TECHNICAL NOTE

Sequencing Guidelines for Single Cell Copy Number Variant Detection

Introduction

The Chromium Single Cell CNV Solution provides a comprehensive, scalable solution for revealing genome heterogeneity and understanding clonal evolution. Chromium Single Cell DNA libraries can be sequenced and analyzed using the Cell Ranger DNA software package. This Technical Note provides guidelines for determining the optimal combination of targeted cell number, sequencing depth, and Copy Number Variation (CNV) resolution for Single Cell DNA libraries.

Genomic heterogeneity is the hallmark of many complex diseases, including cancer, and is characterized by cellular subpopulations evolving into distinct genotypes and phenotypes. Traditional bulk methods, such as whole genome sequencing (WGS) and microarrays, limit our ability to accurately assess and investigate the clonal makeup of complex diseases. The Chromium Single Cell CNV Solution offers the capability to resolve genomic heterogeneity by profiling the sample at single cell resolution.

Figure 1 delineates the factors and choices that impact the design of a single cell DNA experiment. The sequencing depth per cell determines the smallest CNV event that can be detected. Conversely, when the desired CNV resolution is known, it dictates the per cell sequencing that will be required. To detect rare populations of cells an increased target cell number is recommended. For a fixed total sequencing depth, therefore, the CNV resolution and the sensitivity to detect a rare population can be traded off against each other.

The key considerations that determine the sequencing depth for Single Cell DNA libraries are CNV resolution, sample type, and sequencing capability (Figure 1). Based on the experimental hypothesis, sequencing depth for Chromium Single Cell DNA libraries may range from 50,000-750,000 read pairs per cell. Sequencing libraries at 750,000 read pairs per cell enables accurate detection of ~2 Mb events per single cell. However, when detecting larger CNV events, sequencing at less than 750,000 read pairs down to 50,000 read pairs per cell can be performed.

Sequencer Type

Chromium Single Cell DNA libraries can be sequenced on all Illumina sequencers.

Generating & Sequencing Chromium Single Cell DNA Libraries

Experimental Design

Scenario 1	Scenario 2
Desired CNV Resolution	Available Sequencing Capability
↓	↓
Choose Targeted Cell Number (based on sample heterogeneity)	CNV Resolution Required
↓	↓
Determine Sequencing Depth	Determine Targeted Cell Number

Key Considerations

Sample	• 250-5,000 target cells/sample • Higher cell numbers recommended for detection of rare cell populations
Sequencing Depth	• 50,000-750,000 read pairs /cell
CNV Resolution	• Smaller CNV events – Sequence more • Larger CNV events – Sequence less (see Table 1 for details)

- Choose sequencing depth based on target cell number, CNV resolution, and/or sequencing capability
- Compatible with all Illumina sequencers

Figure 1. Guidelines for generating and sequencing Chromium Single Cell DNA libraries.

Table 1 shows the calculated total sequencing read pairs for Single Cell DNA libraries as a function of sequencing read pairs per cell and the targeted cell numbers in a sample. The single cell CNV resolution is determined by the sequencing depth per cell, while the cell numbers determine the minimum clonal fraction sensitivity. The color index in Table 1 shows total sequencing reads relative to a bulk human whole genome sequencing (WGS) experiment (2x100 bp). These sequencing recommendations are for Chromium Single Cell DNA libraries generated from the human genome. If profiling CNVs in another organism, the same CNV resolution can be achieved by using this calculation:

Read pairs per cell (Organism of interest) =
Read pairs per cell (human) x (Organism genome size/ Human genome size)

Impact of Sequencing Depth on CNV Detection: MKN-45 Cell Line

MKN-45 is a gastric cancer cell line with low cellular heterogeneity and a broad spectrum of copy number events, both in terms of copy number change & event size. A comparison of the bulk CNV and Single Cell CNV data generated from MKN-45 cells recapitulates the theoretical trends shown in Table 1.

Chromium Single Cell DNA libraries were generated from 5,220 MKN-45 cells and sequenced to a single cell depth of 1.25 million read pairs per cell (2x100 bp, see Datasets). Figure 2 shows a cluster of 2,122 cells (Group 10114) that was relatively homogeneous and whose CNV profile best matched the pseudo-bulk CNV calls obtained by first aggregating reads from all cells and then performing CNV calling.

Bulk CNV calls were originally determined by generating a TruSeq PCR-free library from DNA extracted from MKN-45 cells and processing the data through the Ginkgo CNV caller (see References). The results in Table 1 were validated by comparing single cell CNV calls for each cell in Group 10114 with bulk CNV calls. This approach enables comparison of the sensitivity but not the positive predictive value, since single cell calls are not expected to be identical to bulk calls.

Table 1. Total sequencing depth for Chromium Single Cell DNA libraries calculated as a function of read pairs per cell and targeted cell numbers, along with corresponding single cell CNV resolutions and minimum clonal fraction sensitivity.

Sequencing Read Pairs/Cell	Single Cell CNV Resolution (Mb)	Targeted Cell Numbers				
		300	500	1,000	2,500	5,000
		Minimum Clonal Fraction Sensitivity				
		2.8%	1.7%	0.8%	0.3%	0.2%
		Total Sequencing Read Pair in Millions (M)				
50,000	13 ± 4	15 M	25 M	50 M	125 M	250 M
100,000	7 ± 2	30 M	50 M	100 M	250 M	500 M
150,000	5 ± 2	45 M	75 M	150 M	375 M	750 M
300,000	2.5 ± 0.7	90 M	150 M	300 M	750 M	1,250 M
500,000	1.8 ± 0.5	150 M	250 M	500 M	1,250 M	2,500 M
750,000	1.4 ± 0.3	225 M	375 M	750 M	1,875 M	3,750 M

Single Cell CNV Resolution: Estimated based on the smallest copy number change from 2 to 3 events that can be theoretically detected in a diploid human sample with 95% sensitivity and 95% positive predictive value (PPV), given the sequencing depth and the properties of the Single Cell DNA library. Although this resolution is calculated based on change from 2 to 3 events, it serves as a ballpark estimate for other copy number changes; e.g., 2 to 1 events are expected to have similar resolution, while smaller 2 to 4 events can be detected.

Cell Numbers: In a sample with a rare cell population, it is possible that no rare cell is observed due to probability of detection. The detection limit is defined as the smallest population of cells as a percentage of the total number of cells that would result in observing at least 3 cells from the rare population 99% of the time. To increase detection, using higher number of cells is recommended.

Color Index

< 20X WGS
20-40X WGS
> 40X WGS

Figure 3 summarizes the correlation between CNV event size resolution and read pairs per cell in Single Cell DNA library data from MKN-45 cells. The plot shows the smallest event sizes that can be detected with at least 90% (orange) and 95% (blue) sensitivity at the indicated sequencing depths. The plot also shows the single cell CNV resolution (green) from Table 1, along with error bars for cell-to-cell variability.

In summary, the data derived from 5,220 MKN-45 cells aligns well with the trend from Table 1, showing that as the sequencing depth per cell is decreased, CNV events can still be detected at the single cell level.

Clustering & Aggregation Increases CNV Resolution

Aggregation of reads from cells that share a CNV event increases read coverage, which in turn enhances the signal to noise level, resulting in improved detection of smaller CNV events that are not detected at the single cell level. For example, at a sequencing depth of 750,000 read pairs per cell, aggregating data from 10 cells (1% at 1,000 targeted cells) allows for the detection of 200 Kb CNV events, which is 10x more than the resolution at the single cell level. This increased resolution for CNV detection is only possible when cells can be clustered accurately at the single cell level.

Table 2 shows one such example with 1,000 target cells. The CNV event resolution (Kb) with 1%, 5 % , and 10% of clones for corresponding sequencing depths is shown. The resolution is calculated as in Table 1 and is a theoretical lower bound on the event size.

Note that the cellranger-dna pipeline (part of Cell Ranger DNA 1.1) bins the data at 20 Kb resolution. Hence, CNV calls <100 Kb can be inaccurate (gray* cells in Table 2). This is not a fundamental limitation of the library but of the current software.

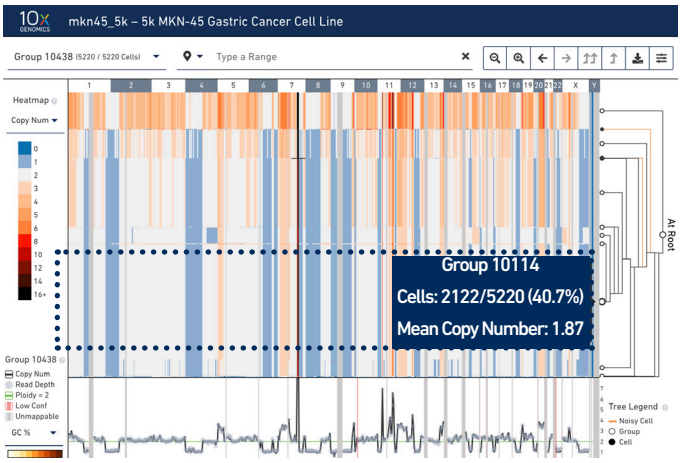


Figure 2. Single Cell CNV data from MKN-45 cells. CNV calls in a cluster of 2,122 cells (Group 10114) CNV calls are in agreement with the pseudo-bulk CNV calls (bottom of figure).

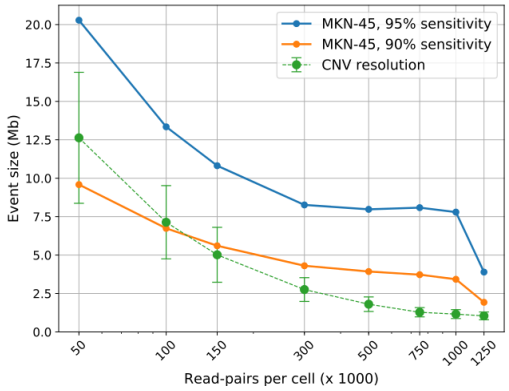


Figure 3. Correlation between CNV event size detection and sequencing read pairs per cell in Single Cell DNA library data from MKN-45 cells.

Sequencing Read Pairs/Cell	1,000 Targeted Cells		
	Clone Fraction		
	10 cells (1%)	50 cells (5%)	100 cells (10%)
	CNV Resolution (Kb)		
50,000	1,516 ± 520	338 ± 117	177 ± 61
100,000	848 ± 288	188 ± 65	*
150,000	593 ± 215	131 ± 48	*
300,000	323 ± 92	*	*
500,000	210 ± 56	*	*
750,000	148 ± 35	*	*

*CNV calls <100 Kb cannot be detected accurately by the software.

Table 2. Cell clustering to detect rare CNV events. Sequencing depth and corresponding CNV resolutions for different clone fractions in a 1,000 target cell experiment are shown below.

Impact of Sequencing Depth on CNV Detection: Breast Cancer Cells

To demonstrate the dependence of CNV detection on sequencing depth, a 5 Mb deletion event (~Chr22:31,680,000-22:36,720,000) was identified in 51 cells from 10,088 cells (0.5% clone) in the public breast cancer cell dataset (see Datasets). A diploid region and a heterozygous duplication flank this deletion event. Data from the 51 cells were processed using cellranger-dna reanalyze, available as part of Cell Ranger DNA 1.1, to examine the ability to retain this call, with appropriate boundaries, as sequencing depth was reduced.

The full dataset, sequenced at 622,000 read pairs per cell, was computationally downsampled to various sequencing depths (Figure 4A-4B). At 500,000 and 300,000 read pairs per cell, the event boundaries were comparable to the original data with the exception of one cell (Figure 4A-4B). At 150,000 read pairs per cell, approximately half of the cells retained the event boundaries, except 7 cells with merged boundaries and 4 cells with no boundaries detected (Figure 4C). At 50,000 read pairs per cell, almost all cells lost the original event but at aggregate resolution, the event was still retained (Figure 4D, bottom ploidy track).

This example demonstrates visually the change in detection quality for a fixed size CNV event as the sequencing depth is reduced. These data are consistent with Table 1 and show that the amount of sequencing can be reliably and predictably modified based on the desired CNV resolution.

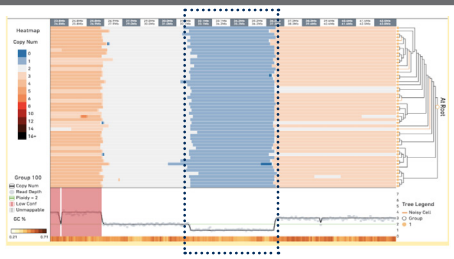
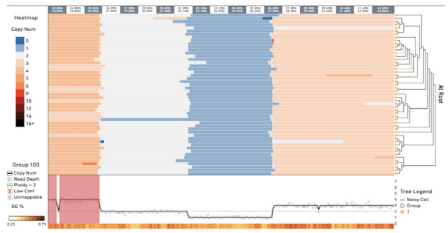
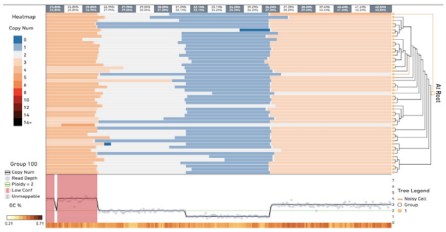
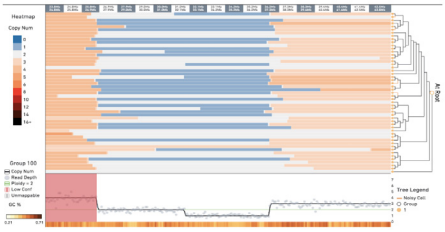
Sequencing Depth	5 Mb Deletion Event Detection	Results
4A 500,000 read pairs/cell	• Event boundaries retained in all cells • Comparable to data sequenced at 622,000 read pairs/cell	
4B 300,000 read pairs/cell	• Event boundaries retained in all cells • Still comparable to original data, except 1 cell	
4C 150,000 read pairs/cell	• Event boundaries retained in ~50% cells • Event boundaries were merged in 7 cells and not detected in 4 cells	
4D 50,000 read pairs/cell	• Event is retained at aggregate resolution (bottom ploidy track)	

Figure 4. Correlation between sequencing depths and CNV resolution demonstrated in dataset derived from breast cancer cells.

Conclusions

In summary, the Chromium Single Cell CNV Solution[1] is a flexible and versatile tool for shallow single cell sequencing of genomic DNA across hundreds to thousands of cells to study genome heterogeneity and understand clonal evolution. As described in this Technical Note, based on the research hypothesis and desired CNV resolution, the targeted cell numbers and sequencing depth can be optimized to achieve the best results.

[1]Available only on the Chromium Controller (PN-120223;120246) or Chromium Single Cell Controller (PN-120263;120212).

References

Garvin, Tyler, Robert Aboukhalil, Jude Kendall, Timour Baslan, *et. al*., **Interactive analysis and assessment of single-cell copy-number variations.** *Nature methods* 12, no. 11 (2015): 1058.

Datasets

https://support.10xgenomics.com/single-cell-dna/datasets

Contact:
support@10xgenomics.com
10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588 USA