NIH Public Access
Author Manuscript
Hematol Oncol Clin North Am. Author manuscript; available in PMC 2014 April 01.

Published in final edited form as:
Hematol Oncol Clin North Am. 2013 April ; 27(2): 157–171. doi:10.1016/j.hoc.2013.01.001.

Genomic Approaches to Chronic Lymphocytic Leukemia

Ma. Reina Improgo, PhD[a,b] and **Jennifer R. Brown, MD, PhD**[c,d]

[a]Research Fellow, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

[b]Research Fellow in Medicine, Harvard Medical School, Boston, Massachusetts, USA

[c]Director, CLL Center, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

[d]Assistant Professor of Medicine, Harvard Medical School, Boston, Massachusetts, USA

Synopsis

This article discusses recent advances in genomic approaches used to understand chronic lymphocytic leukemia (CLL). We describe tools for analyzing DNA sequence level alterations, summarize data obtained from these various platforms, and discuss the clinical relevance of these findings.

Keywords

CLL; Genomics; CGH; SNP Arrays; Linkage Mapping; GWAS; Whole-exome sequencing; Whole-genome sequencing

Genomics refers to the systematic study of an organism's entire DNA sequence (genome). Molecular information derived from genomic techniques has increased our understanding of many complex diseases including CLL. CLL is the most common form of adult leukemia and is characterized by a highly variable clinical course. Little is known about the molecular correlates underlying the different CLL disease patterns that are clinically evident, and despite epidemiologic evidence showing familial aggregation of CLL, no major predisposing genes have been identified. Here, we discuss the historical application of genomic techniques to these various problems in CLL: first, the use of comparative genomic hybridization (CGH) and single nucleotide polymorphism (SNP) arrays to identify somatically acquired genetic alterations in CLL, then the use of genome-wide linkage analysis and genome-wide association studies (GWAS) to attempt to determine the cause of heritable predisposition to CLL. Finally, we discuss next-generation sequencing technology that allows detection of both somatic and germline lesions in CLL at unprecedented speeds.

© 2013 Elsevier Inc. All rights reserved.

[c,d]Corresponding author for proofs and reprints: Jennifer R. Brown, M.D., Ph.D., Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, Phone: 617-632-4564, Fax: 617-632-4422, Jennifer_Brown@dfci.harvard.edu.
[a,b]Co-Author Address: Ma. Reina Improgo, Ph.D., Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02115, Phone: 617-632-4193, Fax: 617-632-3351, Reina_Improgo@dfci.harvard.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

DISCLOSURES
Funding:
Conflict of Interest:

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Tumor Analysis by Comparative Genomic Hybridization (CGH)

The first genome-wide copy number analyses of CLL were made possible by the development of CGH in the 1990s.[1] CGH allows the detection of chromosomal imbalances using differentially labeled tumor and normal DNA that are co-hybridized to normal metaphase chromosomes. Intensities of the fluorescent labels are then used to determine overrepresentation (gain) or underrepresentation (loss) of genomic content in specific chromosomal regions.

In 1995, Bentz and colleagues first applied the CGH method to 28 CLL patients.[2] They observed copy number alterations (CNAs) in 68% of the patients, with the most frequent gains found in chromosome 8q and 12 and the most frequent losses in chromosomes 6q, 11q, 13q, and 17p (Table 1). A similar study was conducted by Karhu and colleagues using 25 CLL patients.[3] They observed chromosomal imbalances in 48% of the cases, with frequent gains in chromosome 12 and frequent losses in chromosome 11q and 13q (Table 1).

To identify chromosomal abnormalities specifically in familial CLL, Summersgill and colleagues used CGH to analyze 24 pedigrees.[4] In this study, the investigators detected at least one chromosomal imbalance in each patient, with an average of seven abnormalities per case. The most common gain and loss was observed in the X chromosome (Table 1). The chromosomal imbalances observed in the X chromosome appeared to be more common in familial CLL and were hypothesized to contribute to the differential survival of male and female CLL patients.

Classical CGH utilized metaphase chromosomes, which have limited resolution. Substitution of the chromosome targets with a matrix or array containing nucleic acids with defined sequences allowed the detection of much smaller gains and losses.[5] This technique, known as matrix-CGH or array-CGH, makes use of sequence pools representative of whole chromosomes or chromosome arms, cloned in bacterial artificial chromosomes (BAC), P1-derived artificial chromosomes (PAC), or other vectors.

An automated matrix-CGH array specific for CLL was developed in 2004 and validated against 106 CLL cases.[6] Array profiles were compared to cytogenetic data and showed high specificity and sensitivity. A total of 27 gains and 95 losses were detected using this approach. Novel recurrent genomic imbalances were identified, namely trisomy 19 and a small copy number gain in the *MYCN* gene on chromosome 2p24 (Table 1). Similarly, Gunn and colleagues analyzed 187 CLL cases using BAC array-based CGH.[7] They identified copy number changes in 90% of the cases, with expected frequencies for the common genomic alterations, deletions of 13q, 11q and 17p with gain of 12. In addition, they observed submicroscopic deletions of chromosome 22q11 in 28 cases (15%).

More recently, our group employed high-resolution array-CGH to investigate differences between 37 sporadic and 38 familial cases.[8] Sporadic cases showed significant association with 11q loss while familial cases showed significant association with 14q11 gain. Alterations in 14q11 were also associated with mutated *IGHV* status and with homozygous deletions in 13q. Homozygous deletion in 13q was associated with mutated *IGHV*, low expression of ZAP-70 and a significantly longer time to first treatment (TTFT).

[Tags: Comparative Genomic Hybridization, CGH, Copy number, Chromosomal Imbalances, Chromosomal gains, Chromosomal losses, Copy Number Alterations, CNA]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Tumor Analysis by High-Resolution Single Nucleotide Polymorphism (SNP) Arrays

The necessity of overcoming the resolution limits of classical CGH (10–20 Mb) and of array-CGH (0.1 Mb) prompted the use of even higher resolution platforms. In 2004, Bignell and colleagues demonstrated the utility of using SNP arrays, originally designed for genotyping, to detect CNAs at a genome-wide level.[9] Using the Affymetrix p501 array as a prototype, they showed that simultaneous genotyping and copy number analysis of cancer cell lines allowed the detection of genomic alterations that would have been missed by array CGH or genotyping alone, including loss of heterozygosity (LOH) in copy neutral regions as in the case of uniparental disomy (UPD) or as acquired in cancer.

SNP arrays were first applied to CLL samples by Pfeifer and colleagues, using the Affymetrix 10K and 50K arrays.[10] They identified chromosomal imbalances in 65.6% and 81.5% of cases, respectively, indicating greater sensitivity of higher-density arrays. Deletion 13q14 was the most frequent aberration found, followed by trisomy 12, del11q22, and del17p13. In addition, they identified 24 regions with LOH without altered gene dosage.

Kujawski and colleagues subsequently used the 50k Affymetrix SNP array to quantify genome-wide allelic imbalances, including LOH, in 178 CLL patients in order to derive a genomic complexity score.[11] The complexity scores correlated well with their clinical endpoints, TTFT and time to subsequent therapy (TTST). Specifically, high genomic complexity was found to be an independent risk factor for disease progression and treatment failure.

Forconi and colleagues used a higher density array (250k Affymetrix) to specifically investigate patients with deletions in 17p, as these exhibit aggressive disease.[12] All cases displayed multiple copy number changes, with frequent losses in chromosome 8p and frequent gains in chromosome 8q and 2p. 8p loss and 2p gain predicted shorter TTFT and poorer overall survival in these 17p patients. We subsequently reported a similar association between 8p loss and 17p deletion, using the SNP6.0 array.[13]

To determine which platforms allowed reliable detection of CNAs, Gunnarsson and colleagues carried out a comparative study of four high-resolution platforms: BAC arrays (32K), oligonucleotide arrays (185K), and two SNP arrays (Affymetrix 250K and Illumina 317K).[14] All platforms could robustly detect large aberrations, with 29 CNAs concordantly detected, including common alterations. However, small CNAs were detected only by the high-density oligonucleotide and SNP arrays. The oligonucleotide array had lower baseline variation compared to the other platforms. The 250K Affymetrix array detected more CNAs than the 317K Illumina array, but the latter detected more LOH events.

To assess the performance of SNP arrays for routine clinical use, Hagenkord and colleagues compared low density (Affymetrix 10K2.0), medium density (Affymetrix 250K *Nsp*), and high density (Affymetrix SNP6.0) SNP arrays.[15] The 10K2.0 array was found to be unsuitable for use in the clinic due to its relatively poor resolution. The SNP6.0 array was superior in detecting small aberrations but it was equivalent to the 250K array for detecting lesions known to be clinically relevant. Further, the 250K array was less costly and easier to manage. The 250K array demonstrated 98.5% concordance with the standard CLL FISH panel but also detected acquired UPD and additional regions of genomic complexity.

Ouillette and colleagues also performed copy number analysis of 255 CLL patients using Affymetrix 6.0 Arrays.[16] They observed 2 CNAs in 39% of all cases and 3 CNAs in 20% of patients. They correlated genomic complexity (defined as the total number of CNAs)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

with clinical outcomes and found elevated genomic complexity to be an independent marker for aggressive CLL and short overall survival. The same group also analyzed SNP 6.0 arrays to specifically define 13q14 deletions in 255 CLL patients.[17] Large 13q14 deletions encompassing the *RB1* gene were detected in 20% of these patients and were associated with decreased survival.

Recently, our group integrated copy number analysis using Affymetrix 6.0 SNP arrays together with gene expression profiling in 161 CLL patients.[13] With matched germline controls, we found a median of only one somatic CNA per sample, suggesting that the CLL genome is relatively stable. We identified recurrent CNAs associated with short TTFT: 8q24 amplification, 3q26 amplification, and 8p deletions. Amplifications of 3q26 were focused on the *PIK3CA* gene and amplifications of 8q were focused on *MYC* and on the regulatory region near *MYC* which has been implicated by GWAS in disease risk in CLL and many other cancers.

Similarly, Edelmann and colleagues used the Affymetrix 6.0 array to analyze 353 untreated CLL samples.[18] They identified an average of 1.8 CNAs per case and found copy neutral LOH in 6% of cases, most frequently in 13q, 17p, and 11q. Chromosome 13q14 was deleted in 61% of cases, with minimally deleted regions refined to the *DLEU1* and *DLEU2* genes. They also found novel lesions including a frequent deletion at 15q15.1 (4%), with the smallest deletion found in the *MAX* gene associated (*MGA*) gene locus.

Altogether, CGH and SNP array studies of CLL tumors have identified recurrently altered loci that are likely involved in the pathogenesis of CLL and could even potentially be involved in CLL susceptibility. The extent of genomic aberrations, as well as the presence of specific CNAs in addition to those classically identified by FISH, such as 2p gain, 3q gain, 8p deletion and 8q gain, have been suggested to have clinical relevance and should be tested in prospective clinical trials to evaluate their true usefulness as predictors for clinical outcomes.

[Tags: SNP Array, BAC Array, Oligonucleotide Array, Affymetrix, Illumina, Loss of Heterozygosity, LOH, Allelic imbalance, Genomic complexity score]

Germline Genome-Wide Linkage Mapping

Genome-wide linkage mapping, also known as genome scanning, is a high-throughput method that uses genetic markers to assess the likelihood that a marker associated with a disease phenotype is linked to a predisposing gene. These studies are typically done on germline DNA in order to identify genes involved in heritability, and this field developed in parallel with the early studies on tumor DNAs. The statistical likelihood that a particular genetic marker is linked to the phenotype in question is usually represented by a logarithm of the odds (LOD) score, which is a measure of the probability that an observed linkage is indeed a true linkage. By convention, a LOD 3.0 is required for significance, as this indicates that the odds are 1000 to 1 in favor of genetic linkage. LOD scores 2.0 are considered suggestive and LOD scores 1.0 indicate regions that require follow-up studies.

Goldin and colleagues first used genome-wide linkage mapping to genotype 18 CLL families.[19] They scanned 359 microsatellite markers in 28 panels using a medium density linkage mapping set. LOD scores 1 were observed for chromosomes 1, 3, 6, 12, 13, and 17 but none of these showed significant or suggestive linkage. A follow-up study was later conducted using 28 families, focusing on markers around the above regions of interest.[20] Similarly, the data did not support linkage in chromosomes 1, 3, 6, 12, and 17. The authors argue that chromosome 13q21.33 remained a region of interest as it was significant at the locus level though not at the genome-wide level. Fine mapping of this region using

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

interphase FISH in 6 CLL families revealed a minimally deleted region in 13q21.33-q22.2 shared by four families.[21] Two asymptomatic siblings who shared this haplotype exhibited monoclonal B-cell lymphocytosis, which is thought to be a precursor of CLL. Sequencing of the 13 genes found in this region revealed 85 polymorphisms though none of these were coding or frameshift mutations. An intronic polymorphism in the *PIBF1* gene cosegregated with the haplotype shared by three affected members of one family.

In 2005, Sellick and colleagues analyzed a bigger cohort of 115 families using the Affymetrix GeneChip Mapping 10kv1 Xba Array.[22] Again, no region of significant linkage was observed in this study. Chromosome 11p11 displayed suggestive linkage and chromosomes 5q22-23, 6p22, 10q25, and 14q32 yielded LOD scores > 1.15. Although none of these regions correspond to those commonly found in cytogenetic studies or in earlier studies focused on tumor analysis, this is not necessarily surprising since the regions defined by Sellick and colleagues should be associated with germline disease predisposition rather than the tumor-related somatic alterations described in most previous studies.[2]

To increase detection power, Sellick and colleagues analyzed an additional 101 pedigrees using the GeneChip Mapping 10Kv2.0 Xba Array, which scans 10,200 SNP markers.[23] They then pooled the results of this study with the results from the 105 families in their previous cohort. Chromosome 2q21.2 emerged as a major susceptibility locus. This locus contains the gene encoding the chemokine receptor 4 (*CXCR4*). The same group analyzed the genotype frequency of rs2228014, a polymorphic variant in *CXCR4,* in 1058 CLL cases and 1807 controls.[24] They found no evidence that rs2228014 influences CLL risk. They did, however, identify three cases with *CXCR4* mutations, a finding that would be interesting to further investigate.

As illustrated by the above, the ability to find a significant LOD score is dependent on the study power, which is dependent in turn on the size of the families and the numbers of affected and unaffected individuals available to be studied. In CLL, study power can be a particular problem due to relatively small families with just a few affected individuals, some of whom may be deceased prior to the study. Individual genetic events that are likely causative in single families have been described[25–27], but as yet no recurrent highly penetrant predisposing gene has been identified. These findings suggest that such a gene may not exist, and that CLL risk may more typically arise from the combination of multiple lower-risk alleles.

[Tags: Genome-wide linkage mapping, Logarithm of the Odds, LOD score, Microsatellite markers, Susceptibility locus]

Genome-Wide Association Studies (GWAS)

The absence of a major highly penetrant disease-causing locus in CLL identifiable by linkage suggests that genetic predisposition to CLL may lie in the coinheritance of multiple lower-risk variants. GWAS allows the identification of such variants for particular diseases. The first GWAS conducted for CLL analyzed 299,983 SNPs in a total of 1,529 cases and 3,115 controls from a European cohort.[28] Seven SNPs representing six CLL risk loci were identified in this study, namely rs17483466, rs13397985, rs872071, rs9378805, rs735665, rs7176508, and rs11083846 (Table 2). The strongest statistical evidence was obtained for rs872071 and rs9378805, both of which map to a region on chromosome 6p25.3 near the interferon regulatory factor 4 (*IRF4*) gene. In addition, the risk genotype of rs872071 was found to correlate with lower expression of *IRF4* in lymphoblastoid cell lines.[29] Fine-scale mapping of the 6p25.3 locus narrowed the association signal to an 18-kb region containing the 3'-untranslated region (UTR) of *IRF4*.[30] This region is predicted to encode a binding site for the trans-acting regulatory element *MZF1*, a growth suppressor in hematopoietic cells,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

but the biology of how this alteration in *IRF4* may predispose to CLL is not yet understood.[31] Furthermore, *IRF4* has since been shown to be somatically mutated in 1.5% of CLLs,[32] leading to increased activity, and to be amplified in the germline of a family with Mendelian-type inheritance of CLL, leading to decreased expression.[27] Indeed, although these various lines of evidence implicate *IRF4* in CLL, the underlying biological mechanisms of *IRF4* involvement in CLL remain obscure.

To further verify the association of these risk SNPs with CLL, Crowther-Swanepoel and colleagues genotyped a Spanish cohort consisting of 424 cases and 450 controls as well as a Swedish cohort consisting of 400 cases and 400 controls.[33] They confirmed association between CLL risk and rs13397985, rs872071, rs735665, rs7176508, and rs11083846 (Table 2). An extension study performed by the same group identified four additional susceptibility loci, with 8q24.21 (a GWAS susceptibility region for multiple solid tumors, which likely functions as a *MYC* enhancer) and 16q24.1 (containing *IRF8*) appearing most interesting (Table 2).[34] Another extension study by the same group, using pooled data from previous work as well as new datasets from Poland, Italy, and the UK, reported additional risk loci at 15q25.2 (near *CPEB1*) and 18q21.1 (near *CXXC1* and *MBD1*).[35]

Slager and colleagues later evaluated risk SNPs in a Caucasian cohort from the United States consisting of 438 cases and 328 controls.[36] Associations were confirmed for the previously described SNPs rs17483466, rs13397985, rs872071, rs735665, rs7176508, and rs9378805 (Table 2). To identify SNPs specifically associated with familial CLL, the same group conducted a GWAS enriched for familial cases.[37] Using a total of 407 CLL patients, 102 of which were familial cases, and 296 controls, they found four SNPs with genome-wide significance in the 16q24.1 locus containing the *IRF8* gene (Table 2). All four risk alleles were found to be associated with decreased *IRF8* mRNA levels in lymphocytes. SNPs found to be specifically associated with familial CLL include rs615672, rs674313, and rs502771 (near *HLA-DRB5*) and rs9272219 and rs9272535 (near *HLA-DQA1*).

Meta-analysis of the aforementioned GWAS, followed by validation in an independent case-control series, identified a novel risk locus at chromosome 6p21.33.[38] The two risk SNPs in this locus are rs210134, which lies 100 kb telomeric to the BCL2 antagonist killer (*BAK1*) gene, and rs210142, which lies in intron 1 of *BAK1*. A strong relationship between the risk allele of rs210134 and reduced *BAK1* expression was also found in lymphoblastoid cell lines. BAK1 is known to promote apoptosis by antagonizing BCL2 and other antiapoptotic proteins.[39]

Most risk SNPs identified by GWAS are located in non-coding or intergenic regions, prompting the hypothesis that their function lies in regulating gene expression. To identify SNPs that alter gene expression in CLL, Sille and colleagues performed expression quantitative trait loci (eQTL) analysis, an approach involving integration of genome-wide SNP data with gene expression profiles to attempt to identify genes that are differentially expressed based on the genotype of GWAS SNPs.[40] Using publicly available databases, they found a total of 19 SNPs associated with differential gene expression in lymphoblastoid cell lines: 16 SNPs associated with expression of *SP140*, a putative tumor suppressor gene; and 3 SNPs linked to expression of *DACT3*, a member of the WNT/β-catenin pathway, and of *GNG8*, a gene involved in G protein-coupled receptor signaling. Of these, 14 were found to lie in predicted regulatory elements, several of which have been implicated in CLL or other hematologic malignancies. These results suggest that these genes may be involved in CLL pathogenesis, but further validation will be required.

[Tags: Genome-wide Association Studies, GWAS, Genetic predisposition, Single Nucleotide Polymorphisms, SNP, Risk Locus, Expression Quantitative Trait Loci, eQTL]

Moving to Sequence-Level Analysis

Direct Sanger sequencing (first generation) has enabled the identification of key mutations in CLL, including somatic mutations in *TP53*[41–44] and *ATM*[45–47] that appear to play important roles in the pathogenesis of CLL. However, somatic mutations in *TP53* and *ATM* are present in only 10–40% of CLL cases depending on prior therapy. Furthermore, systematic sequencing of tyrosine kinases in CLL also revealed no somatic mutations,[48] and sequencing of the entire kinome did not identify recurrently mutated kinase genes in CLL.[49] These experiments reveal the limitations of Sanger sequencing which is slow, labor-intensive and not high-throughput.

Next generation sequencing (NGS) technologies have therefore been employed in an effort to determine the full spectrum of genetic lesions in CLL. Also known as second-generation sequencing, NGS refers to methods that involve simultaneous detection of nucleotides from multiple amplified DNA clones.[50] NGS allows more rapid sequencing of entire exomes, genomes and transcriptomes at lower cost, and can facilitate the discovery of novel chromosomal rearrangements and copy number alterations. NGS is also able to identify genetic lesions occurring at low frequency in clinical samples by sequencing at higher or deeper coverage, that is, by sequencing more amplified clones per region.

Whole-exome sequencing (WES) was first performed in CLL by Fabbri and colleagues, in combination with SNP array analysis.[51] From five CLL patients, 40 somatic non-silent mutations were observed involving 39 distinct genes. Direct Sanger sequencing of coding and splice site regions of all mutated genes were then performed in an independent panel of 48 CLL cases. From the results of the combined analyses, five recurrently mutated genes emerged, namely, *NOTCH1*, *TP53*, *PLEKHG5, TGM7*, and *BIRC3* (Table 3). *NOTCH1* mutations were detected at significantly higher frequency in chemorefractory CLL and during disease progression to Richter's syndrome. Integration of WES and copy number data revealed a total of 52 genetic lesions (range of 7–13 lesions per case), with 40 somatic mutations (ranging from 6–10 mutations per case) and 12 copy number alterations (ranging from 1–5 alterations per case).

Around the same time, Puente and colleagues reported results of their whole-genome sequencing of four CLL cases, composed of two *IGHV* unmutated and two *IGHV* mutated cases.[52] They identified 46 somatic mutations in 45 genes. To validate mutations in 26 expressed genes, they used a combination of PCR amplification and Illumina sequencing of pooled samples from a set of 169 additional patients. They identified four recurrently mutated genes: *NOTCH1*, *XPO1*, *MYD88*, and *KLH6*. *MYD88* and *KLH6* were associated with mutated *IGHV* status whereas *NOTCH1* and *XPO1* were associated with unmutated *IGHV* status. The recurrent mutation in *MYD88* was associated with younger age and clinically advanced stage at diagnosis. This same mutation has been identified in our CLL cohort[53] as well as in diffuse large B-cell lymphoma, where it has been shown to confer cancer cell survival.[54]

Our group recently conducted whole-exome and whole-genome sequencing in 88 and 3 CLL patients, respectively.[53] The patients were selected to reflect the full cytogenetic spectrum of CLL including 17p and 11q deletions. We identified nine significantly mutated genes, namely *TP53*, *ATM*, *MYD88*, and *NOTCH1* (all previously implicated in CLL) as well as *SF3B1*, *ZMYM3*, *MAPK1*, *FBXW7*, and *DDX3X*. *SF3B1* was the second most frequently mutated gene in this cohort, and occurred in association with deletions in chromosome 11q,

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

which is associated with poor CLL prognosis. *SF3B1* had been previously found to be mutated in myelodysplastic syndromes.[55] Tumor samples with mutations in *SF3B1* were found to have alterations in pre-messenger RNA splicing, consistent with *SF3B1* being a component of the core spliceosome. The presence of an *SF3B1* mutation was an independent predictor of short TTFT in multivariable analysis in this cohort. Mutations in *FBXW7* may also be interesting as *FBXW7* is important for proteosomal degradation of NOTCH.[56] We also found novel, non-synonymous mutations in WNT pathway members *DKK2* and *BCL9*, though no individual recurrent driver mutations were found. Somatic mutations in CLL clustered in pathways involved in cell-cycle regulation, DNA repair, NOTCH signaling, inflammation, RNA processing, and WNT signaling.

In the context of the International Cancer Genome Consortium, Quesada and colleagues performed large-scale sequencing involving 105 CLL patients and also identified *SF3B1* among 78 genes recurrently mutated in CLL.[57] Clinical analysis showed that patients with *SF3B1* mutations presented with advanced disease at diagnosis and were characterized by adverse features such as elevated β2-microglobulin levels and mutated *IGHV* status. In addition, patients with *SF3B1* mutations had significantly shorter time to disease progression and lower 10-year survival rates. Other genes found to be recurrently mutated in this study include *NOTCH1*, *POT1*, *CHD2*, and *LRP1B*. *TP53* and *ATM* mutations were rare in this study, likely because these patients represented an untreated cohort.

Thus, sequencing efforts to date have identified *SF3B1* and *NOTCH1* as somatically mutated in 10–15% of CLLs, with other mutations, apart from the previously identified *TP53* and *ATM,* generally occurring in 5% or fewer patients. The latter mutations are likely still significant, but their relatively low frequency and diversity will make understanding and targeting them more complicated. Data to date suggest that *SF3B1* and *NOTCH1* are both associated with poor prognosis but these findings remain to be confirmed in prospective clinical trials.

Conclusions

Genomic approaches have provided a comprehensive understanding of genomic alterations in CLL and most recently of the somatic mutational landscape in CLL. Significant prognostic associations have been suggested for particular chromosomal gains and losses such as 8p deletion and 8q24 gain; for the total number of copy number alterations; and for particular somatic mutations such as *SF3B1* and *NOTCH1*. New recurrently mutated genes have been identified (e.g., *SF3B1*) whose role in cancer was previously unknown and is still poorly understood but is currently under active investigation.

To date, efforts to use genomics to elucidate germline predisposition to CLL either through linkage mapping or GWAS have found associated markers, but the underlying causative pathogenesis has remained elusive. Ongoing efforts applying NGS to the germline of familial CLL cases will be more challenging than the somatic analysis completed so far, but will hopefully lead to meaningful insights in the years to come.

While the advances are encouraging, many avenues of investigation remain. As a community we are now in a position to form a fully integrated model of genomic and biologic prognostic factors and to start to incorporate this into clinical practice. We are also poised to begin targeting specific mutational events with therapies in the clinic. Drugs that target *NOTCH1* are already in the clinic for other disease indications, which may facilitate their evaluation in CLL. Targeting may also become possible for *SF3B1,* as we learn more about its role in the pathogenesis of CLL and as better inhibitory drugs advance into the clinic. Advances in the understanding of germline predisposition to CLL remain slow but

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

will hopefully start to benefit from NGS. The explosion of genomic knowledge in CLL has just begun, and is just starting to expand into the clinic, but in the coming years as our understanding expands and ongoing technological innovation fuels new discoveries, we expect the impact of genomic discovery on prognosis and therapy of CLL to come to fruition.

[Tags: Next-generation sequencing, Whole-genome sequencing, whole-exome sequencing, WES, *NOTCH1*, *SF3B1*, Recurrent Mutations]

References

1. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992; 258(5083):818–821. [PubMed: 1359641]
2. Bentz M, Huck K, du Manoir S, et al. Comparative genomic hybridization in chronic B-cell leukemias shows a high incidence of chromosomal gains and losses. Blood. 1995; 85(12):3610–3618. [PubMed: 7780145]
3. Karhu R, Knuutila S, Kallioniemi OP, et al. Frequent loss of the 11q14-24 region in chronic lymphocytic leukemia: a study by comparative genomic hybridization. Tampere CLL Group. Genes Chromosomes Cancer. 1997; 19(4):286–290. [PubMed: 9258665]
4. Summersgill B, Thornton P, Atkinson S, et al. Chromosomal imbalances in familial chronic lymphocytic leukaemia: a comparative genomic hybridisation analysis. Leukemia. 2002; 16(7): 1229–1232. [PubMed: 12094247]
5. Solinas-Toldo S, Lampel S, Stilgenbauer S, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer. 1997; 20(4):399–407. [PubMed: 9408757]
6. Schwaenen C, Nessling M, Wessendorf S, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. Proc Natl Acad Sci U S A. 2004; 101(4):1039–1044. [PubMed: 14730057]
7. Gunn SR, Bolla AR, Barron LL, et al. Array CGH analysis of chronic lymphocytic leukemia reveals frequent cryptic monoallelic and biallelic deletions of chromosome 22q11 that include the PRAME gene. Leuk Res. 2009; 33(9):1276–1281. [PubMed: 19027161]
8. Setlur SR, Ihm C, Tchinda J, et al. Comparison of familial and sporadic chronic lymphocytic leukaemia using high resolution array comparative genomic hybridization. Br J Haematol. 2010; 151(4):336–345. [PubMed: 20812997]
9. Bignell GR, Huang J, Greshock J, et al. High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Res. 2004; 14(2):287–295. [PubMed: 14762065]
10. Pfeifer D, Pantic M, Skatulla I, et al. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood. 2007; 109(3):1202–1210. [PubMed: 17053054]
11. Kujawski L, Ouillette P, Erba H, et al. Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. Blood. 2008; 112(5):1993–2003. [PubMed: 18436738]
12. Forconi F, Rinaldi A, Kwee I, et al. Genome-wide DNA analysis identifies recurrent imbalances predicting outcome in chronic lymphocytic leukaemia with 17p deletion. Br J Haematol. 2008; 143(4):532–536. [PubMed: 18752589]
13. Brown JR, Hanna M, Tesar B, et al. Integrative Genomic Analysis Implicates Gain of PIK3CA at 3q26 and MYC at 8q24 in Chronic Lymphocytic Leukemia. Clin Cancer Res. 2012; 18(14):3791–3802. [PubMed: 22623730]
14. Gunnarsson R, Staaf J, Jansson M, et al. Screening for copy-number alterations and loss of heterozygosity in chronic lymphocytic leukemia--a comparative study of four differently designed, high resolution microarray platforms. Genes Chromosomes Cancer. 2008; 47(8):697–711. [PubMed: 18484635]
15. Hagenkord JM, Monzon FA, Kash SF, et al. Array-based karyotyping for prognostic assessment in chronic lymphocytic leukemia: performance comparison of Affymetrix 10K2.0, 250K Nsp, and SNP6.0 arrays. J Mol Diagn. 2010; 12(2):184–196. [PubMed: 20075210]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

16. Ouillette P, Collins R, Shakhan S, et al. Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. Blood. 2011; 118(11):3051–3061. [PubMed: 21795749]
17. Ouillette P, Collins R, Shakhan S, et al. The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. Clin Cancer Res. 2011; 17(21):6778–6790. [PubMed: 21890456]
18. Edelmann J, Holzmann K, Miller F, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. Blood. 2012
19. Goldin LR, Ishibe N, Sgambati M, et al. A genome scan of 18 families with chronic lymphocytic leukaemia. Br J Haematol. 2003; 121(6):866–873. [PubMed: 12786797]
20. Ng D, Marti GE, Fontaine L, et al. High-density mapping and follow-up studies on chromosomal regions 1, 3, 6, 12, 13 and 17 in 28 families with chronic lymphocytic leukaemia. Br J Haematol. 2006; 133(1):59–61. [PubMed: 16512829]
21. Ng D, Toure O, Wei MH, et al. Identification of a novel chromosome region, 13q21.33-q22.2, for susceptibility genes in familial chronic lymphocytic leukemia. Blood. 2007; 109(3):916–925. [PubMed: 17047154]
22. Sellick GS, Webb EL, Allinson R, et al. A high-density SNP genomewide linkage scan for chronic lymphocytic leukemia-susceptibility loci. Am J Hum Genet. 2005; 77(3):420–429. [PubMed: 16080117]
23. Sellick GS, Goldin LR, Wild RW, et al. A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia. Blood. 2007; 110(9): 3326–3333. [PubMed: 17687107]
24. Crowther-Swanepoel D, Qureshi M, Dyer MJ, et al. Genetic variation in CXCR4 and risk of chronic lymphocytic leukemia. Blood. 2009; 114(23):4843–4846. [PubMed: 19812382]
25. Calin GA, Trapasso F, Shimizu M, et al. Familial cancer associated with a polymorphism in ARLTS1. N Engl J Med. 2005; 352(16):1667–1676. [PubMed: 15843669]
26. Raval A, Tanner SM, Byrd JC, et al. Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. Cell. 2007; 129(5):879–890. [PubMed: 17540169]
27. Brown JR, Hanna M, Tesar B, et al. Germline copy number variation associated with Mendelian inheritance of CLL in two families. Leukemia. 2012; 26(7):1710–1713. [PubMed: 22382893]
28. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2008; 40(10):1204–1210. [PubMed: 18758461]
29. Stranger BE, Forrest MS, Clark AG, et al. Genome-wide associations of gene expression variation in humans. PLoS Genet. 2005; 1(6):e78. [PubMed: 16362079]
30. Crowther-Swanepoel D, Broderick P, Ma Y, et al. Fine-scale mapping of the 6p25.3 chronic lymphocytic leukaemia susceptibility locus. Hum Mol Genet. 2010; 19(9):1840–1845. [PubMed: 20123861]
31. Gaboli M, Kotsi PA, Gurrieri C, et al. Mzf1 controls cell proliferation and tumorigenesis. Genes Dev. 2001; 15(13):1625–1630. [PubMed: 11445537]
32. Havelange V, Pekarsky Y, Nakamura T, et al. IRF4 mutations in chronic lymphocytic leukemia. Blood. 2011; 118(10):2827–2829. [PubMed: 21791429]
33. Crowther-Swanepoel D, Mansouri M, Enjuanes A, et al. Verification that common variation at 2q37.1, 6p25.3, 11q24.1, 15q23, and 19q13.32 influences chronic lymphocytic leukaemia risk. Br J Haematol. 2010; 150(4):473–479. [PubMed: 20553269]
34. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet. 2010; 42(2):132–136. [PubMed: 20062064]
35. Crowther-Swanepoel D, Di Bernardo MC, Jamroziak K, et al. Common genetic variation at 15q25.2 impacts on chronic lymphocytic leukaemia risk. Br J Haematol. 2011; 154(2):229–233. [PubMed: 21554262]
36. Slager SL, Goldin LR, Strom SS, et al. Genetic susceptibility variants for chronic lymphocytic leukemia. Cancer Epidemiol Biomarkers Prev. 2010; 19(4):1098–1102. [PubMed: 20332261]
37. Slager SL, Rabe KG, Achenbach SJ, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. Blood. 2011; 117(6):1911–1916. [PubMed: 21131588]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

38. Slager SL, Skibola CF, Di Bernardo MC, et al. Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. Blood. 2012; 120(4):843–846. [PubMed: 22700719]
39. Chittenden T, Harrington EA, O'Connor R, et al. Induction of apoptosis by the Bcl-2 homologue Bak. Nature. 1995; 374(6524):733–736. [PubMed: 7715730]
40. Sille FC, Thomas R, Smith MT, et al. Post-GWAS functional characterization of susceptibility variants for chronic lymphocytic leukemia. PLoS One. 2012; 7(1):e29632. [PubMed: 22235315]
41. Gaidano G, Ballerini P, Gong JZ, et al. p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 1991; 88(12):5413–5417. [PubMed: 2052620]
42. Wattel E, Preudhomme C, Hecquet B, et al. p53 mutations are associated with resistance to chemotherapy and short survival in hematologic malignancies. Blood. 1994; 84(9):3148–3157. [PubMed: 7949187]
43. Zenz T, Krober A, Scherer K, et al. Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up. Blood. 2008; 112(8):3322–3329. [PubMed: 18689542]
44. Zenz T, Eichhorst B, Busch R, et al. TP53 mutation and survival in chronic lymphocytic leukemia. J Clin Oncol. 2010; 28(29):4473–4479. [PubMed: 20697090]
45. Bullrich F, Rasio D, Kitada S, et al. ATM mutations in B-cell chronic lymphocytic leukemia. Cancer Res. 1999; 59(1):24–27. [PubMed: 9892178]
46. Schaffner C, Stilgenbauer S, Rappold GA, et al. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. Blood. 1999; 94(2):748–753. [PubMed: 10397742]
47. Austen B, Skowronska A, Baker C, et al. Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. J Clin Oncol. 2007; 25(34):5448–5457. [PubMed: 17968022]
48. Brown JR, Levine RL, Thompson C, et al. Systematic genomic screen for tyrosine kinase mutations in CLL. Leukemia. 2008; 22(10):1966–1969. [PubMed: 18754031]
49. Zhang X, Reis M, Khoriaty R, et al. Sequence analysis of 515 kinase genes in chronic lymphocytic leukemia. Leukemia. 2011; 25(12):1908–1910. [PubMed: 21701494]
50. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010; 11(10):685–696. [PubMed: 20847746]
51. 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(7):1389–1401. [PubMed: 21670202]
52. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011; 475(7354):101–105. [PubMed: 21642962]
53. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011; 365(26):2497–2506. [PubMed: 22150006]
54. Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. Nature. 2011; 470(7332):115–119. [PubMed: 21179087]
55. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011; 478(7367):64–69. [PubMed: 21909114]
56. Welcker M, Clurman BE. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. Nat Rev Cancer. 2008; 8(2):83–93. [PubMed: 18094723]
57. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet. 2012; 44(1):47–52. [PubMed: 22158541]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Key Points

- Genomic approaches have led to a recent explosion in knowledge of the spectrum of genetic aberrations in CLL.
- Comparative genomic hybridization and high-resolution SNP arrays have been used to detect copy number alterations, some of which have been associated with prognosis and/or overall survival in CLL.
- Efforts to identify the genetic basis of familial CLL have included genome-wide linkage studies to identify CLL susceptibility genes as well as genome-wide association studies to determine CLL risk loci.
- Within the last couple of years, next-generation sequencing techniques have identified common recurrent somatic mutations in *SF3B1* and *NOTCH1* in CLL. Ongoing work involves the investigation of lower frequency somatic mutations and clonal evolution over time as well as the identification of germline variants.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Chromosomal imbalances in CLL based on CGH and SNP array studies

	CGH		SNP Arrays	
Chromosome	**Gains**	**Losses**	**Gains**	**Losses**
1	1 (33%)[4]	-	-	1p31 (10%)[14]
2	2p (4%)[3] 2p24 (6%)[6]	2 (25%)[4]	2p14 (28%)[12]; 2p16 (6%)[10]; 2p22 (28%)[12]	-
3	3 (19%)[4]; 3q (7%)[2]	3p (4%)[2]	3q24 (17%)[12]; 3q26 (6%)[13]	-
4	-	4 (25%)[4]; 4q (4%)[2]	-	-
5	-	5 (17%)[4]	-	-
6	Trisomy 6 (4%)[3]	6q (11%)[2], (4%)[4]; 6q15 (4%)[3]		-
7	-	7q (4%)[2]; 7q31 (4%)[3]	Trisomy[7] (10%)[14]	7q34 (4%)[13]
8	8q (11%)[2]	8p (4%)[2]	8q13 (10%)[14]; 8q21 (20%)[14]; 8q23 (28%)[12]; 8q24 (4%)[13]; 8q24.3 (28%)[12]; 8q24.13 (28%)[12]	8p12 (28%)[12]; 8p21 (28%)[12]; 8p23 (10%)[14], (28%)[12], (5%)[13]; 8q12 (10%)[14]
9	-	-	-	9p21 (10%)[14]; 9q21 (22%)[12]
10	-	10q (4%)[2]	-	-
11	11 (38%)[4]	11q (14%)[2], (24%)[3] (17%)[4], (13%)[8]	-	11q22 (13%)[10], (10%)[14], (6%)[13]
12	Trisomy 12 (16%)[3], (7%)[2]; 12p (4%)[2]	12p12 (4%)[3]	Trisomy 12 (13%)[10], (20%)[14], (12%)[13]	-
13	-	13q (11%)[2], (13%)[4], (53%)[8]; 13q13 (4%)[3], 13q14 (4%)[3], 13c (4%)[3]	-	13q14 (51%)[10], (40%)[14], (50%)[12], (51.7%)[17], (57%)[13]
14	14q (15%)[8]	-	14q32 (28%)[12;]	14q31 (3%)[13]; 14q32 (33%)[12]
15	Trisomy 15 (4%)[2]; 15q (4%)[2]	15q (4%)[2]; 15q11 (4%)[3]	15q26 (10%)[14]	15q13 (10%)[14];
16	-	16p (4%)[2]	-	-
17	17q (7%)[2]	17p (29%)[2], (4%)[3], (17%)[4]	17q 21.1 (28%)[12]; 17q21.31 (33%)[12]; 17q21.32 (28%)[12]	17p11 (10%)[14]; 17p12 (6%)[13]; 17p13 (6%)[10], (20%)[14]
18	18q (4%)[2]	18p (7%)[2]	18q21 (10%)[14]	18q11 (10%)[14]; 18q22 (10%)[14]
19	Trisomy 19 (4%)[2], (5%)[6]	-	19p13 (20%)[14]	19p13 (20%)[14]; 19q13 (10%)[14]
20	-	-	-	-

	CGH		SNP Arrays	
Chromosome	Gains	Losses	Gains	Losses
21	-	-	-	-
22	22q (4%)[2];	22q11 (15%)[7]	-	22q11 (10%)[14] (33%)[13]
X	X (42%)[4]; Xq22 (4%)[3]	X (33%)[4]		

*Numbers in parentheses indicate frequencies of gains or losses reported in different studies;4 Study used only familial cases; Abbreviations: CLL-chronic lymphocytic leukemia; CGH-comparative genomic hybridization; SNP-single nucleotide polymorphism

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

GWAS Loci and SNPs associated with CLL risk

Risk Locus	SNP	Risk Allele	Nearest Gene/s	SNP Location	Population
2q13	rs17483466	G	*ACOXL*, *BCL2L11*	Intron 10 of *ACOXL*	European (UK)[28] Caucasian (US)[36,37]
2q37.1	rs13397985	G	*SP140, SP110*	Intron 1 of *SP140*	European (UK)[28] Caucasian (US)[36,37] Spanish[33] Swedish[33]
2q37.3	rs757978	A	*FARP2*	Exon 9	Combined: European (UK), Spanish, Swedish[34] Caucasian (US)[37]
6p21.3	rs210134	G	*BAK1*	~ 100 kb telomeric	Meta-Analysis: European (UK), Spanish, Swedish, Caucasian (US)[38]
	rs210142	C	*BAK1*	Intron 1	Meta-Analysis: European (UK), Spanish, Swedish, Caucasian (US)[38]
6p25.3	rs872071	G	*IRF4*	3'UTR	European (UK)[28] Caucasian (US)[36,37] Spanish[33] Swedish[33]
	rs9378805	C	*IRF4*	~ 10 kb from 3'UTR	European (UK)[28] Caucasian (US)[36,37]
8q24.21	rs2456449	G	*MYC* (enhancer element)	-	Combined: European (UK), Spanish, Swedish[34] Caucasian (US)[37]
11q24.1	rs735665	A	*GRAMD1B*	~ 50 kb centromeric	European (UK)[28] Caucasian (US)[36,37] Spanish[33] Swedish[33]
15q21.3	rs7169431	A	*NEDD4, RFX7*	-	Combined: European (UK), Spanish, Swedish[34] Caucasian (US)[37]
15q23	rs7176508	A	-	-	European (UK)[28] Spanish[33] Swedish[33]
	rs11072110	T	-	-	Caucasian (US)[36]
	rs10220831	G	-	-	Caucasian (US)[36]

Risk Locus	SNP	Risk Allele	Nearest Gene/s	SNP Location	Population
	rs35707742	G	-	-	Caucasian (US)[36]
	rs4777184	T	-	-	Caucasian (US)[36]
15q25.2	rs783540	G	*CPEB1*	Intron 2	Combined: European (UK), Italian, Polish[35]
16q24.1	rs305061	T	*IRF8*	~ 19 kb telomeric	Combined: European (UK), Spanish, Swedish[34] Caucasian (US)[37]
	rs305077	C	*IRF8*	Intron 3	Caucasian (US)[37]
	rs391525	G	*IRF8*	Intron 3	Caucasian (US)[37]
	rs2292982	G	*IRF8*	Intron 3	Caucasian (US)[37]
	rs2292980	C	*IRF8*	Intron 3	Caucasian (US)[37]
18q21.1	rs1036935	T	*CXXC1, MBD1*	Telomeric	Combined: European (UK), Italian, Polish[35]
19q13.32	rs11083846	A	*PRKD2, STRN4*	Intron 3 of *PRKD2*	European (UK)[28] Spanish[33] Swedish[33]

*SNP – single nucleotide polymorphism;; - means no known genes, transcripts, or microRNAs lie within a 250-kb region flanking the marker

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Genes with recurrent mutations in CLL identified in whole-genome/whole-exome sequencing studies

Gene	Frequency	Reference
TP53	16.5% 7.5%	53 51
SF3B1	15.4% 9.7%	53 57
NOTCH1	15.1% 12.2% 12.1% 4.4%	51 52 57 53
ATM	9.9%	53
MYD88	9.8% 2.9%	53 52
CHD2	4.8%	57
LRP1B	4.8%	57
POT1	4.8%	57
FBXW7	4.4%	53
ZMYM3	4.4%	53
PLEKHG5	3.8%	51
BIRC3	3.8%	51
TGM7	3.8%	51
DDX3X	3.3%	53
MAPK1	3.3%	53
XPO1	2.4%	52
KLHL6	1.8%	52