

Highly similar genomic landscapes in monoclonal B-cell lymphocytosis and ultra-stable chronic lymphocytic leukemia with low frequency of driver mutations

Andreas Agathangelidis,^{1*} Viktor Ljungström,^{2*} Lydia Scarfò,¹ Claudia Fazi,¹ Maria Gounari,^{1,3} Tatjana Pandzic,² Lesley-Ann Sutton,^{2,4} Kostas Stamatopoulos,³ Giovanni Tonon,⁵ Richard Rosenquist^{2,4**} and Paolo Ghia^{1**}

¹Strategic Research Program on CLL and B-cell Neoplasia Unit, Division of Experimental Oncology, Università Vita-Salute San Raffaele and IRCCS Istituto Scientifico San Raffaele, Milan, Italy; ²Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden; ³Institute of Applied Biosciences, Center for Research and Technology Hellas, Thessaloniki, Greece; ⁴Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden and ⁵Functional Genomics of Cancer Unit, Division of Experimental Oncology, IRCCS Istituto Scientifico San Raffaele, Milan, Italy

*AA and VL contributed equally as first authors. **RR and PG contributed equally as last authors



Haematologica 2018
Volume 103(5):865-873

ABSTRACT

Despite the recent discovery of recurrent driver mutations in chronic lymphocytic leukemia, the genetic factors involved in disease onset remain largely unknown. To address this issue, we performed whole-genome sequencing in 11 individuals with monoclonal B-cell lymphocytosis, both of the low-count and high-count subtypes, and 5 patients with ultra-stable chronic lymphocytic leukemia (>10 years without progression from initial diagnosis). All three entities were indistinguishable at the genomic level exhibiting low genomic complexity and similar types of somatic mutations. Exonic mutations were not frequently identified in putative chronic lymphocytic leukemia driver genes in all settings, including low-count monoclonal B-cell lymphocytosis. To corroborate these findings, we also performed deep sequencing in 11 known frequently mutated genes in an extended cohort of 28 monoclonal B-cell lymphocytosis/chronic lymphocytic leukemia cases. Interestingly, shared mutations were detected between clonal B cells and paired polymorphonuclear cells, strengthening the notion that at least a fraction of somatic mutations may occur before disease onset, likely at the hematopoietic stem cell level. Finally, we identified previously unreported non-coding variants targeting pathways relevant to B-cell and chronic lymphocytic leukemia development, likely associated with the acquisition of the characteristic neoplastic phenotype typical of both monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia.

Introduction

Chronic lymphocytic leukemia (CLL), the most common adult leukemia in the West, is a clinically heterogeneous disease.¹ At one end of the spectrum, CLL patients present with an indolent disease that does not require therapy for decades. At the other end of the spectrum, patients experience a rapidly progressive disease, need early treatment, and frequently relapse.^{2,3}

High-throughput studies^{14,15} have established that, though displaying a markedly lower mutational burden compared to solid tumors,¹⁶ CLL is characterized by a diverse genetic landscape with driver gene mutations in pathways considered central for disease pathogenesis, e.g. NOTCH and NF-κB signaling.^{7,9,17} The frequency of most driver gene mutations in CLL tends to increase in aggressive/refractory cases supporting their involvement mainly in disease progression.¹⁸⁻²⁰

Chronic lymphocytic leukemia is preceded by a condition termed monoclonal B-cell lymphocytosis (MBL) that is characterized by the presence of circulating monoclonal B cells with a CLL phenotype, however, at a lower concentration than

Correspondence:

ghia.paolo@hsr.it

Received: August 8, 2017.

Accepted: February 7, 2018.

Pre-published: February 15, 2018.

doi:10.3324/haematol.2017.177212

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/103/5/865

©2018 Ferrata Storti Foundation

Material published in *Haematologica* is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>.

Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



required for a clinical diagnosis of CLL ($\geq 5 \times 10^9/L$).²¹⁻²⁴ MBL, found in otherwise healthy individuals, is divided into 2 subtypes based on the number of circulating cells: 'high-count MBL' (HC-MBL: $0.5-5 \times 10^9/L$) that evolves into CLL requiring therapy at a rate of 1%/year,²⁵ and 'low-count MBL' (LC-MBL: $<0.5 \times 10^9/L$) that has not been observed to progress into a clinical disease,²⁶ yet persists over time.^{26,27} Several typical CLL driver gene mutations have been reported in HC-MBL^{9,28,29} even years before the transition to CLL,³⁰ and these correlate with adverse disease course.³¹ Such mutations have been reported in multipotent hematopoietic progenitor CD34⁺ cells from patients with CLL,³² suggesting that such aberrations may also be implicated in CLL onset.

Here, we aimed to gain insight into the genetic lesions that may be involved in the transformation from MBL to CLL, analyzing LC-MBL cases for the first time. To this end, we used whole-genome sequencing (WGS) and targeted re-sequencing to profile LC-MBL, HC-MBL and a particularly indolent subset of CLL, i.e. patients with ultra-stable disease for more than ten years, thus, clinically analogous to MBL. Moreover, in order to explore the possible origin of genetic lesions at the hematopoietic progenitor cell level, we analyzed polymorphonuclear (PMN) cells from the study participants.

We report that the genomic profiles of ultra-stable CLL patients are very similar to individuals with LC-MBL and HC-MBL, characterized by infrequent CLL driver gene mutations that, however, were not associated with disease progression. Furthermore, we observed non-coding variants (NCVs) that target key pathways/cellular processes relevant to normal and neoplastic B-cell development, thus, potentially contributing to the leukemic transformation. We also found shared somatic mutations between MBL/CLL and PMN cells, strengthening the notion that at least a proportion of somatic mutations may occur before the onset of CLL.

Methods

The research protocol was approved by the Institutional Ethics Committee and all participants gave written informed consent in accordance with the Declaration of Helsinki.

Study population

The study cohort comprised 9 subjects with LC-MBL, 13 subjects with HC-MBL, and 7 patients with Rai stage 0 CLL, herein called 'ultra-stable' CLL. Detailed information about the study cohort is provided in the *Online Supplementary Appendix*.

Cell samples

Chronic lymphocytic leukemia cells were stained with anti-CD19, anti-CD5 and anti-CD20 antibodies. CD19⁺CD5⁺CD20^{dim} cells were sorted using a High Speed FACS Sorter MoFlo (Beckman Coulter) according to previously published methods.²⁶ PMN cells were sorted based on physical parameters. Buccal cells were collected with the use of appropriate buccal swabs (Epicentre, Madison, USA).

DNA extraction

The NucleoSpin® Tissue XS kit (Macherey-Nagel, Germany) was used for DNA extraction in samples with less than 5×10^4 cells and the QIAamp DNA Micro kit (Qiagen, Germany) in samples with cell numbers ranging between 5×10^4 and 1×10^6 . The QIAamp

DNA Blood Mini kit (Qiagen, Germany) was used for samples with more than 1×10^6 cells as well as for the buccal samples. DNA quantity and quality were assayed using the Qubit dsDNA HS Assay Kit (Life Technologies, USA).

WGS: library preparation

The Nextera technology was utilized for the library construction (Nextera™ DNA Sample Prep Kit, Illumina, USA) as it requires low input material whilst maintaining library complexity. Fifty ng of genomic DNA were used for the construction of libraries that were sequenced in paired-end mode 2x100bp on a HiSeq 2000 (Illumina, USA).

A variant allele frequency (VAF) of 10% was used as threshold for variant calling. More detailed information regarding the bioinformatics analysis is given in the *Online Supplementary Appendix*.

Targeted re-sequencing: library preparation

Probes targeting all coding exons or hotspot regions of 11 known or postulated CLL driver genes (*ATM*, *BIRC3*, *MYD88*, *NOTCH1*, *SF3B1*, *TP53*, *EGR2*, *POT1*, *NFKBIE*, *XPO1*, *FBXW7*) (*Online Supplementary Table S1*) were designed using Agilent's SureDesign service (<https://earray.chem.agilent.com/suredesign/home.htm>). The target regions were captured using the HaloPlex HS targeting enrichment kit (Agilent Technologies, USA). Paired-end sequencing (150 bp reads) was performed on the NextSeq instrument with the use of the 500/550 High Output Kit (Illumina, USA).

Gene enrichment analysis

The identification of genes/gene pathways (gene enrichment analysis, GEA) enriched within the targets of NCVs and motif-breaking events caused by NCVs was performed with Enrichr³¹ using the KEGG 2016 gene database.

Results

WGS reveals highly similar mutational profiles in MBL and ultra-stable CLL

Whole-genome sequencing was performed on 6 individuals with LC-MBL, 5 individuals with HC-MBL, and 5 patients with ultra-stable CLL. For each individual/patient, samples from MBL/CLL cells and PMN cells were evaluated against buccal (control) cells resulting in a total of 48 samples sequenced with an average autosomal coverage of 32X (*Online Supplementary Table S2*). Basic demographic and biological characteristics of the MBL/CLL cases included in the WGS analysis are provided in *Online Supplementary Table S3*. Overall, 37,033 somatic variants were detected in MBL/CLL samples with an average of 2040 somatic variants in LC-MBL (range: 298-2871), 2558 in HC-MBL (range: 1428-3483), and 2400 in CLL (range: 1650-3176), respectively. Notably, 2792 variants were identified in the 15 PMN control samples compared with buccal DNA, with an average of 186 variants/sample (the PMN sample from case CLL_3 was excluded from the analysis due to tumor cell contamination) (Figure 1A). Highly analogous mutation rates were observed in HC-MBL and CLL (0.79 and 0.74 mutations per Mb, respectively), while a slightly lower rate was seen in LC-MBL (0.63 mutations per Mb); this latter finding was due to sample LC-MBL_1 (excluding this sample, the average mutation rate for LC-MBL would have been 0.74 mutations per Mb) (Figure 1B). The ratio of single-nucleotide variants (SNVs)/small indels was again almost identical in

the three entities (ranging from 11.4 to 12.6), whereas it was significantly lower in the PMN samples (3.9 for LC-MBL and CLL and 4.1 for HC-MBL, respectively) ($P < 0.005$ for all cases) (Figure 1C).

The transition to transversion (Ti/Tv) ratio ranged from 0.99 to 1.13 in the MBL/CLL samples, while it was slightly lower in the PMN samples (0.94) (Figure 2A). No clear differences were observed between MBL/CLL and PMN samples when the distribution of mutations among the six types was examined (Figure 2B). We then evaluated the sequence context of each mutation by incorporating information on the bases immediately upstream and downstream of the mutated base, hence leading to 96 possible mutation types in this classification¹⁶ (Online Supplementary Table S4). Almost all major differences were identified between MBL/CLL samples and PMN samples with the former group exhibiting mainly C>T mutations at NpCpG trinucleotides ($P < 0.05$ for ultra-stable CLL and HC-MBL).

The MutationalPatterns package³³ was used to delineate the mutational signatures in our cohort. Mutational patterns identified in the MBL/CLL samples resembled those reported by Puente *et al.*;⁹ this finding was corroborated by

calculating the pairwise similarity with the 30 previously published signatures, where signature 9 and, to some extent, signature 1 where the main contributors (Figure 2C). The same analysis in the PMN samples gave different results, with a strong impact of mutational signatures 3 and 5 (Figure 2D). Signature 3 had previously been identified in solid tumors and is associated with failure of DNA double-strand break-repair by homologous recombination.¹⁶ Signature 5 exhibits transcriptional strand bias for C>T and T>C mutations at ApTpN context and displays a correlation between smoking history and mutation contribution.¹⁶

MBL and ultra-stable CLL display a paucity of mutations in putative CLL driver genes

Whole-genome sequencing identified 186 non-synonymous exonic variants amongst MBL/CLL samples and 15 amongst PMN samples. The average number was 8.9 for LC-MBL (range: 1-16), 14.8 for HC-MBL (range: 9-27), 11.6 for ultra-stable CLL (range: 7-19), and 0.9 for the PMN samples (range: 0-6), respectively (Figure 3A). In MBL/CLL samples, the vast majority of non-synonymous mutations were missense [LC-MBL: 47 of 53 (88.7%); HC-

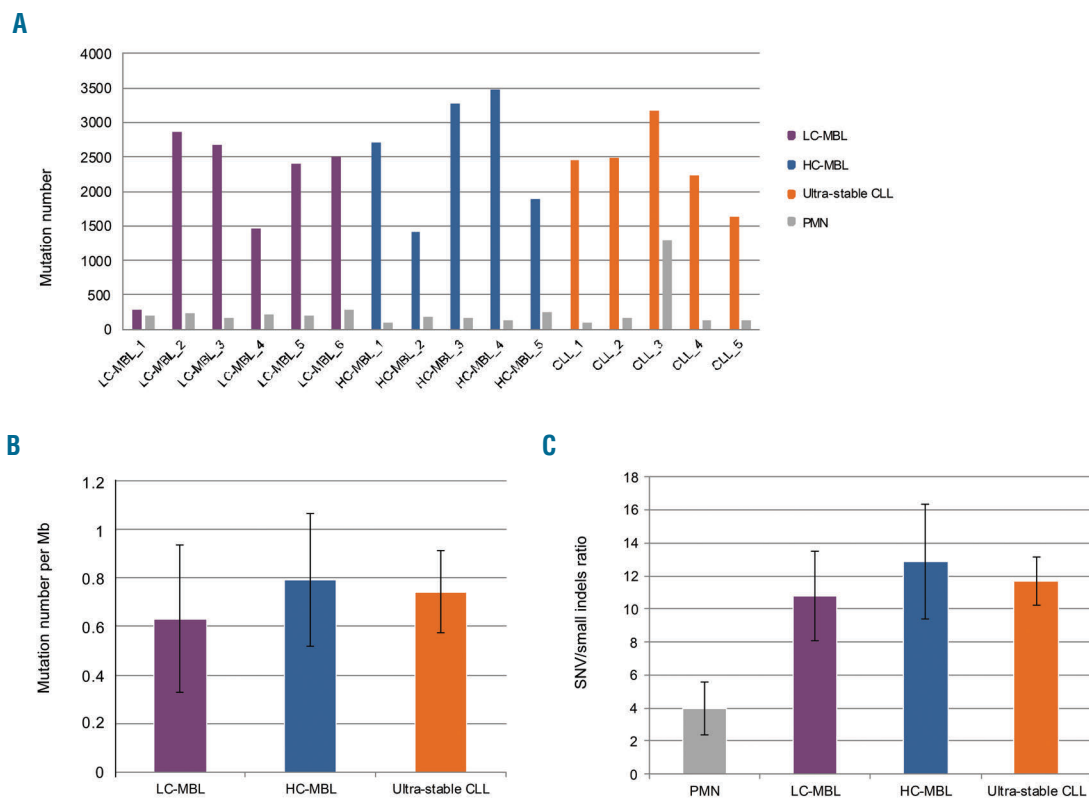


Figure 1. Somatic mutational analysis of ultra-stable chronic lymphocytic leukemia (CLL), high-count monoclonal B-cell lymphocytosis (HC-MBL), low-count monoclonal B-cell lymphocytosis (LC-MBL) and control polymorphonuclear (PMN) cell samples. (A) Total number of somatic mutations identified by whole-genome sequencing (WGS) in CLL cell samples from MBL, CLL and the respective PMN samples. All samples carried similar mutational loads with the exception of a single LC-MBL sample (LC-MBL_1) that displayed a very low number of mutation events; as can be seen, the corresponding PMN sample had a mutation load similar to the other PMN samples, where comparison of mutation profiles between the MBL and PMN sample showed few common hits, thus excluding the likelihood of contamination. Concerning PMN control samples, they were also characterized by high homogeneity regarding the mutational load. There was a single sample with a very high mutational load; detailed comparison against its respective CLL sample showed a high overlap of mutations indicating potential tumor cell contamination, hence this sample was removed from downstream analysis. (B) Average mutation rates \pm Standard Deviation (SD) for LC-MBL, HC-MBL and CLL. Highly analogous mutation rates were observed in the HC-MBL (0.79 mutations per Mb) and CLL (0.74 mutations per Mb) samples, while LC-MBL samples had a slightly lower ratio (0.63 mutations per Mb). (C) Average SNV to small indels ratio \pm SD for all sample groups. All 3 entities displayed similar ratios in clear contrast to the PMN samples where the ratio was much lower.

MBL: 61 of 73 (83.7%); ultra-stable CLL: 50 of 59 (84.7%)], whereas the remainder concerned either nonsense mutations or frameshift indels (Figure 3B and *Online Supplementary Table S5*). Forty-nine of the 186 mutations (26.3%) had a VAF more than 50%. Concerning the 7 PMN samples harboring mutations, only a single mutation (6.7%) had a VAF more than 50% (Figure 3C) (*Online Supplementary Table S6*). The most commonly mutated gene was *IGLL5*, in accordance with a recently reported study,⁶ carrying mutations in 5 different samples (2 LC-MBL, 2 HC-MBL and 1 CLL samples), likely introduced by the somatic hypermutation (SHM) process. Only 6 of 186 mutations (1.6%) detected in the MBL/CLL samples concerned putative CLL driver genes, according to 2 recently reported lists.^{7,9} In detail, 3 were identified in individuals with HC-MBL: i) a single *NOTCH1* p.P2514Rfs*4 deletion (VAF 20%), a known hotspot mutation in CLL^{10,28,34-36} in HC-MBL_4; ii) a single *FBXW7* p.W307S mutation (VAF 26%) in HC-MBL_2; and iii) a single *KIAA0947* p.L2093X (VAF 43%) in HC-MBL_5. Two mutations concerned individuals with LC-MBL: i) a *KLHL6* p.A91D mutation (VAF 45%) in LC-MBL_5; and ii) a single *CD79A* p.E200G mutation (VAF 53%) in LC-MBL_6. Finally, a *CD79B* p.N68S mutation (VAF 41%) was identified in a single CLL sample (CLL_5). Although most of these exact mutations have not previously been reported in CLL, functional

prediction using Polyphen-2 classified all but the *CD79B* mutation as probably damaging. No CLL driver gene mutations were found in the PMN samples.

To assess whether the non-synonymous mutations identified here might be potentially relevant to CLL, we compared our findings to the variants reported by Puente *et al.*⁹ and the International Cancer Genome Consortium (ICGC) database.³⁷ Overall, the vast majority of genes carrying mutations in our series were also reported as mutated in either or both datasets: 94% in LC-MBL, 89% in HC-MBL, and 97% in CLL.

We extended our analysis by performing targeted resequencing of 11 putative CLL driver genes in 8 LC-MBL, 13 HC-MBL and 7 ultra-stable CLL samples as well as 24 corresponding PMN samples. All but one LC-MBL case (LC-MBL_4) subjected to WGS were included in this analysis (*Online Supplementary Table S7*). In total, 5 variants were detected in 3 different HC-MBL samples, including 4 missense variants and 1 frameshift deletion. Two variants (targeting the *NOTCH1* and *FBXW7* genes) had been already identified by WGS, whereas the remaining 3 concerned the *POT1* (n=2) and *SF3B1* (n=1) genes. In detail, a single HC-MBL case (HC-MBL_5) harbored an *SF3B1* mutation (p.K700E; VAF 1.1%), a known hotspot mutation in CLL,^{18,29,38,39} and a *POT1* mutation (p.M1V; 4.3%), while the other *POT1* mutation (p.S38R; 6.7%) was found

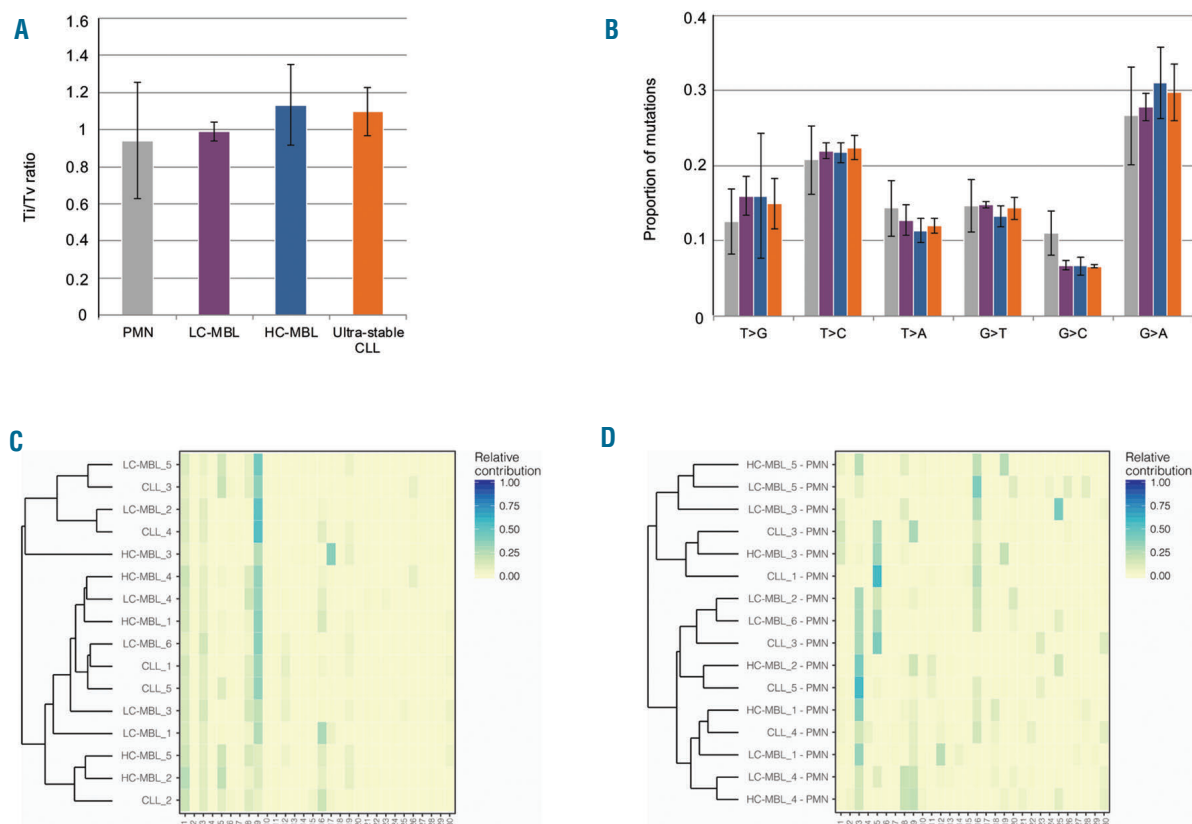


Figure 2. Detailed analysis of mutation types. (A) Transition to transversion (Ti/Tv) ratios were comparable in all monoclonal B-cell lymphocytosis (MBL)/chronic lymphocytic leukemia (CLL) samples and somewhat lower in the polymorphonuclear (PMN) cell samples. (B) Similar distribution of mutations among the 6 mutation classes for each MBL/CLL entity and PMN samples (average values \pm Standard Deviation). Similar profiles were evident for all entities with the G>A mutation predominating in all cases. (C) Mutational signatures that contribute to the somatic mutations observed in the MBL/CLL samples. (D) Mutational signatures that dominate in the PMN samples.

in HC-MBL₂, which also harbored the *FBXW7* mutation. Two somatic non-synonymous variants were identified in 2 different PMN samples: an *ATM* variant (p.R337C; VAF 20%) detected in an LC-MBL case, frequently reported and probably with a low functional impact, and a *TP53* mutation (p.G245A, VAF 3%) found in a HC-MBL case, previously reported in several human cancers and lymphomas⁴⁰ (Online Supplementary Table S8).

Non-coding variants in MBL and ultra-stable CLL target genes in pathways relevant to CLL pathogenesis

Coding and non-coding regions enriched for mutations were detected using Fishhook²⁹ in 10 kilobase windows across the genome and with compensation for replication timing. In line with previous findings,⁹ the analysis revealed highest mutational enrichment in the IG loci and within sites known to be recurrently affected by off-target somatic hypermutation (e.g. *BTG2*, *BCL6* and *TCL1A*) (Online Supplementary Figure S1).

Funseq2,⁴¹ a bioinformatics tool investigating the linkage between NCVs and target genes using integrated bisulfite sequencing, ChIP-Sequencing, and RNA-sequencing data from the Roadmap Epigenomics Project, was used for the examination of the NCVs. This analysis revealed a total of 1517 variants in the MBL/CLL samples and 39 in the PMN samples. After stringent filtering, 106 NCVs of potential relevance to MBL and CLL were identified (Online Supplementary Table S9): 29 in LC-MBL (average 4.8), 45 in HC-MBL (average 9), and 32 in CLL samples (average 6.4), respectively; only 4 NCVs were found in 2 PMN samples. Since we intentionally selected for NCVs in transcription factor (TF) highly occupied regions (see Online Supplementary Appendix), not unexpectedly most variants were located in gene promoter sites (Figure 4A). Twenty-nine variants (26.4%) concerned 16 cancer-associated genes and were evenly distributed amongst the 3 entities: 9 in LC-MBL samples, 11 in HC-MBL, and 9 in ultra-stable CLL samples. Three of these cancer-associated genes were recurrently targeted: 9 variants concerned the *BTG2* gene in 4 samples (2 CLL, 1 HC-MBL and 1 LC-MBL), 5 variants involved the *BCL6* gene in 2 samples (1 HC-MBL and 1 LC-MBL), and 2 variants targeted the *BIRC3* gene in 2 samples (1 CLL and 1 HC-MBL). We also identified 6 variants concerning the *ST6GAL1* gene in 3 CLL samples and the same *NKIRAS1*-related variant in 2 CLL samples (Figure 4B). Moreover, pathway analysis with Enrichr³¹ showed that 30 of 110 (27.3%) of the variants targeted genes were implicated in key CLL pathways and cellular processes, such as the PI3K-AKT pathway (*TCL1A*, *CCND1*, *BCL2*, *PKN1*, *DDIT4* and *SGK3*) ($P < 0.05$), the NF- κ B pathway (*BIRC3*, *BCL2* and *PLAU*) ($P < 0.05$) and the spliceosome machinery (*DDX46* and *HSPA2*). In most of these cases (22 of 30, 73.3%), the variants were located at promoter sites. Comparison to the series by Puente *et al.*⁹ identified 4 common gene targets: *BTG2*, *BCL6*, *BACH2* and *TCL1A*; none of our samples carried variants affecting either the 3' UTR of the *NOTCH1* gene or the *PAX5* gene enhancer.

We also analyzed the predicted impact of the NCVs on TF binding and found that 72 of 110 (65.5%) of the variants could result in a motif-breaking event (LC-MBL: $n=21$, HC-MBL: $n=33$, ultra-stable CLL: $n=18$) (Online Supplementary Table S10). We subsequently investigated genes and gene pathways that may be affected by such TF motif breaks. In 55 of 72 (76.4%) cases, variants disrupted

a DNase I hypersensitive site, while enrichment analysis of the implicated target genes using Enrichr³¹ led to the identification of genes participating in pathways relevant to CLL pathogenesis, such as the MAPK, WNT and AP-1 pathways ($P < 0.0005$) (Online Supplementary Table S11).

Moreover, we examined the potential relation of the NCVs that affect TF binding to AID activity by checking if they occurred in the known hotspots (WRCY, RGYW, WA, TW). According to our findings, 21 of 72 (29.2%) NCVs were located at AID hotspots. Gene enrichment analysis of the remaining 51 target genes revealed similar pathways as in the original analysis (namely AP-1 and DNA damage response pathways).

Shared mutations between CLL and polymorphonuclear cells indicate that somatic variants can arise before CLL onset

Shared mutations between MBL/CLL samples and their respective PMN samples were identified in all samples irrespective of origin. Regarding exonic mutations, the same synonymous *GSE1* mutation was found in an HC-MBL case and its paired PMN sample with comparable VAF (28% vs. 26%). In addition, a LC-MBL sample and its paired PMN sample carried an identical mutation within the ncRNA gene *LOC339874*, though with different VAF (16% vs. 31%). In the case of non-exonic mutations, 179 shared NCVs were identified between MBL/CLL and PMN samples (Online Supplementary Table S12); the average number per sample was 15.8 for LC-MBL, 8.2 for HC-MBL, and 9 for ultra-stable CLL (range: 2-34), respectively. Most of these mutations were intergenic (128 of 179, 71.5%) (Figure 4C). Interestingly, 6 NCVs were recurrently found in more than one MBL/CLL-PMN sample pair: 3 were intergenic and the other 3 were intronic (Online Supplementary Table S13). Finally, we also examined the mutational signatures for shared mutations between MBL/CLL and PMN samples but did not observe clear correlations with any signature (data not shown).

In order to exclude the possibility of contamination of the PMN cell fraction by MBL/CLL DNA, we designed allele-specific primers (Online Supplementary Table S14) and performed PCR amplification of the clonotypic IGH gene rearrangement in both the MBL/CLL and the respective PMN samples in 11 of 16 cases with available material. We identified the clonotypic rearrangement in all 11 MBL/CLL samples but in none of the corresponding PMN samples examined, effectively ruling out the possibility that the observed results were due to contamination.

Somatic copy-number analysis

sCNA analysis was performed in 3 samples from each entity and their respective PMN control samples, as well as in 4 additional PMN samples from 1 HC-MBL and 3 LC-MBL cases. In total, 16 sCNAs were identified in the MBL/CLL samples (average: 1.8, range: 1-6); 7 in LC-MBL, 4 in HC-MBL, and 5 in the ultra-stable CLL samples, all but one concerning deletion events (Online Supplementary Table S15). Of the recurrent cytogenetic aberrations included in the Döhner hierarchical model,⁴² del(17p), del(11q) and trisomy 12 were not identified in any of the samples, whereas del(13q) was detected in 7 of 9 MBL/CLL cases (2 LC-MBL, 3 HC-MBL and 2 CLL cases). FISH analysis gave concordant results in 5 of 7 cases where data from both techniques were available; in the remaining 2 cases del(13q) was detected with a single

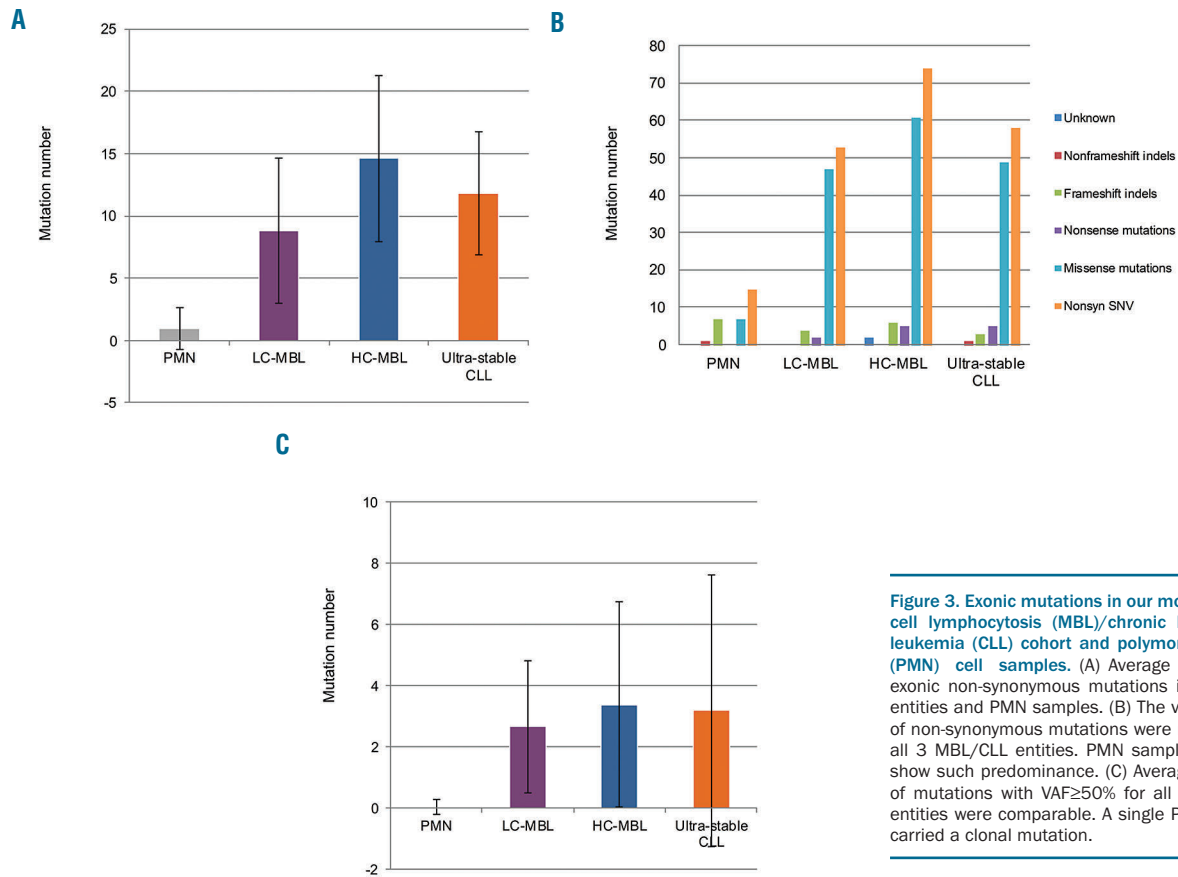


Figure 3. Exonic mutations in our monoclonal B-cell lymphocytosis (MBL)/chronic lymphocytic leukemia (CLL) cohort and polymorphonuclear (PMN) cell samples. (A) Average numbers of exonic non-synonymous mutations in MBL/CLL entities and PMN samples. (B) The vast majority of non-synonymous mutations were missense in all 3 MBL/CLL entities. PMN samples failed to show such predominance. (C) Average numbers of mutations with VAF \geq 50% for all 3 MBL/CLL entities were comparable. A single PMN sample carried a clonal mutation.

technique each (*Online Supplementary Table S16*). All other sCNAs represented unique events. In terms of distribution across the chromosomes, 7 of the 32 (21.9%) sCNAs were found in the vicinity of centromeres, whereas 11 of 32 (34.4%) were located close to a telomere (distance $<10 \times 10^7$ bp). None of the PMN samples demonstrated sCNAs typical of CLL. Only one MBL case showed a shared del(8)(p11.22) between the LC-MBL sample and its paired PMN sample.

Discussion

Limited information is available concerning the genomic landscape at the very early or indolent phases of CLL. To this end, we compared the genomes of ultra-stable CLL cases, defined as those cases stable for more than ten years after diagnosis, to genomes from individuals with: i) LC-MBL, a condition that does not progress into a clinically relevant leukemia;²⁶ and, ii) HC-MBL, a clinically identifiable pre-leukemic state.²⁵

Both types of MBL and ultra-stable CLL exhibited the same low level of genomic complexity, similar genome-wide mutation rates, and average number of exonic mutations, which were distinct from those of the control samples. Reflecting this similarity, analysis relating to published mutational signatures revealed similar patterns in samples from all 3 entities. In more detail, signature 9 that predominated in the MBL/CLL cohort has been previously identified in CLL and B-cell lymphomas and is attributed to polymerase η that is involved in AID-induced

somatic hypermutation.¹⁶ The second ranking signature 1 is an age-related signature stemming from spontaneous deamination of 5-methylcytosine that has been detected in many cancer types.¹⁶ Analogies between MBL and ultra-stable CLL extended also to sCNAs in that all samples, irrespective of origin, carried very few sCNA. Del(13q) predominated in all three entities, as shown in previous studies.²⁶ Most of the sCNAs were located in close proximity to either centromeres or telomeres, in keeping with previous findings reporting significant over-representations in these regions due to duplication rates.⁴³ Thus, most of the sCNAs identified here may not be directly related to the MBL/CLL phenotype.

Interestingly, PMN cells harbored a significantly higher load of mutations compared to buccal cells. Mutations detected in the PMN samples were characterized by the dominance of distinct mutational signatures compared to the MBL/CLL cohort. However, these samples carried shared somatic mutations with the respective MBL/CLL cell samples in all analyzed cases. Most shared mutations concerned intergenic regions, yet we also identified a single shared exonic mutation. This finding supports the notion that some mutations present in the CLL clone could be acquired prior to disease onset, as previously suggested.³²

Almost all genes that were found mutated in HC-MBL and/or LC-MBL had been previously described as recurrently mutated in CLL.^{9,37} In contrast to our recent WES study on relapsing CLL,⁸ where the great majority of cases carried at least one CLL driver mutation, such mutations were relatively scarce in our cohort. Most importantly, the

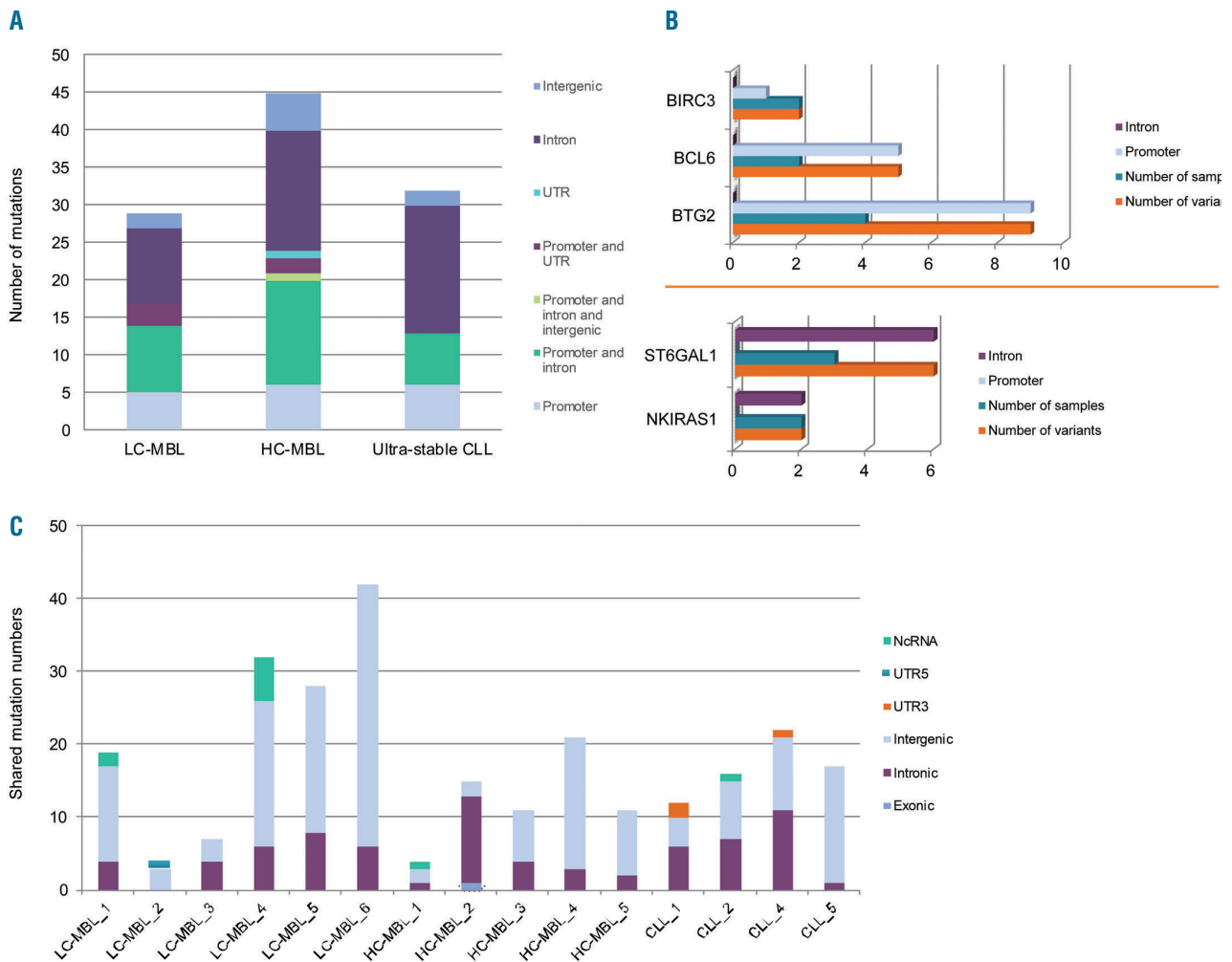


Figure 4. Analysis of non-coding variants and shared mutations between monoclonal B-cell lymphocytosis (MBL)/chronic lymphocytic leukemia (CLL) and polymorphonuclear (PMN) cell samples in the present cohort. (A) Topology of the 106 relevant non-coding variants identified in the present study. The majority concerned gene promoter sites. (B) Recurrent non-coding variants in genes relevant to CLL. The *BIRC3*, *BCL6* and *BTG2* genes are known to be associated with various types of cancer. (C) Topology of shared mutations between MBL/CLL and PMN samples. Intergenic mutations predominated, followed by intronic mutations.

lack of any obvious impact of the identified mutations on disease progression after a prolonged follow up highlights the fact that the mere presence of a given driver mutation does not axiomatically equate with disease progression, as previously reported.^{31,44} Additional studies are required in order to clarify this phenomenon.

Chronic lymphocytic leukemia cases have been shown to harbor detrimental gene mutations in subclones not visible at diagnosis that are progressively selected, e.g. following the use of chemotherapy.^{45,46} Such mutations were identified by targeted re-sequencing, yet only in a minor fraction of the present cohort. In this context, it has been recently proposed that sequence depths greater than 4000X will be essential in order to robustly identify all subclonal mutations and predict aggressive cases.¹² Arguably, the absence of such driver mutations when applying highly sensitive methods may potentially help to identify individuals with very indolent disease for whom less frequent, if any follow up will be warranted.

Whole-genome sequencing revealed that the non-cod-

ing mutome commonly targets gene pathways and cellular processes involved in CLL pathobiology. A sizeable proportion of variants affected the promoter sites of genes previously associated with cancer, e.g. *BTG2*, *BCL6*, *BACH2* and *TCL1A*. Interestingly, all 4 genes have been recently reported as non-IG targets of the SHM process in patients with lymphomas,⁴⁷ implicating this otherwise normal process in the emergence of CLL-like clones. Additional studies need to be performed to address the relevance of such mutations; however, since AID-related mutations are common in CLL,¹⁶ they may indeed be relevant to disease pathogenesis. We did not identify any “poor-prognostic” 3’ UTR *NOTCH1* mutations;⁹ however, we did discover two NCVs targeting indirectly the *BIRC3* gene. We also found a number of variants within the promoter sites of genes implicated in pathways relevant to CLL biology, including the PI3K-AKT and NF-κB pathways as well as the spliceosome machinery. Furthermore, we identified TF binding motif breaking events that may arise due to NCVs; most concerned the

MAPK, WNT and AP-1 signaling pathways. In this context, preliminary results (*data not shown*) from our ongoing high-throughput study on aggressive CLL cases showed a great degree of consistency in the targeting of NCVs: the same “CLL-relevant” gene pathways were again among the most common targets of NCVs, further corroborating our present findings. Having said that, some of these variants could represent bystander SHM targets of unknown significance or minor contributors to disease pathogenesis, therefore requiring further studies before definitive conclusions can be drawn regarding their actual significance.

It is important to note that a recent study on the epigenetic profile of CLL⁴⁸ reported a novel pathogenic role of TF dysregulation in CLL, with increased activity of EGR and NFAT as well as loss of EBF and AP-1, causing imbalances in the normal B-cell epigenetic program. Interestingly, certain members of these networks (e.g. *EBF1*, *JUN* and *FOS*) were among the most commonly affected TFs across all sample types tested. Collectively, our findings support the notion that gene pathways could be indirectly targeted by NCVs with the targets being either the genes themselves or other interacting genes, e.g. TFs.

Limitations of the present work involve the relatively small size of the cohort, mainly due to the rarity of samples meeting the selection criteria. In particular, CLL patients had to have stable disease after a prolonged follow up, whereas all individuals with MBL had to have a persistent monoclonal B-cell population. Concerning LC-MBL, low CLL cell number was an additional challenging factor. Furthermore, although our targeted re-sequencing approach covered almost 50% of reported mutations in putative driver genes (as reported by Puente *et al.*),⁹ by definition this approach is not exhaustive.

In summary, we report that MBL and ultra-stable CLL are virtually indistinguishable at the genomic level. While this may be reflective of a passive and slow accumulation of mutations, we identified both exonic and NCV-targeted pathways central for B-cell biology and CLL development, likely linked to the acquisition of the MBL/CLL phenotype. Importantly, ultra-stable CLL cases carried few known driver gene mutations, even after ten years of follow up, perhaps reflecting the central role of microenvironmental signals rather than cell-intrinsic defects in shaping clonal behavior. In other words, cell-extrinsic triggering, specifically mediated through the B-cell receptor, might represent the major driving force in the early stages of CLL, whereas disease progression will require acquisition of genetic driver mutations.

Funding

This research project was supported by the *Associazione Italiana per la Ricerca sul Cancro*, AIRC (Investigator Grant #15189 to PG and Special Program Molecular Clinical Oncology – 5 per mille #9965), Milano, Italy, *Ricerca Finalizzata 2010* (#2318823 to PG); *Swedish Cancer Society*, the *Swedish Research Council*, *Uppsala University*, *Uppsala University Hospital*, *Lion's Cancer Research Foundation*, and *Selander's Foundation*, Uppsala; and, H2020 “AEGLE, An analytics framework for integrated and personalized healthcare services in Europe”, by the European Union; “MEDGENET, Medical Genomics and Epigenomics Network” (No.692298) by the European Union; “GCH-CLL” funded by the General Secretariat for Research and Technology (GSRT) of Greece and the Italian Ministry of Health (MoH); and IMI2 “HARMONY”, funded by the European Union. AA is a fellow of *Associazione Italiana per la Ricerca sul Cancro* AIRC (Triennial fellowship “Guglielmina Lucatello é Gino Mazzega”).

References

- Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *N Engl J Med*. 2005;352(8):804-815.
- Parikh SA, Shanafelt TD. Prognostic factors and risk stratification in chronic lymphocytic leukemia. *Semin Oncol*. 2016; 43(2):233-240.
- Scarfo L, Ferreri AJ, Ghia P. Chronic lymphocytic leukaemia. *Crit Rev Oncol Hematol*. 2016;104:169-182.
- Fabbri G, Khiabanian H, Holmes AB, et al. Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *J Exp Med*. 2013;210(11):2273-2288.
- Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med*. 2011;208(7):1389-1401.
- Kasar S, Kim J, Improgo R, et al. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun*. 2016; 6(8866).
- Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526(7574):525-530.
- Ljungstrom V, Cortese D, Young E, et al. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood*. 2016;127(8):1007-1016.
- Puente XS, Bea S, Valdes-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015; 526(7574):519-524.
- Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101-105.
- Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 2012;44(1):47-52.
- Rose-Zerilli MJ, Gibson J, Wang J, et al. Longitudinal copy number, whole exome and targeted deep sequencing of 'good risk' IGHV-mutated CLL patients with progressive disease. *Leukemia*. 2016;30(6):1301-1310.
- Schuh A, Becq J, Humphray S, et al. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood*. 2012;120(20):4191-4196.
- Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365(26):2497-2506.
- Burns A, Alsolami R, Becq J, et al. Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHV(mut) and IgHV(unmut) subgroups. *Leukemia*. 2018;32(2):332-342.
- Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500(7463):415-421.
- Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152(4):714-726.
- Rossi D, Brusca A, Spina V, et al. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. 2011;118(26):6904-6908.
- Rossi D, Fangazio M, Rasi S, et al. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood*. 2012;119(12):2854-2862.
- Schnaiter A, Paschka P, Rossi M, et al. NOTCH1, SF3B1, and TP53 mutations in fludarabine-refractory CLL patients treated with alemtuzumab: results from the CLL2H trial of the GCLLSG. *Blood*. 2013;122(7):1266-1270.

21. Landgren O, Albitar M, Ma W, et al. B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med.* 2009; 360(7):659-667.
22. Marti GE, Rawstron AC, Ghia P, et al. Diagnostic criteria for monoclonal B-cell lymphocytosis. *Br J Haematol.* 2005; 130(3):325-332.
23. Dagklis A, Fazi C, Scarfo L, et al. Monoclonal B lymphocytosis in the general population. *Leuk Lymphoma.* 2009; 50(3):490-492.
24. Ghia P, Prato G, Scielzo C, et al. Monoclonal CD5+ and CD5- B-lymphocyte expansions are frequent in the peripheral blood of the elderly. *Blood.* 2004; 103(6):2337-2342.
25. Rawstron AC, Bennett FL, O'Connor SJ, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med.* 2008;359(6):575-583.
26. Fazi C, Scarfo L, Pecciarini L, et al. General population low-count CLL-like MBL persists over time without clinical progression, although carrying the same cytogenetic abnormalities of CLL. *Blood.* 2011; 118(25):6618-6625.
27. Matos DM, Furtado FM, Falcao RP. Monoclonal B-cell lymphocytosis in individuals from sporadic (non-familial) chronic lymphocytic leukemia families persists over time, but does not progress to chronic B-cell lymphoproliferative diseases. *Rev Bras Hematol Hemoter.* 2015;37(5):292-295.
28. Rasi S, Monti S, Spina V, et al. Analysis of NOTCH1 mutations in monoclonal B-cell lymphocytosis. *Haematologica.* 2012; 97(1):153-154.
29. Imielinski M, Guo G, Meyerson M. Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. *Cell.* 2017;168(3):460-472 e414.
30. Kreutzer DA, Essigmann JM. Oxidized, deaminated cytosines are a source of C --> T transitions in vivo. *Proc Natl Acad Sci USA.* 1998;95(7):3578-3582.
31. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016; 44(W1):W90-97.
32. Damm F, Mylonas E, Cosson A, et al. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov.* 2014;4(9):1088-1101.
33. Blokzijl FJ, R; van Boxtel, R; Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *bioRxiv* 071761; doi: <https://doi.org/10.1101/071761>.
34. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia.* 2015;29(2):329-336.
35. Sutton LA, Young E, Baliakas P, et al. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica.* 2016;101(8):959-967.
36. Lionetti M, Fabris S, Cutrona G, et al. High-throughput sequencing for the identification of NOTCH1 mutations in early stage chronic lymphocytic leukaemia: biological and clinical implications. *Br J Haematol.* 2014;165(5):629-639.
37. Zhang J, Baran J, Cros A, et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford).* 2011;2011:bar026.
38. Strefford JC, Sutton LA, Baliakas P, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia.* 2013; 27(11):2196-2199.
39. Jeromin S, Weissmann S, Haferlach C, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia.* 2014;28(1):108-117.
40. Wilda M, Bruch J, Harder L, et al. Inactivation of the ARF-MDM-2-p53 pathway in sporadic Burkitt's lymphoma in children. *Leukemia.* 2004;18(3):584-588.
41. Fu Y, Liu Z, Lou S, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.* 2014; 15(10):480.
42. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med.* 2000;343(26):1910-1916.
43. Nguyen DQ, Webber C, Ponting CP. Bias of selection on human copy-number variants. *PLoS Genet.* 2006;2(2):e20.
44. Hurtado AM, Chen-Liang TH, Przychodzen B, et al. Prognostic signature and clonality pattern of recurrently mutated genes in inactive chronic lymphocytic leukemia. *Blood Cancer J.* 2015;5:e342.
45. Ojha J, Ayres J, Secreto C, et al. Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood.* 2015;125(3):492-498.
46. Rossi D, Khiabani H, Spina V, et al. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood.* 2014;123(14):2139-2147.
47. Khodabakhshi AH, Morin RD, Fejes AP, et al. Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget.* 2012;3(11):1308-1319.
48. Oakes CC, Seifert M, Assenov Y, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nature Genetics.* 2016;48(3):253-264.