mission bio BioLegend®

tapestri
APPLICATION NOTE

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

Takeaways

- The Tapestri Platform is the only single-cell multi-omics platform that measures genotypes and phenotypes simultaneously from thousands of cells
- Simple visualization and data analysis software integrates multi-faceted DNA and protein single-cell data
- The optimized and fully validated 45-plex TotalSeq™-D Heme Oncology Cocktail v1.0 from BioLegend complements Mission Bio's pre-designed and custom targeted DNA panels

Abstract

Cancer is a heterogeneous disease, both genotypically and phenotypically, however, the ability to combine genotype-phenotype relationships at the single-cell level has remained a challenge for researchers. Instead, researchers have relied on bulk measurements from multiple different assays to piece together information for clinical samples. For the first time, we demonstrate, in partnership with BioLegend, a commercially available wet-lab validated protein panel that works specifically with the Tapestri Platform to simultaneously measure DNA variations for up to 1,000 targets and protein expression for 45 targets in thousands of single cells. This novel capability has the power to reveal cell identities and subtle cell states and link genomic variation to protein expression, leading to more informed research on disease and therapeutic development.

Introduction

Acute myeloid leukemia (AML) is a cancer that develops in the myeloid tissue of the bone marrow and has several subtypes with variable outcomes. Various states of leukemic cells, such as leukemic stem cells or progenitors, myeloid blasts, and normal cells, can be defined by cell surface protein markers or morphology. In addition, DNA mutational data and genomic analysis of leukemic cells can define varied clonal populations and lineages. Here, we show for the first time, the ability to combine genotype and phenotype information from the same single cells in order to obtain proteogenomic information with a fast, cost-effective, high-throughput, and wet-lab validated method.

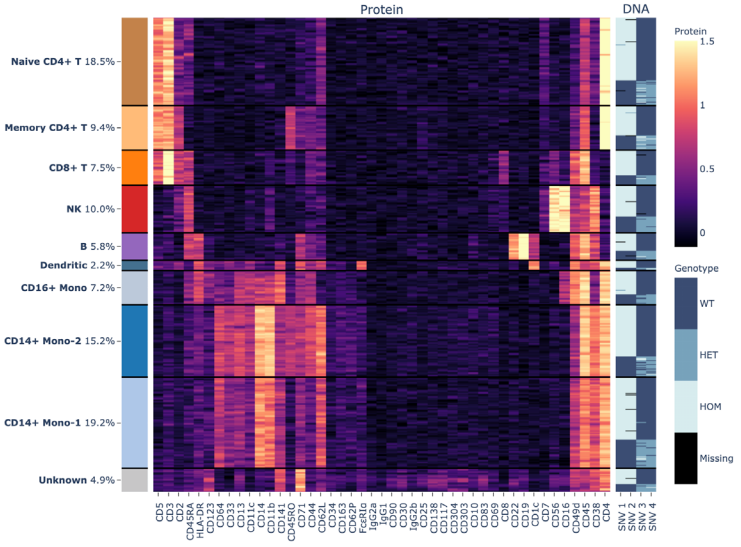

Figure 1 - PBMCs from two donor samples were mixed together. Heatmap visualization reveals the power to obtain genotype and phenotype from the same cells across thousands of cells.

First, we demonstrate in Peripheral Blood Mononuclear Cells (PBMCs) from healthy donors, the ability to cluster thousands of cells by phenotype and genotype using the newly available 45-plex TotalSeq-D Heme Oncology Cocktail v1.0 by BioLegend with the Tapestri Single-cell DNA Myeloid Panel from Mission Bio (Figure 1). Next, to test the system on AML cancer samples and obtain for the first time, proteogenomic information from these samples, we collaborated with Dr. Ross Levine's laboratory at Memorial Sloan Kettering Cancer Center. Six AML patient samples that had previously been analyzed with conventional bulk sequencing were analyzed on the Tapestri Platform for DNA and protein[1]. Each sample was analyzed with a custom DNA panel and a custom oligo-conjugated antibody panel. Over 20,000 cells were sequenced in total, and the data were analyzed with Tapestri Pipeline, Tapestri Insights, and Tapestri Mosaic analysis tools. Overall, the results demonstrate a new single-cell multi-omics approach that integrates genotype and phenotype data for the same cell across thousands of cells.

Materials & Methods

Experiments for DNA and protein were performed using the complete end-to-end workflow that includes a cell staining step with oligo-conjugated antibodies upstream of the Tapestri Platform (Figure 2). PBMCs from two healthy donors were mixed together and analyzed with the Tapestri Single-cell DNA Myeloid Panel and the BioLegend 45-plex TotalSeq-D Heme Oncology Cocktail v1.0. In collaboration with Dr. Ross Levine's laboratory, six AML patient samples were analyzed with a custom DNA panel targeting 32 genes across 109 amplicons together with a custom protein antibody panel for six proteins targeting CD3, CD11b, CD34, CD38, CD45RA, and CD90. Secondary data analysis was performed using Tapestri Pipeline software. Tertiary data analysis was done with Tapestri Insights software and the Tapestri Mosaic analysis package.

Results

45-plex TotalSeq-D Heme Oncology Cocktail v1.0 by BioLegend

Cancer researchers rely heavily on protein markers to identify different cell types and cell states, but they have not been able to combine this phenotypic information with genotypic information on a large scale, particularly in single cells. In conjunction with BioLegend, a 45-plex TotalSeq-D Heme Oncology Cocktail v1.0 was developed that was curated by world leaders in heme oncology across academia and biotech and then extensively wet-lab optimized and validated. Data from a mixture of PBMCs from two healthy donors were clustered on a UMAP plot using the protein expression data from the 45-plex panel with Tapestri Insights software. Results showed distinct clustering of monocytes, CD8+ T cells, CD4+ T cells, NK cells and B cells (Figure 3a) that was verified with cell-specific markers (Figure 3b). Further analysis using the Tapestri Mosaic analysis package to generate a hierarchical clustering heatmap showed even greater resolution of PBMCs into nine cellular subpopulations, including naive versus memory CD4+ T cells, CD14+ versus CD16+ monocytes and

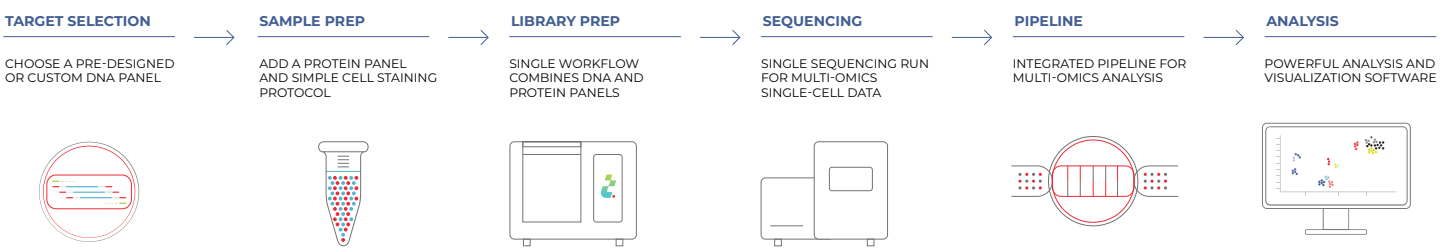

Figure 2 - Multi-omics DNA and protein Tapestri workflow.

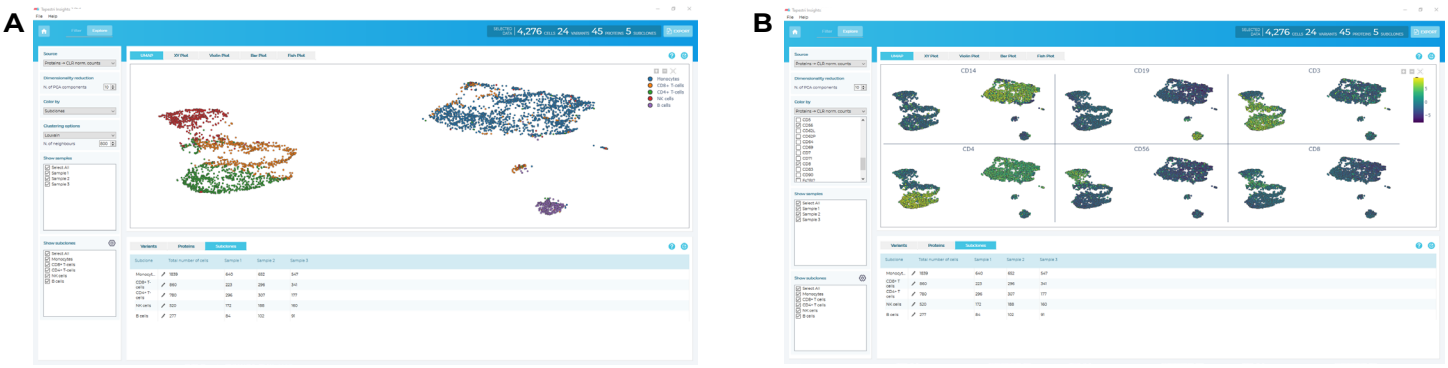

Figure 3 - UMAP plot of PBMCs using the 45-plex TotalSeq-D Heme Oncology Cocktail 1.0 from BioLegend and analyzed with Tapestri Insights software (a-b).

a minor population of dendritic cells (Figure 1). Moreover, protein information was paired with genotype information to identify the PBMCs that came from donor 1 versus donor 2. For the first time, researchers have a large-scale validated protein panel that pairs with large-scale targeted DNA panels to obtain high-throughput single-cell multi-omics insights.

Analysis of an AML patient sample using Tapestri Insights software

While SNVs and indels within clonal populations can be indicative of AML disease status, progression, and relapse, high and low levels of cellular protein expression can be indicative of AML prognosis. An AML patient sample was analyzed using Tapestri Insights software, and the variants *DNMT3A R882H,* and *IDH1 R132C* were identified and selected based on previous bulk sequencing data, their prevalence in the sample, and their likely pathogenicity using ClinVar annotations. UMAP plots were generated using the DNA variant allele frequencies (VAFs) and two clear clusters emerged, one that was wildtype and one that was mutated, containing both pathogenic variants (Figure 4a). Zygosity and co-occurrence status showed that the majority of the mutated cells harbored both mutations in the same single cells in a heterozygous state, while rare populations of homozygous cells were also observed for each variant. From this UMAP plot, protein expression was overlaid, and CD3

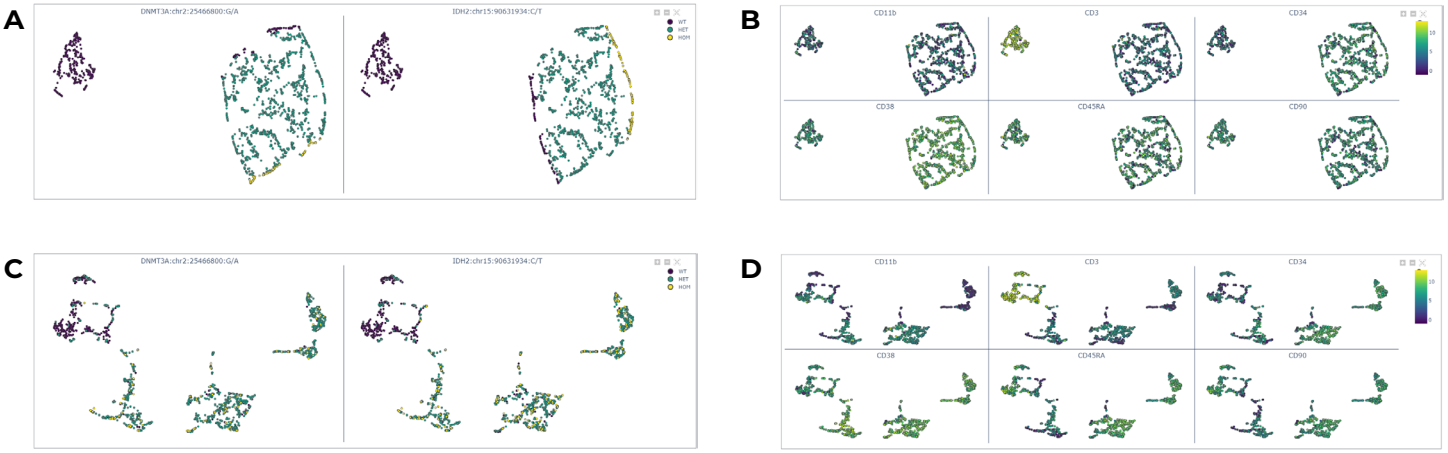

Figure 4 - An AML patient sample analyzed using Tapestri Insights shows UMAP plots clustered by DNA (a-b) or protein (c-d) and identifies distinct cell populations.

expression was observed in the genotypic wildtype population while CD34 and CD38 expression was observed in the genotypic mutated cell population (Figure 4b).

Alternatively, for researchers who want to start with protein data first, UMAP plots can be generated in Tapestri Insights based on clustering from the protein expression. Cells were clustered by the expression of the six protein markers, and the two DNA pathogenic variants were overlaid, showing a clear wildtype population and several mutated populations (Figure 4c). When the protein expression of the six markers was overlaid on the UMAP plot, data showed high, medium, and low expression in the different clusters of cells and revealed how the various combinations of expression profiles for the six markers led to a more refined grouping of the single cells into distinct populations (Figure 4d). Whether researchers want to start first with DNA variants or first with protein expression, Tapestri Insights software supports both analysis pathways and allows researchers to go between the two analyses interchangeably.

In addition to unsupervised clustering, groups or subclones can additionally be defined in Tapestri Insights using k-means or Louvain algorithms. Here we defined two subclones that corresponded with the wildtype and mutated clusters on the UMAP plot colored in by k-means grouping (Figure 5a). In addition to multiple and comprehensive UMAP tools for both DNA and protein analysis, Tapestri Insights also offers additional features for visualizations and analysis. XY scatter plots show two protein markers in relation to each other, giving investigators a readout more akin to fluorescence-activated cell sorting (FACS) analysis (Figure 5b). Multiple violin plots are easily selected that show DNA variant depth of coverage (DP), genomic quality (GQ), and allele frequency (AF), while violin plots based on protein data display inverse

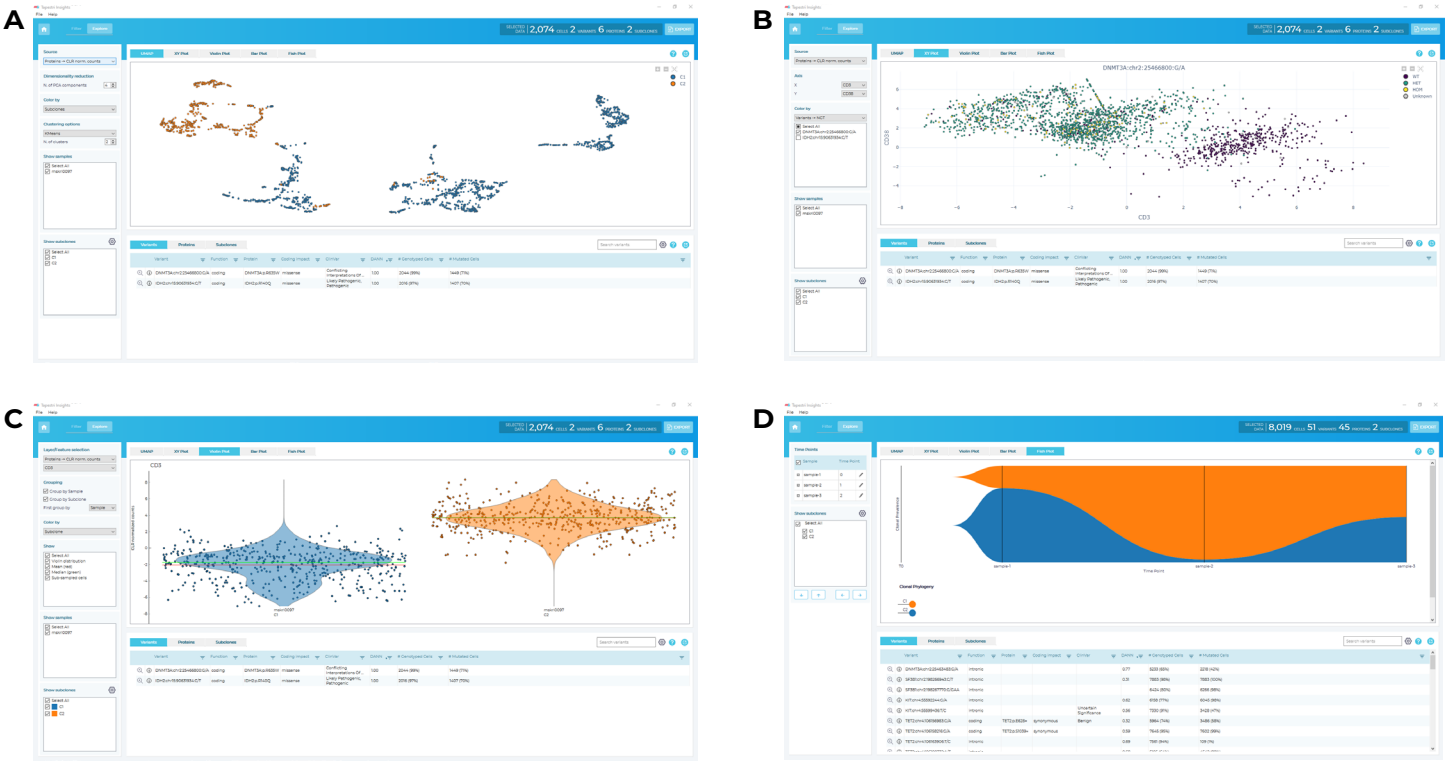

Figure 5 - Multiple analysis and visualization tools in Tapestri Insights software. UMAP plot (a), XY scatter plot (b), violin plot (c), and fish plot (d).

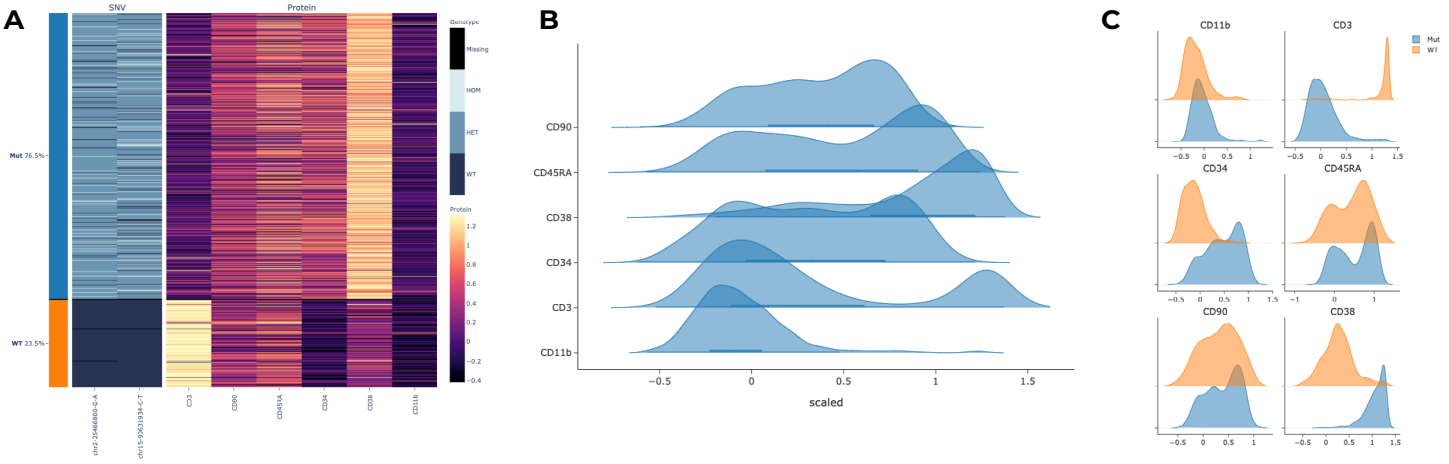

Figure 6 - Heatmap showing unsupervised clustering of an AML patient sample based on SNV (left) and cell surface proteins (right) (a). Ridge plots for each protein marker across the entire population of cells (b). Ridge plots by cell type based on genotype (c).

hyperbolic sine function (Asinh) and centered log ratio (CLR) normalized counts (Figure 5c). Bar plots quantify the number of cells in each subclone while fish plots visualize clonal evolution from time course or multi-sample experiments (Figure 5d). Tapestri Insights software gives researchers a quick and easy way to visualize complex single-cell multi-omics data in multiple formats.

Analysis of AML patient samples using the Tapestri Mosaic analysis package

In addition to the easy and intuitive Tapestri Insights software, more advanced analysis tools are available with the Tapestri Mosaic analysis package of Jupyter Notebooks available on GitHub (https://github.com/MissionBio/mosaic). A heatmap with both DNA and protein data side-by-side visualizes the high CD3 expression in the wildtype cells and the high CD38 and CD34 expression in the mutated cells (Figure 6a). Ridge plots for each protein marker show the distribution of protein expression across all cells (Figure 6b), while ridge plots broken down by genotype reveal the complexity of protein expression within each subclone (Figure 6c). These data exhibit the complex and interesting expression patterns of bimodal and continuum profiles and support a recent report from Dr. Adam Abate

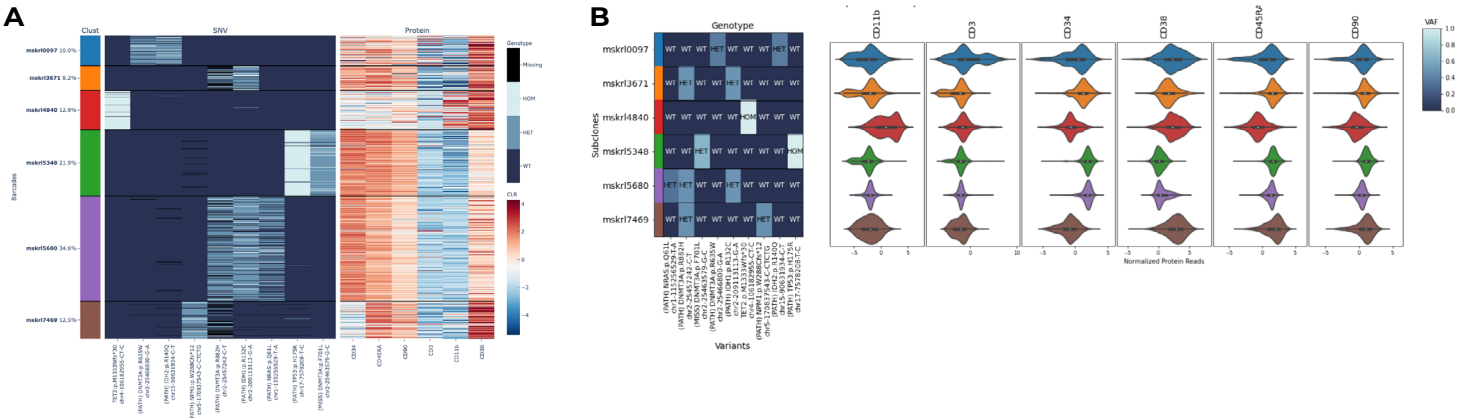

Figure 7 - Heat maps showing six AML patient samples and their SNV (left) and cell surface proteins (right) status (a). A simplified plot of genotype (left) and violin plots of phenotype (right) for each of the six samples (b).

and Dr. Cathy Smith of UCSF that found similar findings in their cohort of AML patient samples across multiple treatment time points and clinical recurrences[2].

Finally, data for the six AML patient samples were merged and simultaneously analyzed for DNA and protein using the Tapestri Mosaic analysis package. Two visualizations tools, heatmaps (Figure 7a) and violin plots (Figure 7b), showed where genotype and phenotype correlated and where they differed for each patient sample as well as between the six different patient samples. These data showed the heterogeneity and complexity within and between each patient sample, and directly linked genetic variation with protein expression in clonal populations using the single-cell multi-omics Tapestri Platform.

Conclusion

The Tapestri Platform has multi-omic capabilities that can determine cell surface protein expression with genotypes simultaneously in thousands of single cells[1-3]. Using a simple cell staining technique, cell identification and subtle cell states are determined, and genomic variants are correlated to protein expression. Here, we validated this new technique on multiple healthy and cancer blood samples. And we present the first commercially available platform from Mission Bio and the first commercially available reagent by BioLegend that detects proteins and DNA simultaneously from the same single cells to unravel complex genotype-phenotype relationships in thousands of cancer cells.

References

1. Miles L, et al. Single cell mutation analysis of clonal evolution in myeloid malignancies. Nature (2020).
2. Demaree B, et al. Joint profiling of proteins and DNA in single cells reveals extensive proteogenomic decoupling in leukemia. BioRxiv (2020).
3. Morita K, et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. Nature Communications (2020).