Poster #126

Subclonal identification of driver mutations and copy number variations from single-cell DNA sequencing of tumors

James Flynn, Adam Sciambi, Sombeet Sahu, Chrysanthi Ainali, Mani Manivannan, Anup Parikh
Mission Bio Inc

Introduction: Single-cell sequencing elucidates unique insights in understanding intratumor heterogeneity and clonal evolution. Both chromosomal structural change/copy number alteration/variation (CNA/CNV) and driver gene mutation events appear somatically at the early stages of oncogenesis and are critical in cancer initiation, tumor progression and therapy response. Previously, we have developed a high-throughput single-cell DNA analysis platform that leverages droplet microfluidics and a multiplex-PCR based targeted DNA sequencing approach (Figure 1). The platform demonstrates high sensitivity detection of SNVs and indels in the same cells and generation of high-resolution maps of clonal architecture based on mutational profiling.

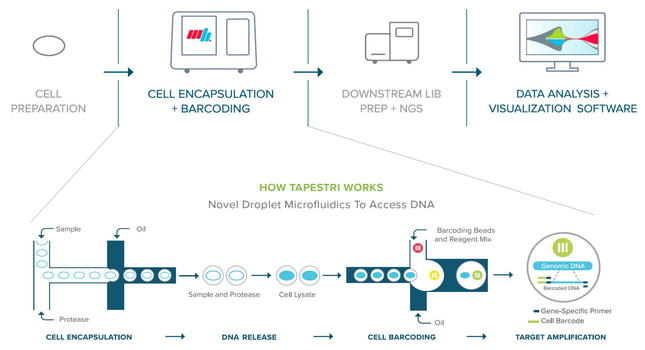

Figure 1: Tapestri Workflow seamlessly integrates into existing workflows

Methods: We developed a dynamic solution to simultaneously characterize point mutations, small indels and gene-level CNVs from the same single-cell. Novel data analysis algorithms detect amplification or loss of function in oncogenes and/or tumor suppressors reliably. Copy number changes are calculated by using Loss Of Heterozygosity (LOH) or the mutation profiles as well as the difference in read information that is gained using a targeted approach. An R package was developed to enable generation of a baseline control population and estimation of the ploidy (Figure 2).

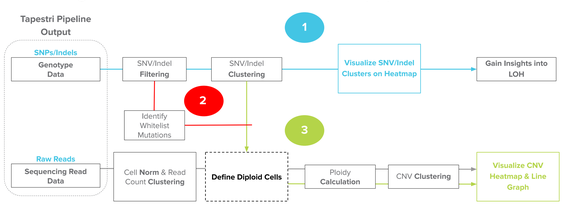

Figure 2: Tapestri R package unlocks CNV information from Tapestri Single-Cell datasets. To infer copy number alterations, there are three approaches: (i) user can use genotype information to infer LOH, define baseline group and then calculate ploidy, (ii) user can identify somatic mutations that define WT clone and then use this to estimate ploidy or (iii) user can use genotype and read information to measure ploidy

Results

We validated this method on clinical samples and admixture samples with cell lines mixed at known ratios. Mutz-8, Raji, K562 and Jurkat cell lines were used with the Tapestri Single-Cell DNA AML. Raw data were first analysed using 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" package in R. The identities of the four cell lines were resolved using unsupervised clustering on the genotype data and then detection of CNVs was done by calculating baseline control and then estimate ploidy.

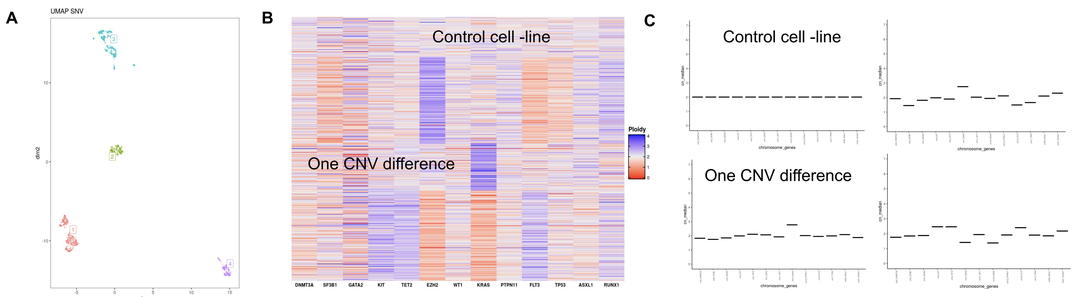

Figure 3: (A) Clustering of SNV data for mixed population of 4 cell lines showed clear clustering of cell lines, (B) heatmap derived from CNV data and (C) CNV plots by gene illustrate the loss or gain of copies for specific genes

In addition, renal cell carcinoma (RCC) samples were analyzed using a 338 amplicon custom panel covering ~67.9 kb. Hierachical clustering and heatmap visualization was performed. Genotypes were plotted for variants present in >5% of cells and were not >99% wildtype. The following steps were applied:

1. **SNV/Indel Filtering:**
   a. apply quality filters (e.g., genotype quality, read depth)
   b. include high-quality synonymous and non-synonymous variants
2. **Clustering**: Hierarchical clustering across cells and heatmap visualisation with chromosomal ordering of variants
3. **Data Interpretation**: Identify normal vs tumor groups. LOH in chr3 and 14

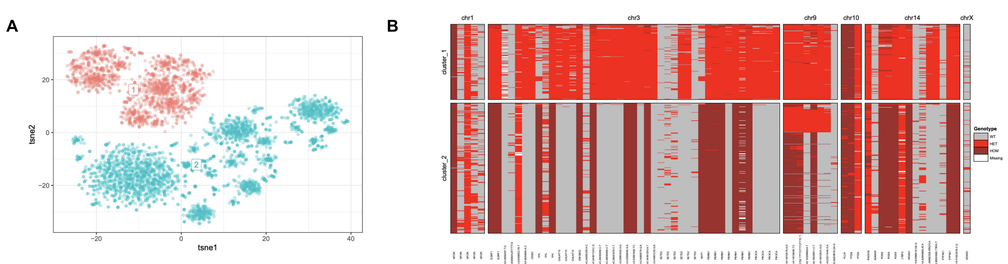

Figure 4: (A) Clustering of SNV data (t-sne plot), (B) heatmap derived from genotype data

From Figure 4, we can define diploid REF cell population by leveraging genotype information from synonymous and non-synonymous variants. Using this reference cell population as baseline group, we calculate ploidy by amplicon median scaling and detect copy number loss or gain across subpopulations

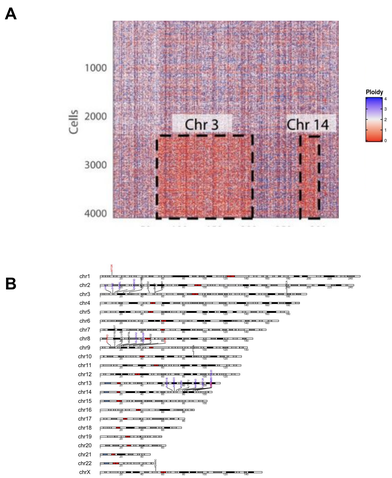

Figure 5: (A) Heatmap derived from read data showing lower number of reads for Chr3 and chr14 (B) karyoplot representing gain or loss for each gene per chromosome. Red color is gain, blue is loss and black is ploidy.

Conclusions:

- Tapestri Platform can simultaneously co-detect CNVs and SNVs in single cells
- Using the R package can easily Calculate gene-level and chromosome-level amplifications and deletions, including LOH
- Integration of CNVs and SNVs facilitates more accurate reconstruction of tumor evolution to better understand cancer progression mechanisms as well for quality control of gene edited cells, to further advance cancer research and therapy.