samples, including single cells, has created an increasingly unmet need for technologies that allow for economical and accurate analysis of those samples.

A primary application for single-cell analysis is Preimplantation Genetic Testing (PGT). PGT has grown dramatically in the last ten years, enabled by improvements in the array and sequencing platforms, as well as by the patented SMARTer® PicoPLEX® WGA (PicoPLEX) quasi-random priming technology—the international gold standard for whole genome amplification (WGA) for subsequent detection of CNV in fixed or unfixed single cells. Initially, PicoPLEX chemistry was optimized to allow for reproducible detection of aneuploidies and CNVs in embryo biopsies. The original versions of the technology are not optimized for other applications such as genetic analysis in cancer screening, diagnosis of disease, or therapeutic drug monitoring due to the frequency of false-positive mutation rates.

To address the need for accurate detection of single nucleotide variants (SNVs), we enhanced the PicoPLEX chemistry using optimized enzymes, primers, and protocols that improve sequencing coverage, uniformity, and accuracy while increasing the resolution for CNV detection and retaining reproducibility. This enhanced chemistry named SMARTer PicoPLEX Gold Single Cell DNA-Seq Kit (PicoPLEX Gold) is a single-cell library-prep kit with a simple, four-step protocol to convert fixed or unfixed single cells into NGS libraries in under three hours with minimum hands-on-time.

Libraries prepared from single GM12878 cells using the PicoPLEX Gold kit were sequenced on an Illumina® NextSeq® platform to a depth of ~35 million read pairs (2 x 150 cycles), generating >50% genome coverage. This coverage represents a 2-fold improvement over the original PicoPLEX kit, along with a 4X reduction in duplication rates. The PicoPLEX Gold kit detected 3.5X more SNVs compared to Multiple Displacement Amplification (MDA) with the same number of reads. Our proprietary high-fidelity polymerases used in the PicoPLEX Gold kit produced up to 50% lower allele drop-in (false-positive) rates than MDA. The increased coverage and low bias of the PicoPLEX Gold kit translated to extremely low allele-dropout rates (ADO), ~5X lower than MDA. Therefore, a single PicoPLEX Gold kit library enables reliable, high-resolution CNV analysis with shallow sequencing, and an accurate and reproducible SNV and CNV analysis with deeper sequencing.

1 PicoPLEX Gold technology—principle and workflow

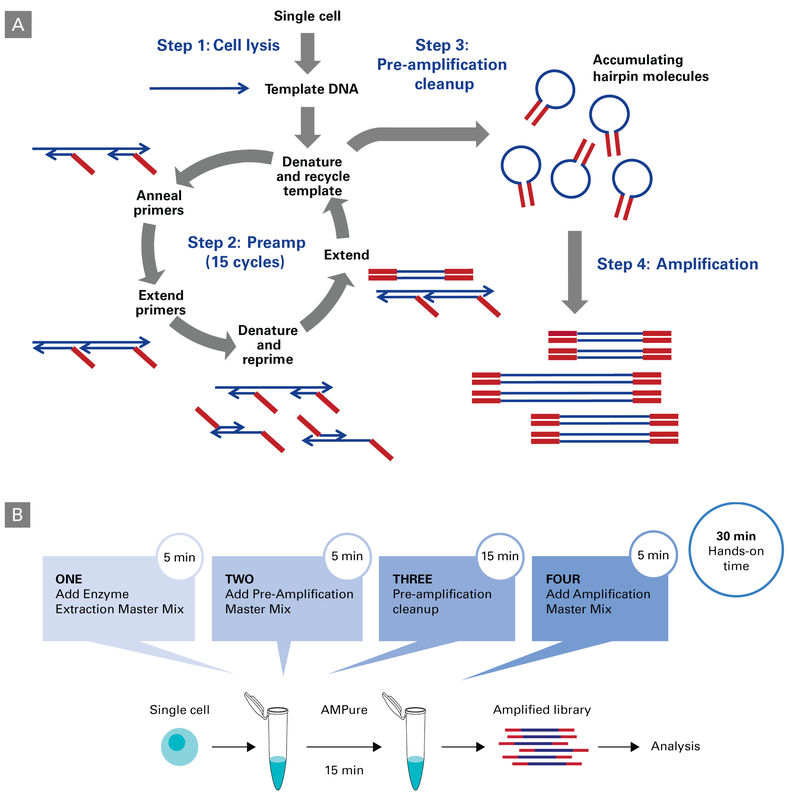

Figure 1. An overview of the principle of PicoPLEX technology and workflow schematic. Panel A. Step 1: A single cell is lysed and the DNA is released free of proteins. Step 2: Multiple rounds of quasi-linear amplification of the single-cell gDNA. The formation of hairpins prevents subsequent amplification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

2 Greatly improved genome coverage and superior reproducibility between two single cells using PicoPLEX Gold

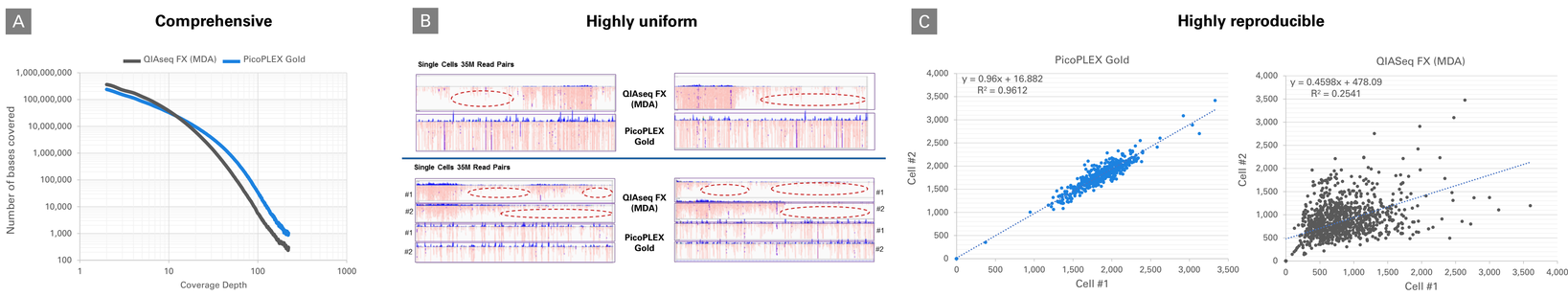

Figure 2: Coverage depth, uniformity and reproducibility of PicoPLEX Gold in comparison to QIAseq FX. Panel A. A log-log plot of the number of bases covered at various depths of sequencing is shown. The coverage of PicoPLEX Gold is similar to QIAseq FX (MDA) at lower depths and greater at higher depths. **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) is shown. As evident from this example, the coverage of PicoPLEX Gold is highly uniform and significantly better than that of QIAseq FX (MDA). **Panel C.** The reproducibility of coverage was evaluated in 100-KB bin sizes. Total reads in each window from the two single-cell libraries were plotted. PicoPLEX Gold shows high reproducibility, which provides a clear advantage for the detection of structural variants (CNVs). In summary, PicoPLEX Gold has far superior and robust coverage compared to QIAseq FX.

3 Low allele drop-out (false-negative) and drop-in (false-positive) rates enable sensitive SNV detection from single cells or five cells

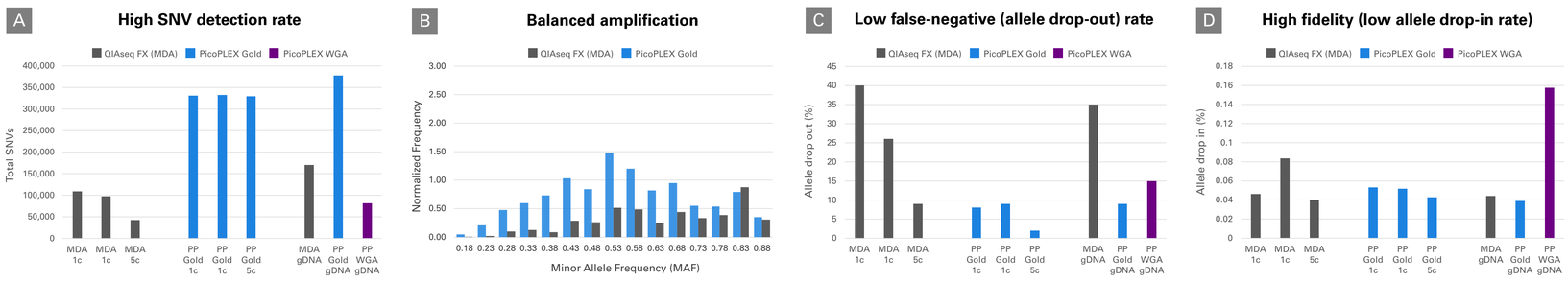

Figure 3: Summary of the ability of PicoPLEX Gold to generate high-quality single nucleotide variants (SNVs). Data were generated using the standard GATK pipeline and filtered at 10X depth and Q75 minimum quality using GM12878 single cells, five cells, or NA12878 gDNA libraries and sequenced to a depth of 35 million read pairs. The results were benchmarked using gold-standard (NIST) bulk-sequencing data from the same cell line. **Panel A.** The high fidelity and robust coverage of PicoPLEX Gold (PP Gold; blue bars) provide a clear advantage in detecting a greater (~3X) number of high-quality SNVs compared to QIAseq FX (MDA; gray bars). The results for gDNA samples were also compared to those obtained using the SMARTer PicoPLEX WGA Kit (PP WGA; purple bars). **Panel B.** The symmetric distribution of the minor allele frequencies (MAF) for PicoPLEX Gold (blue bars) and a peak at 0.53, indicates a balanced recovery of both alleles, clearly outperforming QIAseq FX (MDA)(gray bars). **Panel C.** Unbiased amplification of PicoPLEX Gold generates 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 compared to QIAseq FX (MDA; gray bars).

4 High-resolution and highly reproducible CNV detection from a single cell with low-pass sequencing

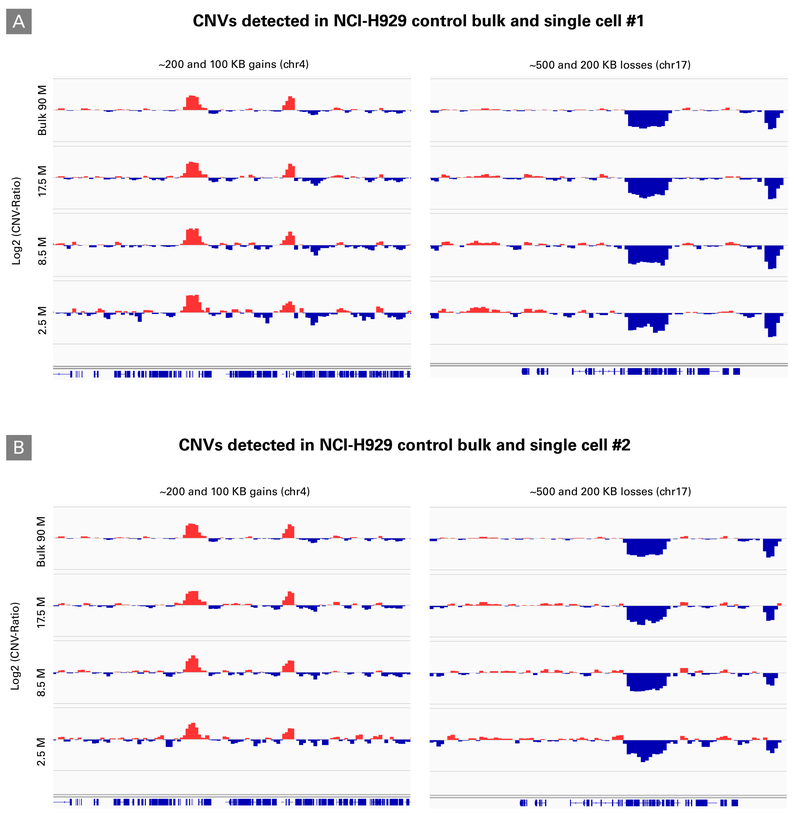

Figure 4: CNVs detected in single cells from a multiple myeloma cell line, NCI-H929, normalized to a reference cell line, 12878. Panels A and B. Two individual single cells (NCI-H929) were amplified with PicoPLEX Gold and analyzed at different read depths for detecting CNVs. The Log2 ratio of the total number of reads in 50 KB bins from H929 and 12828 cells were plotted using CNV-seq. Red bars represent copy-number gains (left panels) while blue bars represent losses (right panels). The highly reproducible coverage of PicoPLEX Gold enables it to retain its ability to accurately detect structural variants as small as 100 KB even at relatively shallow sequencing (2.5–8.5 million read pairs).

5 Target enrichment facilitates significantly higher coverage and SNV recovery rate from single-cell libraries

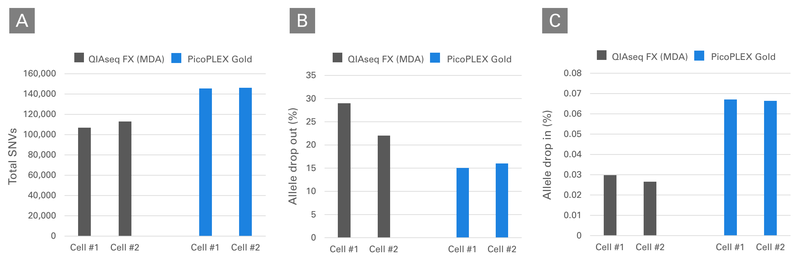

Figure 5: Whole exome target capture in single cells (GM12878) amplified using PicoPLEX Gold and QIAseq FX. Panel A. A greater number of SNVs was detected accurately using PicoPLEX Gold compared to QIAseq FX (MDA) for two individual single cells. **Panel B:** The allele drop-out (false-negative) rate was lower for PicoPLEX Gold compared to QIAseq FX. The data generated with PicoPLEX Gold enabled a higher level of confidence in SNV detection. **Panel C.** The level of allele drop-in (false-positive), was higher for PicoPLEX Gold but in an acceptable range (less than 0.1%).

Conclusions

Metric	SMARTer PicoPLEX Gold	QIAseq FX (MDA)
Coverage (singe cell)	**Good**	Good
Reproducibility (single cell)	**Very high**	Low
Uniformity (single cell)	**High**	Low
Multiple rounds of template utilization	**Yes**	No
Workflow (total time to sequencing)	**~3 hr**	~5 hr
Workflow (number of steps)	**4**	6
Direct NGS prep (Illumina, Ion)	**Yes**	No
Failure with single cells	**Very low**	Up to 50%
Ability to process (Formalin) fixed cells	**Good**	Poor
Single nucleotide detection	**High**	Low
Fidelity (of amplification)	**Very good**	Very good
Allele drop-out (false-positive) rate	**Low**	High
Allele drop-in (false-negative) rate	**Low**	Low
Reproducible variant detection	**Yes**	No
High-resolution structural variant detection	**Yes**	No

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

To view more data visit
www.takarabio.com/PicoPLEX_Gold

800.662.2566
Visit us at www.takarabio.com

Clontech TaKaRa cellartis