Acquired Initiating Mutations in Early Hematopoietic Cells of CLL Patients

Frederik Damm1,3, Elena Mylonas1,3, Adrien Cosson4,5, Kenichi Yoshida12,15, Véronique Della Valle1,3,9, Enguerran Mouly1,3,9, M’boyba Diop1,3, Laurianne Scourzic1,3,9, Yuichi Shiraishi13,14, Kenichi Chiba13,14, Hiroko Tanaka13,14, Satoru Miyano13,14, Yoshikane Kikushige16,17, Frederick Davi4,5,6, Jérôme Lambert7, Daniel Gautheret3,10, Hélène Merle-Béral4,5,6, Laurent Sutton11, Philippe Dessen1,3, Eric Solary2,3,8,9, Koichi Akashi16,17, William Vainchenker2,3,9, Thomas Mercher1,3, Nathalie Dron1,3,9, Seishi Ogawa12,15, Florence Nguyen-Khac4,5,6, and Olivier A. Bernard1,3,8,9
INTRODUCTION

Cancer develops from an individual cell that accumulates acquired mutations. Appropriate medical care requires thorough understanding of the natural history of the disease, including the identification and order of occurrence of the mutations, the cell of origin, and the clonal organization of the tumor cells. In addition, because the transformation process can capture preexisting somatic mutations (1, 2), their driver nature needs to be fully established, based on their recurrence and their functional consequences. Such in-depth investigations identify initial driver mutations, which are relevant as targets for therapy.

Chronic lymphocytic leukemia (CLL), the most frequent adult leukemia in Western countries, is characterized by an accumulation of mature B lymphocytes (3). The CLL tumor cells are clonal, as assessed by rearrangement of the immunoglobulin heavy chain (IGH) gene, and express low levels of surface B-cell receptor (BCR). In a fraction of patients, the IGH variable gene segment (IGHV) rearrangement is mutated, reflecting normal somatic hypermutation triggered by antigen recognition. Patients with IGHV mutations have a better prognosis than those without IGHV mutations.

Investigation of CLL samples by massively parallel sequencing has identified a number of acquired somatic mutations (4, 5), but no individual gene is mutated in more than 20% of the patients. The products of these mutated genes are involved in RNA metabolism, genome stability and cell cycle, control of the Notch pathway, Wnt signaling, and inflammation (4). Transformation may also depend on specific IGH rearrangements and BCR intracellular signaling cascades (6, 7). The cell of origin of CLL is currently debated. Immunophenotype and expression profiling analyses pointed at mature CD5+ B cells (8), but the involvement of early hematopoietic cells in CLL development has been emphasized in xenograft experiments, which showed that the hematopoietic stem/progenitor cells from patients with CLL show biased and abnormal differentiation toward the B-lymphoid lineage in immunodeficient mice (9).

To investigate the natural history of CLL, we embarked on a thorough analysis of CLL samples using massive parallel sequencing and cellular analyses.
RESUL T S

**SF3B1 Mutations Are Detected in Nonlymphocytic Cellular Fractions of CLL Patients**

To search for CLL mutations in the hematopoietic progenitor cell fraction, we first investigated the distribution of SF3B1 mutations in the hematopoietic tree of patients with SF3B1-mutated CLL, because this gene is frequently mutated in both CLL and myelodysplastic syndrome (MDS), a chronic stem cell–derived myeloid tumor (10, 11). Sanger sequencing of the mutational hotspots in the SF3B1 gene using DNA from 50 patients with CLL identified 7 patients carrying an SF3B1 mutation. We next flow-sorted cells according to the expression of mutually exclusive cell-surface markers: CD34 (which marks the immature progenitor cells compartment at the apex of hematopoietic differentiation), and markers of mature cells, including CD3 (T cells), CD14 (monocytes), and CD19 (both normal and tumor B cells). Sequencing analyses of DNA from these cellular fractions showed wild-type SF3B1 sequences in the CD3\(^+\) cells and the mutated SF3B1 sequence in the CD19 fraction in all seven cases. The mutation was also observed in the CD34\(^+\) and/or CD14\(^+\) fractions in two patients (Supplementary Fig. S1), suggesting it was acquired in these patients in an early progenitor cell that was able to participate in both lymphoid and myeloid differentiation.

**Acquired Mutations Are Detected in Multipotent Progenitors in the Majority of CLL Patients**

We next used whole-exome sequencing of DNA from flow-sorted cell populations derived from 24 patients with CLL (17 IGHV-unmutated and 7 IGHV-mutated; Supplementary Table S1). Results of the IGH gene rearrangements were always compatible with monoclonal proliferation. Viable cells were flow-sorted to purities greater than 96% (see flow chart description; Supplementary Fig. S2A and S2B). Comparison of exome sequences from tumor cells and T lymphocytes (essentially spared by CLL mutations as shown for SF3B1 mutation) identified a total of 415 somatic mutations predicted to result in protein-coding changes in 361 different genes with a median of 17 mutations/patient (range, 7–34; Supplementary Table S1; Fig. 1A). Some mutations were present in virtually all CD19\(^+\) cells, whereas the allelic ratio of other mutations showed that they were present only
Early Driver Mutations in CLL

Figure 1. (Continued) B, variant allele frequencies in the hematopoietic fractions, sorted on the basis of the cell-surface expression of the following antigens (in abscissa): CD34+ (progenitors), CD3+ (T cells), CD19 and CD5+ (tumor cells), and CD14+ (monocytes). Column on the right, the genotyping of colony forming units (CFU) from single CD34+ cells. If only wild-type (WT) colonies were observed, the number of analyzed colonies is indicated in parentheses. The ratio of 50% and 4% are indicated by dotted lines, orange and black, respectively. ND, not done.

in a fraction of these cells, indicating that they were secondarily acquired.

We used targeted deep resequencing to simultaneously validate and quantify the mutation burden in DNA from the sorted fractions (Fig. 1B and Supplementary Table S1). Sorting impurity and aberrant antigen expression should be taken into account when analyzing cell-sorted fractions. Mutation burdens below 4% that potentially result from sorting contamination were regarded as negative. Among the 24 patients analyzed, only 3 (CLL03, 15, and 24) were devoid of mutations in the CD34+ progenitor or the CD14+ monocyte fractions. All 3 patients carry mutated IGHV rearrangements in their CLL cells. In the other 21 patients, at least one mutation was detectable in the CD34+ or in the CD34+ fractions. Two patients (CLL14 and 27) showed mutations in the CD34+ and not in the CD34+ fraction and, conversely, 6 patients (CLL10, 11, 16, 17, 18, and 22) showed mutation in the CD34+ but not in the CD14+ fraction. In 13 samples (CLL01, 02, 05, 06, 07, 08, 09, 12, 13, 19, 20, 26, and 28), at least one mutation was detected in both fractions. The presence of a CLL mutation in immunophenotypical progenitor (CD34+) or myeloid (CD14+) primary cells confirmed the involvement of immature cells in CLL pathogenesis. The burden of mutations detected in the immature hematopoietic cells (referred to hereafter as early mutations) was always among the highest mutation burdens in CLL cells, consistent with their occurrence at the initial steps of CLL development. However, in all patients only a subset of the CLL mutations was observed in the progenitors or myeloid fractions.

Because cell-surface expression of myeloid antigens is not sufficient to attest to the myeloid nature of a progenitor cell, we next tested the myeloid differentiation capacities of the mutated progenitors. We sorted single CD34+CD19- progenitor cells and grew them in vitro in myeloid conditions. Viable cells were available for 18 patients. The cloning efficiency was close to 60% for each patient (exceptions were CLL08 and CLL09) and colonies were confirmed as myeloid (CD14+) primary cells (erythroid, megakaryocytic, or/and granulo-monocytes) by FACS immunophenotyping of randomly chosen colonies. Colony genotyping confirmed the presence of CLL mutations in myeloid cells in 13 patients, whereas 5 patients (CLL03, 08, 16, 19, and 26) did not show mutated colonies.
(Fig. 1B and Supplementary Table S1). Although the absence of mutation may be due to the low number of colonies in some patients (as for patient CLL08), the other 4 patients clearly did not show mutated cells in over 50 colonies analyzed. In addition, the frequency of mutated colonies differed from the estimated mutation burden in the sorted CD34+ fractions, supporting the idea that not all mutated progenitors could grow in these myeloid culture conditions. We also investigated the myeloid colonies from 17 patients for the presence of the CLL IGH rearrangement using rearrangement-specific PCR. No IGH-rearranged colonies were detected for 11 patients (CLL01, 02, 03, 08, 09, 11, 12, 13, 16, 20, and 26). A low number of IGH-rearranged colonies was observed for 6 patients: CLL05 (8/96), 07 (2/109), 17 (1/96), 27 (2/96), 18 (3/96), and 19 (1/59). Nucleotide sequence analyses showed that the variable joining gene segments (VJ) junction amplified from the colonies matched the tumor cell rearrangement in patients 05 and 07. Colonies from patients 17 and 19 showed a rearrangement differing from those of their CLL counterpart. Half of the colonies from patients 18 and 27 showed the same rearrangement as the corresponding tumor cells, whereas the other half carried other VJ junctions. Of note, every VJ-positive colony also carried an early mutation. Together, these data demonstrate the presence of CLL mutations in a multipotent hematopoietic progenitor fraction in the majority of patients with CLL. Reasoning that the mutations had been originally acquired in a single cell, the high proportion of mutated cells in the CD34+ or CD14+ fractions demonstrates that the cell carrying the identified mutation had some clonal advantage and accumulated over time. The mutations seem to variably affect hematopoietic differentiation, as judged from the mutation burden detected in the hematopoietic fractions (see Supplementary Fig. S3A).

Some patients showed an overall normal balance between myeloid and B-lymphoid differentiation. They showed multilineage involvement indicative of an unbiased differentiation of the mutated stem/progenitor cells (for example, CLL02, 07, 12, and 20 in Fig. 1B and Supplementary Table S1). In our settings, a mutation would be detected only if it induces the accumulation of the mutated cell in the given fraction. If a mutation induces accumulation at a late step and not at early steps of differentiation, accumulation will occur in the mature cells (CD14+) and not the immature cells (CD34+). For example, patients CLL14 and CLL27 would belong to this first group of patients.

A second type of patient (for example, CLL10, 11, and 22) shows an unbalanced involvement of myeloid cells (a lower mutational burden than in CD34+ progenitors), suggesting that the early mutations bias the mutated stem/progenitor cells toward the lymphoid lineage or specifically allow the accumulation of lymphoid-primed progenitors. A third type of patient (CLL03, 15, and 24 in this series) lacks detectable mutation in either the myeloid or the progenitor compartments, suggesting either a strict commitment toward lymphoid differentiation or the involvement of a lymphoid-primed progenitor. Alternatively, these patients may follow a different transformation pathway. The numbers and burden of mutations did not differ statistically between these 3 patients and the others (Supplementary Table S1).

Early Mutations Affect Genes Recurrently Mutated in CLL and Other Malignancies

Mutations detected in the progenitors of patients with CLL affected genes already known to be mutated in CLL, in other hematologic malignancies, or even in other cancers, supporting their active role in transformation (Supplementary Table S2). Early mutations were observed in the NOTCH1, SF3B1, TP53, and XPO1 genes are among the most frequently mutated genes in CLL (4, 5, 12). Genes such as BRAF and ML22 are mutated in CLL and in other B-cell malignancies (13–15). A few EGR2- and NFKBIE-mutated patients have been reported in CLL (4, 5, 16). To further establish the importance of the early mutations identified in our patients, we investigated the recurrence of some of them by direct Sanger sequencing of the mutational hotspots of BRAF, EGR2, MED12, MYD88, NFKBIE, NOTCH1, SF3B1, TP53, and XPO1 in the 168 untreated patients with stage B and C CLL who were sampled at inclusion in a clinical trial (www.clinicaltrials.gov, NCT00931645; Supplementary Table S3; ref. 17). A total of 113 mutations in 84 patients were identified, and 84 of 168 (50.0%) patients presented with at least one mutation of this nine-gene panel (Fig. 2A and S and Supplementary Table S4). Inactivating mutations of NFKBIE were found in 10.7% (18 of 168; Fig. 2C) of the patients. Missense mutations of EGR2 were observed in 8.3% (14 of 168; Fig. 2C) of the patients and associated with higher CD38+ expression (median, 70% vs. 17%; $P = 0.009$), a known poor prognosis marker, a shorter time to treatment (median, 15.4 vs. 12 months; $P = 0.0006$), and a shorter 5-year overall survival (56.2 vs. 80.4 months; $P = 0.04$; Fig. 2D).

Deregulation of BCR Signaling as a Phenotypic Convergence of Early Mutations in CLL

Normal BCR and pre-BCR signaling occurs through BRAF, which activates ERK proteins (18), which in turn phosphorylate and activate the ternary complex factor–serum response factor (SRF) dimer, resulting in the upregulation of a set of immediate early genes, including EGR2 (19, 20). BRAF and EGR2 mutations may therefore affect the BCR signaling, which is abnormal in CLL (7). BRAF mutations, most frequently V600E, have been described in a variety of human malignancies, including hairy cell leukemia (13, 21), another malignant B-cell disease. In CLL, BRAF mutations target amino acids located in the P-loop of the kinase (Supplementary Table S4; ref. 22), leading to weaker activation than the canonical V600E mutations (18). Ectopic expression of the CLL-mutant BRAF-G469R in Ba/F3 cells showed a constitutive ERK phosphorylation and Egr2 transcription (Fig. 3A and B). To analyze the impact of BRAF-G469R in B-cell differentiation, we transduced hematopoietic progenitors with BRAF–wild-type (WT), BRAF-G469R, or empty murine stem cell virus (MSCV) vector and engrafted the cells in irradiated syngeneic recipients. Animals were analyzed after 5 weeks, before the onset of any gross hematologic disorders. Careful analyses of the immunoglobulin M (IgM)–positive B-cell compartment showed a decrease in the proportion of B cells in the BRAF-G469R mice, as compared with MSCV or BRAF-WT mice (Fig. 3C and D). In addition, the mean fluorescence of IgM was significantly lower in B220+IgM+ cells in the BRAF-G469R expressing cells than in their WT or MSCV counterparts (Fig. 3E). A similar abnormal (IgM low, IgD+) B-cell population was present in the spleen of the BRAF-G469R mice (Fig. 3C and E).
Early Driver Mutations in CLL

We next investigated the consequences of EGR2 mutations. The EGR2 gene encodes a versatile transcription factor that participates in the control of cellular differentiation, including myeloid (23), B-lymphoid, and T-lymphoid differentiation (24, 25). All EGR2 mutations identified in CLL were heterozygous missense mutations, and, with the exception of R426Q, were located within the zinc-finger domains (Fig. 2C). In addition, EGR2 mutations were detected as early molecular events in 2 patients (CLL12 and CLL22; Fig. 1B and Supplementary Table S1).

To investigate the functional consequences of EGR2 mutations, we first expressed GST fusion proteins, including the zinc-finger region of WT or two EGR2 mutants (E356K and H384N). Electrophoretic mobility shift assays (EMSA) using a biotinylated probe corresponding to a high-affinity EGR2 site (26) showed specific binding of the WT and the H384N proteins (Fig. 4A), although H384N binding seemed weaker than WT despite comparable protein amounts (Fig. 4B). The interaction of the E356K protein with the probe was not observed in this assay. To investigate their ability to regulate transcription, we expressed the WT and the mutant forms of EGR2 in the murine multipotent hematopoietic cell line EML. Expression levels of all EGR2 isoforms were comparable (Fig. 4C) and were associated with slower growth (Fig. 4E). Cells expressing WT EGR2 showed a progressive reduction in the expression of the cell-surface markers B220 (B lymphoid) and Gr1 (myeloid; Fig. 4D). Growth slowdown and loss of B220+ and Gr1+ cells occurred even faster in cells expressing mutant EGR2 (Fig. 4D and E), indicating that the mutations had a functional impact. We investigated the expression of known EGR2 target genes using RNA obtained from sorted GFP+ cells, 3 days after transduction, to detect primary transcriptional changes induced by EGR2 expression. As shown in Fig. 4F, WT and mutated EGR2 proteins interfered with the expression of EGR2 targets. The effects of the three EGR2 proteins were similar on Csf1 transcription, whereas WT-EGR2 was stronger than E356K, which was stronger than H384N, in the transactivation of Gadd45b. Taken together, these results indicate that the EGR2 mutations in CLL do not functionally inactivate the protein but rather affect the

---

**Figure 2.** Gene mutation profile in 168 patients with CLL. Frequencies (A) and distributions (B) of identified mutations. C, localization of identified mutations in EGR2 and NFKBIE proteins. Circles, mutations; filled circle, proven somatic mutation. D, overall survival of patients with CLL according to EGR2 mutation status (log-rank test).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutated patients</th>
<th>Overall frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF3B1</td>
<td>27</td>
<td>16.1</td>
</tr>
<tr>
<td>NFKBIE</td>
<td>18</td>
<td>10.7</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>15</td>
<td>8.9</td>
</tr>
<tr>
<td>TP53</td>
<td>14</td>
<td>8.3</td>
</tr>
<tr>
<td>EGR2</td>
<td>14</td>
<td>8.3</td>
</tr>
<tr>
<td>XPO1</td>
<td>14</td>
<td>8.3</td>
</tr>
<tr>
<td>BRAF</td>
<td>6</td>
<td>3.6</td>
</tr>
<tr>
<td>MED12</td>
<td>3</td>
<td>1.8</td>
</tr>
<tr>
<td>MYD88</td>
<td>2</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Patients with at least one mutation: 84/168 (50%).
transcription of EGR2 target genes. Whether the differential activities of the mutants are due to differences in DNA binding or interaction with transcription cofactors at target genes will require additional investigation.

To investigate the functional consequences of EGR2 mutations in patient samples, we analyzed RNA-seq data obtained from 16 CLL samples. Fifteen genes were downregulated, whereas 224 genes were specifically upregulated in EGR2-E356K samples (n = 4) as compared with EGR2 WT or unanalyzed patients (n = 10; P < 0.01; Supplementary Table S5). Hierarchical clustering using the 224 upregulated genes showed clustering of all 5 EGR2-mutated CLL samples, including the EGR2-H384R sample (Fig. 5A). An additional sample, which was not analyzed by exome sequencing and lacked acquired mutations in EGR2, clustered together with the EGR2-mutated samples, suggesting that other alterations mimic the effect of EGR2 mutations. To investigate whether the differentially expressed genes might be direct EGR2 targets, we used ChIP-seq data obtained from primary human monocyte extracts via chromatin immunoprecipitation with anti-EGR2 antibodies (27). Peaks were observed close to 168 of the 224 upregulated genes, indicating that these genes were likely directly regulated by EGR2 (P < 0.001; see Supplementary Table S5 and Methods).

To further investigate this point, we used publicly available CLL expression data (28) to identify 24 predicted EGR2 target genes using the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACne; see Methods and Supplementary Table S6). When used as a surrogate marker of EGR2 transcriptional activity, this signature showed transcriptional activity in EGR2-mutated samples (Fig. 5B). Together, these data confirm that expression of mutated EGR2 proteins interferes with the expression of EGR2 target genes in vivo.

Because EGR2 is downstream of normal BCR signaling, we next determined a BCR signaling signature. For this purpose,
we defined a set of genes upregulated upon BCR stimulation of normal B cells, using available data (29, 30), and used Gene Set Enrichment Analyses (GSEA) to show that this signature is enriched in EGR2-mutated samples, with respect to non-mutated samples (Fig. 5C and D). Reciprocally, the EGR2-E356K signature was markedly enriched in BCR-stimulated samples, when compared with unstimulated B cells (Fig. 5E and F), further establishing the deregulation of intracellular BCR signaling in EGR2-mutated samples.

**DISCUSSION**

Here, we identified acquired mutations in the hematopoietic progenitors of patients with CLL and provided proof-of-principle for the role of these mutations during the natural history of the disease.

Our data identified early-mutated genes in patients with CLL. The high mutation burden observed in some patients, in the progenitor and/or mature myeloid fractions, underscores the notion that the identified mutations are functionally relevant and lead to the accumulation of mutated cells, in the progenitor and/or mature fractions. Some of those early drivers are well-known CLL oncogenes (i.e., NOTCH1, XPO1, and SF3B1). We also identified recurrent inactivating mutations of the NFKBIE gene in 10% of patients and as an early event in 1 patient. NFKBIE encodes an inhibitor of NF-kB activity with a specific role in B-lymphocyte biology (31–33). In addition, we showed that acquired missense
mutations of the EGR2 transcription factor are associated with a negative prognostic impact on patient outcome and occurred as an early event in 2 patients with CLL. Our functional data and global expression analyses also point at a common functional consequence of several mutations found in human CLL. EGR2 mutations alter the transcriptional activity of the protein to different extents depending on the mutation. A similar variability has also been observed for the EGR2 mutants observed in congenital neuropathies (34, 35).

EGR2 is a downstream target of the BCR and pre-BCR complexes, through an intracellular signaling cascade involving BRAF, ERK, ELK–SRF, and finally upregulation of EGR2 transcription (19). EGR2 plays an important role in the fine-tuning of early B-cell differentiation (19, 24, 36, 37). Expression of a CLL BRAF mutant in murine progenitors induced abnormal B-cell maturation in mice, including low expression of IgM, a feature of human CLL. Abnormal BCR signaling and EGR2 deregulation are observed in CLL (7, 30, 38), and our observations provide a molecular basis for these observations. We have not been able to investigate the involvement of the progenitor fractions in our series of patients. In a different patient with CLL relapsing from allograft treatment, we have detected an acquired SF3B1 mutation in the lymphoid-primed multipotent progenitor fraction (defined by expression of CD34+/CD38–/CD45RA–; ref. 39; Supplementary Fig. S3B). Together with a previous report of differentiation bias of CLL progenitor cells in xenograft experiments (9), our results suggest that abnormalities in hematopoietic progenitors and early B-cell differentiation are an early step during CLL pathogenesis. They also support the hypothesis that early CLL mutations, despite their diversity, show a convergent phenotype through the impairment of B-cell differentiation upon deregulation of (pre-)BCR signaling. CLL would then develop from progenitors undergoing aberrant B-cell differentiation.

Finally, the diverse early CLL mutations may all induce a pre-leukemic stage devoid of overt clinical signs, conceptually similar to the one proposed for acute leukemia or observed in chronic myeloid neoplasms (1, 2). These observations may therefore have an impact on the follow-up and treatment of patients with CLL. It will therefore be important to understand how these findings relate to the clinical evolution of the patients and to what extent they also apply to other mature lymphoid malignancies (40–43).

**METHODS**

Patient samples were provided by the tumor bank at Pitié-Salpêtrière Hospital (Paris, France), and the study was performed under the supervision of Institutional Review Boards of the participating institutions. Samples were chosen on the basis of the availability of sufficient viable cells. Patients gave informed consent according to the declaration of Helsinki and most of them were enrolled in a clinical trial (www.clinicaltrials.gov; NCT00931645; ref. 17).

**Statistical Analysis**

Clinical and laboratory variables were compared across patients with or without mutation using the Wilcoxon rank-sum test (for quantitative variables) or the Fisher exact test (for qualitative variables). Time to treatment was defined as time between diagnosis and first treatment and compared across groups using the Wilcoxon rank-sum test.
Figure 5. Expression analyses of 15 patients with CLL. A, heatmap representation of the clustering of patients with CLL based on the 239 genes differentially expressed in EGR2-mutated samples. EGR2-mutated patients are indicated by asterisks. B, the expression of genes predicted as target genes of EGR2 in CLL was used as a surrogate marker of EGR2 transcriptional activity and indicated a higher activity of the transcription factor in EGR2-mutated samples. C, class comparison identified a list of genes upregulated upon BCR stimulation of normal human B lymphocytes, which was then used to compare transcription of EGR2-mutated and EGR2-nonmutated CLL samples using GSEA. Shown is the enrichment plot (ES, 0.7718; normalized ES (NES), 1.7360; FDR q value, 0.0049). D, heatmap of the expression of the leading-edge genes (from analyses shown in C) is shown. E, the gene list differentially expressed in EGR2-E356K samples shows enrichment in BCR-stimulated normal human B cells, versus unstimulated samples (ES, 0.6971; NES, 1.8360; FDR q value, 0.0010). F, heatmap of the expression of the leading-edge genes (from analyses shown in E).
rank-sum test. Overall survival was defined as survival since study enrollment; a Kaplan–Meier estimator was used and survival curves were compared using the log-rank test. All tests were two-sided, with \( P \) value less than 0.05 considered as statistically significant. The SAS 9.3 (SAS, Inc.) and R 3.0.2 (R Development Core Team, 2006) software packages were used.

**Exome Sequence Analyses**

We used sorted CD19+ tumor cells (and CD5+ when appropriate) and nontumor (CD3+) cells to extract DNA for exome capture with the SureSelect V4 Mb All Exon Kit (Agilent Technologies) following the standard protocols. We performed paired-end sequencing (100 bp) using HiSeq2000 sequencing instruments at IGR or University of Tokyo. We mapped reads to the reference genome hg19 using the Burrows–Wheeler Aligner (BWA) alignment tool version 0.5.9. PCR duplicates were removed using SAMtools (0.1.18). The detection of candidate somatic mutations was performed according to the previously described algorithms with minor modifications (44). Briefly, the number of the reads containing single-nucleotide variations (SNV) and indels in both tumor and reference samples was enumerated using SAMTools, and the null hypothesis of equal allele frequencies between tumor and reference was tested using the two-tailed Fisher exact test. For candidate somatic mutations, those variants were adopted as candidate mutations whose \( P \) value was \( <0.01 \) and allele frequency was \( <0.1 \) in the reference sample. Finally, the list of candidate somatic mutations was generated by excluding synonymous SNVs and other variants registered in either dbSNP131 (http://www.ncbi.nlm.nih.gov/projects/SNP/) or an in-house SNP database constructed from 180 individual samples (Genomon-exome: http://genomon.hgc.jp/exome/en/index.html) as previously described (44).

**RNA Sequencing, Mapping, and Identification of Differentially Expressed Genes**

RNA was extracted from flow-sorted CD19+ fraction using Qiagen columns, based on material availability. The cDNA libraries were prepared using the ScriptSeq Complete Kit (Epigenetec). We performed paired-end sequencing as described for exome analysis. We removed ribosomal RNA reads (average 2, 11% of total reads) using alignment to the GenBank database. We removed low-quality bases and adapters using Trimomatic version 0.32. The remaining paired reads were mapped to the human reference genome hg19 using Tophat aligner version 2.0.9. The mapped reads were sorted according to their name using SAMTools version 0.1.18. We used the HTSeq python library version 0.5.4p5 to count the number of reads per gene based on the gtf annotation file from the UCSC browser (hg19; ref. 45). Genes with no count in all the samples were discarded and technical replicates were summed. Read numbers and normalization were performed using DESeq version 1.14.0 in the R environment version 3.0.2. To test for differential expression between EGR2 WT (10 samples) and EGR2-E356K (four samples), we used the R package DESeq with negative binomial distribution and a shrinkage estimator for ples) and EGR2-E356K (four samples), we used the R package DESeq version 1.14.0 in the R environment version 4.5. Genes with no count in all the samples were discarded and technical replicates were summed. Read numbers and normalization were performed using DESeq version 1.14.0 in the R environment version 3.0.2. To test for differential expression between EGR2 WT (10 samples) and EGR2-E356K (four samples), we used the R package DESeq with negative binomial distribution and a shrinkage estimator for the distribution's variance. \( P \) values (adjusted by the Benjamini and Hochberg procedure) lower than 1 \times 10^{-2} and fold changes higher than 2 were considered significant. Genes located on sex chromosomes were not considered.

**GSEA Analysis**

The CEL files of the GSE39411 (30) and GSE22762 (28) sets have been normalized with a Robust Multi-Array Average (RMA) procedure. A list of 63 genes was obtained from normalized GSE39411 by a Class Comparison at a \( P \) value of 0.001 with BrB Array Tools (http://linux.nci.nih.gov/BRB-ArrayTools.html) by comparing IgM-stimulated and unstimulated normal B cells at 90 minutes.

A first GSEA analysis was performed by comparing this signature with the log_{2} expression of RNA-seq data of patients with CLL with and without an EGR2-E356K mutation. Reciprocally, a second GSEA analysis was performed by comparing the 239 genes signature obtained by differential expression of genes (Supplementary Table S5) between samples with and without an EGR2-E356K mutation with the log_{2} expression of IgM-stimulated and unstimulated normal B cells at 90 minutes.

**EGR2 Activity Level**

EGR2 targets were predicted using the reverse-engineering algorithm ARACNe (adaptive partitioning, 100 bootstraps; \( P < 1 \times 10^{-9} \); ref. 46) using CLL expression profiles from GSE22762 (28). EGR2 targets were used to compute the activity of the transcription factor across samples. For this purpose, we first defined activated and repressed targets of EGR2 using the Spearman correlation sign between EGR2 and each target using the GSE22762 dataset. The RNA-seq–CLL gene expression profiles were centered and scaled so as to define a comparable rank of expression of each gene across samples. Then, for each independent sample, we computed the activity level of EGR2 defined as the enrichment score (ES), as defined in GSEA (47), computed with EGR2 targets as the gene set and the ranked list of genes in the sample as the reference set. EGR2 activity will be high when EGR2-activated and EGR2-repressed targets are respectively among the most and the least expressed across samples. This will be reflected as a high ES, here computed as the subtraction of the ES of the activated and the ES of the repressed targets.

Peaks identified from an EGR2 ChIP-seq experiment on human monocytes [Gene Expression Omnibus (GEO) accession GSM785503; ref. 27] were associated with neighbor transcripts (corresponding to 9,651 genes) and were obtained by annotation with the coordinates at \(-5\) kb around the transcription start site. Assuming a normal distribution of the peaks (16,558 total peaks), 1,000 tests sampling 224 genes within the 24,910 genes known in hg19 result in a distribution with an average of 80.6 ± 6.85. A deviation from the average of 12.4 leads to a probability of \( P = 9.86 \times 10^{-10} \) to identify 165 genes among the 9,596 genes detected in the ChIP-seq experiment.

**Mutational Analyses in 168 CLL Patients**

Genomic DNA was extracted from peripheral blood mononuclear cells collected at the time of study enrollment using the DNA/RNA Kit (Qiagen) and amplified using the REPLI-G Kit (Qiagen). Genomic regions of **BRAF** (exons 11, 12, and 15), **EGFR** (total coding sequence), **MYD88** (exons 1 and 2), **MYD88** (exons 4 and 5), **XPO1** (exons 1 and 2), **NOTCH1** (partially exon 34), **SF3B1** (exons 13–16), **TP53** (exons 4–10), and **XPO1** (exons 14 and 15) were amplified using intron-flanking primers tagged with M13 universal primers at the 3' or 5' ends. All abnormalities were validated on nonamplified DNA. The list of used primers can be provided upon request. Statistical analyses comparing patients' baseline characteristics, such as age, gender, Binet stage, blood counts, and cytogenetics analysis, have been performed as previously described (48).

**Flow Cytometry and Cell Sorting or Cloning**

Peripheral blood samples were stained with FITC anti-CD3, allophycocyanin (APC) anti-CD14, PerCP-Cy5.5 anti-CD5, PE-Cy7 anti-CD19, and phycoerythrin (PE) anti-CD34, all from BD Pharmgen, Inc. For patients with sufficient available material, additional fractions using FITC anti-CD56, PE anti-igk, and APC anti-igλ were collected. A representative flow chart of the sorting procedure is shown in Supplementary Fig. S1. CD34+ cells were sorted as CD34+CD19+ and were then cloned at 1 cell per well in 96-well plates (Supplementary Fig. S1). Single-cell culture of CD34+ clones was performed as described (41) for 10 to 12 days in MEM-x milieu (Life Technologies).
supplemented with 10% fetal bovine serum (FBS; STEMCELL Technologies, Inc.) and recombinant human cytokines: stem cell factor (SCF; 50 ng/ml); FLT3-Ligand (50 ng/ml); pegylated thrombopoietin (TPO; 10 ng/ml); IL3 (10 ng/ml); IL6 (10 ng/ml); granulocyte macrophage colony-stimulating factor (GM-CSF; 5 ng/ml); erythropoietin (EPO; 1 IU/ml); and G-CSF (10 ng/ml). All cytokines from Peprotech, Inc.

Targeted Resequencing and Mutation Validation

Sorted cell fractions were subjected to DNA/RNA extraction using the AllPrep DNA/RNA Kit (Qiagen) according to the manufacturer’s recommendations. We designed primers flanking exons containing candidate somatic variants using Primer3 (http://frodo.wi.mit.edu/primer3/). Short fragments of 100 to 200 bp were PCR-amplified from genomic DNA of sorted fractions and were subsequently pooled for library construction using the Ion Xpress Plus Fragment Library Kit (Life Technologies). Template preparation was performed with the OneTouch System v37 (Life Technologies). Bar-coded libraries were run on a 1-Gb chip on an Ion PGM Sequencer (Life Technologies). Analysis of acquired data was performed with the Ion Torrent v2.2 software (Life Technologies). Only high-quality reads with a phred score ≥20 were included for further analysis. At least 250 reads were obtained per PCR fragment.

Colony Genotyping

DNA from CD34+ colonies was prepared as described previously (49). Mutational status and VJ rearrangement were analyzed by Sanger sequencing. The complete list of primers will be provided upon request.

Cellular Methods

The IL3-dependent Ba/F3 cell line (from the American Type Culture Collection) is a kind gift from P. Dubreuil (INSERM U1068, Marseille, France); the SCF-dependent cell line EML is a kind gift from G. Mouchiroud (CNRS U5534, Lyon, France). Cells were repeatedly tested for their growth factor dependency and checked to be of murine origin by FACS. EML cells were grown in Iscove’s Modified Dulbecco’s Medium (IMDM), 20% horse serum, and 1% penicillin/streptomycin/glutamine, and supplemented with 10% of BHK cells supernatant. BaF3 cells were grown in RPMI medium, 10% bovine serum, and 1% penicillin/streptomycin/glutamine, and supplemented with 10 ng/ml of IL3. Retroviruses were produced and transduction was performed as described previously (50).

Growth Curve

Twelve hours after transduction, cells were washed and seeded at 5 × 10^5 cells per well. Cells were counted and analyzed by flow cytometry every 2 days. PE-conjugated antibodies were Gr1 (RB-8C5) and B220 (RA3-6B2) from eBioscience and Kit CD117 (2B8) from BD Pharmingen. Experiments were done at least twice in triplicate.

EMSA

The cDNA portion of EGR2 encoding zinc-finger domain (AA 1-2) was amplified by PCR and cloned into PGEX vector (GE Healthcare Life Sciences). Protein production was induced by IPTG stimulation, and the fusion proteins were purified using Glutathione–Sepharose beads, and eluted from the beads with reduced glutathione following the manufacturer’s instructions. SDS-PAGE gel migration followed by Coomassie blue staining and image scanning was used for qualitative and quantitative assessment.

Double-stranded probes were prepared by annealing complementarily oligonucleotides harboring one EGR2-consensus binding site. To generate low-affinity and non-binding sites, base changes were introduced in the core sequence (bold case) of the EGR2 consensus site (underlined) of the strong binding probe 5′-CTCTGT TACGGGCGGTTA-3′. Nonspecific competitor was 5′-CTCTGT TACCCCGGGCGTTA-3′ (26). The LightShift Chemiluminescent EMSA Kit (Thermo Scientific; cat. no. 20148) was used to detect DNA–protein complexes, following the instructions of the manufacturer. Briefly, 2 μL (–2 μg) of purified GST-EGR2 protein extracts were incubated with 50 fmol of double-stranded biotinylated probes in Binding Buffer supplemented with 50 mmol/L KCl, 10 mmol/L MgCl2, 1 mmol/L EDTA, 1 mmol/L DTT, and 1 μg poly dIdC for 10 minutes at room temperature. For competitive assays, a 200× excess of double-stranded nonlabeled probes was added to the mixture.

Binding reactions were loaded in 5% nondenaturing polyacrylamide gels and electrophoresed in 0.5× TBE buffer at 200 V for 30 minutes. DNA and protein complexes were transferred to HyBond N+ membranes (Amersham) in 0.5× TBE buffer at 300 mA for 30 minutes. After UV cross-linking, the membranes were blocked, hybridized with streptavidin–horseradish peroxidase (HRP) conjugated, and revealed following the manufacturer’s instructions. Images were recorded using an ImageQuant detector (GE Healthcare Life Sciences).

Western Blot and Expression Analysis

Two days after transduction, GFP+ cells were flow-sorted and the RNA and protein were extracted using the RNA/DNA/Protein Purification Plus Kit (47700; Norgen Biotek Corp.). Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes. Anti-EGR2 (P100880, Aviva Systems Biology) and anti-ACTINE (A3853, Sigma), Phospho-p44/42 MAPK and p44/42 MAPK antibodies (Cell Signaling Technology), and Raf-B (C-19; Santa Cruz Biotechnology) were used as primary antibodies. Secondary HRP-conjugated antibodies [anti-rabbit IgG (NA934V, GE) and anti-mouse IgG (NA931V, GE)], and ECL Plus Kit (RPN2132, GE) were used for detection.

The following TaqMan probes were purchased from Applied Biosystems: Ab1l: Mm00802038_g1, Gadd45b Mm00435121_g1, Csf1l Mm00432686_m1, Ccl1 Mm00441236_m1, Gapdh Mm999999_g1, Gusb Mm00446956_m1, Egrl Mm0065672_m1, Dtxl Mm00492297_m1, and EGR2 Mm00456650_m1.

Retroviruses

All cDNAs (EGR2: NM_001136177; BRAF: NM_004333) were subcloned into MSCV-GFP backbone. Mutations were introduced using the Quick Change Kit, following the manufacturer’s instructions. Every PCR-amplified or mutagenized fragment was checked by sequencing. Viral particles and transduction procedures were as described previously (50).

Bone marrow transplantation assays and hematopoietic differentiation analyses were performed as described previously (41), except that the mice were analyzed 5 weeks after transplantation. Antibodies used for analyzing B-cell differentiation are anti-mouse CD45.2 V450 (BD Horizon); anti-mouse CD19 APC-eFluor 780, anti-mouse CD43 PE, and anti-mouse IgM PerCP-eFluor 710 (eBioscience); and anti-mouse CD45R/B220 PE-Cy7 and anti-mouse IgD APC (BD Pharmingen).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: F. Damm, Y. Kikushige, K. Akashi, F. Nguyen-Khac, O.A. Bernard

Development of methodology: F. Damm, V. Della Valle, W. Vainchenker, T. Mercher, N. Dronin, S. Ogawa

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.) F. Damm, E. Mylonas, K. Yoshida,


Writing, review, and/or revision of the manuscript: F. Damm, E. Mylonas, Y. Kikutusuge, P. Dessen, E. Solary, K. Akashi, F. Nguyen-Khac, O.A. Bernard

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): V. Della Valle, H. Merle-Béral, L. Sutton, E. Solary

Study supervision: O.A. Bernard

Acknowledgments

The authors thank the present and past IGR platforms team members for skillful help; Sylvie Chevet, Julie Lejeune (SBIM, St-Louis, Paris, France), and Claude Lesty (Pitié-Salpetrière Hospital) for help in statistical analyses; F. Norol and H. Trebeden-Negre (Pitié-Salpetrière Hospital) for help with patient material; K. Maloum, D. Roos-Weil, O. Tournilhac, and L. Veronese for patient material and biologic data; and Patrick Charnay and Pascale Gilardi for helpful discussions.

Grant Support

This work was funded by grants from INSERM, Institut National du Cancer (INCa), Ligue Nationale Contre le Cancer (LNCC; équipe labélisée à E. Solary and O.A. Bernard), INCa-DGOS-INSERM (6043), Fondation Gustave Roussy, KAKENHI (23240052 and 22134006), and the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science. L. Scourzic is the recipient of a fellowship from the Région Ile de France. F. Damm was the recipient of a fellowship from the Région Ile de France. O. A. Bernard was the recipient of a fellowship from the Japanese Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 31, 2014; revised May 20, 2014; accepted June 2, 2014; published OnlineFirst June 11, 2014.

REFERENCES

Early Driver Mutations in CLL

Acquired Initiating Mutations in Early Hematopoietic Cells of CLL Patients

Frederik Damm, Elena Mylonas, Adrien Cosson, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-14-0104

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2014/06/10/2159-8290.CD-14-0104.DC1

Cited articles
This article cites 50 articles, 21 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/4/9/1088.full#ref-list-1

Citing articles
This article has been cited by 20 HighWire-hosted articles. Access the articles at:
http://cancerdiscovery.aacrjournals.org/content/4/9/1088.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.