

# Single-Cell DNA Sequencing Resolves the Genetic Complexity Underlying CLL Progression

## Takeaways

- Data generated using the Tapestri Single-Cell DNA CLL Panel reveals clonal heterogeneity that exists in different genes across multiple CLL patients
- Aggregated single-cell variant allele frequencies of 30+ disease-relevant mutations highly correlates with bulk sequencing data
- Co-occurrence and zygosity of mutations in subclonal populations was uncovered in the single-cell data across all patient samples.

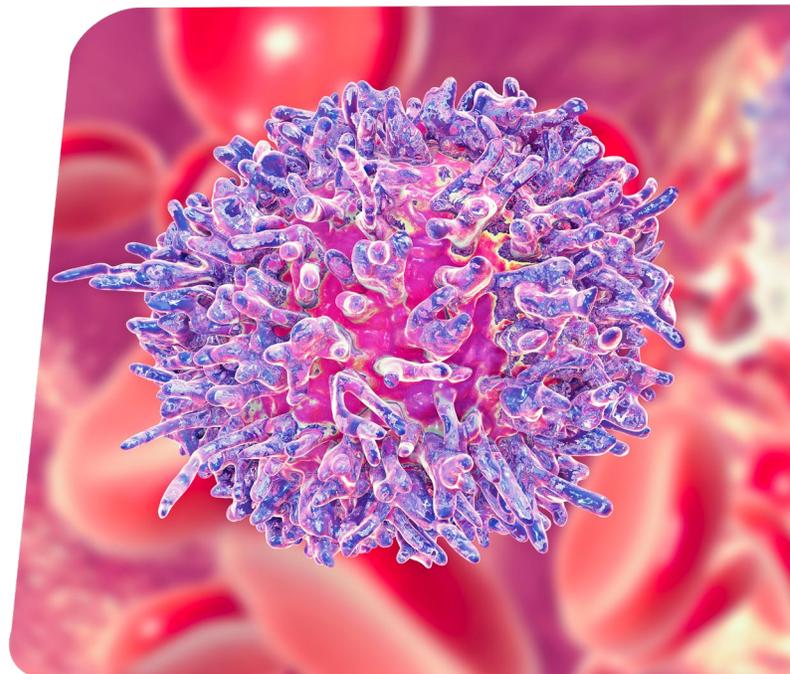
## Abstract

Early detection of cancer and precancerous states are becoming a reality thanks to newly established sequencing technologies. High-count monoclonal B-cell lymphocytosis (MBL) is an asymptomatic state that can evolve into chronic lymphocytic leukemia (CLL) when B-cells gain a cancer-inducing combination of driver mutations. In a recently published study<sup>1</sup>, MBL patient samples, taken at several time points before treatment for CLL, were analyzed using bulk targeted deep-sequencing. The researchers' important finding was that in a subset of cases, there were a significant number of mutations, in many cases affecting the same genes. Moreover, these mutations were detectable on average 41 months prior to CLL progression in the MBL samples. To further gain insights into these MBL patient samples, single-cell DNA sequencing was performed on the Mission Bio Tapestri® Platform together with a 33-gene CLL amplicon panel. Here we show that previously generated data from bulk sequencing were highly correlative to the newly generated single-cell data. We

were able to unambiguously resolve co-occurrence and zygosity of all detected mutations to track clonal evolution and population expansion over time. These results show that single-cell DNA sequencing is a powerful tool for resolving clonal heterogeneity. Furthermore, these subclones can be analyzed over time to identify the subclone that becomes dominant within the patient sample.

## Introduction

Chronic lymphocytic leukemia (CLL) is a genetically heterogeneous disease preceded by an asymptomatic expansion of clonal B-cells in a state known as high-count monoclonal B-cell lymphocytosis (MBL). Approximately 1-2% of high-count MBLs evolve to CLL per year<sup>2</sup>. While significant gains have been made in our understanding of the genetic mutations that contribute to an aggressive CLL disease state, identifying the biologic and genetic events that determine which MBLs will progress to CLL is less understood. By gaining a better understanding of clonal evolution that drives the continuum from MBL to CLL, patient prognosis could be predicted long before CLL symptoms occur, and patients could receive regular disease monitoring and earlier treatment leading to overall better outcomes.



In a recent study (Barrio et al.<sup>1</sup>), bulk next-generation sequencing (bulk NGS) of 48 high-count MBL patient samples revealed that MBL has a mutation profile similar to that of patients who progress to CLL, and that the mutations are detectable years before disease progression. At the initial time point analyzed on the MBL/CLL continuum, driver mutations were found in over 50% of MBL cases tested. 27% had more than one mutation, and 8% showed multiple mutations in the same gene. Several time points showed an increase in mutation frequency over time as patients progressed to CLL. While this illustrated strong evidence that the tumor mutation profile is present in the pre-malignant MBL stage, the data is limited by bulk NGS analysis.

Cancer forms through coordination and serial acquisition of several driver mutations. Bulk NGS analysis of MBL patient samples may reveal a total number of mutations found within a population of cells, but it cannot distinguish whether mutations co-occur in the same cells or whether the cells are heterozygous or homozygous for those mutations. Here, in collaboration with Dr. Esteban Braggio at the Mayo Clinic, we retest two MBL samples used in the Barrio et al. study that were part of the 8% cohort that had multiple mutations in the same gene, and two new unpublished MBL samples using the Tapestri Single-Cell DNA CLL Panel and Tapestri Platform. We show high correlation to previous bulk NGS analysis, and resolution of mutational co-occurrence and zygosity.

## Experiment & Methods

### Cells and sample preparation

Four frozen MBL patient samples (Patient 1 - 4) at two time points (TP1, TP2) were obtained.<sup>1</sup> Samples consisted of enriched peripheral blood mononuclear cells (PBMCs) with a high count of B cells. Patient samples were chosen for this study due to their previously identified mutational status using bulk NGS analysis. Cells were thawed and washed before centrifugation and resuspended in Tapestri cell buffer before proceeding to cell encapsulation in droplets.

## Analysis with Tapestri Single-Cell DNA CLL Panel

Using the Tapestri Platform workflow, cells were encapsulated in droplets, nuclei lysed, DNA barcoded, and CLL genes amplified. The Tapestri Single-Cell DNA CLL Panel was used, targeting 272 amplicons covering ~52.3 kilobases across 33 oncogenes relevant in CLL (Table 1). An average of 8,400 cells were sequenced for each sample. Patient 1 and Patient 2 were sequenced using Illumina's HiSeq 4000 and Patient 2, Patient 3 and Patient 4 sequenced using Illumina's Novaseq S1. Read coverage averaged 56x and panel uniformity which is the percentage of amplicons covered at >20% of the mean coverage, averaged 93.8% (Table 2). Sequencing results were analyzed by Tapestri Pipeline and Tapestri Insights Software.

### 33-GENE CLL PANEL

ATM	CD79B	EZH2	MED12	PLCG2	TP53
BCOR	CHD2	FAT1	MYD88	POT1	XPO1
BIRC3	CREBBP	FBXW7	NFKBIE	RPS15	ZMYM3
BRAF	CXCR4	KRAS	NOTCH1	SETD2	-
BTK	DDX3X	LRP1B	NRAS	SF3B1	-
CARD11	EGR2	MAP2K1	PAX5	SPEN	-

Table 1. Tapestri CLL Panel gene content.

## Results

### Single-cell DNA sequencing reveals clonal heterogeneity across MBL patient samples

Multiple mutations in several genes were identified in each sample using single-cell analysis and matched the mutations found in the previous bulk NGS experiments<sup>1</sup>. Each patient presented with different driver mutations, showing the heterogeneity underlying CLL disease progression (Table 3) and also the benefits of using a targeted CLL panel to capture a multitude of potential genotypes.

Sample ID	Sequencer	Number of cells before filtering	Number of reads	Reads per Cell	Panel Uniformity
Patient 1 - TP1	HiSeq 4000	9,094	163M	10,455 (37x)	93.0%
Patient 1 - TP2		7,939	136M	6,089 (21x)	94.4%
Patient 2 - TP1		5,227	138M	11,270 (39x)	94.4%
Patient 2 - TP2	Novaseq S1	6,154	317M	15,920 (56x)	93.4%
Patient 3 - TP1		5,445	332M	37,328 (131x)	93.0%
Patient 4 - TP1		12,759	299M	14,585 (51x)	94.4%
Patient 4 - TP2		12,503	340M	16,656 (58x)	94.1%
Average		8,446	246M	16,043 (56x)	93.8%

Table 2. Single-cell DNA sequencing details for patient samples tested

	Patient 1		Patient 2		Patient 3		Patient 4	
Clone 1	BIRC3 p.Q547Nfs21*	het	BIRC3 p.Q547Nfs*21	het	BRAF p.G466E	het	ATM p.W164X	het
Clone 2	BIRC3 p.E553Efs14*	het	BIRC3 p.E554Rfs*5	het	BRAF p.G469A	het	ATM p.W164X	hom
Clone 3	KRAS p.K117N	het	BIRC3 p.R555Nfs*3	het	BRAF p.G469V	het	ATM p.V2757G	het
Clone 4	KRAS p.G12D	het	BIRC3 p.E564K	het	BRAF p.L485W	het	ATM p.I2888T	het
Clone 5			BIRC3 p.K596*	het	BRAF p.N581S*	het	TP53 p.F231S	hom
Clone 6					BRAF p.N581S*	het	ATM p.W164X / ATM p.I2888T	het / het
Clone 7					BRAF p.K601E	het	ATM p.W164X / ATM p.V2757G	het / het
Clone 8					BRAF p.D594G	het		
Clone 9					KRAS p.Q22K	het		
Clone 10					KRAS p.K117N	het		
Clone 11					NRAS p.Q61R	het		
Clone 12					SF3B1 p.P780L	het		

Table 3. Results from mutational analysis of the final timepoint for four MBL patients using the TapeStri CLL gene panel and single-cell DNA sequencing platform. Genetic heterozygosity (het) and homozygosity (hom) are noted for each clone.

(\*) Represents a variant with the same protein change but caused by a different nucleotide mutation.

## Single-cell data is highly correlated with bulk NGS data over multiple time points

Next, we focused more closely on data from Patient 4 across the 2 timepoints. Correlation of the mutations and variable allele frequency (VAF) identified by single-cell and bulk NGS was extremely high, with an  $R^2$  of 0.994 (Figure 1). Additionally, the Tapestri platform was able to detect ATM p.V2757G and TP53 p.F231S mutations in the earlier TP1 sample, while bulk NGS results, using a VAF cutoff of 1%, did not detect these mutations until a later time point (Table 4).

### CORRELATION BETWEEN BULK SEQUENCING AND SINGLE-CELL PSEUDO BULK SEQUENCING

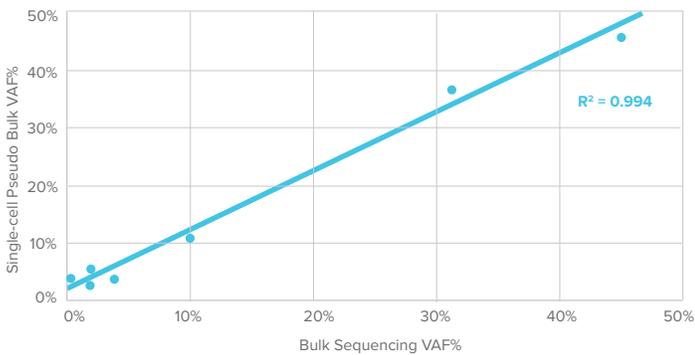


Figure 2. VAF% for each mutation identified (point on graph) in bulk-NGS1 (x-axis) and Tapestri single-cell sequencing (Y-axis) for TP1 and TP2 samples from Patient 4.

Variant	TP 1		TP 2	
	Bulk VAF %	scVAF %	Bulk VAF %	scVAF %
<b>ATM p.W164X</b>	31%	36%	45%	46%
<b>ATM p.V2757G</b>	ND	2%	1%	3%
<b>ATM p.I2888T</b>	3%	3%	10%	11%
<b>TP53 p.F231S</b>	ND	2%	1%	2%

Table 4. Mutations and VAF% identified in bulk-NGS (bulkVAF%) and Tapestri single-cell sequencing (scVAF%) for TP1 and TP2 samples from Patient 4.

ND-not detected.

## Resolution of co-occurrence of mutations and zygosity

Unlike bulk NGS, single-cell DNA sequencing enables the identification of distinct clones within a sample population by resolving co-occurrence of mutations and their zygosity. Using the Tapestri Platform, we found heterozygous populations in ATM W164X, ATM V2757G, and ATM I288T in Patient 4 samples, while ATM W164X and TP53 F231S displayed homozygous populations. The homozygous populations could also be due to the generation of independent monosomies in chromosome 11 and 17 occurring in the clones with ATM W164X and TP53 F231S mutations respectively. Three of the four mutations grew in VAF over time from TP1 to TP2 (Table 5 and Figure 2).

Variant	TP 1			TP 2		
	WT	Het	Hom	WT	Het	Hom
<b>ATM p.W164X</b>	44.4%	37.7%	17.9%	31.6%	45.6%	22.8%
<b>ATM p.V2757G</b>	97.4%	2.6%	0%	95.8%	4.2%	0%
<b>ATM p.I2888T</b>	95.4%	4.6%	0%	78.6%	21.4%	0%
<b>TP53 p.F231S</b>	98.5%	0%	1.5%	98.5%	0%	1.5%

Table 5. VAF% of wild-type (WT), heterozygous (het), and homozygous (hom) for mutated genes identified by the Tapestri single-cell platform for TP1 and TP2 samples from Patient 4.

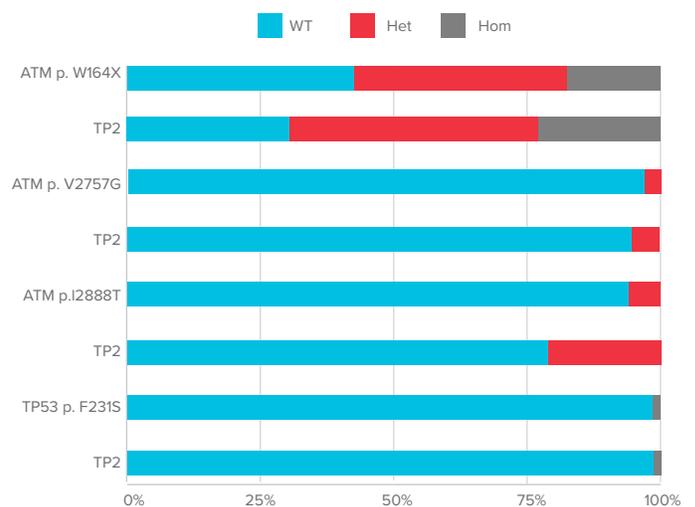


Figure 2. Bar graph of VAF% of wild-type (WT), heterozygous (het), and homozygous (hom) for mutated genes identified by the Tapestri single-cell platform for TP1 and TP2 for samples from Patient 4.

Additionally two clonal populations were discovered that had 2 co-occurring mutations in the same cells: heterozygous ATM W164X with heterozygous ATM V2757G, and heterozygous ATM W164X with heterozygous ATM I2888T. Overall, 6 clonal populations were established using single-cell sequencing for samples from Patient 4 (Table 3). We were also able to generate a phylogenetic tree to track the evolution of mutations from the single-cell data (Figure 3). Clone 2, which had a heterozygous ATM W164X mutation, evolves into Clone 3 acquiring a homozygous mutation of the same gene. Clone 2 also gains a co-occurring

heterozygous ATM V2757G mutation forming Clone 4, and in another subclone gains a heterozygous ATM I2888T mutation forming Clone 5. Furthermore, expansion of clonal populations can be tracked over time by comparing VAF at TP1 and TP2, and this is represented to scale by the size of the circles in Figure 3. It suggests that subclones with a biallelic inactivation of a tumor suppressor gene tend to expand more quickly in disease progression. This could be indicative of mutational gains that lead to increased rates of cell division and survival fitness, features that are characteristic of precancerous cells.

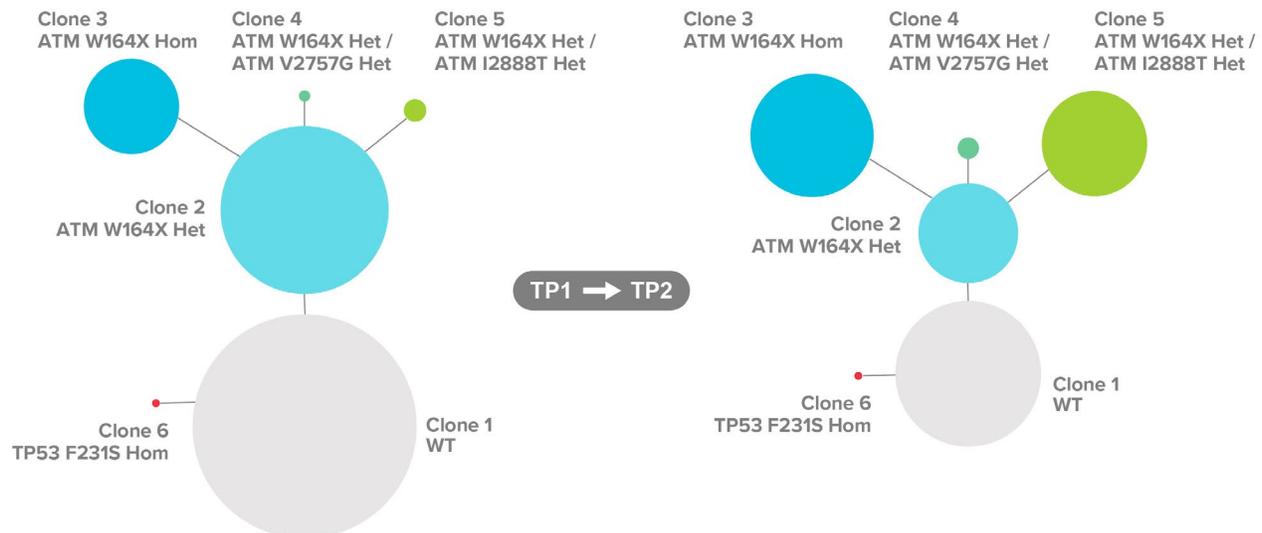


Figure 3. - Phylogenetic tree for TP1 and TP2 representing the evolution of clonal population defined by differentiating mutations. Circles are drawn to scale to represent the percent fraction of the subclone versus the total population.

## Conclusions

- Single-cell analysis of MBL/CLL patient samples correlated strongly with bulk NGS analysis, enabling confident comparison with previously-acquired results.
- The Tapestri Platform proved highly sensitive and detected mutations at <2% VAF.
- Single-cell analysis unambiguously identified co-occurrence and zygosity of mutations enabling the classification of many clones in each sample.
- Clonal population growth dynamics were able to be tracked over time using patient samples from multiple time points.

## References

1. Barrio *et al.*, Genomic Characterization of High-Count MBL Cases Indicates that Early Detection of Driver Mutations and Subclonal Expansion are Predictors of Adverse Clinical Outcome. *Leukemia*, **31**(1): 170–176 (2017)
2. Rawstron *et al.*, Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med.* **359**(6):575–83 (2008).



### QUESTIONS?

missionbio.com  
info@missionbio.com

6000 Shoreline Court, Suite 104, South San Francisco, CA 94080 USA  
+1 (415) 854-0058

For research use only. Not for use in diagnostic procedures. Mission Bio, Inc. makes no representation or warranty as to the usefulness, completeness, or accuracy of the information contained in this technical note. The products, services, and specifications set forth in this brochure are subject to change without notice and Mission Bio, Inc. disclaims any and all liability for such changes. The information contained herein is provided without warranties of any kind, either express or implied, and Mission Bio, Inc. disclaims any and all liability for typographical, printing, or production errors or changes affecting the products, services, and/or the specifications contained herein. It is the responsibility of the customer to thoroughly analyze all aspects of the customer's proposed application for the products and services. MISSION BIO, INC. DISCLAIMS ALL WARRANTIES, STATUTORY, EXPRESS, IMPLIED OR OTHERWISE, INCLUDING, BUT NOT LIMITED TO, QUALITY, CONDITION, AND SAFETY, AND THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.

IN NO EVENT WILL MISSION BIO, INC. BE LIABLE FOR ANY SPECIAL, INCIDENTAL, DIRECT, INDIRECT, PUNITIVE, OR CONSEQUENTIAL DAMAGES OF ANY KIND ARISING OUT OF THE RELIANCE ON THE CONTENTS OF THIS TECHNICAL NOTE OR FROM THE SALE OR USE OF ANY PRODUCT OR SERVICE DESCRIBED HEREIN.