SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia


From the Cancer Vaccine Center (L.W., Y.W., L.Z., W.Z., Q.L.S., C.J.W.) and the Departments of Medical Oncology (L.W., Y.W., L.Z., W.Z., A.R.V., S.M.F., N.R.G., B.T., Q.L.S., M.M., J.R.B., C.J.W.), Biostatistics and Computational Biology (K.S., L.W., D.N.), and Pediatric Oncology (E.G.F., R.R., T.R.G.), Dana–Farber Cancer Institute; the Departments of Cell Biology (E.G.F., R.R., T.R.G.) and Medicine (J.R.B., C.J.W.), Harvard Medical School; the Division of Allergy, Immunology, and Rheumatology, Department of Medicine, Massachusetts General Hospital (N.H.); and the Department of Medicine, Brigham and Women’s Hospital (J.R.B., C.J.W.) — all in Boston; the Broad Institute, Cambridge, MA (M.S.L., P.S., C.S., A.S., D.S.D., K.C., E.S., S.G., N.H., M.M., T.R.G., E.S.L., G.G.); and the Howard Hughes Medical Institute, Chevy Chase, MD (T.R.G.). Address reprint requests to Dr. Wu at the Dana–Farber Cancer Institute, Harvard Institutes of Medicine, Rm. 416B, 77 Ave. Louis Pasteur, Boston, MA 02115, or at cwu@partners.org.

Drs. Wang, Lawrence, and Wan and Drs. Brown, Getz, and Wu contributed equally to this article.

This article (10.1056/NEJMoa1109016) was published on December 12, 2011, at NEJM.org.


ABSTRACT

BACKGROUND

The somatic genetic basis of chronic lymphocytic leukemia, a common and clinically heterogeneous leukemia occurring in adults, remains poorly understood.

METHODS

We obtained DNA samples from leukemia cells in 91 patients with chronic lymphocytic leukemia and performed massively parallel sequencing of 88 whole exomes and whole genomes, together with sequencing of matched germline DNA, to characterize the spectrum of somatic mutations in this disease.

RESULTS

Nine genes that are mutated at significant frequencies were identified, including four with established roles in chronic lymphocytic leukemia (TP53 in 15% of patients, ATM in 9%, MYD88 in 10%, and NOTCH1 in 4%) and five with unestablished roles (SF3B1, ZMYM3, MAPK1, FBXW7, and DDX3X). SF3B1, which functions at the catalytic core of the spliceosome, was the second most frequently mutated gene (with mutations occurring in 15% of patients). SF3B1 mutations occurred primarily in tumors with deletions in chromosome 11q, which is associated with a poor prognosis in patients with chronic lymphocytic leukemia. We further discovered that tumor samples with mutations in SF3B1 had alterations in pre–messenger RNA (mRNA) splicing.

CONCLUSIONS

Our study defines the landscape of somatic mutations in chronic lymphocytic leukemia and highlights pre-mRNA splicing as a critical cellular process contributing to chronic lymphocytic leukemia.
CHRONIC LYMPHOCYTIC LEUKEMIA IS AN incurable disease characterized by extensive clinical heterogeneity despite a common diagnostic immunophenotype (surface expression of CD19+, CD20⁺dim, CD5+, CD23+, and sigM⁺dim). Whereas the course of disease is indolent in some patients, it is steadily progressive in approximately half of patients, leading to substantial morbidity and mortality.¹ Our ability to predict a more aggressive disease course has improved with the use of tests for biologic markers (degree of somatic hypermutation in the variable region of the immunoglobulin heavy chain [IGHV] gene and expression of ZAP70) and the detection of cytogenetic abnormalities (deletions in chromosomes 11q, 13q, or 17p and trisomy 12).² Still, even with these advances, prediction of the disease course is not highly reliable.

Massively parallel sequencing technology now provides a means of systematically discovering the genetic alterations that underlie disease and identifying new therapeutic targets and clinically predictive biomarkers. To date, most studies designed to discover tumor-associated mutations have relied on sequencing the genome or exome of only a few tumors; the newly discovered mutations detected are then further studied in an expanded cohort. These efforts have led to the identification of several important disease-associated mutations.⁴ A more powerful approach to the process of initial discovery is to sequence a much larger set of samples.² The approach increases the chances that the full range of mutated genes will be detected, allows reconstruction of the genetic pathways underlying disease pathogenesis, and reveals associations between genetic events and the clinically important features of a disease. We therefore sequenced DNA samples of leukemia cells from 91 patients with chronic lymphocytic leukemia (88 exomes and 3 genomes), representing the broad clinical spectrum of the disease.

METHODS

STUDY DESIGN

Samples of DNA were obtained from normal tissues and tumors in 91 patients (discovery cohort) and 101 patients (extension cohort) with chronic lymphocytic leukemia, all of whom provided written informed consent before sample collection. DNA was extracted from blood- or marrow-derived lymphocytes in tumors and from autologous epithelial cells, fibroblasts, or granulocytes in normal tissue.

GENOME AND EXOME SEQUENCING

Libraries were constructed and sequenced on an Illumina Genome Analyzer II with the use of 101-bp paired-end reads for whole-genome sequencing and 76-bp paired-end reads for whole-exome sequencing. Output from Illumina software was processed by the PICARD data-processing pipeline to yield BAM files containing well-calibrated, aligned reads.⁷,⁸ BAM files were processed by the Broad Institute’s Firehose Pipeline, which provides quality control and identifies somatic point mutations, insertions or deletions, and other structural chromosomal rearrangements. Pre–messenger RNA (pre-mRNA) splicing alteration in leukemic samples was identified with the use of quantitative reverse-transcriptase–polymerase-chain-reaction assays to detect spliced and unspliced forms of representative spliceosome targets BRD2 and RIOK3.¹⁰ Associations between mutation rate and clinical features were assessed with the use of the Wilcoxon rank-sum test, Fisher’s exact test, or the Kruskal–Wallis test, as appropriate. A stepwise Cox proportional-hazards model was used to identify features with a significant effect on time to initial treatment. The materials and methods used in the study are more fully described in the Supplementary Appendix, available with the full text of this article at NEJM.org.

RESULTS

SOMATIC MUTATION RATE

We sequenced DNA derived from CD19⁺CD5⁺ leukemia cells and matched germline DNA derived from autologous skin fibroblasts, epithelial cells in saliva, or blood granulocytes. Samples were obtained from patients with a broad range of clinical characteristics, including those with del(11q) and del(17p) (indicating a poor prognosis) and with either unmutated or mutated IGHV status (Fig. 1A in the Supplementary Appendix). Deep sequence coverage was obtained to provide high sensitivity in identifying mutations (Table 1). To detect point mutations and insertions or deletions, we compared sequences in each tumor

---

sample with corresponding normal sequences by means of well-validated algorithms.\textsuperscript{7,8,11,12}

We detected 1838 nonsynonymous and 539 synonymous mutations in protein-coding sequences, corresponding to a mean (±SD) somatic mutation rate of 0.72±0.36 per megabase (range, 0.08 to 2.70), and an average of 20 nonsynonymous mutations per patient (range, 2 to 76) (Table 1, and Table 1 in the Supplementary Appendix). This rate is similar to that previously reported in chronic lymphocytic leukemia and other hematologic cancers.\textsuperscript{5-7,13,14} We observed no significant difference in the rates of nonsynonymous mutation between tumors with mutated genes and those with unmutated genes or in tumors in different clinical stages of disease (Table 2 in the Supplementary Appendix). Prior exposure to chemotherapy (in 30 of 91 patients) was not associated with an increased rate of nonsynonymous mutation (P=0.14) (Fig. 1B in the Supplementary Appendix).\textsuperscript{15}

**Identification of Genes with Significant Mutation Frequencies**

To identify genes whose mutations were associated with leukemic tumorigenesis (“driver” mutations), we examined all 91 leukemic and normal pairs with the use of the MutSig algorithm for genes that were mutated at a rate significantly higher than the background rate, given their sequence composition (see the Supplementary Appendix). Nine such genes were identified (Q≤0.1 after correction for multiple-hypothesis testing): TP53, SF3B1, MYD88, ATM, FBXW7, NOTCH1, ZMYM3, DDX3X, and MAPK1 (Fig. 1). Whereas the overall ratio of nonsynonymous to synonymous mutations was 3.1:1, the mutations in these nine genes were exclusively nonsynonymous (65:0, P<5×10\textsuperscript{−6}) (Table 1 in the Supplementary Appendix), a finding that further supports their functional importance. Moreover, these mutations occurred exclusively in conserved sites across species (Fig. 2 in the Supplementary Appendix).

Four of the genes with significant mutation frequencies, TP53, ATM, MYD88, and NOTCH1, have been described previously in chronic lymphocytic leukemia.\textsuperscript{6,16-18} We found 15 TP53 mutations in 14 of 91 patients (15%; Q≤6.3×10\textsuperscript{−8}), most of which were localized to the DNA-binding domain that is critical for its tumor-suppressor activity\textsuperscript{17} (Fig. 3A in the Supplementary Appendix). In 8 patients (9%), we detected 9 ATM mutations (Q≤1.1×10\textsuperscript{−5}) scattered across this large gene, including in regions where mutation has been associated with defective DNA repair in patients with chronic lymphocytic leukemia.\textsuperscript{16} MYD88, a critical adaptor molecule of the interleukin-1 receptor–toll-like receptor (TLR) signaling pathway, harbored missense mutations in 9 patients (10%) at three sites localized within 40 amino acids of the interleukin-1 receptor–TLR domain. One site was novel (P258L), whereas the other two were identical to those recently described as activating mutations of the nuclear factor κB (NF-κB)–TLR pathway in patients with diffuse large B-cell lymphoma (M232T and L265P) (Fig. 3C in the Supplementary Appendix).\textsuperscript{19} Finally, we detected a recurrent frameshift mutation (P2514fs) in the C-terminal PEST domain of NOTCH1 in 4 patients (4%) that was identical to that recently reported in other investigations of chronic lymphocytic leukemia.\textsuperscript{5,6} This mutation is associated with unmutated IGHV and a poor prog-

### Table 1. Summary Metrics of Whole-Genome and Whole-Exome Sequencing Studies.\textsuperscript{*}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Whole genomes</strong></td>
<td></td>
</tr>
<tr>
<td>No. Bases covered per genome — %</td>
<td>70</td>
</tr>
<tr>
<td>Genome coverage</td>
<td></td>
</tr>
<tr>
<td>CLL samples</td>
<td>38x</td>
</tr>
<tr>
<td>Normal samples</td>
<td>33x</td>
</tr>
<tr>
<td><strong>Whole exomes</strong></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>88</td>
</tr>
<tr>
<td>Bases covered per exome — %</td>
<td>81</td>
</tr>
<tr>
<td>Exome coverage</td>
<td></td>
</tr>
<tr>
<td>CLL samples</td>
<td>132x</td>
</tr>
<tr>
<td>Normal samples</td>
<td>146x</td>
</tr>
<tr>
<td><strong>Nonsynonymous mutations</strong></td>
<td></td>
</tr>
<tr>
<td>No. of mutations per megabase</td>
<td>0.7±0.36</td>
</tr>
<tr>
<td>Coding mutations — no. (range)</td>
<td>20 (2–76)</td>
</tr>
<tr>
<td><strong>Synonymous mutations</strong></td>
<td></td>
</tr>
<tr>
<td>No. of mutations per megabase</td>
<td>0.2±0.16</td>
</tr>
<tr>
<td>Coding mutations — no. (range)</td>
<td>5.8 (0–31)</td>
</tr>
</tbody>
</table>

*Plus–minus values are means ±SD. CLL denotes chronic lymphocytic leukemia.
† In 38x, 33x, 132x, and 146x, “x” denotes the average number of reads covering each nucleotide base that was sequenced.
nosis, and it is predicted to cause impaired degradation of NOTCH1, leading to pathway activation.

Five of the genes with significant mutation frequencies (SF3B1, FBXW7, DDX3X, MAPK1, and ZMYM3) do not have established roles in chronic lymphocytic leukemia. Strikingly, the second most frequently mutated gene in our cohort was splicing factor 3b, subunit 1 (SF3B1), with missense mutations occurring in 14 of 91 patients (15%). SF3B1 is a component of the SF3b complex, which is associated with the U2 small nuclear ribonucleoprotein (snRNP) at the catalytic center of the spliceosome. SF3B1, other U2 snRNP components, and defects in splicing are not typically implicated in the biology of chronic lymphocytic leukemia. Remarkably, all 14 mutations were localized within C-terminal PP2A-repeat regions 5 through 8, which are highly conserved from humans to yeasts (Fig. 2 and 3 in the Supplementary Appendix), and in 7 instances, an identical amino acid change was produced (by the K700E mutation). The clustering of heterozygous mutations within specific domains and at identical sites suggests that mutations in SF3B1, like those in MYD88 and NOTCH1, cause specific functional changes. Whereas the N-terminal domain of SF3B1 is known to interact directly with other spliceosome components, the precise role of its C-terminal domain remains unknown. Only 6 mutations have been reported in SF3B1, all in solid tumors and in the PP2A-repeat region (Table 4 in the Supplementary Appendix).

The four remaining genes with significant mutation frequencies have not, to our knowledge, been reported in previous studies of chronic lymphocytic leukemia and appear to have functions that interact with the five commonly mutated genes cited above (Fig. 3 in the Supplementary Appendix). FBXW7 (four distinct mutations) is a ubiquitin ligase that is known to be a tumor-suppressor gene, with loss of expression in a wide range of cancers. Its targets include important oncoproteins such as Notch1, c-Myc, c-Jun, cyclin E1, and MCL1. DDX3X (three distinct mutations) is an RNA helicase that functions at multiple levels of RNA processing, including RNA splicing, transport, and translation initiation as well as regulation of an RNA-sensing proinflammatory pathway. DDX3X interacts directly with XPO1, which was recently reported as being mutated in 2.4% of patients with chronic lymphocytic leukemia.
mia. MAPK1 (three distinct mutations), also known as ERK, is a kinase that is involved in core cellular processes, such as proliferation, differentiation, transcription regulation, and development, and is a key signaling component of the TLR pathway. Two of three distinct MAPK1 mutations localize to the protein kinase domain; these mutations, to our knowledge, are the first reported examples of somatic mutations within the protein kinase domain of an ERK family member in a human cancer. Finally, we identified four distinct mutations in ZMYM3, a component of multiprotein complexes containing histone deacetylase that function to silence genes by modifying chromatin structure.

We validated the three most frequently recurring mutations — SF3B1–K700E, MYD88–L265P, and NOTCH1–D2514fs — in 101 independent paired tumor and germline DNA samples, with similar detection frequencies in the discovery and extension cohorts (P = 0.20, P = 0.58, and P = 0.38, respectively) (Table 5 in the Supplementary Appendix).

The nine genes with mutations at significant frequencies appear in five core signaling pathways, in which the genes play well-established roles: DNA repair and cell-cycle control (TP53 and ATM), Notch signaling (FBXW7 and NOTCH1), inflammatory pathways (MYD88, DDX3X, and MAPK1), and RNA splicing and processing (SF3B1 and DDX3X) (Fig. 2). We also noticed that additional genes are mutated in these pathways — in 8 of 22 del(11q) samples (36%, P = 0.004). Of the 6 leukemic samples with mutated SF3B1 and without del(11q), 2 also harbored a heterozygous mutation in ATM. These findings strongly suggest that there is an interaction between del(11q) and SF3B1 mutation in the pathogenesis of this clinical subgroup of chronic lymphocytic leukemia.

We further observed that the mutations in NOTCH1 and FBXW7 were associated with trisomy 12 (P = 0.009 and P = 0.05, respectively). As in previous studies, NOTCH1 mutations were consistently associated with unmutated IGHV. The NOTCH1 and FBXW7 mutations were present in independent samples, suggesting that they may lead to aberrant Notch signaling in patients with trisomy 12 and unmutated IGHV.

All MYD88 mutations were present in samples that were heterozygous for del(13q) (P = 0.009). As indicated in recent reports, we found that the MYD88 mutation was always associated with mutations in the IGHV region (P = 0.001), which suggests a postgerminal-center origin. We speculate that in chronic lymphocytic leukemia, as in
diffuse large B-cell lymphoma, in which MYD88 is frequently mutated, the constitutive activation of the NF-κB pathway may have its greatest effect in the germinal center.

**SF3B1 Mutations**

In the 192 leukemic samples in the discovery and extension sets, mutations in NOTCH1 were associated with unmutated IGHV status and mutations in MYD88 were associated with mutated IGHV status. The mutation SF3B1–K700E was associated with unmutated IGHV status (P=0.048), but it was also detected in samples with mutated IGHV, which suggests that it is an independent risk factor for chronic lymphocytic leukemia (Fig. 5A in the Supplementary Appendix). Indeed, a Cox multivariable regression model designed to test for clinical factors contributing to an earlier initiation of treatment in the 91 patients in the study revealed that an SF3B1 mutation was predictive of an earlier need for treatment (hazard ratio, 2.20; P=0.03), independent of other established predictive markers, such as IGHV mutation status, del(17p), or an ATM mutation (Fig. 4A).

Consistent with the results of these analyses was the finding that the time to initial treatment for patients with the SF3B1 mutation alone — without del(11q) — was similar to that for patients with del(11q) alone or with both del(11q) and an SF3B1 mutation. All three groups had significantly shorter times to initial treatment than patients without an SF3B1 mutation or without del(11q) (Fig. 5B in the Supplementary Appendix) (P<0.001). Similarly, shorter times to initial treatment were observed among 3 patients in the extension cohort whose tumors harbored the SF3B1–K700E mutation as compared with those whose tumors did not show this mutation. Because SF3B1 encodes a splicing factor that lies at the catalytic core of the spliceosome, we looked for functional evidence of alterations in splicing associated with an SF3B1 mutation. Kotake et al. previously used intron retention in the endogenous genes BRD2 and RIOK3 to test the function of the SF3b complex. We confirmed that E7107, which targets this complex, inhibits the splicing of BRD2 and RIOK3 in both normal cells and chronic lymphocytic leukemia cells (Fig. 6A in the Supplementary Appendix). Using this assay, we found aberrant endogenous splicing activity in tumor samples from 13 patients with mutated SF3B1 as compared with 17 patients with wild-type SF3B1, and the ratio of unspliced to spliced mRNA forms of BRD2 and RIOK3 was significantly higher in patients with SF3B1 mutations (median ratio, 4.3:1 vs. 1.5:1 [P=0.001], and 4.6:1 vs. 2.1:1 [P=0.006], respectively) (Fig. 4B).

In contrast, no splicing defects were detected in samples with the del(11q) defect and wild-type SF3B1 as compared with samples with the del(11q) defect and mutated SF3B1 (Fig. 6 in the Supplementary Appendix). These studies indicate that splicing function in chronic lymphocytic leukemia is altered as a result of a mutation in SF3B1 rather than del(11q).

**Discussion**

Massively parallel sequencing technology has dramatically accelerated the discovery of genetic alterations in cancer. Our analysis of samples from 91 patients with chronic lymphocytic leukemia provided the statistical power to identify the involvement of nine driver genes and to suggest the involvement of six distinct pathways in the pathogenesis of this disease. Moreover, we discovered novel associations with prognostic markers that shed light on the biology underlying this clinically heterogeneous disease.

The data led us to several general conclusions. First, like other hematologic cancers, chronic lymphocytic leukemia has a lower rate of somatic mutation than most solid tumors. Second, the rate of nonsynonymous mutation was not strongly affected by therapy. Third, in addition to finding the expected mutations in cell-cycle and DNA-repair pathways, we found
genetic alterations in Notch signaling, inflammatory pathways, and RNA splicing and processing. Fourth, driver mutations showed striking associations with standard prognostic markers, suggesting that particular combinations of genetic alterations may act in concert to drive cancer.
A major surprise was the finding that a core spliceosome component, SF3B1, was mutated in about 15% of the study patients. Further analysis revealed that samples with SF3B1 mutations had enhanced intron retention within two specific transcripts previously shown to be affected by compounds that disrupt SF3b spliceosome function.10,31 Studies of these compounds have suggested that rather than inducing a global change in splicing, SF3b inhibitors alter the splicing of a narrow spectrum of transcripts derived from genes involved in cancer-related processes, including cell-cycle control (p27, CCA2, STK6, and MDM2),31-33 angiogenesis, and apoptosis.34 Our results suggest that SF3B1 mutations lead to mistakes in the splicing of these and other specific transcripts that affect the pathogenesis of chronic lymphocytic leukemia. Ongoing studies will focus on determining how mutations in SF3B1 alter its function in the processing of critical mRNAs.

Since mutations in SF3B1 are highly enriched in patients with del(11q), these mutations may be synergistic with loss of ATM, a hypothesis that is supported by the observation that two patients had point mutations in both ATM and SF3B1 but did not have del(11q). Providing further support for this hypothesis, a recent unbiased functional screen showed that core spliceosome components were required for DNA repair in mammalian cells.35

As illustrated by our findings regarding SF3B1 mutations, identification of coding mutations in chronic lymphocytic leukemia can lead to the development of mechanistic hypotheses, novel prognostic markers, and potential therapeutic...
targeting. In addition, this information provides a starting point for the systematic analyses needed to address several fundamental questions concerning chronic lymphocytic leukemia, including which genes within chromosomal deletions and amplifications are essential, how each mutation alters cellular networks and phenotypes, which combinations of mutations are critical in the development of cancer, and how genetic events in the host may affect the importance of specific mutations and their combinations.

Supported by the Broad Institute through a grant from the National Human Genome Research Institute at the National Institutes of Health (US4HG003067) and by grants from the National Institutes of Health (GM43375, to Drs. Folco and Reed, and K23 CA115602, to Dr. Brown), the National Cancer Institute (SR2CA115043-2), the Melton and Rosenbach Funds (to Dr. Brown), the Blavatnik Family Foundation (to Dr. Wu), the Early Career Physician-Scientist Award of the Howard Hughes Medical Institute (to Dr. Wu), and the Damon Runyon Cancer Research Foundation (CI-38-07, to Dr. Wu). Dr. Brown is a Scholar of the American Society of Hematology and a Scholar in Clinical Research of the Leukemia and Lymphoma Society.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank all members of the Biological Samples Platform, Genetic Analysis Platform, and Genome Sequencing Platform at the Broad Institute, without whom this work would not have been possible; Dan-Avi Landau and Jerome Ritz for their critical reading of an earlier draft of the manuscript; James Rheinwald and Patricia Barron of the Cell Culture Core at the Harvard Skin Disease Research Center for initiating the primary human dermal fibroblast cell lines used in this study; and the Chronic Lymphocytic Leukemia Research Consortium Biorepository for providing information on some of our samples.

REFERENCES


Figure 4. Mutations in SF3B1 and Altered mRNA Splicing.

In Panel A, a Cox multivariable regression model designed to test for clinical factors contributing to the need for earlier initiation of treatment showed that an SF3B1 mutation was an independent predictor of a shorter time to treatment, regardless of the status of several other independent predictive markers. Panel B shows the relative amounts of spliced and unspliced spliceosome target messenger RNA (mRNA) in BRD2 and RIOK3 in normal CD19+ B cells and chronic lymphocytic leukemia B cells with wild-type or mutated SF3B1, as measured by means of a quantitative polymerase-chain-reaction assay. The ratios of unspliced to spliced mRNA were normalized to the percentage of leukemia cells per sample, and comparisons were calculated with the use of the Wilcoxon rank-sum test. CI denotes confidence interval.
34. Massiello A, Roesser JR, Chalfant CE. SAP155 binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 5’ splice site selection of Bcl-x pre-mRNA. PASEB J 2006;20:1680-2.

Copyright © 2011 Massachusetts Medical Society.