Demystifying tumor heterogeneity with a fully automated, high-throughput single-cell DNA-seq (scDNA-seq) workflow

Xuan Li*, Mike Covington, Hima Anbunathan, Raymond Mendoza, Samantha Leong, Shuwen Chen, Bryan Bell, Mohammad Fallahi, Takara Bio Engineering, Patricio Espinoza, Yue Yun, Andrew Farmer

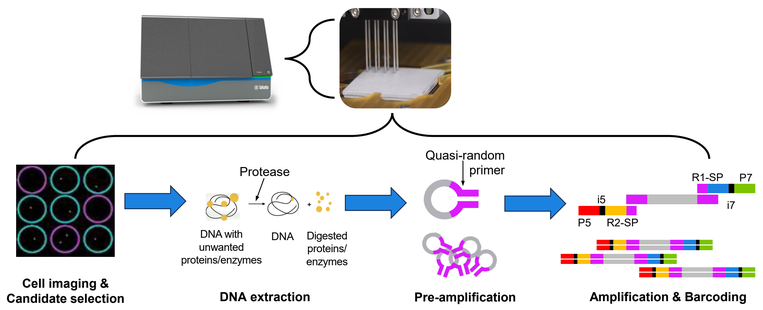
Takara Bio USA, Inc., 2560 Orchard Pkwy, San Jose, CA 95131 USA *Corresponding Author

Introduction

Genetic heterogeneity is a key factor underlying tumor drug resistance and metastatic potential. Understanding this heterogeneity is therefore vital to improving both prognosis and treatment. Next-generation sequencing is a valuable tool for analyzing the genetic makeup of tumors. However, bulk sequencing methods lack the sensitivity to fully resolve tumor heterogeneity. While single-cell methods provide a powerful approach to dissect such heterogeneity, to date, these methods have been limited in their throughput. To address this bottleneck, we have developed a fully automated workflow for generating scDNA-seq libraries based on PicoPLEX® whole-genome amplification (WGA) technology. This high-throughput method, which has been optimized on our Shasta™ Single-Cell System, enables the generation of WGA libraries for >1,000 single cells within one day.

Methods

We benchmarked the new high-throughput method to generate WGA libraries against the standard PicoPLEX workflow for genome coverage, GC bias, and other typical quality metrics. We additionally assessed copy number variant (CNV) detection sensitivity using two cell lines with well-characterized small segmental CNVs: GM22601 (~25 Mb deletion on chromosome 4) and GM05067 (~45 Mb gain on chromosome 9). We also analyzed GM12878 as a control cell line for CNV calling and a lymphoblastoid line (K562) that carries a range of chromosomal aneuploidies. A total of 1,288 single-cell WGA libraries were generated and sequenced to a depth of 250,000 paired-end reads per cell. Data analysis was done using the Ginkgo CNV pipeline with an average bin size of 500 Kb. As a final proof of principle, we generated single-cell data using cells dissociated from tumor and tumor-adjacent tissue from Stage I and Stage III clear cell renal cell carcinoma (ccRCC) samples purchased from BioIVT.



Shasta WGA workflow. Hoechst-stained single cells are dispensed into a 5,184-nanowell chip and screened by imaging. Reagents are deposited into nanowells via a series of dispenses: a lysis mix to release DNA from nuclei, a pre-amplification mix during which quasi-random primers amplify random locations of the genome, a PCR mix, and two indexing primer dispenses to amplify the libraries while simultaneously incorporating a unique barcode for each single cell. The pooled barcoded libraries are ready for Illumina® sequencing after off-chip purification.

Results

Libraries generated using our high-throughput workflow had a high mapping rate, with 94.1% of the reads being uniquely mapped. Additionally, the libraries were comparable to those generated with the standard PicoPLEX workflow in terms of coverage uniformity, GC bias, and other metrics. Moreover, the segmental aneuploidies in both GM22601 and GM05067 were reliably detected in >90% of cells at a read depth as low as 250,000 reads per cell. Analysis of the ccRCC samples revealed subclonal heterogeneity with various CNVs common to ccRCC, including deletion of chr. 3p, amplification of chr. 5q, and duplication of chr. 2.

1 Primary sequencing quality metrics of the WGA library of single cells from four cell lines generated using the Shasta WGA kit

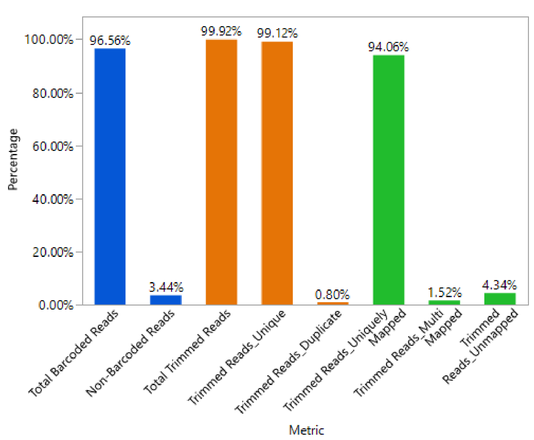


Figure 1. QC metrics of Shasta WGA libraries generated from single cells from four different cell lines. Library sequencing was performed using a NextSeq® 500/550 High Output Kit with a 150-cycle cartridge (read length 2 x 75 bp) with a PhiX spike-in percentage of 18.81%. 96.56% of the reads were successfully barcoded. For barcoded single cells, after adaptor trimming, 99.12% of the barcoded reads were unique reads, and 94.06% of the barcoded reads were uniquely mapped to the hg38 reference genome.

2 Copy number profile (CNP) heatmap of single cells from four cell lines

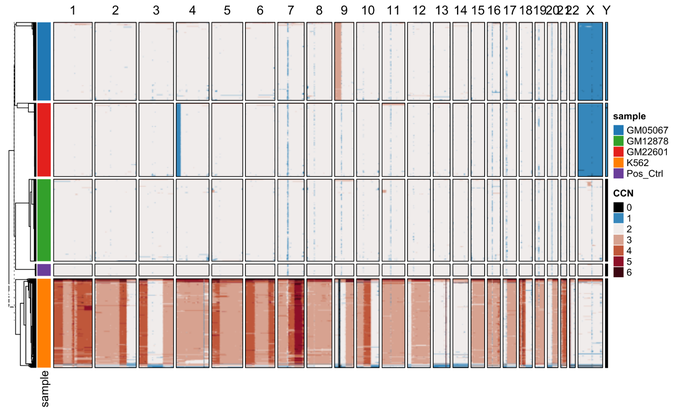


Figure 2. CNP heatmaps generated single cells of four different cell lines using the Shasta WGA kit. The WGA libraries of 1,124 single cells with known segmental and chromosomal aneuploidies were generated via the Shasta WGA kit. Ginkgo was used as the pipeline for CNV calling. The average read depth for a single cell was 250,000 and the average bin size for segmentation was 500 Kb. GM12878 cells showed no CNVs, similar to the NA12878 gDNA used in the positive controls. Both the segmental deletion (~25 Mb at chromosome 4p) in GM22601 cells and the segmental amplification (~45 Mb at chromosome 9p) in GM05067 cells were successfully detected at shallow sequencing depth. K562 cells, which are derived from leukemia cells, showed hyperploidy and more complex chromosomal alterations.

3 Example CNPs of single cells from each cell line detected using the Shasta WGA kit

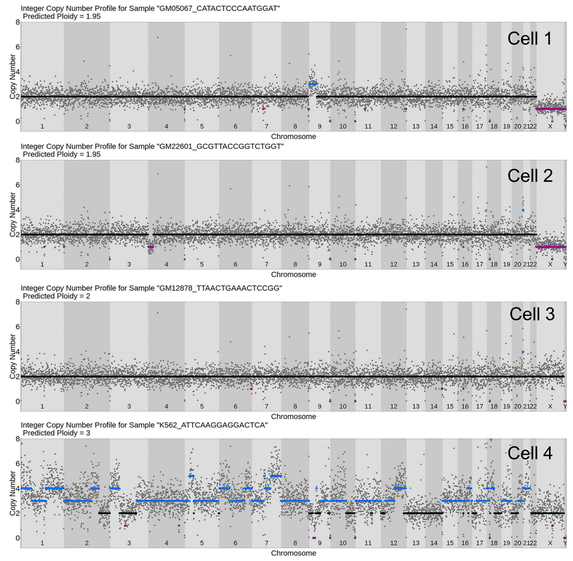


Figure 3. Example CNPs generated from individual cells with the Shasta WGA kit. Cell 1 is a GM05067 cell, and its signature ~45 Mb segmental amplification at chromosome 9p is specifically detected using the Shasta WGA chemistry. Cell 2 is a GM22601 cell, which shows a signature ~25 Mb segmental deletion at chromosome 4p. Cell 3 is a GM12878 cell. The single-cell WGA library showed a ploidy of 2 with no CNVs, which demonstrated comparable performance as the WGA library generated using 15 pg NA12878 gDNA. Cell 4 is a K562 cell, which has a predicated ploidy around 3. Hyperploidy is a common phenomenon for rapidly proliferating cancer cells.

4 GC content distributions of Shasta WGA libraries compared to other scWGA chemistries

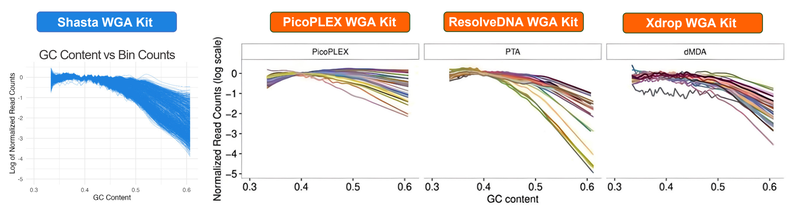


Figure 4. GC content of Shasta WGA libraries compared to other scWGA chemistries. The high-throughput single-cell WGA libraries generated using the Shasta WGA Kit have GC content distribution comparable to that of the plate-based PicoPLEX WGA Kit and the Xdrop WGA Kit, and a better GC content distribution than the ResolveDNA WGA Kit. PicoPLEX, ResolveDNA, and Xdrop data are from reference Kalef-Ezra et al. 2023.

5 Lorenz curves of Shasta WGA libraries compared to other scWGA chemistries

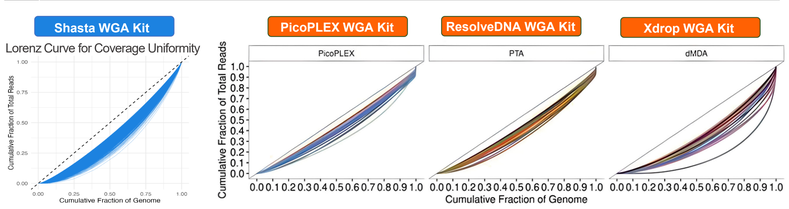


Figure 5. Coverage uniformity of Shasta WGA libraries compared to other scWGA chemistries. The high-throughput single-cell WGA libraries generated using the Shasta WGA kit have coverage uniformity comparable to that of the plate-based PicoPLEX WGA Kit or ResolveDNA WGA Kit, and a better coverage uniformity than the Xdrop WGA Kit. PicoPLEX, ResolveDNA, and Xdrop data are from reference Kalef-Ezra et al. 2023.

6 Pseudo-bulk SNV analysis for cell clusters

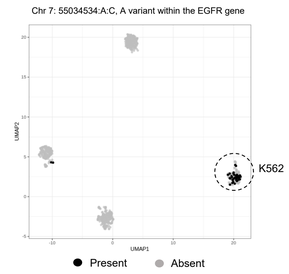


Figure 6. Pseudo-bulk SNV analysis from CNV profile clustering. While each cell only receives shallow sequencing, we developed a method to achieve sufficient read depth for SNV analysis by clustering single cells by their CNV profiles and performing pseudo-bulk SNV analysis using the Monopogen algorithm. The germline variants were called for each cluster and the putative somatic variants were called for each single cell. Single cells were clustered based on their CNV profiles. Each cell cluster has ~100 million paired-reads. Single cells carrying chr7: 55034534: A:C, a variant within the EGFR gene, are highlighted on the UMAP plot. Most of the cells are in the K562 cell cluster.

7 CNP heatmaps of single cells disassociated from tumor tissue and adjacent normal tissue

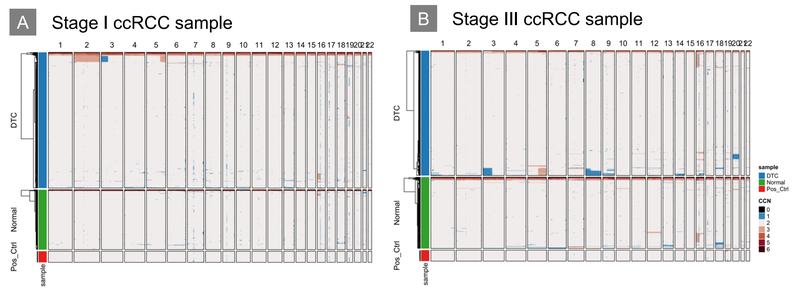


Figure 7. CNP heatmaps generated from clear cell renal carcimona cancer (ccRCC) tumor and tumor-adjacent normal tissue with the Shasta WGA kit. Ginkgo was used as the pipeline for CNV calling; the average bin size for segmentation was 500 Kb. The average read depth for the Stage I ccRCC sample (**A**) was 352,000 read pairs, and the average read depth for the Stage III ccRCC sample (**B**) was 114,000 read pairs. Tumor cells in Stage I showed +Chr 2, whereas tumor cells in Stage III sample showed other CNV events, including -8p, -Chr 8, -Chr 20, but the majority of single cells from the tumor tissue or the normal adjacent tissue were diploid cells with no extreme CNVs and there were common CNV events shared between the tumor cells and the normal adjacent cells (+16p, -Chr 18, and -Chr X). The tumor cells in the Stage I and Stage III samples both showed -3p and +5q alternations, which are characteristic CNV events for ccRCC. A portion of the tumor cells also showed unique CNV features. As the percentage of cells with CNVs was low, these features could have been masked with a bulk DNA-seq approach, however, the single-cell WGA approach allowed for discovery of the heterogeneity of the tumor cell population and rare cells with abnormal CNV features.

8 Example copy number profiles for single tumor cells

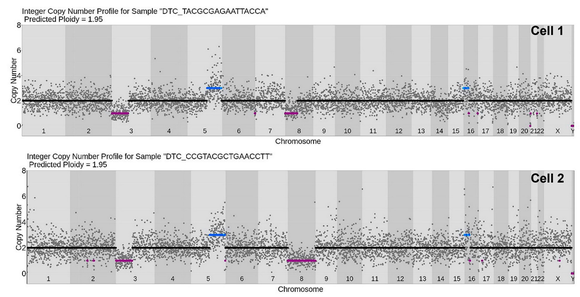


Figure 8. Example CNPs generated from individual ccRCC tumor cells with the Shasta WGA kit identify CNVs commonly associated with ccRCC. Cell 1 and Cell 2 both had a loss of the p-arm on chromosome 3 (-3p), a gain of the q-arm on chromosome 5 (+5q), and a partial gain of the p-arm on chromosome 16 (+16pter). -3p is a cytogenic hallmark of ccRCC, and 91% of 417 patients in a 2013 study had the -3p alternation (Cancer Genome Atlas Research Network Analysis 2013). This region encompasses the 4 of the most commonly mutated genes in ccRCC: *VHL*, *PBRM1*, *BAP1*, and *SETD2*. +5q is also a common CNV in ccRCC and 67% of the 417 patients from the 2013 study had the a +5q alternation. In addition, Cell 1 had a loss of the p-arm on chromosome 8 (-8p) and Cell 2 had a loss of an entire chromosome 8. The partial or complete loss of chromosome 8 has been reported in previous studies of ccRCC patients as well and is typically associated with *TCEB1* mutations (Sato et al. 2013).

Conclusions

- By adapting PicoPLEX technology to high throughput using the Shasta Single-Cell System, we have obtained single-cell WGA libraries from up to 1,500 cells at once.
- Libraries created using the Shasta WGA workflow enabled the reliable detection of CNVs and tumor subclones at a shallow sequencing depth.
- Leveraging the automated nanoliter-dispensing capabilities of the system this method provides a significant reduction in reagent use and labor compared to plate-based methods.

References

Cancer Genome Atlas Research Network Analysis. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43-49 (2013).

Kalef-Ezra, E. et al. Single-cell somatic copy number variants in brain using different amplification methods and reference genomes. *BioRxiv* (2023).

Sato, Y. et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature Genetics* 45, 860-867 (2013).