Resolving tumor heterogeneity with a fully automated, high-throughput single-cell WGA 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

Objective

Understanding genetic heterogeneity is crucial for improving the prognosis and treatment of cancer. Bulk sequencing methods lack the sensitivity to fully dissect heterogeneity, while single-cell methods have been limited by throughput. To address this, we developed Shasta™ Whole-Genome Amplification (WGA), an automated workflow for generating WGA libraries of up to 1,500 single cells of 1-8 samples in a single day.

Methods

Shasta WGA library was benchmarked against standard PicoPLEX® workflow, with its copy number variation (CNV) detection sensitivity assessed using cell lines with well-characterized segmental deletion (GM22601), segmental gain (GM05067), and hyperploid cancer cells (A-498, SK-BR-3). Both germline and putative somatic single nucleotide variants (SNVs) were called using a pseudo-bulk approach, Monopogen (Dou et al. 2024), based on CNV clusters. Additionally, Shasta WGA was applied to clear cell renal cell carcinoma (ccRCC) tumor samples. All analyses were done via Takara Cogent™ NGS Analysis Pipeline.

1 Shasta WGA workflow

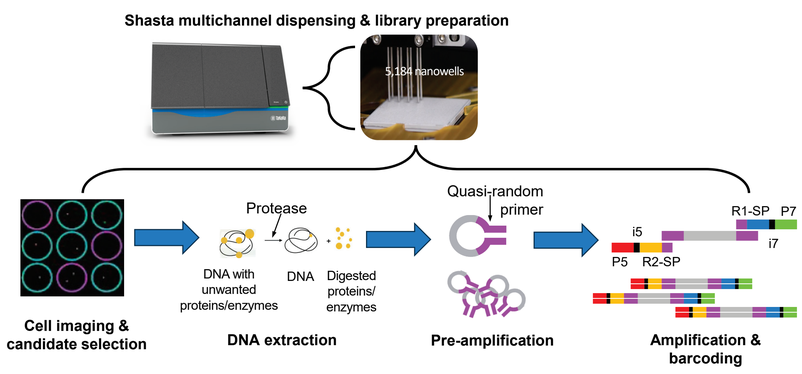

Figure 1. Shasta WGA workflow. Hoechst-stained single cells are dispensed into a 5,184-nanowell chip and screened by imaging. Reagents are deposited into nanowells via an additional five-step series of equal-volume (50 nL) dispenses: a lysis mix to release DNA from nuclei, a pre-amplification mix of quasi-random primers to 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

A Shasta WGA library containing 1,185 cells was sequenced and yielded comparable performances to standard PicoPLEX in coverage uniformity and GC bias, with a high unique mapping rate. Segmental aneuploidies were reliably detected in >95% of GM22601 and GM05067 cells at a shallow sequencing depth of 3.26 x 10^5 paired-end (PE) reads per cell. Using a pseudo-bulk SNV analysis approach, we identified ~1.4–1.8 x 10^6 germline SNVs and ~1–2 x 10^5 putative somatic variants for each cluster. Analysis of ccRCC single-cell samples revealed subclonal heterogeneity with signature ccRCC CNVs such as Chromosome 3p deletion and Chromosome 5q amplification.

2 Primary sequencing quality metrics of a Shasta WGA library generated from single cells of four different cell lines

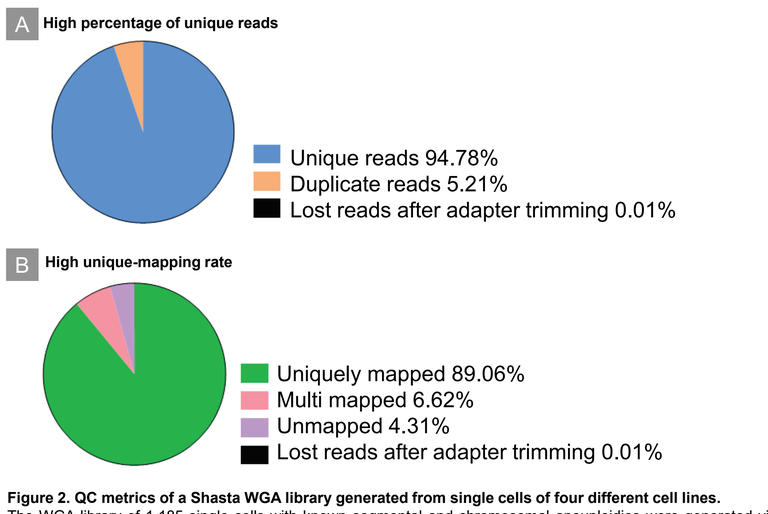

Figure 2. QC metrics of a Shasta WGA library generated from single cells of four different cell lines. The WGA library of 1,185 single cells with known segmental and chromosomal aneuploidies were generated via the Shasta WGA kit. Library was sequenced on NextSeq® 2000 at a read length of 2 x 151 bp, with a ~15% PhiX spike in. The average read depth is 326,000 PE reads per cell. Excluding PhiX, 93.41% of reads were successfully barcoded. After adapter trimming, 94.78% of the barcoded reads were unique reads (**Panel A**), and 89.06% of the barcoded reads were uniquely mapped to the hg38 reference genome (**Panel B**).

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

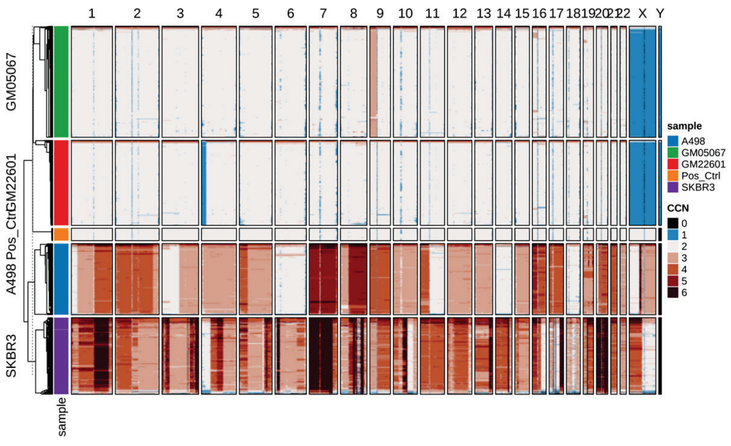

Figure 3. CNP heatmap of 1,185 single cells of four different cell lines. Ginkgo was used as the pipeline for CNV calling. The average read depth for a single cell was 326,000 and the average bin size for segmentation was 500 kb. These 4 cell lines have distinct copy number profiles. Within each cell line, cell-to-cell heterogeneity is also shown in the heatmap. GM22601 and GM05067 are diploid, A-498 has an intermediate ploidy, and SK-BR-3 is hyperploid. The segmental deletion (~26 Mb at chromosome 4p) in GM22601 cells and the segmental amplification (~45 Mb at chromosome 9p) in GM05067 cells were successfully detected at this shallow sequencing depth. NA12878 gDNA positive control wells showed uniform CNV profiles, confirming good well-to-well reproducibility.

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

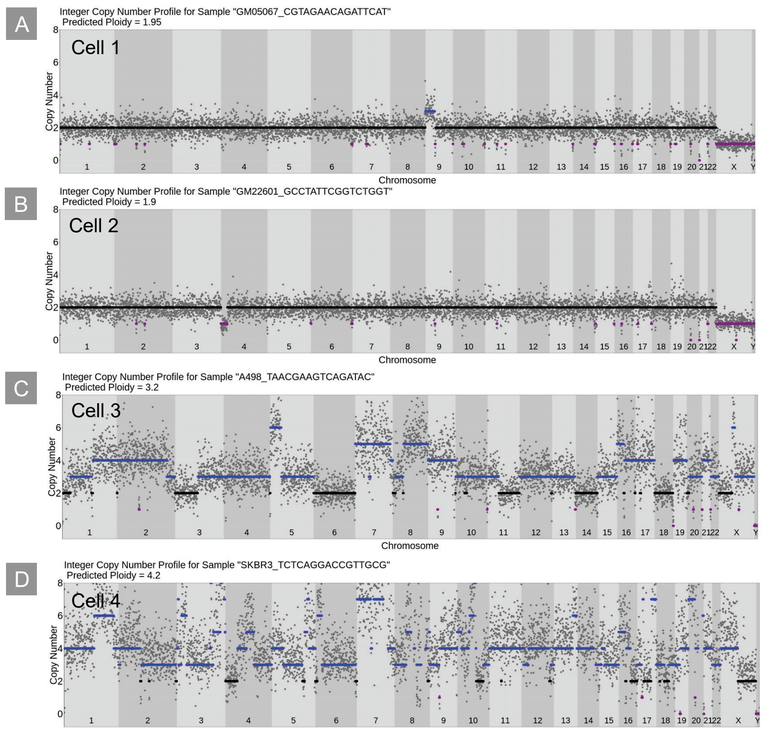

Figure 4. Example CNPs from individual cells of each cell line detected using the Shasta WGA kit.
Panel A. Cell 1 is a GM05067 cell, and its signature ~45 Mb segmental amplification at chromosome 9p is specifically detected using the Shasta WGA chemistry. **Panel B.** Cell 2 is a GM22601 cell, which shows a signature ~26 Mb segmental deletion at chromosome 4p. **Panel C.** Cell 3 is an A-498 cell. It has a predicated ploidy around 3 which is common for cancer cells. The 3p-arm loss is a marker feature for ccRCC. **Panel D.** Cell 4 is an SK-BR-3 cell which is hyperploid. SK-BR-3 also has complex chromosome compositions and structures.

5 Pseudo-bulk SNV analysis for cell clusters

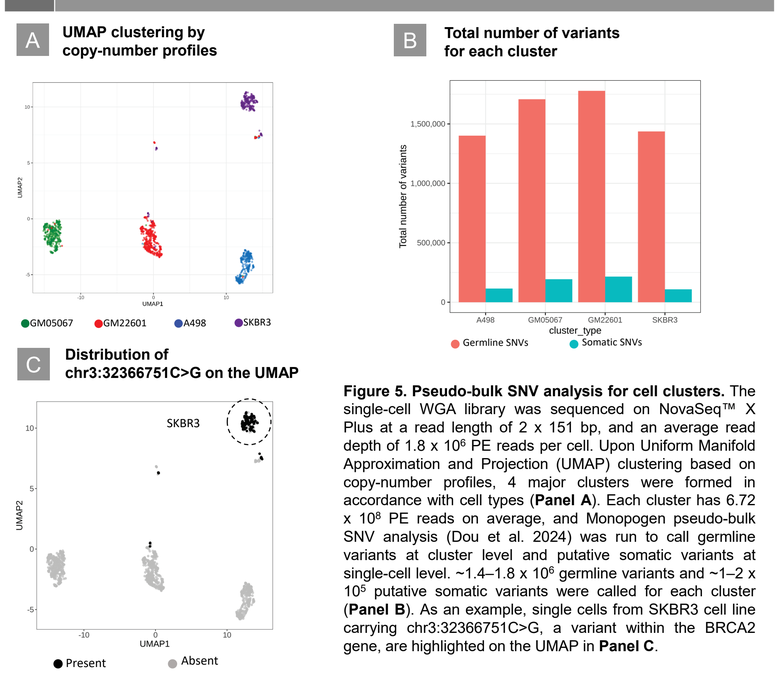

Figure 5. Pseudo-bulk SNV analysis for cell clusters. The single-cell WGA library was sequenced on NovaSeq™ X Plus at a read length of 2 x 151 bp, and an average read depth of 1.8 x 10^6 PE reads per cell. Upon Uniform Manifold Approximation and Projection (UMAP) clustering based on copy-number profiles, 4 major clusters were formed in accordance with cell types (**Panel A**). Each cluster has 6.72 x 10^8 PE reads on average, and Monopogen pseudo-bulk SNV analysis (Dou et al. 2024) was run to call germline variants at cluster level and putative somatic variants at single-cell level. ~1.4–1.8 x 10^6 germline variants and ~1–2 x 10^5 putative somatic variants were called for each cluster (**Panel B**). As an example, single cells from SKBR3 cell line carrying chr3:32366751C>G, a variant within the BRCA2 gene, are highlighted on the UMAP in **Panel C**.

6 CNP heatmaps of single cells disassociated from tumor tissue & adjacent normal tissue

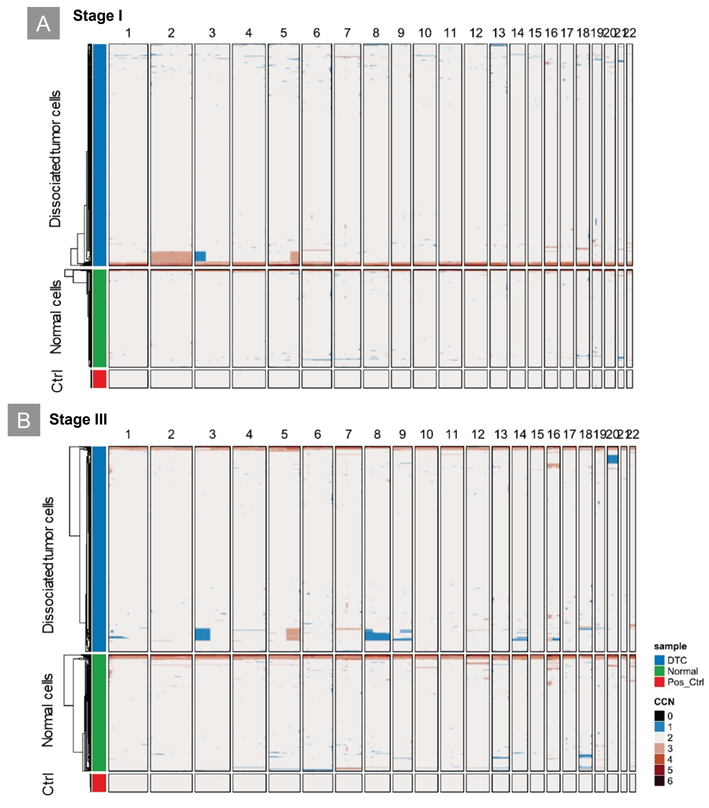

Figure 6. CNP heatmaps of single cells disassociated from the tumor tissues and the normal adjacent tissues of ccRCC. Dissociated cells from the tumor and normal adjacent tissues of two ccRCC patients were processed via Shasta WGA. The single-cell WGA library was generated for 815 cells from the Stage I patient, and 858 cells from the Stage III patient. The libraries were sequenced on NextSeq 500, at a read length of 2 x 151 bp, and an average read depth of 370,000 PE reads per cell. Ginkgo was used as the pipeline for CNV variation calling, and the average bin size for segmentation was 1 Mb. As shown in the heatmap, the majority of the single cells from either tumors or the normal adjacent tissues were diploid cells with no extreme CNVs. Intriguingly, a portion of the tumor cells showed unique CNV features.

As the percentage of cells with CNVs was low, these features could have been masked in bulk DNA seq data. However, the single-cell WGA approach allowed for discovery of heterogeneity in the tumor cell population and for the identification of rare cells with abnormal CNV features. The tumor cells in the Stage I and Stage III samples showed -3p and +5q, which are characteristic CNV events for ccRCC. The 4 most commonly mutated genes in ccRCC are included in -3p: VHL, PBRM1, BAP1, and SETD2. The two types of tumor cells also had unique CNV events. Tumor cells in the Stage I sample showed +Chr 2, whereas the tumor cells in the Stage III sample showed other CNV events such as -8p, -Chr 8, -Chr 20, +Chr 16, -9q, -Chr 9, etc.

7 Example copy number profiles for single tumor cells

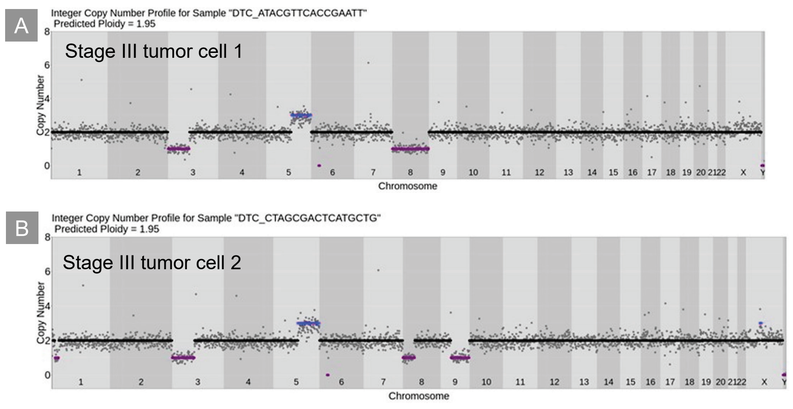

Figure 7. Example CNPs generated from individual ccRCC tumor cells with the Shasta WGA kit identify CNVs commonly associated with ccRCC. Tumor Cell 1 (**Panel A**) and Tumor Cell 2 (**Panel B**) both had loss of the p-arm on chromosome 3 (-3p), and gain of the q-arm on chromosome 5 (+5q). A comprehensive study from The Cancer Genome Atlas (TCGA) found that -3p is a cytogenic hallmark of ccRCC, that was present in 91% of 417 patients (Cancer Genome Atlas Research Network 2013). This region encompasses the 4 most commonly mutated genes in ccRCC: VHL, PBRM1, BAP1, and SETD2. +5q is also a common CNV in ccRCC. In the same TCGA study, 67% of the 417 patients showed +5q, which was associated with better patient survival. Tumor Cell 1 had loss of a p-arm on chromosome 8 (-8p) and Tumor Cell 2 had loss of an entire chromosome 8. The partial or complete loss of chromosome 8 has been reported in previous studies of ccRCC patients and is typically associated with TCEB1 mutations (Sato et al. 2013).

8 GC content distributions of Shasta WGA libraries compared to other scWGA kits

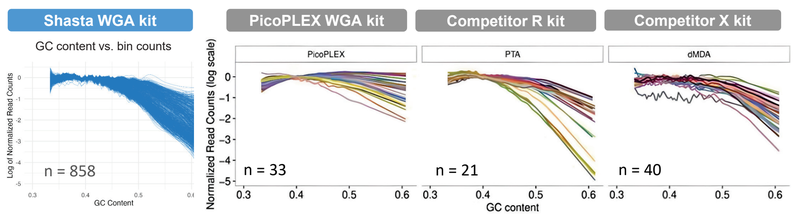

Figure 8. GC content distributions of Shasta WGA libraries compared to plate-based scWGA kits. The high-throughput single-cell WGA libraries generated using the Shasta WGA kit have a comparable GC content distribution as the plate-based PicoPLEX WGA Kit or Competitor X WGA Kit, and a better GC content distribution than the Competitor R WGA Kit. Figure adapted from Kalef-Ezra et al. 2023, *bioRxiv,* under a CC BY 4.0 license.

9 Lorenz curves of Shasta WGA libraries compared to other scWGA kits

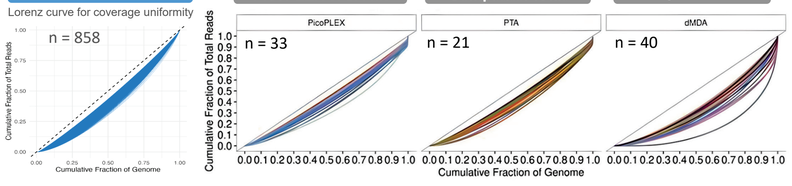

Figure 9. Coverage uniformity of Shasta WGA libraries compared to in-tube scWGA chemistries. The high-throughput single-cell WGA libraries generated using the Shasta-WGA kit have a comparable coverage uniformity as the plate-based PicoPLEX WGA kit or Competitor R WGA kit, and a better coverage uniformity than the Competitor X WGA kit. Figure adapted from Kalef-Ezra et al. 2023, *bioRxiv,* under a CC BY 4.0 license.

Conclusions

- The Shasta WGA workflow enables the construction of single-cell WGA libraries for up to 1,500 cells per chip.
- Shasta WGA libraries have demonstrated a high unique mapping rate and good coverage uniformity. It enables reliable detection of single-cell CNVs and pseudo-bulk SNVs. This capability aids in identifying tumor subclones even with shallow sequencing depth.
- Additionally, Shasta WGA utilizes automated nanoliter dispensing to substantially reduce reagent and labor costs compared to plate-based methods.

References

Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-9 (2013).

Dou, J. et al. Single-nucleotide variant calling in single-cell sequencing data with Monopogen. *Nature Biotechnology* **42**, 803-12 (2024).

Sato, Y. et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature Genetics* **45**, 860-7 (2013).

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

Learn more:
takarabio.com/shastaWGA

800.662.2566
www.takarabio.com