Author Manuscript Published OnlineFirst on June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Acquired initiating mutations in early hematopoietic cells of CLL patients

Frederik Damm1,2*, Elena Mylonas1,2*, Adrien Cosson 3,4, Kenichi Yoshida 5, Véronique Della Valle 1,2,7, Enguerran Mouly 1,2,7, M'boyba Diop1,2, Laurianne Scourzic 1,2,7, Yuichi Shiraishi 8, Kenichi Chiba 8, Hiroko Tanaka 8, Satoru Miyano 8, Yoshikane Kikushige 9,10, Frederick Davi 3,4,11, Jérôme Lambert 12, Daniel Gautheret 13, Hélène Merle-Béral 3,4,11, Laurent Sutton 14, Philippe Dessen 1,2, Eric Solary2,7,15,18 Koichi Akashi 9,10, William Vainchenker 2,7,15, Thomas Mercher 1,2,7, Nathalie Droin 2,7,15, Seishi Ogawa 5,6#, Florence Nguyen-Khac 3,4,11#, Olivier A. Bernard 1,2,7,18#

1 INSERM U985, Villejuif 94805, France

2 Institut Gustave Roussy, Villejuif 94805, France

3 INSERM U1138, Paris 75006, France

4 Université Pierre et Marie Curie-Paris 6, Paris 75005, France

5 Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655 Japan

6 Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan

7 Université Paris-Sud, Orsay 91400, France

8 Laboratory of DNA Information Analysis; Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

9 Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka 812-8582, Japan

10 Center for Cellular and Molecular Medicine, Kyushu University Graduate School of Medical Sciences, Fukuoka 812-8582, Japan

11 Service d'Hématologie Biologique, Hôpital Pitié -Salpêtrière, APHP, Paris 75013, France

12 Université Paris Diderot, Hôpital Saint-Louis, SBIM, Paris, France

13 Université Paris-Sud, Institut de Génétique et Microbiologie, CNRS UMR 8621, Orsay F-91405, France; and Institut de recherche intégrée en Cancérologie à Villejuif, Institut Gustave Roussy, 94805 Villejuif, France

14 Service d'Hématologie Clinique, Centre Hospitalier Victor Dupouy, Argenteuil, France

15 INSERM, U1009, Villejuif 94805, France

18 Ligue Nationale Contre le Cancer, Equipe labellisées (OAB and ES)

* These authors contributed equally to this work

# These authors share senior authorship.

Correspondence:

Seishi Ogawa mail : sogawa-tky@umin.ac.jp

Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan

Florence Nguyen-Khac mail : florence.nguyen-khac@psl.aphp.fr

Unité de Cytogénétique Hématologique, Service d'Hématologie Biologique, GH Pitié-Salpêtrière/Charles Foix, 83 Bd de l'Hôpital, 75013 Paris

Tel : 33 1 42162451 / 2460

Fax : 33 1 42162453

Olivier A. Bernard mail: olivier.bernard@inserm.fr

INSERM U985, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif France

Tel: 33 (0) 1 42 11 42 33

Fax: 33 (0) 142 11 51 01

The authors have no conflict of interest to disclose.

Word count:

Figures/tables: 5

Abstract (147/150 words)

Appropriate cancer care requires thorough understanding of the natural history of the disease, including the cell of origin, the clonal organization of the proliferation and the functional consequences of the mutations. Using deep sequencing of flow-sorted cell populations from chronic lymphocytic leukemia (CLL) patients, we established the presence of acquired mutations in multipotent hematopoietic progenitors of CLL patients. Mutations affected known lymphoid oncogenes including *BRAF, NOTCH1* and *SF3B1*. *NFKBIE* and *EGR2* mutations were observed at unexpected high frequencies, 10.7% and 8.3% of 168 advanced stage patients, respectively. *EGR2* mutations were associated with a shorter time to treatment and poor overall survival. Analyses of BRAF and EGR2 mutation suggest they affect deregulation of BCR intracellular signaling. Our data propose deregulation of hematopoietic and early B-cell differentiation through the deregulation of pre-BCR signaling as a phenotypic convergence of CLL mutations and show that CLL develops from a pre-leukemic phase.

Statement of Significance: The origin and pathogenic mechanisms of chronic lymphoblastic leukemia (CLL) are not fully understood. The current work indicates that CLL develops from pre-leukemic multipotent hematopoietic progenitors carrying somatic mutations. It advocates for abnormalities in early B-cell differentiation as a phenotypic convergence of the diverse acquired mutations observed in CLL.

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 pre-existing 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 an accumulation of mature B-lymphocytes (3). The CLL tumor cells are clonal, based on the immunoglobulin heavy chain (IGH) gene rearrangement and express low levels of surface B-cell receptor. In a fraction of the patients, the IGHV rearrangement is mutated, reflecting normal somatic hyper mutation triggered by antigen recognition. The IGHV-mutated patients have a better prognosis than IGHV-non-mutated ones.

Investigation of CLL samples by massively parallel sequencing has identified numbers of acquired somatic mutations (4, 5), but no individual gene is mutated in more than 20% of the patients. The products of the 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 B-cell receptor intracellular signaling cascades (6, 7). The cell of origin of CLL is currently debated. Immunophenotype and expression profiles analyses pointed at mature $CD5^+$ B-cells (8), but the involvement of early hematopoietic cells in CLL development has been revitalized following xenograft

experiments: the hematopoietic stem/progenitor cells from CLL patients show biased and abnormal differentiation toward the B-lymphoid lineage in immunodeficient mice. (9)

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

Author Manuscript Published OnlineFirst on June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

RESULTS

SF3B1 mutations are detected in non-lymphocytic 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 *SF3B1*-mutated CLL patients since 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 of the *SF3B1* gene on DNA from 50 CLL patients identified 7 patients carrying a *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, CD3 (T-cell), 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 $CD3^+$ and the mutated sequence in the CD19 fraction, in all seven cases. The mutation was also observed in $CD34^+$ and/or $CD14^+$ fractions in two of them (Supplementary Fig. S1), suggesting it was acquired, in these patients, in an early progenitor, 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 of 24 CLL patients (17 IGHV-unmutated and 7 *IGHV*-mutated, Supplementary Table S1). Results of the immunoglobulin heavy chain gene rearrangements were always compatible with a monoclonal proliferation. Viable cells

were flow-sorted to purities greater than 96% (see flow chart description Supplementary Fig. S2A-B). 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 of 361 different genes with a median of 17 mutations/patient (range: 7 – 34; Supplementary Table S1 and Fig. 1A). Some mutations were present in virtually all the CD19-positive cells, whereas the allelic ratio of other mutations indicated that they were present only in a fraction of them, indicating 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 have to be taken into account when analyzing cell-sorted fractions. Mutation burdens below 4% potentially due to sorting contamination are 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 IGH rearrangements in their CLL cells. In the other 21 patients, at least one mutation was detectable in the $CD14^+$ or in the $CD34^+$ fractions. Two patients (CLL14, 27) showed mutations in the $CD14^+$ and not in the $CD34^+$ fraction and conversely, six patients (CLL10, 11, 16, 17, 18, 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, 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 (called hereafter early

mutations) is 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 sub-set of the CLL mutations was observed in the progenitors or myeloid fractions.

Because cell surface expression of myeloid antigens is not sufficient to attest 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: CLL08 and CLL09) and colonies were confirmed as myeloid (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, 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, 26). A low number or IGH-rearranged colonies were 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 VJ junction amplified from the colonies matched the patient's tumor cell

rearrangement in patients 05 and 07. Colonies from patient 17 and 19 showed a rearrangement differing from those from their CLL counterpart. Half of the colonies from patients 18 and 27 showed the tumor cells rearrangement whereas the other half carried other VJ junctions. Of note, every VJ-positive colony also carries an early mutation.

Together, these data demonstrate the presence of CLL-mutations in a multipotent hematopoietic progenitor fraction of the majority of CLL patients. Reasoning that the mutations have 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 appear to variably affect hematopoietic differentiation, as judged from the mutation burden detected in the hematopoietic fractions (see supplemental Fig. S3A).

Some patients show an overall normal balance between myeloid and B-lymphoid differentiation. They show multi-lineage 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 will be detected only if it induces the accumulation of the mutated cell in the given fraction. If a mutation induces accumulation at a late and not at early steps of differentiation, accumulation will occur in the mature cells (CD14+) and not the immature cells (CD34+). For example patient CLL14 and CLL 27 are in this situation and would belong in this first group of patients.

A second type of patients (for example CLL10, 11, 22) shows an unbalanced involvement of myeloid cells (a lower mutational burden than in $CD34^+$-progenitors), suggesting that the early mutations biase the mutated stem/progenitor cells toward

the lymphoid lineage or specifically allow the accumulation of lymphoid-primed progenitors.

A third type of patients (CLL03, CLL015, CLL024 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, this third type of patients may follow a different transformation pathway. The numbers and burden of mutations did not differ statistically between these three patients and the others (Supplementary Table S1).

The early mutations affect genes recurrently mutated in CLL and other malignancies

Mutations detected in the progenitors of CLL patients affected genes already known to be mutated in CLL, other hematological 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 that belong to the most frequent mutated ones in CLL (4, 5, 12). Genes such as *BRAF* or *MLL2* 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 stage B and C CLL untreated patients sampled at inclusion in a clinical trial (www.clinicaltrials.gov: NCT00931645 and Supplementary Table S3) (17). A total of 113 mutations in 84 patients were identified and 84/168 (50.0%) patients presented at least one mutation of this 9 genes panel (Fig. 2A and B and Supplementary Table

S4). Inactivating mutations of *NFKBIE*, were found in 10.7% (18/168, Fig. 2C) of the patients. Missense mutations of *EGR2* were observed in 8.3% (14/168, 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 months *vs* 1.2 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-SRF dimer resulting in the up regulation of a set of immediate early genes, including *EGR2* (19, 20). *BRAF* and *EGR2* mutations may therefore impact on 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 rather target amino acids located in the P-loop of the kinase (Supplementary Table S4 and (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-WT, BRAF-G469R or empty MSCV vector and engrafted the cells in irradiated syngenic recipients. Animals were analyzed 5 weeks after, before the onset of any gross hematological disorders. Careful analyses of the IgM-positive B-cell compartment showed a decrease in the proportion of B-cells in the BRAF-G469R

mice, when compared to MSCV or BRAF-WT mice (Fig. 3 C and D). In addition, the mean fluorescence of IgM was significantly lower in B220-IgM positive BRAF-G469R-expressing cells than their wild type or MSCV counterparts (Fig. 3 E). A similar abnormal (IgM low,IgD-) B-cell population is present in the spleen of the BRAF-G469R mice (Fig.3 C and E).

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- 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 two 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 wildtype or two EGR2 mutants (E356K and H384N). Electromobility shift analyses 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 appears weaker than wildtype despite comparable protein amounts (Fig. 4B). The interaction of the E356K protein with the probe was not observed in this assay. To investigate their transcriptional regulation properties, we expressed the wild type 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 associated with a slower growth (Fig. 4E). Cells expressing wildtype EGR2 showed a progressive reduction in the expression of the cell surface markers B220 (B-lymphoid) and Gr1 (myeloid) (Fig. 4D). Growth slow-down and loss of $B220^{+}$ and

$Gr1^+$ cells was 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 Fig. 4F, wild type 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 transcription of EGR2-target genes. Whereas the differential activities of the mutants is due to differences in EGR2 binding site or interaction with transcription co-factors at the target genes will require additional investigations.

To investigate the functional consequences of EGR2 mutations in patient samples, we analyzed RNA-seq data obtained from 16 CLL samples. Fifteen genes were down regulated, whereas 224 genes were specifically up regulated in *EGR*2-E356K samples (n=4) as compared to *EGR*2 wild type or unanalyzed patients (n=10; p<0,01; Supplementary Table S5). Hierarchical clustering using the 224 up regulated 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 mutation 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 using anti-EGR2 antibodies-chromatin immunoprecipitation (27) from primary human monocytes extracts. Peaks

were observed close to 168 out of the 224 up regulated genes, indicating that theses genes were likely directly regulated by EGR2 ($p<0.001$; Supplementary Table S5 and material & method section). To further investigate this point, we used publicly available CLL expression data (28) to identify 24 predicted EGR2-target genes using ARACne (see material and 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.

Since EGR2 is downstream of normal BCR signaling, we next determined a BCR-signaling signature. In that purpose, we defined a set of genes up regulated upon BCR-stimulation of normal B-cells, using available data (29, 30) and used GSEA analyses 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 CLL patients 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 CLL patients. The high mutation burden observed in some patients, in the progenitor and/or mature myeloid fractions underscores that the identified mutations are functionally relevant and lead to 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, SF3B1). We also identified recurrent inactivating mutations of the *NFKBIE* gene in 10% of the patients and as an early event in one patient. *NFKBIE* encodes an inhibitor of NFKB activity, with a specific role in B-lymphocyte biology (31-33). In addition, we show that acquired missense mutations of the *EGR2* transcription factor are associated with a negative prognostic impact on patients' outcome and occur as an early event in two CLL patients.

Our functional data and global expression analyses also point at a common functional consequences of several mutations found in human CLL. EGR2 mutations alter transcriptional regulation activity of the protein, differently between mutants. 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 *EGR2* transcription upregulation (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 mouse, 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 ground for these observations. We have not been able to investigate the involvement of the progenitor fractions in our series of patient. In a different patient, relapsing from allograft treatment, we have detected an acquired SF3B1 mutation in the lymphoid primed multipotent progenitor (defined by expression of CD34+/CD38-/CD45RA+/CD90- (39)) fraction of a patient suffering from relapsed CLL (Supplementary Fig. S3B). Together with the 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 is 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-Salpétrière 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 signed informed consent, according to the declaration of Helsinki and most of them were enrolled in a clinical trial (www.clinicaltrials.gov: NCT00931645).(17)

Statistical analysis

Clinical and laboratory variables were compared across patients with or without mutation using Wilcoxon rank sum test (for quantitative variables) or Fisher's 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. Overall survival was defined as survival since study enrollment; 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., NC, USA) and R 3.0.2 (R Development Core Team, 2006) software packages were used.

Exome sequence analyses

We used sorted tumor CD19+ cells (and CD5+ when appropriate) and non-tumor (CD3+) cells to extract DNA for exome capture with the SureSelect V4 Mb All Exon kit (Agilent Technologies) following standard protocols. We performed paired-end sequencing (100 bp) using HiSeq2000 sequencing instruments at IGR or 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 (SNVs) 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's 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 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 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 (Epicentre). 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 Trimmomatic 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 use the HTSeq python library version 0.5.4p5 to count the number of reads per gene using the gtf annotation file from UCSC (hg19) (45). Genes with no count in all

the samples were discarded and technical replicates were summed. Read numbers and normalization was performed using DESeq version 1.14.0 in the R environment version 3.0.2. To test for differential expression between EGR2 wild type (10 samples) and EGR2-E356K (4 samples), we used the R package DESeq with negative binomial distribution and a shrinkage estimator for the distribution's variance. P-values (adjusted by Benjamini and Hochberg procedure) lower than $1x10^{-2}$ 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 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://linus.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 log2 expression of RNA-Seq data of CLL patients 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 with the log2 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 (46) (adaptive partitioning, 100 bootstraps, p<1e-9) using CLL expression profiles from

GSE22762 (28). EGR2 targets were used to compute the activity of the transcription factor across samples. For that, we first defined activated and repressed targets of EGR2 using the Spearman correlation sign between EGR2 and each target using the GSE22762 data set. The RNASeq-CLL gene expression profiles were centred and scaled so that each gene in order to define a comparable rank of expression of each gene across samples. Then for each independent sample, we can compute 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 -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 activated and ES of the repressed targets.

Peaks identified from EGR2- Chip Seq experiment on human monocytes (27) (GEO Accession GSM785503) were associated with neighbor transcripts (corresponding to 9651 genes) was obtained by an annotation with the coordinates at -5/+5kb around the transcription start site. Assuming a normal distribution of the peaks (16558 total peaks), 1000 tests sampling 224 genes within the 24910 genes known in hg19, results in a distribution with an average of 80.6 + / - 6.85. A deviation from the average of 12.4 leads to a probability of $p=9.86\ 10^{-10}$ to identify 165 genes among the 9596 genes detected in the ChipSeq experiment.

Mutational analyses in 168 CLL patients.

Genomic DNA was extracted from peripheral blood mononuclear cells collected at time point 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), *EGR2*

(total coding sequence), *MED12* (exons 1 and 2), *MYD88* (exons 4 and 5), *NFKBIE* (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' prime ends. All abnormalities were validated on non-amplified 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, APC anti-CD14, PerCP-Cy5.5 anti-CD5, PE-Cy7 anti-CD19 and PE anti-CD34 all from BD Pharmigen, Inc. For patients with sufficient available material, additional fractions using FITC antiCD56, PE anti-Igκ and APC anti-Igλ were collected. A representative flow chart of the sorting procedure is shown in Supplementary Fig. S1. $CD34^+$ were sorted as $CD34^+CD19^-$ and were then cloned at one cell per well in 96-well plates (Supplementary Fig. S1). Single-cell culture of $CD34^+$ clones was performed for 10-12 days in MEM-alpha milieu (Life Technologies) supplemented with 10% fetal bovine serum (FBS, StemCell Technologies, Inc.) and recombinant human cytokines: Stem Cell Factor (SCF, 50 ng/mL, Biovitrum AB, Inc.), FLT3-Ligand (50 ng/mL, Celldex Therapeutics, Inc.), pegylated thrombopoietin (TPO, 10 ng/mL, Kirin Laboratories, Inc.), interleukin-3 (IL-3, 10 ng/mL, Miltenyi Biotec), interleukin-6 (IL-6, 10 ng/mL, gift from S. Burstein, Oklahoma City, OK, USA), granulocyte-macrophage colony–stimulating factor (GM-CSF, 5 ng/mL, Peprotech, Inc.), erythropoietin (EPO,

1 IU/mL, Amgen, Inc.) and granulocyte colony–stimulating factor (G-CSF,10 ng/mL, Amgen).(41)

Targeted resequencing and mutation validation.

Sorted cell fractions were subjected to DNA/RNA extraction using the All Prep 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 100-200bp 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 ≥Q20 were included for further analysis. At least 250 reads were obtained per PCR fragment.

Colonies genotyping.

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

Cellular methods

The IL3-dependent Ba/F3 cell line (from ATCC) is a kind gift of P Dubreuil (Marseille, France) ; the SCF dependent cell line EML is a kind gift of Guy Mouchiroud (Lyon, France). Cells were repeatedly tested for their growth factor-dependency and checked to be from murine origin by FACS. EML cells were grown in IMDM medium, 20% horse serum, 1% Penicilin/Streptomycine/Glutamine and supplemented with 10% of BHK cells supernatant. BaF3 cells were growth in RPMI medium, 10% bovine serum, 1% Penicilin/Streptomycine/Glutamine and supplemented with 10 ng/mL Il-3. Retroviruses were produced and transduction performed as described (50).

Growth curve

Twelve hours after transduction cells are washed, and seeded at 5 10^6 cells per well. Cells were counted and analyzed by flow cytometry every 2 days. PE conjugated antibodies were Gr1 (RB6-8C5), B220 (RA3-6B2) from eBiosciences and Kit CD117 (2B8) from BD Pharmigen. Experiments were done at least twice in triplicate.

Electromobility shift assays (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 purified using Glutathione-Sepharose beads, and eluted from the beads with reduced glutathione following the manufacturer 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 complementary oligonucleotides harbouring 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 (under lined) of the strong binding probe 5' CTCTGTAC*GCG**GGG**GCG*GTTA 3'. Non-specific competitor is 5' CTCTGTAC*GCG**CCC**GCG*GTTA 3'.(26). LightShift® Chemiluminescent EMSA Kit (Thermo Scientific Cat 20148) was used to detect DNA-protein complexes, following the instructions of the manufacturer. Briefly, 2 uL (~2ug) of purified GST-EGR2 protein extracts were incubated with 50 fmol of double strand biotinylated probes in Binding Buffer supplemented with 50 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1 ug poly dIdC, during 10 minutes at room temperature. For competitive assays, a 200X excess of double stranded non-labelled probes was added to the mixture.

Binding reactions were loaded in 5% non-denaturing polyacrylamide gels and electrophoresed in 0.5X TBE buffer at 200 V for 30 minutes. DNA and AND-proteins-complexes were transferred to HyBond N+ membranes (Amersham) in 0.5X TBE buffer at 300 mA for 30 min. After UV crosslinking, the membranes were blocked, hybridized with streptavidin-HRP conjugated and revealed following the manufacturer instructions. Images were recorder using an ImageQuant® detector (GE Healthcare Life Sciences).

Western blot and expression analysis

Two days after transduction GFP-positive cells were flow- sorted and the RNA and protein were extracted using the RNA/DNA/Protein purification Plus Kit (47700, Norgen Biotek Corp). Protein 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), Raf-B(C-19), (Santa Cruz Biotechnology) were used as primary antibodies. Secondary HRP conjugated antibodies (anti rabbit IgG (NA934V, GE), anti mouse IgG (NA931V, GE) and ECL Plus Kit (RPN2132, GE) were use for detection.

The following Taqman probes were purchased from Applied Biosystems: Abl1: Mm00802038_g1, Gadd45b Mm00435121_g1, Csf1 Mm00432686_m1, Ccl1 Mm00441236_m1, Gapdh Mm999999_g1, Gusb Mm00446956_m1, Egr1 Mm0065672_m1, Dtx1 Mm00492297_m1, EGR2 Mm00456650_m1.

Retroviruses

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

Bone marrow transplantation assays and hematopoietic differentiation analyses were performed as described (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™); Anti-Mouse CD45R/B220 PE-Cy™7 and Anti-Mouse IgD APC (BD Pharmingen™)

ACKNOWLEDGMENT:

The work was funded by grants from INSERM, Institut National du Cancer (INCa),

Ligue Nationale Contre le Cancer (LNCC; équipe labélisée to ES and OAB), INCa-DGOS-INSERM (6043), Fondation Gustave Roussy, KAKENHI (23249052 and 22134006), the Japan Society for the Promotion of Science through the Funding Program for World-Leading Innovative R&D on Science. We want to thanks the present and past IGR platforms team members for skillful help; Sylvie Chevret, Julie Lejeune (DBIM, St-Louis), Claude Lesty (Pitie-Salpetriere hospital) for help in statistical analyses; F Norol, H Trebeden-Negre (Pitie-Salpetriere hospital) for help with patients material; K Maloum, D Roos-Weil, O Tournilhac, L Veronese, for patients material and biological data and Patrick Charnay and Pascale Gilardi for helpful discussion. LS is the recipient of a fellowship from the Région Ile de France. FD is the recipient of a Deutsche Krebshilfe fellowship (grant 109686).

AUTHOR CONTRIBUTIONS

F D, EM, AC, VDV EM, LS performed research. M D analyzed exome and RNA-seq data under the supervision of DG. PD analyzed microarray data. Y S, K C, H T, S M analyzed exome data. FD, H M-B, L S, W V provided clinical samples and biological data. JL and FD performed statistical analyses. YK, KA, WV, ES, ND and TM provided critical intellectual inputs and reagents. S O, F N-K and OAB designed and supervised research. OAB wrote the manuscript with the help of FD, KA, WV, SO, TM and F N-K. All authors agreed with the manuscript. None of the authors has competing financial interests.

References

1. Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, et al. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Science translational medicine. 2012;4:149ra18.
2. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150:264-78.
3. Zenz T, Mertens D, Kuppers R, Dohner H, Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat Rev Cancer. 2009.
4. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. The New England journal of medicine. 2011;365:2497-506.
5. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature genetics. 2011;44:47-52.
6. Burger JA, Chiorazzi N. B cell receptor signaling in chronic lymphocytic leukemia. Trends Immunol. 2013.
7. Herishanu Y, Perez-Galan P, Liu D, Biancotto A, Pittaluga S, Vire B, et al. The lymph node microenvironment promotes B-cell receptor signaling, NF-kappaB activation, and tumor proliferation in chronic lymphocytic leukemia. Blood. 2011;117:563-74.
8. Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Durig J, et al. Cellular origin and pathophysiology of chronic lymphocytic leukemia. The Journal of experimental medicine. 2012;209:2183-98.
9. Kikushige Y, Ishikawa F, Miyamoto T, Shima T, Urata S, Yoshimoto G, et al. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. Cancer Cell. 2011;20:246-59.
10. Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. Blood. 2012;119:3211-8.
11. Ebert B, Bernard OA. Mutations in RNA splicing machinery in human cancers. N Engl J Med. 2011;365:2534-5.
12. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011;475:101-5.
13. Tiacci E, Trifonov V, Schiavoni G, Holmes A, Kern W, Martelli MP, et al. BRAF mutations in hairy-cell leukemia. The New England journal of medicine. 2011;364:2305-15.
14. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476:298-303.
15. Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. The Journal of experimental medicine. 2012;209:1537-51.

16. Domenech E, Gomez-Lopez G, Gzlez-Pena D, Lopez M, Herreros B, Menezes J, et al. New mutations in chronic lymphocytic leukemia identified by target enrichment and deep sequencing. PloS one. 2012;7:e38158.
17. Sutton L, Chevret S, Tournilhac O, Divine M, Leblond V, Corront B, et al. Autologous stem cell transplantation as a first-line treatment strategy for chronic lymphocytic leukemia: a multicenter, randomized, controlled trial from the SFGM-TC and GFLLC. Blood. 2011;117:6109-19.
18. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell. 2004;116:855-67.
19. Yasuda T, Sanjo H, Pages G, Kawano Y, Karasuyama H, Pouyssegur J, et al. Erk kinases link pre-B cell receptor signaling to transcriptional events required for early B cell expansion. Immunity. 2008;28:499-508.
20. Chavrier P, Zerial M, Lemaire P, Almendral J, Bravo R, Charnay P. A gene encoding a protein with zinc fingers is activated during G0/G1 transition in cultured cells. The EMBO journal. 1988;7:29-35.
21. Dhomen N, Marais R. New insight into BRAF mutations in cancer. Current opinion in genetics & development. 2007;17:31-9.
22. Jebaraj BM, Kienle D, Buhler A, Winkler D, Dohner H, Stilgenbauer S, et al. BRAF mutations in chronic lymphocytic leukemia. Leukemia & lymphoma. 2013;54:1177-82.
23. Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee HJ, Sciammas R, et al. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. Cell. 2006;126:755-66.
24. Li S, Symonds AL, Zhu B, Liu M, Raymond MV, Miao T, et al. Early growth response gene-2 (Egr-2) regulates the development of B and T cells. PloS one. 2011;6:e18498.
25. Zheng Y, Zha Y, Driessens G, Locke F, Gajewski TF. Transcriptional regulator early growth response gene 2 (Egr2) is required for T cell anergy in vitro and in vivo. The Journal of experimental medicine. 2012;209:2157-63.
26. Nardelli J, Gibson T, Charnay P. Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucleic acids research. 1992;20:4137-44.
27. Pham TH, Benner C, Lichtinger M, Schwarzfischer L, Hu Y, Andreesen R, et al. Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. Blood. 2012;119:e161-71.
28. Herold T, Jurinovic V, Mulaw M, Seiler T, Dufour A, Schneider S, et al. Expression analysis of genes located in the minimally deleted regions of 13q14 and 11q22-23 in chronic lymphocytic leukemia-unexpected expression pattern of the RHO GTPase activator ARHGAP20. Genes, chromosomes & cancer. 2011;50:546-58.
29. Vallat LD, Park Y, Li C, Gribben JG. Temporal genetic program following B-cell receptor cross-linking: altered balance between proliferation and death in healthy and malignant B cells. Blood. 2007;109:3989-97.
30. Vallat L, Kemper CA, Jung N, Maumy-Bertrand M, Bertrand F, Meyer N, et al. Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2013;110:459-64.
31. Whiteside ST, Epinat JC, Rice NR, Israel A. I kappa B epsilon, a novel member of the I kappa B family, controls RelA and cRel NF-kappa B activity. The EMBO journal. 1997;16:1413-26.

32. Memet S, Laouini D, Epinat JC, Whiteside ST, Goudeau B, Philpott D, et al. IkappaBepsilon-deficient mice: reduction of one T cell precursor subspecies and enhanced Ig isotype switching and cytokine synthesis. Journal of immunology (Baltimore, Md : 1950). 1999;163:5994-6005.
33. Emmerich F, Theurich S, Hummel M, Haeffker A, Vry MS, Dohner K, et al. Inactivating I kappa B epsilon mutations in Hodgkin/Reed-Sternberg cells. The Journal of pathology. 2003;201:413-20.
34. Musso M, Balestra P, Taroni F, Bellone E, Mandich P. Different consequences of EGR2 mutants on the transactivation of human Cx32 promoter. Neurobiology of disease. 2003;12:89-95.
35. Warner LE, Svaren J, Milbrandt J, Lupski JR. Functional consequences of mutations in the early growth response 2 gene (EGR2) correlate with severity of human myelinopathies. Human molecular genetics. 1999;8:1245-51.
36. Brummer T, Shaw PE, Reth M, Misawa Y. Inducible gene deletion reveals different roles for B-Raf and Raf-1 in B-cell antigen receptor signalling. The EMBO journal. 2002;21:5611-22.
37. Li S, Miao T, Sebastian M, Bhullar P, Ghaffari E, Liu M, et al. The transcription factors egr2 and egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells. Immunity. 2012;37:685-96.
38. Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, Villamor N, et al. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. Genome Res. 2014;24:212-26.
39. Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell. 2011;19:138-52.
40. Couronne L, Bastard C, Bernard OA. TET2 and DNMT3A Mutations in Human T-Cell Lymphoma. The New England journal of medicine. 2012;366:95-6.
41. Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, et al. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. Cancer Cell. 2011;20:25-38.
42. Vicente-Duenas C, Fontan L, Gonzalez-Herrero I, Romero-Camarero I, Segura V, Aznar MA, et al. Expression of MALT1 oncogene in hematopoietic stem/progenitor cells recapitulates the pathogenesis of human lymphoma in mice. Proceedings of the National Academy of Sciences of the United States of America. 2012;109:10534-9.
43. Weigert O, Kopp N, Lane AA, Yoda A, Dahlberg SE, Neuberg D, et al. Molecular ontogeny of donor-derived follicular lymphomas occurring after hematopoietic cell transplantation. Cancer discovery. 2012;2:47-55.
44. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478:64-9.
45. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010;11:R106.
46. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC bioinformatics. 2006;7 Suppl 1:S7.
47. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting

genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102:15545-50.
48. Nguyen-Khac F, Lambert J, Chapiro E, Grelier A, Mould S, Barin C, et al. Chromosomal aberrations and their prognostic value in a series of 174 untreated patients with Waldenstrom's macroglobulinemia. Haematologica. 2013;98:649-54.
49. Dupont S, Masse A, James C, Teyssandier I, Lecluse Y, Larbret F, et al. The JAK2 617V>F mutation triggers erythropoietin hypersensitivity and terminal erythroid amplification in primary cells from patients with polycythemia vera. Blood. 2007;109:71-7.
50. Malinge S, Ragu C, Della-Valle V, Pisani D, Constantinescu SN, Perez C, et al. Activating mutations in human acute megakaryoblastic leukemia. Blood. 2008;112:4220-6.

FIGURE LEGENDS

Figure 1. Somatic mutations identified in 24 CLL patients.
A, Variant allele frequencies in the tumor fraction. Genes analyzed in the extension cohort are colored. ZMYM3, DDX3X, SMARCA1 and HEPH genes are on chromosome X. See Supplementary Table S1 for details. The patient numbers are in abscissa.
B, Variant allele frequencies in the hematopoietic fractions, sorted based on the cell surface expression of the following antigens (in abscissa): CD34+ (progenitors), CD3+ (T-cells), CD19+ and CD5+ (tumor cells), CD14+ (monocytes). The column on the right refers to the genotyping of Colony Forming Units (CFU) from single CD34+ cells. If only wild type colonies were observed, the number of analyzed colonies is indicated in brackets. Ratio of 50% and 4% are indicated by dotted lines, orange and black, respectively. ND: not done.

Figure 2. Gene mutation profile in 168 CLL patients.
A, Frequencies and **B,** Distributions of identified mutations.
C, Localization of identified mutations in EGR2 and NFKBIE proteins. The circles indicate mutations; Filled circle: proven somatic mutation. (D) Overall survival of CLL patients according to *ERG2* mutation status (log-rank test).

Figure 3. **Functional analyses of BRAF-G469R.**
A, Expression of BRAF-G469R from MSCV results in constitutive ERK activation. The right panel shows ERK phosphorylation is detected in IL3-starved Ba/F3 cells expressing mutated BRAF, but not BRAF-WT or empty control. Antibodies are indicated and also evaluation of BRAF expression normalized to B-Actin and BRAF-WT expression.
B, RQ-PCR evaluation of Egr2 expression in transduced Ba/F3 cells, normalized with respect to Gapdh.
C, Representative FACS analysis of bone marrow (left side) and spleen (right side) cells from MSCV, BRAF-WT and BRAF-G469R mice analyzed 5 weeks after engraftment. Plots are gated on donor (CD45.2+) GFP-positive cells and the percentages of gated cells is shown.
D, Mean percentage (±SD n=5 mice) of bone marrow GFP+ donor cells expressing B220 and negative (left panel) or positive (right panel) for membrane IgM.
E, Mean fluorescence intensity (MFI) of IgM in GFP+ donor B-cells, immature (IgM+ IgD-, left panel) and matures (IgM+ IgD+, right panel). Bone marrow (BM) or spleen origins are indicated.
(p values: ****<0.0001, 0.0001<***<0.001, 0.001<**<0.01, 0.01<*<0.05, ns >0.05)

Figure 4. Functional analyses of EGR2 mutants.
A, The Zinc finger regions of mutant EGR2 are less efficient than wildtype at binding EGR2 high affinity sequences (CTCTGTACGCGGGGGCGGTTA) in electrophoretic mobility shift assays. Shifted complexes (indicated by an arrow) are observed in lanes containing GST-EGR2-WT protein (lanes 3 and 5). Formation of this complex is inhibited in the presence of a 200-time molar excess of unlabeled probe, used as

specific competitor (SC: lanes 4, 7, 10), but not in the presence of a same excess of a non-specific competitor (NSC: lanes 5, 8, 11). Shifted complexes are also detected in lanes corresponding to GST-EGR2-H384N (lanes 9 and 11), but are virtually absent in lanes corresponding to GST-EGR2-E356K (lanes 6 and 8). Lane 1: probe only; lane 2: native GST protein.
B, Qualitative and quantitative assessment of native and fusion GST proteins. MW: molecular weight markers.
C, Western blot analyses using EGR2-specific antibodies show comparable levels of expression in transduced cells. Ratio was normalized to B-Actin and EGR2-WT expression.
D, Expression of EGR2 proteins is associated with decrease of Gr1 and B220 membrane expression. Analyses were gated on GFP-positive cells. FACS analysis at day 1 (blue) and day 4 (black) post-transduction are shown.
E, Growth curve showing of EML cells transduced by the indicated vector. Monitoring the proportions of GFP-positive cells show a decrease of the proportion of cells expressing WT and mutant EGR2 proteins, with a stronger decrease for the mutants.
F, Transcriptional modulation of endogenous Egr2 target genes in transduced cells. Expression was normalized with respect to GAPDH and expression in empty vector (MSCV) transduction. (p values: ****<0.0001, 0.0001<***<0.001, 0.001<**<0.01, 0.01<*<0.05, ns >0.05). only significant differences are shown.

Figure 5. Expression analyses of 15 CLL patients.
A, Heat map representation of the clustering of the CLL patients based on the 239 genes differentially expressed in EGR2-mutated samples.
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 up regulated upon BCR-stimulation of normal human B-lymphocytes, which was then used to compare transcription of EGR2-mutated and EGR2-not mutated CLL samples using GSEA. Shown is the enrichment plot (Enrichment Score ES=0.7718; Normalized Enrichment Score (NES)1.7360; FDR q-value=0.0049),
D, Heat map 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, Heat map of the expression of the leading edge genes (from analyses shown in E).

Figure 1

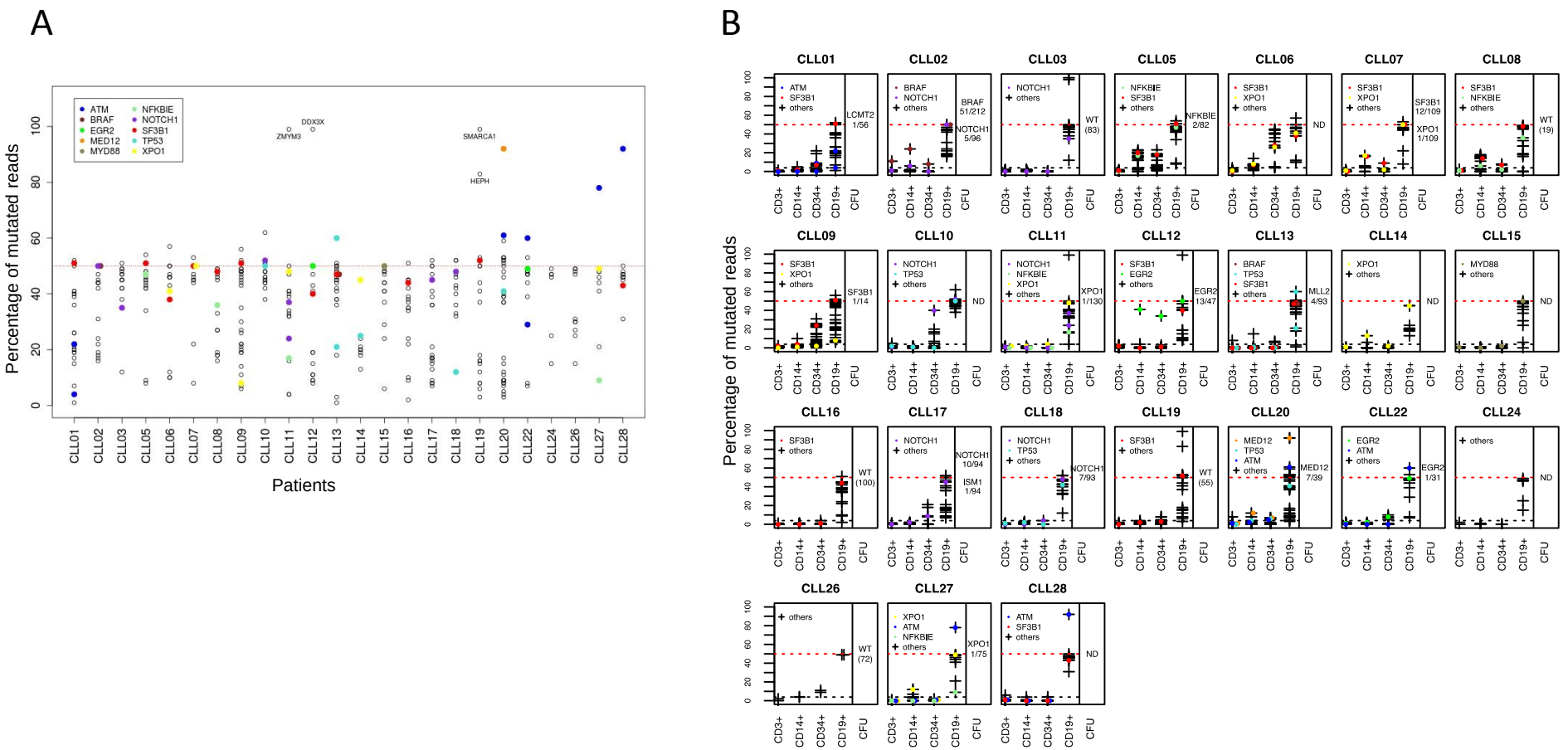


Figure 3

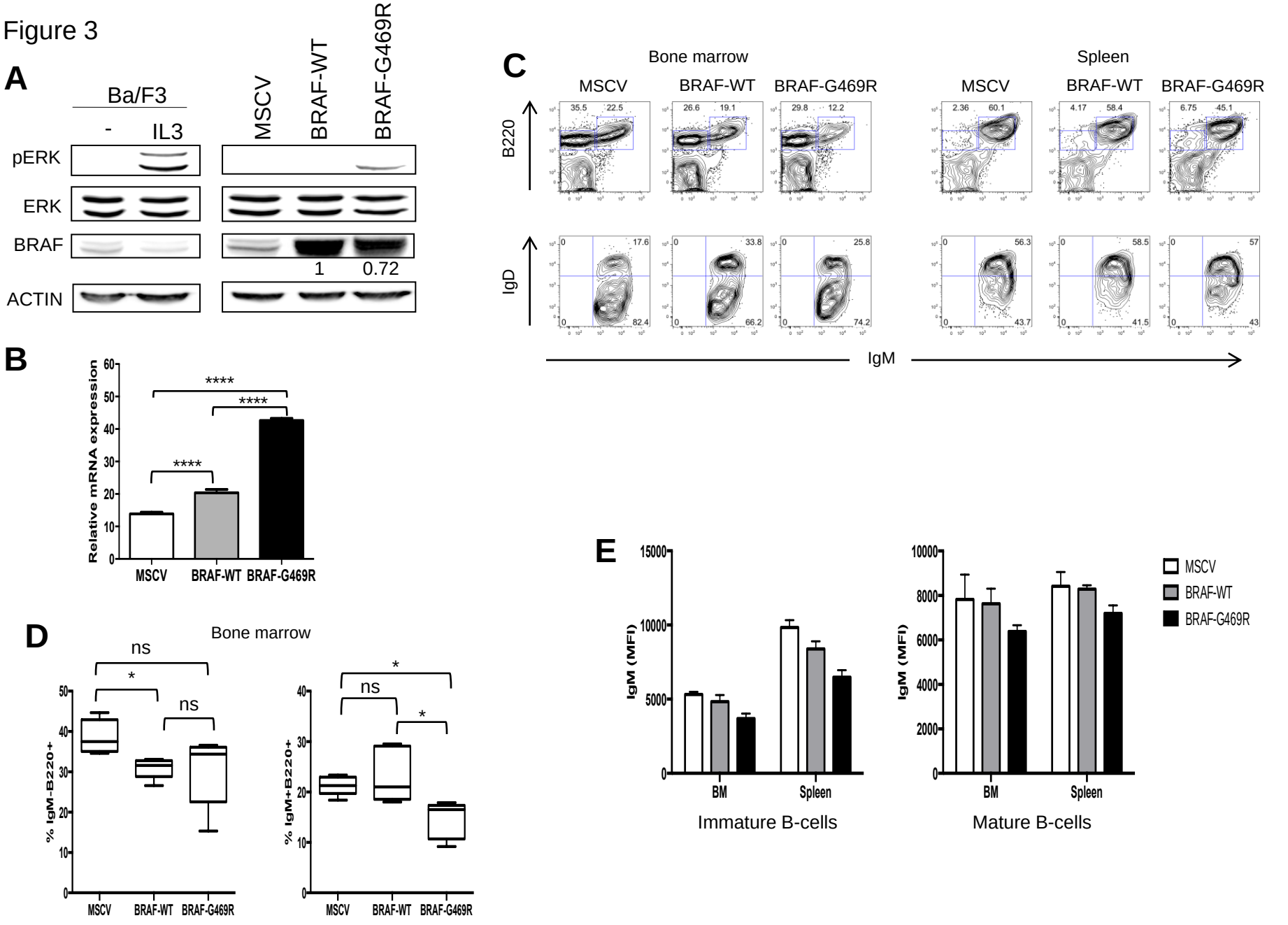
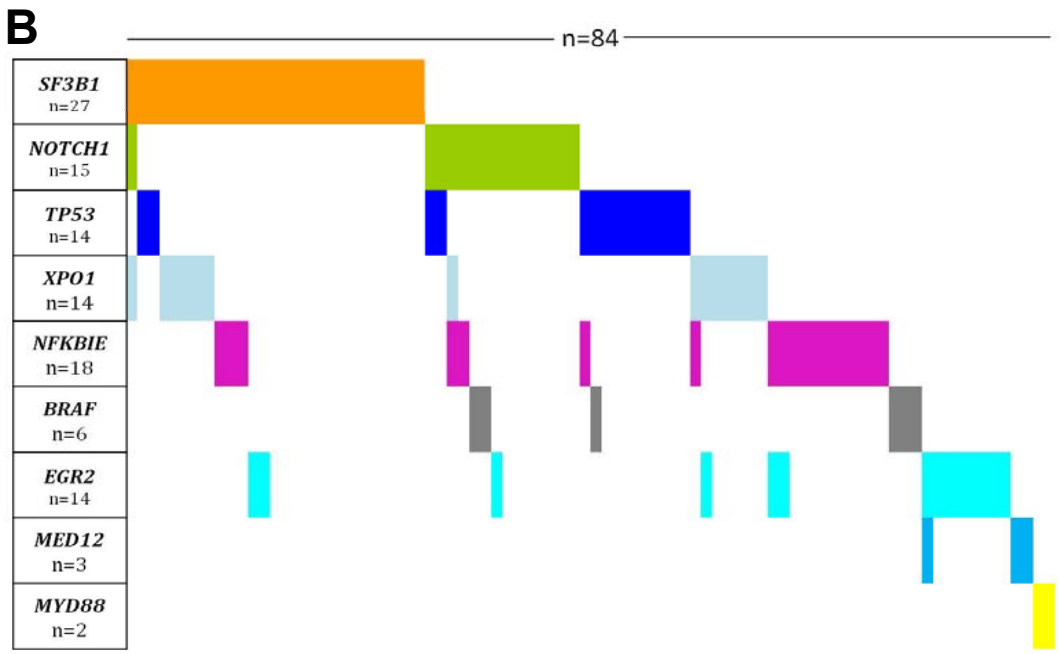
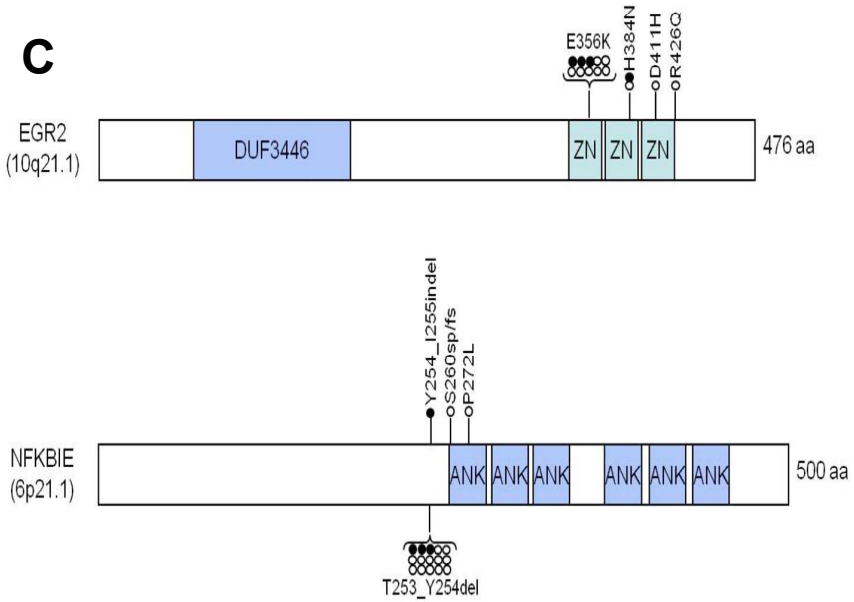


Figure 2

A

Gene	Mutated patients	Overall frequency (%)
SF3B1	27	16.1
NFKBIE	18	10.7
NOTCH1	15	8.9
TP53	14	8.3
EGR2	14	8.3
XPO1	14	8.3
BRAF	6	3.6
MED12	3	1.8
MYD88	2	1.2
Patients with at least one mutation	84/168	50%





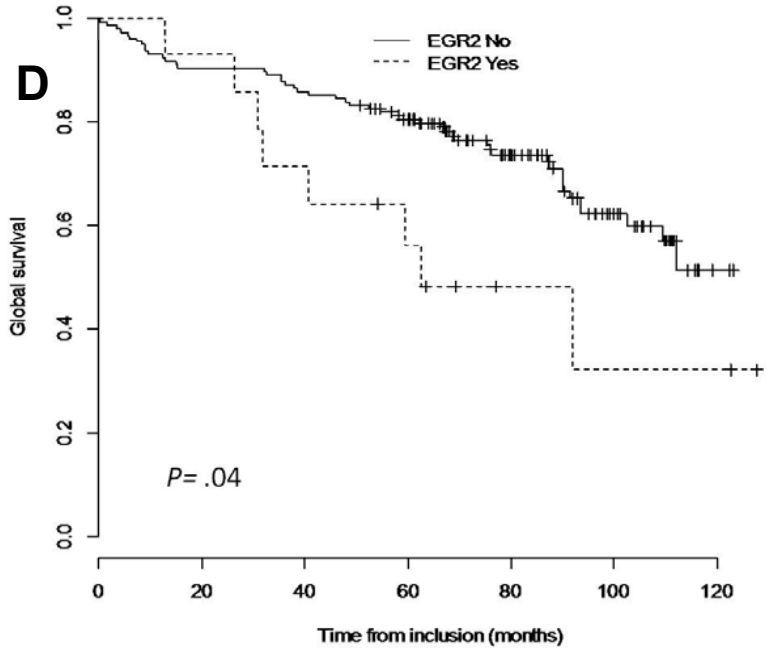


Figure 4

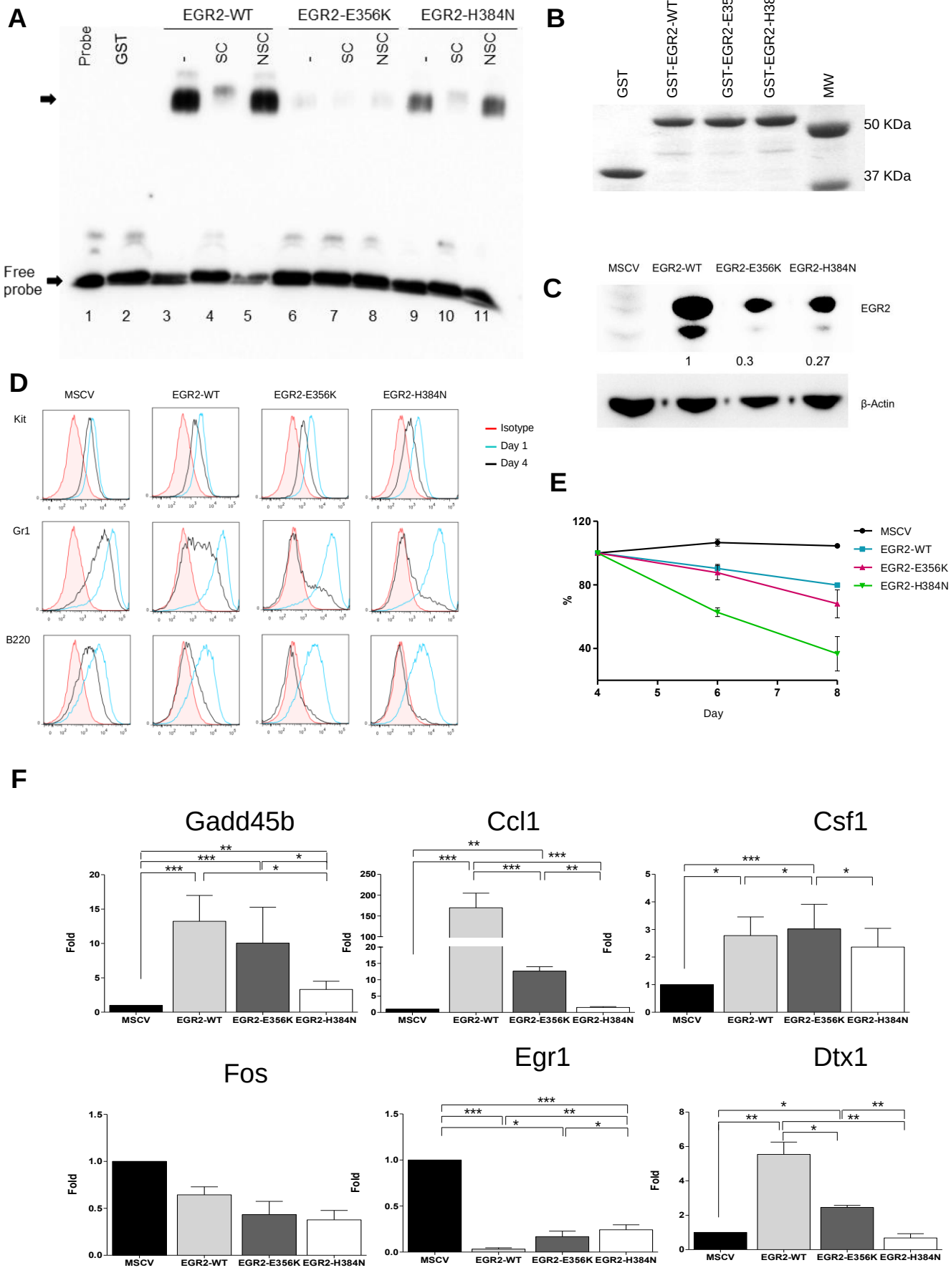
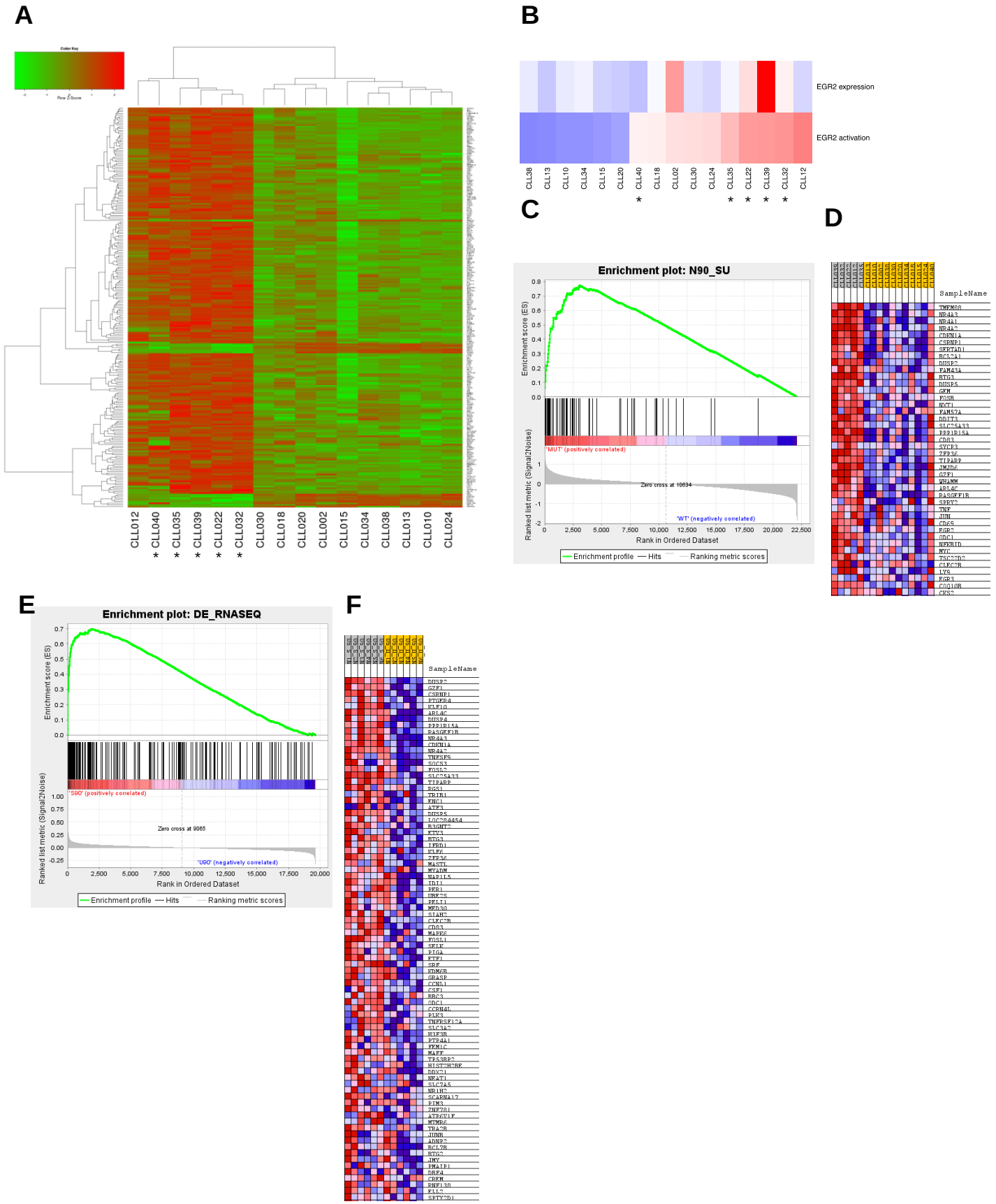


Figure 5



AACR American Association for Cancer Research

CANCER DISCOVERY

Acquired initiating mutations in early hematopoietic cells of CLL patients

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

Cancer Discovery Published OnlineFirst June 11, 2014.