Abstract #833

Linking phenotype and genotype: multimodal analysis of surface proteins, intracellular proteins, and SNVs in single cells

Prithvi Singh, Saurabh Parikh, Tim Lai, Dalia Dhingra, Adam Sciambi, and **Aik Ooi**. Mission Bio, Inc. South San Francisco, CA, USA.

Introduction

Cancers evolve via processes of clonal expansion, selection, and somatic variation, which contribute to tumor heterogeneity. While bulk analysis has improved our understanding of cancer, the heterogeneity of a tumor is masked with the average readout provided by a bulk measurement. This problem could be overcome by employing single-cell technologies.

One of the most common single-cell technologies is single-cell RNA sequencing (scRNA-seq). Transcriptomic data collected by scRNA-seq are used to imply protein expression levels based on RNA. Signaling pathways related to cancer initiation and progression have been mapped solely based on scRNA-seq results. This approach, however, is not an accurate measurement of proteins that are the functional molecules responsible for key events in cancer. Recent breakthroughs of multimodal approaches such as CITE-seq and REAP-seq partially addressed this problem by allowing simultaneous analysis of transcriptomes and surface protein expression levels. These methods still lack a couple components that are required to fully dissect the biology of cancer. First, the methods are limited to the measurements of cell surface proteins, whereas proteins involved in cancer-related signal transduction and transcriptional pathways are mostly localized in the intracellular compartments. Second, the methods do not allow direct analysis of DNA sequences, hence unable to provide an accurate readout of genotypic information such as single-nucleotide variants (SNVs) and copy number variations (CNVs).

Here we describe a technology that could overcome these hurdles. To achieve this, cells are first treated with oligonucleotide-barcoded antibodies targeting surface proteins. Cells are then fixed and permeabilized, followed by incubation with barcoded antibodies for intracellular proteins. The resulting cells are processed on the Mission Bio Tapestri® platform, a device that enables encapsulation of single cells in droplets. Targeted DNA sequencing libraries are generated from the single cells using a multiplex panel of primers targeting regions of interest. Protein sequencing libraries are separately generated from the oligonucleotides off the antibodies.

The inclusion of intracellular protein detection enables measurement of pivotal proteins in cancer mechanisms, including apoptosis (BCL2 family proteins), transcription factors (GATA3), tumor suppressors (TP53), as well as phosphorylated proteins involved in cell growth signaling pathways (phosphorylated ERK, STAT3 proteins). This is the first ever method that provides a solution to effectively link surface and intracellular protein measurement with targeted DNA analysis. With this approach, single-cell readout of genotypic and phenotypic information can be collected together, allowing concurrent analyses of cancer clonal evolution and driver protein expression.

Tapestri platform and reagent kit

CONTACT:
Aik Ooi
Mission Bio, Inc.
ooi@missionbio.com

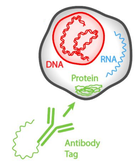
The multi-omic workflow

The Tapestri platform is able to simultaneously analyze both genotypic (SNVs, indels, and CNVs) and phenotypic (proteins) factors from thousands of cells individually. This comprehensive, single-cell analysis is performed by a series of analyte-specific processing steps to generate associated DNA sequences: proteins are tagged with oligonucleotide-conjugated antibodies, and gDNA in the nucleus is freed from chromatin via protease digestion. The resulting DNA sequences can then be barcoded with a cell-specific sequence for later identification.

Mission Bio enables single-cell DNA and protein detection.

Multi-omic strategy:

- Tag proteins with oligonucleotide-barcoded antibodies.
- Free DNA from nucleus.
- Barcode all.



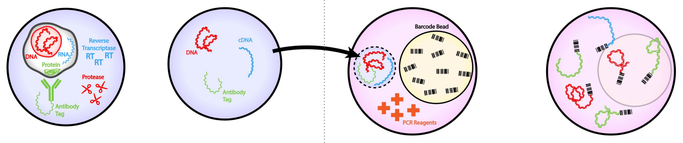
The workflow is enabled by its two-droplet steps (below). Cells are individually encapsulated in a first droplet (**1A**) where analyte preparation can occur, including aggressive digestion and subsequent protease heat inactivation (**1B**). Afterwards, sensitive enzymes that are incompatible with the earlier preparation can be added to make a new drop (**2A**) where barcoding and amplification proceed (**2B**).

1A Encapsulate cells in droplets

1B Prepare analytes, deactivate proteases

2A Add barcodes, polymerase

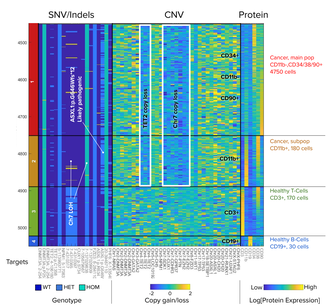
2B Barcode and amplify



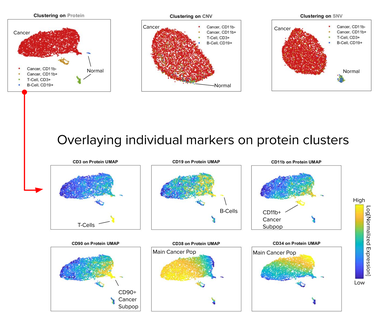
An AML sample: SNVs, copy loss, loss of heterozygosity, and surface protein expression

In this experiment, an acute myeloid leukemia (AML) sample is analyzed and 5000+ cells are recovered. The sample is found to be 95% cancerous as identified by any of the three analytes. Most striking is the Chromosome 7 copy loss, a known pathologic mechanism of AML, as detected in CNV and orthogonally validated by loss of heterozygosity in the SNV data (see heatmap, **3A**). In addition, the cancerous population is identified by a pathogenic mutation (ASXL1 p.G646V) and higher expression of CD11b, CD34, CD38, and CD90 in two sub populations (**3B**).

3A Heatmap: cell types vs. SNV, CNV, and protein



3B Clustering on protein, CNV, and SNV



Enabling simultaneous analysis of DNA with both cell surface and intracellular proteins

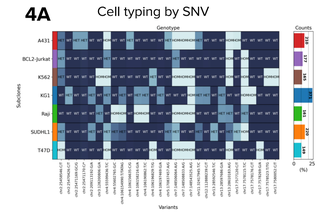
To perform surface and intracellular protein analysis, cells are first incubated with barcoded antibodies for surface proteins followed by cell washing to remove unbound antibodies. Cells are then fixed, blocked, and permeabilized before they are incubated with barcoded antibodies targeting intracellular proteins. After washing away the unbound antibodies, cells are introduced into the Tapestri instrument where single cells are encapsulated into individual emulsion droplets. Targeted DNA and protein tag amplification occur in the droplets, and cell identification barcode is added to each amplicon. Protein libraries are separated from the DNA libraries by size and prepared separately.

In the following example (**4A**–**4E**), a mixture of 7 cell lines are analyzed using the Tapestri Single-Cell DNA AML Panel, which surveys 127 amplicons across 20 genes. TotalSeq™-D Heme Oncology Cocktail from BioLegend, which contains 45 antibodies including 3 isotype controls, is used for the cell surface protein detection. The TotalSeq-D panel is designed specifically for the Tapestri platform. As a proof of concept, 5 intracellular protein targets are included in the experiment.

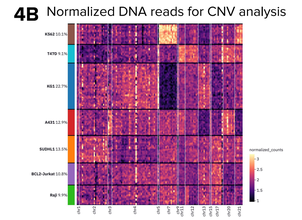
Using SNV analysis from the DNA data, cells are identified based on their genotypic signatures (**4A**). Normalized read counts from the DNA data can also be analyzed to infer CNV information (**4B**). In this example, BCL2-Jurkat is used as the diploid reference for the analysis. The read depth analysis shows that there is a copy number gain in chromosome 7 for K562 cells, and a copy number loss in the same chromosome for KG1 cells.

UMAP analysis is performed on normalized protein reads, and the resulting plot is colored based on the cell type identified by SNV (**4C**). To visualize differential protein expression among cell types, expression profiles are displayed in the UMAP plots for each protein target (**4D**), or in a heatmap (**4E**). In **4D**, 36 out of 50 protein targets are plotted; in **4E**, 37 out of 50 protein targets are plotted.

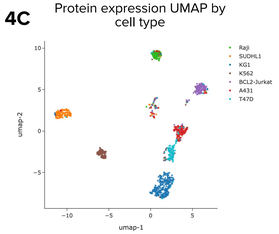
4A Cell typing by SNV



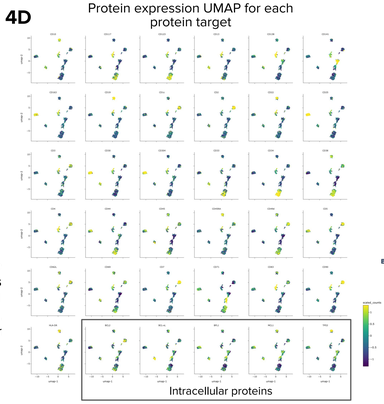
4B Normalized DNA reads for CNV analysis



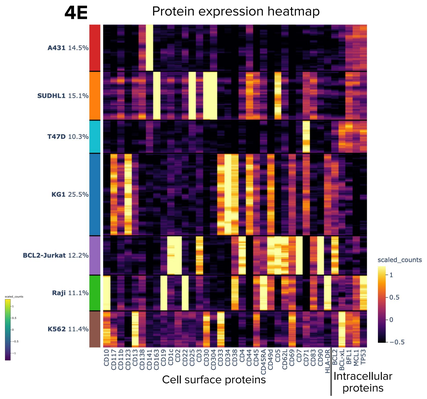
4C Protein expression UMAP by cell type



4D Protein expression UMAP for each protein target

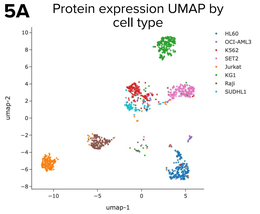


4E Protein expression heatmap

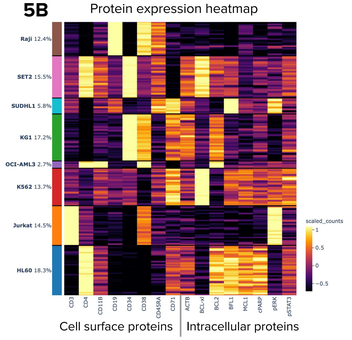


To further validate the capability of this method, we show that it is possible to apply the workflow on post-translationally modified proteins, such as phosphorylated proteins (phospho-ERK, phospho-STAT3) and cleaved proteins (cleaved PARP), as shown in **5A**–**5B**. In the final example (**6A**–**6B**), the workflow is tested on PMBC along with cell lines, targeting nuclear proteins (GATA3, TP53) and more phosphorylated proteins (phospho-AKT, phospho-RPS6).

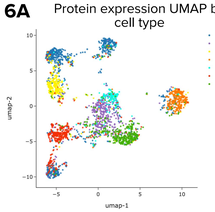
5A Protein expression UMAP by cell type



5B Protein expression heatmap



6A Protein expression UMAP by cell type



6B Protein expression heatmap

Conclusions

Utilizing the Mission Bio Tapestri platform, we successfully implemented concurrent SNV, CNV, surface protein, and intracellular protein analyses from single cells. The ability to perform this single-cell multi-omic analysis enables researchers to precisely pair DNA data (information on driver mutations, clone tracing, drug resistant clones, and CNVs) with intracellular protein expression (information on signal transduction pathways, cell death mechanisms, transcription factors, tumor suppressors, oncoproteins, and drug targeted proteins). This novel technology will provide a solution to properly couple genotype and phenotype relevant to cancer biology, which is not possible by most currently available multi-omic methods of surveying RNA and cell surface proteins.