Abstract #1663

Analytical Methods to Identify Tumor Heterogeneity and Rare Subclones in Single-Cell DNA Sequencing Data from Targeted Panels

Sombeet Sahu[1],Manimozhi Manivannan[1] ,Shu Wang[1], Dong Kim[1], Adam Sciambi[1], Nianzhen Li[1],Nigel Beard[1]
[1]Mission Bio, South San Francisco, CA, USA

Conflicts of interest: S.S., M.M., S.W., D.K., A.S., N.L., N.B. are employees and shareholders of Mission Bio, Inc.

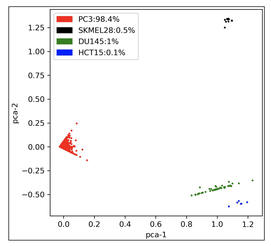
Abstract

Background: With the advancements of single-cell sequencing technologies it is now possible to interrogate thousands of cells in a single experiment. Single-cell RNA-Seq has been available for several years but high-throughput single-cell DNA analysis is in its infancy. Therefore, it is essential to develop new capabilities for assessing genetic variation present in rare cells and to better understand the role that these cells play in the evolution of tumor progression. To address these challenges and enable the characterization of genetic diversity in cancer cell populations, we developed a novel approach to identify mutation signatures which define subclones present in a tumor population.

Methods: Here we present subclone identification method using data generated on the Mission Bio Tapestri® Platform and analyzed by Tapestri analytical workflow. The pipeline steps involve obtaining raw reads from the sequencer, removing adapters, aligning and mapping the reads, calling individual cells, and identifying genetic variants within each cell. After filtering for high quality variants, we then filter for data completeness to ensure only high quality data is used in downstream processing. The variant-cell matrix is then subjected to identification of subclones. Top variants that define the signature of each subclone are also identified. To validate our methodology, we used two different targeted sequencing panels on model systems with known truth mutations. Our pipeline shows the distinct clusters correlating with titration and cell-line ratios. Cluster associated signature mutations were also identified. The pipeline can be used for multi sample analysis with time-series data from diagnosis to relapse or from primary site to metastasis to understand clonal diversity. These data demonstrate the utility of the Tapestri platform, the analytical pipeline, and associated data visualization capability. Our approach addresses key issues of identifying rare subpopulations of cells down to 0.1%, and transforms the ability to accurately characterize clonal heterogeneity in tumor samples. This high-throughput method advances research efforts to improve patient stratification and therapy selection for various cancer indications.

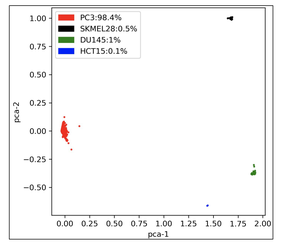
Up to 0.1% High Sensitivity in Subclone Identification and Clonal Signature

A



Latin Square 4 cell lines mixture, AML Panel

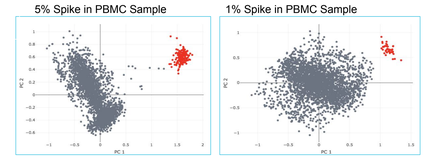
Variants	PC3 (98.6%)	DU145 (0.95%)	SKMEL28 (0.35%)	HCT15 (0.11%)
SORCS3:chr10:106721610	0	0	2	0
TP53:chr17:7577610	0	0	0	2
TP53:chr17:7578115	0	2	0	2
TP53:chr17:7578210	0	0	0	2
KIT:chr4:55599436	2	1	0	1



Latin Square 4 cell lines mixture, THP Panel

Variants	PC3 (98.7%)	DU145 (0.88%)	SKMEL28 (0.4%)	HCT15 (0.04%)
TP53:chr17:7578515	2	0	0	0
JAK3:chr19:17946054	2	0	0	0
SMO:chr7:128846557	2	2	2	0
BRAF:chr7:140453136	0	0	2	0
ERBB2:chr17:37879588	2	0	0	2
TP53:chr17:7578115	0	2	0	2
TP53:chr17:7578210	0	0	0	2
SMAD4:chr18:48586344	0	0	0	2

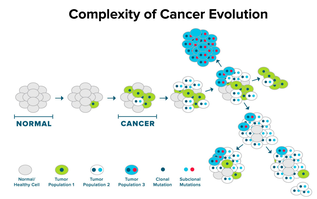
B



Cell Line Spike into PBMC Sample

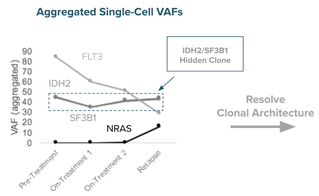
To validate the analytical methods, Latin square 4 cell lines assay was performed using two targeted panels AML and THP. **A).** Distinct clusters assigned by truth variants correlates well with titration and cell line ratios in both AML and THP assays. Rare subclone and clonal signatures were detected at ~0.1%. **B).** Successful cluster and distinguish spiking-in cell line in PBMC sample at 1%.

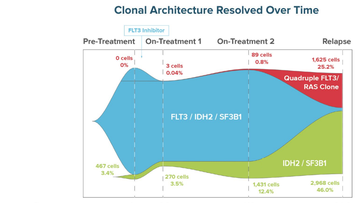
Why Single-Cell ?



Tapestri Workflow Overview

Longitudinal Analysis of AML Samples and Detection of Remission Clones

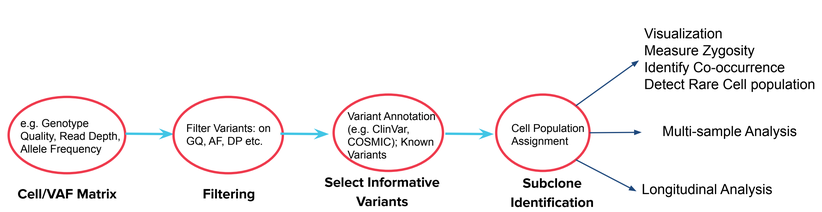




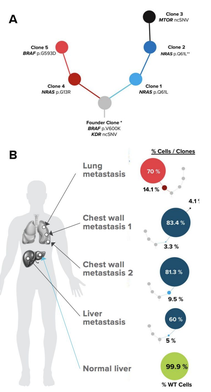
- Expansion of **double-mutant relapse clone** never detected w/bulk NGS
- Acquisition of **RAS clone** detected too late w/bulk NGS

To understand which mutations are real drivers of AML versus ones that are just passengers or contributors, longitudinal analysis using AML panel was performed to reveal the therapy resistant clones. Pre-treatment leukemia sample, on-treatment and relapse BM samples were analyzed by single-cell sequencing. The analysis resolves evolution of 3 subclones basing on combination of 4 mutations, helps to understand clonal composition of cancer for making dynamic changes in treatment.

Primary Analysis and Subclone Identification



Clonal Distribution and Phylogeny Analysis to detect Rare Subclones and Tumor Purity



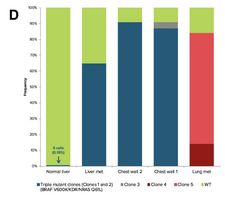
C

Clones	Wild-type Clone	Founder Clone	Clone 1	Clone 2	Clone 3	Clone 4	Clone 5
Tissues containing clone	All	N/A	Liver Chest wall 1 Chest wall 2	Liver Chest wall 1 Chest wall 2	Chest wall 1	Lung	Lung
MTOR ncSNV	WT	WT	WT	WT	HET	WT	WT
NRAS p.Q61L	WT	WT	HET	HOM	HOM	WT	WT
KDR ncSNV	WT	HET	HET	HET	HET	HET	HET
BRAF p.V600K	WT	HET	HET	HET	HET	HET	HET
NRAS p.G13R	WT	WT	WT	WT	WT	HOM	HOM
BRAF p.G593D	WT	WT	WT	WT	WT	WT	HET

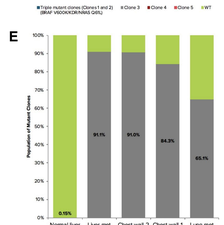
A). A phylogenetic tree illustrates the evolutionary path of each clone. **B).** Clonal distribution in all tissue samples. *founder clone. **Unlike in clone 1, the wild-type allele was not detected in clone 2. **C).** All clones identified across all tissue samples using the Tapestri Platform. The inferred putative founder clone is based on the presence of the mutational pattern in all clones. WT is wild-type for that allele, HET is heterozygous for that mutant allele, and HOM is homozygous for that mutant allele. **D).** Triple-mutant clone distribution in all tissue samples. Rare subclones detected in normal liver at 0.15% (5 cells). **E).** The distribution of mutant cells in all samples (shown in gray) indicate varying levels of tumor purity. Using single-cell genotype data the wild-type cells can be directly detected and removed from downstream analysis.

Francis Crick Institute

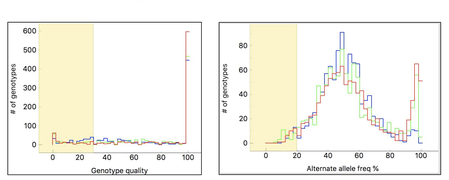
D

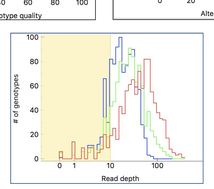


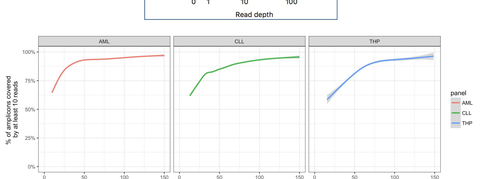
E



Ensuring Data Completeness







Selecting Informative Variants

Keeping Clinically Relevant Variants

Variant	Function	Protein	Coding impact	ClinVar	Y	DANN	Whitelist
ASXL1:chr20:31022959:T/C	coding	ASXL1:p.L815P	missense	not provided		0.31	False
TP53:chr17:7578211:C/T	coding	TP53:p.R174Q	missense	Pathogenic		1.00	False
TP53:chr17:7577581:A/G	coding	TP53:p.Y196H	missense	Conflicting interpretations of pathogenicity		1.00	False
TP53:chr17:7578115:T/C	intronic			Benign		0.65	False

	Variant 1	Variant 2	Variant 3	Variant 4
Cell1	0	0	1	0
Cell2	0	1	1	0
Cell3	0	0	0	0
Cell4	1	1	0	0

Subclones	C1	C2
EZH2:chr7:148504854:A/AGACTT	Het	Hom
TP53:chr17:7577581:A/G	Het	WT
TP53:chr17:7578211:C/T	Het	WT
Total	10557 (77.79%)	3015 (22.21%)
1% K562	6467 (99.11%)	58 (0.89%)
50% K562	4090 (58.04%)	2957 (41.96%)

Tapestri INSIGHTS: RAJI-K562 Cell Line Mix

Results and Conclusions

With the analytical method demonstrated above, we show the distinct clusters correlating with titration and cell line ratio. We also identify the cluster associated signature mutations. These data demonstrate the utility of the Tapestri Platform, and the analytical pipeline, and associated data visualization capability. Our approach has the potential to address the key issues of identifying rare subpopulations of cells and transforms our ability to accurately characterize clonal heterogeneity in tumor samples. This high-throughput method of accurately characterizing clonal populations should lead to improved patient stratification and therapy selection for various cancer indications.

Tapestri Solution and Future Work

Catalog Panels for Hematology and Solid Tumor

TAPESTRI SINGLE-CELL DNA PANELS

	# genes	# amplicons	Target regions coverage (Kbp)	Panel Uniformity
AML Panel	19	50	~8.5	>90%
CLL Panel	34	286	~54.9	>90%
Myeloid Panel	47	330	~66.0	>90%
Tumor Hotspot Panel	59	244	~40.0	>90%

Learn more about Mission Bio at our other posters at ISMB

Poster	Date/Time	Title
#923	Session A	Methods to detect large indels and tandem duplication in acute myeloid leukemia using single cell DNA sequencing
#940	Session B	Detecting doublets in Single Cell DNA-Sequencing using Deep Learning
#956	Session B	Error Correction in single-cell DNA sequencing: Finding that rare allele for MRD clone
#982	Session B	Using machine learning to optimize assays for single cell targeted DNA sequencing
#1167	Session B	Analytical Methods to Identify Tumor Heterogeneity and Rare Subclones in Single Cell DNA Sequencing Data from Targeted Panels