mission bio

tapestri
APPLICATION NOTE

# Copy Number Variants and Single Nucleotide Variants Simultaneously Detected in Single Cells

## Takeaways

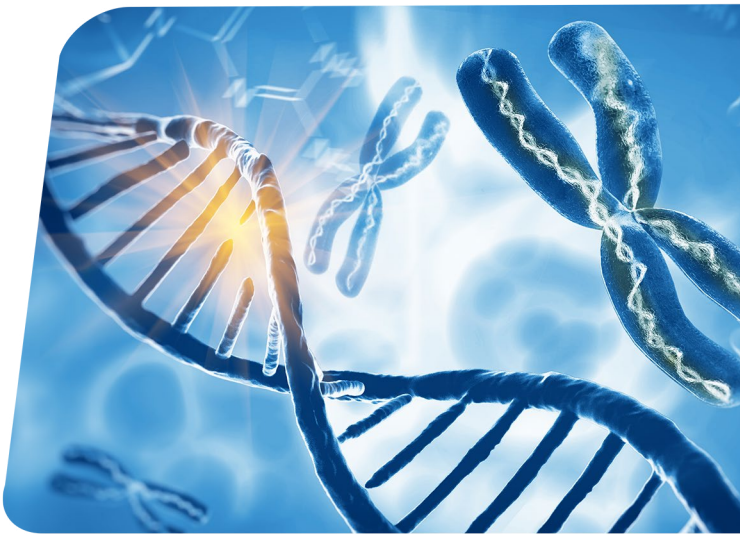
- The Tapestri Platform is the first and only single-cell platform to simultaneously detect CNVs and SNVs from the same cell
- Use the existing Tapestri workflow and Single-Cell DNA Panels to analyze samples for both CNVs and SNVs across thousands of cells
- Calculate gene-level and chromosome-level amplifications and deletions, including LOH


## Abstract

Cancer is a clonal disease developing from a single cell, however, because of its inherent genetic instability cancer cell populations tend to evolve to be genetically heterogeneous. In addition to genetic mutations often caused by single nucleotide variants (SNVs), gene copy number variants (CNVs) including loss of heterozygosity (LOH), play a large role in driving cancer progression and evolution. Differences in SNVs and CNVs contribute to cancer heterogeneity making some clonal populations more virulent and drug-resistant than others. Accurately defining clonal populations and reconstructing clonal phylogenies can only be achieved through single-cell analysis and is critical for informed clinical research. Previously, SNV and CNV status was unable to be determined in the same single cell at scalable cost and throughput. Here we show this capability for the first time using the existing Mission Bio Tapestri® Platform and reagents. Single-cell CNV analysis is a new capability for the Tapestri Platform - the only commercially available system that detects CNVs at the gene-level or chromosome-level using gene-specific amplicons versus whole-genome amplification approaches. Using several cancer cell lines and tumor samples, the Tapestri Platform co-detects CNVs and SNVs in the same cell starting from whole single cells or isolated single nuclei. Data correlates with publicly available known CNVs from the Catalogue of Somatic Mutations in Cancer (COSMIC) database[1] as well as bulk sequencing. The single-cell data enables the detection and definition of clonal subpopulations with distinct CNV/SNV signatures and identifies rare clones with high sensitivity.


## Experimental Methods

Mutz-8, Raji, K562 and Jurkat cell lines were used with the Tapestri Single-Cell DNA AML and Myeloid Panels. In addition, two renal cell carcinoma (RCC) samples from a previously published study[2] were analyzed using a 338 amplicon custom panel covering ~67.9 kb. The standard Tapestri workflow was followed starting from whole cells for the AML and Myeloid panels, and nuclei isolated from solid tumor for the RCC panel. Illumina sequencing results were analyzed with Tapestri Pipeline software. Loom and barcode distribution files were exported from Tapestri Pipeline software and further analyzed to identify SNVs, indels, CNVs and LOH. Each sample was analyzed by running the “tapestri-cnv” method in R:



1. Cell normalization: The reads of each cell were first normalized by the cell's total read count and cells were grouped by hierarchical clustering based on amplicon read distribution or SNVs. A control cell cluster with known CNVs was identified and amplicon counts from all cells were divided by the median of the corresponding amplicons from the control group.

2. Detection of LOH: Variants were excluded if they were a. mutated in <5% of the cells and b. called wildtype reference (WT) in >96% of the cells. Cells were clustered according to the grouping of the variants and CNVs were identified where heterozygous (HET) variants became consistently homozygous (HOM) or WT across large regions.

# Results

## Reliable CNV analysis and clonal clustering using the Tapestri Platform

Mutz-8, Raji, K562 and Jurkat cells were mixed together at 43%, 26%, 20% and 11% respectively and analyzed using the Tapestri Platform and Single-Cell DNA AML Panel. Normalized percentage of sequencing reads using the amplicons in the AML panel were used to calculate CNVs for each gene tested. Jurkat cells were used as a control cell line with a known diploid status for all genes tested. The identities of the 4 cell lines were resolved using unsupervised clustering of the CNV data and visualized with a heat map (Figure 1a). In addition, CNV data were grouped on a t-SNE plot (Figure 1b) and data were color-coded based on SNV genotypes previously established from pure cell lines (Table 1). The data showed clear correlations between CNV-based clustering and cell-line specific SNV signatures, and showed for the first time, CNV and SNV measurements from the same cells. Plotting the calculated CNV value across all genes shows averaged integer gene-specific CNV calls in the Mutz-8, Raji and K562 cell lines (Figure 2a). Observed CNVs across 13 genes in the Raji and K562 cells compared to known CNVs obtained from the COSMIC database showed high correlation with $R^2$ = 0.92 (Figure 2b). COSMIC information was not available for Mutz-8.

| SNV | K562 | RAJI | MUTZ-8 | JURKAT |
|---|---|---|---|---|
| EZH2-148504817 | WT | HET | WT | WT |
| EZH2-148504853 | HOM | HET | HOM | WT |
| KIT-55599435 | HOM | WT | HET | HET |
| TP53-7578114 | HOM | HET | HOM | HOM |
| TP53-7578210 | WT | HET | WT | WT |
| TP53-7578522 | HOM | WT | WT | WT |
| TP53-7577580 | WT | HET | WT | WT |

Table 1: Cell-line specific SNV data determined from previous experiments.

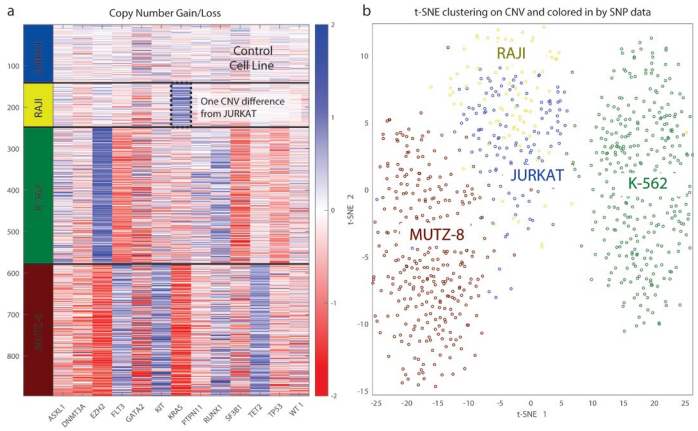



Figure 1: (a) heat map and (b) t-SNE plot derived from CNV data of a mixed population of 4 cell lines showed clear clustering of cell lines based on CNV status.

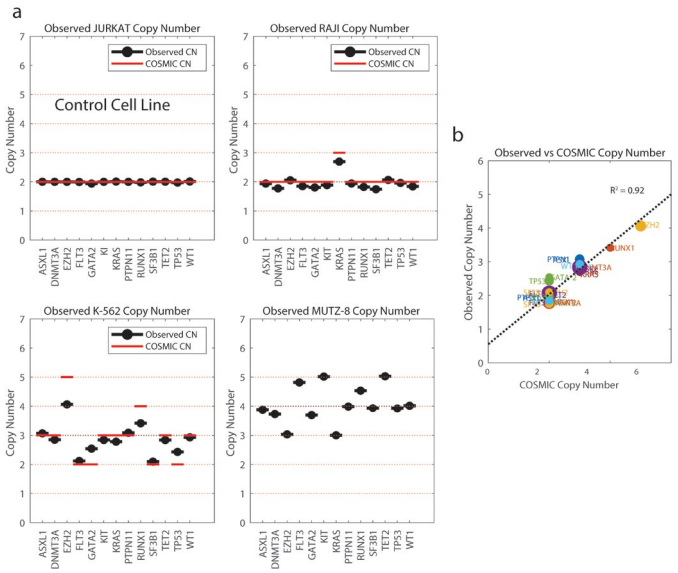



Figure 2: (a) CNVs determined using the Tapestri platform (b) compared with COSMIC data illustrates high correlation between observed and known CNVs.

To illustrate the sensitivity of CNV detection, K562 cells were mixed with Raji cells at ratios of 50%, 10%, and 5% and analyzed for SNVs and CNVs using the Tapestri Single-Cell DNA Myeloid Panel. Heat maps (Figure 3a) and t-SNE plots (Figure 3b) clearly showed clustering of 2 different cell types using CNV data, and illustrate the ability of the platform to detect rare populations of cells based on CNVs.

Together, results using the Tapestri Single-Cell DNA AML and Myeloid Panels show that CNVs can be determined on a single gene basis using small panels (125 amplicons for AML) or large panels (312 amplicons for Myeloid) on the existing Tapestri Platform.

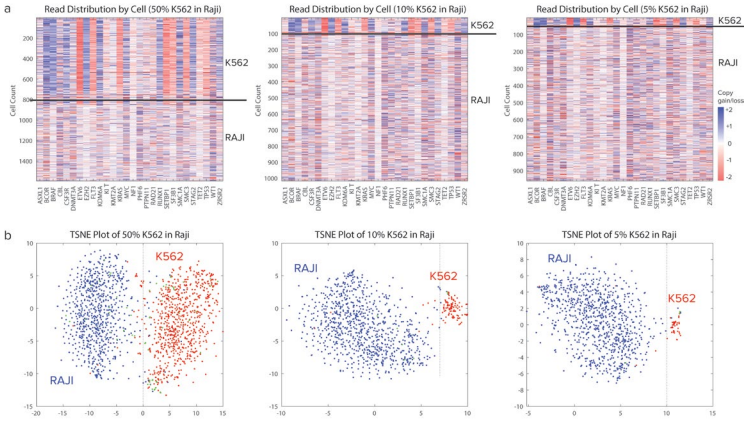



Figure 3: a) heat maps and (b) t-SNE plots derived from CNV data of a mixed population of Raji and K562 cells show clustering based on CNV status from a 50% to 5% ratio. t-SNE plots are colored based on SNV data.

## LOH shown in RCC samples

RCCs have a high prevalence of LOH in several chromosomal regions, including chr. 3, 9 and 14[3]. These chromosome deletions can result in the loss of critical tumor suppressor genes and enhance the progression of cancer. To test if the Tapestri Platform can infer LOH in RCC samples, two samples from different regions of the same tumor[2] were analyzed using a custom panel targeting regions within chromosomes 1, 3, 9, 10, 14 and X. Plotting the calculated CNV values across amplicon position along the chromosomes showed potential areas of LOH in chromosomes 3, 9 and 14 in Sample 1 (Figure 4a), and LOH in chromosomes 3 and 14 in Sample 2 (Figure 4b). CNV status was calculated and plotted across chromosomes. Data showed that cells associated to LOH in Sample 1 lost VHL, SETD2, BAP1 and PBRM1 among other genes on chr. 3 and RAD51B, PTPN21 and others on chr. 14 (Figure 4c). Sample 2 lost VHL and SETD2 on chr. 3, ADAMTSL1 and CDKN2B on chr. 9, and several genes on chr. 14 (Figure 4d).

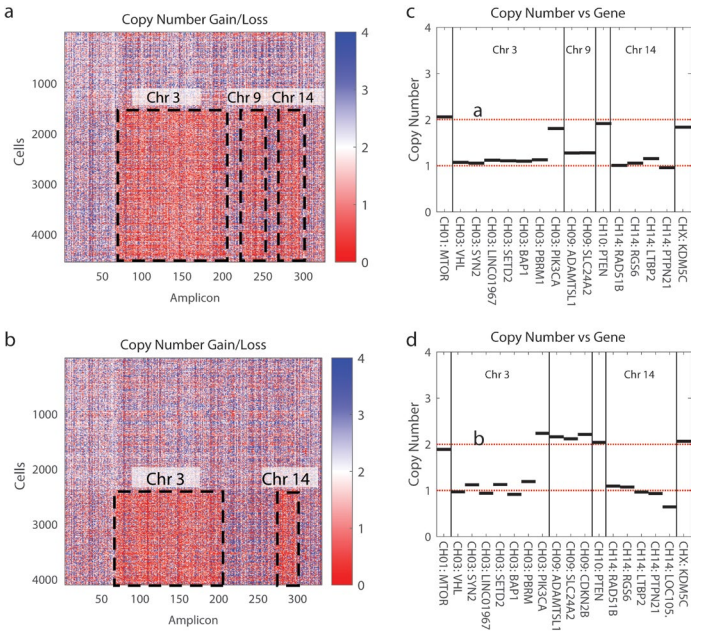



Figure 4: (a-b) Heat maps of CNV clustering from 2 RCC patient samples showed grouping of normal diploid cells and cancer cells containing various chromosomal losses. (c-d) CNV plots by gene illustrate the loss of one copy for specific genes in each sample.

Using the genotypes of WT, HET, and HOM called by Tapestri Pipeline software, clones with normal diploid copy numbers compared to clones with loss of copy number in each sample were distinguishable using hierarchical clustering and heatmap visualization. Again, Sample 1 showed a population that had LOH in chr. 3, chr. 9 and chr. 14 (Figure 5a), while Sample 2 showed a population identified by LOH on chr. 3 and chr. 14 (Figure 5b). Finally, in addition to the LOH calls, SNVs and indels were detected and matched 100% (12/12) with bulk data published from the same samples[2] (Table 2). Data show for the first time, the ability to detect LOH and SNV and indels in the same clonal population in single cells.

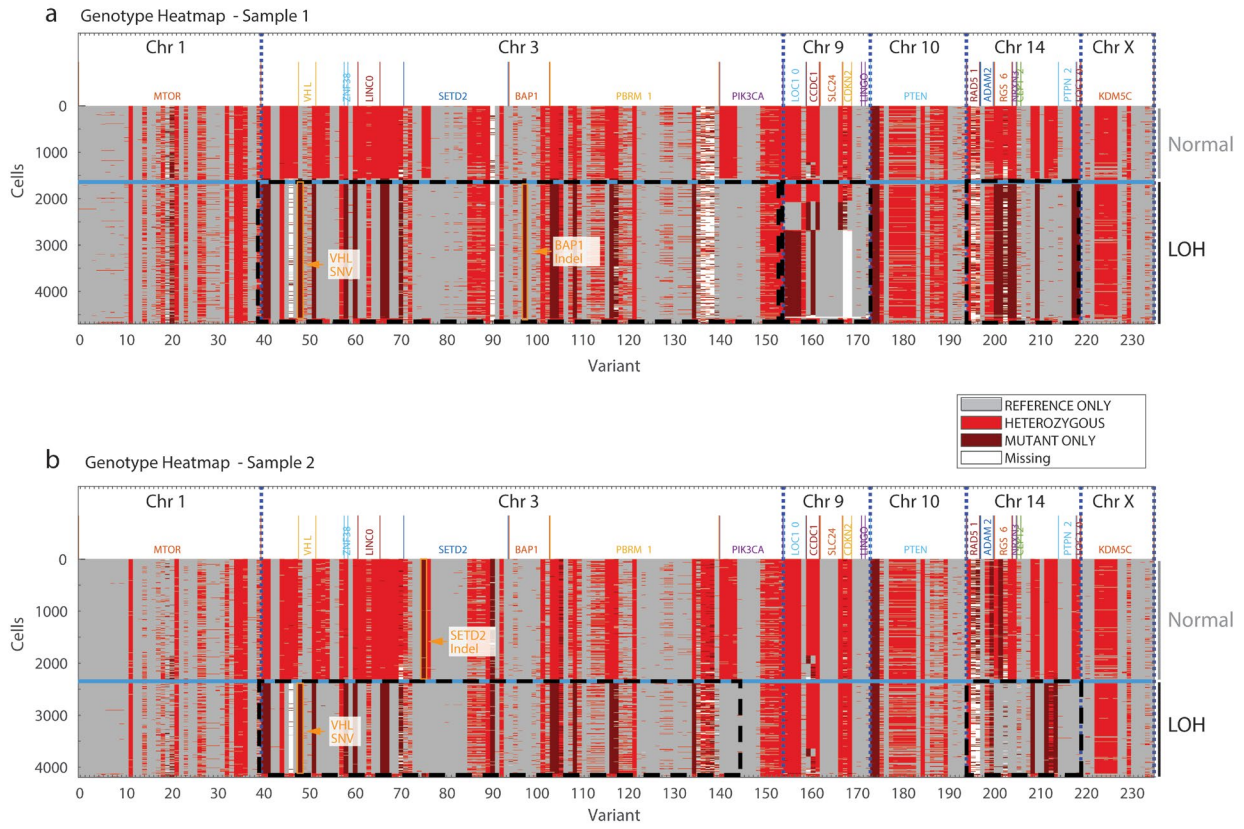



Figure 5: (a-b) Clustering by genotypes shows LOH, SNVs, and indels.

| SNV/Indel or LOH | Sample 1 | Sample 2 |
|---|---|---|
| 3p LOH | YES | YES |
| VHL SNV | YES | YES |
| SETD2 Indel | NO | YES |
| BAP1 Indel | YES | NO |
| 9p LOH | YES | NO |
| 14p LOH | YES | YES |

Table 2: Somatic driver variants found from previously published bulk sequencing data.

## Conclusions

The Tapestri Platform co-detects CNVs and SNVs simultaneously in single cells from cell lines and tumor samples. Using catalog or custom Tapestri Single-Cell DNA Panels, loss or gain of gene copies is reliably determined. Since the same amplicons used to determine SNVs are used to calculate CNVs, researchers may apply the present CNV analysis strategy to previously generated Tapestri single-cell data. Here we present the first commercially available platform that detects CNVs, LOH, SNVs, and indels in the same cells.

## References

1. Tate et al., COSMIC: the catalogue of somatic mutations In cancer. Nucleic Acids Res, 47(D1):D941-D947 (2019)
2. Turajic S et al., Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal, Cell, 173, 595-610 (2018)
3. Toma et al., Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density affymetrix 10K single nucleotide polymorphism mapping array, Neoplasia, 10(7): 634-642 (2008)