Chronic lymphocytic leukemias with trisomy 12 show a distinct DNA methylation profile linked to altered chromatin activation

by Maria Tsagiopoulou, Vicente Chapaprieta, Martí Duran-Ferrer, Theodoros Moysiadis, Fotis Psomopoulos, Panagoula Kollia, Nikos Papakonstantinou, Elias Campo, Kostas Stamatopoulos, and Jose I. Martin-Subero

Haematologica 2020 [Epub ahead of print]

Citation: Maria Tsagiopoulou, Vicente Chapaprieta, Martí Duran-Ferrer, Theodoros Moysiadis, Fotis Psomopoulos, Panagoula Kollia, Nikos Papakonstantinou, Elias Campo, Kostas Stamatopoulos, and Jose I. Martin-Subero. Chronic lymphocytic leukemias with trisomy 12 show a distinct DNA methylation profile linked to altered chromatin activation. Haematologica. 2020; 105:xxx
doi:10.3324/haematol.2019.240721

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in print on a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.
Chronic lymphocytic leukemias with trisomy 12 show a distinct DNA methylation profile linked to altered chromatin activation

Maria Tsagiopoulou, Vicente Chapaprieta, Martí Duran-Ferrer, Theodoros Moysiadis, Fotis Psomopoulos, Panagoula Kollia, Nikos Papakonstantinou, Elias Campo, Kostas Stamatopoulos, Jose I. Martin-Subero

1. Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece.
2. Department of Biology, National and Kapodistrian University of Athens, Athens, Greece.
3. Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain.
5. Departamento de Fundamentos Clínicos, Universitat de Barcelona, Barcelona, Spain.
6. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Spain.
7. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

RUNNING TITLE: Epigenomic profiling of CLLs with trisomy 12

CORRESPONDING AUTHOR: Jose I. Martin-Subero (imartins@clinic.cat)

WORD COUNT: 1604

AUTHOR CONTRIBUTIONS: M.T. performed the study and wrote the manuscript; V.P. performed data analysis and assisted in data interpretation; M.D-F., T.M., F.P., P.K., N.P., and E.C. assisted in research, data analysis, and interpretation; K.S. designed and supervised the research and revised the manuscript, J.I.M.S. designed
and supervised the research, assisted in data interpretation and wrote the manuscript.

**ACKNOWLEDGEMENTS**

Data used in the present study have been generated through funding from the European Union’s Seventh Framework Programme through the Blueprint Consortium (grant agreement 282510), the International Cancer Genome Consortium (Chronic Lymphocytic Leukemia Genome consortium to E.C.), the World Wide Cancer Research Foundation Grant No. 16–1285 (to J.I.M.-S.), the Generalitat de Catalunya Suport Grups de Recerca AGAUR 2017-SGR-736 (to J.I.M.-S.) and 2017-SGR-1142 (to E.C.), and CIBERONC (CB16/12/00225). E.C. is an Academia Researcher of the Institució Catalana de Recerca i Estudis Avançats (ICREA) of the Generalitat de Catalunya. MT is a recipient of a fellowship from the State Scholarships Foundation of Greece, and a short term collaboration award by the European Hematology Association. This work was partially developed at the Centro Esther Koplowitz (CEK, Barcelona, Spain).
Chronic lymphocytic leukemia (CLL) is a neoplasm derived from mature B cells showing a broad spectrum of clinico-biological features. The landscape of genetic alterations of CLL is currently well characterized and found to be extremely heterogeneous, with multiple chromosomal aberrations and dozens of driver genes mutated in relatively small proportions of the cases. In spite of this heterogeneity, four cytogenetic alterations, i.e. del(13q) (>50% of the patients), del(11q) (18%), +12 (16%), and less frequently del(17p) (7%), are collectively detected in at least 80% of patients. These copy number changes are part of the routine risk assessment of CLL, as they are robustly associated with treatment choices and the clinical course of the patients. At one end of the prognostic spectrum, isolated del(13q) is related to favorable prognosis, +12 with intermediate prognosis, del(11q) with poor prognosis and del(17p) with the worst prognosis of all groups. This latter subgroup identifies patients with particular resistance to chemoimmunotherapy who, instead, benefit considerably from biological agents.

Although the targets of del(11q), del(13q) and del(17p) have been narrowed down to ATM/miR-34b-c, the DLEU2–mir-15-16 cluster, and TP53, respectively, how +12 contributes to CLL pathogenesis remains elusive. This group of CLLs is characterized by high rates of cell proliferation as well as clinical and biological heterogeneity, since it is linked to additional genomic aberrations such as trisomy 19, high frequency of NOTCH1 mutations and enrichment in aggressive stereotyped subset and IgG-switched heavy chains. Moreover, +12 was recently identified as an important modulator of response to kinase inhibitors in CLL.

In addition to genetic changes, various layers of the CLL epigenome have also been characterized and these analyses have revealed novel insights into the cellular origin.
and molecular mechanisms underlying disease pathophysiology. However, the relationship between genetic and epigenetic changes in CLL is still unclear. Therefore, the goal of the present study is to identify the epigenetic correlates of the three most frequent cytogenetic subgroups of CLL and to shed light into the molecular pathogenesis of +12 cases.

We used 450K arrays to analyze the DNA methylomes from 255 CLL cases in whom del(13q), del(11q) and +12 were mutually exclusive. In particular, we investigated 29 CLL cases with del(11q) [3 IGHV-mutated CLL (M-CLL) and 26 IGHV-unmutated CLL (U-CLL)], 45 with +12 (17 M-CLL and 28 U-CLL) and 181 with del(13q) (131 M-CLL and 50 U-CLL), as well as 5 biological replicates each of normal naive (NBC) and memory (MBC) B cells sorted from healthy donors (Supplemental Table 1).

We initially applied a series of unsupervised methods to characterize the DNA methylome of the cases. We found that the mean methylation levels of the three CLL subgroups was lower than in normal B cells, and that +12 cases globally had a significantly hypomethylated genome as compared to the other CLL cases (+12 vs. del(11q), p=0.006; +12 vs. del(13q), p=7.2e-08; del(13q) vs. del(11q), p=0.01) (Figure 1A). Next, we analyzed the CpGs with most variable methylation levels (sd≥0.1, n=185936) by Principal Component Analysis (PCA) and studied the information provided by each of the 5 principal components (Supplemental Figure 1). As previously reported, the first two components revealed different fractions of the variability to reflect different concepts: (i) overall CLLs are epigenetically more similar to memory B cells (PC1, 7.9%); (ii) U-CLL and M-CLL are epigenetically different (PC1 and PC2); and, (iii) CLLs as a whole are different from naive and memory B cells (PC2, 3.7%). The third and fourth components (1.9 and 1.3%) were
less clear and showed subtle differences between naive and memory B cells as well as between U-CLL and M-CLL. Remarkably, the fifth component (1%) was unrelated to the IGHV gene somatic hypermutation status and revealed clear differences between +12 cases and CLL cases from the other cytogenetic subgroups (Figure 1B, Supplemental Figure 1), suggesting that +12 cases may indeed show a specific epigenetic configuration.

We then aimed at identifying the +12-specific CpGs by two complementary approaches. First, we correlated the eigenvalues of the PC5 and the methylation beta-values and found that 1760 CpGs were significantly associated with this source of variability and confirmed to be related with +12 by hierarchical clustering (Supplemental Figure 2). To more specifically detect +12-related CpGs, we performed a supervised differential methylation (DM) analysis in cases with and without +12 considering U-CLL and M-CLL separately (absolute mean beta value difference of at least 0.25 and a FDR<0.05). In +12 U-CLLs (n=28), we observed a signature of 646 DMCpGs, which was mostly composed by hypomethylated sites in the +12 cases (80.1%) (Figure 1C-D, Supplemental Table 2). This hypomethylation signature was present in +12 CLLs lacking and showing NOTCH1 mutations, and therefore, was specifically associated with +12. The +12 M-CLLs (n=17) were more heterogeneous and a lower number of DMCpGs were detected (Figure 1C). However, the 646 DMCpGs in +12 U-CLLs showed the same trend in M-CLL cases (Supplemental Figure 3). Furthermore, analyzing B cells spanning the entire B cell maturation program revealed that the hypomethylation signature was mostly acquired de novo in the +12 cases, as for the great majority of the CpGs it was not present in any B cell subpopulation (Figure 1D, Supplemental Figure 3). Based on
these findings, downstream analyses in the +12 cases were focused on the U-CLL subgroup. We also performed a similar analysis for del(13q) and del(11q) cases, which showed few DMCpGs (Figure 1C). The predominantly de novo acquired hypomethylation signature in +12 U-CLLs was related to gene bodies and UTR regions outside CpG islands, and targeted both transcribed regions and enhancer elements (Figure 1E).

The link between +12-specific hypomethylation and enhancer elements prompted us to evaluate the association between DNA methylation and histone 3 lysine 27 acetylation (H3K27ac), a chromatin mark related to active regulatory elements. In a previous study, we generated H3K27ac ChIP-seq data in CLL patients, which included 21 of the 255 cases whose DNA methylome is studied in the present report.

We initially observed that 48.3% of the DMCpGs (312/646) specific for +12 U-CLLs were located within 246 H3K27ac peaks (Supplemental Table 3). Remarkably, studying the H3K27ac signal within the overlapping peaks, we noticed that the regions losing methylation showed a trend towards higher H3K27ac levels, i.e. more activation, and those regions gaining methylation showed lower H3K27ac levels and, therefore, less active chromatin (Figure 2A-B). A differential analysis of the 246 H3K27ac peaks in U-CLLs with and without +12 (adjusted p<0.05) revealed 35 regions with significantly increased H3K27ac levels (Figure 2C). Although a previous report indicated that the overall H3K27ac pattern of +12 cases was similar to normal B cells, the increased statistical power of our targeted analysis of de novo hypomethylated sites focused on +12 U-CLLs revealed the presence of chromatin activation. As transcription factor (TF) binding has been described to be related to DNA methylation and chromatin activity, we postulated that TFs could be
implicated in inducing the +12-specific epigenomic signature. As TFs bind to DNA lacking nucleosomes, we used previously reported chromatin accessibility data generated by ATAC-seq (available data for 20/255 examined cases) \(^\text{12}\) (Figure 2D).

We found that the 35 regions of interest contained 52 sites of accessible chromatin where TF potentially bind (Supplemental Table 4). The DNA sequences within these 52 accessible sites were significantly enriched \((p<0.05)\) in binding sites of the NFIB-C, MYCN, TFCP2 and XBP1 TFs (Figure 2E). Interestingly, the TFCP2 gene is located in chromosomal band 12q13 and show moderate but significant overexpression cases with +12 as compared to CLLs lacking this genetic change \((\text{log2FC}=0.284, \text{FDR}=0.024)\) (Supplemental Figure 4). Although detailed functional studies are needed to establish the role of TFCP2 in CLL, our data suggest that its overexpression may be one of the potential mechanisms through which +12 exerts its pathogenic effect in CLL.

In order to identify the potential target genes of the regulatory elements targeted by hypomethylation in +12 cases, we analyzed the gene expression arrays of U-CLLs with del(11q), del(13q) or +12. As regulatory elements can affect the expression of distant genes within topologically associating domains \((\text{TAD})\) \(^\text{12}\), the +12-specific hypomethylated regions associated with increased acetylation \((n=35)\) were integrated with the TADs from GM12878, a lymphoblastoid B cell line widely used to characterize epigenetic patterns in CLL \(^\text{12}-\text{14}\). Using gene expression arrays from an extended series of 20 U-CLLs with +12 and 54 U-CLLs lacking this genetic change, we identified 25 genes with increased expression in +12 as compared to del(11q) and del(13q) U-CLLs (Figure 2F, Supplemental Table 5). Amongst them, RUNX3 was a remarkable example. This gene is a master regulator of gene expression during
development that has been reported to act as tumor suppressor or oncogene in cancer. We identified that, +12 U-CLLs show a cluster of 7 hypomethylated CpGs plus increased H3K27ac levels in a regulatory region 35 Kb upstream the RUNX3 promoter, a gene that was significantly overexpressed (Figure 2G). No other region across the RUNX3 gene changed DNA methylation levels and the gene promoter showed similar H3K27ac levels in cases showing and lacking +12 (Figure 2G). These results suggest that epigenetic activation of dozens of genes, and in particular the activation of a distant RUNX3 regulatory element leading to gene overexpression, may account for the distinct biological background of this CLL subtype.

Taken together, our findings further support the unique biological features of +12 CLLs from the epigenetic perspective. This group of CLL is associated with a subset of epigenetically-upregulated genes that may account for its distinct biological background. These novel insights into +12 CLLs may provide a biological rationale to identify specific therapies to treat this unique subtype of CLL.
CONFLICT OF INTEREST DISCLOSURES

KS received research support from Janssen Pharmaceuticals and Roche SA. All other authors declare no conflict of interest.
REFERENCES

FIGURE LEGENDS

**Figure 1. DNA methylation analysis of del(13q), del(11q) and +12 CLLs.**

A. Barplot showing the overall DNA methylation levels in the examined cytogenetic subgroups and normal B cells (+12 versus del(11q), p=0.006 | +12 versus del(13q), p=7.166e-08 | del(13q) versus del(11q), p= 0.01)

B. Principal component analysis showing components 1 and 5 in the cytogenetic subgroups and normal B cells.

C. Number of DMCpGs resulting from the comparison of each cytogenetic subgroup versus all the others CLL cases. The white bar represents the U-CLL while the grey bar represents the M-CLL cases.

D. Hierarchical clustering based on 646 DMCpGs of +12 U-CLL cases as compared to other U-CLLs and normal B cells.

E. Enrichment analysis of DMCpGs hypomethylated in +12 U-CLLs on gene location, location related to islands and chromatin states of memory B cells. The density represents the log2FC of the enriched CpGs compared to the background in each condition (hypo- and hyper-methylation). Each color is linked to a particular subgroup: orange for del(11q), blue for del(13q), red for +12, green for memory B cells and yellow for naïve B cells. PC: principal component, FC: fold change.

**Figure 2. Characterization of the chromatin and transcriptional features of regions de novo hypomethylated in +12 cases.**

A. DNA methylation levels of 312 CpGs that overlap with H3K27ac-containing regions.

B. Heatmap of signal intensities of H3K27ac regions associated with CpGs from panel A.

C. Heatmap of the 35 hypomethylated regions showing significantly higher H3K27ac levels in U-CLLs as compared with cases without +12.

D. Scheme of the strategy used to detect accessible sites within regions with DMCpGs and increased chromatin activity.

E.
Transcription factors whose binding sites are significantly enriched in accessible regions within hypomethylated and active regulatory elements. F. Heatmap showing the relative expression levels of the 25 genes epigenetically-upregulated in +12 U-CLLs as compared with cases lacking +12. G. On the left, a genome browser display of the RUNX3 region including H3K27ac and gene expression of two representative +12 and non +12 U-CLLs, as well as DNA methylation data from all U-CLLs studied. The red square points to the distant regulatory region that becomes hypomethylated and active in +12 U-CLLs. On the right, a zoom in panel of the H3K27ac levels of the distant enhancer region as well as box plots of H3K27ac and gene expression in all U-CLLs studied. Each color is linked to a particular subgroup: orange for del(11q), blue for del(13q), red for +12.
Supplemental Methods

Patient samples

The study group included samples of 232 CLL patients from the Hospital Clinic of Barcelona that belong to the International Cancer Genome Consortium. In addition, we included a series of 23 CLL patients from the Greek cohort in Thessaloniki, Greece. All cases were diagnosed with CLL according to the guidelines of the International Workshop Chronic Lymphocytic Leukemia/National Cancer Institute (iwCLL/NCI). All patients were untreated and did not carry a complex karyotype. The main clinico-biological features of the two cohorts are summarized in Supplementary Table 1. The study was approved by the local Ethics Review Committee of the participating institutions.

Illumina 450k DNA methylation arrays and data analysis

Methylation estimates of 255 CLL cases, analyzed by the 450k Human Methylation Array (Illumina), were mined. We excluded the CpGs that overlapped with the chromosomal abnormalities of the study group. The above filtering resulted in a total number of 426627 CpG sites. CpG sites were considered as differentially methylated by applying criteria such as: i) a minimum absolute difference between mean beta-values (beta difference) of the two subgroups and ii) the statistical significance was evaluated with Benjamini& Hochberg. The 232 microarray data are available at ICGC and the 23 were available at: https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7575.
Genomic location, chromatin states and TFBS enrichment analysis

The DMCpGs were characterized according to their genomic locations and chromatin states for memory B cells from healthy donors, recently published (18). The p-value was calculated using the hypergeometric distribution, when enrichment analysis was performed. We performed TFBS analysis using the MEME suite and AME motif enrichment based on the criterions of average odds score and Ranksum test.

Linking DNA methylation signature with active regions, chromatin accessibility and gene expression

The interpretation of the CpGs were performed by previously published data from expression arrays, ChiP-seq for H3K27ac antibody and ATAC-seq of the same samples and they are part of the BLUEPRINT epigenome project, at the European Genome-Phenome Archive (EGA) which is hosted at the European Bioinformatics Institute. We investigated the overlap of the genomic ranges between the CpGs of interest with the peaks of H3K27ac and ATAC which were detected in a previous study of our group(1). The differential expression and peak analysis were examined using limma and DeSeq2, respectively.
Supplemental Figures

Supplemental Figure 1. The boxplots represent the PC values in each of the 5 principal components in relation to cytogenetic markers aberrations, somatic hypermutation status of the IGHV genes and normal B cells. Each color points to a particular subgroup: orange for del(11q), blue for del(13q), red for tris12, green for memory B cells, yellow for naïve B cells, white for M-CLL, grey for U-CLL
Supplemental Figure 2. Hierarchical clustering based on 1760 CpGs whose PC5 values correlated with the b-values (|r|=0.3)
Supplemental Figure 3. DNA methylation levels of 646 DM CpGs in +12 U-CLLs in the context of M-CLLs and the entire B cell differentiation program.

Supplemental Figure 4. A. Heatmap of expression levels in U-CLL cases of the 5 TFs highlighted from the TF binding site analysis. The expression levels are indicated as row z-scores. B. Dot plot with median showing the expression levels of the TFCP2 gene in each CLL cytogenetic subgroup.
Supplemental Tables

- **Supplemental Table 1:** Demographic, clinical and biological data for the patient cohort and B cell subpopulations from healthy donors.

- **Supplemental Table 2:** 646 DMCPGs revealed after differential methylation analysis between trisomy 12 vs non-trisomy 12 U-CLL.

- **Supplemental Table 3:** 312 CpG sites which showed overlap with H327ac-associated regions and the results of DeSeq2 for the differences on H3K27ac.

- **Supplemental Table 4:** 52 accessible and acetylation-associated regions which showed hypomethylation between tris12 and non-tris12 U-CLL.

- **Supplemental Table 5:** List of the differentially expressed genes (FDR<0.05) associated with the TADs on analysis of the tris12 vs non-tris12 U-CLL.

References