Molecular pathogenesis of chronic lymphocytic leukemia

Gianluca Gaidano,1 Robin Foà,2 and Riccardo Dalla-Favera3,4,5,6

1Division of Hematology, Department of Translational Medicine, Amedeo Avogadro University of Eastern Piedmont, Novara, Italy. 2Division of Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy. 3Institute for Cancer Genetics, Department of Pathology and Cell Biology, 4Department of Genetics and Development, and 5Department of Microbiology and Immunology, Columbia University, New York, New York, USA.

Chronic lymphocytic leukemia (CLL) is the most common leukemia in adults. Here, we highlight important genetic alterations that contribute to tumorigenesis, clinical progression, and chemorefractoriness of CLL. All CLLs share a common gene expression profile that suggests derivation from antigen-experienced B cells, a model supported by frequent B cell receptor repertoire skewing and stereotypy. Many CLL patients carry mutated immunoglobulin heavy-chain variable genes, while approximately 35% harbor unmutated IgV genes, which are associated with an inferior outcome. Deletion of chromosome 13q14, which is the most common genetic mutation at diagnosis, is considered an initiating lesion that frequently results in disruption of the tumor suppressor locus DLEU2/MIR15A/MIR16A. Next-generation sequencing has revealed additional recurrent genetic lesions that are implicated in CLL pathogenesis. These advancements in the molecular genetics of CLL have important implications for stratifying treatment based on molecular prognosticators and for targeted therapy.

Introduction

Chronic lymphocytic leukemia (CLL) is the most common leukemia in adults (1, 2). Historically, CLL was viewed as a tumor caused by the accumulation of long-lived but mainly resting lymphocytes with a very low proliferation index (3). However, this model has been challenged in recent years. Heavy-water experiments have shown that CLL contains a small fraction of actively proliferating cells, with approximately 2% of cells newly generated each day (4). Though lymphocytes in the peripheral blood (PB) are predominantly resting (5), specific structures known as proliferation centers, which are localized in the lymph nodes and in the bone marrow, replenish the CLL cell population (1, 6, 7). The small proportion of CLL cells in the PB with a proliferative phenotype likely represent cells that have recirculated through the lymph node microenvironment and have exited the proliferation centers before becoming quiescent again (1, 6).

The natural history of many CLLs involves progression toward a more malignant disease. Most, if not all, cases of CLL are preceded by monoclonal B cell lymphocytosis (MBL), a very indolent cell expansion defined by less than 5,000 monoclonal B cells in the PB (8–12). MBL is detectable in approximately 5% of the elderly population (13). Clinically, CLL is characterized by a marked degree of heterogeneity, ranging from patients that harbor highly stable disease with a nearly normal life expectancy to patients with rapidly progressive disease who are destined to succumb in a short time (15, 16). The variable course of CLL is driven, at least in part, by heterogeneity in the disease biology. Over time, a small fraction of CLL cases transform into a very aggressive form known as Richter syndrome (RS), which morphologically mimics diffuse large B cell lymphoma (DLBCL) (1, 17, 18).

This review focuses on recent advances in the analysis of the genome of CLL. Genome-wide analysis has allowed the characterization of the spectrum of genetic lesions present in the CLL coding genome, thus providing further insights into CLL pathogenesis and disease progression.

Phenotype and putative cell of origin

CLL expresses a distinct immunophenotype, characterized by co-expression of CD19, CD5, and CD23 coupled with low levels of surface immunoglobulins (1). This phenotypic profile is different from any normal B cell subset, concealing any indication of a normal counterpart. Although the precise cell of origin of CLL is still under investigation (19), immunogenetic studies and gene expression profiling (GEP) analyses have provided important information regarding the putative CLL progenitor (Figure 1).

Extensive molecular investigations of the B cell receptor (BCR) indicate that 60%–65% of CLLs carry immunoglobulin heavy-chain variable (IGHV) genes with evidence of somatic hypermutation in their variable regions, a process that occurs in the germinal center (GC) and may modify BCR affinity for antigens (refs. 20–23 and Figure 1). Conversely, 35%–40% of CLLs are devoid of IGHV somatic mutations (20–22). The association with IGHV gene mutations suggests that a fraction of cases (CLL with mutated IGHV genes [M-CLL]) derive from GC-experienced B cells, while the remaining cases (CLL with unmutated IGHV genes [U-CLL]) derive from B cells that have undergone differentiation in a GC-independent fashion (19, 24, 25).

Both M-CLL and U-CLL share a largely similar GEP, with few genes, such as ZAP70, differentiating the two categories (26, 27). This observation suggests that M-CLL and U-CLL derive from progenitors that are reminiscent of antigen-experienced B cells (26). In humans, antigen-experienced B cells include memory B cells and marginal zone B cells, whose IGHV genes can be somatically mutated or unmutated (28–30). Both these B cell subsets have been proposed to be the cell of origin of CLL (19). The definition of antigen-experienced B cells includes both GC-experienced cells as well as memory-like B cells generated in a T cell–independent reaction not requiring the GC microenvironment and not necessarily involving IGHV somatic hypermutation (31). This notion accounts for the fact that lymphocytes from U-CLLs resemble...
The role of antigen stimulation is corroborated by the skewed BCR repertoire expressed by many CLL patients and by the association of approximately 30% of cases with a stereotyped BCR. Certain IGHV genes are utilized at a significantly higher frequency than expected based on the representation of these same IGHV genes among normal B cells (20, 32, 33). In addition to preferential usage of specific IGHV genes, the BCRs of approximately 30% of CLLs are clustered into stereotyped subsets, each of which is characterized by a high degree of homology of the CDR3 region (34–38). By chance alone, the possibility is extremely low ($<1 \times 10^{-12}$) that the same BCR rearrangement is shared by different CLL patients, suggesting that stereotyped BCRs have been positively selected for their interactions with specific antigens (34–38). Importantly, stereotyped BCRs have also been detected in B cells of normal individuals, though with features distinct from those associated with CLL (39). Although the antigens involved in CLL have not been fully elucidated, a potential role of autoantigens such as non-muscle myosin has been proposed (40, 41).

The distinction between M-CLL and U-CLL is clinically relevant because germline IGHV genes predict poor outcome (21, 22). The worse prognosis predicted by unmutated IGHV genes is conceivably due to the enrichment of some genetic lesions that confer higher aggressiveness (e.g., mutations of NOTCH1) among antigen-experienced B cells even though their IGHV genes have not undergone somatic hypermutation in the GC (26, 27).

The genome of CLL

Karyotypic investigations revealed the association of CLL with del13q14, trisomy 12, del11q22-q23, and del17p13 (refs. 1, 2, 49, 50, and Figure 2). The advent of next-generation sequencing (NGS) technologies, coupled with gene copy number analysis, has enabled exploration of the CLL genome (45, 46, 51, 52), uncovering genetic lesions that recurrently target this leukemia (Figure 2).

The CLL genome is largely devoid of the chromosomal translocations and aberrant somatic hypermutations that are involved in several B cell non-Hodgkin lymphomas (B-NHLs) (1, 2, 49, 50, 53). These observation are consistent with a post-GC or GC-independent derivation of CLL because B-NHL-associated translocations are caused by errors during somatic hypermutation and class switch recombination, two mechanisms that are active in GC B cells (23).

NGS studies have further elucidated the genomic complexity of CLLs and have shown that the average number of non-silent mutations (i.e., mutations that alter the protein sequence) per case is 10–12 at diagnosis, whereas the average number of copy number abnormalities is approximately two (45, 46, 51, 52). The order of magnitude of lesions detected in the coding genome of CLL appears considerably lower than in common epithelial cancers (54). Among hematologic malignancies, the complexity of the CLL genome is also on the lower side, markedly smaller than DLBCL and multiple myeloma, and comparable to some acute leukemias (55–58).

Del13q14 is the most frequent alteration and occurs in 50%–60% of cases (2, 49, 50). Because this lesion is found at a similar frequency in MBL and is often detectable as a single lesion (11, 13), this alteration as well as the greater predisposition of U-CLLs, compared with M-CLLs, to undergo clonal evolution (42–47).

In contrast with the model suggesting a B cell origin of the entire CLL pathogenesis, recent data suggest that the initiating event may occur in a hematopoietic stem cell (48). Hematopoietic stem cells obtained from CLL patients and transplanted into xenograft models have a propensity to generate clonal B cells with a CLL-like phenotype and IGHV usage similar to that of CLL. Nonetheless, the lymphoid expansions that occur in mice are clonally unrelated to those from the original patient, indicating de novo generation, presumably initiated by genetic lesions that are present in the donor stem cells. Although the nature of the initiating event(s) in the hematopoietic precursors is unknown, and the above results require independent confirmation, these findings challenge current theories on the nature of the cell of origin of CLL and its pathogenesis.
also downregulated in CLL in the absence of the del13q14 deletion, possibly as a consequence of epigenetic mechanisms (61, 62). The DLEU2 locus generates a long non-coding RNA that has only one known function, to serve as the primary RNA from which the mature miR-15a/miR-16a microRNAs are processed (59–61, 63). Based on animal models, MIR15A/MIR16A are considered the main candidate tumor suppressor genes involved with del13q14, although a role is likely for additional genetic elements including DLEU2 (refs. 64, 65, and see below). In normal cells, these two microRNAs appear to inhibit expression of multiple genes, including BCL2, the cyclins CCND1 and CCND3, and cyclin-dependent kinase 6 (CDK6) (66–71). Consequently, deletion of MIR15A/MIR16A abrogates this inhibitory effect and favors the constitutive cycling of B cells. In addition, miR-15a/miR-16a may also participate in a microRNA/tumor protein p53 (microRNA/TP53) feedback circuitry associated with CLL pathogenesis and prognosis (72).

Trisomy 12 occurs in approximately 15% of CLL cases and is thought to alter the gene dosage of one or more proto-oncogenes, although the precise molecular mechanism of this alteration is not currently understood (2, 49). Del11q22-q23 in most cases affects the ataxia telangiectasia mutated (ATM) gene, the deficiency of which causes genomic instability and, if inherited in the germline, disrupts the tumor suppressor gene (49, 79–87). Many cases of del11q22-q23 are selectively restricted to CLL, whereas these same mutations are very rare in MBL (92). NOTCH1 mutations at CLL diagnosis preferentially occur among U-CLLs and cluster with trisomy 12 (45–47, 93, 94). NOTCH1 mutations identify a high-risk subgroup of patients showing poor survival comparable to that associated with TP53 abnormalities and exert a prognostic role that is independent of widely accepted risk factors, as confirmed in multiple consecutive series from different institutions (45–47).

NOTCH1 mutations in CLL are highly consistent among cases and disrupt the C-terminal PEST domain that normally limits the intensity and duration of NOTCH1 signaling (45–47). Removal of the PEST domain results in impaired NOTCH1 degradation and accumulation of an active NOTCH1 isoform (90). One recurrent mutation (c.7544_7545delCT) accounts for approximately 80% of all NOTCH1 mutations (45–47) and can be rapidly detected by a simple PCR-based strategy, providing a potential approach for a first-level screening of NOTCH1 alterations (47). Although the precise role of NOTCH1 activation in CLL pathogenesis is still under scrutiny, the relevance of these mutations is highlighted by the distinctive GEP signature of CLL carrying NOTCH1 mutations (46, 94).

Mutations of the SF3B1 gene. SF3B1 is a critical component of both major (U2-like) and minor (U12-like) spliceosomes, which enact the precise excision of introns from pre-mRNA (95). At diagnosis, SF3B1 is mutated in 5%–10% of CLLs (51, 52, 89). Conversely, SF3B1 mutations are exceptional in MBL (96). The pathogenicity of SF3B1 mutations in CLL is supported by the clustering of these mutations in evolutionarily conserved hotspots localized within HEAT domains (51, 52, 89, 97). Although the precise mechanistic aspects of SF3B1 mutations are still under investigation, the observation that SF3B1 regulates the alternative splicing program of genes controlling cell cycle progression and apoptosis points to a potential contribution of SF3B1 mutations in modulating tumor cell proliferation and survival (51, 52). Across the spectrum of mature B cell neoplasias, mutations of SF3B1 are selectively restricted to CLL, whereas these same mutations are highly recurrent in myelodysplastic syndromes (98). At CLL diagnosis, SF3B1 mutations predict reduced survival independent of other clinical and biological risk factors (51, 52, 89).

BIRC3 gene inactivation. BIRC3, along with TRAF2 and TRAF3, cooperates in the same protein complex that negatively regulates
MAP3K14, an activator of the non-canonical pathway of NF-κB signaling (99). All BIRC3 mutations in CLL are predicted to disrupt the C-terminal RING domain, which is essential for proteasomal degradation of MAP3K14 by BIRC3 (78). Consistently, CLLs harboring BIRC3 disruption display constitutive NF-κB activation (78). At CLL diagnosis, BIRC3 disruption associates with unfavorable clinical and genetic features and predicts poor outcome independent of other risk factors (78). Among B cell neoplasms, mutations of BIRC3 appear to be specific for CLL and for splenic marginal zone lymphoma (78, 100).

Other mutations revealed by NGS studies. Whole exome sequencing studies have revealed mutations of several other genes in CLL, albeit at low frequency (45, 46, 51, 52). Of these, MYD88 is mutated in 3%–5% of CLL at diagnosis (46). MYD88 is a critical adaptor molecule of the TLR complex that is also mutated in other B cell malignancies, namely DLBCL and marginal zone lymphoma (100, 101). In CLL, mutations of MYD88 target specific hotspots also implicated in other B cell malignancies (46, 101, 102).

**Pathways to progression and chemorefractoriness**

The predisposition to clinical progression of CLL is influenced by the genetic lesions of the leukemic clone. At CLL diagnosis, TP53 disruption, ATM deletion, NOTCH1 mutations, SF3B1 mutations, and inactivation of BIRC3 have all been shown to predict poor outcome (45–47, 49, 51, 52, 78, 81–89). While the prognostic values of TP53 and ATM have been assessed both in consecutive series and in the context of clinical trials, the roles of NOTCH1, SF3B1, and BIRC3 as clinical predictors have been tested only in consecutive series. Therefore, it remains to be determined whether specific treatments, such as the fludarabine/cyclophosphamide/rituximab regimen when administered to ATM-deleted CLL patients (103), are able to overcome the unfavorable prognostic value of NOTCH1, SF3B1, and BIRC3.

Molecular knowledge of the CLL genotype has significantly advanced the understanding of the genetic basis of chemorefractoriness (Figure 2). TP53 disruption predicts treatment failure with many regimens and mandates the use of alternative therapeutic strategies that are more efficacious (81–88, 103, 104). TP53 disruption, however, accounts for only 30%–40% of CLLs refractory to fludarabine, the mainstay drug for this leukemia. In the remaining cases, the molecular basis of chemorefractoriness is unknown. The newly identified genetic lesions of NOTCH1, SF3B1, and BIRC3 have clarified, at least in part, this issue. All these alterations are enriched in patients requiring therapy and refractory to fludarabine (NOTCH1 mutations: 20%; SF3B1 mutations: 17%; BIRC3 disruption: 24%) compared with CLL at diagnosis or compared with cases requiring therapy but sensitive to the drug (45–47, 78, 89). Among chemorefractory patients, alterations of NOTCH1, SF3B1, and BIRC3 appear to be largely exclusive with disruption of TP53, indicating that multiple pathways can lead to treatment failure (45, 47, 78, 89).

**Pathways to Richter transformation**

A fraction of CLLs develops RS, which represents the clinicopathologic transformation of CLL into DLBCL (17, 18, 105); the cumulative incidence of RS at 5 and 10 years after CLL diagnosis is approximately 10% and 15%, respectively (106).

TP53 disruption (50%–60%), NOTCH1 activation (30%), and MYC abnormalities (20%) have emerged as the most frequent recurrent genetic lesions in RS (refs. 45, 107, and Figure 2). These abnormalities are frequently acquired at transformation, thus pointing to the pathogenetic relevance of these genetic lesions in acquiring the DLBCL phenotype (45, 107).

In RS, NOTCH1 mutations are largely mutually exclusive with MYC oncogenic activation (45). This is consistent with the observation that NOTCH1 directly stimulates MYC transcription and suggests that activation of oncogenic MYC may be a common pathway in RS transformation (45, 108). Conversely, both NOTCH1 mutational activation and MYC deregulation often coexist with TP53 inactivation. The concurrence of TP53 inactivation with MYC activation in tumors may be selected to prevent the apoptotic effects and the response to genomic instability induced by MYC overexpression (45). Genetic lesions of TP53, NOTCH1, and MYC account for many (~60%), though not all, cases of RS (45, 107). Additional genetic alterations acquired at the time of CLL transformation to RS are likely to be revealed by NGS studies.

The high prevalence of TP53 disruption in RS reflects the selection of a chemorefractory clone under the pressure of previous CLL treatments, thus suggesting an explanation for the poor outcome and the limited sensitivity to conventional drugs (17, 18, 105). Consistent with this hypothesis, TP53 status is an independent predictor of RS survival after transformation, segregating RS into prognostically meaningful categories (107).

Despite morphologic and phenotypic similarities with non-GC DLBCL arising de novo, the molecular profile of RS is distinct (107). Most genetic features that are recurrent in de novo DLBCL are rare or absent in RS (107). Moreover, the high frequency of unmutated IGHV genes observed in RS is in sharp contrast with the immunogenetic features of de novo DLBCL, which carry mutated IGHV genes in virtually all cases (107). Such molecular and immunogenetic differences between RS and de novo DLBCL reflect differences in the cell of origin and/or the microenvironment in which the lymphoma develops.

Early recognition of RS may be clinically useful to avoid the exposure of patients to multiple lines of therapy that are of little efficacy. A number of biological risk factors help to independently predict RS at the time of CLL diagnosis (106, 109–113). Biological predictors of RS reflect both the molecular features of the CLL clone — namely, stereotyped CDR3 of specific IGHV genes, telomere length, and mutations of NOTCH1 — as well as the genotype of the host, as exemplified by SNPs of the CD38 and LRP4 genes (109–113). CLL patients exhibiting activation of a specific IGHV gene such as IGHV4-39 in a stereotyped fashion (so-called BCR subset 8) have a very high risk of RS transformation (110). Curiously, a stereotyped IGHV4-39 BCR tends to associate with CLLs expressing surface IgG (35). The affinity of the BCR expressed by these patients for vimentin and possibly other apoptosis-associated autoantigens points to the role of BCR stimulation in RS development (41). Importantly, up to 40%–50% of CLL patients carrying NOTCH1 mutations at diagnosis are ultimately projected to develop and die of RS (47, 113). Consistent with the enrichment of NOTCH1 mutations in IGHV4-39–positive CLL, patients harboring both NOTCH1 mutations and IGHV4-39 constitute the disease category with the highest risk of RS transformation (113).

Since immunogenetic investigations have shown that the subclone from which RS emerges is already identifiable at CLL diagnosis in most patients who subsequently develop RS (114), deep sequencing studies performed in sequential samples are needed to assess whether traces of the genetic lesions manifestly acquired at RS transformation are already present at CLL diag-
nosis. Preliminary evidence suggests that this might be the case, at least in some patients (45).

Animal models of CLL

Mouse models have formally proven the pathogenetic role of del13q14 in CLL development. Three different lines of transgenic mice designed to mimic del13q14 develop CLL and other del13q14-associated lymphoproliferative disorders (64, 65). One transgenic line mimics the MDR common to all cases with del13q14, a second line of transgenic mice contains a deletion selectively targeting the Mir15a/16-1 cluster within intron 4 of the \(\text{Deux}^+\) gene, and the third line carries a relatively large deletion spanning the genomic region most commonly deleted in 13q14 and including several genes (64, 65). All three models develop clonal lymphoproliferative disorders that occur late in life, are mostly indolent, and arise in a B cell–autonomous fashion. The severity of the disease appears to be proportional to the extent of the deletion, suggesting that genetic elements in addition to \(\text{Mir15a}/16-1\) may contribute to the pathogenetic role of del13q14 (64, 65). The observed lymphoproliferations are characterized by low penetrance in all transgenic lines (40% MDR, 25% \(\text{Mir15a}/16-1\) deletion) (64). This observation, coupled with the long latency associated with the disease phenotype, suggests that the development of a full-blown phenotype in mice might require accumulation of additional genetic alterations.

In addition to mice mimicking del13q14, other mouse models of CLL have been reported, including TCL1-transgenic mice and New Zealand black (NZB) mice (67, 115, 116). However, neither TCL1-mice nor NZB mice genetically mimic lesions in human CLL. Transgenic mice overexpressing TCL1 specifically in B cells develop lymphoproliferative disorders that follow an indolent course and display CD5 expression and stereotyped BCR, similar to mouse models of del13q14 (115, 116). Although TCL1 is highly expressed in human CLL, no structural genetic lesions are associated with this phenotype (117, 118). In NZB mice, three loci have been linked to CLL development, including a point mutation that affects the processing of Mir16-1, though this mutation is not detectable in human CLL (67).

Future mouse models combining del13q14 with mutations of NOTCH1, SF3B1, and BIRC3 might clarify the synergy of these genetic lesions in CLL pathogenesis and may more closely recapitulate the human disease.

Therapeutic implications

Advancing the knowledge of CLL molecular genetics may have important therapeutic implications including: (a) therapeutic stratification based on molecular prognosticators; and (b) targeted therapy aimed at correcting the genetic defect or its functional consequences. The guidelines for CLL management that are currently in use consider only the prognostic value of TP53 disruption for the stratification of CLL therapy (88, 104). Because CLL patients with TP53 disruption have an extremely high likelihood of failing standard immunotherapy, it is recommended that these patients receive treatment with alternative regimens that are able to overcome the chemorefractoriness induced by TP53 disruption (104). Given their recent identification (45–47, 51, 52, 78, 89), it has not yet been established whether genetic alterations of NOTCH1, SF3B1, and BIRC3, all of which predict poor outcome in consecutive series, will also be useful for therapeutic stratification.

Some of the genetic lesions that have been recently identified by NGS represent attractive candidates for targeted therapy of CLL. In particular, NOTCH1 is a well-established therapeutic target with some drugs already available and others under active development (118, 119). Compounds inhibiting the enzymatic conversion of NOTCH1 to an active transcription factor have already entered clinical trials for T lineage acute lymphoblastic leukemia (119, 120). The identification of BIRC3 disruption might also bear potential therapeutic implications, because NF-κB pathway inhibitors may potentially target NF-κB activation resulting from BIRC3 abrogation (78).

Apart from the potential targeting of genetic lesions with specific inhibitors, the biological advances in CLL have revealed that the BCR plays a central role in disease pathogenesis and that, consequently, BCR signaling might represent a suitable target for therapy in many patients (121). Currently, two drugs that act as inhibitors of BCR signaling, i.e., the Bruton tyrosine kinase inhibitor ibrutinib and the PI3K-δ inhibitor GS-1101, appear to be particularly promising and are under advanced clinical investigation (121).

Improved understanding of the genetic basis of CLL continues to refine the stratification of patients, with a goal of improving CLL management through targeted therapy. The clinical heterogeneity of CLL is underscored by the variety of genetic lesions observed, from classical markers of CLL to new genetic lesions uncovered by whole genome and whole exome sequencing. Translating these molecular prognosticators into the clinic, along with continued development of targeted therapies, will be critical to improving treatment options.

Acknowledgments

Work by the authors described in this review was supported by the Italian Association for Cancer Research (AIRC) Special Program in Clinical Molecular Oncology 5 × 1000 (number 10007).

Address correspondence to: Riccardo Dalla-Favera, Institute for Cancer Genetics and the Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York 10032, USA. Phone: 212.851.5273; Fax: 212.851.5256; E-mail: rd10@columbia.edu.

13. Rossi D, et al. The prognosis of clinical monoclo- nal B cell lymphocytosis differs from prognosis of Rai 0 chronic lymphocytic leukemia and is rea-
review series


59. Stevenson FK, et al. Chronic lym


