Published in final edited form as:
Clin Cancer Res. 2012 July 15; 18(14): 3791–3802. doi:10.1158/1078-0432.CCR-11-2342.

Integrative Genomic Analysis Implicates Gain of *PIK3CA* at 3q26 and *MYC* at 8q24 in Chronic Lymphocytic Leukemia

Jennifer R Brown[1,10,*], **Megan Hanna**[3,5], **Bethany Tesar**[1], **Lillian Werner**[2], **Nathalie Pochet**[5,8], **John M. Asara**[7,10], **Yaoyu E Wang**[4], **Paola dal Cin**[9], **Stacey M Fernandes**[1], **Christina Thompson**[1], **Laura MacConaill**[3,9], **Catherine J Wu**[1,10], **Yves Van de Peer**[8], **Mick Correll**[4], **Aviv Regev**[5,6], **Donna Neuberg**[2], and **Arnold S Freedman**[1,10]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA

[2]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA

[3]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, MA

[4]Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA

[5]Broad Institute of Harvard and MIT, Cambridge, MA

[6]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Boston, MA

[7]Beth Israel Deaconess Medical Center, Boston, MA

[8]Department of Plant Systems Biology, VIB, Department of Biotechnology and Genetics, Ghent University, Gent, Belgium

[9]Department of Pathology, Brigham and Women's Hospital

[10]Department of Medicine, Harvard Medical School

Abstract

Purpose—The disease course of chronic lymphocytic leukemia (CLL) varies significantly within cytogenetic groups. We hypothesized that high resolution genomic analysis of CLL would identify additional recurrent abnormalities associated with short time to first therapy (TTFT).

Experimental Design—We undertook high resolution genomic analysis of 161 prospectively enrolled CLLs using Affymetrix 6.0 SNP arrays, and integrated analysis of this dataset with gene expression profiles.

Results—Copy number analysis (CNA) of nonprogressive CLL reveals a stable genotype, with a median of only 1 somatic CNA per sample. Progressive CLL with 13q deletion was associated with additional somatic CNAs, and a greater number of CNAs was predictive of TTFT. We identified other recurrent CNAs associated with short TTFT: 8q24 amplification focused on the

[*]To Whom Correspondence Should Be Addressed: Jennifer R Brown MD PhD, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215; tel 617-632-4564; fax 617-582-7909.

Conflicts of Interest: JRB reports that she has served as a consultant for Calistoga Pharmaceuticals.

Author Contributions

The study was designed, funded and managed by JRB and ASF. JRB enrolled the patients. MH, BT, JA, PDC, SMF, CT and LM performed the experiments. Data were analyzed by JRB, MH, BT, NP, JA, YEW, YVDP, MC, AR and DN. Statistical analysis was performed by LW and DN. The paper was written by JRB with input from MH, BT, NP, LM, CJW, DN and ASF.

Statistical Analysis and Remaining Methods are detailed in the Supplementary Methods.

cancer susceptibility locus near *MYC* in 3.7%; 3q26 amplifications focused on *PIK3CA* in 5.6%; and 8p deletions in 5% of patients. Sequencing of *MYC* further identified somatic mutations in two CLLs. We determined which catalytic subunits of PI3K were in active complex with the p85 regulatory subunit, and demonstrated enrichment for the alpha subunit in three CLLs carrying *PIK3CA* amplification.

Conclusions—Our findings implicate amplifications of 3q26 focused on *PIK3CA* and 8q24 focused on *MYC* in CLL.

Keywords

CLL; *PIK3CA*; *MYC*; genomics; copy number

Introduction

Chronic lymphocytic leukemia is the most common leukemia of adults but still incurable. Prognosis at diagnosis is widely variable, and the key cytogenetic abnormalities determined by FISH remain one of the best predictors of prognosis and treatment response(1). However disease course still varies significantly within these cytogenetic groups, and our ability to predict prognosis remains limited. Once treated, CLL inevitably relapses and each subsequent remission gets shorter.

The advent of high resolution array-based technologies lends itself to the detailed characterization of cancers on multiple levels including but not limited to copy number, gene expression, protein expression and modification, and methylation. Copy number analyses have been undertaken extensively in cancer, allowing the recent publication of a paper that surveys the landscape of somatic copy number alterations (CNAs) primarily in solid tumors but including ALL(2). A small high resolution study in CLL described wide variability in the number of CNAs observed(3), though a larger study looking at newly diagnosed patients(4) found a generally stable genome in previously untreated patients. Recent efforts have dissected the structure of 13q and 11q deletions in detail(5, 6, 7) and associated the number of CNAs with overall survival(8). We hypothesized that high-risk CLLs would likely harbor additional recurrent CNAs not part of the canonical CLL FISH panel but which would likely reflect on disease pathogenesis. We therefore undertook a large integrative study of CLL, using both copy number analysis with direct comparison to cognate germline and gene expression profiling, in order to characterize the CLL genome at very high resolution.

Methods

CLL Patients

161 patients with CLL enrolled on a prospective cohort natural history study at Dana Farber Cancer Institute were studied. The diagnosis of CLL according to WHO criteria was confirmed in all cases. Any individual 18 years or older seen at DFCI with CLL/small lymphocytic leukemia (SLL) was eligible. The study was approved by the Dana-Farber Institutional Review Board and all subjects signed written informed consent. CLL cells were collected from peripheral blood as detailed in the Supplementary Methods. The median follow-up from diagnosis is 81 months.

Genome-Wide DNA Profiling

Genome-wide DNA profiles were obtained using the Genome-wide Human SNP Array 6.0 (Affymetrix), run on the Genetic Analysis Platform at the Broad Institute of Harvard and MIT, according to the manufacturer’s protocol. Tumor DNA (PBMCs or isolated B cells)

and at least one matched germline control (saliva, granulocytes or both) was run for every sample. The quality of all DNAs was verified in three independent PCR reactions prior to use. The data were initially analyzed by the GISTIC method, which identifies significant deletions and amplifications based on analysis of the frequency of occurrence and the amplitude of each aberration in the tumor samples alone, as previously described(2, 9). The GISTIC analysis used the five nearest neighbor normalization method and removed all catalogued germline copy number variants (CNVs)(10). Statistical significance was assessed using a permutation test based on the overall pattern of aberrations across the genome and accounted for multiple hypothesis testing using a false discovery rate framework, with q values < = 0.05 considered significant. To confirm that all abnormalities identified by GISTIC were somatic, the paired somatic and germline samples were manually reviewed using the Integrative Genomics Viewer (IGV) from the Broad Institute (http://www.broadinstitute.org/software/igv/)(11). In addition we compared all CNAs in each tumor with its cognate germline. For details of this analysis, please see the Supplementary Methods. Results are in Supplementary Table 1.

Gene Expression Profiling

All 146 CLLs for which adequate quality RNA was obtained were assessed for their mRNA expression profile by hybridization to the Affymetrix U133 Plus 2.0 array. RNA was isolated from viably frozen tumor cells using the Qiagen RNeasy Midi kit. RNA quality was assessed by A260:280 ratios and by RIN analysis on a Bioanalyzer prior to submission for gene expression profiling. All expression profiles were processed using RMA, implemented by the PreprocessDataset module in GenePattern (http://www.broadinstitute.org/cancer/software/genepattern/) (12, 13). Probes were collapsed to unique genes by selecting the probe with the maximal average expression for each gene. Batch effects were further removed using the ComBat module in GenePattern(13).

To examine gene expression in relation to each CNA, differential gene expression was determined between samples with and without the CNA. Genes were selected according to a t-test p-value < = 0.05, using the ComparativeMarkerSelection and ExtractComparativeMarkerResults modules in GenePattern(13, 14). The significance (nominal p value) of each marker gene was computed using a permutation test based on 1000 replications. Samples were ordered in heat maps based on the correlation of their expression phenotype to that of the samples with the given CNA. Those with a positive correlation are displayed to the right of the gap in the heat maps, and were analyzed in gene set enrichment analysis (GSEA), as detailed in the Supplementary Methods.

Immunoprecipitation of p85 and Shotgun LC/MS/MS

Approximately 10^8 previously frozen PBMCs from individuals with and without the 3q26 amplification were thawed, ficolled and lysed, and a minimum of 0.5 mg cell lysate was subjected to immunoprecipitation with anti-p85 (N-SH2 antibody (Millipore)) followed by mass spectrometry analyses to identify p85 interacting proteins. Further details of this procedure and the quantification technique can be found in the Supplementary Methods.

Results

Characteristics of Patient Cohort

The clinical characteristics of the patients are shown in Supplementary Table 2. 78% were untreated at the time of sampling, while 22% had been previously treated. An additional 21% of patients were treated in the subsequent follow-up period, for a total of 43.6% treated, with an overall median time from diagnosis to treatment of 41 months (0.4–161.2 months). Unmutated *IGHV* and ZAP-70 positivity were both highly predictive of reduced TTFT and

overall survival (OS), as previously reported, although data were missing for a subset of patients (Supplementary Tables 3 and 4).

82% of the 161 CLLs showed at least one CNA by high resolution SNP array (Table 1, Figure 1). GISTIC analysis on the entire population identified the known common CLL abnormalities at frequencies that would be expected in a largely untreated cohort: 57% del 13q, 6.2% deletion 11q, 5.6% deletion 17p, and 12% trisomy 12 (Figure 1A“B and Table 1). Deletions 17p and 11q are known to have an adverse prognosis with short TTFT and OS(1) and this finding was confirmed in our study for TTFT, while OS was adversely impacted at this point only for 17p deletion (Supplementary Tables 3 and 4).

Number of CNAs Predicts TTFT

GISTIC analysis revealed a paucity of CNAs in comparison to most solid tumors (Figure 1A, B; Table 1), as previously reported(8), and some of the identified regions were actually rare germline CNVs when compared to cognate germline. Our analysis revealed that one of four significant amplifications and 14 of 29 significant deletions identified by GISTIC were also present in the corresponding germline. Therefore, we assessed total somatic CNAs per sample using not only the GISTIC analysis but also direct tumor-normal comparison and manual review. The median number of acquired somatic CNAs in the overall population was 1, but in treated patients the median was 2 (Supplementary Tables 1 and 5; see Supplementary Methods). Choosing a cut-off at the median or one CNA greater served to dichotomize groups with markedly different TTFTs (Figure 1D). In CLLs with deletions of 11q or 17p, a median of 3 CNAs per patient was observed; 41% of patients with 3 CNAs or more, and 70% of patients with 4 CNAs or more, also had 11q or 17p deletions (Supplementary Table 5). Even in patients without 11q or 17p deletions, however, increasing number of CNAs remained predictive of short TTFT, suggesting that number of CNAs is an independent adverse predictor (Figure 1E), as previously reported(5).

Progression in 13q Deletion is Associated with Presence of Other CNAs

Ninety-one patients (57%) had 13q deletions, and GISTIC analysis categorized them into three groups based on decreasing size, labeled R1, R2, R3 (Table 1 and Figure 2A). As shown in Figure 2A, all 91 patients were at least partially deleted in the R3 region, which includes the previously described minimally deleted region extending from *DLEU2* and *miR-15a/16-1* to *DLEU7* (6, 15). Interestingly we were unable to define a universally present minimally deleted region (MDR) in our dataset, since three patients had partial deletion of *DLEU2* but lacked deletion of *miR-15a/16-1* (Supplementary Figure 1 section 2), and one patient completely lacked deletion of the *DLEU2* and *miR-15a/16-1* region, instead showing only deletion of *DLEU7* (Supplementary Figure 1 section 3). A second patient showed biallelic deletion of *DLEU7* with only monoallelic deletion of the *DLEU2 miR-15a/16-1* region (Supplementary Figure 1 section 3). These findings are consistent with recent results suggesting that multiple genes in this region contribute to the CLL phenotype(16–19).

Longer 13q deletions were seen in 30 patients and were subdivided into two groups by GISTIC, with a subset having deletions extending just past the *RB* gene (and thereby deleting regions R2 and R3 in Figure 2A and Table 1), and the remainder having very long deletions including *RB* and extending up to 40 Mb (thereby deleting regions R1, R2 and R3 in Figure 2A and Table 1). Longer 13q deletions that delete *RB* have been labeled type II deletions and reported to carry a poor prognosis(5–7), but in our cohort, no significant difference in TTFT was observed between patients with short 13q deletions confined to R3 and those with longer 13q deletions (Figure 2B), defined as either those extending to R2 only, or those extending through R1 and R2.

Our analysis further demonstrated that 24% of the del 13q patients carried biallelic 13q deletions. 17 of 61 patients (28%) with short deletions confined to the R3 region had biallelic loss, as compared to 5 of 30 patients (16.7%) with long deletions extending to R1 or R2 ($p = 0.3$). Of the latter 5 patients, the regions of biallelic deletion were confined to the much smaller R3 region around *miR-15a/16-1*, while the longer deletion regions were monoallelic. In previous work we and others have reported that biallelic 13q deletions are associated with a longer TTFT(4, 20). However, in this study TTFT was similar for both groups (Figure 2C), although biallelic 13q deletion was significantly associated with mutated *IGHV* and negative ZAP-70 (100% in all evaluable cases, compared to 70% for monoallelic deletion 13q, $p = 0.03$ for both).

We evaluated the impact of additional CNAs defined by high-density SNP arrays on the predictive value of 13q deletion. Dohner and colleagues originally established that deletion of 13q as a sole aberration identified by FISH was prognostically favorable(1). In this dataset we observed only a trend toward shorter TTFT of 13q deletion when associated with other abnormalities identified by FISH (data not shown). However, when we used high-density SNP array as the standard to define additional somatic CNAs, we found that any other single somatic CNA along with 13q deletion resulted in a TTFT that was comparable to those patients lacking 13q deletion altogether (Figure 2D). This finding was true for the entire cohort as well as for those patients untreated at time of sampling (Figure 2D). Thus, additional CNAs identified by high density SNP array can most effectively identify patients with 13q deletion who will progress rapidly to treatment. Furthermore, no association of number of CNAs with long vs short 13q deletions was observed ($p = 0.24$).

Gene expression profiling analysis was used to assess genes located in the 13q deletion region, focusing in particular on *DLEU2* and *TRIM13* (deleted in R3) and *RB* (deleted in R2). *DLEU2* and *TRIM13* at 13q14 showed significant downregulation only in samples with biallelic deletion (Figure 2E and data not shown). Samples with monoallelic deletion at GISTIC regions R1 and R2, resulting in deletion of one copy of the *RB* gene, did show significant downregulation of *RB* compared to those samples intact at 13q or deleted only at GISTIC R3, which does not include *RB* (Figure 2E).

CNAs Associated with Disease Progression

In order to identify recurrent CNAs associated with high risk disease, we compared the molecular profiles of patients who remained untreated to the overall cohort (Figure 1B–C). The genomes of untreated patients were significantly more stable, showing predominantly 13q deletion or trisomy 12 (Figure 1C). Patients who had undergone therapy either prior to or after sampling showed three highly significant additional abnormalities: amplification at 3q26.32 and 8q24.21, and deletion at 8p (Figures 1A–B). The amplification events include focal amplifications at the known oncogenes *PIK3CA* and *MYC*, respectively.

Focal Amplification of 3q26.32 Affects *PIK3CA*

Nine patients (9/161 or 5.6%) had amplification of 3q26.32. Three of these patients demonstrated focal somatic amplification of a small region (< 350 bp) corresponding to exon 21 of *PIK3CA* (Figure 3A), while the remainder had large gains (Figure 3A). *PIK3CA* encodes the p110 alpha isoform of the PI3 kinase catalytic subunit (PI3K), one of four class 1 isoforms which also include beta, gamma and delta, and *PIK3CA* is mutated by amplification or activating mutation in many solid tumors(21, 22). A subset of the amplifications we observed were confirmed by FISH, with a probe to the *MECOM* gene at 168.8 Mb (near *PIK3CA* at 180 Mb); FISH showed that three patients had 3 copies, and one patient had 3–6 copies (data not shown). These amplifications were significantly associated with positivity for ZAP-70 (78% vs 30%, $p = 0.007$) and CD38 (44% vs 15%, $p = 0.045$), as

well as with a higher total number of therapies ($p = 0.009$). These amplifications also appeared to be associated with a mildly reduced TTFT ($p < 0.0001$; Figure 3B); the numbers are too small to assess focal and broad amplification patients separately (Supplementary Figure 2). Five of these nine patients are deceased, as compared to 18 of 152 without this amplification, suggesting a possible effect on overall survival, although again the numbers are small.

The significance of the focal *PIK3CA* amplifications compared to the broad amplifications is unclear. The focal amplifications affect the kinase domain of the protein, which is a hotspot for somatic mutation in solid tumors, but would not be expected to increase RNA or protein expression. In fact when we analyze *PIK3CA* RNA expression by GEP, we see increased expression in the broad but not the focal amplification patients (Figure 3C). Immunoprecipitation of the alpha isoform of PI3K from CLLs with focal and broad 3q26 amplifications also shows increased protein in the broad amplification samples but not in the focal amplification samples (Figure 3D). In order to assess the functional significance of increased expression in the samples with broad gains, we determined which catalytic subunits of PI3K were in complex with the p85 regulatory subunit in CLL cells. To accomplish this we performed immunoprecipitation experiments with an antibody to the p85 regulatory subunit and then used mass spectrometry to identify the proteins in complex with p85, in three patients with broad amplifications and six control patients with no copy number gain. We found that in the control CLL patients, the delta subunit of PI3K was the predominant p110 catalytic subunit associated with p85 (48%). In contrast, in the three CLLs carrying broad amplifications of *PIK3CA* that we were able to assess, delta represented only 33% of associated catalytic subunits ($p = 0.04$; Figure 3E), and the alpha subunit was enriched in complex relative to delta (alpha:delta ratio for gain samples 1.24 vs 0.43 for controls, $p = 0.02$; Figure 3F). These results suggest that at least the broad 3q26 amplifications result in altered PI3K subunit composition in these CLLs.

Given that mutations in the PI3K pathway (*PIK3CA* and *PTEN* in particular) are well-described in solid tumors, we sequenced the entire coding regions of *PIK3CA*, *PIK3CD*, *PIK3CG*, *PTEN* and *PIK3R1* and genotyped *AKT* E17K in 188 CLLs. No somatic mutations were identified, suggesting that point mutation is not a mechanism of activation of these genes in CLL, even though the PI3 kinase pathway has been shown to be constitutively activated(23, 24).

We evaluated whether the gene expression profiling data showed any pattern associated with 3q26 gain (Figure 3G). Supervised analysis identified 2981 genes that were differentially expressed between CLLs with and without 3q26 gain (Supplementary Table 6). The CLL samples were then ordered based on the correlation of their gene expression with that of the 3q26 gain samples (Figure 3G). In the 3q26 gain samples themselves, as well as those samples with gene expression that positively correlated with the 3q26 gain samples, an exploratory analysis using gene set enrichment analysis (GSEA) identified increased expression of gene sets comprised of genes repressed by Polycomb complexes in embryonic stem (ES) cells (Supplementary Table 7). This finding is discussed further below.

Amplification of 8q24 Affects *MYC*

Amplification at 8q24 was present in 6 of 161 CLLs (3.7%; Figure 4A). Amplification of *MYC* was confirmed by FISH, with two patients harboring three intact copies of *MYC*, one patient four copies and one patient had one rearranged copy with two intact copies (data not shown). Two patients had focal amplification of the “gene desert” regulatory region approximately 360 kb centromeric to *MYC*, indicated by the GISTIC plot at the bottom of Figure 4A. This 8q24 “gene desert” region contains multiple single nucleotide polymorphisms (SNPs) that have been implicated by genome wide association studies

(GWAS) in susceptibility to multiple solid tumors, as well as CLL(25–27). 8q24 amplification was associated with short TTFT which appeared to be independent of co-occurrence with high risk deletions of 11q and 17p, although the numbers in each group are very small ($p = 0.0001$; Figure 4B). Western blot demonstrated increased MYC expression in samples with gain compared to several controls without gain (Figure 4C).

Since we observed focal gains affecting a region previously implicated as a *MYC* regulatory region, the likely target of 8q24 amplification appeared to be the *MYC* locus. Since known mutations in exon 1 of *MYC* lead to Burkitt's lymphoma (BL), we sequenced exon 1 of *MYC* in 188 CLL samples. We found one sample with a MYC Thr58Ala mutation, which has been previously described in BL and shown to abrogate a regulatory phosphorylation site, leading to activation of MYC(28). Interestingly this mutation impairs FBXW7-mediated degradation of MYC by the proteasome(29), and we have recently identified recurrent mutations in FBXW7 in CLL(30), suggesting that MYC may be a target of FBXW7 in CLL. We also identified a second somatic mutation in *MYC*, a heterozygous insertion mutation that duplicates nine amino acids of the N-terminal interaction and transactivation domain (Supplementary Figure 3); the patient carrying this mutation had a very short TTFT and died 49 months from diagnosis. Although infrequent, these somatic mutations suggest another possible mechanism of *MYC* involvement in CLL.

Analysis of GEP data comparing patients with 8q24 amplification to patients without identified 5307 genes whose expression was significantly different, and the samples were again ordered based on the correlation of their gene expression pattern with that of the 8q24 amplified samples (Figure 4D, Supplementary Table 8). 8q24 samples and the samples with gene expression similar to 8q24 samples again demonstrated enrichment for gene sets repressed by Polycomb complexes in ES cells, similar to the findings in the 3q26 gain samples. Interestingly, many of the genes differentially regulated in the 3q26 and 8q24 gain samples were shared between them (1290 genes, FDR-corrected p value 3.4×10^{-116}) (Supplementary Table 10).

Deletion of 8p Predicts Short TTFT

Deletion at the 8p locus was observed in 8 of 161 samples (5.0%). The common region of deletion was broad, spanning 11.0–29.6 Mb (Figure 5A). A previous report found that 28% of a small number of 17p-deleted patients also harbored deletion at the 8p locus(31). In our study, 3 of 8 patients with 8p deletion had a co-existent 17p deletion and a 4th had a co-existent 11q deletion. Six of the eight patients had not been treated at the time of sampling, indicating that the deletion occurred de novo. Deletion at 8p was associated with short TTFT, with 7 of 8 patients subsequently undergoing therapy, and rapidly ($p < 0.0001$; Figure 5B). TTFT was short independent of deletions 17p or 11q, although the numbers are small (Figure 5B). Overall survival also appeared poor, with four of eight of these patients deceased, compared to 19 of 153 in the overall cohort, although again the numbers are small.

Co-analysis with GEP available for seven of the eight CLLs with 8p loss identified 807 genes that were differentially regulated, including 63 located on chromosome 8p itself (Figure 5C, Supplementary Table 11). Using GSEA on samples with gene expression correlated with the 8p deletion samples again identified upregulation of genes that are repressed by Polycomb in ES cells, although a significant overlap with the differentially regulated genes in the 3q26 and 8q24 samples was not observed (Supplementary Table 12).

Single Sample GSEA (SSGSEA) Identifies Common Expression Signature Previously Associated with HSCs

Exploratory GSEA analysis of all three CNA groups identified increased expression of gene sets composed of targets of Polycomb-based silencing in ES cells(32). Previous work in ES cells has identified several core components of their gene expression signature: a Polycomb cluster of genes bound by Polycomb complex factors; a Core cluster of genes bound by the pluripotency factors Oct4, Sox2 and Nanog; and a Myc cluster of genes targeted by Myc(32). Differential expression of these clusters has been described in hematopoietic stem cells and a variety of cancers(33). Therefore, in order to further characterize the finding of Polycomb cluster overexpression in our CNA groups, we used SSGSEA to test these gene sets previously reported as ES cell gene sets(32, 34, 35). SSGSEA showed that the CNA group in each case, which included samples with the CNA itself as well as those samples with a positively correlated gene expression pattern, was enriched for gene modules previously associated with self-renewing long-term hematopoietic stem cells (HSCs), specifically showing induction of ES Polycomb (SUZ12, EED, H3K27ME3) and 'core ES' gene sets (WEINBERG_ES_CORE_NINE and WEINBERG_ES_2), together with repression of ES cell gene sets reflecting proliferation, such as MYC and proliferation gene sets (Supplementary Tables 13–15)(32, 34). Control samples in contrast showed the opposite pattern of enrichment, with induction of ES MYC and proliferation modules and repression of ES Polycomb modules, a pattern previously associated with short-term HSCs(32, 34). Interestingly, the same SSGSEA analysis performed on samples with 17p deletion, 11q deletion, high number of CNAs (> 2) and unmutated *IGHV* failed to identify any consistent pattern between the high and low risk groups (Supplementary Tables 16–19).

Therefore, given the similar results among the three CNA groups, we assessed for overlap among the samples with GEP that correlated with each CNA. Here we found substantial overlap among the CLL samples related to each CNA (FDR-corrected p-values: 6.4×10^{-12} for the overlap between 3q26 and 8p, 3.1×10^{-21} for the overlap between 3q26 and 8q24, and 3.2×10^{-7} for the overlap between 8p and 8q24). We therefore investigated whether these samples sharing gene expression patterns correlated with all three CNAs showed any common clinical features or shorter TTFT. We found that the distribution of clinical features in this group was similar to the overall group (Supplementary Figure 4A). However, the samples sharing the CNA gene expression pattern showed a significantly shorter TTFT compared to control samples, with median 61 months compared to 161 months for the control group ($p = 0.03$; Supplementary Figure 4B). These results suggest that patterns of gene expression previously associated with HSC biology may play a heretofore uninvestigated role in CLL. Future work will need to test this hypothesis in other CLL cohorts.

Discussion

We report the results of a large integrated analysis of SNP array screening and gene expression profiling of the CLL genome. Significant advantages of our dataset include the use of an extremely high resolution platform and comparison to matched germline DNA, allowing clear determination of somatic events and filtering of previously undescribed germline CNVs. We find that CLL is quite genomically stable compared with most solid tumors, with a median of only 1 CNA per genome in stably untreated patients, often 13q deletion or trisomy 12. This estimate is lower than earlier studies without matched germline controls(4, 36), but similar to more recent studies that included matched germline analysis(5). The genomic stability of indolent CLL is unsurprising given that many cases display a benign course with minimal progression for years.

We also found that an increasing number of somatic CNAs was predictive of short TTFT, as previously reported(8, 36). This finding was true in the entire cohort and in those lacking 17p and 11q deletions, suggesting that increasing CNAs is independently associated with short TTFT. Similarly, additional CNAs were the major predictor of short TTFT in the context of 13q deletion. The size of 13q deletion and whether it was mono- or bi-allelic have both been reported to have prognostic significance, but in this cohort neither feature was predictive. The most important predictor of TTFT was the presence of any additional somatic CNA defined by SNP array. Our data are therefore an extension of Doehner and colleagues' original observation using FISH(1) but our conclusions are based on a much higher resolution platform, and thus may allow more definitive prognostic prediction.

Although the number of CNAs has prognostic significance, it remains likely that specific recurrent CNAs may target genes important in CLL pathogenesis. We were therefore interested in identifying other genomic regions targeted recurrently in CLL and found three that were significantly associated with requiring therapy in our dataset: amplification of 3q26 focused on *PIK3CA*, amplification of 8q24 focused on the known GWAS cancer risk region near *MYC*, and 8p loss. All three broad chromosomal regions have been reported previously, but the targets of these broad events in CLL were not previously hypothesized. The very high resolution platform used here allowed us to identify very focal CNAs that suggest likely targets in two of these cases, *PIK3CA* and *MYC*. In fact, a recent study by Beroukhim and colleagues that catalogued the CNAs observed in more than 3000 cancers found that amplifications do most commonly involve either the whole chromosome arm or are focal(2), similar to what we observe in CLL. In the Beroukhim study, both *PIK3CA* and *MYC* were targets of recurrent gains in cancer.

To date genomic alterations in PI3K have not been reported in CLL, although both amplifications and activating point mutations occur in solid tumors(21, 22). Here we report amplifications of the *PIK3CA* locus in CLL, but we did not observe activating point mutations. A similar pattern of *PIK3CA* amplification without mutation has been described in mantle cell lymphomas(37), and amplifications in endometrial cancer show a distinct phenotype compared to somatic mutation, suggesting that amplification and somatic mutation may have distinct consequences(38). The PI3 kinase pathway is constitutively(23, 24) and inducibly activated by multiple cell surface signals in CLL, including the B cell receptor pathway(39, 40). Some interest has focused on the question of which PI3 kinase catalytic isoform is most important in CLL, given that the delta isoform is highly expressed and the delta knockout mouse shows impairment in the B cell compartment(39, 41, 42). Our data suggest that the delta isoform of p110 is most commonly in active complex with the p85 regulatory subunit in CLL, but that at least in several samples with alpha amplification, this balance shifts toward alpha. Our data suggest that *PIK3CA* amplification may be one of many mechanisms contributing to PI3K activation in CLL. Currently a delta-specific PI3K inhibitor is showing marked clinical activity in CLL(43); whether pan-PI3K or PI3K alpha inhibitors will have similar potency remains to be determined, as does the effect of *PIK3CA* amplification on the activity of the delta inhibitor. Ultimately, prospective validation of the frequency of *PIK3CA* amplification in CLL will be required to determine its importance in the disease.

A role for *MYC* in the initiation or progression of CLL has been much less clear, although transgenic mice expressing *MYC* together with *BAFF* have recently been reported to develop a CLL-like disease(44). This study also found that higher *MYC* expression in CLL patient samples was associated with shorter TTFT(44). Genomic analyses of Richter's transformation, namely CLL that has transformed to a higher grade lymphoma, have identified *MYC* amplification as a common event thought to be acquired at the time of transformation(45). Here we report *MYC* amplifications in CLL without transformation,

through whole chromosome arm amplification or focal amplification of the 8q24 risk region near *MYC*. We also identify rare somatic mutations in *MYC*. These findings suggest multiple mechanisms of *MYC* activation in CLL, albeit at low frequency. The 8q24 gains identified here while uncommon are associated with short TTFT, and therefore likely a poor prognosis.

We identify two CLLs with focal gains of the 8q24 risk region, in which a SNP (rs2456449) has been associated with germline risk of CLL(25). Multiple studies have shown that this region can act as an enhancer for *MYC* (46–48). These focal amplifications may therefore represent somatic amplification of a germline risk allele, as described previously in neuroblastoma(49). If alleles identified by GWAS truly promote the risk of malignancy, additional similar instances will likely be identified over time.

Interestingly, gene expression analyses of our three CNA groups identified an overlapping set of CLLs with a shared expression pattern associated with induction of ES Polycomb gene sets, repression of ES MYC and proliferation gene sets, and induction of small specific modules of ES ‘core’ factors and targets, all of which have been previously associated with long-term self-renewing HSCs(32). These findings raise the possibility of a role for histone methylation in CLL pathogenesis and an association with HSC biology, but further work will be required to validate this finding and determine its significance to CLL.

In summary, our comprehensive integrated analysis of CLL has characterized three recurrent CNAs associated with reduced TTFT. Two of these CNAs affect *PIK3CA* and *MYC* focally. These CNAs will require validation in uniformly treated prospective cohorts in order to better define their incidence and prognostic significance. These studies together with emerging sequencing data will hopefully define key molecular subgroups of CLL that will clarify prognosis and inform novel therapeutic avenues in the coming years.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the patients who participated in this study as well as the clinic and research staff who assist with sample collection. We thank the Genetic Analysis Platform of the Broad Institute of Harvard and MIT for running the SNP arrays, the DFCI Microarray core facility for gene expression profiling, and the CLL Research Consortium tissue bank for *IGHV* and ZAP-70 results. MH and LM are supported through the CCGD and the Dana-Farber Strategic Plan Initiative. YEW and MC are supported through the CCCB and the Dana-Farber Strategic Plan Initiative. CJW acknowledges support from the Leukemia and Lymphoma Translational Research Program and is a Damon-Runyon Clinical Investigator supported in part by the Damon-Runyon Cancer Research Foundation (CI-38-07). NP is a postdoctoral research fellow of the Fund for Scientific Research-Flanders (FWO Vlaanderen) and a Broad Fellow of the Broad Institute. AR is supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, an NIH Pioneer award, HHMI and the Merkin Foundation for Stem Cell Research at the Broad Institute. JMA is supported by 5PO1-CA120964 and 5P30-CA006516 from the National Institutes of Health. ASF is supported in part by NIH 5 PO1 CA092625. JRB was supported by K23 CA115682 from the National Institutes of Health, and is a Scholar of the American Society of Hematology as well as a Scholar in Clinical Research of the Leukemia and Lymphoma Society. These studies were supported by the Okonow-Lipton Fund, the Melton Fund and the Rosenbach Fund.

References

1. Dohner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. The New England journal of medicine. 2000; 343:1910–6. [PubMed: 11136261]

2. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905. [PubMed: 20164920]
3. Grubor V, Krasnitz A, Troge JE, Meth JL, Lakshmi B, Kendall JT, et al. Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). Blood. 2009; 113:1294–303. [PubMed: 18922857]
4. Gunnarsson R, Isaksson A, Mansouri M, Goransson H, Jansson M, Cahill N, et al. Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK. 2010; 24:211–5.
5. Ouillette P, Collins R, Shakhan S, Li J, Li C, Shedden K, et al. The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. Clinical cancer research : an official journal of the American Association for Cancer Research. 2011; 17:6778–90. [PubMed: 21890456]
6. Ouillette P, Erba H, Kujawski L, Kaminski M, Shedden K, Malek SN. Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14. Cancer research. 2008; 68:1012–21. [PubMed: 18281475]
7. Ouillette P, Fossum S, Parkin B, Ding L, Bockenstedt P, Al-Zoubi A, et al. Aggressive chronic lymphocytic leukemia with elevated genomic complexity is associated with multiple gene defects in the response to DNA double-strand breaks. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010; 16:835–47. [PubMed: 20086003]
8. Ouillette P, Collins R, Shakhan S, Li J, Peres E, Kujawski L, et al. Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. Blood. 2011; 118:3051–61. [PubMed: 21795749]
9. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:20007–12. [PubMed: 18077431]
10. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nature genetics. 2008; 40:1166–74. [PubMed: 18776908]
11. Robinson J, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E, Getz G, et al. Integrative Genomics Viewer Nature Biotechnology. 2010 In Press.
12. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. [PubMed: 12925520]
13. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nature genetics. 2006; 38:500–1. [PubMed: 16642009]
14. Gould J, Getz G, Monti S, Reich M, Mesirov JP. Comparative gene marker selection suite. Bioinformatics. 2006; 22:1924–5. [PubMed: 16709585]
15. Mosca L, Fabris S, Lionetti M, Todoerti K, Agnelli L, Morabito F, et al. Integrative genomics analyses reveal molecularly distinct subgroups of B-cell chronic lymphocytic leukemia patients with 13q14 deletion. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010; 16:5641–53. [PubMed: 20947517]
16. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, et al. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell. 2010; 17:28–40. [PubMed: 20060366]
17. Palamarchuk A, Efanov A, Nazaryan N, Santanam U, Alder H, Rassenti L, et al. 13q14 deletions in CLL involve cooperating tumor suppressors. Blood. 2010; 115:3916–22. [PubMed: 20071661]
18. Parker H, Rose-Zerilli MJ, Parker A, Chaplin T, Wade R, Gardiner A, et al. 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK. 2011; 25:489–97.

19. Lia M, Carette A, Tang H, Shen Q, Mo T, Bhagat G, et al. In Vivo Functional Dissection of the Chromosome 13q14 Tumor Suppressor Locus Reveals That Extent of Deletions Impacts Disease Course and Phenotype. ASH Annual Meeting Abstracts. 2011; 118:468.
20. Setlur SR, Ihm C, Tchinda J, Shams S, Werner L, Cho EK, et al. Comparison of familial and sporadic chronic lymphocytic leukaemia using high resolution array comparative genomic hybridization. British journal of haematology. 2010; 151:336–45. [PubMed: 20812997]
21. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, et al. High frequency of mutations of the PIK3CA gene in human cancers. Science. 2004; 304:554. [PubMed: 15016963]
22. Shayesteh L, Lu Y, Kuo WL, Baldocchi R, Godfrey T, Collins C, et al. PIK3CA is implicated as an oncogene in ovarian cancer. Nature genetics. 1999; 21:99–102. [PubMed: 9916799]
23. Cuni S, Perez-Aciego P, Perez-Chacon G, Vargas JA, Sanchez A, Martin-Saavedra FM, et al. A sustained activation of PI3K/NF-kappaB pathway is critical for the survival of chronic lymphocytic leukemia B cells. Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, UK. 2004; 18:1391–400.
24. Ringshausen I, Schneller F, Bogner C, Hipp S, Duyster J, Peschel C, et al. Constitutively activated phosphatidylinositol-3 kinase (PI-3K) is involved in the defect of apoptosis in B-CLL: association with protein kinase Cdelta. Blood. 2002; 100:3741–8. [PubMed: 12393602]
25. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nature genetics. 2010; 42:132–6. [PubMed: 20062064]
26. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nature genetics. 2008; 40:1204–10. [PubMed: 18758461]
27. Slager SL, Rabe KG, Achenbach SJ, Vachon CM, Goldin LR, Strom SS, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. Blood. 2011; 117:1911–6. [PubMed: 21131588]
28. Chang DW, Claassen GF, Hann SR, Cole MD. The c-Myc transactivation domain is a direct modulator of apoptotic versus proliferative signals. Molecular and cellular biology. 2000; 20:4309–19. [PubMed: 10825194]
29. Yada M, Hatakeyama S, Kamura T, Nishiyama M, Tsunematsu R, Imaki H, et al. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. The EMBO journal. 2004; 23:2116–25. [PubMed: 15103331]
30. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. The New England journal of medicine. 2011; 365:2497–506. [PubMed: 22150006]
31. Forconi F, Rinaldi A, Kwee I, Sozzi E, Raspadori D, Rancoita PM, et al. Genome-wide DNA analysis identifies recurrent imbalances predicting outcome in chronic lymphocytic leukaemia with 17p deletion. British journal of haematology. 2008; 143:532–6. [PubMed: 18752589]
32. Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell. 2010; 143:313–24. [PubMed: 20946988]
33. Rothenberg ME, Clarke MF, Diehn M. The Myc connection: ES cells and cancer. Cell. 2010; 143:184–6. [PubMed: 20946977]
34. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nature genetics. 2008; 40:499–507. [PubMed: 18443585]
35. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008; 134:521–33. [PubMed: 18692474]
36. Kujawski L, Ouillette P, Erba H, Saddler C, Jakubowiak A, Kaminski M, et al. Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. Blood. 2008; 112:1993–2003. [PubMed: 18436738]
37. Psyrri A, Papageorgiou S, Liakata E, Scorilas A, Rontogianni D, Kontos CK, et al. Phosphatidylinositol 3′-kinase catalytic subunit alpha gene amplification contributes to the

pathogenesis of mantle cell lymphoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009; 15:5724–32. [PubMed: 19723646]

38. Salvesen HB, Carter SL, Mannelqvist M, Dutt A, Getz G, Stefansson IM, et al. Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of PI3 kinase activation. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:4834–9. [PubMed: 19261849]
39. Herman SE, Gordon AL, Wagner AJ, Heerema NA, Zhao W, Flynn JM, et al. Phosphatidylinositol 3-kinase-delta inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. Blood. 116:2078–88. [PubMed: 20522708]
40. Niedermeier M, Hennessy BT, Knight ZA, Henneberg M, Hu J, Kurtova AV, et al. Isoform-selective phosphoinositide 3′-kinase inhibitors inhibit CXCR4 signaling and overcome stromal cell-mediated drug resistance in chronic lymphocytic leukemia: a novel therapeutic approach. Blood. 2009; 113:5549–57. [PubMed: 19318683]
41. Lannutti BJ, Meadows SA, Herman SE, Kashishian A, Steiner B, Johnson AJ, et al. CAL-101, a p110{delta} selective phosphatidylinositol-3-kinase inhibitor (PI3K) for the treatment of B cell malignancies inhibits PI3K signaling and cellular viability. Blood.
42. Okkenhaug K, Bilancio A, Farjot G, Priddle H, Sancho S, Peskett E, et al. Impaired B and T cell antigen receptor signaling in p110delta PI 3-kinase mutant mice. Science. 2002; 297:1031–4. [PubMed: 12130661]
43. Furman R, Byrd J, Brown J, Coutre S, Benson D, Wagner-Johnston N, et al. CAL-101, an isoform-selective inhibitor of phosphatidylinositol 3 kinase p110δ, demonstrates clinical activity and pharmacodynamic effects in patients with relapsed or refractory chronic lymphocytic leukemia. Blood. 2010; 116:31.
44. Zhang W, Kater AP, Widhopf GF 2nd, Chuang HY, Enzler T, James DF, et al. B-cell activating factor and v-Myc myelocytomatosis viral oncogene homolog (c-Myc) influence progression of chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:18956–60. [PubMed: 20956327]
45. Rossi D, Spina V, Capello D, Forconi F, Martini M, Rasi S, et al. Molecular History of Richter Syndrome: Origin From a Common Ancestor Cell Already Present at Chronic Lymphocytic Leukemia Diagnosis. Blood. 2010; 116:2425.
46. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nature genetics. 2009; 41:882–4. [PubMed: 19561607]
47. Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, et al. Long-range enhancers on 8q24 regulate c-Myc. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:3001–5. [PubMed: 20133699]
48. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nature genetics. 2009; 41:885–90. [PubMed: 19561604]
49. Wang K, Diskin SJ, Zhang H, Attiyeh EF, Winter C, Hou C, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. Nature. 2011; 469:216–20. [PubMed: 21124317]
50. Asara JM, Christofk HR, Freimark LM, Cantley LC. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. Proteomics. 2008; 8:994–9. [PubMed: 18324724]

Translational Relevance

CLL is a heterogeneous disease in which genetic markers are the most effective determinants of prognosis. In our study we have identified and characterized three recurrent genomic abnormalities that are associated with progressive CLL. Our very high resolution platform allowed us to demonstrate that the targets of two of these abnormalities are *MYC* and *PIK3CA*, both well known oncogenes in solid tumors but lacking a well described role in CLL. Our work provides the foundation for assessing the potential importance of these genetic abnormalities as prognostic markers in prospective clinical trials. Furthermore, drugs that target *PIK3CA* are already in clinical trials in CLL, so determining whether the *PIK3CA* amplifications identified in this work will predict sensitivity to treatment with these drugs will be critically important.

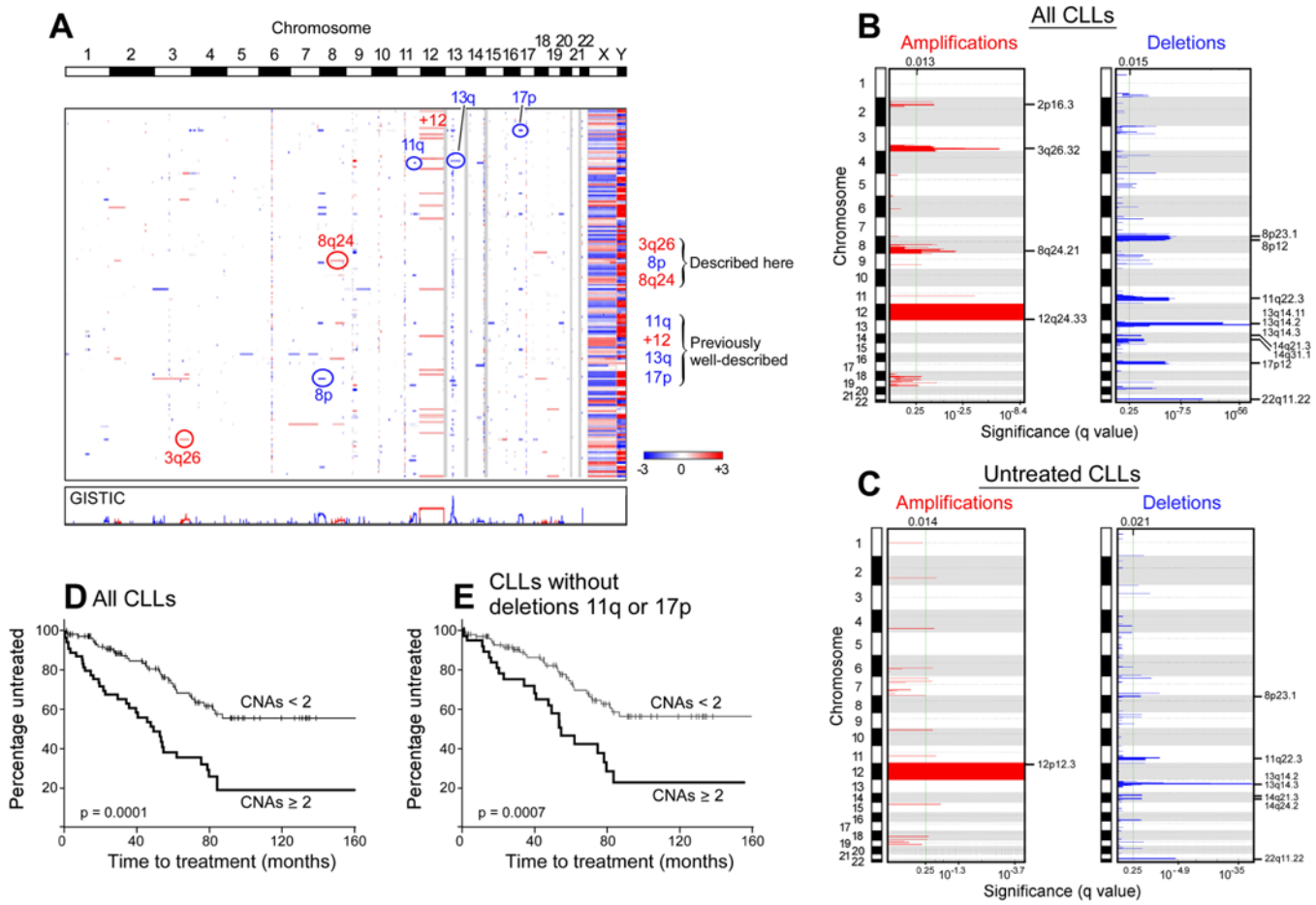

Figure 1.

A. Overview of CNAs identified in CLL. GISTIC track at bottom shows statistically significant recurrent gains and losses. Gradations of gain (red) to loss (blue) are presented as previously described(9). **B**. GISTIC plots of amplifications and deletions identified in all CLL tumor samples, with chromosomal locations labeled on the right. The size of each bar shows the G-score which is the frequency times average amplitude of the aberrations. **C**. GISTIC plots of amplifications and deletions identified in CLL tumor samples from patients who remained untreated, with chromosomal locations labeled on the right. **D**. TTFT based on number of CNAs in the entire cohort. Median 161.2 months for CNAs < 2 (n = 107), and 49.5 months for CNAs > = 2 (n = 53) (p< 0.0001). **E**. TTFT based on number of CNAs in patients without deletions of 11q or 17p. Median 161.2 months for CNAs < 2 (n = 105), and 54.8 months for CNAs > = 2 (n = 38) (p = 0.0007).

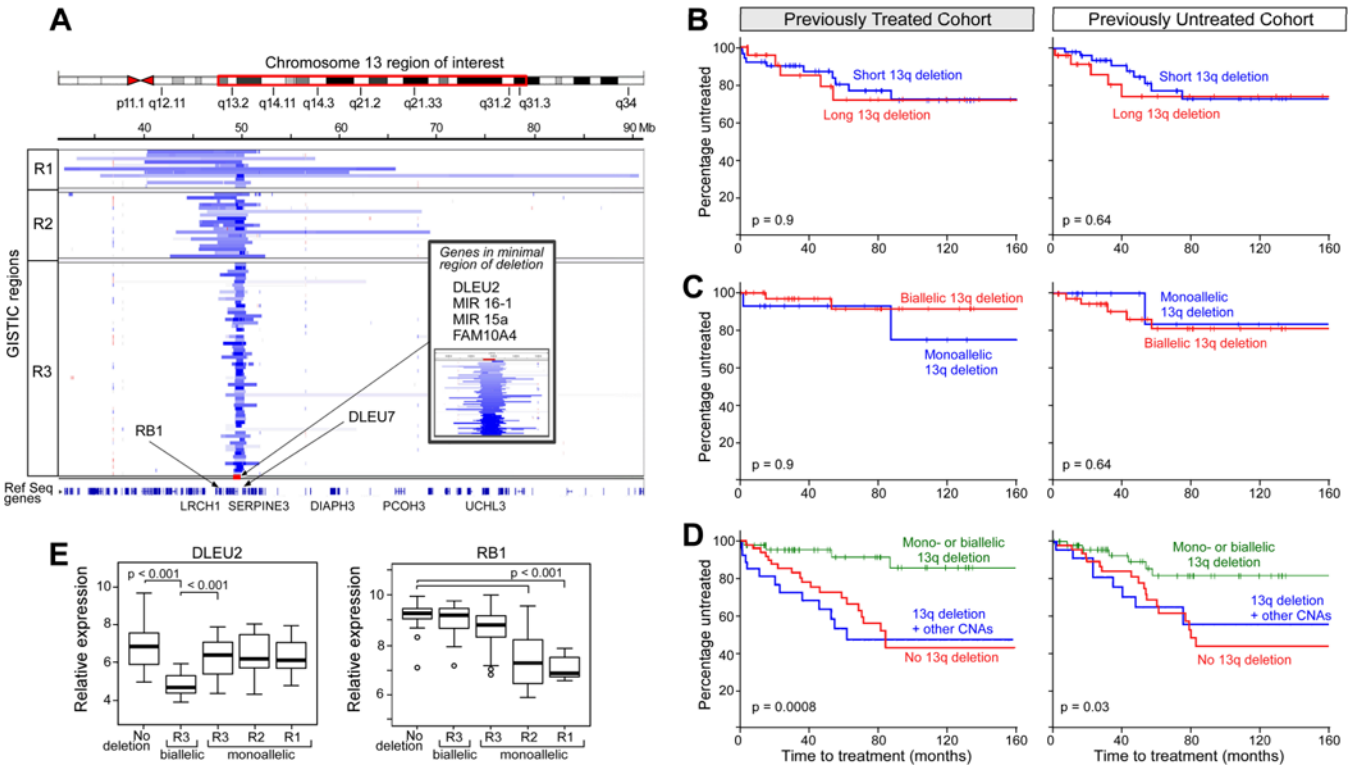

Figure 2.

A. Detailed view of 13q deletion showing GISTIC regions R1, R2 and R3. The red bar on the chromosome indicates the region shown in the expanded figure below. **B**. TTFT for long vs short 13q deletions, divided by whether patients were treated prior to or after sampling. **C**. TTFT for biallelic vs monoallelic 13q deletions, divided by whether patients were treated prior to or after sampling. **D**. TTFT for sole 13q deletion as compared to 13q deletion with any other somatic CNA defined by SNP array, divided by whether patients were treated prior to or after sampling. Previously treated cohort: median not reached for mono- or biallelic 13q deletion (n = 50); 62 months for 13q deletion with other CNAs (n = 27); 84 mos for no 13q deletion (n = 52) (p = 0.0008). Previously untreated cohort: median not reached for mono- or biallelic 13q deletion (n = 52); 40 months for 13q deletion with other CNAs (n = 23); 80 mos for no 13q deletion (n = 49) (p = 0.03). **E**. Boxplots of *DLEU2* and *RB1* gene expression, based on 13q deletion status.

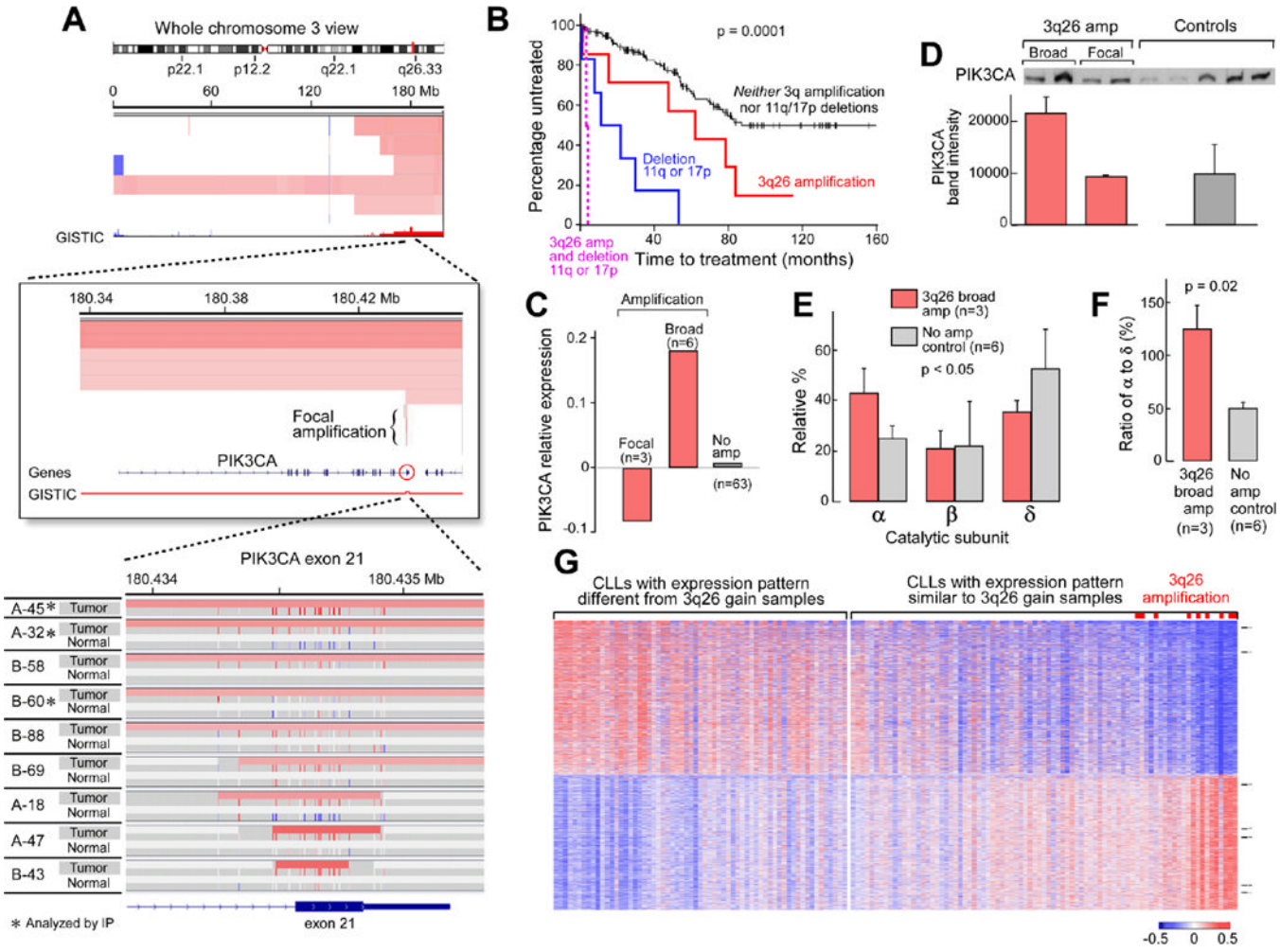

Figure 3.

A. Amplification of 3q26.32, focused on *PIK3CA* exon 21. Lower panel shows segmented (top) and raw (bottom) data for tumor and normal as noted. **B**. TTFT for 3q26 patients with or without deletions of 11q or 17p. Neither, median 87 mos (n = 136); 3q26 amplification alone, 62 mos (n = 7); deletions 11q or 17p alone, 27 mos (n = 15); 3q26 amplification with deletions 11q or 17p, 3.5 mos (n = 2) (p = 0.0001). **C**. *PIK3CA* expression determined by gene expression profiling. **D**. Immunoprecipitation of PIK3CA from two CLLs with broad and two with focal 3q26 amplifications, and 5 controls, with quantitation of bands below. **E**. Percentage of alpha, beta and delta p110 catalytic subunits in complex with p85, in samples with broad 3q26 amplification (n = 3) and controls without amplification (n = 6), as measured by mass spectrometry ($p < 0.05$). **F**. Ratio of alpha:delta p110 catalytic subunit, in complex with p85 regulatory subunit (p = 0.02). N = 3 samples with broad amplifications and 6 control samples, one of which had lower spectral counts(50). **G**. Heat map showing the CLL cohort ordered by correlation of gene expression with the 3q26 gain samples. Samples with 3q26 gain are labeled in red at right, and samples to the right of the gap show positive correlation with the 3q26 expression pattern. Tick marks at right indicate the locations of genes present on chromosome 3q26 (see Supplementary Table 20 for names).

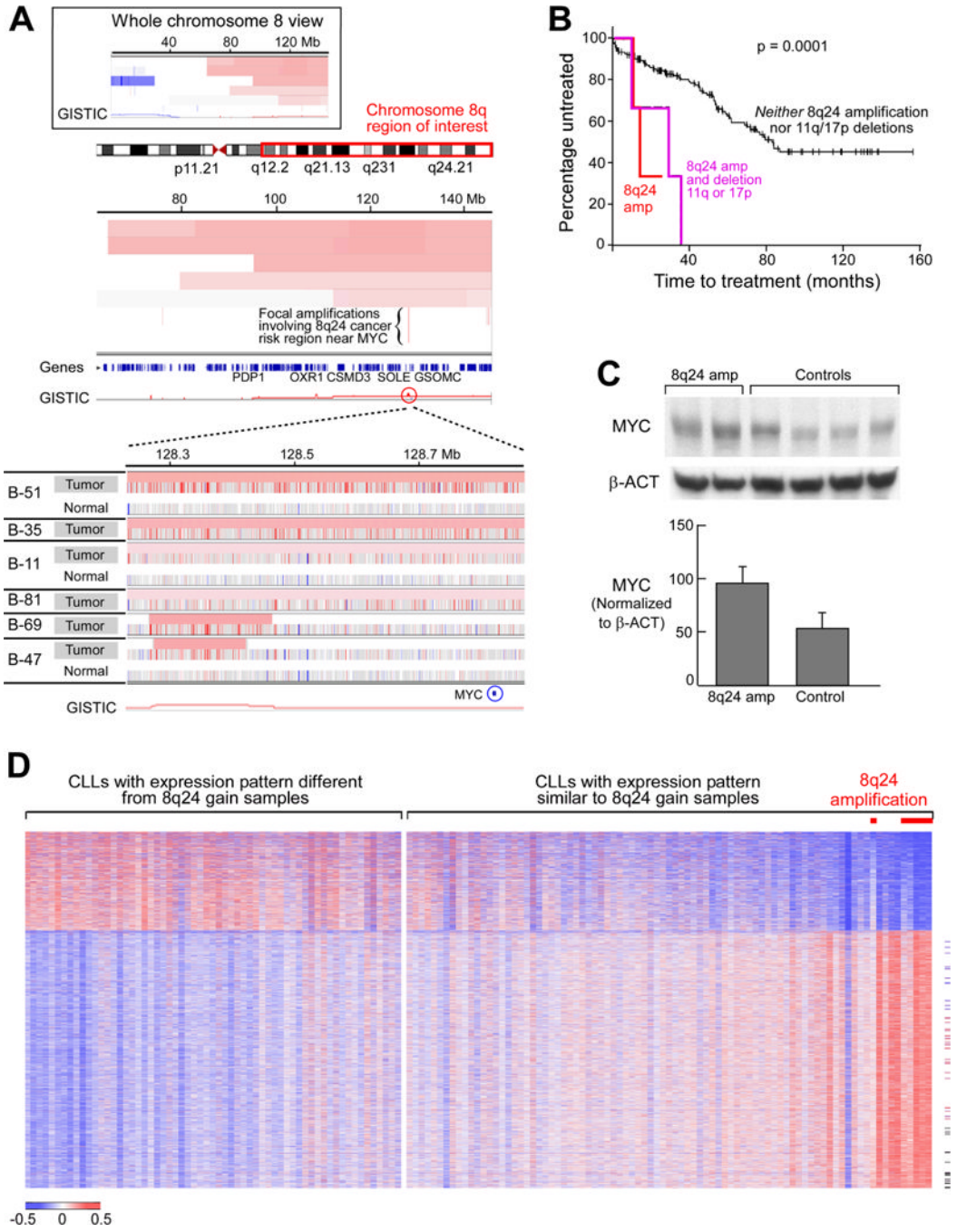

Figure 4.

A. Amplification of 8q24.21, whole chromosome view and focal region. The red bar on the chromosome indicates the region shown in the expanded figure below. Lower panel shows segmented (top) and raw (bottom) data for tumor and normal as noted. **B**. TTFT for 8q24 patients with or without deletions 11q or 17p. Neither, median 83.1 mos; 8q24 amplification with deletions 11q or 17p, 30 mos; 8q24 alone, 13 mos (p = 0.0001). **C**. Western blot showing MYC protein levels in 2 samples with 8q24 gain compared to 4 controls, with quantitation of bands below. **D**. Heat map showing the CLL cohort ordered by correlation of gene expression with the 8q24 gain samples. Samples with 8q24 gain are labeled in red at

right, and samples to the right of the gap show positive correlation with the 8q24 expression pattern. Tick marks at right indicate the locations of genes present on chromosome 8q24 (see Supplementary Table 20 for names).

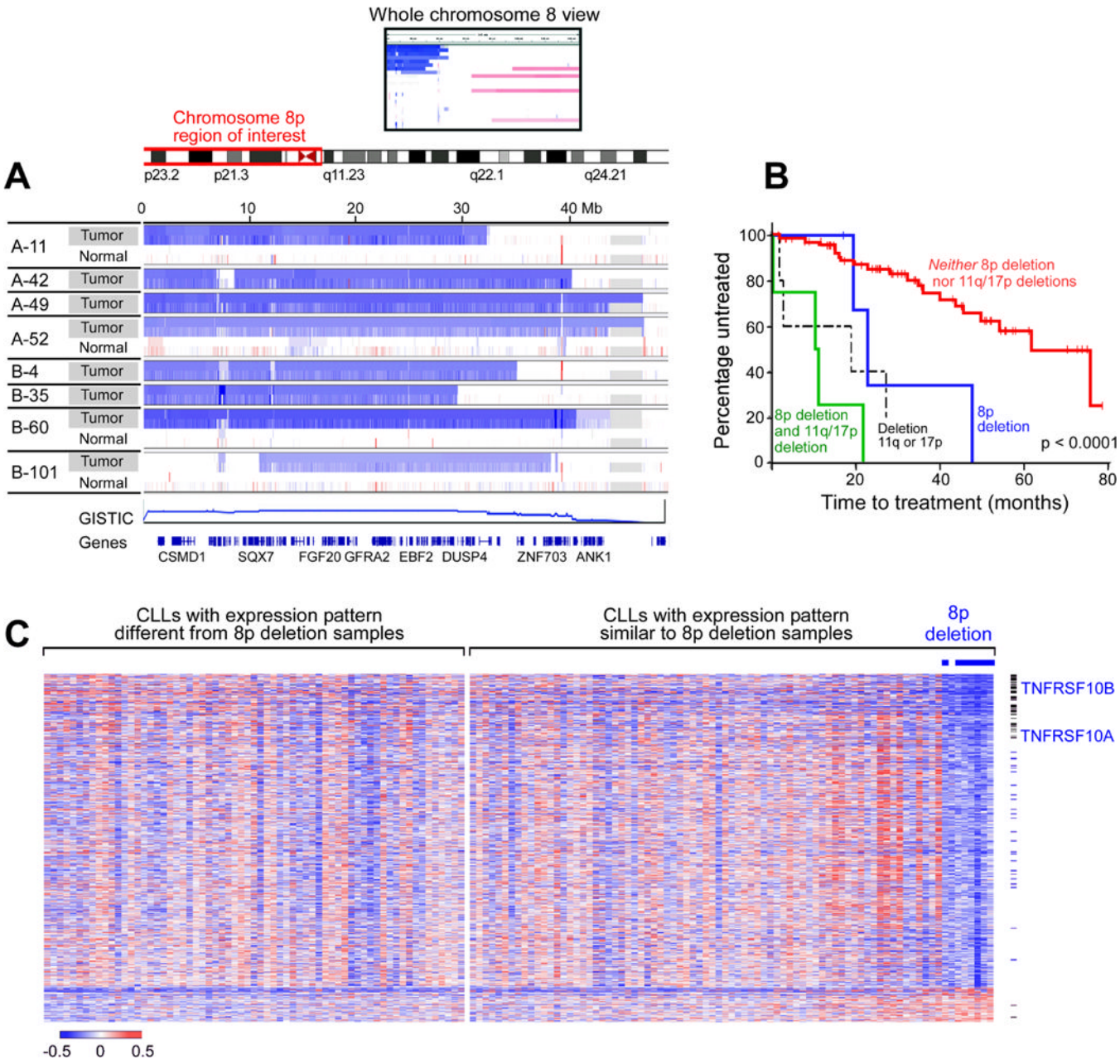

Figure 5.

A. Deletion of 8p, whole chromosome view and focal region. The data show segmented data above raw data for tumor or normal as indicated. **B**. TTFT for deletion 8p patients with or without deletions 11q or 17p. Neither, median 61 mos; 8p deletion alone, 23 mos; deletions 11q or 17p alone, 19 mos; 8p deletion and deletions 11q or 17p, 11 mos ($p < 0.0001$). **C**. Heat map showing the CLL cohort ordered by correlation of gene expression with the 8p deletion samples. Samples with 8p deletion are labeled in blue at right, and samples to the right of the gap show positive correlation with the expression pattern of the 8p deletion samples. Tick marks at right indicate the locations of genes present on chromosome 8p (see Supplementary Table 20 for names).

Table 1

Somatic MDRs Identified by GISTIC

		Genomic Location of CNA	
	N (%)	**Start**	**End**
GISTIC Gains			
2p16.3	3 (1.9)	32,764,999	70,027,926
3q26	9 (5.6)	180,434,350	180,434,928
8q24	6 (3.7)	128,267,747	128,426,358
Trisomy 12	19 (11.8)		
18q	3 (1.9)	50,065,017	76,116,030
Trisomy 19	1 (0.6)		
Trisomy 7	1 (0.6)		
13q	1 (0.6)	18,902,673	19,673,569
16p	1 (0.6)	2,467,076	23,835,347
16q	1 (0.6)	77,763,344	77,909,776
17p	1 (0.6)	21,618,001	21,819,902
GISTIC Losses			
1q44 Loss	4 (2.5)	243,904,769	244,959,448
3p14.2 Loss	1 (0.6)	34,513	85,197,968
3p26.2 Loss	1 (0.6)	34,513	7,303,192
4q35.2 Loss	1 (0.6)	187,977,019	190,299,620
5q14.3 Loss	1 (0.6)	86,702,690	132,661,780
7q34 Loss	7 (4.3)	141,945,240	142,213,197
8p23.1 Loss	8 (5.0)	10,970,612	29,620,652
11q22.3 Loss	10 (6.2)	102,877,806	113,811,014
13q14.11 R1	11 (6.8)	40,438,503	40,478,258
13q14.2 R2	19 (11.8)	47,405,627	47,786,494
13q14.3 R3	61 (37.9)	49,538,817	50,081,506
Biallelic	22 (13.7)		
14q31.1	5 (3.1)	77,003,257	77,535,100
		91,704,917	93,186,479
17p12	9 (5.6)	7,258,530	7,990,010
20p Arm	2 (1.2)	9,756	26,223,105
22q11.22	53 (32.9)	21,507,073	21,555,877