SF3B1 mutations in chronic lymphocytic leukemia
Youzhong Wan1,2 and Catherine J. Wu1-3

1Cancer Vaccine Center and 2Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; and 3Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

SF3B1 is a critical component of the splicing machinery, which catalyzes the removal of introns from precursor messenger RNA (mRNA). Next-generation sequencing studies have identified mutations in SF3B1 in chronic lymphocytic leukemia (CLL) at high frequency. In CLL, SF3B1 mutation is associated with more aggressive disease and shorter survival, and recent studies suggest that it can be incorporated into prognostic schema to improve the prediction of disease progression. Mutations in SF3B1 are predominantly subclonal genetic events in CLL, and hence are likely later events in the progression of CLL. Evidence of altered pre-mRNA splicing has been detected in CLL cases with SF3B1 mutations. Although the causative link between SF3B1 mutation and CLL pathogenesis remains unclear, several lines of evidence suggest SF3B1 mutation might be linked to genomic stability and epigenetic modification. (Blood. 2013;121(23):4627-4634)

Introduction

The advent of next-generation sequencing has been transformative for understanding the molecular underpinnings of cancer, including the blood malignancies. One of the most surprising findings arising from the recent next-generation sequencing data has been the discovery of mutated splicing factor 3b1 (SF3B1) as a putative candidate driver gene of the common lymphoid malignancy, chronic lymphocytic leukemia (CLL). This discovery has been particularly revealing in CLL because it had been long suspected that distinct subgroups of the disease exist based on its highly variable clinical course, ranging from indolent disease in some patients to a rapidly fatal course despite aggressive therapy in others.1 By now, massively parallel DNA sequencing and targeted DNA sequencing have been undertaken in >250 and 3800 CLL cases, respectively, and have consistently identified SF3B1 as one of the most frequently mutated genes in CLL in 5% to 18% of patients, depending on the composition of the various cohorts in terms of the proportion of cases at presentation, at the time of treatment or at disease progression.2-4 Contemporaneously, a high frequency of mutation in SF3B1 was identified in a variant of myelodysplastic syndrome (MDS) (refractory anemia with ringed sideroblasts), suggesting that altered splicing may be an important pathogenic mechanism in blood malignancies and cancer. Herein, we review the emerging understanding of the impact of SF3B1 mutations on the clinical outcome of CLL patients and stepwise transformation of CLL, and discuss the potential mechanisms by which SF3B1 mutation is linked to the pathobiology of CLL.

Identification of mutated SF3B1 as a putative CLL driver

Initial studies to characterize the mutational landscape of CLL using whole-genome sequencing and whole-exome sequencing (WES) unexpectedly revealed a high frequency of nonsilent heterozygous mutations in CLL.2-4 Subsequent studies in CLL, summarized in Table 1,3-5-8 have confirmed that mutated SF3B1 consistently ranks among the most commonly identified somatic mutations in CLL, together with other putative CLL-associated driver mutations in TP53, ATM, MYD88, BIRC3, and NOTCH1.

Because SF3B1 is a critical component of the splicing machinery, the discovery of a high frequency of SF3B1 mutations in CLL has highlighted the essential pathway (in which the introns from precursor messenger RNAs [pre-mRNAs] are removed) as involved in CLL. Splicing is catalyzed by the spliceosome, which consists of a set of small nuclear ribonucleoprotein particles (snRNPs), namely U1, U2, U4, U5, and U6, as well as numerous splicing factors.9 SF3B1 is an essential component of the U2 snRNP.

The vast majority of mutations in SF3B1 are localized to its highly conserved C-terminal domain, which comprises 22 Huntingdon Elongation Factor 3 PR65/A TOR (HEAT) repeats. As shown in Figure 1, most mutations have been detected between the fifth and the eighth HEAT repeats (encoded by exons 14-16), with K700E as the most frequently mutated site (50% of reported cases). A second hot spot has been detected at G742 (19%), with additional common sites at K666 (12%) and H662 (4%). All of these hot-spot mutations occur within highly conserved sequences, even to Caenorhabditis elegans, suggesting that they occur in functionally important regions. As aforementioned, SF3B1 mutations have been also predominantly identified in MDS (76% of patients with ringed sideroblasts), and in breast (2%) and pancreatic cancers (3%).10-15 At sites similar to those in CLL (ie, most commonly at K700E, Figure 1). Of note, the G742D mutation is very rare in cancers other than CLL. A separate recent exome sequencing study in uveal melanoma reported an unexpectedly high frequency of mutations (18.6% of 102 cases) occurring exclusively at codon 625 of the SF3B1 gene.15 Altogether, these observations demonstrate that mutations at different sites of SF3B1 can be differentially enriched depending on the specific cancer type, and suggest these mutations might have different functional impact on the respective diseases. For CLL, the striking positional clustering of SF3B1 mutations suggests that they are positively selected during CLL pathogenesis, and therefore are most likely driver mutations in CLL. Consistent with the
distribution of affected cancers, SF3B1 is ubiquitously expressed, although at higher levels in cells from hematopoietic lineages (Figure 2A).16 Compared with normal B cells, CLL cells have increased SF3B1 expression (Figure 2B).3 DNA hypomethylation in the gene body of SF3B1 has been observed in CLL (Figure 2C),17 although it remains unclear whether this hypomethylation is related to the increased SF3B1 expression in CLL.

Table 1. Incidence of SF3B1 mutation in CLL varies depending on disease status

<table>
<thead>
<tr>
<th>Disease status</th>
<th>Incidence (%)</th>
<th>No. of subjects in cohort</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newly diagnosed</td>
<td>3.6</td>
<td>360</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>301</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>9.3</td>
<td>1124</td>
<td>6</td>
</tr>
<tr>
<td>Firstline treatment</td>
<td>17</td>
<td>437</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>18.4</td>
<td>621</td>
<td>8</td>
</tr>
<tr>
<td>Treatment refractory</td>
<td>17</td>
<td>59</td>
<td>3</td>
</tr>
</tbody>
</table>

Cases with mutated SF3B1 at the DNA level also have detectable expression of the mutated transcript at the mRNA level (Figure 2D).

SF3B1 mutation is associated with poorer clinical outcome

Two of the initial studies that reported associations with mutated SF3B1 and CLL were conducted using case cohorts that were heterogeneous with respect to their clinical characteristics, and which were collected at variable points in disease course.2,4 These studies found SF3B1 mutations to associate with poorer clinical outcome. Wang et al reported that SF3B1 mutation is significantly associated with an earlier time-to-treatment initiation,2 while Quesada et al observed associations with faster disease progression and shorter overall survival.4 Importantly, both studies found SF3B1 mutation to provide independent prognostic information from other known CLL prognostic markers, such as ZAP70.
Rossi et al identified that SF3B1 mutations were associated with more advanced disease, as 17% of fludarabine-refractory cases had SF3B1 mutations, compared with 5% of cases at diagnosis. More recent studies have provided a consistent picture that SF3B1 mutation is associated with more aggressive disease and poorer clinical outcome (summarized in Table 1 and Table 2). Greco et al found that SF3B1 mutations were present in only 1.5% of individuals with the CLL precursor condition, monoclonal B-cell lymphocytosis. Consistent with these reports, Mansouri et al examined 360 newly diagnosed Scandinavian patients, and identified only a small percentage of cases with mutated SF3B1 (3.6%). However, multivariate analysis revealed that those patients harboring SF3B1 mutation demonstrated significantly worse overall survival and earlier time to treatment. Independently, a large study of 1124 newly diagnosed CLL patients from the German CLL Study Group (GCLLSG) has confirmed that SF3B1 mutations associate with a shorter time to first therapy, and with unmutated IGHV status. An analysis of ~500 cases from a prospective and controlled clinical trial (the UK LRF CLL4 trial) demonstrated a high rate of SF3B1 mutation (17%) in patients requiring frontline treatment, and that presence of mutation associated with significantly worse overall survival and with a trend toward shorter progression-free survival. Finally, a recent analysis of cases enrolled in the GCLLSG CLL8 trial, in which patients were randomized to frontline therapy with either the combination of fludarabine and cyclophosphamide alone or with rituximab demonstrated that while SF3B1 mutations did not impact response to therapy, they were associated with significantly lower progression-free survival and a trend toward inferior overall survival.

Taken together, these findings suggest that a molecular marker such as SF3B1 mutation can potentially help refine the existing framework for prognostication in CLL. The current gold standard uses common fluorescence in situ hybridization (FISH) cytogenetic abnormalities, and includes favorable cytogenetic abnormalities such as del(13q), and poorer features such as trisomy 12, del(11q), and del(17p). Recently, Rossi et al evaluated the potential of mutations in SF3B1, NOTCH1, and BIRC3 to more finely stratify patients into different risk groups beyond what could be achieved with FISH cytogenetics alone. In this integrated analysis of mutations and cytogenetics, patients with SF3B1 mutation and/or NOTCH1 mutation and/or del(11q) were classified into the intermediate-risk group with 37% 10-year survival. Although SF3B1 mutation does not confer the same degree of poor prognosis as TP53 mutation,

**Figure 2.** SF3B1 expression and methylation in normal and CLL cells. (A) SF3B1 relative expression in different human tissues. (B) SF3B1 relative expression in normal B-cell subpopulations and CLL samples. CB, centroblast; CC, centrocyte. Data generated from Affymetrix HG-U133Plus2 arrays. (C) DNA methylation on SF3B1 in CLL samples with or without SF3B1 mutation and normal B cells. (D) Targeted pyrosequencing of a SF3B1 K700E site of cDNA from normal peripheral blood mononuclear cells (top, NL PBMC) compared with cDNA from a CLL sample with known K700E mutation in SF3B1 (bottom). K700E mutation is generated from an A2098G transition on the sense strand; this pyrosequencing assay was designed to detect the T to C transition at the corresponding site on the antisense strand. cDNA, complementary DNA. Panel A adapted from the Novartis Gene Atlas with permission. Panel B adapted from Rossi et al with permission. Panel C adapted from Kulis et al with permission.
Table 2. Associations of SF3B1 mutation in patients with CLL and MDS

<table>
<thead>
<tr>
<th>Disease</th>
<th>Initial discovery of association, reference</th>
<th>Association with a disease subtype</th>
<th>Further clinical associations</th>
<th>Clonal status</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLL</td>
<td>2</td>
<td>del(11q)²,⁶</td>
<td>• Shorter treatment-free or overall survival²,⁴,⁶,⁷,¹⁸</td>
<td>Predominantly subclonal:</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>• Faster disease progression²,⁴,⁵,¹⁸</td>
<td>likely a later event¹⁹</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>• Fludarabine refractoriness²</td>
<td></td>
</tr>
<tr>
<td>MDS</td>
<td>10</td>
<td>Refractory anemia with ring sideroblasts¹⁰,¹¹,②₀-②²</td>
<td>• Longer event-free or overall survival¹¹,②₁-②₃</td>
<td>Predominantly clonal:</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td></td>
<td>• Lower risk of evolution into AML¹¹</td>
<td>likely an earlier event¹⁹</td>
</tr>
</tbody>
</table>

these results overall demonstrate that SF3B1 mutation status can improve the accuracy of prediction of disease progression in combination with conventional biologic markers.²⁵ Interestingly, a recent analysis shows that CLL patients with poor prognostic markers, such as SF3B1, TP53, and NOTCH1 mutation, do not show decreased 6-year survival after reduced-intensity allogeneic hematopoietic stem cell transplantation (HSCT), demonstrating these mutations do not have high impact on the long-term disease control of HSCT.²⁶

SF3B1 and the temporal order of mutations in CLL

The genetic studies described above consistently demonstrate that mutated SF3B1 is associated with poor prognosis. We considered that the common cytogenetic abnormalities in CLL can often be present in only leukemia cell subpopulations and yet still hold prognostic significance. For example, up to 80% of CLL samples have detectable FISH abnormalities, and multiple FISH abnormalities often occur at varying frequencies within the same bulk sample, implying the coexistence of subclones harboring different chromosome-level changes. Acknowledgment of the existence of subpopulations harboring prognostically significant cytogenetic abnormalities raises the question of whether and how specific genetic alterations, including driving somatic mutations, such as SF3B1, that are present within subclones, can impact clinical course.

Recently, whole-genome sequencing and WES studies have provided the opportunity to address this question, as subpopulations can be detected and tracked at unprecedented resolution through identification and clustering of mutations with similar allelic frequencies.²⁷,²⁸ Using this approach, a striking degree of intratumoral heterogeneity has been detected across blood malignancies.²⁹,³⁰ In the case of CLL, Schuh et al performed whole-genome sequencing on sequential samples collected from 3 CLL cases over the course of years and following various treatments to investigate clonal evolution of somatic mutations in CLL. Their longitudinal analysis revealed that the clonal evolution patterns between individual patients are quite heterogeneous.³¹ Specifically, a remarkable equilibrium among leukemia subpopulations over the course of years, even in the presence of cytoreductive therapy, was observed in some cases, while acquisition of new mutations and change in the percentage of affected cells over time was observed in others. These data suggest that subclonal mutations expand over time as a function of their fitness-integrating intrinsic factors (eg, proliferation and apoptosis) and extrinsic pressures (eg, interclonal competition and therapy).³²

In a separate analysis, Landau et al developed an analytic approach to enable the study of clonal evolution of CLL using WES data. In this study, WES and copy number data were integrated to estimate the fraction of cancer cells harboring each somatic mutation, and this approach was applied to study clonal evolution of somatic mutations in 149 CLL samples.¹⁹

Analysis of the clonal and subclonal mutations in this large CLL series by Landau et al revealed that SF3B1 is typically a subclonal mutation and hence likely a later event in the progression of CLL.¹⁹ This result was inferred by comparing the aggregate frequencies of recurrent driver mutations across this large series of cases. Focusing on a set of 24 CLL driver events that were derived by performing a significance analysis for frequent somatic single nucleotide variants and recurrent somatic copy number alterations, Landau et al inferred the fraction of cancer cells and the percentage of clonality of these driver events across 149 samples.¹⁹ As shown in Figure 3A,¹⁹ 3 driver mutations (MYD88 [n = 12 samples], trisomy 12 [n = 24], and heterozygous del(13q) [n = 62]) were cloned in 80% to 100% of samples, a significantly higher level than for other driver events (q < 0.1), suggesting that they arise earlier in typical CLL development. Moreover, in CLL samples that harbored 1 of these 3 early driver mutations, additional driver alterations were found at either similar or lower cancer cell fraction (CCF) suggesting that the proposed order of events holds across individual patients, not only in aggregate. Other drivers (eg, ATM, TP53, and SF3B1) were often observed at subclonal frequencies, indicating that they often arise later in leukemic development. Analysis of co-occurrence of the 19 SF3B1 mutations within this cohort with other driver alterations revealed that SF3B1 was subclonal in 10 of 19 cases (53%). In 7 of 10 cases (70%), subclonal SF3B1 was present together with clonal del(13q), and in 3 of 10 cases (30%), with clonal del(11q) (Figure 3B). These data together suggest that mutations that selectively affect B cells may contribute more to the initiation of disease and precede selection of more generic cancer drivers that underlie disease progression.

Within this conceptual framework, SF3B1 mutations are mostly subclonal and these results suggest that SF3B1 mutations are not typically initiators of CLL, but rather promoters of disease progression. As such, it is likely that SF3B1 mutations cooperate with other mutations to affect CLL. Consistent with this framework, Landau et al assessed the evolution of somatic mutations in 18 patients, in which data from 2 distant time points were available.¹⁹ Clonal evolution was observed in 11 of 18 patients (10 of 12 who received intervening treatment, but only 1 of 6 without intervening treatment, P = .012) and confirmed that subclonal mutations (eg, del(11q), SF3B1 and TP53) shifted toward clonality over time (Figure 3C).

One notable observation by Landau et al from the analysis of these 18 longitudinally studied cases was that the presence of subclonal driver mutations (defined by CLL significance analysis)³⁰ or by mutations within highly conserved sites of genes within the Cancer Gene Census³¹ in the pretreatment sample appeared to be indicative of clonal evolution. Further analysis of the 149 patients
revealed that presence of subclonal (but not clonal) driver mutations, such as SF3B1, was indeed associated with poorer clinical outcome, and was independent of confounding factors such as IGHV mutation status, the identity of driver mutations and presence of cytogenetic abnormalities.19 Thus, an important finding with potential clinical implications is that detection of subclonal (but not clonal) driver mutations such as SF3B1 marks ongoing clonal evolution, which is in turn associated with more aggressive disease in CLL.

Potential functional effects of SF3B1 mutations in CLL

At present, it remains unclear how SF3B1 mutation impacts CLL at the cellular level. However, it is constructive to consider the known cellular functions of SF3B1, as these might offer important clues. SF3B1 is an essential component of U2 snRNP, which contains U2 small nuclear RNA and ~20 proteins that associate in a highly organized complex. As shown in Figure 4A, SF3B1 directly interacts with both the 5' and 3' sides of the branch point sequence (BPS) (Figure 4A). In addition to the components within the U2 snRNP, SF3B1 also interacts with other proteins in the spliceosome. For instance, SF3B1 directly interacts with U2AF2, which binds to the polypyrimidine tract (PY tract) adjacent to the SF3B1 binding site.34,35 It has been suggested that such multiple interactions are crucial for the recognition and stabilization of the spliceosome at the 3' splice sites.36 In fact, the formation of a duplex between the U2 small nuclear RNA and the BPS is an early step in the splicing pathway (Figure 4A), and results in the bulging of a conserved adenosine in the BPS. This adenosine is used as the nucleophile for the first catalytic step of the splicing reaction (Figure 4B).37 The rough structure of the C-terminal 22 HEAT repeats domain of SF3B1 has been characterized by single-particle electron cryomicroscopy techniques, and these studies suggest that the HEAT repeats region might undergo significant conformational change during the formation of the U2 snRNP complex37,38 Quesada et al constructed a tentative homology model of the C-terminal domain of SF3B1.4 According to this model, the mutation hot spots in SF3B1 seem to cluster at the inner surfaces of the structure, at the supposed sites of interaction with RNA and cofactors. These characterizations suggest that mutation in SF3B1 could possibly mediate alteration in its normal function through change in the physical interactions of SF3B1 with its binding partners.

Consistent with the critical role of SF3B1 in splicing, evidence of altered splicing has been detected in CLL cases with SF3B1 mutations. Wang et al reported intron retention in known target genes of spliceosome inhibition.2,39,40 Quesada et al compared splicing in CLL cases with and without mutated SF3B1 by exon arrays and RNA sequencing. They detected relatively few transcripts with altered splicing in CLL cases with mutated SF3B1, suggesting that the SF3B1 mutation does not exert a global effect on splicing, but rather only affects few specific target transcripts.4 One of the identified
splicing variants associated with mutant SF3B1 affects the FOXP1 forkhead transcription factor, and expression of truncated FOXP1 variants has been implicated in diffuse large B-cell lymphoma. This observation suggests that the splicing variants associated with SF3B1 mutation might affect factors involved in malignant transformation of B cells.

One hint of the potential impact of SF3B1 mutation on CLL comes from the observation that SF3B1 mutation significantly co-occurs with del(11q). As the minimal deleted region of 11q(del) contains the ATM gene, one intriguing possibility is that SF3B1 mutation could alter responses to DNA damage. In support of this idea, te Raa et al demonstrated that CLL samples with SF3B1 mutation show decreased expression of P53 target genes in response to irradiation, resembling CLL samples with ATM mutation or del(11q). Of note, a large unbiased small interfering RNA screen identified a group of splicing factors, including multiple components in U2 snRNP, as key factors involved in DNA damage response. These results implicate a general role of splicing in maintenance of genomic stability.

These observations suggest that SF3B1 mutation might have broader functions beyond RNA splicing alone. An open question is whether such effects are mediated through splicing variant intermediaries or other direct effects (schematically shown in Figure 5). For example, gene expression involves a highly complex network of interactions, and splicing is coupled with other steps in gene expression, such as transcription, capping, polyadenylation, and mRNA nuclear export. Thus, mutations in splicing factors, such as SF3B1, might affect gene expression through mechanisms other than splicing itself. Recently, Ramsay et al reported frequent somatic mutations in components of the RNA processing machinery in CLL, raising the question of whether disruption of not only splicing but also the entire RNA processing pathway might be involved in CLL. Finally, heterozygous Sf3b1 knockout mice show decreased interaction with the Polycomb protein Zfp144 and manifest defects in bone development concomitant with ectopic expression of Hox genes. These associations suggest a potential role of SF3B1 in epigenetic modification. Recently, Kulis et al reported global DNA hypomethylation in CLL, and demonstrated a significant role of epigenetic alteration in the disease. It is conceivable that SF3B1 mutation might also lead to some epigenetic modifications, which impacts CLL, but this has yet to be fully demonstrated.

**SF3B1 mutations in MDS**

SF3B1 mutation in MDS has been extensively reviewed recently. In striking contrast to CLL, SF3B1 mutations in MDS are
associated with favorable prognosis and appear to be mostly clonal events in MDS (Table 2). The basis of these differences between MDS and CLL are currently unknown. SF3B1 mutations are highly prevalent in a mild MDS variant with presence of ring sideroblasts. Visconte et al reported that bone marrow aspirates of Sf3b1 heterozygous knockout mice show ring sideroblasts, suggesting that SF3B1 haploinsufficiency leads to the formation of ring sideroblasts. Despite this finding, whether SF3B1 mutations cause loss of function in MDS remains unclear.

Conclusion

SF3B1 mutation is associated with rapid disease progression and shorter survival in CLL, which can be used as a prognosis marker to improve the prediction of disease progression. In CLL, SF3B1 mutations are mostly subclonal events, and therefore likely involved in disease progression. Dedicated efforts are now required to investigate the causative link between SF3B1 mutation and CLL pathogenesis. Mutated SF3B1 might be a good therapeutic target for the treatment of CLL if the underlying mechanism of the functional impact of SF3B1 in CLL can be clarified.

Acknowledgments

The authors thank Robin Reed and Angela Brooks for critical reading of this manuscript. They also thank Peter Stojanov for analysis data. Y.W. is supported by a fellowship from the Leukemia and Lymphoma Society. C.J.W. acknowledges support from the Blavatnik Family Foundation, the Lymphoma Research Foundation, and the National Institutes of Health National Heart, Lung, and Blood Institute (1RO1HL103532-01, 1RO1HL116452-01) and National Cancer Institute (1R01CA155010-01A1), and is a recipient of a Leukemia Lymphoma Translational Research Program Award and an American Association of Cancer Research: Stand Up To Cancer Innovative Research Grant.

Authorship

Contribution: Y.W. and C.J.W. conceived the article and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Catherine J. Wu, Dana-Farber Cancer Institute, Harvard Institute of Medicine, Room 420, 77 Avenue Louis Pasteur, Boston, MA 02115; e-mail: cwu@partners.org.

References


