Restrictions in the T-cell repertoire of chronic lymphocytic leukemia: High-throughput immunoprofiling supports selection by shared antigenic elements

A Vardi, E Vlachonikola, M Karypidou, E Stalika, V Bikos, K Gemenetzi, C Maramis, A Siorenta, A Anagnostopoulos, S Pospisilova, N Maglaveras, I Chouvarda, K Stamatopoulos, A Hadzidimitriou

Cite this article as: A Vardi, E Vlachonikola, M Karypidou, E Stalika, V Bikos, K Gemenetzi, C Maramis, A Siorenta, A Anagnostopoulos, S Pospisilova, N Maglaveras, I Chouvarda, K Stamatopoulos, A Hadzidimitriou, Restrictions in the T-cell repertoire of chronic lymphocytic leukemia: High-throughput immunoprofiling supports selection by shared antigenic elements, *Leukemia* accepted article preview 1 December 2016; doi: 10.1038/leu.2016.362.

Received 30 August 2016; revised 23 October 2016; accepted 7 November 2016; Accepted article preview online 1 December 2016

RESTRICTIONS IN THE T-CELL REPERTOIRE OF CHRONIC LYMPHOCYTIC LEUKEMIA: HIGH-THROUGHPUT IMMUNOPROFILING SUPPORTS SELECTION BY SHARED ANTIGENIC ELEMENTS

Anna Vardi[1,2,3], Elisavet Vlachonikola[1], Maria Karypidou[1], Evangelia Stalika[1], Vasilis Bikos[4], Katerina Gemenetzi[1], Christos Maramis[1,5], Alexandra Siorenta[6], Achilles Anagnostopoulos[2], Sarka Pospisilova[4], Nikos Maglaveras[1,5], Ioanna Chouvarda[1,5], Kostas Stamatopoulos[1,7], Anastasia Hadzidimitriou[1,7].

[1]Institute of Applied Biosciences, CERTH, Thessaloniki, Greece

[2]Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece

[3]Medical School, University of Crete, Heraklion, Greece

[4]CEITEC, Masaryk University and University Hospital Brno, Brno, Czech Republic

[5]Laboratory of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Greece

[6]Immunology and National Tissue Typing Center, General Hospital of Athens 'G. Gennimatas', Athens, Greece

[7]Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden

RUNNING TITLE: High-throughput immunoprofiling of the T-cell repertoire in CLL

KEYWORDS: TR, CLL, stereotyped BcR, oligoclonality

FINANCIAL SUPPORT. Supported in part by H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe", by the EU; "MEDGENET, Medical Genomics and Epigenomics Network" (No.692298) by the EU. Part of the work was carried out with the support of core facilities of CEITEC – Central European Institute of Technology under CEITEC – open access project, ID number LM2011020, funded by the Ministry of Education, Youth and Sports of the Czech Republic under the activity "Projects of major infrastructures for research, development and innovations". PS and BV are financially

supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601) and by the Horizon2020 Programme Twinning (MEDGENET/2016-2018/no.692298).

CORRESPONDING AUTHOR:

Kostas Stamatopoulos, MD, PhD

Institute of Applied Biosciences, CERTH

57001 Thessaloniki, GREECE

Phone: +30.2310.498271

Fax: +30.2310.498270

e-mail: kostas.stamatopoulos@gmail.com

CONFLICT-OF-INTEREST DISCLOSURE

The authors have no relevant conflicts of interest to disclose.

ABSTRACT

Immunoglobulin (IG) gene repertoire restrictions strongly support antigen selection in the pathogenesis of chronic lymphocytic leukemia (CLL). Given the emerging multifarious interactions between CLL and bystander T cells, we sought to determine whether antigen(s) are also selecting T cells in CLL. We performed a large-scale, next-generation sequencing (NGS) study of the T-cell repertoire, focusing on major stereotyped subsets representing CLL subgroups with undisputed antigenic drive, but also included patients carrying non-subset IG rearrangements to seek for T-cell immunogenetic signatures ubiquitous in CLL. Considering the inherent limitations of NGS, we deployed bioinformatics algorithms for qualitative curation of T-cell receptor rearrangements, and included multiple types of controls. Overall, we document the clonal architecture of the T-cell repertoire in CLL. These T-cell clones persist and further expand overtime, and can be shared by different patients, most especially patients belonging to the same stereotyped subset. Notably, these shared clonotypes appear to be disease-specific, as they are found in neither public databases nor healthy controls. Altogether, these findings indicate that antigen drive likely underlies T-cell expansions in CLL and may be acting in a CLL subset-specific context. Whether these are the same antigens interacting with the malignant clone or tumor-derived antigens remains to be elucidated.

INTRODUCTION

Chronic lymphocytic leukemia (CLL) is characterized by remarkable skewing of the B-cell receptor immunoglobulin (BcR IG) repertoire, culminating in the existence of subsets of patients with quasi-identical, stereotyped BcR IGs. This implies antigen selection in the natural history of CLL, ultimately affecting clonal behavior and clinical outcome. Indeed, immunogenetic characteristics of the clonal BcR IGs serve as established disease prognostic markers, typically exemplified by the subdivision of CLL cases into those with no or limited somatic hypermutation (SHM) within the clonotypic IGs ("unmutated" U-CLL) who generally experience an aggressive disease course versus those with a significant SHM load within their IG genes ("mutated", M-CLL) who display a considerably more indolent disease (1-3). Furthermore, certain stereotyped CLL subsets are associated with either aggressive (i.e. subset #1 and subset #2) or indolent clinical course (i.e. subset #4) (2, 4-6). This immunogenetic evidence for antigen involvement in CLL pathogenesis has been recently complemented by clinical evidence where drugs inhibiting key elements of BcR signaling have proven highly efficacious, leading to their accelerated approval for the treatment of patients with CLL (7-10).

Extracellular cues that are vital for CLL cell survival and proliferation are also provided by accessory cells, which, together with cytokines and chemokines, form the tumor microenvironment (11). T cells appear to be a critical component of this microenvironment, as evidenced by animal model experiments demonstrating that CLL cell engraftment and clonal expansion depend on trophic signals provided by autologous T cells (12). However, despite accumulating evidence of an intricate crosstalk between T cells and CLL B cells (13-15), little is known regarding the role of antigen(s) in the selection and activation of cognate T cells. This is highly relevant in light of CLL and T cell interactions inducing T-cell tolerance, especially since reversal of this tolerance by immunomodulating drugs acting at the level of the immune synapse between clonal B and T cells shows promising results in clinical trials (16-18).

Using low-throughput subcloning techniques followed by Sanger sequencing in 58 CLL patients, we recently reported T-cell receptor beta chain (TRB) gene repertoire restriction, pointing to antigenic selection (19). However, due to the inherent limitations of low-throughput analysis, definitive conclusions were not possible. Here, we sought to advance the analytical depth of our approach by employing high-throughput, next generation sequencing (NGS). Our analysis was intentionally biased towards patients expressing stereotyped BcR IGs, therefore cases most evidently selected by antigen. Additionally, we

included M-CLL and U-CLL cases not assigned to stereotyped subsets in order to obtain a comprehensive view of the T-cell repertoire in CLL.

We report pronounced T-cell oligoclonality, with clonotypes that persist and expand over time and shared clonotypes between different patients, especially those belonging to stereotyped subsets, which appear to be disease-specific. These findings implicate selection by antigenic elements that may be acting in a CLL subset-specific context. Whether these are the same antigens interacting with the malignant clone or tumor-derived antigens remains to be elucidated.

METHODS

Patient group

We analyzed 57 samples from 32 untreated CLL patients (**Supplemental Table 1**), and 2 control samples from 2 healthy individuals (ages 70 and 80 year-old). Twenty-four of the 32 CLL cases were selected based on their expression of stereotyped BcR IGs which assigned them to well-annotated CLL subsets. More specifically, these cases belonged to subset #4 (n=12), subset #1 (n=8), and subset #2 (n=4) (**Supplemental Table 2**). The remaining 8 CLL cases were not assigned to stereotyped subsets, carrying either mutated (<98% IGHV identity to germline, n=5) or truly unmutated (100% IGHV identity to germline, n=3) BcR IGs. Two subset #4 and one subset #2 case were analyzed over time (two timepoints, median sampling interval 23 months - all patients remained untreated throughout the sampling period). No case had evidence of infection (neither signs nor symptoms) at sampling. The great majority of CLL patients were CMV and EBV seropositive, and so were both healthy individuals (**Supplemental Table 3**). The local Ethics Review Committee approved the study and written informed consent form was obtained from all individuals.

PCR amplification of TRBV-TRBD-TRBJ gene rearrangements and library preparation

Total cellular RNA was primarily isolated from peripheral blood mononuclear cells (n=40), but also purified $CD4^+$ and $CD8^+$ T-cell subpopulations (n=9 and n=7, respectively), bone marrow (n=2) or fresh lymph node (LN) tissue (n=1). Isolation of $CD4^+$ and $CD8^+$ T-cell subpopulations was performed by negative selection (RosetteSep™ Human $CD4^+$/$CD8^+$ T-Cell Enrichment Cocktail, Stemcell Technologies), resulting in >95% purity by flow cytometry. In all samples, starting absolute T-cell count exceeded $0.5x10^6$ cells, in order to determine actual repertoire diversity instead of over-amplifying the same TRBV-TRBD-TRBJ rearrangements. For repertoire analyses, each category of samples was studied separately (PBMCs, BM, LN, $CD4^+$ and $CD8^+$ T cells).

Except for two samples for which gradient RNA quantities were tested, in all other samples 1 μg of RNA was reverse transcribed to complementary DNA (cDNA) using the SuperScript II RT kit (Invitrogen Life technologies, USA). TRBV-TRBD-TRBJ gene rearrangements were RT-PCR amplified according to the BIOMED-2 collaborative protocol (20). Amplicons were gel-purified with the QIAGEN DNA purification columns (QIAGEN, Hilden, Germany) and used for library preparation (TruSeq LT, Illumina) according to the manufacturer's instructions. PhiX was used at a 20% concentration to optimize library diversity. Four runs were performed on the Illumina platform using the MiSeq Reagent Kit v2 (2x250bp).

Paired-end protocol allowed sequencing of the TRB complementarity-determining region 3 (CDR3) twice/read, thus increasing the accuracy of results. Still, considering the inherent limitations of PCR-based NGS, the experimental setup included multiple internal controls: (i) "sample replicates", i.e. starting from aliquots of a single peripheral blood sample containing the same as well as varying T-cell counts (2 duplicates and 1 quadruplicate, range 0.5-18x10^6 T cells), (ii) "PCR replicates", i.e. starting from the same RNA sample [2 duplicates and 1 quadruplicate, containing the same as well as varying RNA quantity (range 1.0-5.8 μg)], and, (iii) "sequencing replicates", i.e. starting from the same PCR product but sequenced in separate MiSeq runs (5 duplicates).

Bioinformatics processing of raw data, definitions and interpretation

Initial data filtering was performed by the Illumina signal-processing software, leading to the rejection of low quality and erroneous sequences. Sequences were assigned to samples based on incorporated indexes, and sequence segments corresponding to the adapters were trimmed. In order to further increase the accuracy of results, raw NGS reads were subjected to a purpose-built, bioinformatics algorithm performing: (i) length and quality filtering of raw reads; (ii) merging of filtered-in paired reads via local alignment; (iii) length and quality filtering of stitched sequences. Detailed length, quality and overlap rules that were applied for the analysis are provided in **Supplemental Table 4**.

Filtered-in sequences were submitted to IMGT/HighV-QUEST tool (http://www.imgt.org), and metadata was processed by an in-house bioinformatics pipeline designed for clonotype computation and repertoire analysis. Only productive TRBV-TRBD-TRBJ gene rearrangements were included in the analysis. TRBV-TRBD-TRBJ gene rearrangements carrying TRBV genes with <95% germline identity were also discarded as sequences with unacceptable error rate, given the lack of somatic hypermutation in T cells.

Clonotypes were computed as unique pairs of TRBV genes and CDR3 amino acid sequences within a sample. Clonotypes were considered expanded (forming clusters) when they contained ≥2 sequences, with the ten most expanded clonotypes within a sample referred to as major; otherwise they were considered as "singletons". The relative frequency of each clonotype/sample was calculated as the number of rearrangements corresponding to the clonotype divided by the total number of productive, filtered-in rearrangements for that particular sample. For repertoire analysis, clonotypes rather than single rearrangements were considered in order to avoid potential biases due to expansion following antigenic stimulation, i.e. individual TRBV gene frequencies within a sample were calculated as the

number of clonotypes using particular TRBV genes over the total number of clonotypes.

Comparison to healthy controls and public data mining

We performed clonotype comparison across all 32 CLL patients, as well as against the two healthy controls. Clonotypes from replicate samples as well as $CD4^+$ and $CD8^+$ T-cell clonotypes belonging to the same patient were concatenated, so that all clonotypes of each patient would be included in the comparison. The 10 most expanded clonotypes of each CLL sample, referred to as "major clonotypes", were compared to all clonotypes of the healthy controls (n=248,593), as well as a panel of 6,318 non-redundant, well-annotated, unique clonotypes from various entities, available to our group (n=2,316) or retrieved from the IMGT/LIGM-DB sequence database (http://www.imgt.org/IMGTindex/LGM.html) (n=4,002).

HLA typing

Typing of the HLA-A, -B, -C loci for low resolution and HLA-DRB1 locus for allelic level high-resolution determination was performed as described previously (19).

Statistical analysis and visualization tools

Descriptive statistics for discrete parameters included counts and frequency distributions. For quantitative variables, statistical measures included means, medians, and min-max values. The significance of bivariate/multivariate relationships between variables was assessed using the unpaired t-test, and the non-parametric Kruskal Wallis and Mann-Whitney tests. Cox-regression and hazard ratio tests were applied for assessing the predictive value of T-cell clonality regarding time-to-first-treatment (TTFT). TTFT was evaluated from the sampling date until the date of initial treatment; untreated cases were censored at the time of last follow-up. For all comparisons a significance level of p=0.05 was set. All statistical analyses were performed using the statistical Package GraphPad Prism version 5.03 (GraphPad Software, Inc., San Diego, USA).

Data access

Raw TR sequence data discussed can be found under accession number SRR3737053 in GenBank sequence database (www.ncbi.nlm.nih.gov/genbank/).

RESULTS

The T-cell repertoire of CLL is skewed, especially in CD8$^+$ T cells

Overall, 19,513,078 filtered-in sequences were obtained. Only productive TRBV-TRBD-TRBJ rearrangements were evaluated (85.7% of filtered-in sequences, median 324,058/sample). For repertoire analysis, clonotypes (i.e. TRB rearrangements with identical TRBV gene usage and amino acid CDR3 sequence) were considered. The median number of distinct clonotypes/sample was 34,554 (range: 2,193-166,078); the median numbers of expanded clonotypes and singletons/sample were 12,507 (range: 812-57,035) and 21,778 (range: 1,381-122,586), respectively.

Only 6 genes [TRBV12-3/12-4 (8.4%), TRBV29-1 (6.9%), TRBV27 (6.7%), TRBV19 (6.2%), TRBV7-9 (5.4%), and TRBV6-5 (5.1%)] accounted for more than one-third of the repertoire. No significant TRBV gene repertoire differences were identified between U-CLL versus M-CLL or amongst different stereotyped subsets. Of note, the frequency of individual TRBV genes differed significantly when comparing the CLL CD8$^+$ versus CD4$^+$ T-cell repertoire. In particular, the TRBV27 and TRBV7-9 genes predominated in CD8$^+$ versus CD4$^+$ T cells (6.4% versus 3.9%, p=0.001 and 4.6% versus 2.6%, p=0.001, respectively), whereas the opposite held for the TRBV5-1 and TRBV7-2 genes (4.0% versus 6.8%, p=0.001 and 3.1% versus 6.0%, p=0.002, respectively) (**Supplemental Tables 5-6**).

Asymmetric TRBV gene usage characterized healthy samples as well, although the small number of samples analyzed precludes comprehensive healthy TRBV gene repertoire analysis (**Supplemental Table 7**). However, the CLL TRB repertoire was significantly more oligoclonal compared to healthy controls as reflected in a median cumulative frequency of the 10 most expanded clonotypes/sample of 23.6% in CLL versus 5.1% in healthy controls (p<0.05) (**Figure 1A; Supplemental Table 8**).

To investigate whether T-cell clonality is associated to disease course, we evaluated TTFT in 22/24 CLL patients for whom PBMC samples were tested (measured from the date of sampling, when T-cell clonality was assessed, until the date of treatment or last follow-up date for untreated patients). The cumulative frequency of the 10 most expanded clonotypes/patient did not reach statistical significance as an independent predictor for TTFT when treated as a quantitative variable (p=0.54). When we divided our patients into 2 clonality categories, above (n=11) or below (n=11) the median cumulative frequency of the 10 most expanded clonotypes (22.3%), we found that the former category was associated

with higher probability for treatment need, albeit without reaching statistical significance (Hazard ratio 1.7, p=0.34).

The distinct CLL immunogenetic subgroups analyzed did not exhibit major differences regarding oligoclonality. Clonality stemmed mainly from the $CD8^+$ T-cell compartment, where the median cumulative frequency of the 10 most expanded clonotypes/sample was 43.6% versus only 5.3% for the $CD4^+$ T-cell samples, respectively ($p<0.001$) (**Figure 1B, Supplemental Table 8**).

Longitudinal analysis reveals clonal persistence

To assess clonal dynamics over time, one subset #2 and two subset #4 CLL patients were studied longitudinally, in two successive time points (median sampling interval 23 months). For the patient assigned to subset #2, PBMCs were analyzed in both time points. The TRBV gene repertoire remained remarkably stable overtime (**Figure 2A**). For the two subset #4 patients, the first time point involved PBMC analysis, whereas the second time point involved analysis of isolated $CD4^+$ and $CD8^+$ T-cell subpopulations, therefore TRBV gene repertoire comparison was not performed.

Major clonotypes (i.e. the 10 most expanded clonotypes/sample) were found to persist in all three studied cases. In the subset #2 case, 10.4% (3970/38054) of all clonotypes and 17.4% (2286/13134) of all expanded clonotypes from the first time point persisted over time. Interestingly, in the second time point the persisting clonotypes constituted a larger fraction of the total repertoire (31.9% of all clonotypes and 40.9% of all expanded clonotypes, respectively). In the two subset #4 patients, 14.9% (2576/17250) and 27.6% (9918/35962) of all clonotypes identified in the PBMC samples were found in subsequent time points (11.8% and 21.0% retained in the $CD4^+$ T-cell compartment, and 3.1% and 6.6% retained in the $CD8^+$ T-cell compartment, respectively). Similarly, 16.7% (1247/7455) and 35.3% (4984/14125) of all expanded clonotypes identified in the PBMC samples persisted in the following time points (13.1% and 26.8% retained in the $CD4^+$ T-cell compartment, and 3.6% and 8.5% retained in the $CD8^+$ T-cell compartment, respectively). Most major persisting clonotypes were contained within the $CD8^+$ T-cell subpopulation **(Figure 2B)**.

Shared T-cell receptor clonotypes in CLL: mostly disease-specific

Comparisons across CLL patients and against the two healthy controls revealed that 162,179/2,464,321 (6.6%) CLL clonotypes were shared by two or more patients, 96.8% of which were not found in either healthy control. Most (79.2%) public CLL clonotypes

concerned pairs of patients, and few were found in 3 (13.3%) or ≥4 (7.5%) CLL patients. No pattern of similar disease characteristics was discerned among patients sharing clonotypes. Highly similar results were obtained when performing the same analysis excluding singletons (data not shown).

In most cases, public CLL clonotypes were low-frequent. Hence, we performed the same analyses considering only the major clonotypes of each CLL sample and found: (i) two major clonotypes shared among 2 pairs of subset #4 patients, (ii) one major clonotype shared among a pair of subset #1 patients (iii) one major clonotype shared among a subset #1 and a subset #2 patient, and, (iv) two major clonotypes shared among CLL patients belonging to different immunogenetic subgroups (i.e. subset and non-subset cases); one of these was identified in healthy controls, albeit at very low frequencies (0.04% and 0.0005%, respectively). In all subset cases sharing major clonotypes, HLA typing revealed relevant HLA restrictions (**Table 1, Supplemental Table 9**).

We next performed cross-entity comparisons among the identified CLL major clonotypes and a panel of 6,318 non-redundant, unique clonotypes from various entities deposited in the IMGT/LIGM-DB sequence database or available to our group. As shown in **Table 1**, only a single public major CLL clonotype matched with a public database entry, corresponding to a T-large granular lymphocytic leukemia (LGL) patient. This clonotype was not subset-specific, but rather shared between a subset #4 patient and a M-CLL patient not assigned to any stereotyped subset. Among the patient-specific ('private') CLL major clonotypes, we found matches with: (i) two EBV-specific clonotypes, (ii) a T-cell clonotype identified in a T-LGL patient, and, (iii) a T-cell clonotype from an individual with CLL-like monoclonal B-cell lymphocytosis (MBL), a condition considered as premalignant to CLL. Additionally, we found a match between a $CD4^+$ T-cell clonotype belonging to a subset #1 patient (P17) with a T-cell clonotype from another subset #1 CLL patient previously analyzed by our group through classic subcloning and Sanger sequencing **(Table 2)**: this pair of subset #1 patients shared the HLA-DRB*13:03 allele. Hence, overall, public clonotypes shared between different CLL cases were mostly disease-specific, i.e. not found amongst other entities.

Clonotype sharing in different tissue microenvironments

For one subset #1 patient, synchronous peripheral blood and lymph node samples were tested. The TRBV gene repertoire was almost identical in both samples (**Figure 3A**). The extent of clonality was also similar. In particular, the median cumulative frequency of the 10 most expanded clonotypes/sample was 8.8% in PBMCs versus 7.3% in the lymph node and

the percentage of expanded clonotypes over the total repertoire of each sample was 37.2% (15072/40503) versus 38.1% (15639/41010), respectively. Overall, 18.7% (7576/40503) clonotypes and 24.5% (3696/15072) expanded clonotypes identified in the PBMC sample were also identified in the lymph node sample, including two of the major clonotypes that were common for both tissue samples (**Figure 3B**).

Replicates

First, we examined "sample replicates" i.e. used aliquots of peripheral blood containing varying numbers of T cells (1 quadruplicate and 2 duplicates). Overall, despite limited fluctuation of the TRBV gene repertoire, the dominant T-cell clonotypes showed remarkable consistency for the tested T-cell range (0.5-18 x 10^6 cells), proving that sample quantity ensured adequate repertoire profiling depth (**Supplemental Figures 1,2**).

We then examined "PCR replicates", i.e. starting from the same RNA sample (1 quadruplicate and 2 duplicates), containing varying RNA quantity. Again, limited fluctuations of the TRBV gene repertoire were noted in few instances, however the dominant T-cell clonotypes were practically identical, ensuring that both cDNA and NGS library preparation protocols are robust, at least for the RNA range tested (1.0-5.8 μg) (**Supplemental Figures 3,4**).

Finally, we examined "sequencing replicates", i.e. starting from the same PCR product but sequenced in separate MiSeq runs, and actually different MiSeq machines (5 duplicates). Both the TRBV gene repertoire and the dominant clonotypes were remarkably consistent, proving that the sequencing procedure is indeed reproducible (**Supplemental Figures 5,6**).

DISCUSSION

NGS immunoprofiling holds the potential to reveal the architecture of complex repertoires, however still presents limitations and pitfalls. For the present analysis of the TRB repertoire in CLL, we followed a systematic and stringent approach including multiple controls, and focused on CLL cases assigned to the most populated and best-characterized CLL stereotyped subsets representing distinct entities with opposite clinical course and outcome (2). We also included CLL patients carrying non-subset BcR IG rearrangements investigating whether certain T-cell immunogenetic features may be ubiquitous in CLL, regardless of the particular IG receptor that is expressed.

Our findings provide solid evidence for T-cell oligoclonality in all CLL patients analyzed, stemming mainly from the $CD8^+$ T-cells. A similar extent of TR oligoclonality was noted between patients with distinct BcR IG, thus sharply contrasting the polyclonal profile of age-matched healthy controls, suggestive of antigenic selection. That said, the TRBV gene frequency distribution did not differ significantly between CLL patients and healthy controls, prompting speculations that it might reflect either differential primer amplification efficacies or naturally occurring preferential TRBV gene rearrangements (21, 22). Nevertheless, the statistically significant differences observed between $CD4^+$ versus $CD8^+$ TRBV gene repertoire favor the latter scenario.

A major finding of the present study concerned clonal persistence over time in CLL patients belonging to different immunogenetic subgroups of the disease (i.e. subsets #2 and #4). Importantly, the persisting clonotypes, found within both the $CD4^+$ and $CD8^+$ T-cell compartments, constituted a larger fraction of the total repertoire in subsequent time points in all tested cases. Most major persisting clonotypes corresponded to $CD8^+$ T-cells, however this might be attributed to the greater extent of clonality in the $CD8^+$ T-cell repertoire. On these grounds, it is reasonable to argue that the selecting antigens are also persistent, driving the respective T-cell clones to further expansion. Despite the small number of samples analyzed, clonal persistence was evident in two distinct immunogenetic subgroups of the disease with sharp clinicobiological differences (i.e. subsets #2 and #4), thereby alluding to a generic CLL characteristic. It would be interesting to test this hypothesis in more CLL samples, including cases that are not assigned to any stereotyped subset. That said, ongoing interactions with the selecting antigen(s) emerge as a recurring theme in CLL pathobiology, being evident in both the clonotypic IGs (most notably in subset #4) (23-26) and the TRs of the reactive T cells.

As expected for individuals who are not HLA-matched, most of the identified clonotypes were patient-specific (“private”) and only few were "public", i.e. shared by ≥2 patients. Moreover, most "public" clonotypes were represented at low frequencies, therefore had dubious biological significance. However, and perhaps of most biological relevance, we also identified major clonotypes shared between different patients. These public major clonotypes tended to cluster among pairs of patients assigned to the same stereotyped subset, where the CLL cells are most likely selected by shared antigenic elements (27, 28). Interestingly, they were not identified in the examined healthy controls, despite both expressing HLA-A*24, i.e. fulfilling the identified HLA restriction for at least one major clonotype shared among two subset #1 patients. Overall, the aforementioned findings allude to selection of the T-cell repertoire by restricted antigenic epitopes in a subset-specific context. Thus, they raise the possibility that the implicated antigens might be those selecting the CLL progenitors or even the malignant cells themselves. Alternatively, they could correspond to CLL subset-specific IG structures acting as idiotypic neoantigens (29). Recent work combining crystallographic and functional studies provided exciting novel results regarding the epitopes involved in homotypic CLL BcR interactions. Such epitopes appear to indeed be subset-specific, thus elucidating the structural basis of autonomous activation of CLL B-cells and reconciling the existence of a shared pathogenetic mechanism with the distinct clinicobiological profiles of CLL subsets (30). In light of these findings, these subset-specific epitopes could be tested as to whether they might be recognized/bound by cognate T cells. If successful, such endeavors will assist in clarifying the nature of antigenic elements that are responsible for T-cell repertoire skewing in CLL.

In conclusion, massive parallel sequencing documents the pronounced skewing of the TR repertoire in CLL, supporting selection by restricted antigenic elements. These elements appear to persist over time, driving the respective T-cell clones to further expansion. Expanded clonotypes are shared among patients, most especially among those assigned to the same stereotyped subset. Whether they recognize the same antigens interacting with the malignant clone or tumor-associated epitopes, possibly contained within the idiotypic IG, remains to be elucidated and may be of particular benefit in light of the clinical testing of immune checkpoint inhibitors, designed to boost endogenous T-cell anti-tumor responses (18, 31-32).

REFERENCES

1. Stamatopoulos K, Belessi C, Moreno C, Boudjograh M, Guida G, Smilevska T, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood. 2007;109(1):259-70.
2. Baliakas P, Hadzidimitriou A, Sutton LA, Minga E, Agathangelidis A, Nichelatti M, et al. Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. Lancet Haematol. 2014;1(2):e74-84.
3. Tobin G, Thunberg U, Johnson A, Eriksson I, Soderberg O, Karlsson K, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood. 2003;101(12):4952-7.
4. Agathangelidis A, Darzentas N, Hadzidimitriou A, Brochet X, Murray F, Yan XJ, et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. Blood. 2012;119(19):4467-75.
5. Baliakas P, Agathangelidis A, Hadzidimitriou A, Sutton LA, Minga E, Tsanousa A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. Blood. 2015;125(5):856-9.
6. Gounari M, Ntoufa S, Apollonio B, Papakonstantinou N, Ponzoni M, Chu CC, et al. Excessive antigen reactivity may underlie the clinical aggressiveness of chronic lymphocytic leukemia stereotyped subset #8. Blood. 2015;125(23):3580-7.
7. Burger JA, Tedeschi A, Barr PM, Robak T, Owen C, Ghia P, et al. Ibrutinib as Initial Therapy for Patients with Chronic Lymphocytic Leukemia. N Engl J Med. 2015;373(25):2425-37.
8. Brown JR, Barrientos JC, Barr PM, Flinn IW, Burger JA, Tran A, et al. The Bruton tyrosine kinase inhibitor ibrutinib with chemoimmunotherapy in patients with chronic lymphocytic leukemia. Blood. 2015;125(19):2915-22.
9. Brown JR, Byrd JC, Coutre SE, Benson DM, Flinn IW, Wagner-Johnston ND, et al. Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110delta, for relapsed/refractory chronic lymphocytic leukemia. Blood. 2014;123(22):3390-7.
10. Byrd JC, Furman RR, Coutre SE, Flinn IW, Burger JA, Blum KA, et al. Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. N Engl J Med. 2013;369(1):32-42.
11. Burger JA, Gribben JG**.** The microenvironment in chronic lymphocytic leukemia (CLL) and other B cell malignancies: insight into disease biology and new targeted therapies. Semin Cancer Biol. 2014;24:71-81.
12. Bagnara D, Kaufman MS, Calissano C, Marsilio S, Patten PE, Simone R, et al. A novel adoptive transfer model of chronic lymphocytic leukemia suggests a key role for T lymphocytes in the disease. Blood. 2011;117(20):5463-72.
13. Gorgun G, Holderried TA, Zahrieh D, Neuberg D, Gribben JG**.** Chronic lymphocytic leukemia cells induce changes in gene expression of CD4 and CD8 T cells. J Clin Invest. 2005;115(7):1797-805.
14. Os A, Burgler S, Ribes AP, Funderud A, Wang D, Thompson KM, et al. Chronic lymphocytic leukemia cells are activated and proliferate in response to specific T helper cells. Cell Rep. 2013;4(3):566-77.
15. Ramsay AG, Johnson AJ, Lee AM, Gorgun G, Le Dieu R, Blum W, et al. Chronic lymphocytic leukemia T cells show impaired immunological synapse formation that can be reversed with an immunomodulating drug. J Clin Invest. 2008;118(7):2427-37.

16. Ramsay AG, Clear AJ, Fatah R, Gribben JG. Multiple inhibitory ligands induce impaired T-cell immunologic synapse function in chronic lymphocytic leukemia that can be blocked with lenalidomide: establishing a reversible immune evasion mechanism in human cancer. Blood. 2012;120(7):1412-21.
17. Ramsay AG, Evans R, Kiaii S, Svensson L, Hogg N, Gribben JG. Chronic lymphocytic leukemia cells induce defective LFA-1-directed T-cell motility by altering Rho GTPase signaling that is reversible with lenalidomide. Blood. 2013;121(14):2704-14.
18. Shanafelt TD, Ramsay AG, Zent CS, Leis JF, Tun HW, Call TG, et al. Long-term repair of T-cell synapse activity in a phase II trial of chemoimmunotherapy followed by lenalidomide consolidation in previously untreated chronic lymphocytic leukemia (CLL). Blood. 2013;121(20):4137-41.
19. Vardi A, Agathangelidis A, Stalika E, Karypidou M, Siorenta A, Anagnostopoulos A, et al. Antigen Selection Shapes the T-cell Repertoire in Chronic Lymphocytic Leukemia. Clin Cancer Res. 2016;22(1):167-74.
20. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia. 2003;17(12):2257-317.
21. Venturi V, Kedzierska K, Price DA, Doherty PC, Douek DC, Turner SJ, et al. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. Proc Natl Acad Sci U S A. 2006;103(49):18691-6.
22. Quigley MF, Greenaway HY, Venturi V, Lindsay R, Quinn KM, Seder RA, et al. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. Proc Natl Acad Sci U S A. 2010;107(45):19414-9.
23. Sutton LA, Kostareli E, Hadzidimitriou A, Darzentas N, Tsaftaris A, Anagnostopoulos A, et al. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen. Blood. 2009;114(20):4460-8.
24. Kostareli E, Sutton LA, Hadzidimitriou A, Darzentas N, Kouvatsi A, Tsaftaris A, et al. Intraclonal diversification of immunoglobulin light chains in a subset of chronic lymphocytic leukemia alludes to antigen-driven clonal evolution. Leukemia. 2010;24(7):1317-24.
25. Sutton LA, Kostareli E, Stalika E, Tsaftaris A, Anagnostopoulos A, Darzentas N, et al. Temporal dynamics of clonal evolution in chronic lymphocytic leukemia with stereotyped IGHV4-34/IGKV2-30 antigen receptors: longitudinal immunogenetic evidence. Mol Med. 2013;19:230-6.
26. Sutton LA, Rosenquist R. Clonal evolution in chronic lymphocytic leukemia: impact of subclonality on disease progression. Expert Rev Hematol. 2015;8(1):71-8.
27. Murray F, Darzentas N, Hadzidimitriou A, Tobin G, Boudjogra M, Scielzo C, et al. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. Blood. 2008;111(3):1524-33.
28. Chu CC, Catera R, Zhang L, Didier S, Agagnina BM, Damle RN, et al. Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. Blood. 2010;115(19):3907-15.
29. Duhren-von Minden M, Ubelhart R, Schneider D, Wossning T, Bach MP, Buchner M, et al. Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. Nature. 2012;489(7415):309-12.

30. Gounari M, Minici C, Ubelhart R, Duehren-von Minden M, Schneider D, Tasdogan A, et al. Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukemia. 21st Congress of EHA; 2016; Copenhagen, Denmark: Haematologica. 2016; 9.
31. McClanahan F, Hanna B, Miller S, Clear AJ, Lichter P, Gribben JG, et al. PD-L1 checkpoint blockade prevents immune dysfunction and leukemia development in a mouse model of chronic lymphocytic leukemia. Blood. 2015;126(2):203-211.
32. McClanahan F, Riches JC, Miller S, Day WP, Kotsiou E, Neuberg D, et al. Mechanisms of PD-L1/PD-1–mediated CD8 T-cell dysfunction in the context of aging-related immune defects in the Eμ-TCL1 CLL mouse model. Blood. 2015;126(2):212-221.

FIGURE LEGENDS

Figure 1. CLL T-cell repertoire is oligoclonal, with clonality stemming mainly from CD8$^+$ T cells. A. T-cell clonality analysis for each immunogenetic CLL subgroup (subset #1, subset #2, subset #4 and cases with heterogeneous BcR IG rearrangements) versus healthy individuals. Each lane corresponds to a single PBMC sample (replicate PBMC samples have been excluded). B. T-cell clonality analysis in CD8$^+$ versus CD4$^+$ T-cells. Here, each lane corresponds to a CD4$^+$ or CD8$^+$ T-cell sample from CLL patients.

The colored tiles of each lane depict the relative size of the 10 most expanded clonotypes within the respective sample. The outer tile (shown in white in all instances) depicts the polyclonal background, i.e. the cumulative frequency of all other clonotypes within the sample.

Figure 2. Over time analysis of the TRBV repertoire and clonal drift. A. The TRBV gene repertoire in a subset #2 CLL patient remains stable over time. The 10 most frequent TRBV genes are depicted. B. Each column represents the 10 most expanded clonotypes within the tested sample. The persisting clonotypes for each patient are depicted with the same color. The lower tile (shown in white in all instances) contains major clonotypes that were not identified within the 10 most expanded clonotypes of the respective longitudinal sample. Major clonotypes persist in all 3 CLL cases tested.

Figure 3. Repertoire comparisons in synchronous PBMC and lymph node samples. A. The TRBV gene repertoire is almost identical in the two tissues. The 10 most frequent TRBV genes are depicted. B. Two major clonotypes are common among the two tissue samples (shown in color). The lower tile (shown in white) contains the remaining 8 major clonotypes of each sample.

Patient ID	Immunogenetic subgroup	Sample type	TRBV GENE	CDR3 AMINO ACID SEQUENCE	matches with public database clonotypes or healthy controls	HLA restriction
P8	subset #4	PBMCs	TRBV7-2	ASSLSGRGPLGSNQPQH	no hit with PDB/healthy	HLA-A*02
P9	subset #4	CD8+				
P8	subset #4	PBMCs	TRBV19	ASTRQGTGELF	no match with PDB/healthy	HLA-A*02 or HLA-B*51
P3	subset #4	PBMCs				
P15	subset #1	BM	TRBV12-3	ASSSTGGTGELF	no match with PDB/healthy	HLA-A*24
P13	subset #1	LN				
P16	subset #1	PBMCs	TRBV19	ASSQGAGNTIY	no match with PDB/healthy	HLA-A*02
P32	subset #2	PBMCs				
P18	M-CLL, IGHV4-34	CD8+	TRBV29-1	SVGTGGTNEKLF	Large granular lymphocytic leukemia	ND
P3	subset #4	PBMCs				
P16	subset #1	PBMCs	TRBV5-1	ASRGGEKLF	Match with healthy	ND
P22	subset #2	BM				
P25	UM-CLL, IGHV4-34	PBMCs				
P30	UM-CLL, IGHV3-21	PBMCs				

Table 1. Shared major clonotypes amongst CLL patients.

Patient ID	Immunogenetic subgroup	TRBV gene	CDR3 AMINO ACID SEQUENCE	Entity description	PDB Accession No
P23	M-CLL, IGHV4-34	TRBV10-3	AIGTGDSNQPQH	EBV	AM041168, AM041175
P31	M-CLL, IGHV3-21	TRBV10-3	ASGTGDSNQPQH	EBV	AM041172
P11	subset #4	TRBV19	ASSIGTGELF	LGL	
P32	subset #2	TRBV12-3	ASSPNYSNQPQH	MBL	
P17	subset #1	TRBV6-5	ASSRAGQPQH	CLL, subset #1	

Table 2. CLL major clonotype matches with public database sequences.

Figure 1A.

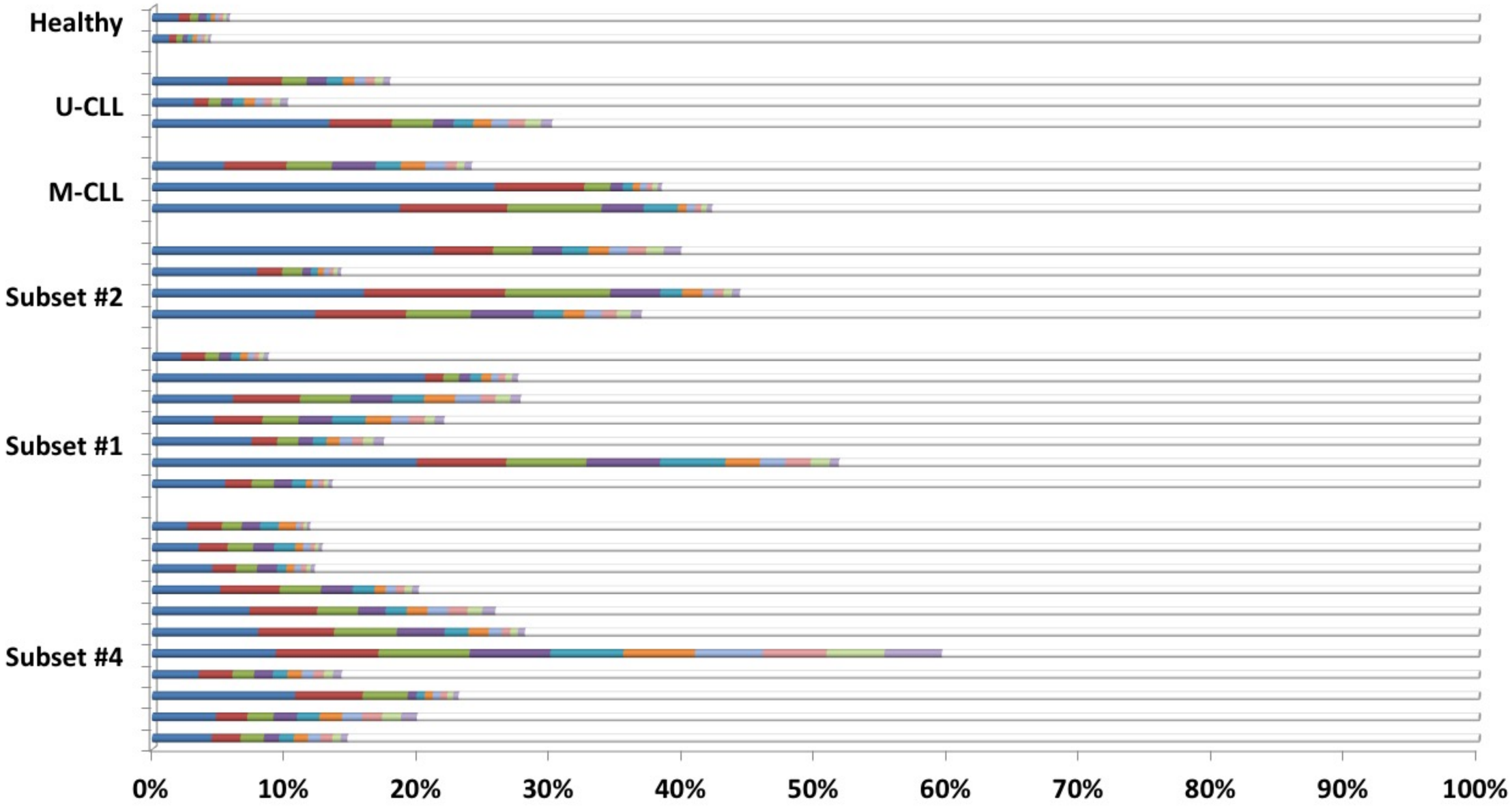


Figure 1B.

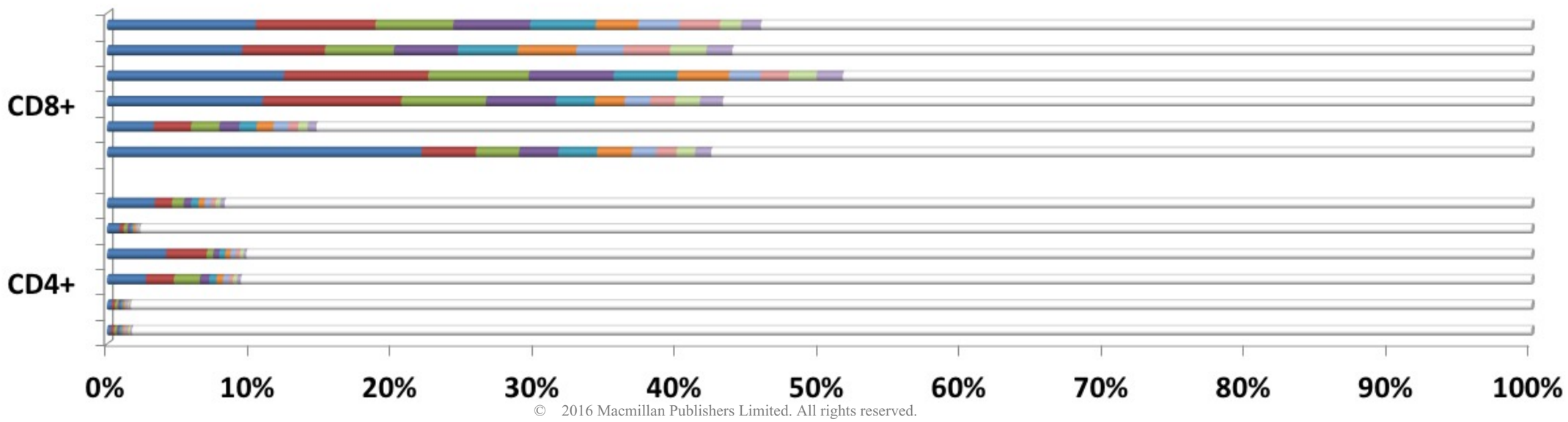


Figure 2A.

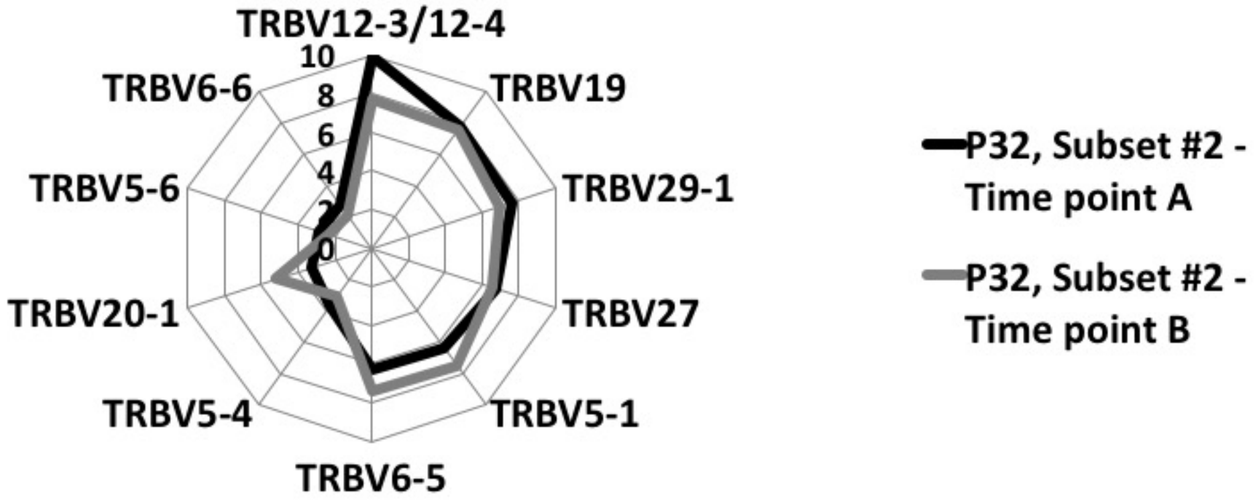
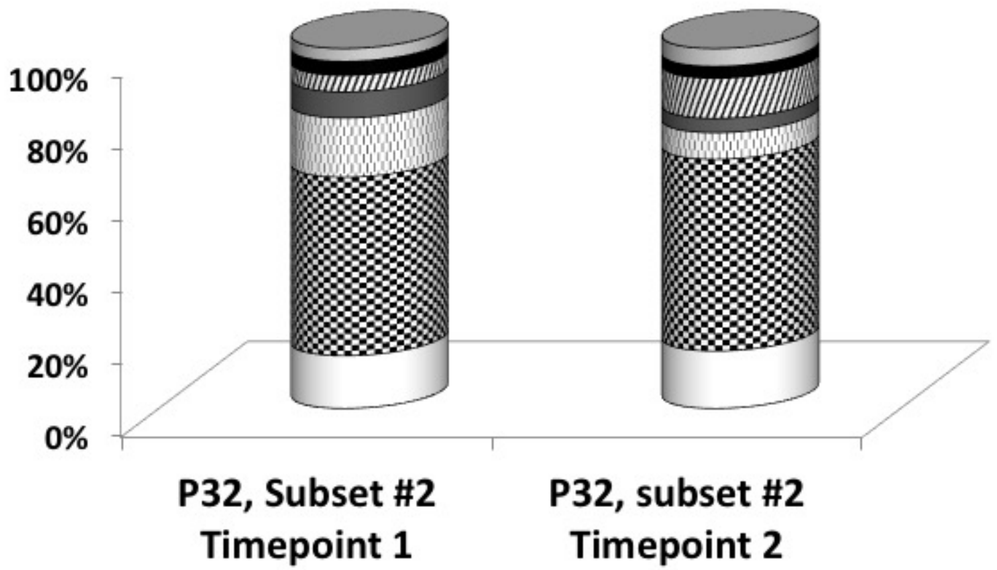
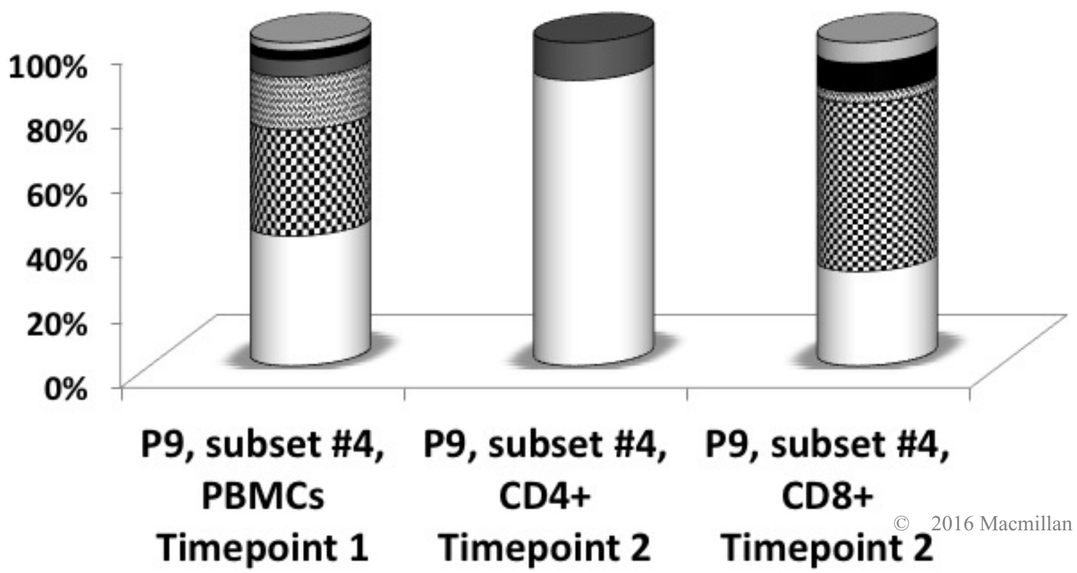


Figure 2B.





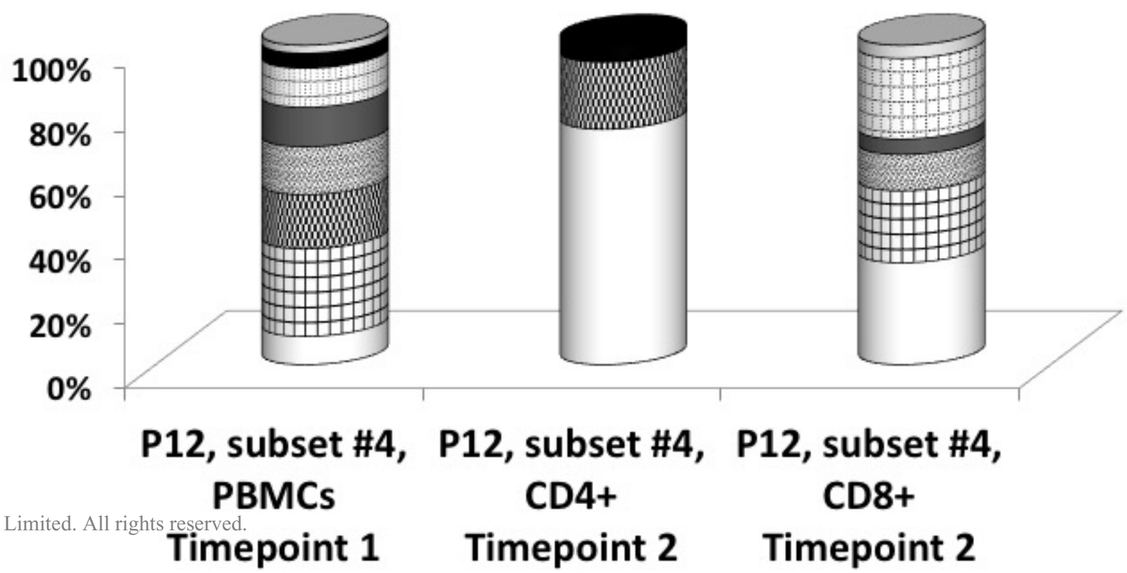


Figure 1A.

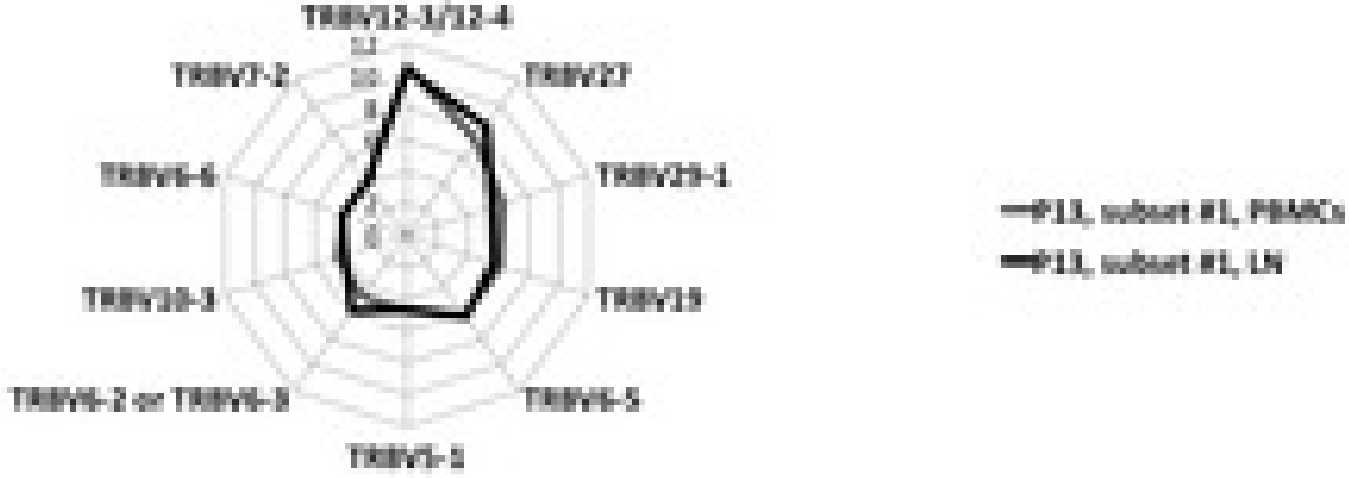


Figure 1B.