2259

Single-cell multi-omic analysis of SNV, CNV, and protein expression

Aik Ooi[1], Dalia Dhingra[1], Adam Sciambi[1], Kate Thompson[1], Jackie Marin[1], Saurabh Parikh[1], and David Ruff[1].
[1] Mission Bio, South San Francisco, CA, USA.

Introduction

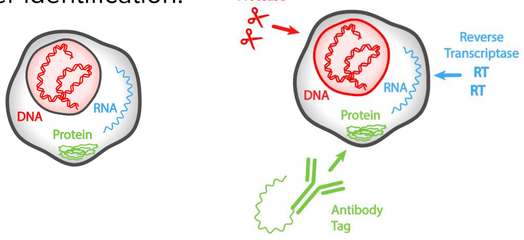
Recent advancements in precision medicine, while highly promising, presents a major technical challenge to researchers due to disease heterogeneity. The emergence of single-cell technologies has greatly refined the resolution in which sample diversity can be investigated, enhancing the efficiency of selecting appropriate molecular targets. Additionally, applying multi-omics analysis on single cells would further improve the understanding of cell-to-cell heterogeneity by providing unique insights on cellular and genetic composition. Using a two-step droplet microfluidic technology, the Mission Bio Tapestri Platform enables multiplex-PCR based high-throughput targeted DNA sequencing in single cells to obtain single-nucleotide variation (SNV) and copy number variation (CNV) information. 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. 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. The number of protein targets can be in the range from a few targets to over 40, which is beyond the limit for a single flow cytometry run. This method has been successfully performed on cell lines, fresh and frozen PBMCs, as well as clinical samples. In an acute myeloid leukemia (AML) sample, combined single-cell SNV, CNV, and protein expression data illustrated the heterogeneity within the sample. The data clearly identified CD3+ T cells and CD19+ B cells without pathogenic SNVs and CNVs. $CD34^{hi}CD11b^{lo}$ and $CD34^{lo}CD11b^{hi}$ subpopulations were also identified within the cells carrying the same pathogenic SNVs and CNVs. We believe that this novel multi-omic technology will facilitate new discoveries in the complex relationship between genotype and phenotype, enable a better understanding of disease biology, and subsequently improve the design of diagnostics and therapies.

The multi-omic workflow

The Tapestri Platform is able to simultaneously analyze both genotypic (SNV, Indels, and CNV) and phenotypic (RNA and Protein) 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 DNA-conjugated antibodies, RNA is reverse transcribed into cDNA, and gDNA in the nucleus is freed from chromatin via protease. The resulting DNA sequences can then be barcoded with a cell-specific sequence for later identification.

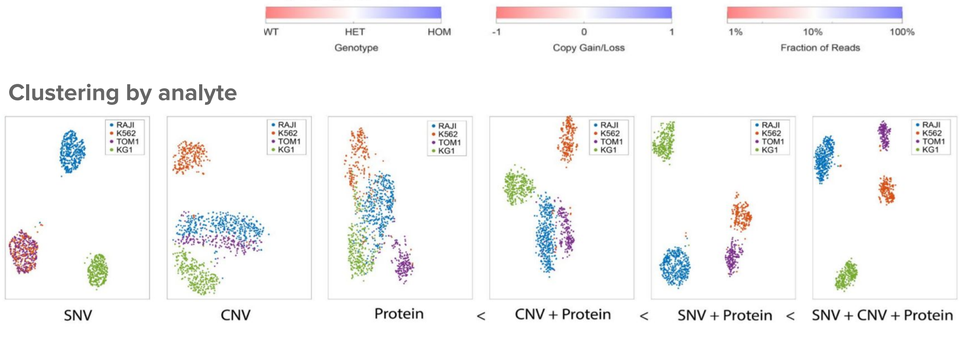
Multi-omic strategy:

- Tag proteins with antibodies.
- Reverse transcribe RNA.
- 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).

The two-step workflow



The power of single-cell multi-omics

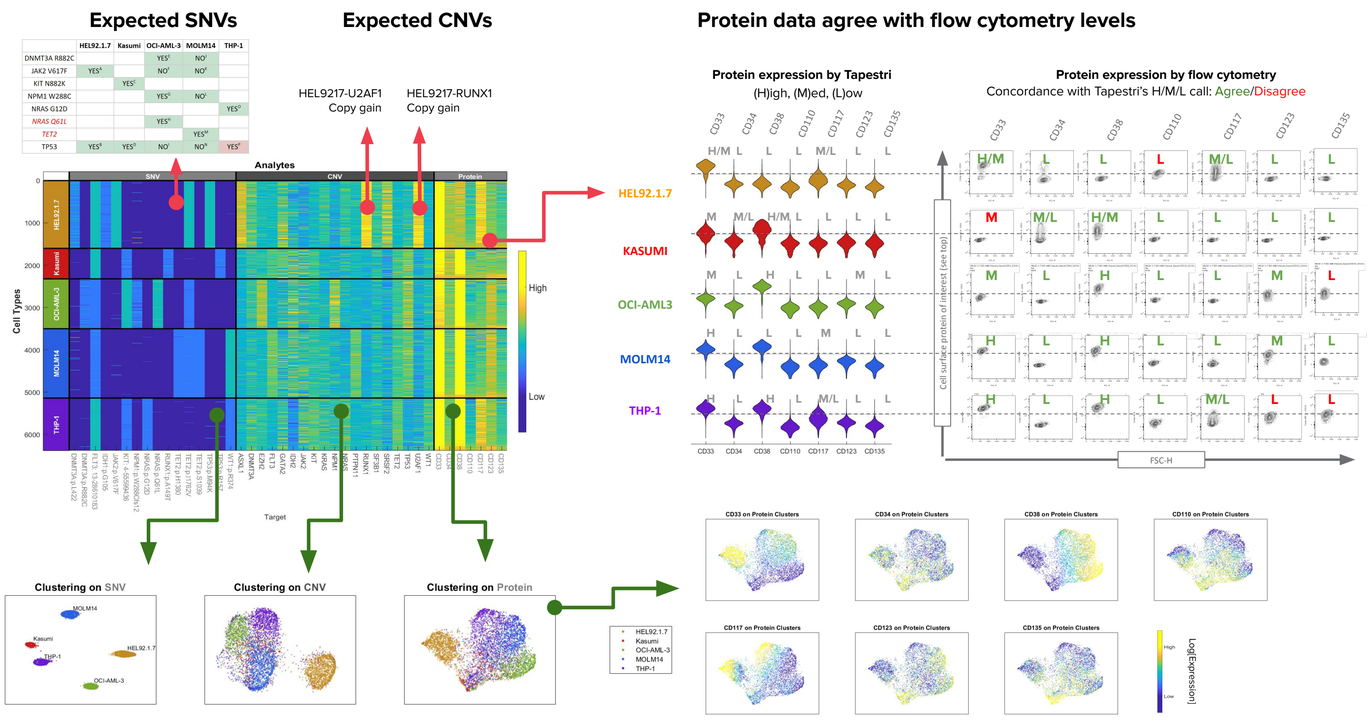
Multi-omics tell a complete biological story. Any particular analyte might contain a critical disease signature that would otherwise be missed. Taken together, multiple analytes can show how a mutation or copy change directly alters RNA and protein expression. Yet another benefit is the extra power to resolve subpopulations. As an example of the latter, the plots below show four cell lines that have been combined in one run. Single-cell SNV, CNV, and protein clustering do not fully reconstruct the four subtypes, but together they can.

Raw single-cell data: SNV, CNV, protein

Clustering by analyte

Validating the platform

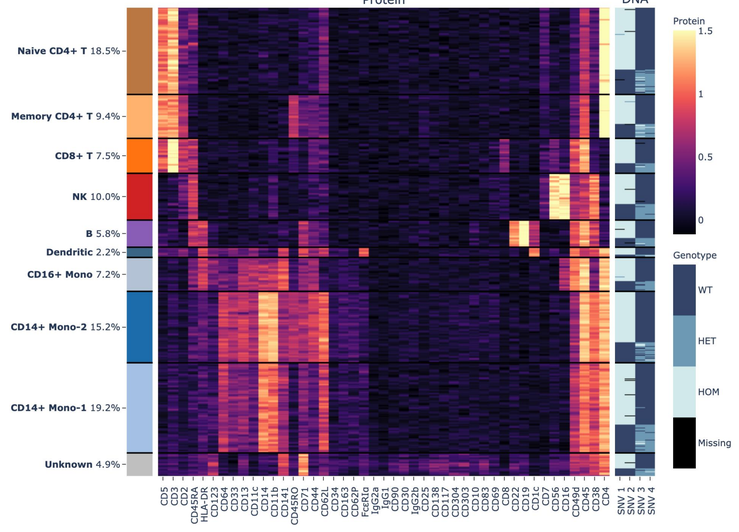
To demonstrate the accuracy of the system, many well known samples have been run and found to have concordance with literature. Given below is a 6000-cell customer Tapestri run where five cell lines were combined and evaluated against a targeted SNV, CNV, and Protein panel. From the raw data (center heatmap), the customer validated 15/16 expected SNVs, 2/2 expected CNVs, and 30/35 protein levels measured on those same cell lines with flow cytometry (above, right). At the same time, each set of analytes could be investigated de novo (below in green) to see how they organize. A product of that analysis is a mapping of individual protein expression on top of panel-wide protein clustering to resolve cell lines (bottom right).



Protein and DNA analysis on PBMCs

To identify the multiple cell types in PBMCs, a higher number of protein targets is needed. Here we performed single-cell analysis on a mixed population of PBMCs from two different individuals using a 45-plex protein target panel and 312-plex DNA target panel. The protein expression data clearly resolve the major cell types and their subtypes within PBMCs, whereas the SNV data are used to differentiate the two PBMCs donors.

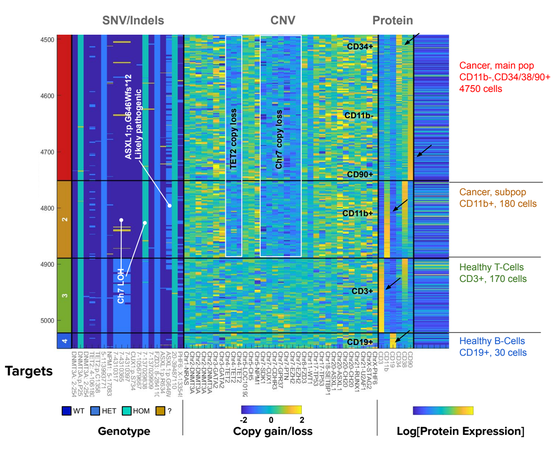
Heatmap: cell populations resolved by protein expression; cell donors differentiated by SNVs



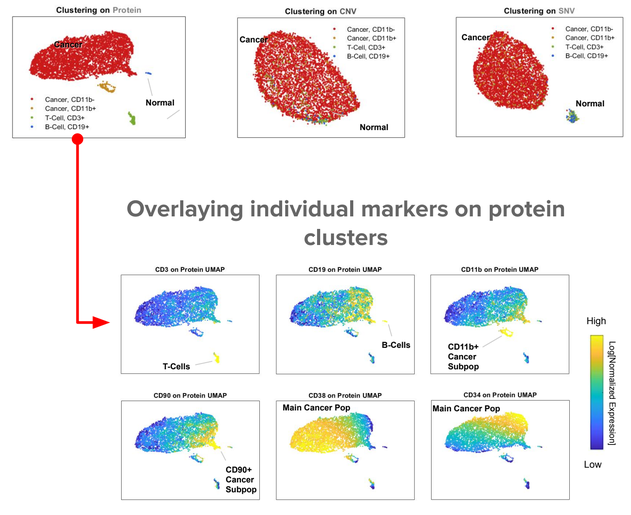
An AML patient sample: SNVs, CNV, loss of heterozygosity, and protein expression

In this experiment, an AML patient sample was analyzed and 5000+ cells were recovered. The sample was found to be 95% cancerous as identified by any of the three analytes. Most striking was 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, lower left). In addition, the cancerous population was identified by a pathogenic mutation (ASXL1 p.G646V) and higher expression of CD11b, CD34, CD38, and CD90 in two sub populations (lower right.)

Heatmap: cell types vs. SNV, CNV, protein



Clustering on SNV, CNV, and protein



Conclusions

Within a heterogeneous sample various populations are defined by inherent patterns, which may be resolved by both genotype and phenotype at single-cell resolution. Taking a single-cell multi-omic approach ensures the coverage of several parameters to capture the most relevant characteristics in defining critical populations and understanding cellular biology. Our data illustrate the power of integrated single-cell genotype and phenotype data to differentiate cell types and subpopulations, and strengthen the conclusions made from individual analytes alone.