High-throughput analysis of the T cell receptor gene repertoire in low-count monoclonal B cell lymphocytosis reveals a distinct profile from chronic lymphocytic leukemia

Andreas Agathangelidis¹,², Chrysi Galigalidou²,³, Lydia Scarfò¹, Theodoros Moysiadis²,⁴, Alessandra Rovida¹, Elisavet Vlachonikola², Electra Sofou², Fotis Psomopoulos², Anna Vardi², Pamela Ranghetti¹, Alexandra Siorenta⁵, Alex Galanis³, Kostas Stamatopoulos²,⁴, Anastasia Chatzidimitriou²,⁴*, Paolo Ghia¹*

*A.C. and P.G. contributed equally to this work.

¹Strategic Research Program on CLL and B-cell neoplasia Unit, Division of Experimental Oncology, Università Vita-Salute San Raffaele and IRCCS Ospedale San Raffaele, Milan, Italy
²Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece
³Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece
⁴Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
⁵Immunology and National Tissue Typing Center, General Hospital of Athens ‘G. Gennimatas’, Athens, Thessaloniki, Greece

Running title: T cell receptor repertoire analysis in LC-MBL


**Corresponding author**

Anastasia Chatzidimitriou  
Institute of Applied Biosciences  
Centre for Research and Technology Hellas  
57001 Thermi, Thessaloniki, Greece  
Phone: +30.2310.498474  
e-mail: achatzidimitriou@certh.gr

Text word count: 1500  
Number of figures and tables: 3  
Number of references: 15  
Number of supplemental files: 1
Monoclonal B-cell lymphocytosis (MBL) is an asymptomatic condition of monoclonal B cell expansions in the blood of healthy, mostly elderly, individuals.\textsuperscript{1,2} MBL is classified into 3 distinct subtypes: (i) “CLL-like” MBL (CD5\textsuperscript{+}CD23\textsuperscript{+}) that represents the vast majority of cases, (ii) “atypical CLL-like” MBL (CD5\textsuperscript{+}CD23\textsuperscript{−}CD20\textsuperscript{bright}) and (iii) “non CLL-like” MBL (CD5\textsuperscript{−}).\textsuperscript{3} “CLL-like” MBL is subdivided into 2 different categories based on clonal size; cases with 0.5-5\texttimes10\textsuperscript{9} cells/L are categorized as “high-count MBL” (HC-MBL), whereas those with <0.5\texttimes10\textsuperscript{9} cells/L as “low-count MBL” (LC-MBL).\textsuperscript{4} HC-MBL progresses to CLL requiring treatment at a rate of 1-2\% per year,\textsuperscript{5} whereas the risk of progression for “CLL-like” LC-MBL is negligible despite persisting over time.\textsuperscript{6,7}

Recently, we reported that the genomic profiles of LC-MBL, HC-MBL and ‘ultra-stable’ CLL (no disease progression for ≥10 years) are very similar.\textsuperscript{8} On these grounds, we proposed that the cross-talk between CLL progenitor cells and the microenvironment might represent a major driver at early disease stages. Relevant microenvironmental triggers might be provided by T cells, considering ample evidence for their implication in CLL pathogenesis\textsuperscript{9-11} and the existence of T cell expansions.\textsuperscript{12} Such clonal expansions were also evident in HC-MBL and LC-MBL\textsuperscript{6,13} however, the relevant studies faced important limitations, e.g. limited coverage of the expressed repertoire\textsuperscript{6} and sequencing depth\textsuperscript{13}, hence precluding firm conclusions from being drawn.

In order to overcome these limitations, we characterized the T cell receptor beta (TRB) chain gene repertoire using a high-throughput sequencing approach. Blood samples were collected from individuals from Val Borbera, Italy, where a LC-MBL cohort is regularly followed up (Supplemental Material).
Samples from aged-matched, healthy individuals without MBL from the same region were analyzed as controls; comparisons to CLL\textsuperscript{12} were also performed. The research protocol was approved by the Ethics Committee of the San Raffaele Institute and all participants gave written informed consent in accordance with the Declaration of Helsinki.

Five ml of blood were obtained from all individuals and processed within 24 hours. LC-MBL was diagnosed following a standardized flow cytometry approach\textsuperscript{6} (Supplemental Methods). We analyzed 48 samples from individuals with LC-MBL (“CLL-like” LC-MBL, n=41; “other” LC-MBL subtypes, n=7), and 17 samples from healthy controls (Supplemental Table 1). TRBV-TRBD-TRBJ gene rearrangements were PCR-amplified, sequenced on a MiSeq Sequencer and bioinformatically processed, as previously described.\textsuperscript{12} Overall, 2,357,648 distinct TRB clonotypes were identified. Of these, 1,006,126 (42.7%) were expanded (>1 read), whereas the remainder (1,351,522, 57.3%) concerned singletons (=1 read).

The relevance of age in shaping the TRB repertoire reported in a recent NGS study\textsuperscript{14} was also evident in our cohort; clonal expansions were present in all sample categories, including healthy donors. However, significantly different mean relative frequencies of expanded clonotypes were evident between sample categories (ANOVA, P<0.05). In more detail, expanded clonotypes were larger in “CLL-like” LC-MBL (0.018%, range: 0.003-0.14%) compared to “other” LC-MBL (0.007%, range: 0.001-0.04%) and the “healthy” category (0.01%, range: 0.002-0.012%) (Table 1 ; Supplemental Figure 1).

Two different approaches were followed to assess clonality in each sample: (i) the 10 “major” (most expanded) clonotypes; and, (ii) all expanded clonotypes
with an individual frequency of >1%. The first approach led to the identification of slightly higher clonality levels in LC-MBL versus healthy individuals, yet lower than those reported in CLL\textsuperscript{12} (ANOVA, P=0.3) (Figure 1 ; Supplemental Table 2). When considering clonotypes with a relative frequency >1%, the average number of expanded clonotypes for the “CLL-like” LC-MBL, “other” LC-MBL and healthy groups was 5 (range 0-27), 4 (range 2-6) and 4 (range 1-11), respectively. The average sum of relative frequencies was 20% in “CLL-like” LC-MBL, 17% in “other” LC-MBL and 15% in the healthy group (Supplemental Table 3). The average number of expanded clonotypes in CLL was 8 (range 2-27) and their average cumulative frequency was 28% (ANOVA, P=0.1). This trend raises the possibility of more persistent antigenic pressure, leading to more pronounced T cell expansions in CLL versus LC-MBL versus healthy individuals.

Moreover, we investigated whether the level of T cell clonality could be related to the size of the MBL clone and analyzed the bivariate correlation between the clonal cell size of LC-MBL cases and the cumulative frequency % of: (i) the 10 major clonotypes, and (ii) all clonotypes with an individual frequency of >1% by computing 2 correlation indices (Pearson Correlation Coefficient, Spearman’s rho Correlation Coefficient). A significant correlation was detected regarding (b) (Spearman’s rho=0.286, p<0.05), in line with our previous claim in CLL\textsuperscript{12} that T cell expansions could be driven, at least in part, by B-cell clone-associated antigen(s); alternatively, the expanded T cells and the CLL/MBL cells could recognize the same antigens. However, presently, the nature of the antigens selecting T cells in MBL remains largely unknown.
We also explored the relation between TRB clonality and age in LC-MBL and CLL in comparison to healthy donors. Following the approach of a recent NGS study, we divided the samples of each category into 3 age subgroups: “middle-aged” (average 39 years, range: 41-54 years), “aged” (average 62 years, range: 56-68 years), and “long-lived” (average 82 years, range: 69-93 years) (Supplemental Table 4). We observed a clear tendency of TRB clonality to increase with age in all categories (Supplemental Figure 2), being more pronounced in CLL (particularly) and LC-MBL compared to healthy individuals. This would indicate that T cell expansions are more relevant in these conditions.

TRBV gene repertoire analysis revealed strong biases in all sample categories (Figure 2), alluding to antigen selection. That said, in principle, this finding could also be attributed, at least in part, to the existence of a restricted gene pool in the study population, given that all samples were obtained from a restricted geographical area.

In “CLL-like” LC-MBL, 5 TRBV genes were identified with a frequency greater than 5%, collectively accounting for a significant fraction of the entire cohort (35.5%) (Supplemental Figure 3A). A similar scenario was noted in “other” LC-MBL, where 7 genes were dominant (>5%) and collectively represented 44.9% of the repertoire (Supplemental Figure 3B). The healthy group exhibited somewhat weaker selection biases: 4 genes had an individual frequency of >5% and a cumulative frequency of 27.9% (Supplemental Figure 3C).

Subsequently, we performed comparisons to CLL, where 6 genes displayed an individual frequency of >5% and collectively accounted for 38.7% of the
repertoire. Statistically significant differences concerned the expression of 12/46 (26.1%) TRBV genes (ANOVA, P<0.001), of which 3 were among the most frequent (TRBV28, TRBV5-1, TRBV19). Pairwise comparisons between all sample groups (Supplemental table 5) revealed distinct TRBV gene repertoire biases, which might reflect different antigen selection pressures between LC-MBL, CLL and healthy individuals.

Next, we searched for common TRB clonotypes across all sample categories of the present cohort and CLL\(^{12}\) considering the 10 “major”, hence, more relevant clonotypes of each sample. Most “major” TRB clonotypes (841/900, 93.4%) were present only in a single sample. Amongst the remainder, shared clonotypes, 28 were detected in 2 samples and 1 in 3 samples (Supplemental Table 6). Most shared clonotypes (15/28, 53.6%) were detected in samples of the same category; 8 in “CLL-like” LC-MBL and 7 in healthy individuals; the remaining 13 were found in samples from different categories. The average frequency of shared clonotypes was 4.4% with more than half (31/59, 52.5%) displaying a frequency of >1%.

In order to validate the relevance of shared clonotypes, we HLA-typed 10 LC-MBL cases with available material that altogether expressed 8 shared clonotypes. HLA gene restrictions were found in all sample pairs, indicating that T cells expressing shared clonotypes might recognize the same antigenic epitopes (Supplemental Table 7). These findings support 2 main conclusions: (i) shared clonotypes may recognize similar antigenic epitopes; and, (ii) LC-MBL and CLL display distinct T cell repertoires, given the scarcity of shared clonotypes between the two entities.
In order to better understand the selection processes that drive T cell expansions in MBL and CLL, we compared TRB clonotypes from the present study and CLL with: (i) 6,425 distinct clonotypes from different entities deposited in the IMGT/LIGM-DB and, (ii) expanded TRB clonotypes expressed by virus-specific T cells (VSTs) generated against EBV, CMV and BKV overlapping peptides.\textsuperscript{15} We identified 13 hits in the first case; 6 of these concerned viral infections, whereas the remainder were mostly reported in autoimmune conditions. All shared clonotypes were minor (mean relative frequency: 0.04\%, range: 0.01-0.7\%)(Supplemental Table 8). Only a single clonotype present in 2 samples of our cohort (from the “CLL-like” and “other” LC-MBL sample categories) was also expressed by VSTs. Hence, “major” TRB clonotypes appear to be associated with the intrinsic LC-MBL microenvironment rather than a microbial infection.

In conclusion, the existence of more pronounced T cell expansions in LC-MBL compared to aged-matched healthy individuals indicates that antigen drive could have an effect on the T cell repertoire, yet more limited compared to CLL. The nature of the implicated antigen(s) and whether they are related to the LC-MBL clone remains to be clarified. Furthermore, despite overall similar TRBV gene usage and degree of oligoclonality, shared clonotypes were scant, thus indicating that the antigenic stimuli and/or immune processes shaping the TR profiles in LC-MBL and CLL are probably different.
References


### Tables

**Table 1.** Detailed information regarding the size of TRB raw, processed and clonotype data for each sample category of the present cohort.

<table>
<thead>
<tr>
<th></th>
<th>“CLL-like” LC-MBL</th>
<th>“other” LC-MBL</th>
<th>“healthy”</th>
<th>all samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total number</td>
<td>average</td>
<td>range</td>
<td>total number</td>
</tr>
<tr>
<td><strong>Total raw reads</strong></td>
<td>7,603,175</td>
<td>185,443</td>
<td>24,672 - 504,220</td>
<td>1,502,577</td>
</tr>
<tr>
<td><strong>Filtered-in reads</strong></td>
<td>6,489,775</td>
<td>158,287</td>
<td>18,073 - 445,426</td>
<td>1,279,733</td>
</tr>
<tr>
<td><strong>Filtered-out reads</strong></td>
<td>1,113,400</td>
<td>27,156</td>
<td>3,540 - 72,911</td>
<td>222,844</td>
</tr>
<tr>
<td><strong>Unique clonotypes</strong></td>
<td>1,271,117</td>
<td>31,003</td>
<td>1,914 - 72,394</td>
<td>264,287</td>
</tr>
<tr>
<td><strong>Expanded clonotypes (&gt;1 read)</strong></td>
<td>567,560</td>
<td>13,843</td>
<td>664 - 34,575</td>
<td>120,442</td>
</tr>
<tr>
<td><strong>Singletons</strong></td>
<td>703,557</td>
<td>17,160</td>
<td>1,250 - 37,934</td>
<td>143,845</td>
</tr>
</tbody>
</table>
Figure Legends

**Figure 1.** TRB repertoire is oligoclonal in all entities of the present cohort, yet to different extent. “CLL-like” LC-MBL was more oligoclonal, followed by the “other” LC-MBL subtypes and the “healthy” compartment. Each lane corresponds to a single sample. Bright colors represent the 10 most expanded clonotypes based on their frequency ranking whereas the polyclonal background is depicted with light grey (i.e. the dominant clonotype in all samples is colored in blue).

**Figure 2.** Total TRBV gene repertoire for all entitied of the present cohort. Statistically significant differences in frequency of utilization were found for 12 of 46 identified TRBV genes. (ANOVA, * P<0.05, ** P<0.01 and *** P<0.001). Concerning the most frequently expressed TRBV genes, significant differences were identified for the TRBV28, TRBV19 and TRBV5-1 genes.