mission bio

tapestri
APPLICATION NOTE

Single-cell Multi-omics Enables Granular Resolution Into AML For Powerful Insight Into the Correlation Between Genotype and Phenotype

Takeaways

- Tapestri Platform is the only single-cell multi-omics platform capable of simultaneously detecting SNVs, CNVs, and protein expression from the same single cell
- Bulk sequencing, cell-line databases, and flow cytometry orthogonally verified Tapestri multi-omics results
- The Tapestri Platform enables in-depth multi-omics analysis of cell lines and blood samples

Abstract

Modern medicine has given rise to an array of treatment options for diseases such as acute myeloid leukemia (AML). Cancer is a heterogeneous mixture of cells with varied states, and the best treatment options to an individual patient requires an understanding of the disease state at the cellular level. A single-cell multi-omics approach is the only way to achieve full resolution of the disease at the cellular level, revealing the interplay between genotype and phenotype. Here, we show that the Tapestri Platform enables comprehensive identification of cell subpopulations through single-cell multi-omics profiling. The platform discerned single nucleotide variants (SNVs), copy number variations (CNVs), and cell surface protein expression in single cells from a mixture of AML cell lines and an AML research sample. Our results demonstrate that subpopulations are not consistently defined by one genetic or phenotypic factor alone, but by multiple parameters in conjunction and irretrievable by bulk sequencing methods. We also show that multi-omics data generated by the Tapestri Platform are consistent with the gold-standard techniques currently used for traditional omics analysis of AML samples. Taken together, our results demonstrate the power of the Tapestri Platform to uncover and define various cell states of AML using single-cell multi-omics.

Experiment & Methods

In collaboration with Dr. Saar Gill's laboratory at the University of Pennsylvania, THP-1, MOLM14, OCI-AML-3, Kasumi, and HEL92.1.7 AML cell lines were mixed together at equal ratios to simulate a heterogeneous AML sample and analyzed for SNVs, indels, CNVs and proteins using the Tapestri Platform. Cells were processed on the Tapestri Platform to simultaneously access protein expression using a panel of 7 antibodies conjugated to analyte barcoded oligo tags. The targets consisted of CD33, CD34, CD38, CD110, CD117, CD123, and CD135.

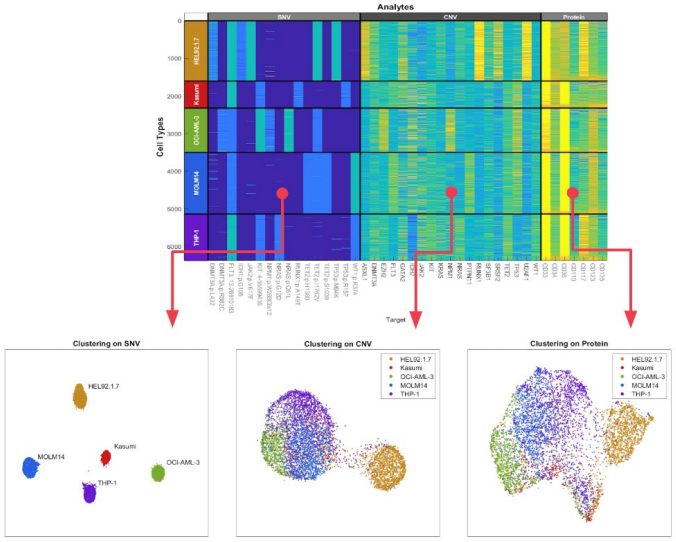

Figure 1 - Heat map and t-SNE plots derived from SNV, CNV, and protein data of a mixed population of five AML cell lines. t-SNEs were colored in by SNV genotypes previously established with pure cell lines.

Next, in collaboration with Dr. Chris Hourigan's laboratory at the National Heart, Blood, and Lung Institute, an AML research sample was analyzed on the Tapestri Platform for DNA and protein. The sample was analyzed with a Tapestri Custom DNA panel kit containing 45 amplicons targeted to genes and a custom oligo-conjugated antibody panel of 6 proteins: CD3, CD11b, CD19, CD34, CD38, and CD90. The custom DNA panel included coverage of chromosome 7, known to have CNVs in AML. Data were analyzed with Tapestri Pipeline software. SNVs and indels were identified using Tapestri Insights software, CNVs were analyzed using the Mission Bio "tapestri-cnv" package for R, and DNA and protein data were integrated and analyzed using the Mission Bio "tapestri-protein" package for R.

Results

Single-cell multi-omics of mixed AML cell lines reveals variable clustering on different analytes

After running the AML cell lines at equal ratios, a heat map and t-SNE projections of SNV, CNV and protein data showed varied resolutions of the 5 cell lines (Figure 1). Genomic SNV data resulted in very distinct clustering of the cell lines, which was expected based on the differences in the known genotypes of the cell lines. Clustering with phenotypic CNV or protein data alone resulted in much less resolved subpopulations, illustrating that analysis of only one analyte may not be enough information to resolve a heterogeneous population of cells. Some samples may contain cells with very similar genotype, but different protein expression patterns. Therefore, Tapestri single-cell multi-omics provides a powerful solution to efficiently resolve subpopulations within a heterogeneous sample.

Tapestri multi-omics results verified using gold-standard pathology methods

To demonstrate the reliability of Tapestri single-cell multi-omics, we compared Tapestri results with flow cytometry, the current gold-standard methods used to analyze AML samples. Analysis of the AML cell line mixture by multi-parameter flow cytometry showed high correlation with Tapestri results for the same proteins (Figure 2). Bulk sequencing of the cell mixture also showed correlation with Tapestri Platform results with 15 out of 16 mutations identified by single-cell analysis. To validate CNV results, we consulted the Cancer Cell Line Encyclopedia (CCLE)[1] and found that single-cell analysis was able to identify known copy gains of RUNX1 and U2AF1 genes in HEL92.1.7 cells. Taken together, these results demonstrate that SNV, CNV, and protein data generated using the Tapestri Platform correlates well with established gold-standard orthogonal methods.

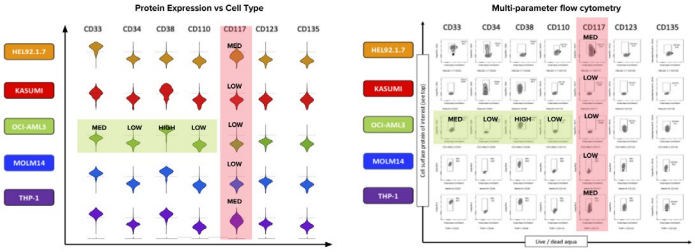

Figure 2 - Comparison of protein expression of AML cell line mixture, as determined by the Tapestri platform (left) and flow-cytometry (right), illustrating high corroboration of results.

AML subpopulations identified through single-cell protein analysis, unresolved through SNV or CNV

With the AML research sample, in total, over 5,000 cells were sequenced and analyzed.

Clustering by SNV, CNV, or protein data on t-SNE projections had variable resolution of subpopulations (Figure 3a). Clustering by protein gave the highest resolution, distinguishing four subpopulations, two of which were normal T-cells (0.6%) and B-cells (3.4%) positive for CD3 and CD19, respectively. The other two subpopulations were cancerous with one positive for CD11b (3.6%) expression and the other negative (95%). Interestingly, clustering by SNV data only distinguished two subpopulations, normal and cancerous, illustrating that the genotype may not correlate directly with phenotype. Also, clustering by CNV data showed limited clustering even though heat map data confirmed copy number losses of several genes in the cancerous subpopulations (Figure 3b).

Further analysis revealed that wild type protein markers (CD3, CD19) positively correlated with a wild type genotype, and that the hematopoietic lineage markers CD11b and CD34 correlated with a cancerous genotype (Figure 3b). We also identified ASXL1 G646V as a heterozygous SNV in the cancer cell populations, which is likely to be pathogenic.

Because we were able to identify normal and cancerous populations from the protein expression data, we were then able to quantify CNVs among the population (Figure 3b). In AML, deletions in chromosome 7 are a recurrent feature contributing to the pathology of the disease. We find that in this patient sample, there are indeed deletions in chromosome 7 affecting many genes. The SNV data corroborated with the CNV data showing loss of heterozygosity (LOH) of genes found on chromosome 7. Specifically, we see several genes shifting from heterozygous in normal cells to wild type in cancerous cells, and one variant changing its genotype from heterozygous to homozygous mutant. This perceived change in zygosity results from genes lost from deletions in chromosome 7. Moreover, our CNV data has also revealed an additional copy loss of TET2 on chromosome 4, although it does not change the wild type status of the TET2 gene (Figure 3b).

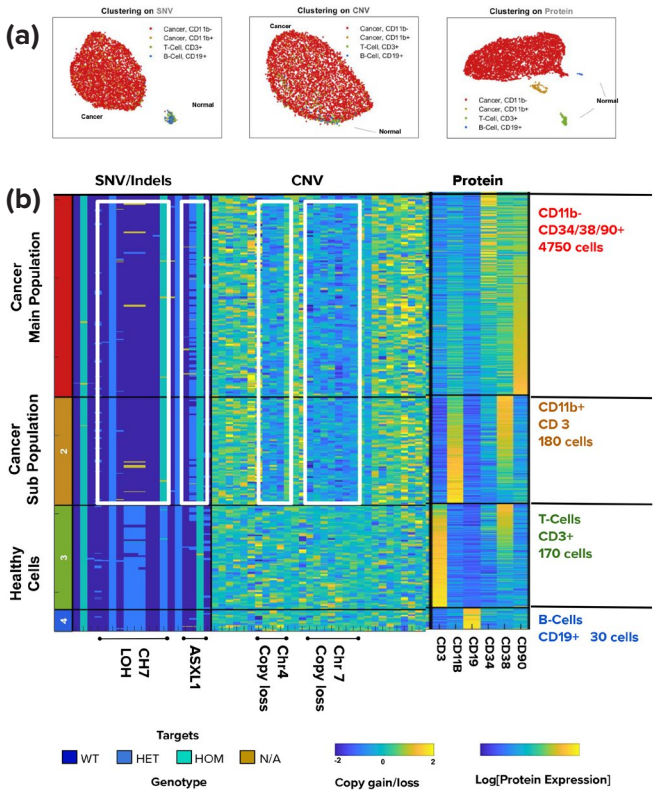

Figure 3 - t-SNE plots (a) and heat map showing unsupervised clustering (b) of SNV/indel, CNV, and protein results from AML research sample, illustrating more distinct clustering based on protein expression versus genotype. t-SNE plots are colored in by subpopulations identified through protein expression clustering.

Conclusion

Within a heterogeneous sample various populations are defined by inherent patterns, which may be defined by both genotype and phenotype at single-cell resolution revealing additional biological insight not visible with conventional bulk analysis. Taking a multi-omics approach ensures coverage of both parameters to capture the characteristic with the strongest signal to define critical subpopulations. Our data illustrates the power of integrated single-cell , CNV, and protein expression data to differentiate cancerous subpopulations and strengthen the conclusions made from individual analytes alone

References

1. https://www.nature.com/articles/nature11003#citeas