Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

ABSTRACT Appropriate cancer care requires a thorough understanding of the natural history of the disease, including the cell of origin, the pattern of clonal evolution, and the functional consequences of the mutations. Using deep sequencing of flow-sorted cell populations from patients with chronic lymphocytic leukemia (CLL), we established the presence of acquired mutations in multipotent hematopoietic progenitors. Mutations affected known lymphoid oncogenes, including *BRAF*, *NOTCH1*, and *SF3B1*. *NFKBIE* and *EGR2* mutations were observed at unexpectedly 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* mutations suggest that they result in deregulation of B-cell receptor (BCR) intracellular signaling. Our data propose disruption of hematopoietic and early B-cell differentiation through the deregulation of pre-BCR signaling as a phenotypic outcome of CLL mutations and show that CLL develops from a pre-leukemic phase.

SIGNIFICANCE: The origin and pathogenic mechanisms of 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. *Cancer Discov; 4(9); 1088–1101. ©2014 AACR.*

See related commentary by Jiang and Elemento, p. 995.

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.

Authors' Affiliations: [1]Institut National de la Santé et de la Recherche Medicale (INSERM) U985; [2]INSERM U1009; [3]Institut Gustave Roussy, Villejuif; [4]INSERM U1138; [5]Université Pierre et Marie Curie-Paris 6; [6]Service d'Hématologie Biologique, Hôpital Pitié-Salpêtrière, APHP; [7]Université Paris Diderot, Hôpital Saint-Louis, SBIM; [8]Ligue Nationale Contre le Cancer, Equipe labellisée, Paris; [9]Université Paris-Sud; [10]Université Paris-Sud, Institut de Génétique et Microbiologie, CNRS UMR 8621, Orsay; [11]Service d'Hématologie Clinique, Centre Hospitalier Victor Dupouy, Argenteuil, France; [12]Cancer Genomics Project, Graduate School of Medicine; [13]Laboratory of DNA Information Analysis and [14]Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo; [15]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto; [16]Medicine and Biosystemic Science; and [17]Center for Cellular and Molecular Medicine, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan

Note: Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

F. Damm and E. Mylonas contributed equally to this article.

S. Ogawa, F. Nguyen-Khac, and O.A. Bernard share senior authorship of this article.

Corresponding Authors: Olivier A. Bernard, INSERM U985, 39 rue Camille Desmoulins, Villejuif 94805, France. E-mail: olivier.bernard@inserm.fr; Florence Nguyen-Khac, Unité de Cytogénétique Hématologique, Service d'Hématologie Biologique, GH Pitié-Salpêtrière/Charles Foix, 83 Bd de l'Hôpital, Paris 75013, France. E-mail: florence.nguyen-khac@psl.aphp.fr; and Seishi Ogawa, Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. E-mail: sogawa-tky@umin.ac.jp

doi: 10.1158/2159-8290.CD-14-0104

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

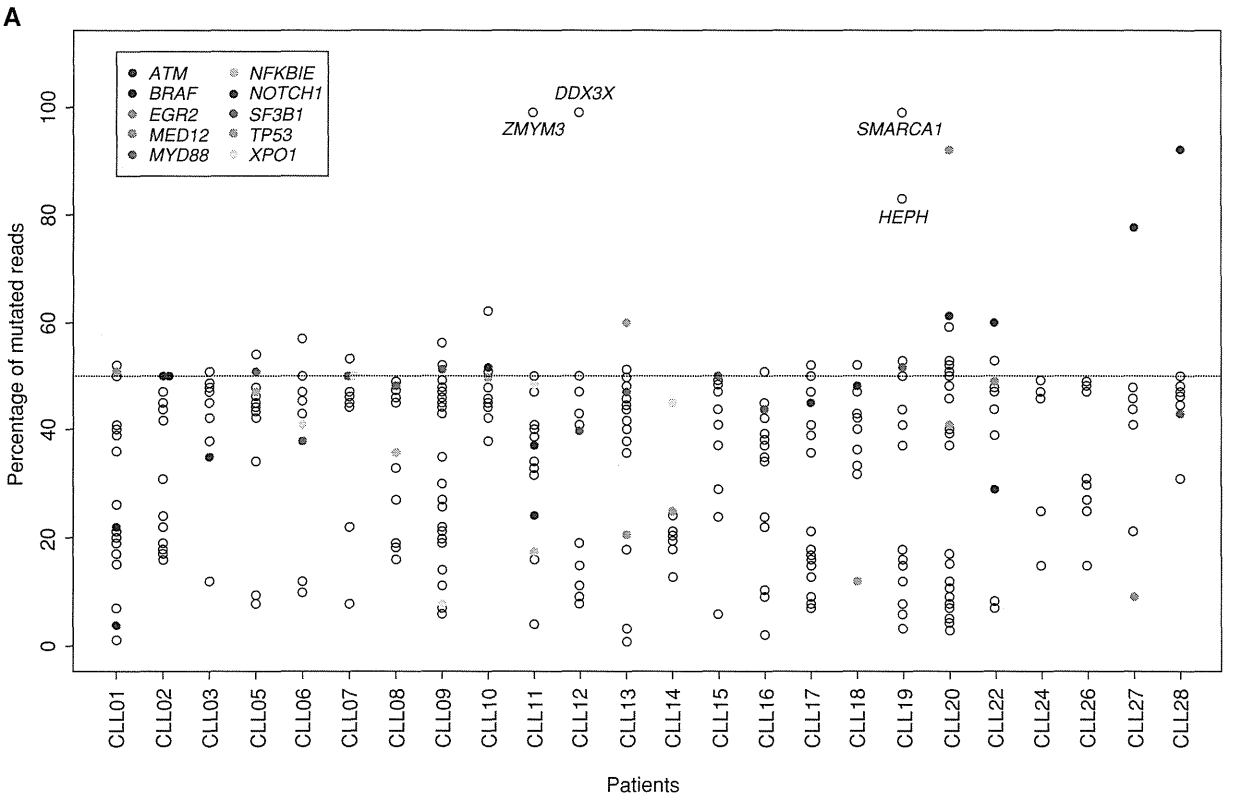

Figure 1. Somatic mutations identified in 24 patients with CLL. **A,** variant allele frequencies in the tumor fraction. Genes analyzed in the extension cohort are colored. *ZMYM3*, *DDX3X*, *SMARCA1*, *MED12*, and *HEPH* genes are on chromosome X. See Supplementary Table S1 for details. The patient numbers are in abscissa. (*continued on following page*)

RESULTS

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

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

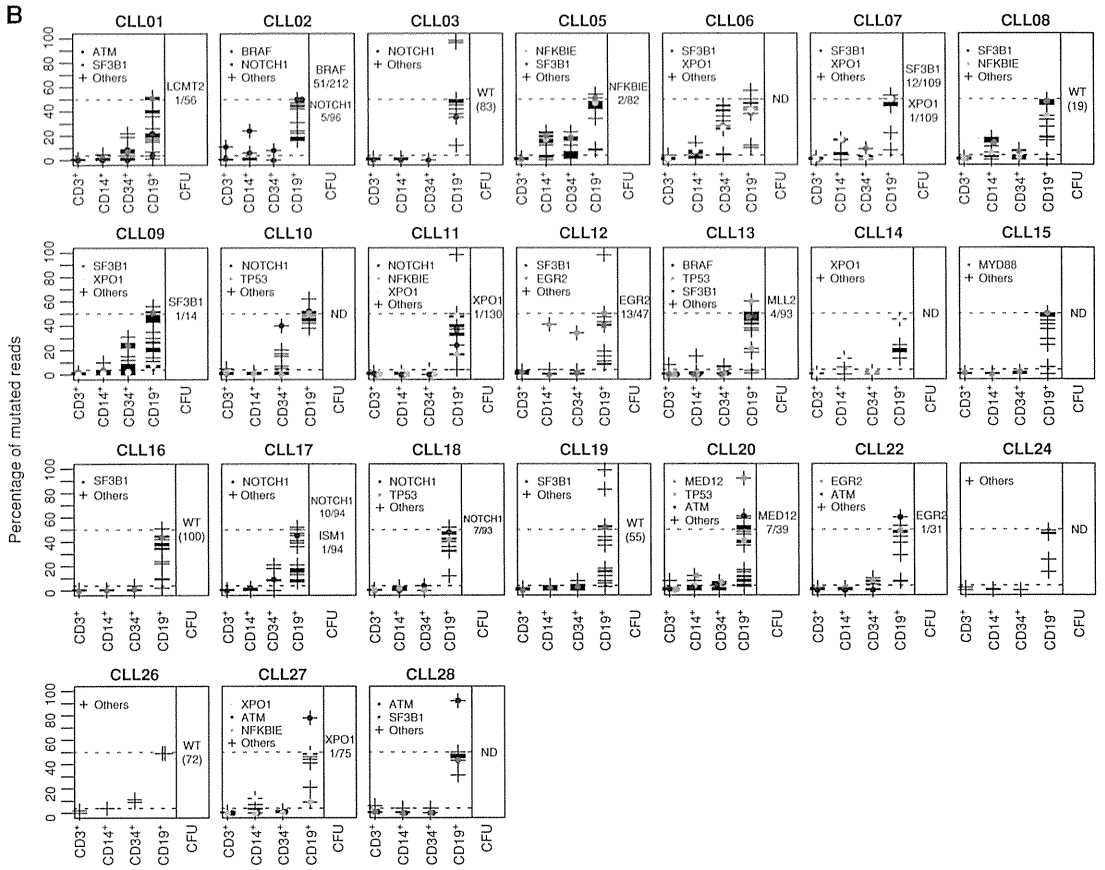

Figure 1. ***(Continued)*** **B,** variant allele frequencies in the hematopoietic fractions, sorted on the basis of the cell-surface expression of the following antigens (in abscissa): CD34 (progenitors), CD3 (T cells), CD19 and CD5 (tumor cells), and CD14 (monocytes). Column on the right, the genotyping 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 $CD14^+$ or in the $CD34^+$ fractions. Two patients (CLL14 and 27) showed mutations in the $CD14^+$ 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 (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

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

(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 *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 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 B 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. 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–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–positive 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).

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

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%

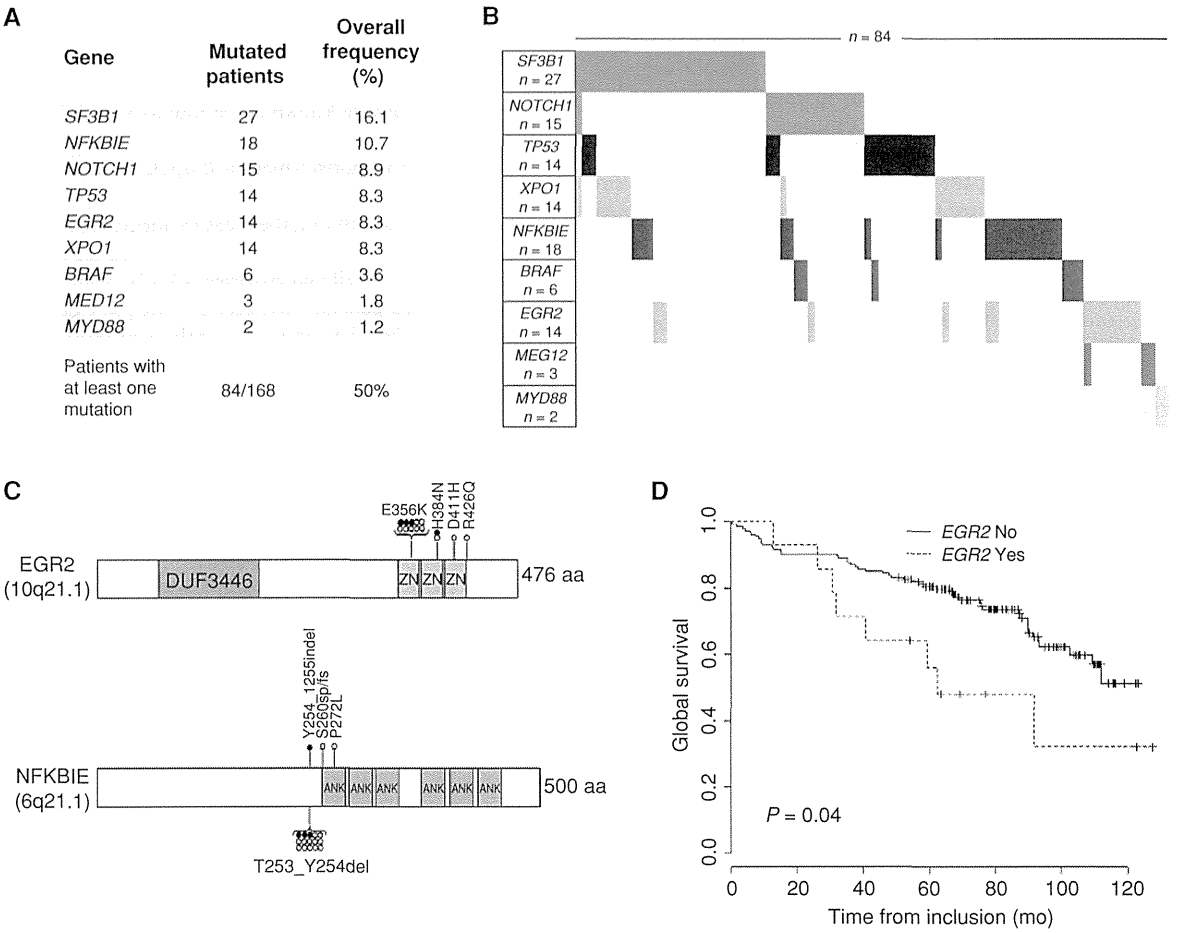

Figure 2. Gene mutation profile in 168 patients with CLL. Frequencies (**A**) and distributions (**B**) of identified mutations. **C,** localization of identified mutations in EGR2 and NFKBIE proteins. Circles, mutations; filled circle, proven somatic mutation. **D,** overall survival of patients with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

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

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

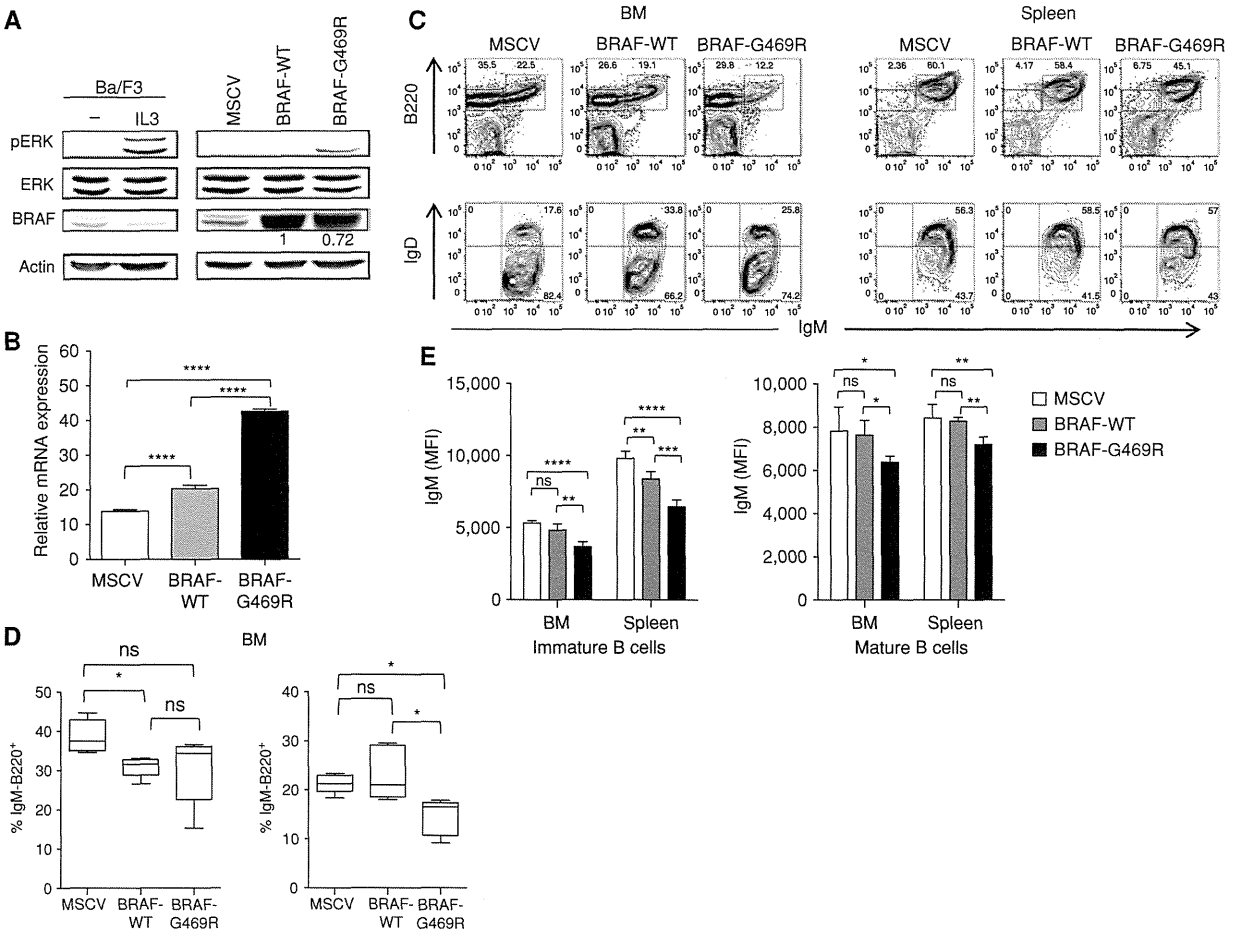

Figure 3. Functional analyses of BRAF-G469R. **A,** expression of BRAF-G469R from MSCV results in constitutive ERK activation. Right, ERK phosphorylation was detected in IL3-starved Ba/F3 cells expressing mutated BRAF, but not BRAF-WT or empty control. Antibodies are indicated and BRAF expression normalized to β-actin and BRAF-WT expression. **B,** real-time quantitative polymerase chain reaction (RQ-PCR) evaluation of *Egr2* expression in transduced Ba/F3 cells, normalized with respect to *Gapdh*. **C,** representative FACS analysis of bone marrow (BM; left) and spleen (right) cells from MSCV, BRAF-WT, and BRAF-G469R mice analyzed 5 weeks after engraftment. Plots are gated on donor (CD45.2$^+$) GFP$^+$ cells and the percentages of gated cells are shown. **D,** mean percentage (±SD; $n = 5$ mice) of bone marrow GFP$^+$ donor cells expressing B220 and negative (left) or positive (right) for membrane IgM. **E,** mean fluorescence intensity (MFI) of IgM in GFP$^+$ donor B cells, immature (IgM$^+$ IgD$^-$; left) and mature (IgM$^+$ IgD$^+$; right). BM or spleen origin is indicated (****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, $P > 0.05$).

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,

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

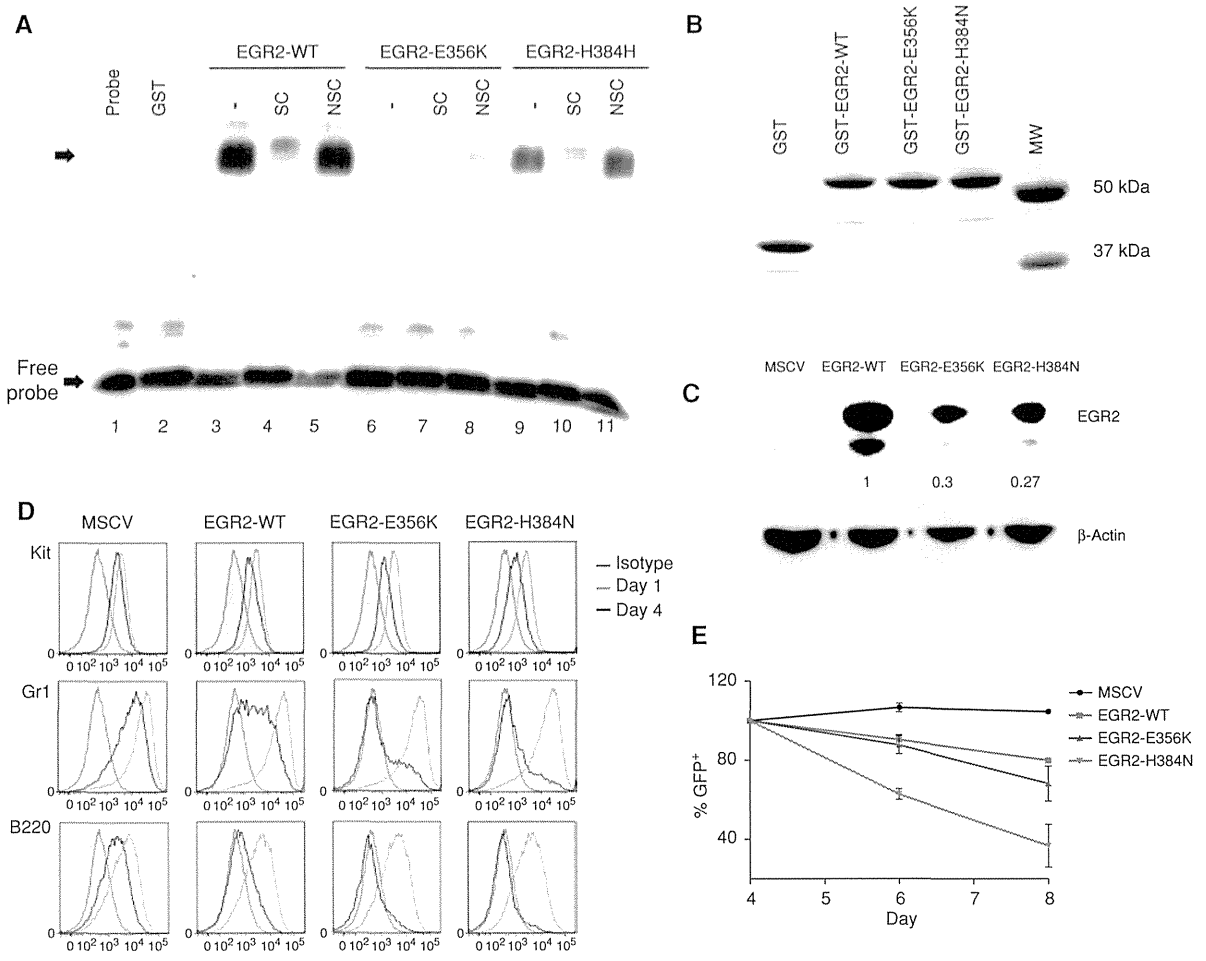

Figure 4. Functional analyses of EGR2 mutants. **A,** the zinc-finger regions of mutant EGR2 are less efficient than WT at binding EGR2 high-affinity sequences (CTCTGTACGCGGGGGCGGTTA) in EMSA. Shifted complexes (indicated by an arrow) are observed in lanes containing GST-EGR2-WT protein (lanes 3 and 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, and 10), but not in the presence of the same excess of a nonspecific competitor (NSC; lanes 5, 8, and 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 β-actin and EGR2-WT expression. **D,** expression of EGR2 proteins is associated with a decrease of Gr1 and B220 membrane expression. Analyses were gated on GFP+ cells. FACS analysis at day 1 (blue) and day 4 (black) posttransduction are shown. **E,** growth curve showing EML cells transduced by the indicated vector. Monitoring the proportions of GFP+ cells shows a decrease in the proportion of cells expressing WT and mutant EGR2 proteins, with a stronger decrease for the mutants. (*continued on following page*)

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

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

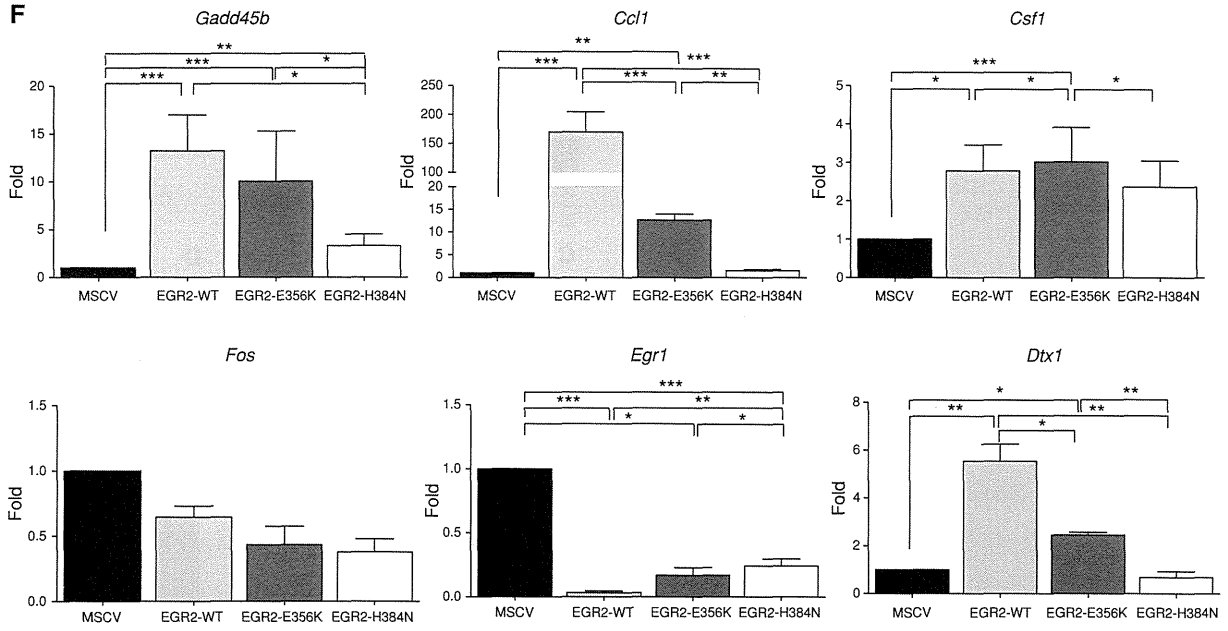

Figure 4. ***(Continued)*** **F,** transcriptional modulation of endogenous EGR2 target genes in transduced cells. Expression was normalized with respect to GAPDH and expression in empty vector (MSCV) transduction (****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$). Only significant differences (<0.05) are shown.

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 $CD34^+/CD38^-/CD45RA^+/CD90^-$; 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-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 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

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

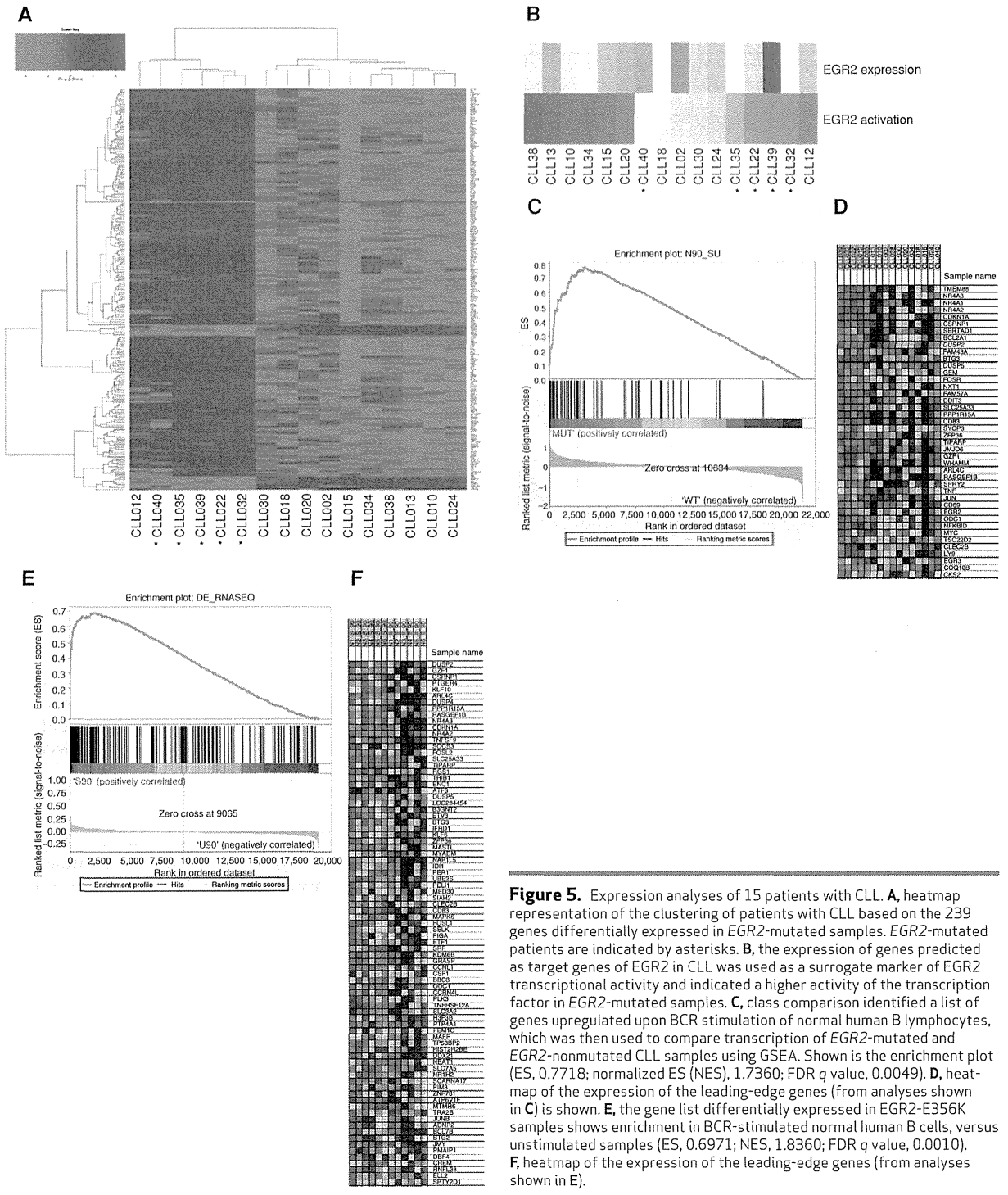

Figure 5. Expression analyses of 15 patients with CLL. **A,** heatmap representation of the clustering of patients with CLL based on the 239 genes differentially expressed in *EGR2*-mutated samples. *EGR2*-mutated patients are indicated by asterisks. **B,** the expression of genes predicted as target 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**).

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

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 (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 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 the distribution's variance. P values (adjusted by the Benjamini and Hochberg procedure) lower than 1×10^{-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://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 $\log_2$ 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; $P < 1e{-}9$; 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/+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), *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′ 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 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$^+$ 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-α milieu (Life Technologies)

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

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 ≥Q20 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 Guy 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% penicilin/streptomycine/glutamine, and supplemented with 10% of BHK cells supernatant. BaF3 cells were grown in RPMI medium, 10% bovine serum, and 1% penicilin/streptomycine/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 5×10^6 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 Pharmigen. 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 complementary 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′-CTCTG TAC**GCG**<u>**GGG**</u>**GCG**GTTA-3′. Nonspecific competitor was 5′-CTCTG TAC**GCG**<u>**CCC**</u>**GCG**GTTA-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 $MgCl_2$, 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 wexre 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, 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. Droin, S. Ogawa

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

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

V. Della Valle, E. Mouly, L. Scourzic, F. Davi, H. Merle-Béral, L. Sutton, W. Vainchenker, N. Droin, S. Ogawa, F. Nguyen-Khac

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): F. Damm, A. Cosson, K. Yoshida, E. Mouly, M. Diop, L. Scourzic, Y. Shiraishi, K. Chiba, H. Tanaka, S. Miyano, J. Lambert, D. Gautheret, P. Dessen, T. Mercher, S. Ogawa, O.A. Bernard

Writing, review, and/or revision of the manuscript: F. Damm, E. Mylonas, Y. Kikushige, 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 Chevret, Julie Lejeune (SBIM, St-Louis, Paris, France), and Claude Lesty (Pitie-Salpetriere Hospital) for help in statistical analyses; F. Norol and H. Trebeden-Negre (Pitie-Salpetriere 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 to E. Solary and O.A. Bernard), INCa-DGOS-INSERM (6043), Fondation Gustave Roussy, KAKENHI (23249052 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 is the recipient of a Deutsche Krebshilfe fellowship (grant 109686).

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

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. Sci Transl Med 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 2010;10:37–50.
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. N Engl J Med 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. Nat Genet 2011;44:47–52.
6. Burger JA, Chiorazzi N. B cell receptor signaling in chronic lymphocytic leukemia. Trends Immunol 2013;34:592–601.
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. J Exp Med 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. N Engl J Med 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. J Exp Med 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 G_0/G_1 transition in cultured cells. EMBO J 1988;7:29–35.
21. Dhomen N, Marais R. New insight into BRAF mutations in cancer. Curr Opin Genet Dev 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. Leuk 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*. J Exp Med 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 Res 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.

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

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. Proc Natl Acad Sci U S A 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. EMBO J 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. J Immunol 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. J Pathol 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. Neurobiol Dis 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. Hum Mol Genet 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. EMBO J 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. N Engl J Med 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. Proc Natl Acad Sci U S A 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 Discov 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 Biol 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. Proc Natl Acad Sci U S A 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.

Published OnlineFirst June 11, 2014; DOI: 10.1158/2159-8290.CD-14-0104

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 2014;4:1088-1101. Published OnlineFirst June 11, 2014.

Deep sequencing reveals stepwise mutation acquisition in paroxysmal nocturnal hemoglobinuria

Wenyi Shen,[1,2] Michael J. Clemente,[1] Naoko Hosono,[1] Kenichi Yoshida,[3] Bartlomiej Przychodzen,[1] Tetsuichi Yoshizato,[3] Yuichi Shiraishi,[4] Satoru Miyano,[4,5] Seishi Ogawa,[3] Jaroslaw P. Maciejewski,[1] and Hideki Makishima[1]

[1]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, USA. [2]Department of Hematology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China. [3]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [4]Laboratory of DNA Information Analysis, Human Genome Center and [5]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Paroxysmal nocturnal hemoglobinuria (PNH) is a nonmalignant clonal disease of hematopoietic stem cells that is associated with hemolysis, marrow failure, and thrombophilia. PNH has been considered a monogenic disease that results from somatic mutations in the gene encoding PIGA, which is required for biosynthesis of glycosylphosphatidylinisotol-anchored (GPI-anchored) proteins. The loss of certain GPI-anchored proteins is hypothesized to provide the mutant clone with an extrinsic growth advantage, but some features of PNH argue that there are intrinsic drivers of clonal expansion. Here, we performed whole-exome sequencing of paired PNH$^+$ and PNH$^-$ fractions on samples taken from 12 patients as well as targeted deep sequencing of an additional 36 PNH patients. We identified additional somatic mutations that resulted in a complex hierarchical clonal architecture, similar to that observed in myeloid neoplasms. In addition to mutations in *PIGA*, mutations were found in genes known to be involved in myeloid neoplasm pathogenesis, including *TET2*, *SUZ12*, *U2AF1*, and *JAK2*. Clonal analysis indicated that these additional mutations arose either as a subclone within the *PIGA*-mutant population, or prior to *PIGA* mutation. Together, our data indicate that in addition to *PIGA* mutations, accessory genetic events are frequent in PNH, suggesting a stepwise clonal evolution derived from a singular stem cell clone.

Introduction

Paroxysmal nocturnal hemoglobinuria (PNH), a prototypical disease of hematopoietic stem cells, is characterized by the clinical triad of intravascular hemolysis, thrombophilia, and bone marrow failure (1). While clonal, PNH has not been considered a malignancy. However, PNH does exhibit similarities to myelodysplastic syndrome (MDS), a chronic preleukemic myeloid neoplasm, including clonal hematopoiesis, persistence of an aberrant stem cell clone, and frequent derivation as a late complication of immune-mediated aplastic anemia (AA). The presence of a singular somatic mutation of the *PIGA* gene in hematopoietic stem cells leads to the defective biosynthesis of glycophosphatidylinositol (GPI) anchors, resulting in the deficiency and absence of GPI-anchored proteins on the cell surface, a hallmark of the PNH phenotype (2). These phenotypic features of affected PNH stem cells are believed to be responsible for an extrinsic growth advantage, which occurs in the context of immune-mediated hematopoietic suppression of hematopoiesis, as seen in AA (3). While immune privilege leading to clonal escape is plausible, it does not completely explain the evolution of PNH. Various observations suggest that intrinsic factors are also involved. For example, PNH often persists for many years after successful immunosuppression in AA, or PNH can present in a pure form without AA. Furthermore, the detection of tiny *PIGA* mutant clones in healthy individuals suggests a need for additional putative intrinsic factors that promote the expansion of GPI-deficient cells within the hematopoietic compartment (4, 5). These factors may include secondary genetic events such as somatic mutations. In support of this hypothesis, chromosomal abnormalities in the form of microdeletions involving the *PIGA* locus have been identified in a small proportion of cases of otherwise classical PNH (6, 7). Occasional clonal chromosomal abnormalities and somatic mutations including *NRAS* and *JAK2* mutations have also been reported in PNH (8, 9), supporting the notion that *PIGA* mutations and additional somatic events correlate with each other and may be responsible for maintenance and expansion of the PNH clone.

Recently, the application of next-generation sequencing (NGS) to study malignant clonal diseases has revealed clonal architecture at a much higher level of complexity than previously anticipated, demonstrating both the stepwise acquisition of mutations and expansion of the most permissive subclones. In this study, we present data from whole-exome sequencing (WES) of clonal (GPI-deficient) and nonclonal cells from PNH patients to examine the mutational history of PNH. We hypothesized that the evolution of a PNH clone may be associated with additional somatic mutational events and that such events may be either of an ancestral or a facilitating nature. These additional somatic mutations, if present, may help to further clarify the mechanism of clonal expansion and persistence of the mutated PNH stem cell, as well as explain the clinical diversity of PNH and distinct behavior of the PNH clones. Our data suggest that PNH, a nonmalignant yet clonal disorder, displays clonal architecture in a manner similar to that of leukemia (10, 11).

▶ Related Commentary: p. 4227

Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: December 16, 2013; **Accepted:** July 10, 2014.
Reference information: *J Clin Invest.* 2014;124(10):4529–4538. doi:10.1172/JCI74747.

Table 1. Patient characteristics

Variable	Whole cohort (n = 60)	WES cohort (n = 12)
Age		
Median, yr	44	43
Range, yr	10–77	20–72
Sex		
Male	27	6
Female	33	6
Symptoms at diagnosis	30 (50%)	5 (42%)
Neutropenia	15 (25%)	3 (25%)
Anemia	35 (58%)	8 (67%)
Thrombocytopenia	15 (25%)	4 (33%)
Hemolysis	31 (52%)	9 (75%)
Thrombosis	12 (20%)	4 (33%)
PNH clone size		
Mean	61%	87%
Range,%	0.8–99.8	60–99

Neutropenia: absolute neutrophil count (ANC) ≤1.5 × 10^9/l; anemia: Hg ≤10g/dl; thrombocytopenia: platelets ≤100 × 10^9/l.

Results

Mutational search using WES. We analyzed a total of 60 patients with PNH (Table 1). After immunomagnetic sorting of wbc into PNH ($CD59^-$) and non-PNH ($CD59^+$) fractions, we verified purity by flow cytometry (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI74747DS1) and performed WES on 12 patients. Application of a bioanalytic algorithm designed to detect somatic mutations led to the selection of a total of 107 nonsilent alterations (98 were single nucleotide variants [SNVs] and 9 were indels). Through validation by Sanger sequencing and targeted deep NGS of candidate alterations (61 genes, Supplemental Table 2), we confirmed the presence of 38 somatic events in a total of 31 genes in PNH-derived DNA. The average coverage of exome sequencing and targeted deep NGS was 103 times and 316 times, respectively. Our stringent bioanalytic platform favored avoidance of false-negatives, and, consequently, the accuracy of the initial calling algorithm was 22% for SNVs and 89% for indels. In addition to 3 cases with Xp22.2 *PIGA* microdeletions, we identified a total of 31 missense, 8 nonsense, 26 frameshift, and 13 splice site mutations in cases analyzed by WES (n = 12) and targeted (n = 36) NGS sequencing (Table 2).

PIGA mutations and deletions. We detected somatic *PIGA* mutations (3 SNVs and 10 indels) in the PNH fractions of 9 of 12 cases analyzed by WES, while paired non-PNH $CD59^+$ DNA samples were negative, although a small cross-contamination artifact was detected in some samples. For example, in PNH patient 5 (PNH5), we identified a single splice site mutation with a high variant allelic frequency in sorted PNH^+ ($CD59^-$) cells that was generally absent in normal ($CD59^+$) cells (Figure 1A). To further explore the frequency of *PIGA* mutations in our cohort aside from the WES cases (n = 12), we used Sanger sequencing (n = 36) and targeted *PIGA* (n = 10) deep sequencing. Overall, *PIGA* mutations were detected in 60% (36 of 60) of patients (Table 2). As reported previously in the literature (12), some PNH patients harbor more than 1 *PIGA* mutation. In 14% (5 of 36) of *PIGA*-mutated cases, 2 independent mutations were found, and 1 case (PNH45) contained 3 mutations (Table 2 and representative patients depicted in Figure 1C and Supplemental Figure 1). In an index female PNH case (PNH1), the somatic nature of 2 *PIGA* mutations (p.G68E and intron 5 splice site) was confirmed, since deep sequencing indicated that both were confined to the PNH fraction (Figure 1C). The 2 mutations were located 431 nucleotides apart, and bacterial subcloning analysis demonstrated that they were not present in the same clone (Figure 1D). Two independent experiments involving Sanger sequencing of single colonies further validated the biclonal nature of the 2 *PIGA* mutations, as no colony contained both mutations (Figure 1E). Deep sequencing provided variant allelic frequencies (VAFs) of 18% (p.G68E) and 12% (intron 5 splice site) with a corresponding clonal size of 36% and 24%, respectively, as the mutations were heterozygous. Overall, in each of the 6 PNH cases with 2 or more *PIGA* mutations, we identified 2 (or 3) independent PNH clones with 1 unique *PIGA* mutation per clone, and all were codominant.

In total, 9 of 12 (75%) WES, 9 of 10 (90%) targeted deep sequencing, and 15 of 36 (42%) Sanger sequencing cases were positive for *PIGA* defects, including 3 nonsense, 21 frameshift, 5 missense, and 11 splice site mutations (Figure 2A). One WES case and an additional 2 cases contained a microdeletion involving the *PIGA* locus (delXp22.2, spanning an average of 559 kb, range 506–616 kb) as detected by SNP array. These microdeletions were confined to PNH cells and were not found in non-PNH fractions. In summary, we were able to detect either *PIGA* mutations or microdeletions in 60% (36 of 60) of patients studied, with 3% (1 of 36) triclonal, 14% (5 of 36) biclonal, 75% (27 of 36) monoclonal, and 8% (3 of 36) microdeletions (Figure 2B). We found no mutations or microdeletions in 40% (24 of 60) of the cases studied, with Sanger sequencing demonstrating a much lower detection rate than WES or targeted deep sequencing. Flow cytometric analysis of PNH clone size using Alexa 488 Proaerolysin Variant (FLAER) indicated that the only patient with 3 *PIGA* mutations also had a significant (>3%) wbc type II PNH clone (Figure 2C). Based on the evolution of PNH clone size as assessed longitudinally by flow cytometry in patients with at least a 4-year follow-up and a detected *PIGA* mutation (n = 16), the vast majority of patients tended to equilibrate at a clone size greater than 80% of all leukocytes regardless of the nature of the *PIGA* mutation (Figure 2D).

Additional somatic mutations in PNH. In addition to the pathognomonic *PIGA* mutations, 21 somatic mutations in 21 other genes were found in 83% (10 of 12) of cases tested by WES (Table 2). All of these mutations were confirmed by independent testing using both Sanger and targeted NGS. In contrast to what would be expected from a benign condition, the number of such novel gene mutations other than *PIGA* ranged between 0 and 6 per case, with an average of 2 additional events. In an index case of PNH with thrombocytosis (PNH1), we detected a novel somatic heterozygous mutation primarily in the PNH fraction in the *NTNG1* gene (Figure 1C). While *NTNG1* mutations or dysregulation of receptor ligand inter-

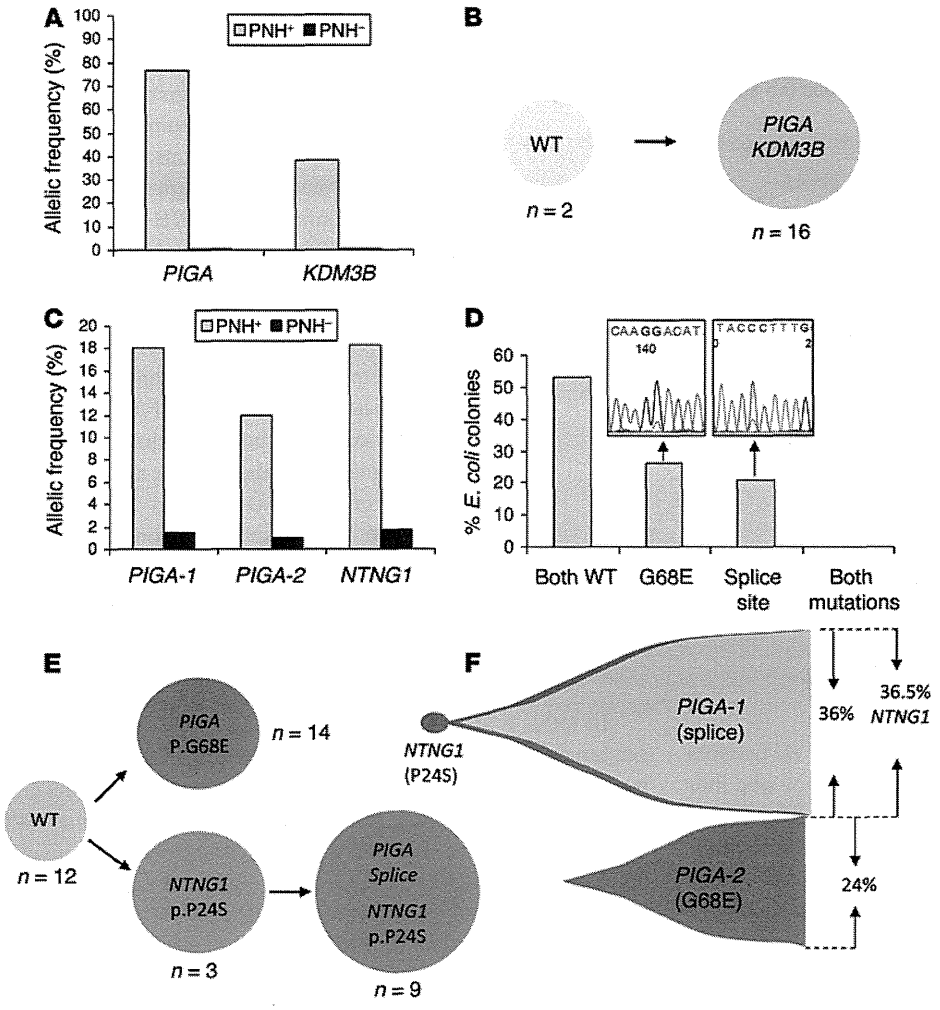

Figure 1. *PIGA* mutations can be either primary or secondary events. The primary event (PNH5) is represented in **A** and **B**; the secondary event (PNH1) is represented in **C**–**F**. (**A**) Analysis of VAFs of the mutations identified in case PNH5 indicated that the *KDM3B* mutation was present at a lower frequency than the *PIGA* mutation, and both mutations were almost exclusively confined to the sorted PNH$^+$ (CD59$^-$) fraction. (**B**) Single-colony sequencing results confirmed that the *PIGA* and *KDM3B* mutations were present in the same cell population. (**C**) Deep sequencing VAFs for *PIGA*-1 (G68E), *PIGA*-2 (splice site), and *NTNG1* (P24S) mutations, all of which were primarily present in the PNH fraction in the PNH1 case. (**D**) Bacterial subcloning and Sanger sequencing results demonstrated that the *PIGA* mutations in this case were independent, suggesting the presence of 2 separate PNH clones. (**E**) Single-colony sequencing further confirmed that 2 independent PNH clones were present and also suggested that the *PIGA* splice site mutation appeared to be a secondary event preceded by a *NTNG1* mutation. (**F**) The combination of deep sequencing data with single-colony sequencing allowed for a representation of the clonal architecture in PNH1.

actions have been described in various disorders including colorectal cancer (13–16), the p.P24S mutation has not been previously reported. Other somatic mutations discovered in PNH include those in *TET2*, *MAGEC1*, *BRPF1*, *KDM3B*, and *STAC3* genes, all found in the PNH fraction and not in phenotypically normal cells (Table 2). All of these mutations were heterozygous without loss of heterozygosity (LOH) encompassing the affected gene locus. To assess whether these gene mutations were frequently recurrent in PNH, we screened an additional 36 cases by targeted deep sequencing, including a variety of genes ($n = 61$) that are frequently affected in MDS (Supplemental Table 2). In addition to 2 somatic homozygous *JAK2* (p.V617F) mutations, both of which were present in PNH cases with microdeletions of Xp22.2, *SUZ12*, *DHX29*, *MECOM*, *BCOR*, *U2AF1*, ASXL1, *BRCC3*, *ETV6*, *KDM6A*, *NTNG1*, *BRPF1*, *MAGEC1*, *CCR9*, *ALDH1B1*, *WDR96*, *TMC1*, *CPD*, *NRXN3*, *CELSR1*, *KDM3B*, *STAC3*, *SLC20A1*, *MUC7*, *RBP3*, *C11orf34*, *MAN1A2*, *PEX14*, *SYNE*, and *FBN1* mutations were found in a single PNH case each, with *RIT1* and *MECOM* in 2 cases and *TET2* in 3 cases, one of which harbored 2 different *TET2* mutations (Table 2).

Clonal architecture. In order to evaluate the clonal composition of each patient, we used single-colony sequencing assays, bacterial subcloning, and analysis of VAF data obtained from targeted deep sequencing. For example, in PNH5, a *KDM3B* mutation was also discovered, albeit at a lower frequency than that of the *PIGA* mutations (Figure 1A). Clonogenic assays were used to determine whether the mutations were independent (biclonal) or co-occuring (subclonal, Figure 1B). In this case, all colonies were either wild-type or had both mutations. Taken in context with allelic frequency data indicating that the *KDM3B* mutation was present at a lower frequency than the *PIGA* mutation, these results suggest that the *PIGA* mutation was the initial event.

Our first indication that a *PIGA* mutation may not always be the initial event in PNH came during our analysis of the WES data from PNH1. VAF of the *NTNG1* mutation (18.3%) was slightly greater than that of either of the *PIGA* mutations (18% vs. 12%) detected (Figure 1C), prompting subcloning and single-colony sequencing experiments to determine the clonal composition of the observed mutations. These experiments confirmed the suspicion that the *NTNG1* mutation preceded the *PIGA* intron 5 splice site mutation, as we observed a number of colonies with a *NTNG1* mutation but without a *PIGA* mutation (Figure 1C). Furthermore, no colonies demonstrated the presence of a *PIGA* intron 5 splice site mutation in the absence of a concurrent *NTNG1* mutation, leading to our current understanding of clonal architecture in this patient (Figure 1D), who clearly has 2 independent *PIGA* mutations, one of which evolved as a secondary event subsequent to the initial NTNG1 alteration.

Table 2. Gene mutations identified by WES (PNH1–10) and targeted deep NGS (PNH15–60)

Patient	No. of mutations	*PIGA*	Other gene mutations	Ancestral mutated gene	Subclonal gene
PNH1	3	G68E, intron 5 splice	*NTNG1* (P24S)	*NTNG1* (P24S)	*PIGA*
PNH2	4	L83fs, S127X	*MAGEC1* (C747Y), *BRPF1*(N797S)	*MAGEC1* (C747Y)	*PIGA*
PNH3	7	L76fs	*TMC1* (E80G), *WDR96* (T1115I), *NRXN3* (Y9C), *CCR9* (M188T), *ALDH1B1* (K81X), *CPD*(P472S)	*TMC1* (E80G), *WDR96* (T1115I), *NRXN3* (Y9C)	*CCR9* (M188T), *ALDH1B1* (K81X), *CPD* (P472S)
PNH4	2	Microdeletion Xp22.2	*CELSR1* (E2404D)	–	–
PNH5	2	Intron 4 splice	*KDM3B* (L125I)	*PIGA*	*KDM3B*
PNH6	3	L243fs	*STAC3* (F97V), *TET2* (E1250X)	*STAC3* (F97V), *TET2* (E1250X)	*PIGA*
PNH7	3	T192fs, R233fs	*SLC20A1* (I671L)	*SLC20A1* (I671L)	*PIGA*
PNH8	4	Q18X	*MUC7* (R358X), *RBP3* (G560S), *C11orf34* (F15L)	*PIGA*	*MUC7* (R358X), *RBP3* (G560S), *C11orf34* (F15L)
PNH9	3	E395fs, R119fs	*MAN1A2* (splice)	*MAN1A2* (splice)	*PIGA*
PNH10	4	T192fs	*PEX14* (Y290C), *SYNE2* (K5198N), *FBN1* (T1340A)	*PEX14* (Y290C), *SYNE2* (K5198N)	*PIGA*, *FBN1* (T1340A)
PNH15	1	–	*SUZ12* (intron 2 splice)	–	–
PNH16	2	–	*TET2* (S1556fs, A671fs)	–	–
PNH17	1	Exon 3 splice	–	–	–
PNH19	1	S130fs	–	–	–
PNH22	2	V150fs	*BCOR* (Q1606X)	–	–
PNH24	1	I53fs	[B]	–	–
PNH25	1	V150fs	–	–	–
PNH26	1	Exon 3 splice	–	–	–
PNH27	2	I269fs	*ASXL1* (M1345L)	–	–
PNH31	1	H154R	–	–	–
PNH34	1	Q252X	–	–	–
PNH37	1	S368fs	[B]	–	–
PNH38	2	M326fs	*DHX29* (K498N)	–	–
PNH39	1	V193fs	[B]	–	–
PNH40	1	V67fs	–	–	–
PNH41	1	I417fs	–	–	–
PNH42	1	H128R	[B]	–	–
PNH43	1	T71fs	[B]	–	–
PNH44	1	–	*MECOM* (P18S)	–	–
PNH45	5	L171R[A], exon 3 splice[A], G146fs[A]	*BRCC3* (A53fs)[A], *RIT1* (Q212X)[A]	–	–
PNH47	2	G263R	*MECOM* (K613fs)	–	–
PNH49	2	Microdeletion Xp22.2	*JAK2V617F*	–	–
PNH50	2	Microdeletion Xp22.2	*JAK2V617F*	–	–
PNH51	2	Intron 4 splice	*U2AF1* (Q157P)	–	–
PNH52	2	Q100fs, intron 4 splice	–	–	–
PNH53	1	Intron 4 splice	–	–	–
PNH54	2	Intron 4 splice	*KDM6A* (K151R)	–	–
PNH57	2	Intron 4 splice	*ETV6* (A52V)	–	–
PNH59	2	Exon 3 splice[A]	*RIT1* (V58G)[A]	–	–
PNH60	1	–	*TET2* (L1514fs)	–	–

[A]Not confirmed by Sanger sequencing; [B]DNA quality insufficient for targeted deep sequencing; –, not detected or undetermined.

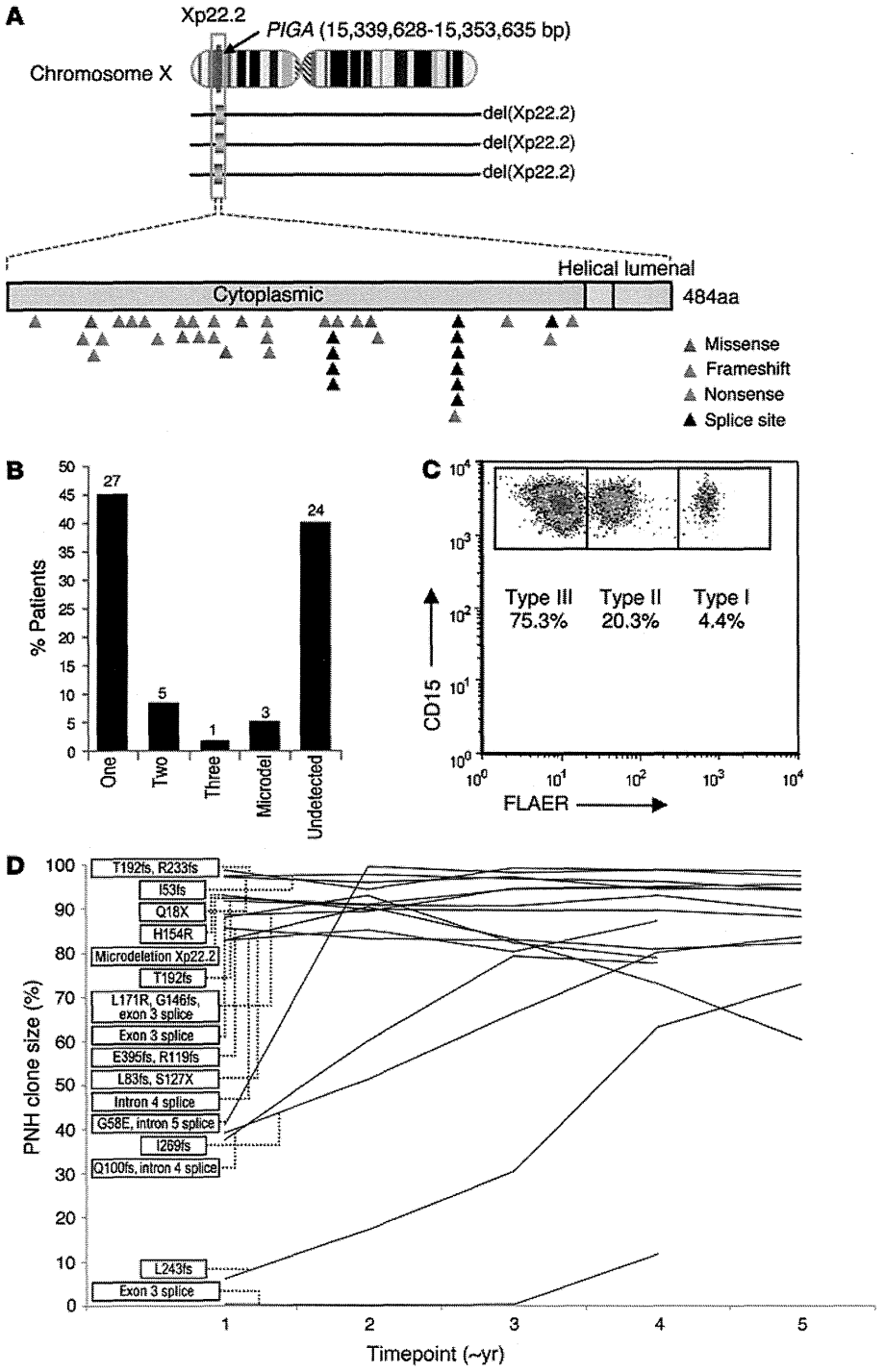

Figure 2. *PIGA* mutations and longitudinal analysis. (**A**) Distribution of *PIGA* missense, frameshift, nonsense, splice site, and microdeletions. (**B**) Proportion and absolute number of patients with 1, 2, or 3 mutations as well as microdeletions and those patients with the PNH phenotype in which a *PIGA* mutation was not detected. (**C**) Representative flow cytometric plot quantifying the number of type I (normal), type II (intermediate GPI anchor loss), and type III (complete GPI anchor loss) cells. (**D**) Longitudinal flow cytometry quantifying the wbc PNH clone size in patients with approximately 4 years or more of follow-up (n = 16). Each time point corresponds to approximately 1 year.

As we continued our analysis of the data, it became evident that additional mutations further increase the complexity of clonal architecture in PNH. In an illustrative case (PNH8), we identified 1 *PIGA* mutation and an additional 3 somatic mutations (*C11orf34*, *RBP3*, *MUC7*; Figure 3A). Again, based on targeted deep NGS and single-colony sequencing (Figure 3B), we deduced that the *PIGA* mutation constituted an initial ancestral event followed by subclonal defects in 3 other genes (Figure 3C). The diversity of clonal architecture in PNH became apparent when more cases were analyzed. In PNH2, another case with 2 *PIGA* mutations, additional somatic events including *MAGEC1* and *BRPF1* mutations, were found and independently confirmed. Comparison of VAFs obtained through targeted deep NGS revealed that the corresponding clonal size for the *MAGEC1* mutation was at least comparable to, if not larger than, that of the largest *PIGA* mutation, suggesting that the initial event may not have been a *PIGA* mutation, but indeed a *MAGEC1* mutation (Supplemental Figure 1). In PNH9, we found that the initial event that may be either a malignant or a passenger mutation, again was not a *PIGA* mutation, but a novel mutation in *MAN1A2*, which was followed by the appearance of 2 *PIGA* mutations, thus creating 2 independent clones, both of which carried the original *MAN1A2* mutation. In a somewhat similar case (PNH10, Supplemental Figure 2), somatic *SYNE2* and *PEX14* gene mutations were the initial events, followed by a *PIGA* frameshift mutation and an additional *FBN1* mutation. Perhaps the most complex case, PNH3, demonstrated 7 somatic gene mutations that were all confined to the PNH fraction; single-colony sequencing confirmed that the *PIGA* mutation occurred after 2 other clonal mutations had already been acquired (*TMC1* and *WDR96*) and, again, was not the initial event (Supplemental Figure 3).

In addition to the detection of concomitant mutations within the PNH clone, we found somatic events that predated the *PIGA* mutations and were present in both PNH and non-PNH myeloid fractions but not in germline DNA derived from T cells. These mutations included *TET2*, *SUZ12*, and *JAK2*. In PNH6, we observed concomitant *TET2* and *STAC3* mutations. However, mutant fractions were larger than those of the *PIGA* mutant, and the *TET2* mutation was also present in the non-PNH fraction, which, in turn, was negative for *STAC3* and *PIGA* mutations (Supplemental Figure 4). These results indicate that PNH, in this case,

Table 3. Clinical characteristics of PNH cases according to the presence of additional somatic mutations

Characteristic	PNH with additional mutations (n = 24)	PNH without additional mutations (n = 26)	P value
Age at diagnosis, yr Mean (range)	37 (16–71)	36 (5–77)	0.77
Age at sequencing, yr Mean (range)	45 (20–75)	40 (10–77)	0.48
wbc PNH clone size, % Mean (range)	68.9 (0.8–98.9)	49.8 (1.3–95)	0.04[A]
Absolute neutrophil count, /mm^3 Mean (range)	2,845 (850–8,130)	3,259 (1,050–9,530)	0.56
Hemoglobin, g/dl Mean (range)	10.8 (8.6–16)	11 (7–17)	0.97
Platelets, 109/l Median(range)	95.3 (14–255)	117 (6–467)	0.37
Hemolysis, n (%)	12 (52)	14 (54)	1
Thrombosis, n (%)	5 (25)	6 (16)	0.4866
Cytogenetics			
Normal, n (%)	14 (58)	7 (26)	1
Abnormal, n (%)	1 (4)	1 (4)	1
NA	9 (38)	18 (69)	

[A]Statistical significance at $\alpha = 0.05$.

Discussion

One of the discoveries of cancer NGS projects was the realization of genetic heterogeneity due to a combination of somatic mutations and the complexity of clonal architecture reflective of the sequential acquisition of genetic defects (11, 17–21). While general principles of clonal dynamics have been derived from the study of leukemia, they have not been explored in nonmalignant diseases such as PNH. The results of our study lead to a conceptual understanding of this disease: PNH is subject to clonal dynamics and selection forces similar to those observed in hematopoietic neoplasms. Consequently, in PNH, additional clonal and subclonal mutations corroborate with the pathognomonic *PIGA* mutation responsible for the prevailing cell phenotype. An important conclusion of this new concept is that the nature and composition of additional clonal somatic mutations may modify the behavior of the *PIGA* clone and have the potential to explain various distinct clinical courses seen in individual PNH patients. Thus, the analogy to MDS is apparent with multiple mutations that arise within the clonal population and undergo clonal selection along with *PIGA* mutations. Furthermore, the mutational events discovered in our research are not unique to PNH, but overlap significantly with the spectrum of mutations seen in typical MDS. Many of these mutations, or other mutations in the same genes, have been identified as key drivers of clonal evolution in MDS or cancer, including *U2AF1* (22, 23), *TET2* (2, 24), *MAGEC1* (25), *BRPF1* (26), *NRXN3* (27), *KDM3B* (28), *SLC20A1* (29), *MUC7* (30), *PEX14* (31), *FBN1* (32), *SUZ12* (33), *ASXL1* (34), *BCOR* (35), *DHX29* (36), *MECOM* (37), *RIT1* (38), and *JAK2* V617F (39). In some ways, PNH appears genetically similar to, but pathogenetically distinct from, MDS. While additional mutations have been described in PNH (6), our results illustrate for the first time in a large PNH cohort that genetic clonal selection and evolution may be operative not only in malignant conditions, but also in otherwise benign hematologic diseases.

Our stringent bioanalytic approach has led to effective exclusion of germline variants based on comparatively similar high VAFs in both experimental and control fractions. If variant reads that derived from clonal mutations were present in the CD59$^+$ cell population due to contamination, their VAF was low and was correspondingly high in the PNH$^+$ population. As such, this would be incompatible with germline alterations, which would be expected to oscillate around 50% of reads in both fractions. With this in mind, all of the mutational events we report in this article are exclusively somatic. While we found that some mutations had similar VAFs in both fractions, they were clearly somatic, since the VAF was much lower than would be expected in heterozygous germline variants. In such situations, we verified the absence of the mutation in the CD3$^+$ fraction and concluded that the *PIGA* mutation arose in the context of a preexisting clone. Future prospective studies would benefit from the use of nonhematopoietic

evolved as a subclone after a clonal *TET2* mutation was acquired. However, in the bone marrow of another case (PNH3), we identified dysplastic changes along with trisomy 8 in 20 of 20 metaphases. FISH analysis resolved the origin of the trisomy 8, which was present only in the non-PNH fraction (data not shown). These results suggest that PNH in patient 3 evolved independently of the acquisition of trisomy 8.

Overall, our clonal analysis of 9 WES cases suggested that *PIGA* mutations were often acquired at a later stage (6 of 9 cases, Table 2), and mutations in other genes were the initial clonal events. Of note is that in a number of cases, the clonal composition showed significant overlap, preventing precise recapitulation of clonal hierarchy. In summary, our experiments indicate that the cohort can be stratified into 4 different scenarios (Figure 4): *PIGA* as the initial ancestral event accompanied by secondary mutations (Figure 4B, patients 5 and 8); *PIGA* as an event secondary to other mutations (Figure 4C, patients 1, 2, 6, 7, 9, and 10); *PIGA* as the lone mutation (Figure 4D, patients 17, 19, 25, 26, 31, 34, 40, 41, 52, and 53); and a *PIGA* mutation coexisting with other mutations responsible for the development of an MDS clone (Figure 4E, patient 3).

Clinical correlations. The discovery of additional somatic mutations in PNH prompted further investigation into the potential clinical impact these mutations may have. Analysis of various relevant clinical parameters with regard to the presence (n = 24) or absence (n = 26) of an additional mutation largely failed to find significant differences (Table 3). However, the presence of an additional mutation beyond *PIGA* was associated with a larger wbc PNH clone size as assessed by flow cytometry (P = 0.04; Table 3). Furthermore, based on the instructive discovery that PNH45 harbored 3 *PIGA* mutations (triclonal PNH) and a relatively large PNH type II wbc population (Figure 2C), we hypothesized that the presence of more than 1 *PIGA* mutation leads to the presence of both type II and type III PNH cells. Indeed, patients with more than 1 *PIGA* mutation had a higher incidence of a type II wbc population of greater than 3% than those with only 1 mutation (4 of 5 vs. 4 of 12, respectively; P = 0.048, data not shown).