CLL intraclonal fractions exhibit established and recently acquired patterns of DNA methylation

B. A. Bartholdy
X. Wang
X. Yan
M. Pascual
M. Fan

See next page for additional authors

Follow this and additional works at: https://academicworks.medicine.hofstra.edu/articles

Part of the Internal Medicine Commons

Recommended Citation

This Article is brought to you for free and open access by Donald and Barbara Zucker School of Medicine Academic Works. It has been accepted for inclusion in Journal Articles by an authorized administrator of Donald and Barbara Zucker School of Medicine Academic Works. For more information, please contact academicworks@hofstra.edu.
CLL intraclonal fractions exhibit established and recently acquired patterns of DNA methylation

Boris A. Bartholdy,1 Xiahoua Wang,1,* Xiao-Jie Yan,2,* Marién Pascual,3-5 Manxia Fan,1 Jacqueline Barrientos,2,6 Steven L. Allen,2,6 Jose Angel Martinez-Climent,3-5 Kanti R. Rai,2,6 Nicholas Chiorazzi,2,6,7 Matthew D. Scharff,1 and Sergio Roa3-5

1Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY; 2The Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY; 3Hemato-Oncology Program, Cima Universidad de Navarra, Pamplona, Spain; 4Navarra Institute for Health Research (IdiSNA), Pamplona, Spain; 5Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Instituto de Salud Carlos III, Madrid, Spain; 6Department of Medicine, Zucker School of Medicine at Hofstra/Northwell, Manhasset, NY; and 7Department of Molecular Medicine, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY

Key Points

• Circulating CLL fractions share a dominant intraclonal DNA methylation hallmark that reflects cell-of-origin and IGHV mutation of leukemia.
• Intraclonal epigenetic and transcriptional changes support a progressive life cycle of CLL that is potentially connected to clinical outcome.

Intraclonal subpopulations of circulating chronic lymphocytic leukemia (CLL) cells with different proliferative histories and reciprocal surface expression of CXCR4 and CD5 have been observed in the peripheral blood of CLL patients and named proliferative (PF), intermediate (IF), and resting (RF) cellular fractions. Here, we found that these intraclonal circulating fractions share persistent DNA methylation signatures largely associated with the mutation status of the immunoglobulin heavy chain locus (IGHV) and their origins from distinct stages of differentiation of antigen-experienced B cells. Increased leukemic birthrate, however, showed a very limited impact on DNA methylation of circulating CLL fractions independent of IGHV mutation status. Additionally, DNA methylation heterogeneity increased as leukemic cells advanced from PF to RF in the peripheral blood. This frequently co-occurred with heterochromatin hypomethylation and hypermethylation of Polycomb-repressed regions in the PF, suggesting accumulation of longevity-associated epigenetic features in recently born cells. On the other hand, transcriptional differences between paired intraclonal fractions confirmed their proliferative experience and further supported a linear advancement from PF to RF in the peripheral blood. Several of these differentially expressed genes showed unique associations with clinical outcome not evident in the bulk clone, supporting the pathological and therapeutic relevance of studying intraclonal CLL fractions. We conclude that independent methylation and transcriptional landscapes reflect both preexisting cell-of-origin fingerprints and more recently acquired hallmarks associated with the life cycle of circulating CLL cells.

Introduction

Chronic lymphocytic leukemia (CLL), the most common adult leukemia in the western world, is characterized by progressive accumulation of immunophenotypically distinct CD5⁺ lymphocytes.1,2 Clinical staging systems by Rai3 and Binet4 are partially based on the accumulation of CLL cells in lymphoid tissues, highlighting the relevance of leukemic cell homing mechanisms. Indeed, the study of surface membrane molecules, including chemokine receptor 4 (CXCR4), and in vivo measurements of proliferation, based on deuterium (²H)-labeling of dividing CLL cells, showed that the peripheral blood contains intraclonal cellular fractions with different trafficking potentials and proliferative histories.5,6 Specifically, differential surface densities of CXCR4 and CD5 suggested a heterogeneous continuum of...
CLL cells from those that had recently divided and migrated out the lymphoid tissues into the circulation (CXCR4\textsuperscript{Dim}CD5\textsuperscript{Bright}; proliferative fraction [PF]) to older cells with increased expression of CXCR4 (CXCR4\textsuperscript{Bright}CD5\textsuperscript{Dim}; resting fraction [RF]) that may be attempting to home back to solid tissues. This is consistent with the finding that the impairment of signaling and expression of CXCR4 by the Bruton tyrosine kinase inhibitor ibrutinib\textsuperscript{7,8} promotes the mobilization of CLL cells into the peripheral circulation and blocks their homing to solid tissues.\textsuperscript{5,10}

Molecularly, CLL patients can be subdivided into 2 subsets with distinct clinical and biological characteristics based on the presence or absence of somatic mutations in the variable region of the immunoglobulin heavy chain (IGHV).\textsuperscript{1,11} Compared with patients with IGHV mutations (M-CLL), IGHV-unmutated patients (U-CLL) have a more aggressive disease course that is more rapidly fatal without treatment.\textsuperscript{12,13} Additional molecular profiling of CLL has revealed not only complex genetic interpatient differences\textsuperscript{14-17} but also epigenetic differences in chromatin landscape\textsuperscript{18,19} and DNA methylation,\textsuperscript{20} reflecting complex molecular and microenvironmental interactions during the development and evolution of CLL.\textsuperscript{2,21-23} Specifically, global DNA methylation patterns can predict distinct CLL epigenetic subtypes with prognostic implications, possibly reflecting their cellular origin, IGHV mutational experience, and their relationship to later stages of normal B-cell development.\textsuperscript{20,24-26} In addition, more aggressive disease progression is often associated with high leukemic birth rates (BRs)\textsuperscript{27,28} and intracranial genetic heterogeneity,\textsuperscript{20,30} indicating that clonal evolution is a key factor in the disease. Likewise, further evolution of DNA methylation may occur in high-risk clinically progressive cases, coevolving with genetic aberrations.\textsuperscript{31} Such somatic epigenetic heterogeneity in CLL has been shown to accumulate intracranially, potentially facilitating the accrual of additional subclonal mutations and promoting shorter remission times after treatment.\textsuperscript{32} It remains unknown whether molecular heterogeneity might also correlate with BR of CLL cells in vivo or contribute to the appearance of the observed CXCR4/CD5 intraclonal subpopulations.

To address these questions, we determined DNA methylation and gene expression changes occurring within each of the more homogeneous intraclonal CLL fractions, which were highly enriched in recently born or in older quiescent leukemic cells. Patient samples were obtained independently from a subgroup of patients who participated in the CRC011 Heavy Water CLL Research Consortium trial,\textsuperscript{28} which examined the usefulness of leukemia cell BR in the prognosis of CLL. This allowed us to associate the epigenetic profiling of CLL fractions with patient BR, clinical outcome, and typical biomarkers of prognosis. Studying sorted intraclonal subpopulations provided us the opportunity to examine the extent to which the fractions reflect cyclic events occurring in circulating leukemic cells that could reveal differences possibly obscured by the heterogeneity of the bulk CLL clone.

**Methods**

**Patients**

Twenty-one previously untreated early-stage (Rai stage 0, 1, or 2) patients with CLL followed at the Northwell Health Cancer Institute and The James Cancer Center, Ohio State University who participated in the CRC011 Heavy Water trial (NCT00481858)\textsuperscript{28} were studied (Table 1). This cohort of 8 U-CLL and 13 M-CLL cases was selected based on additional sample material availability and in general was a good representation of the overall trial cohort (Figure 1). Expression of ZAP70 and CD38, IGHV mutational status, and leukemic BRs of fluorescence-activated cell sorter (FACS)-sorted CD19\textsuperscript{+} CD5\textsuperscript{+} peripheral cells were extracted from the published findings.\textsuperscript{28}

**Isolation of cell fractions**

Peripheral blood mononuclear cells were separated by Ficoll-Paque density gradient centrifugation from heparinized venous blood obtained 3 months before, during, or 3 months after the 6-week \textsuperscript{2} H\textsubscript{2}O labeling period and cryopreserved. Before sorting, CLL peripheral blood mononuclear cells were thawed and incubated with the murine anti-human monoclonal antibodies CD5-PE-Cy7, CXCR4-PE, and CD19-Pacific Blue (BD Biosciences). After gating on CD19\textsuperscript{+} CD5\textsuperscript{+} cells, equal numbers of viable cells falling into the PF, intermediate (CXCR4\textsuperscript{Int}CD5\textsuperscript{Int}; intermediate fraction [IF]), and RF based on reciprocal densities of CXCR4 and CD5\textsuperscript{+} (~5% of the clone for each fraction) were sorted with a BD FACSaria (Figure 1A). After sorting, cells were washed, pelleted, and stored at −80\degree for DNA or RNA extraction was performed.

**DNA methylation profiling**

We used the EZ DNA Methylation kit (Zymo Research) for bisulfite conversion of 500 ng of genomic DNA obtained from at least 0.5 × 10\textsuperscript{6} cells of each fractionated CLL sample. Subsequently, methylation high-density profiling was carried out by using the Illumina Infinium HD Human Methylation 450K BeadChip according to manufacturer’s instructions. See supplemental Methods for additional information regarding DNA methylation analysis and annotation.

**Gene expression profiling**

We extracted total RNA from sorted cells using the RNeasy mini kit (Qiagen) following manufacturer’s recommendations and assessed RNA quality with a Bioanalyzer (Agilent) using Agilent RNA 6000 Kit. We hybridized RNA samples to the Illumina HumanHT-12 v4 Expression BeadChip following Illumina standard protocols. See supplemental Methods for additional information regarding gene expression analysis and annotation.

**Time to first treatment (TTFT) analysis**

TTFT was calculated according to CRC011 trial\textsuperscript{28} from the date of the beginning of heavy water treatment to clinical progression, censored at the end of the follow-up period of 7 years or at the time of exclusion of patients from the study for non–treatment-related reasons. See supplemental Methods for additional information regarding Cox proportional hazard and log-rank tests evaluated with TTFT.

**Results**

High leukemic BRs associate with poor outcome

We analyzed cryopreserved cells from 21 of 97 recently diagnosed, untreated CLL patients (Table 1) who participated in the CRC011 Heavy Water clinical trial.\textsuperscript{28} To gain further insight into the origin,
CD38 to the larger trial. As expected, supporting that our patient cohort generally extrapolates (Figure 1D) were significantly associated with outcome as fractions. The relationship of these fractions to the previously mutation status, and percentage of CD38. Furthermore, log-rank tests showed that BR (Figure 1C), several previously established predictors of disease progression, and life cycle of CLL, we further subdivided the characteristics that were representative of this trial. Specifically, slow BRs.27,28 This patient subset had clinical and genetic cycle of circulating CLL cells, we compared the 3 fractions to... progression, and life cycle of CLL, we further subdivided the leukemic CD19^+CD5^+ population (total fraction) into CXCR4^Dm^Bright (PF), CXCR4^Dm^Int (IF), and CXCR4^Bright^CD5^Dim (RF) fractions. The relationship of these fractions to the previously proposed life cycle of CLL cells25,26 is shown in Figure 1A. We dichotomized these patients (8 U-CLL and 13 M-CLL) using 0.35% cells/day (Figure 1B) as the established threshold between fast and slow BRs.27,28 This patient subset had clinical and genetic characteristics that were representative of this trial. Specifically, these patients showed significant association of high BR with several previously established predictors of disease progression requiring early treatment (increased numbers of ZAP70^+ and CD38^+ cells and unmutated IGHV status) (Figure 1B; Table 1). Furthermore, log-rank tests showed that BR (Figure 1C), IGHV mutation status, and percentage of CD38^+ or ZAP70^+ B cells (Figure 1D) were significantly associated with outcome as expected, supporting that our patient cohort generally extrapolates to the larger trial.

**A dominant IGHV mutation–associated DNA methylation signature is stable in all intraclonal CLL fractions**

To explore DNA methylation dynamics throughout the intraclonal life cycle of circulating CLL cells, we compared the 3 fractions to unfractonated (total/bulk) CLL cells (analyzed in most studies) using high-density DNA methylation microarrays.33 Unsupervised hierarchical clustering (HC) of the methylation revealed that fractions uniformly clustered by patient (Figure 2A), suggesting strong intraclonal epigenetic similarities. In addition, both HC and principal component analysis (PCA) (Figure 2B) revealed that the clearest differences existed between M-CLL and U-CLL patients. Such patient-specific clustering persisted even after focusing on the top (~5%) most variable cytosine guanine dinucleotides (CpGs) (supplemental Figure 1A).

Next, we explored DNA methylation differences between U-CLL and M-CLL samples in the individual sorted fractions and identified a large number of differentially methylated CpGs (DMCs) between the 2 prognostic groups in each of the fractions (Figure 2C). Venn diagram comparisons demonstrated that 9966 of these DMCs were common to all 4 fractions (c-DMCs) (Figure 2D; supplemental Table 1). Notably, 2763 DMCs initially appeared to be fraction specific (f-DMCs) at the given statistical cutoff (941 in PF, 291 in IF, and 1531 in RF) and were not detected in unseparated tumor cells because of the underlying heterogeneity (Figure 2D; supplemental Table 2). However, these f-DMCs exhibited identical directionality in methylation differences in all intraclonal fractions (Figure 2E). In fact, IGHV mutation–associated c-DMCs and f-DMCs similarly included both hypo- and hypermethylated sites that were significantly associated c-DMCs and f-DMCs similarly included both hypo- and hypermethylated sites that were significantly associated.
underrepresented in CpG islands and enriched in “open sea” regions (supplemental Figure 1B). Using functional chromatin segmentation,34 we observed that M-CLL displayed hypomethylation of enhancers and weak transcriptional regulation sites, while hypermethylation occurred in some enhancers but was most evident in heterochromatic regions (supplemental Figure 1C). These results confirm previous established associations of DNA methylation with the IGHV mutational status in CLL20,25 and expand on those findings by demonstrating that such methylation patterns are stably maintained across the circulating fractions. However, as will be confirmed below, these cells do appear to undergo subtle epigenetic changes as they move further away from the time of their last cell division.
Figure 2. Intraclonally stable methylation patterns in CLL reflect signatures of the potential normal B-cell of origin. (A) Unsupervised HC of the fractions by Manhattan distance measurement of DNA methylation (unscaled β-values). Colored horizontal labels indicate IGHV mutation status (U- or M-CLL) and BR (low or high BR, at the threshold of 0.35% cells/day). (B) PCA of DNA methylation data from all fractions, colored by either IGHV mutational status or BR. (C) Scatter plots for pairwise comparisons of DNA methylation between U-CLL and M-CLL (top) or between low BR and high BR (bottom) patients by CXCR4/CD5 intraclonal fractions. Sites were highlighted at false discovery rate (FDR) <0.05 (black dots), and further considered significant at FDR <0.01 and absolute mean methylation difference of ≥10% (red dots). Colors in the cartoon representation of cell fractions are maintained throughout the panels. (D) Venn diagram showing relations between common overlap (c-DMC) and fraction-specific differentially methylated CpGs (f-DMC), which were identified as red dots in panel C after pairwise comparison of U- and M-CLL patients. (E) Heatmap of c-DMCs and f-DMCs between U- and M-CLL, ordered by IGHV status and by fractions. Colored horizontal labels indicate other established predictors of CLL progression: CD38+ (>30% positive), ZAP70+ (>20% positive). Framed quadrants in the heatmap highlights f-DMCs initially identified in the corresponding fraction, despite largely comparable methylation profiles. IF, intermediate fraction; PC, principal component; PF, proliferative fraction; RF, resting fraction.
Established intraclonal methylation patterns reflect those of late B-cell maturation stages

Next, we sought to compare the epigenetic state of our CLL samples with published methylomes from normal human primary B-cell subtypes.20 Supplemental Figure 1D shows a phylogenetic tree based on Manhattan distances between the individual patients based on the total fractions. Consistent with the highly conserved methylation state across the CXCR4/CD5 fractions and the bulk CLL samples, similar results were obtained using the fractions (not shown). A common maturation axis suggested a continuum of DNA methylation, where M-CLL samples exhibited methylation states closer to those of the more differentiated memory B cells or bone marrow plasma cells, while U-CLL samples clustered between naive and germinal center (GC) or early-post-GC stages (tonsil plasma cells), which was suggestive of an early-GC epigenetic hallmark. This was consistent with previous postulations that the IGHV-related DNA methylation pattern likely derives from the founder cell and that U-CLL cells are representative of a less mature B-cell differentiation stage than M-CLL cells.20,25,31,35

Further supporting this notion, a set of 5 previously described CpG biomarkers that can be used to classify CLL methylomes into 3 epigenetic subgroups related to normal counterpart B-cell differentiation stages24 was found among the c-DMCs identified here (supplemental Table 1) and suggested that U-CLL cells bear a methylation phenotype closer to naive B cells than M-CLL cells (Table 1). In addition, this classifier successfully predicted outcome in our population (supplemental Figure 1E). Interestingly, even though patient CLL1039 was classified as M-CLL, its epigenetic profile was more similar to the U-CLL subgroup (supplemental Figure 1D), and this was more consistent with its poor clinical outcome, reflected in a short TTFT, high BR, and chromosomal abnormalities (Table 1). Indeed, the 5-CpG algorithm24 classified CLL1039 as an intermediate case, confirming that the epigenetic cell-of-origin categorization can complement the prognostic value of IGHV mutational status, BR, and other well-known factors, especially for the M-CLL subgroup.

BRs show limited impact on DNA methylation of circulating CLL fractions

Since plasticity and changes of the DNA methylome (mainly hypomethylation) accompany B-cell proliferation,26,27,35 we next investigated whether different leukemia BRs associated with changes in DNA methylation. As discussed above, unsupervised HC and PCA analysis failed to reveal a distinctive global association of DNA methylation and BR in any of the cellular fractions (Figure 2A-B), mainly because of the predominant differences between U- and M-CLL samples. Indeed, a marked overlap of IGHV mutation status and BR can be observed (Figure 1B and 2A), complicating the identification of an independent impact of proliferation rates on the CLL methylome. However, results from the larger clinical trial to which this study is linked demonstrated that increased BR of CLL is an intermediate case, confirming that the epigenetic differences when comparing PF and RF have a limited impact on CLL DNA methylation levels, which is ultimately obscured in the bulk clone.

Increased DNA methylation heterogeneity further accumulates in resting CXCR4brightCD5dimCD19+ cells

Active transition of B cells through the GC may also be revealed by the accumulation of DNA methylation heterogeneity.36-38 To explore whether U-CLL and M-CLL fractions accumulated different levels of DNA methylation heterogeneity as a consequence of a distinct cell-of-origin history, we evaluated interpatient heterogeneity through calculation of pairwise Manhattan distances of their mean methylation levels (Figure 3A) and their correlation coefficients (Figure 3B). We observed that all fractions from M-CLL patients exhibited a significantly increased interpatient methylation heterogeneity compared with U-CLL and, correspondingly, poorer pairwise correlations. Such variegation in DNA methylation is consistent with the proposed GC/early-post-GC origin of the M-CLL subtype (supplemental Figure 1D).

To further identify differentially variable CpG sites (DVCs) in both U- and M-CLL groups, where the fractions in one group have consistent methylation values and the fractions in the other group have highly variable methylation values, we applied the DiffVar method (see supplemental Methods) to our data. We observed a significant increase in the number of variable loci in all M-CLL fractions (Figure 3C), again supporting the intrinsically higher heterogeneity of M-CLL patients compared with U-CLL. Indeed, despite the expected effect of stochastic loss of DNA methylation and increased variance at partially methylated sites (Figure 3D), we still could observe a higher variance in M-CLL patients across the full range of β-values. In both patient subtypes, however, this variability seemed especially evident within the longer-lived CLL cells of the RF, but less pronounced in the recently born CLL cells of the PF (Figure 3C). Consistent with this, both the distributions of variances and the Figner-Killeen test of homogeneity of variances between PF and RF for each CpG (pairwise) showed a significantly higher variance in the RF in both U- and M-CLL groups (Figure 3E), suggesting that DNA methylation heterogeneity further accumulates as cells age in the circulation without dividing.

Longevity-linked DNA methylation changes accumulate in recently born CXCR4brightCD5dimCD19+ cells

The evidence for "time since last division" as a potential source of intraclonal heterogeneity in both U-CLL and M-CLL patients prompted us to investigate the nature of these methylation changes and whether they might occur recurrently at specific loci. To explore this, we focused on intraclonal differences in CpG methylation, rather than interpatient variability, by using paired analysis that compared fractions within each patient (Figure 4). M-CLL and U-CLL subgroups were analyzed separately to investigate whether their apparently different cell-of-origin and IGHV status might also be associated with distinct epigenetic instability as cells age. U-CLL patients evidenced 533 loci that recurrently accumulated significant epigenetic differences when comparing PF and RF intraclonal fractions (Figure 4A). However, these 533 DMCs identified in
U-CLL cells (Figure 4B) showed the same tendency of methylation change in the M-CLL cells (Figure 4C; supplemental Table 4), which may have been obscured by the intrinsic heterogeneity of M-CLL (Figure 3) and is consistent with a common linear epigenetic advancement in the circulating fractions of both U-CLL and M-CLL patients.

In the PF, most of these intraclonal DMCs accumulated hypomethylation in open sea and heterochromatic regions, while hypermethylation was mainly enriched in CpG islands and Polycomb-repressed regions (Figure 4D-E). Interestingly, this signature has been previously linked to normal B-cell differentiation and ongoing GC-like experience.26,35 Furthermore, when we used a previously described epigenetic clock based on DNA methylation biomarkers of aging (DNAm Age)40 as a surrogate marker for contrasting the ages of PF and RF in each patient, we observed that in 80% of cases (17 out of 21), the estimated DNAm Age of the PF was higher or at least comparable to that of the RF (Figure 4F). Altogether, these results reveal that intraclonal CLL fractions in U-CLL and, to a lesser extent, M-CLL patients exhibit subtle epigenetic changes as they advance in a progressive life cycle of CLL, with recently born PF cells being more ancient and carrying acquired longevity-linked DNA methylation changes.

Intraclonal differential expression confirms the proliferative histories of the fractions and reveals unique associations with clinical outcome

Despite the poor overall correlation between gene expression and IGHV mutation status or BR (supplemental Figure 2), we expected to confirm our previous observations that the fractions reveal expression differences related to their proliferative histories.6 Indeed, we found 448 differentially expressed genes (DEGs)
between paired PF and RF fractions (supplemental Table 5), which intersected with and followed a similar directionality with those originally reported (Figure 5A). As expected, messenger RNA levels for CXCR4 and CD5 were consistent with the protein expression of these receptors used for performing the FACS-based fractionation (Figure 5B).

To investigate whether specific expression patterns in the fractions might be more predictive of clinical outcome than looking at the bulk cells, we used Cox regression and log-rank analyses to identify intraclonal DEGs that preserved their prognostic potential in both analyses (Figure 5C; supplemental Table 6). Some exhibited more than twofold difference in expression between PF and RF (Figure 5D). As examples, high levels of cyclin D2 in the PF or of tight junction associated protein 1 (TJAP1) in the RF (Figure 5E), were able to dichotomize clinical outcome of the 21 CLL patients only when the corresponding fraction was interrogated (Figure 5F).

Consistent with this, it has been shown that cyclin D2 is critical for CD5+ B cells and is overexpressed in proliferation centers of CLL, while TJAP1 has been observed in B-CLL cell lines.

---

**Figure 4.** CLL fractions revealed an intraclonal program of accumulated epigenetic changes linked to the longevity of the cells. (A) Scatter plots for pairwise comparisons of DNA methylation between fractions. Differentially methylated CpGs are highlighted as black (FDR-adjusted P < .05) or red (FDR-adjusted P < .05 and \( \beta \)-value difference ≥10%) dots, based on paired linear models generated with the R/Bioconductor package limma. (B) Venn diagram showing \( IGHV \)-specific differential methylation between PF and RF (red dots in panel A). (C) Scatter plot illustrating similar tendencies in U-CLL (x-axis) and M-CLL (y-axis) patients for those differences in DNA methylation between paired PF and RF fractions (supplemental Table 5), which intersected with and followed a similar directionality with those originally reported (Figure 5A). As expected, messenger RNA levels for CXCR4 and CD5 were consistent with the protein expression of these receptors used for performing the FACS-based fractionation (Figure 5B).
Figure 5. Differentially expressed genes between CLL fractions show unique features associated with cell survival and patient outcome. (A) Venn diagrams showing overlaps between our set of DEGs in RF vs PF cells and those previously described. Those overlaps were significantly enriched as measured by Fisher’s exact tests.
Together, these results suggest that transcriptional regulation of some genes in the fractions, especially in the PF, might have a significant impact in the overall clinical progression of the disease that would not be appreciated looking at the bulk CLL clone.

**Discussion**

The majority of leukemic cells in CLL reside in peripheral lymphoid organs, where almost all tumor cell division occurs. However, a constant egress from solid tissues generates easily accessible circulating CD19 CD5+ tumor cells routinely collected to study patient prognosis and disease etiology and pathogenesis. We hypothesized that the analysis of intraclonal cellular subpopulations and the transitions between them would reveal new prognostic markers, patterns of pathogenesis, or insights into the cell of origin of the tumor, which might have been obscured previously by the heterogeneity of the total circulating population. Here, we used the reciprocal expression of CXCR4 and CD5 surface markers on circulating tumor cells from newly diagnosed untreated CLL patients to obtain fractions enriched for cells that had recently divided, presumably in the tissues, and older cells that were quiescent, having divided earlier.

Unsupervised HC of the fractions and of the bulk tumor cells based on DNA methylation separated samples consistent with their IGHV mutation status. Many of the observed methylation differences between U-CLL and M-CLL were shared by all intraclonal fractions and had identical directionality, which in combination with other evidence supports the existence of a remarkably stable overall pattern of DNA methylation in CLL. This most often consisted of hypomethylation at “open sea” CpG sites in enhancers and weakly transcribed regions in M-CLL patients compared with U-CLL. Interestingly, this hallmark had been previously observed in CLL and had been compared with the loss of methylation that characterizes normal B-cell development, suggesting a more mature cell of origin for M-CLL than U-CLL patients. Furthermore, evaluation of interpatient methylation heterogeneity revealed a slightly increased variability and poorer pairwise correlations in fractions obtained from M-CLL patients compared with U-CLL. Such increased variability resembles the accumulation of DNA methylation heterogeneity observed during normal transition through the GC and further supports the pre- or early-GC cell of origin suggested for U-CLL cells. Indeed, our phylogenetic analysis comparing CLL fractions to normal B cells also positioned M-CLL methylomes closer to memory or bone marrow plasma cells, while U-CLL methylomes suggested a slightly less differentiated state. The fact that these differences in methylation profiles were found in all fractions of U-CLL and M-CLL patients and included the 5 previously described CpG biomarkers with the potential to assign a normal B-cell counterpart to the leukemic cell further supports the idea that the dominant methylation hallmark reflects the cell of origin. It is important to note that these interpatient epigenetic differences were not clearly associated with the BR determined by heavy water labeling, even in the PF that had recently undergone cell division. This could be either because DNA methylation is independent of BR or because any potential proliferation-related epigenetic signature disappears very soon after leukemic cells leave the proliferation centers and move into the circulation. Analogously, normal naive B cells from tonsils or peripheral blood exhibit comparable DNA methylation profiles, even though they have significantly different transcriptomes. On the basis of these observations, we conclude that the rate of proliferation and physical egression out of the lymphoid niche into the bloodstream have a limited impact on the methylome of the CLL cell.

Paired analysis of differentially methylated CpGs between fractions from each of the U-CLL and M-CLL patients revealed that in both subtypes, but more clearly in the U-CLL group, the cells of the PF differed from those of the RF based on hypomethylation in open-sea and heterochromatic regions and hypermethylation mainly in CpG islands and Polycomb-repressed regions. Interestingly, this signature has been previously linked not only to solid and hematological tumors but also to aging, extended cellular lifespan, and normal B-cell differentiation. As cells in the RF are thought to derive from the PF after time in circulation without dividing, this represents a paradox that PF methylome is “older” than RF. Especially since we observed that long-lived RF cells carried more CpG sites with heterogeneous methylation than recently born (or re-born) cells of the PF, and such acquired heterogeneity might become a potential source for additional epigenetic aging and genetic evolution as time passes in circulation since the last division. One possibility to reconcile this is in a CLL life cycle where PF cells may be linear descendants, therefore inheriting the epigenetic age, of some resting cells that reentered proliferation centers and further acquired aging-associated epigenetic changes upon cell division (see model in Figure 1A and visual abstract). In every life cycle of the leukemic cells, it might be expected that any RF-related epigenetic signature would either be reset as leukemic cells pass through the proliferation centers, or stay and accumulate in every round of recently born PF. The former seems less likely, since the dominant IGHV mutation–related epigenetic signature remains quite stable, supporting the pre-existence and retention of these epigenetic states in all fractions as constitutive cell-of-origin hallmarks that are not acquired or reshaped during the life cycle of the leukemic cell. Cumulative hallmarks, on the other hand, become hard to identify without comparison of sequential samples from the same patients during progression of the disease. However, our evidence suggests a progressive epigenetic aging of the newly proliferated PF, consistent with cells of the PF being linear descendants of some
cells from the RF that reentered lymphoid solid tissue and received prosurvival stimuli to proliferate again. However, based on the available data, we cannot rule out the possibility that the PF derives from a distinct set of cells, independent of the RF, that continuously undergo rounds of replication. In this nonexclusive scenario, DNA methylation may be increasing at an accelerated rate in the proliferation centers/lymphoid niche due to the stimulation and proliferation of CLL cells, and once released from this environment, the mAge of CLL cells advances at a slower rate (in most patients). Thus, RF cells at a given sampling time point would have been produced from PF cells that were released from the proliferation centers at an earlier date and thus exposed to the mAge-accelerating environment for less time. PF cells present in the sample would have been released at a later date thus exposed to the DNA methylation environment for more time, contributing to PF being epigenetically “older” than RF in most patients. While future research is still required to solve these questions, our observations are consistent with whole-genome sequencing of indolent CLL that has recently revealed an ongoing ageing-related signature based on spontaneous deamination of CpG sites that results in genetic mutations.48,49

Largely independent of DNA methylation, CLL fractions revealed intracranial gene expression differences that confirmed the expected variation of CXCR4/CD5 surface levels and supported the advancement of the leukemic cell from the proliferative to the resting compartment in a clonally related continuum of CD5 states.50 Importantly, a substantial number of differentially expressed genes in the PF associated with TTFT that could not be detected in bulk cells; supported (1) by the transcriptional differences that have been observed previously6 and here between fractions, which affect survival and cell movement pathways and reveal clinical associations with an earlier TTFT that could not be detected in bulk cells; and (2) by the observation that the PF seems epigenetically more homogeneous, with additional aging-related epigenetic changes acquired either as the cell divides just before entering the circulation or in circulation and that might then be carried back into the next round of reboim PF cells. These findings highlight the importance of further research and exploration of therapeutic strategies for targeting the PF while accelerating the senescence and death of the RF to block the reentrance of CLL resting cells into a new life cycle.

Acknowledgments
This work was supported in part by a Lymphoma Research Foundation CLL/SLL collaborative grant (N.C. and M.D.S.) and National Institutes of Health, National Cancer Institute grants R01-CA081554 (N.C.) and 5R01CA072649 (M.D.S.) and National Institute of Allergy and Infectious Diseases grants 5R01AI132507 and 9R01AI112335 (M.D.S.) S.R. was funded by Ministerio de Economía y Competitividad (SAF2013-45787-R and SAF2017-82309-R) cofunded by European Regional Development Fund, the Ramón y Cajal program (RYC-2014-16399) cofunded by European Social Fund, Gobierno de Navarra (GN-106/2014), and contributions from Fundación La Caixa. The authors acknowledge John Byrd, Jeffrey Jones, and Thomas Lin from The Ohio State University for contributing samples for analysis; these samples were collected during the mentioned study by Murphy et al28 and were cryopreserved and stored in the Tissue Core (Laura Rassenti, head) of the National Institutes of Health, National Cancer Institute–funded PO1 program “CLL Research Consortium” (Thomas J. Kipps, principal investigator). The authors also acknowledge philanthropic contributions from The Nash Family Foundation, The Marks Foundation, the Karches Family, the Mona and Edward Albert Foundation, and the Jean Walton Fund for Leukemia, Lymphoma, and Myeloma Research.

Authorship
Contribution: B.A.B. and S.R. performed computational studies and analyzed and interpreted the data; X.W., X.-J.Y., N.C., M.D.S., and S.R. designed the research; X.W., X.-J.Y., M.P., and M.F. performed experiments; J.B., S.L.A., J.A.M.-C., and K.R.R. discussed the results and revised the manuscript; N.C., M.D.S., X.W., and S.R. conceived the study; and N.C., M.D.S., B.A.B., and S.R. cowrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: B.A.B., 0000-0002-7401-8591; M.P., 0000-0002-1678-1054; N.C., 0000-0003-1023-6650; M.D.S., 0000-0003-0368-2433; S.R., 0000-0003-1095-0715.

Correspondence: Matthew D. Scharff, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461; e-mail: matthew.scharff@einsteinmed.org; Nicholas Chiorazzi, The Feinstein Institute for Medical Research, 350 Community Dr, Manhasset, NY 11030; e-mail: nchizzi@northwell.edu; and Sergio Roa, Cima Universidad de Navarra, 55 Av Pio XII, 31008 Pamplona, Navarra, Spain; e-mail: sroa@unav.es.

References


13. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. 


