Published in final edited form as:
N Engl J Med. 2011 December 29; 365(26): 2497–2506. doi:10.1056/NEJMoa1109016.

SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia

Lili Wang, M.D., Ph.D., **Michael S. Lawrence, Ph.D.**, **Youzhong Wan, Ph.D.**, **Petar Stojanov, B.A.**, **Carrie Sougnez, B.S.**, **Kristen Stevenson, M.S.**, **Lillian Werner, M.S.**, **Andrey Sivachenko, Ph.D.**, **David S. DeLuca, Ph.D.**, **Li Zhang, Ph.D.**, **Wandi Zhang, M.D.**, **Alexander R. Vartanov, B.A.**, **Stacey M. Fernandes, B.S.**, **Natalie R. Goldstein, B.A.**, **Eric G. Folco, Ph.D.**, **Kristian Cibulskis, B.S.**, **Bethany Tesar, M.S.**, **Quinlan L. Sievers, B.A.**, **Erica Shefler, B.S.**, **Stacey Gabriel, Ph.D.**, **Nir Hacohen, Ph.D.**, **Robin Reed, Ph.D.**, **Matthew Meyerson, M.D., Ph.D.**, **Todd R. Golub, M.D.**, **Eric S. Lander, Ph.D.**, **Donna Neuberg, Sc.D.**, **Jennifer R. Brown, M.D., Ph.D.**, **Gad Getz, Ph.D.**, and **Catherine J. Wu, M.D.**
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.).

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 are 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.

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.

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

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).[2,3] 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.[4-6] A more powerful approach to the process of initial discovery is to sequence a much larger set of samples.[7,8] This 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.[7,9] 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.[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.[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).[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, $P<5\times10^{-6}$) (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.[6,16-18] We found 15 *TP53* mutations in 14 of 91 patients (15%; $Q\leq6.3\times10^{-8}$), most of which were localized to the DNA-binding domain that is critical for its tumor-suppressor activity[17] (Fig. 3A in the Supplementary Appendix). In 8 patients (9%), we detected 9 *ATM* mutations ($Q\leq1.1\times10^{-5}$) scattered across this large gene, including in regions where mutation has been associated with defective DNA repair in patients with chronic lymphocytic leukemia.[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).[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.[5,6] This mutation is associated with unmutated *IGHV* and a poor prognosis,[5,6] 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.[20] *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,[20] 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.[21,22] Its targets include important oncoproteins such as Notch1, c-Myc, c-Jun, cyclin E1, and MCL1.[21,22] Two of the four mutations in *FBXW7* cause constitutive Notch signaling in T-cell acute lymphoblastic leukemia.[23] *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.[24] *DDX3X* interacts directly with *XPO1*,[24] which was recently reported as being mutated in 2.4% of patients with chronic lymphocytic leukemia.[6] *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.[25,26] 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.[27]

We validated the three most frequently recurring mutations — *SF3B1*–K700E, *MYD88*–L265P, and *NOTCH1*–P2514fs — 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*[23]), 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[28] (Fig. 2, and Fig. 4 and Table 6 in the Supplementary Appendix). Although the frequency of mutations in these genes does not reach statistical significance, whether they are considered alone or as a set, it might do so in a larger collection of samples. On the other hand, in our cohort, 19 of 59 genes classified as members of the Wnt signaling pathway, which has been implicated in chronic lymphocytic leukemia in gene-expression studies,[29,30] were mutated. Although no individual gene had a frequency of mutation that reached significance, the genes involved in the Wnt pathway, as a set, had a high mutation frequency (P = 0.048) (Fig. 2).

ASSOCIATION OF DRIVER MUTATIONS WITH DISTINCT CLINICAL GROUPS

To examine the association between driver mutations and particular clinical features, we assessed the cytogenetic aberrations associated with chronic lymphocytic leukemia and *IGHV* mutation status in samples harboring mutations in the nine genes with significant mutation frequencies. We ordered the samples in accordance with their cytogenetic features on fluorescence in situ hybridization (FISH), using an established model of hierarchical risk[3]; del(13q) was associated with the most favorable prognosis when present alone, trisomy 12 was associated with a less-favorable prognosis, and del(11q) and del(17p) were both associated with aggressive disease that was refractory to chemotherapy (Fig. 3, and Tables 7 and 8 in the Supplementary Appendix).

The distinct prognostic implications of these cytogenetic abnormalities suggest that each abnormality may reflect a distinct pathogenesis. Our data show that different driver mutations are associated with different key abnormalities detected on FISH, providing support for this hypothesis. Most of the *TP53* mutations (11 of 17) were present in samples that also harbored del(17p) (P<0.001), resulting in homozygous p53 inactivation; this finding is consistent with previous observations.[17] Mutations in *ATM*, which lies in the minimally deleted region of chromosome 11q, were marginally associated with del(11q) — that is, in 4 of 22 del(11q) samples (P = 0.09). Strikingly, mutations in *SF3B1* were associated with del(11q) — 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,[5,6] *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,[5,6] 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,[19] 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), independently 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.[10] 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, 2.0:1 vs. 0.55:1 [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.[5,6,14] 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,[14] chronic lymphocytic leukemia has a lower rate of somatic mutation than most solid tumors.[5,6] 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 targets. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by the Broad Institute through a grant from the National Human Genome Research Institute at the National Institutes of Health (U54HG003067) and by grants from the National Institutes of Health (GM43375, to Drs. Folco and Reed, and K23 CA115682, to Dr. Brown), the National Cancer Institute (5R21CA115043-2), 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.

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

1. Zenz T, Mertens D, Küppers R, Döhner H. Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat Rev Cancer. 2010; 10:37–50. [PubMed: 19956173]
2. Rassenti LZ, Huynh L, Toy TL, et al. ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. N Engl J Med. 2004; 351:893–901. [PubMed: 15329427]
3. Döhner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med. 2000; 343:1910–6. [PubMed: 11136261]
4. Tiacci E, Trifonov V, Schiavoni G, et al. *BRAF* mutations in hairy-cell leukemia. N Engl J Med. 2011; 364:2305–15. [PubMed: 21663470]

5. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med. 2011; 208:1389–401. [PubMed: 21670202]
6. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011; 475:101–5. [PubMed: 21642962]
7. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471:467–72. [PubMed: 21430775]
8. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–15. [PubMed: 21720365]
9. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–8. [PubMed: 21478889]
10. Kotake Y, Sagane K, Owa T, et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol. 2007; 3:570–5. [PubMed: 17643112]
11. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214–20. [PubMed: 21307934]
12. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–6. [PubMed: 21221095]
13. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009; 361:1058–66. [PubMed: 19657110]
14. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008; 456:66–72. [PubMed: 18987736]
15. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–8. [PubMed: 18772890]
16. Austen B, Powell JE, Alvi A, et al. Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. Blood. 2005; 106:3175–82. [PubMed: 16014569]
17. Zenz T, Eichhorst B, Busch R, et al. TP53 mutation and survival in chronic lymphocytic leukemia. J Clin Oncol. 2010; 28:4473–9. [PubMed: 20697090]
18. Trbusek M, Smardova J, Malcikova J, et al. Missense mutations located in structural p53 DNA-binding motifs are associated with extremely poor survival in chronic lymphocytic leukemia. J Clin Oncol. 2011; 29:2703–8. [PubMed: 21606432]
19. Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. Nature. 2011; 470:115–9. [PubMed: 21179087]
20. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009; 136:701–18. [PubMed: 19239890]
21. Yada M, Hatakeyama S, Kamura T, et al. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. EMBO J. 2004; 23:2116–25. [PubMed: 15103331]
22. Babaei-Jadidi R, Li N, Saadeddin A, et al. FBXW7 influences murine intestinal homeostasis and cancer, targeting Notch, Jun, and DEK for degradation. J Exp Med. 2011; 208:295–312. [PubMed: 21282377]
23. O'Neil J, Grim J, Strack P, et al. FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. J Exp Med. 2007; 204:1813–24. [PubMed: 17646409]
24. Rosner A, Rinkevich B. The DDX3 subfamily of the DEAD box helicases: divergent roles as unveiled by studying different organisms and in vitro assays. Curr Med Chem. 2007; 14:2517–25. [PubMed: 17979704]
25. Pepper C, Thomas A, Hoy T, Milligan D, Bentley P, Fegan C. The vitamin D3 analog EB1089 induces apoptosis via a p53-independent mechanism involving p38 MAP kinase activation and suppression of ERK activity in B-cell chronic lymphocytic leukemia cells in vitro. Blood. 2003; 101:2454–60. [PubMed: 12446453]
26. Muzio M, Apollonio B, Scielzo C, et al. Constitutive activation of distinct BCR-signaling pathways in a subset of CLL patients: a molecular signature of anergy. Blood. 2008; 112:188–95. [PubMed: 18292287]

27. Lee MG, Wynder C, Cooch N, Shiekhattar R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature. 2005; 437:432–5. [PubMed: 16079794]

28. Subramanian A, Tamayo P, Mootha VK, 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. [PubMed: 16199517]

29. Gutierrez A Jr, Tschumper RC, Wu X, et al. LEF-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal B-cell lymphocytosis. Blood. 2010; 116:2975–83. [PubMed: 20595513]

30. Klein U, Tu Y, Stolovitzky GA, et al. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J Exp Med. 2001; 194:1625–38. [PubMed: 11733577]

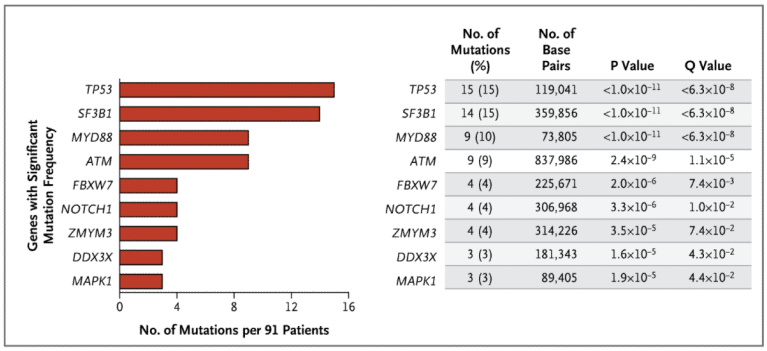
31. Kaida D, Motoyoshi H, Tashiro E, et al. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol. 2007; 3:576–83. [PubMed: 17643111]

32. Corrionero A, Miñana B, Valcárcel J. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev. 2011; 25:445–59. [PubMed: 21363963]

33. Fan L, Lagisetti C, Edwards CC, Webb TR, Potter PM. Sudemycins, novel small molecule analogues of FR901464, induce alternative gene splicing. ACS Chem Biol. 2011; 6:582–9. [PubMed: 21344922]

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. FASEB J. 2006; 20:1680–2. [PubMed: 16790528]

35. Paulsen RD, Soni DV, Wollman R, et al. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol Cell. 2009; 35:228–39. [PubMed: 19647519]



	No. of Mutations (%)	No. of Base Pairs	P Value	Q Value
TP53	15 (15)	119,041	$<1.0\times10^{-11}$	$<6.3\times10^{-8}$
SF3B1	14 (15)	359,856	$<1.0\times10^{-11}$	$<6.3\times10^{-8}$
MYD88	9 (10)	73,805	$<1.0\times10^{-11}$	$<6.3\times10^{-8}$
ATM	9 (9)	837,986	2.4×10^{-9}	1.1×10^{-5}
FBXW7	4 (4)	225,671	2.0×10^{-6}	7.4×10^{-3}
NOTCH1	4 (4)	306,968	3.3×10^{-6}	1.0×10^{-2}
ZMYM3	4 (4)	314,226	3.5×10^{-5}	7.4×10^{-2}
DDX3X	3 (3)	181,343	1.6×10^{-5}	4.3×10^{-2}
MAPK1	3 (3)	89,405	1.9×10^{-5}	4.4×10^{-2}

Figure 1. Genes with Significant Mutation Frequencies in 91 Patients with Chronic Lymphocytic Leukemia

The figure shows the number of mutations per gene that were found at a significant frequency, the percentage of patients who had each mutated gene, and for each gene, the total territory in numbers of base pairs with sufficient sequencing coverage across normal and malignant samples from the 91 patients. The P values and Q values were calculated by comparing the probability of the observed constellation of mutations with the background mutation rates across the data set.

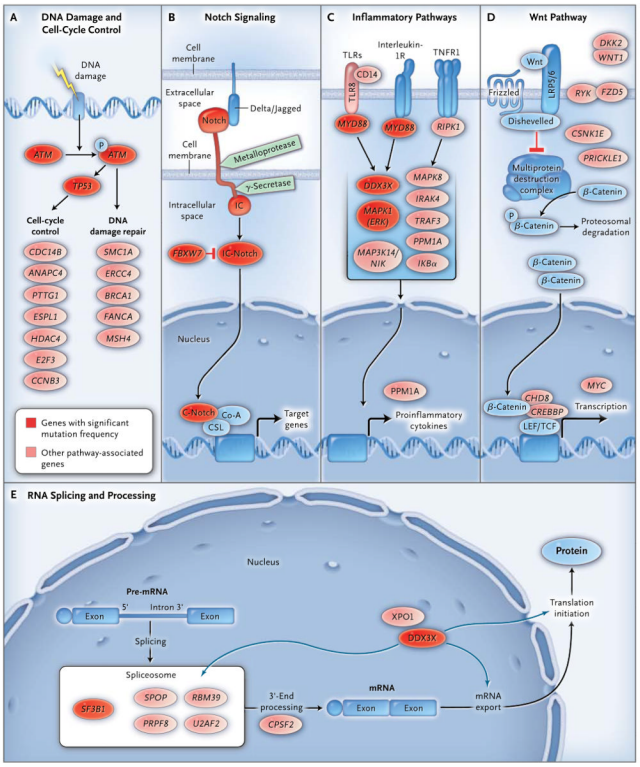


Figure 2. Core Signaling Pathways in Chronic Lymphocytic Leukemia

The nine genes with significant mutation frequencies fall into five core signaling pathways, in which the genes play well-established roles: DNA damage repair and cell-cycle control (Panel A), Notch signaling (Panel B), inflammatory pathways (Panel C), Wnt signaling (Panel D), and RNA splicing and processing (Panel E). Genes with significant mutation frequencies are shown in red, and genes with mutations that are in a signaling pathway related to chronic lymphocytic leukemia are shown in pink. A list of additional pathway-associated genes with mutations is provided in Table 6 in the Supplementary Appendix. Co-A denotes coenzyme A, TLR toll-like receptor, and TNFR tumor necrosis factor receptor.

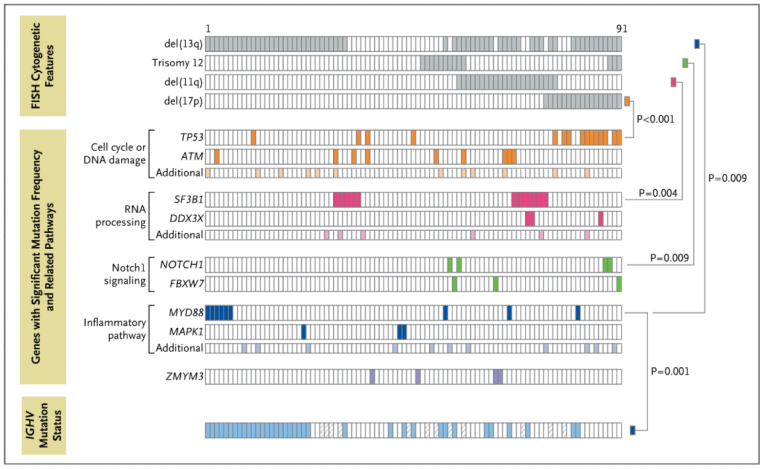


Figure 3. Associations between Gene Mutations and Clinical Characteristics

Samples from the 91 patients in the study were sorted by means of fluorescence in situ hybridization (FISH), with the use of a model of hierarchical risk established by Döhner et al.[3] Samples were scored for the presence or absence of mutations in the nine genes with significant mutation frequencies (darker colors), as well as for mutations in additional pathway-associated genes (lighter colors) and in the immunoglobulin heavy-chain variable (*IGHV*) mutation status (blue indicates mutation, white no mutation, and hatched unknown status). (A list of additional mutated pathway-associated genes is provided in Table 6 in the Supplementary Appendix.) Associations between gene mutation status and FISH cytogenetic features or IGHV status were calculated with the use of Fisher's exact test and corrected for multiple hypothesis testing (Q≤0.1 for all comparisons shown).

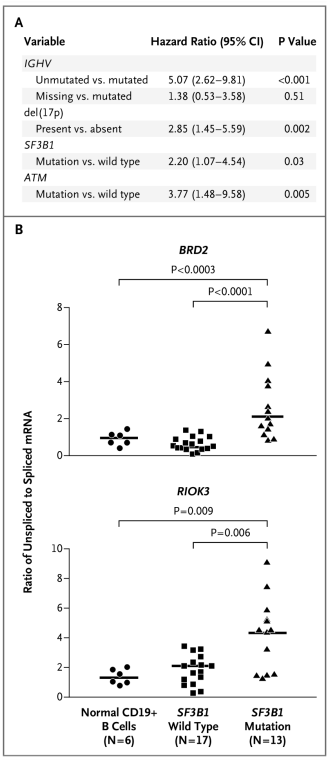


Figure 4. Mutations in *SF3B1* and Altered mRNA Splicing

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.

Table 1

Summary Metrics of Whole-Genome and Whole-Exome Sequencing Studies.*

Variable	Value†
Whole genomes	
No.	3
Bases covered per genome — %	70
Genome coverage	
CLL samples	38x
Normal samples	33x
Whole exomes	
No.	88
Bases covered per exome — %	81
Exome coverage	
CLL samples	132x
Normal samples	146x
Nonsynonymous mutations	
No. of mutations per megabase	0.7±0.36
Coding mutations — no. (range)	20 (2–76)
Synonymous mutations	
No. of mutations per megabase	0.2±0.16
Coding mutations — no. (range)	5.8 (0–31)

*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.