mission bio

tapestri
APPLICATION NOTE

Single-Cell Multi-Omics Reveals Novel Correlations Between Genomic Variants and Protein Expression in AML Patient Samples

Takeaways

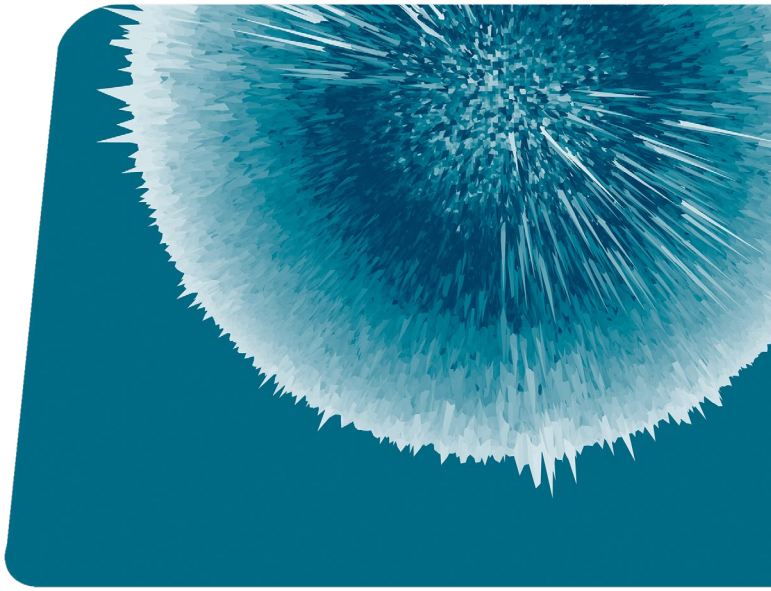
- Tapestri Platform is the only single-cell multi-omics platform capable of uncovering SNVs, CNVs and protein expression from the same single cell
- Simple antibody staining protocol layers on protein data to Tapestri single-cell experiments
- Obtain genotypes and phenotypes from thousands of single cells

Abstract

Genomics data alone provides only partial insights into the cell's type, state or function. To obtain a more complete picture and reveal the cell's phenotypic character, a multi-omics approach is needed that expands beyond DNA profiling. Here, we demonstrate for the first time, that the Tapestri Platform enables the detection of genomic variants and protein expression simultaneously from the same cell. Genomic variations such as single nucleotide variants (SNVs) and copy number variations (CNVs) are now co-detected with proteins. This novel capability has the power to reveal rare cell populations, subtle cell states, and link genomic variation to protein expression leading to more informed research on disease and therapeutic development.

As a proof of concept, four populations of cancer cell lines were mixed and analyzed at the single-cell level using the Tapestri Single-Cell DNA AML Panel together with a panel of 6 oligo-conjugated protein antibodies. Integrated SNV, CNV, and protein data analysis led to more distinct clustering of the four samples versus each analyte alone. This result highlights the power of concurrent single-cell genomic variation and protein expression data to identify distinct cellular subpopulations within a heterogeneous sample.

To extend the feasibility of the new technique to clinical samples, peripheral blood mononuclear cells (PBMCs) and bone marrow (BM) samples were analyzed from six patients with acute myeloid leukemia (AML) for both DNA variants and protein expression. Using a custom panel, the Tapestri Platform, and Tapestri Insights software, pathogenic variants were identified and mutational co-occurrence patterns revealed, allowing the identification of several clones in each of the six samples. Single-cell protein detection of six markers led to the identification of clones in different cellular states. Overall, these results demonstrate a new single-cell multi-omics approach, and for the first time, integrate genotype and phenotype data from the same cell across thousands of cells.



mission bio

tapestri
APPLICATION NOTE

Experiment & Methods

Raji, K562, TOM1 and KG1 cell lines were analyzed using the Tapestri Single-Cell DNA AML Panel for both SNVs/indels and CNVs. Cells were processed on the Tapestri Platform to simultaneously access protein expression using a panel of 6 antibodies conjugated to analyte barcoded oligo tags. The targets consisted of CD19, CD33, CD45, CD90, HLA-DR and mouse IgG1κ. For downstream analysis, only a select few SNVs/indels, CNVs and proteins were included.

Next, six AML patient samples were analyzed with a custom DNA panel of 31 genes relevant to AML, MPN, and MDS across 109 amplicons. In addition, a custom protein antibody panel was used targeting the following 6 proteins: CD3, CD11b, CD34, CD38, CD45RA and CD90. Data were analyzed with custom Tapestri Pipeline software. SNVs and indels were identified using Tapestri Insights software, CNVs were analyzed using the Mision Bio "tapestri-cnv" package for R, and DNA + protein data were integrated and analyzed using the Mission Bio "tapestri-protein" package for R.

Results

Single-cell multi-omics provides greater resolution for clonal clustering

Raji, K562, TOM1 and KG1 cells were mixed together at equal ratios and analyzed for SNVs, indels, CNVs and proteins using the Tapestri Platform. Unsupervised clustering and heat map visualization of each individual analyte resolved 3 cell lines using the SNV data (based on 4 variants), and 4 cell lines using the CNV or protein data (Figure 1a). t-SNE projections using combinations of SNV, CNV and protein data showed varied resolutions of the 4 cell lines (Figure 1b). Resolution of the cell lines increased when SNV or CNV were combined with protein data respectively, while combined SNV, CNV and protein data together led to the most distinct resolution of the 4 cell line populations on the t-SNE. This result illustrates the power of using more data from the same cells with a multi-omics approach to gain the greatest resolution between cell types.

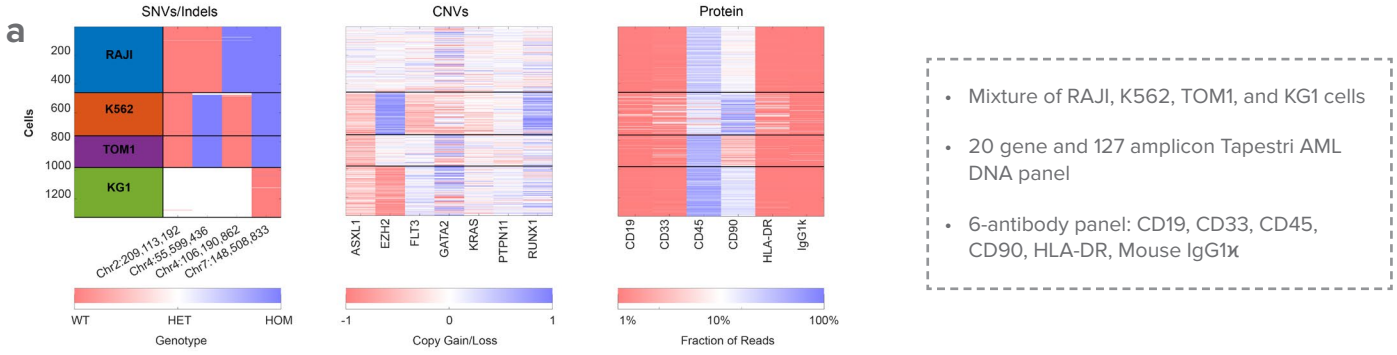


Figure 1 - (a) heat maps and (b) t-SNE plots derived from SNV, CNV, and protein data of a mixed population of four cell lines. t-SNEs were colored in by SNV genotypes previously established with pure cell lines.

SNV analysis of AML patient samples using Tapestri identifies clones with single mutations and co-occurring mutations

AML is a cancer that evolves from the proliferation of undifferentiated myeloid lineage cells that acquire sequential mutations[1]. Various states of leukemic cells, such as leukemic stem cells or progenitors, can be defined by cell surface marker protein expression. In addition, using DNA mutational data, genomic analysis of leukemic cells can define varied clonal populations and lineages, but fails to reveal cell states. For the first time we show that genomic and protein analysis at the single-cell level correlates genomic variance, cell surface protein expression, and inferred cell state. Here we tested the multi-omics capabilities of the Tapestri Platform on samples from patients with AML.

In collaboration with Dr. Ross Levine's laboratory at Memorial Sloan Kettering Cancer Center, six AML patient samples that had previously been analyzed with conventional next-generation sequencing (NGS), were analyzed on the Tapestri Platform for DNA and protein. Each sample was analyzed with a custom DNA panel of 109 amplicons to 31 genes and a custom oligo-conjugated antibody panel of 6 proteins. In total, over 20,000 cells were sequenced and analyzed. In addition to all bulk-sequencing confirmed mutations covered by the custom panel, additional pathogenic variants were identified in a subset of the samples that remained undetected previously. Co-occurrence of SNVs/indels was determined in each population of cells, revealing several clonal populations within each sample and a subset of the clones are shown in Figure 2. Clones harboring one mutation are shown in blue, clones with two co-occurring mutations are shown in green, and clones with three co-occurring mutations in the same single cells are shown in red. DNMT3A R882H and IDH1 R132C were found in multiple samples while an additional seven variants in DNMT3A, IDH2, TP53, TET2, NRAS, and NPM1 were unique to individual samples.

Tapestri analysis of single-cell protein expression reveals distinct subpopulations

In addition to SNV/indel analysis, data for the six patient samples were merged and simultaneously analyzed for DNA and protein. Clustering by genotype data on a t-SNE projection distinguished each patient sample, as expected, and sub-clustered each sample into unique clones (Figure 3a). Clustering by protein data on a t-SNE resulted in lower resolution of all patient samples, highlighting the shared protein expression patterns across genotypically different cell populations (Figure 3b).

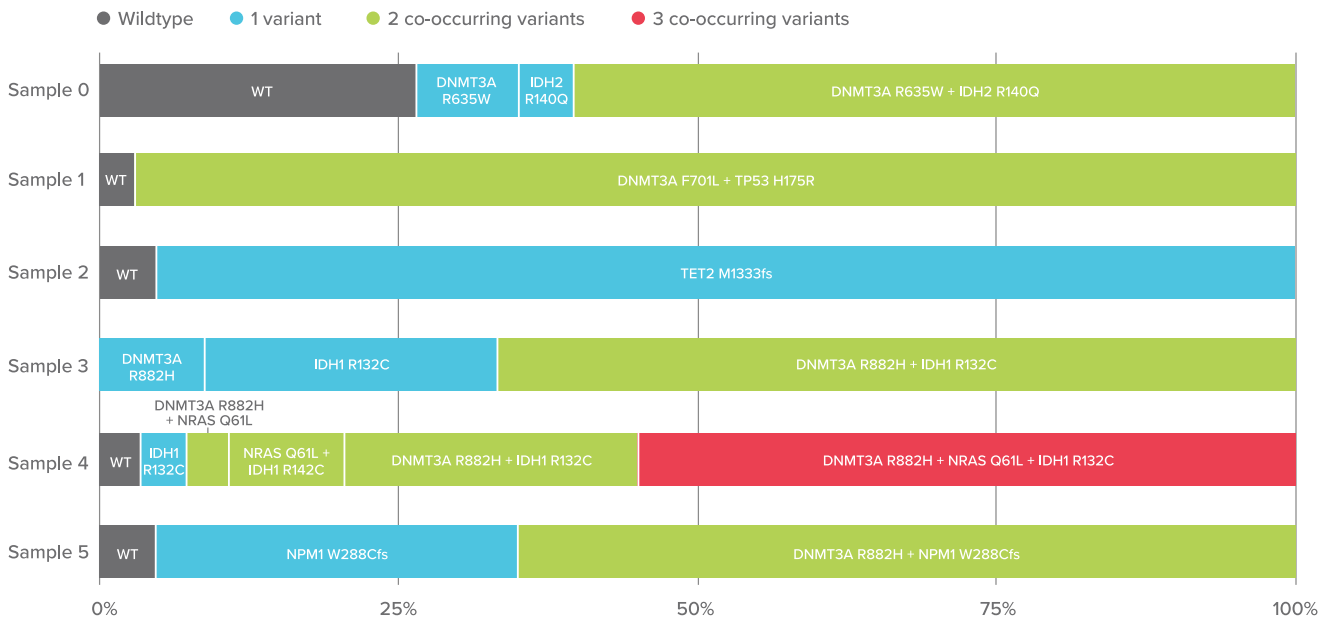


Figure 2 - Co-occurring SNVs and indels were identified in each patient sample using the Tapestri platform and custom DNA panel, establishing distinct clonal populations.

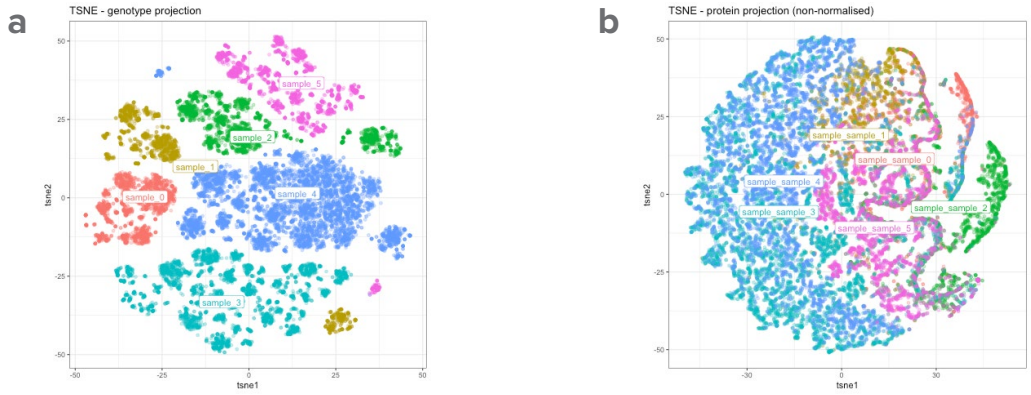


Figure 3 - t-SNE plots of SNV/indel (a) and protein (b) results from 6 AML patient samples illustrating more distinct clustering of patient samples based on genotype versus protein expression. Plots are colored in by sample.

SNVs/indels within clonal populations can be indicative of AML disease status, progression, and relapse, while high and low levels of cellular protein expression can be indicative of AML prognosis. Using an additional dimensional reduction tool, we used protein expression data from the six patient samples to generate a UMAP projection. Overlaying a subset of variants on the UMAP projection revealed distinct distributions of clonal populations (Figure 4a), and together with the protein expression data (Figure 4b-g) enabled us to define different patterns of relationships between DNA mutation status and protein expression levels.

The six protein markers used in this study helped define the phenotypic state of myeloid cells that cause AML, such as stem and progenitor cells versus wildtype T-cells. CD11b is selectively expressed in leukocytes and is associated with poor prognosis for AML patients[2]. CD3 is a T-cell marker regulating cytotoxity while CD45RA is expressed in naive T cells. High levels of CD34 expression and low levels of CD38 expression - associated with stem cell-like states - are also associated with poor AML prognosis[3]. CD90 has been associated with leukemic stem cells. The UMAP showed different cell populations in different cellular states as identified by their protein expression. Importantly, CD3+ T-cells and CD11b+ leukocytes were distinctive from CD34 and CD38 positive stem-cell/progenitor cell populations. Moreover, the CD3+ T-cells were associated with cells that had a wildtype genotype, while the CD34+ and CD38+ stem cells contained cells with mutations. These data show exciting genotype and phenotype information for over 20,000 single cells.

Single-cell multi-omics reveals correlations between AML genomic variants and protein expression

To examine more closely the relationship between genotypes and phenotypes, we generated heat maps of the expression of the six cell proteins along with a heat map of the SNVs/indels identified earlier with Tapestri

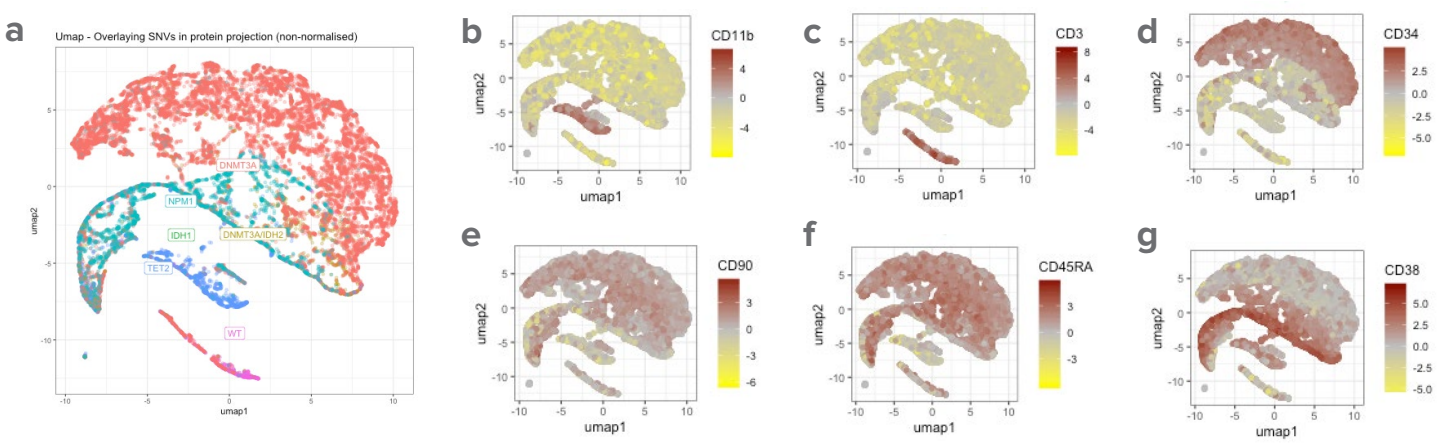


Figure 4 - UMAP clustering by cell surface protein identifies distinct cell types.

Insights. Using unsupervised clustering, we show detailed corollary expression of proteins and SNVs/indels as well as zygosity and co-occurrence of the SNVs/indels within each sample (Figure 5a). Importantly, we verify the expected correlation between NPM1 W288Cfs mutational status and low CD34 expression[4].

To more easily visualize the relationship between mutational status and protein expression we then combined all six samples and generated a correlation heat map relating SNVs/indels to protein expression (Figure 5b, red = positive correlation, blue = negative correlation). We found that TET2 M133fs correlates with $CD34_{low}/CD38_{high}$ phenotype, while NRAS Q61L, IDH1 R132C, and DNMT3A F701L correlates with $CD34_{high}/CD38_{low}$ phenotype. Interestingly, the three DNMT3A mutations showed different correlations with the CD38 marker, with R882H showing no correlation, R635W showing high correlation and F701L showing low correlation. Whether it is the different variants within the same gene or the co-occurrence of these variants with other variants, this result highlights the different nuances that genotype changes have on phenotypic expression. Together, these data highlight the deeper biological insight gained into cancer indications by integrating multi-omics data from the same single cells.

Conclusion

The Tapestri Platform has multi-omic capabilities with the ability to measure cell protein expression with genomic SNVs, indels, and CNVs from the same single cells. Using a simple cell staining technique with the standard Tapestri protocol, subtle cell states are determined and genomic variants correlated with protein expression. We validated this new technique with cell lines and blood cancer samples. And we present the first commercially available platform that detects proteins, SNVs/indels, and CNVs from the same single cells to unravel genotype to phenotype relationships.

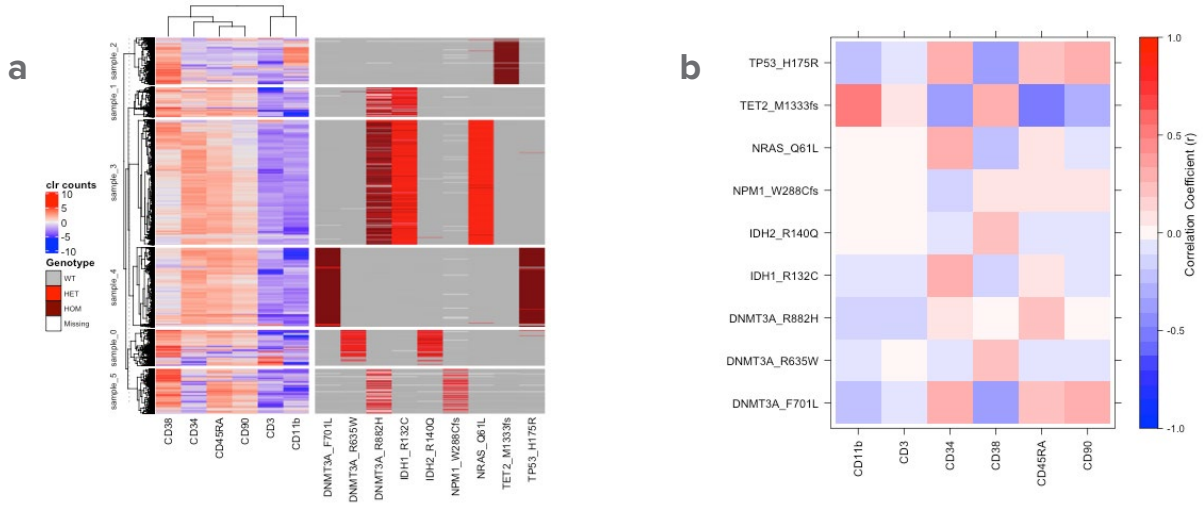


Figure 5 - (a) Heat maps showing unsupervised clustering of patient samples based on cell surface proteins (left) and SNVs (right). (b) Heat map of correlated expression of cell surface proteins to SNVs, indicating strong correlations between certain variants and protein levels (red, high correlation; blue, low correlation).

References

1. Viny AD et al., Roads Diverge--A Robert Frost view of leukemia development, N Engl J Med, 374(23), 2282-2284 (2016)
2. Xu S et al., Prognostic value of CD11b expression level for Acute Myeloid Leukemia patients: A meta-analysis. PLoS One, 10(8):e0135981 (2015)
3. Plesa A et al., High frequency of CD34+CD38-/low immature leukemia cells is correlated with unfavorable prognosis in acute myeloid leukemia, World J Stem Cells, 9(12), 227-234 (2017)
4. Martelli et al., CD34+ cells from AML with mutated NPM1 harbor cytoplasmic mutated nucleophosmin and generate leukemia in immunocompromised mice, Blood, 116(19): 3907-3922 (2010)