Single-cell Simultaneous Detection of DNA Genotype and Protein Expression

Aik Ooi[1], Pedro Mendez[1], Dalia Dhingra[1], Ben Demaree[1,2], Cyrille Delley[2], Adam Abate[2], Linde Miles[3], Ross Levine[3], Shu Wang[1], Manimozhi Manivannan[1], Adam Sciambi[1], Nianzhen Li [1], Nigel Beard[1], and David Ruff[1]

[1]Mission Bio, Inc., South San Francisco, CA. [2]University of California, San Francisco, Department of Bioengineering and Therapeutic Sciences, San Francisco, CA. [3]Memorial Sloan Kettering Cancer Center, Human Oncology & Pathogenesis Program, New York, NY.

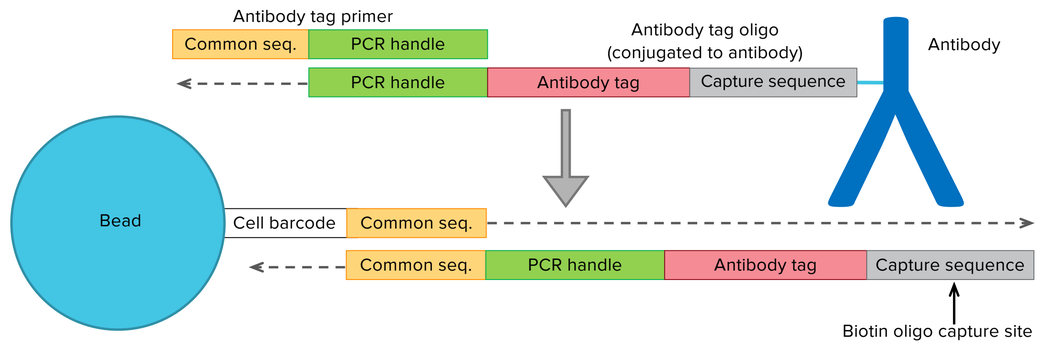
Introduction

Recent advancements in genomic analysis of tumors have revealed that cancer disease evolves by a reiterative process of somatic variation, clonal expansion and selection. Therefore, intra- and inter-tumor genomic heterogeneity have become a major area of investigation. While next-generation sequencing has contributed significantly to our understanding of cancer biology, the genetic heterogeneity of a tumor at the individual cellular level is masked with the average readout provided by a bulk measurement. Very high bulk sequence read depths are required to identify lower prevalence mutations. Rare events and mutation co-occurrence within and across select population of cells are obscured with such average signals. The emergence of single-cell technologies has greatly refined the resolution in which sample diversity can be investigated, enhancing the efficiency of selecting appropriate molecular targets.

Using a two-step droplet microfluidic technology, the Mission Bio Tapestri® Platform enables multiplex-PCR based high-throughput targeted DNA sequencing in single cells. The workflow unlocks access to gDNA, and has been utilized to investigate tumor heterogeneity, disease clonal evolution, and therapy resistance mechanisms

By leveraging this technology, a new workflow is developed to detect protein expression in addition to DNA genotype in the same single cells. In this approach, cells are labeled with a pool of oligonucleotide-conjugated antibodies prior to loading the cells into the Tapestri Instrument for targeted DNA analysis. Sequencing libraries are then prepared from both antibody oligonucleotides and the amplified DNA sequences, followed by identification of single-cell DNA genotypes and protein signatures from the sequencing readout. In a mixed population of three cell lines, single nucleotide variations (SNVs) and protein data independently classified the cells into appropriate clusters.

We believe that this novel multi-omic technology enables high dimensional studies of the complex relationship between genotype and phenotype, leading to a better understanding of disease biology, and subsequently better design of diagnostics and therapies.



Simultaneous measurement of cell surface proteins and genotypes in single cells

Tapestri Single-Cell DNA AML V2 Panel

ASXL1	GATA2	KIT	PTPN11	TET2
DNMT3A	IDH1	KRAS	RUNX1	TP53
EZH2	IDH2	NPM1	SF3B1	U2AF1
FLT3	JAK2	NRAS	SRSF2	WT1

Single-Cell Protein Targets

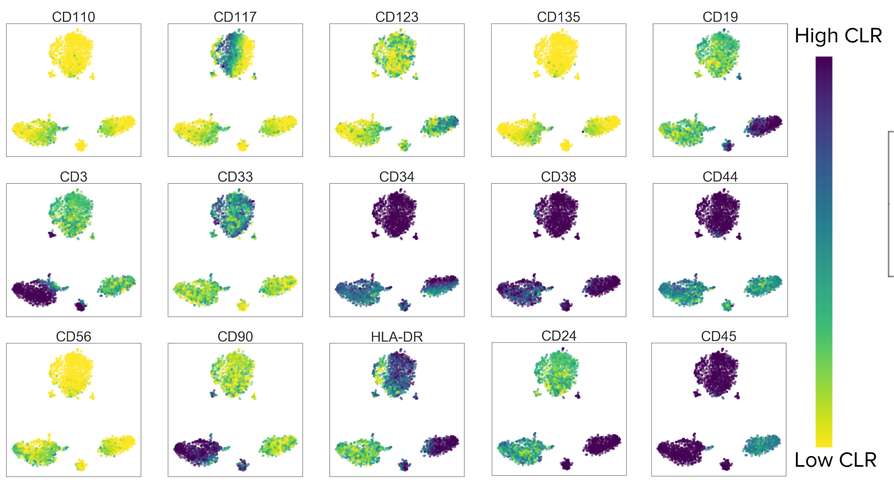
CD3	CD19	CD24	CD33
CD34	CD38	CD44	CD45
CD56	CD90	CD110	CD117
CD123	CD135	HLA-DR	IgG1κ

A mixed population of KG-1, TOM-1, and Jurkat cells were treated with a pool of 15 oligonucleotide-conjugated antibodies of interest plus mouse IgG1κ antibody that served as negative control. Cells were then washed and loaded onto the Tapestri Platform to be analyzed with the Single-Cell DNA AML V2 Panel (128 amplicons covering 20 genes).

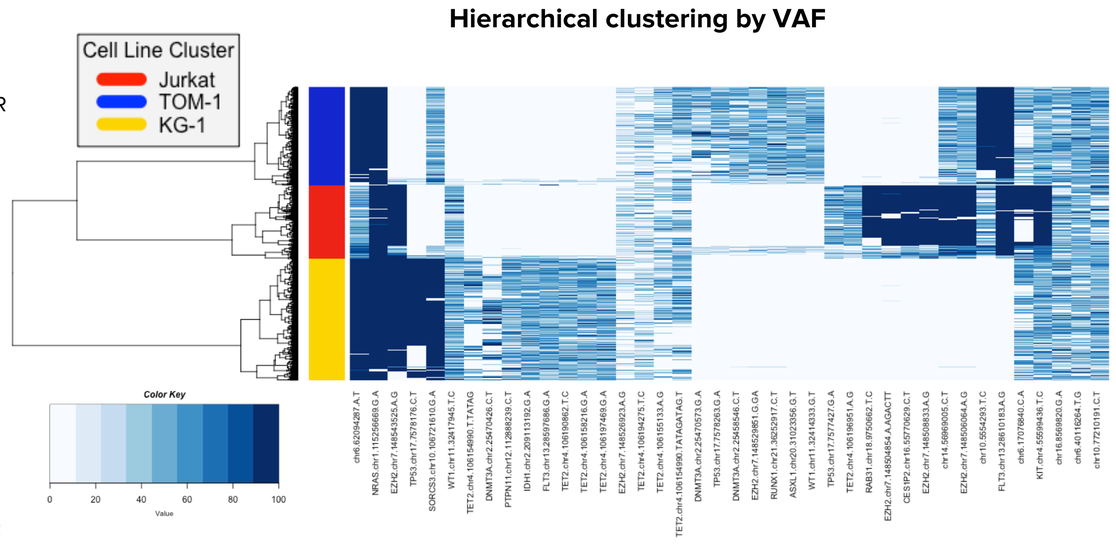
Antibody tag counts were normalized using centered log ratio (CLR) transformation. A t-SNE plot was generated using the CLR values from all protein targets. Single-cell protein marker expression data independently clustered the cells into groups that matched up with the cell genotype data.

Sequencing data for DNA genotype was processed with the Tapestri Pipeline software and further analyzed with the Tapestri Insights software to determined SNVs. Unsupervised hierarchical clustering was performed based on the variant allele frequency (VAF) values.

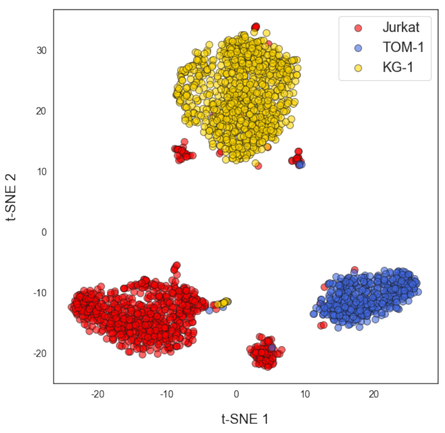
Expression level of each protein represented by CLR-transformed values



Hierarchical clustering by VAF



t-SNE by protein expression; cell color by genotype

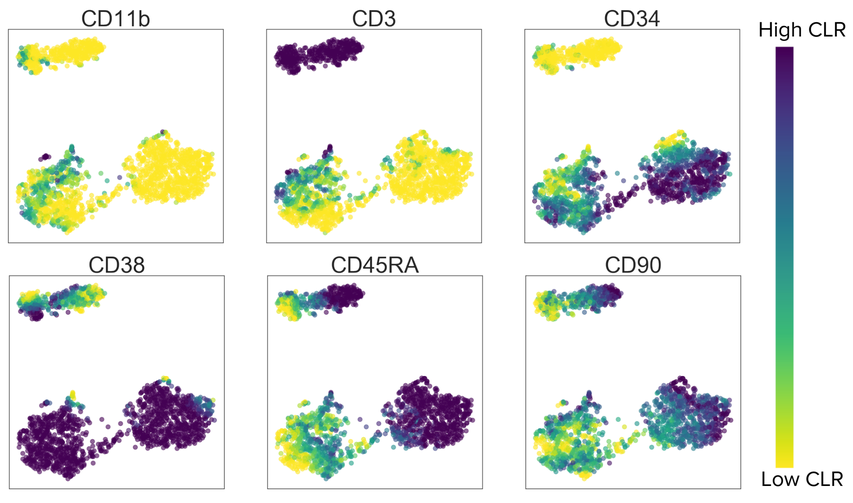


Single-cell protein detection and DNA genotyping in clinical samples

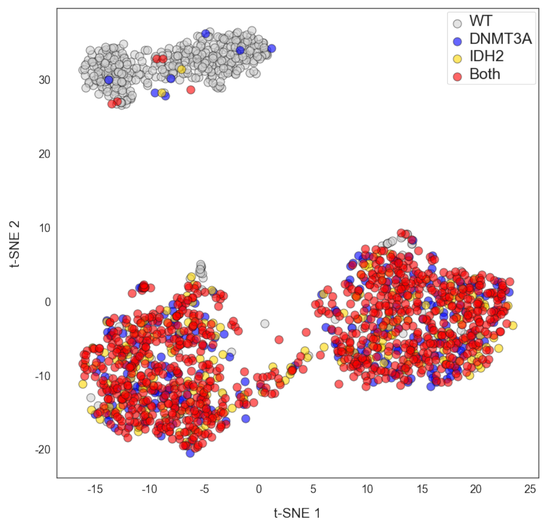
To extend this novel technology beyond cell lines, we applied the workflow to samples with various myeloid malignancies. Samples were first flow-sorted based on viability. A panel of oligonucleotide-conjugated antibodies of 6 cell surface protein markers was used to label the cells. Concurrently, targeted DNA sequencing was performed using a custom-designed panel consisting of 109 amplicons in 50 genes, covering the most frequently mutated regions in myeloid malignancies.

Sequencing data processing and analysis for targeted DNA genotyping were performed based on standard Mission Bio Tapestri Pipeline to identify variants. Antibody tag counts were normalized by CLR transformation. A t-SNE plot was generated using CLR values from all 6 antibody tags to visualize clustering of cells based on cell surface markers, which represent the immunophenotypes of the cells.

Expression level of each protein represented by CLR-transformed values



t-SNE by protein expression; cell color by genotype



Data from one myeloid malignancy sample are presented here. Color gradient based on CLR values for each protein marker was projected on t-SNE plots where cells were clustered based on the expression level of 6 cell surface protein markers (left panel). The cell surface markers clearly categorized the cells by their immunophenotypes.

The DNA genotype analysis identified cells harboring *DNMT3A* or *IDH2* mutations, or both. The mutation information was color coded and projected on the protein t-SNE plot to simultaneously visualize both sets of results (right panel). We found that the CD3+ T cells population did not carry the mutation for either of these genes. Interestingly, two cell clusters identified by protein markers could not be distinguished by DNA genotypes, indicating sample heterogeneity that could not be resolved by protein or DNA information alone.

Summary

- **The multi-omic results generated from this novel workflow allow for high dimensional data analysis of protein expression and SNV from the same single cells.**
- **Expression data from 15 cell surface proteins and SNV information successfully classified cells from a mixture of 3 cell lines into appropriate groups.**
- **This technology has been successfully performed on clinical samples with myeloid malignancies, revealing both genotype and phenotype information.**
- **This is a breakthrough in single-cell technology that will enable more in depth and complex studies of the relationship between genotype and phenotype.**