Evaluation of LymphoTrack® Enterprise Software for Quantitative MRD, CLL Subset Classification, and Ig/TCR Repertoire Diversity Analysis

Jacob Wisotsky [1] , Tran Nguyen[1], Jonathan Guterres[1], Jelena Jablanovic[1], Angelo Arias[1], Julian D'Angelo[1] | [1]Invivoscribe, San Diego, CA

Background:

Accurate characterization of immunoglobulin (Ig) and T-cell receptor (TCR) repertoires by next-generation sequencing (NGS) is central to the study of lymphoid malignancies, including clonality assessment, measurable residual disease (MRD) monitoring, and immune repertoire profiling. However, standardized interpretation of these data often requires separate analytical approaches for quantitative normalization, chronic lymphocytic leukemia (CLL) classification, and repertoire diversity assessment. Here we introduce RUO LymphoTrack® Enterprise Software (LTE) v4.0.0 paired with the latest version of the LymphoTrack Flex assay. With the latest version of LTE, we introduce three new features: enhanced MRD analysis, CLL classification of the detected clones, and a method to assess repertoire diversity.

In CLL, the dominant clone identified at diagnosis is the basis for disease tracking and prognostic assessment. LTE reports the top clones with their V and J gene usage, CDR3 sequence, and supporting statistics that can be used to evaluate somatic hypermutation status. LTE v4.0.0 introduces a feature to classify clones into the 19 most prevalent stereotyped subsets in CLL, which represent approximately 12% of CLL cases[1]. Certain subsets like Subset #2 are indicative of aggressive mutation status and poor outcomes for patients, while others, like Subset #4, are indicative of slow-progression of the disease[1,2].

MRD analysis is used in hematological cancers to detect malignant clone sequences during treatment, to monitor for relapse, as a clinical trial endpoint, and to guide treatment decisions, all of which require high levels of analytical sensitivity. LymphoTrack Flex allows users to screen MRD clones across seven immunological gene targets (*IGH* FR1, FR2, FR3, IGHV Leader, *IGK*, *TRB*, and *TRG*). When analyzing MRD samples, LTE lets users enter the patient-specific clonotype, then reports exact and near-match read counts for that sequence enabling time point studies. LymphoQuant® (LQ) internal controls normalize inputs to enable quantification.

Diversity metrics capture the full distribution of sequence frequencies, allowing a user to see if there is a shift toward one or more clones[3]. LTE now utilizes the diversity measures Shannon, Simpson, Gini, and Pielou, each capturing a different property of the distribution: Shannon (richness and evenness), Simpson (probability two reads come from different clones), Gini (inequality across read counts), and Pielou (evenness independent of clone count). We validated these metrics by mixing 25 samples of varying clonality in four configurations.

Methods:

CLL Subsetting:

An *in-silico* sample was generated containing all 19 CLL subsets in the top 19 clones by percent of total reads. This was done using a background of a heterogeneous normal sample (i.e., non-clonal). These data were processed with the LTE v4.0.0 software with the CLL subsetting option enabled.

MRD Analysis of Clinical Sample:

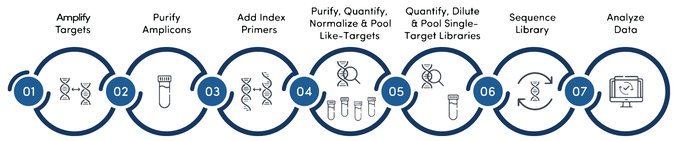
Twenty-nine follow-up timepoints from 11 patients with multiple myeloma were analyzed from archived bone marrow specimens, spanning three LymphoTrack® (LT) assays: *IGK* (n=9), *IGH* FR2 (n=8), and *IGH* FR3 (n=12). Library DNA input was 200 ng for most samples, with one patient's four timepoints measured at 55–229 ng per reaction, one FR3 library tested at 80 ng due to limiting mass.

Samples were originally sequenced on an Illumina MiSeq using the LymphoTrack® assay chemistry, with LymphoQuant® internal control spiked into every reaction. FASTQ files were reprocessed using LTE v4.0.0 generating combined MRD and LymphoQuant (LQ) internal control output for each sample/replicate. Cell equivalents, percent clonal cells, and LymphoQuant frequency were calculated for each timepoint using the clonal and internal control sequences.

Diversity Analysis of Clonal samples:

Libraries were prepared from 25 samples isolated from either peripheral blood or cell lines. A total of fifteen clinical samples were prepared at 50 ng or 200 ng DNA input, while 6 TRG samples were prepared as part of a dilution series, including a positive control, a negative control, and four other background samples. These samples were previously characterized for their clonality. Samples were sequenced on an Illumina NextSeq1000 using the LymphoTrack® Flex (LTF) IGHV Leader assay or the *TRG* assay. Sample FASTQ data was then combined manually to create various clonality scenarios. Sequencing data was then processed through LymphoTrack® Enterprise Software v4.0.0. Repertoire diversity metrics were calculated for each of the twenty-five samples by the software and then processed through plotting software to create the heat maps shown.

LymphoTrack® Flex Workflow



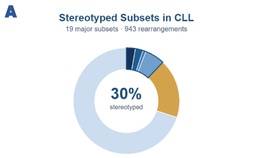
LTE Detects 19 of the Most Prevalent CLL Subsets

Table 1

Subset	IGHV	IGHJ	aa	CDR3 amino acid sequence
#2	IGHV1-69D	IGHJ6	22	ARGSGSLLVDGAADSGYYYGMDV
#5	-	IGHJ6	20	ARGGSGSGLGGSYYYYGMDV
#6	-	IGHJ3	21	ARGGVDYVWGSYRGGDAFDI
#7C	-	IGHJ6	24	ARGSGSGDYVSGYGSDYYYYGMDV
#99	-	IGHJ4	14	ARGQWLGSGYYDY
#1	IGHV1-3	IGHJ4	13	ARDNALGGSYFDY
#2?	IGHV3-21	IGHJ6	9	ARDANGMDV
#4	IGHV4-34	IGHJ6	28	ARGSGSGSGYEYYYYYGMDV
#16	-	IGHJ6	24	AGRFYCSGSGCGSGSGYYGSDV
#201	-	IGHJ3	17	ARRGSGWGSGSGDAFDI
#6	IGHV4-39	IGHJ3	19	ARGSGYSSSWGSGDWFDP
#12	IGHV1-46	IGHJ4	19	ARDGSYYGSGVYSGYTDY
#14	IGHV4-4	IGHJ4	18	ARGGSWGFDY
#20A	IGHV1-2	IGHJ6	17	ARGSGSGSYYYYYYGMDV
#31	IGHV3-23D	IGHJ6	21	ARGGSGSGYSGGDYYYYGMDV
#59	IGHV5-51	IGHJ4	12	ARGSDAYDSGFDY
#44B	IGHV3-48	IGHJ3	21	AKDGSAVVGSGADYYYYGMDV
#77	IGHV4-59	IGHJ4	14	ARDGSGSGWDEYFDY
#202	IGHV3-38	IGHJ4	14	ARSGGGSGLGEYFDY

ASSIGNMENT CRITERIA: V gene | J gene (shaded by family) | CDR3 length (aa) | CDR3 amino acid sequence | Shared V gene

A



Subset #2	2.7%
Subset #1	2.3%
Subset #4	0.9%
16 other major subsets	6.0%
Other stereotypes	18%
Heterogeneous	70%

Agathangelidis et al., Blood 2012;119(19), Fig. 1 · per-subset share derived from case counts

LTE Can Detect 19 of the Most Prevalent CLL Subsets

Table 1 depicts the 19 CLL subsets detected by LTE, broken out by shared V and J Family information. Most subsets share V and J information the CDR3 sequence is what allows LTE to differentiate one subset from another. Subset #2 is marked as putative because D gene information would be needed for definitive identification. CDR3s shown are examples.

Figure A. shows the relative size of each CLL subset identified by Agathangelidis et al.[1], based on VH CDR3 stereotype analysis of 7,596 Ig heavy-chain sequences from 7,424 CLL patients.

LTE tracks MRD from low-level disease to non-detection

Table 2

MRD Across Follow-up Timepoints

Sample	TP	MRD	% clonal cells	LymphoQuant freq.
IGH FR2				
Sample1	1	●	1.62E-01	1.47E-01
"	2	●	1.30E+00	4.12E-02
"	3	●	3.14E+00	1.51E-01
"	4	●	3.71E-04	7.10E-02
Sample2	1	●	4.56E-03	1.22E-01
"	2	●	3.53E-03	1.16E-01
Sample9	1	●	1.18E+00	1.19E-01
"	2	●	3.48E-02	2.57E-01
IGH FR3				
Sample4	1	○	ND	7.12E-02
Sample7	1	●	2.78E+00	5.38E-02
"	2	●	3.45E-01	1.55E-02
Sample8	1	●	1.14E-02	4.14E-03
"	2	●	4.39E-02	9.98E-03
"	3	●	6.99E-02	1.77E-02
"	4	●	8.60E-01	1.55E-02
"	5	●	8.05E-02	2.38E-02
"	6	●	2.42E+01	7.45E-03
Sample10	1	○	ND	2.93E-01
"	2	●	1.81E-03	1.51E-02
Sample12	1	○	ND	3.07E-01
IGK				
Sample3	1	●	1.77E-03	2.63E-01
"	2	●	7.75E-04	2.71E-02
Sample5	1	●	1.72E+00	1.17E-01
"	2	●	9.28E-02	1.44E-01
Sample6	1	●	1.28E-03	2.15E-01
"	2	●	4.05E-03	7.35E-02
"	3	●	4.36E-03	9.63E-02
"	4	●	1.88E-02	1.05E-01
"	5	○	ND	1.25E-01

● Detected ○ Not detected (ND) 11 patients · 29 timepoints

Table 3

Per-Timepoint Statistics Sample1 · IGH FR2

Metric	Timepoint 1	Timepoint 2
INPUT		
DNA input (ng)	55	229
Cell input (cells)	8,462	35,231
Read count	758,820	882,884
CLONE DETECTION		
Detected	Yes	Yes
Exact count	11,103	125,424
Exact frequency	1.5%	14.2%
Cum. mismatch-1 read count	14,538	159,197
Cum. mismatch-1 frequency	1.9%	18.0%
Cum. mismatch-2 read count	15,293	166,019
Cum. mismatch-2 frequency	2.0%	18.8%
CONFIDENCE		
Confidence −3	99.999%	99.999%
Confidence −4	95.389%	95.390%
Confidence −5	26.483%	26.486%
Confidence −6	3.030%	3.030%
QUANTIFICATION		
Clonal frequency	0.1623%	1.2954%
Cell equivalents (CE)	13.73	456.38
Cell equivalents per million (CEPM)	1,622.63	12,953.79
LYMPHOQUANT INTERNAL CONTROL		
LymphoQuant exact frequency	9.4%	2.8%
LymphoQuant cum. mismatch-2 frequency	14.7%	4.1%

Clonal frequency is derived from the LymphoQuant internal control; DNA input differs between timepoints (55 ng vs 229 ng).

LTE Tracks MRD Across Timepoints

Table 2 depicts MRD results from 3 LT MRD targets across 29 follow-up timepoints from 11 patients with multiple myeloma. The % clonal cells at each timepoint was calculated by normalizing to LQ. Open circles mark Not Detected (ND) timepoints.

Table 3 is a stylized version of the flat text output that LTE produces. LTE reports each result with its supporting context: DNA and cell input, exact and near-match read counts, confidence levels, cell equivalents, and the LQ control frequency.

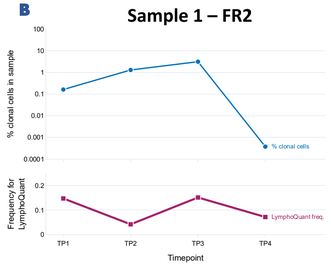
Figures B–D. Clonal burden over time for three patients. Top: % clonal cells (log scale). Bottom: LQ frequency at the same timepoints.

B. Low-level detection. Sample 1 dropped about four logs between TP3 and TP4 and was still detected at TP4 (34 clone reads).

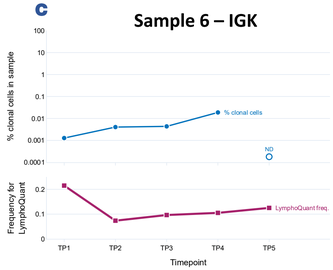
C. Dropout reported as ND. Sample 6 stayed between 0.001% and 0.02% across four timepoints and was Not Detected at TP5. LTE reports ND separately from zero.

D. Tracking rising burden. Sample 8 rose about three logs, from 0.011% at TP1 to 24.2% at TP6.

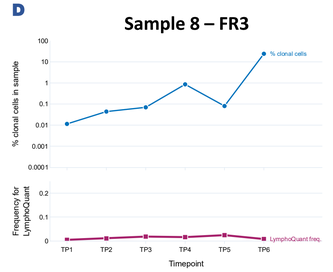
B



C



D

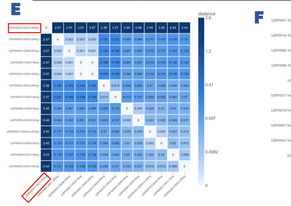


Diversity Metrics for Distinguishing Clonality

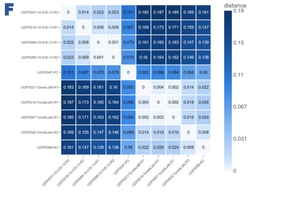
Table 4

Sample	Input DNA	Shannon Index	Simpson Index	Gini Index	Pielou's Evenness
UDP00044	200ng	2.653	0.892	0.787	0.735
UDP00046	200ng	2.522	0.862	0.807	0.688
UDP00048	200ng	2.650	0.890	0.790	0.718
UDP00050	200ng	2.651	0.887	0.788	0.714
UDP00058	50ng	2.514	0.870	0.810	0.696
UDP00035	200ng	**0.084**	**0.024**	0.963	**0.040**
UDP00076	50ng	2.452	0.862	0.821	0.721
UDP00036	200ng	2.466	0.874	0.823	0.705
UDP00062	50ng	2.650	0.889	0.789	0.729
UDP00078	50ng	2.534	0.879	0.810	0.697
UDP00064	50ng	2.564	0.875	0.802	0.700
UDP00042	200ng	2.536	0.871	0.808	0.678
UDP00080	50ng	2.559	0.856	0.793	0.704

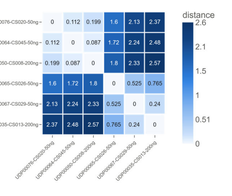
E



F



C



Diversity Metrics Allow Inference of Clonality – Table 4 shows aggregated statistics for an *IGH* Leader run of 13 samples. In this case only UDP00035 is a clonal sample and is statistically significantly ($p < 0.001$) (one-sided left tail for normal-distribution) less diverse than the other samples. **Figures E–G.** Distance matrices of Simpson's Index across run compositions. **E.** Same samples as Table 4. UDP00035 is highlighted with a red box. **F.** A run of 6 previously identified clonal samples and 6 non-clonal samples. **G.** A TRG-targeted run of 3 bi-clonal samples and 3 polyclonal samples.

Conclusion:

We present a data-driven evaluation of LTE v4.0.0, showing how its new features extend clonality assessment, CLL stereotyped subset classification, and longitudinal MRD tracking. **Clonality analysis is expanded through the addition of repertoire diversity metrics.** These metrics provide users another avenue to discern how diverse a sample is in isolation as well as in comparison to other samples. We demonstrate how diversity can be used to discern clonal patterns via distance matrices. **LTE accurately identifies the 19 most prevalent CLL subsets.** These classifications allow users to determine if their samples contain a CLL subset that could impact patient or clinical outcomes, providing another method to characterize a sample's immune repertoire in finer detail**. MRD performance is improved via LQ spike-in normalization and replicate combining**. While not demonstrated here, LTE v4.0.0 allows replicate combining for higher levels of MRD sensitivity and accuracy to patient clones. Applied to multiple myeloma samples from 11 patients, LTE v4.0.0 tracked each patient's clone across up to six follow-up timepoints and quantified clonal burden against LymphoQuant controls. **Together these results demonstrate LymphoTrack Enterprise v4.0.0 as a powerful suite of tools for researchers performing MRD, clonality, or immune repertoire studies.**

References:

1. Agathangelidis, A. *et al.* Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood* 119, 4467–4475 (2012).
2. Agathangelidis, A. *et al.* Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL. *Blood* **137**, 1365–1376 (2021).
3. Yu, H. *et al.* Spatial TCR clonality and clonal expansion in the in situ microenvironment of non-small cell lung cancer. *J. Immunother. Cancer* **13**, e012089 (2025).