A droplet-based microfluidic platform combined with primary template-directed amplification (PTA) enables the analysis of clonal heterogeneity and mosaicism in leukemia at single-cell genome level

Nan Zong[1], Kyle Hukari[2], Joseph Dahl[2], Victor Weigman[2], Jay West[2], Stavros Stravrakis[1]and Andrew deMello[1]

[1]ETH Zurich, Switzerland , [2]BioSkryb Genomics, Durham, NC, USA

Abstract

Mosaicism in leukemia is driven by dynamic genetic instability within a heterogeneous population of cells. This introduces complexity into diagnosis and treatment. The primary challenge is the detection of rare clones within a tumor. A secondary hurdle is detecting point mutations in rare clones that inform treatment regimes.

To address the first challenge, we leverage droplet microfluidics and Primary Template-directed Amplification (PTA) to develop a high throughput platform for whole genome amplification. Specifically, we encapsulate single cells within hydrogel beads to allow amplification of each cell's genome. We demonstrate this using a patient-derived chronic myelogenous leukemia line, K-562, known for its genomic instability. Successful encapsulation of single cells within hydrogel beads allows for the formation of up to 1 million single-cell hydrogel beads/hour. These hydrogel beads were re-encapsulated with the PTA mixture within droplets for whole genome amplification, with real-time fluorescence monitoring of individual bead amplification. Detection of thousands of amplified single-cell genomes occurs in less than 1 hour. We then employ a microfluidic device to fragment genomes and attach unique barcodes to all fragments in the workflow.

To address the secondary challenge, we developed a targeted approach for sample sequencing. First, we assess the performance of libraries at low sequencing depth to determine the presence of discrete barcodes and to ensure accurate insert structure. High-performing libraries are then sequenced to a greater depth to determine cellular mosaicism through copy number variation (CNV) at approximately 0.5X per cell. To determine single-nucleotide variant (SNV) status, clusters of cells with common aneuploidy structures are bioinformatically grouped to establish sub-clonal populations. The accuracy of CNV-clustered SNV calling is verified with deeper whole genome sequencing of a subset of cells. Data delineates the existence of single-cell CNV and SNV mosaicism at a scale that allows detection of rare emergent clones, which could be applied to clinical samples to determine disease evolution and define pathology mitigation strategies.

Methods

- K562 cell line – CCL-24, cells were cultured in IMDM media
- Hydrogel media was 3% low temperature agarose in 1X PBS Maintained as a liquid prior to cell encapsulation. Cells and hydrogel media are then combined to create a droplet
- The hydrogel mix was merged with a perfluorinated oil using a microfluidic chip to form discrete droplets ~28 um in diameter
- Encapsulation of single cells within agarose droplets was achieved according to Poisson distribution
- Cooling on ice at ~ 4oC allows hydrogel mix to harden into a porous network, 6-10% of which contain a single K652 cell. These were identified by fluorescence during genome amplification
- A droplet generation device was used to re-encapsulate with reagents from the BioSkryb Genomics ResolveDNA v2 Whole Genome Amplification (WGA) kit Powered by PTA[1].
- WGA single-cell hydrogel merged with droplets containing Tn5 transposase to accomplish WGA product tagmentation, merging of one barcode droplet with one hydrogel through electrofusion
- NGS analysis at ~2 million reads/cell
- Custom analysis pipeline in the BaseJumper® to perform single cell CNV clustering and SNV calling from clonal groups
- We have termed this method ***Super Low Pass Sequencing*** **analysis** or **slpSeq**.
  Briefly, slpSeq corrects for gaps in 2 steps: a) leverage the high number of cells to build tissue-specific haplotype blocks and b) missing regions are then Haploboosted based on determined ancestry. This limits overfitting alleles based on population and letting rare cellular clones to contribute meaningful alleles. The data indicate that the method can detect ~ 85-90% of the mutations across the single-cells with equivalent of 3x WGS sequencing, and critically reducing the needed depth while reducing false-positive calling

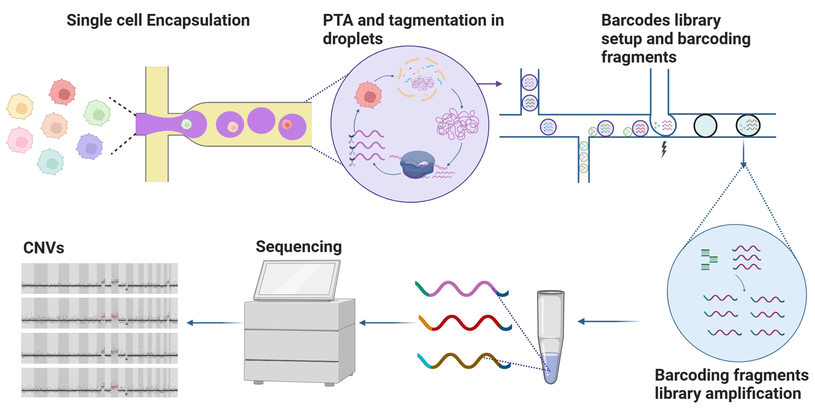


Figure 1: Method Schematic for Generation of Single Cell Droplets for ResolveDNA and slpSeq Analysis

Results

PRIMARY TEMPLATE-DIRECTED AMPLIFICATION (PTA)

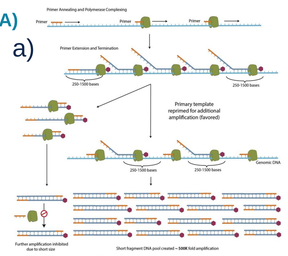


Figure 2: The PTA mechanism.

a) ResolveDNA makes use of proprietary nucleotides prevents the production of long amplicons, which are kinetically unfavored to be re-copied during amplification.

b) By limiting the size of the produced amplicon, primers are re-directed to the primary template, the performance of the assay outperforms many previously development WGS chemistry processes.

b)

Method	ResolveDNA	Mixed Method A	MDA A	MDA B	Mixed Method B	Mixed Method C	DOP-PCR
Genome Mapping	97%	91%	88%	55%	88%	55%	52%
Genome Recovery	97%	73%	65%	59%	50%	33%	20%
CV of coverage	0.8	1.3	1.8	2.3	2.6	3.2	3.6
SNV sensitivity	95-97%*	70%	65%	55%	45%	30%	19%
SNV Precision	99%	88%	87%	88%	28%	35%	35%

Sensitivity are based on positions that are 25X coverage for ResolveDNA, Values for alternaive methods taken from Gonzalez, etal. PNAS[1]
CV: Coefficient of Variation SNV: Single Nucelotide Variant

HYDROGEL ENCAPSULATION OF SINGLE-CELL AND CHEMISTRY INTEGRATION

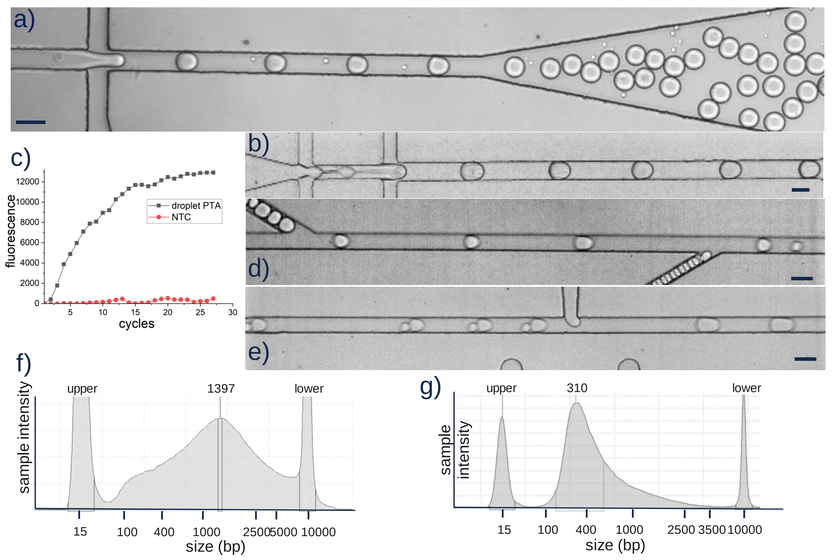


Figure 3. Encapsulation and Library Quality Control.

a) Single K562 cells were encapsulated in hydrogels in a droplet generation microfluidic.
b) Following lysis of cells inside the hydrogels, we use a second droplet generation device to re-encapsulate the hydrogels together with reagents from the Resolve DNA for whole genome amplification.
c) Amplification of single genomes contained within the hydrogel droplets as judged from the fluorescence signal variation of droplets.
d) A microfluidic device that creates pairing of each hydrogel droplet that contains fragmented DNA with a barcode droplet.
e) By using a picoinjector we load the PCR mixture efficiently in the merged droplets containing fragmented genomes and barcodes.
f) Histograms of DNA sizes extracted form gel images showing the distribution of the amplified DNA with a peak at 1397 bp.
g) Droplets containing fragmented genome and barcodes are collected, PCR amplified, and the genomes are purified. The size of the fragmented DNA after this process is at 310 bp. Scale bars: 50 µm

DEVELOPMENT OF SUPER LOW PASS NGS (slpSeq)

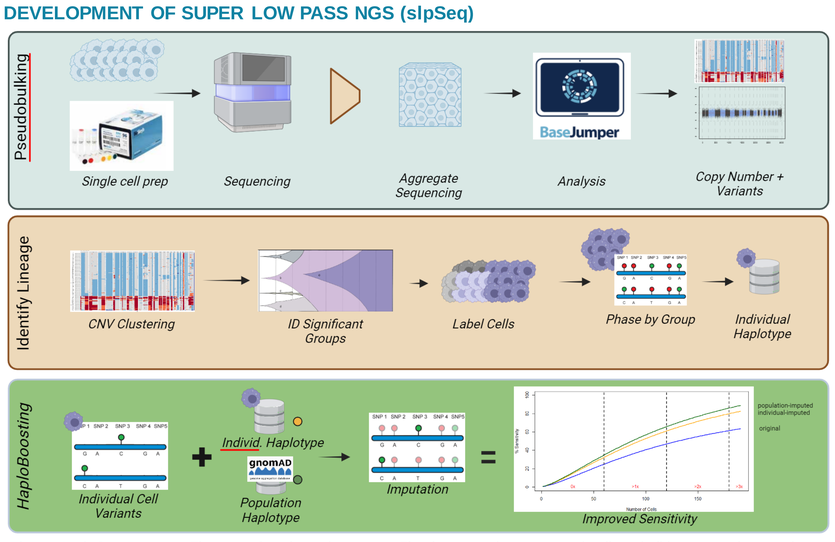


Figure 4: slpSeq Process. There are 3 main workstreams. First is to use the sequenced cells to build a composite sample to build allele frequencies across cells as well as clones present. Second is to use the cells associated to each clone to build a haplotypes and use those types to impute individual cells with alleles specific to the tissue. Lastly remaining gaps in individual cells are filled by ethnicity-based imputation. This allows alleles from rare cells to be preserved and greatly reduce false positives

SINGLE-CELL CNV ANALYSIS OF K562 CELLS

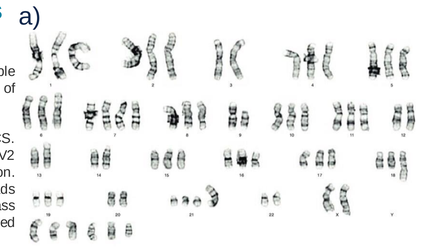
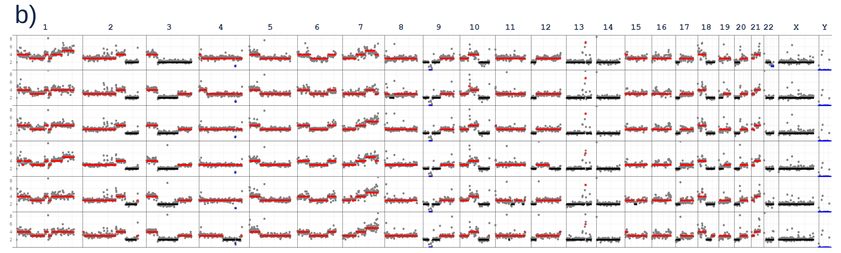


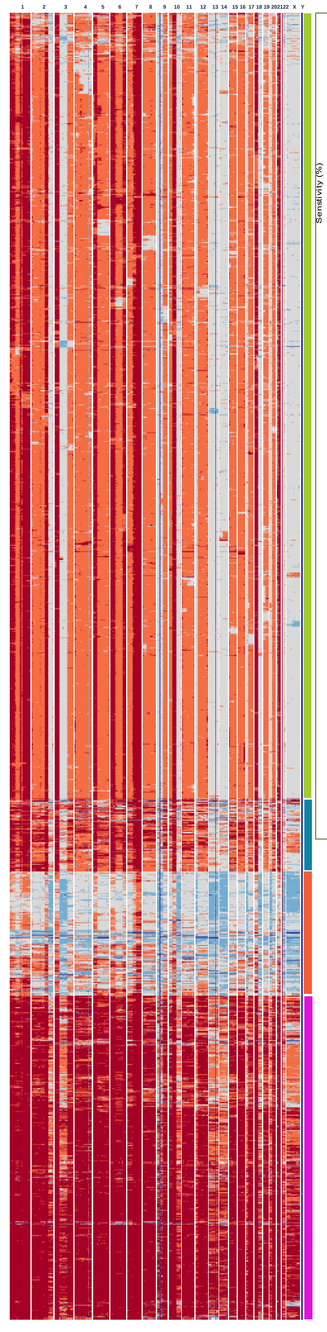
Figure 5. Baseline of K562 genome mosaicism.

a) Karyotyping of K562 cells[2]. Ploidy is characterized by multiple chromosomes with full duplication ranging from 3 to 5 copies of a given chromosome.

b) Single cells and beads containing single cells sorted by FACS. Single genomes were prepared using standard ResolveDNA V2 and library preparation. Here we show 3 representative single cells (top 3) or beads (bottom 3). The quality of genomic coverage with lowpass sequencing (~2M reads) is indistinguishable for encapsulated beads and cells.



HIGH THROUGHPUT SINGLE CELL GENOME ANALYSIS



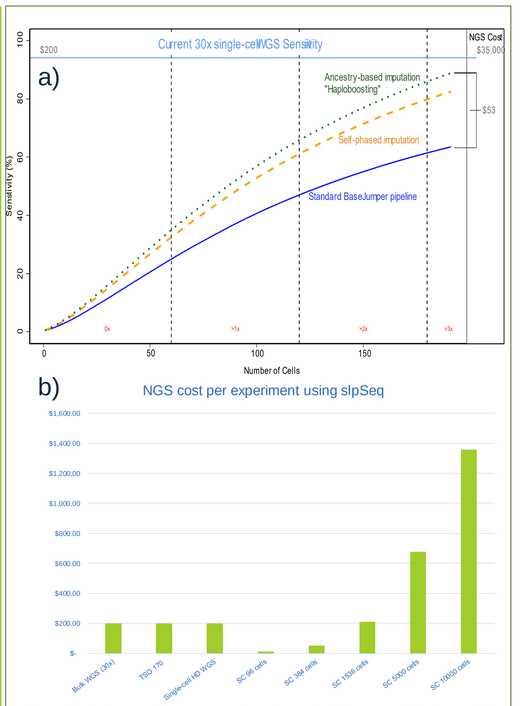


Figure 6: slpSeq Performance and estimated NGS cost. a) Allele sensitivity improvements through 3 parts of the slpSeq method. Blue represents typical incremental detection of individual cells. Orange curve from imputation based on tissue-based haplotyping. Green represents haploboosting, where remaining gaps in individual cells are imputed on ethnicity-based haplotype blocks B) per sample sequencing costs through different panel and WGS sequencing modalities of high volumes of single cell sequencing

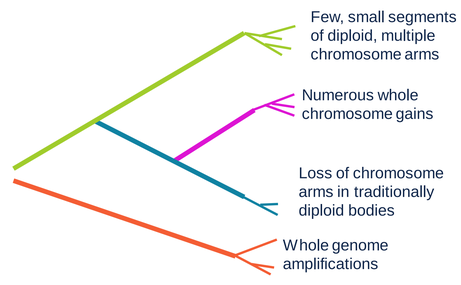


Figure 7: High throughput CNV: A sub sample of encapsulated k562 leukemia cells (~ 3K) was processed by NGS (2M reads/cell) reveals significant clonal diversity (A). We found ~ 4 large clonal populations, with additional diversity within each group. As showing in Fig 5A we confirm diploid of chromosomes 13, 14 and 22 and the overall 3 genome copies, B) Reduced lineage schematic, showing the further evolution of the major clones within K562 cells (represented by colors from group labels in a. The group subclones are designated by completely different mechanisms for observed changes

Conclusions

- Encapsulation of leukemia cells in hydrogel beads allows high throughput processing of single cells and integration of amplification and NGS library prep.
- Clustering of individual cells by CNV and SNV delineates the sub-clonal structure of the population at a fraction of the current high depth NGS costs.
- Reducing the cost of SNV analysis of single cells from ~ $600 to approximately $0.135 to $1.00 per cell.

References

1. Proc Natl Acad. Sci U S A. 2021 Jun 15;118(24):e2024176118.doi:10.1073/pnas.2024176118.
2. Genome Res. 2019 Mar; 29(3):472-484. doi: 10.1101/gr.234948.118