blood

Prepublished online Oct 15, 2008;
doi:10.1182/blood-2008-05-158865

Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA)

Vladimir Grubor, Alex Krasnitz, Jennifer E. Troge, Jennifer L. Meth, B. Lakshmi, Jude T. Kendall, Boris Yamrom, Garrick Alex, Deepa Pai, Nicholas Navin, Lisa A. Hufnagel, Yoon-Ha Lee, Kerry Cook, Steven L. Allen, Kanti R. Rai, Rajendra N. Damle, Carlo Calissano, Nicholas Chiorazzi, Michael Wigler and Diane Esposito

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published semimonthly by the American Society of Hematology, 1900 M St, NW, Suite 200, Washington DC 20036.

Novel genomic alterations and clonal evolution in Chronic Lymphocytic Leukemia revealed by Representational Oligonucleotide Microarray Analysis (ROMA)

Short title: ROMA analysis of CLL

Vladimir Grubor[1], Alex Krasnitz[1], Jennifer E. Troge[1], Jennifer L. Meth[1], B. Lakshmi[1], Jude T. Kendall[1], Boris Yamrom[1], Garrick Alex[1], Deepa Pai[1], Nicholas Navin[1], Lisa A. Hufnagel[l], Yoon-Ha Lee[1], Kerry Cook[1], Steven L. Allen[2,3], Kanti R. Rai[2,4], Rajendra N. Damle[2], Carlo Calissano[2], Nicholas Chiorazzi[2,3], Michael Wigler[1] and Diane Esposito[1].
[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY;
[2]The Feinstein Institute for Medical Research, North Shore – LIJ Health System, Manhasset, NY;
[3]North Shore University Hospital, North Shore – LIJ Health System, Manhasset, NY;
[4]Long Island Jewish Medical Center, North Shore-LIJ Health System, New Hyde Park, NY.

Address correspondence to: Diane Esposito (esposito@cshl.edu; Ph:1-516-367-8378) or Vladimir Grubor (grubor@cshl.edu; Ph:1-516-367-8382),
Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY, 11724, USA
Fax:1-516-367-8381

Scientific Heading: NEOPLASIA

Abstract

We examined copy number changes in the genomes of B cells from 58 patients with chronic lymphocytic leukemia (CLL) using representational oligonucleotide microarray analysis (ROMA), a form of comparative genomic hybridization (CGH), at a resolution exceeding previously published studies. We observed at least one genomic lesion in each CLL sample and considerable variation in the number of abnormalities from case to case. Virtually all abnormalities previously reported were also observed here, most of which were indeed highly recurrent. We observed the boundaries of known events with greater clarity and identified previously undescribed lesions, some of which were recurrent. We profiled the genomes of CLL cells separated by the surface marker CD38, and found evidence of distinct subclones of CLL within the same patient. We discuss the potential applications of high resolution CGH analysis in a clinical setting.

Introduction

Chronic lymphocytic leukemia (CLL), the most common form of adult onset leukemia in the western world, [1,2] is typically an indolent disease. While not all disease will progress to an advanced stage within the otherwise normal lifespan of the patient, CLL can evolve over time into a more dangerous and lethal disease. Patients who first present with CLL are usually not treated, as no statistically significant benefits from treatment in early stages of the disease have yet been demonstrated. Therefore, the ability to identify patients at greatest risk, i.e., those harboring lesions associated with poor prognosis, but who have not yet progressed, would offer an opportunity for selective and more effective therapy. Survival might be increased by treating patients with markers of advanced disease and sparing those not in need the toxic effects of therapy.

The heterogeniety of disease progression led to the Rai and Binet staging systems which remain the standard for tracking the disease and evaluating conditions for treatment. [3,4] However, neither method can project the course of the disease in those patients diagnosed at early stages. Currently, several molecular markers, such as the presence or absence of IgVH somatic mutations, and expression of CD38 and 70-kDa zeta-associated protein (ZAP-70), appear to have prognostic value, [1,5-8] though each are limited in their ability to predict disease progression, survival, and resistance to therapy.

Improved prognosis might also be achievable by genome analysis. Cytogenetics, [9-11] fluorescent in situ hybridization (FISH), [12,13] and comparative genome hybridization (CGH) [14] have revealed DNA segment gains (e.g., partial or complete trisomy 12) and deletions (e.g., 13q14.2, 11q22-q23, 17p13, and 6q21), which occur sporadically in CLL.

Some of these loci correlate with prognostic outcomes, but to varying degrees. [10,12,15,16] The sparcity of evidence linking these loci with specific genes indicates our incomplete understanding of the disease, and reflects the inadequacy of present tools for assessing chromosomal damage. We therefore conducted the present study with the aim of better describing the genomic abnormalities that occur in CLL, and to enhance the understanding of the ongoing evolution of genetic lesions in patients with CLL.

We began our initial study of the genomic landscape in close to sixty samples of CLL. We compared the leukemic genome to the patient's normal DNA using a high resolution CGH technique called representational oligonucleotide microarray analysis (ROMA).[17,18] We designed oligonucleotide hybridization microarrays of 85,000 and 390,000 probes. On average, the resolution of the 85K and 390K arrays is a probe every 35kb and 9kb, respectively. In principle, each probe is a detector capable of measuring the relative "gene copy number" in a leukemic genome, however to infer true copy number changes with higher confidence we used no fewer than six consecutive probes. The resolution of our study still exceeds previously published CGH studies on CLL. We have also examined some CLL samples at a resolution of 2.1 million probes to understand how the landscape changes when using an even-higher sensitivity In comparison to previous studies, we observed far more cases of CLL with lesions, and more lesions per case.

The resolution by ROMA is so high and the method is so sensitive that we can examine the clonal heterogeneity of CLL within the same patient from mixed sub-populations. The presence of greater than 30% of B-cells with the CD38 cell surface marker has been associated with poor outcome in CLL.[5] It is an open question whether

this reflects genetic heterogeneity and possibly clonal evolution. To investigate this possibility, we analyzed $CD38^+$ and $CD38^-$ fractions from individual patients and demonstrated that three out of the four patients examined had undergone intraclonal diversification leading to new subclones of appreciable size.

Our studies indicate that complete analyses of genome stability and prognosis require high-resolution comparative genomic hybridization.

Patients and Methods

Patient Samples

The Institutional Review Board of the North Shore - LIJ Health System approved these studies. After obtaining informed consent in accordance with the Declaration of Helsinki, 58 patients with CLL, diagnosed according to NCI Working Group criteria, were studied. Venous blood was taken and peripheral blood mononuclear cells (PBMCs) were separated by Ficoll-Hypaque density gradient centrifugation. Next, B cells were isolated by negative selection using a B-cell isolation kit (Miltenyi, Auburn, Ca), yielding fractions with >92% $CD19^+$ cells.

CLL clones were analyzed for IgVH gene mutations and CD38 expression as described.[5,19] CLL clones expressing IgVH genes differing by ≥2% from the most similar germline gene were defined as "Mutated CLL" (M-CLL), and clones expressing IgVH genes with <2% difference from germline gene as "Unmutated CLL" (U-CLL). Clones containing ≥30% CD38-expressing cells were considered "$CD38^+$" and those with <30% "$CD38^-$''. In some instances, PBMCs from CLL patients were labeled with mouse monoclonal IgG1 anti-human CD5 FITC , CD38 PE and CD19 APC (BD Biosciences)

and a BD FACSAria ™ was used to collect in parallel $CD19^+CD5^+CD38^+$ and $CD19^+CD5^+CD38^-$ gated fractions. After sorting, cells were washed three times in PBS, pelleted and stored at -80ºC until DNA extraction for ROMA was performed.

DNA extraction

Genomic DNA was extracted from purified B cells and PMN cells using Puregene genomic DNA purification kits (Gentra Systems, Qiagen Inc., Valencia, CA), according to manufacturer's instructions. Genomic DNA was subsequently stored at –20ºC until used.

ROMA

We examined all CLL genomes with ROMA, a form of CGH that utilizes genomic representations. [18,20] Complexity-reducing representations of genomic DNA were hybridized to microarrays of 50-mer oligonucleotide probes designed from the sequence of the human genome. [21] Samples were mainly hybridized on two platforms: 85K arrays based on Bgl II representations, and 390K arrays based on DpnII representations depleted of DpnII fragments containing AluI sites ("depleted" representations). [18] Array probes were chosen to be complementary to the complexity-reduced representations. All arrays were manufactured by NimbleGen (NimbleGen, Madison, WI).

ROMA greatly increases signal-to-noise ratios in CGH, and diminishes the amount of sample needed for analysis. All hybridizations were performed in color reversal to prevent color bias and assure data quality. [22,23] A few samples were hybridized without representation, and without color reversal, on NimbleGen's high-density, 2.1 million probe, prototype array (HD2).

The preparation of genomic representations, labelings, and hybridizations were performed as described previously. [18,20,24] Briefly, complexity-reduced representations, consisting of small (200-1200 bp for the 85K and 150-400 bp for 390K) fragments were amplified by adaptor-mediated PCR of genomic DNA. DNA samples (2 µg) were labeled either with Cy5-dCTP or Cy3-dCTP using Amersham-Pharmacia MegaPrime labeling kit (Amersham Biosciences, Piscataway, NJ), and competitively hybridized to each other on the same slide. Each sample genome was analyzed in duplicate, swapping the Cy5 and Cy3 dyes with the control (i.e. "color reversal"). Hybridizations consisted of 35 µL of hybridization solution (37% formamide, 4x SSC, 0.1%SDS, and labeled DNA). Samples were denatured in an MJ Research Tetrad (Bio-Rad, Hercules, CA) at 95°C for 5 min, and then pre-annealed at 37°C for no more than 30 min. The solution was then applied to the microarray and hybridized under a coverslip in an oven at 42°C for 14 to 16 h. Thereafter, slides were washed 1 min in 0.2% SDS/0.2x SSC, 30 sec in 0.2x SSC, and 30 sec in 0.05x SSC. Slides were dried by centrifugation and scanned immediately. An Axon GenePix 4000B scanner (MDS Analytical Technologies, Toronto, Canada) was used with a pixel size of 5 µm.

A limited number of samples were hybridized to a prototype HD2 array. Briefly, 1µg of both CLL cells and corresponding control PMN DNAs were mixed with either 5' Cy5- or 5'Cy3- labeled random nanomers (TriLink, San Diego, CA) to a final concentration of 9pg/µL, in 100 µL of the 9-mer buffer (50 mM Tris, 5mM $MgCl_2$, 1.75 µL/mL β-mercaptoethanol). Samples were denatured for 10 minutes at 100^{o}C, followed by the addition of 20 µL of labeling buffer (10mM Tris, 1mM EDTA, 200 µM dNTPs, 100 units of Klenow). The samples were incubated at 37^{o}C for three hours and then

isopropanol precipitated. 30 µg of each Cy5- and Cy3-labeled DNAs were competitively hybridized at 42°C for three days in a NimbleGen 12-bay hybridization system (NimbleGen, Madison, WI). The slides were washed and scanned as per procedure described above.

Informatics

Microarrays were scanned and gridded using GenePix Pro 4.0 software (MDS Analytical Technologies, Toronto, Canada) and data were imported into S-Plus 2000 analysis software (Insightful, Seattle, WA). The data were normalized using a lowess curve-fitting algorithm, followed by a local normalization (previously described in Hicks et al.).[20] After placement in genome order, the mean of log ratios was computed for color reversal experiments for each sample. All microarray data has been deposited with Gene Expression Omnibus (GEO) under accession number GSE12794.

Segmentation was performed on the above-described data. Segments are defined as non-overlapping, genomic regions where copy number has changed. Our segmentation method is based on the minimization of "the square-sum of differences between log-ratios and means" (squared deviation) over segments larger than 4 probes in size. Initially, the segmenter searches for breakpoints that might be boundaries of segments. The first known breakpoint on a given chromosome is its first probe. For a given breakpoint, a 100-probe window to its right is selected. The sum of squared deviations of the flanking probes is calculated for each probe within this window. A probe whose squared deviation value produces a local minimum with respect to its neighbors, and is below a threshold of 95% of the square deviation within a window, is accepted as a new, known breakpoint. Whenever a probe is found below the threshold in

the newly defined region, the segmenter recursively breaks said region into two pieces until it cannot find any further breakpoints therein. If no breakpoints are found, the 100-probe window is shifted by half its size and this procedure continues until a chromosome end is reached.

Initial segments are constructed using found breakpoints. Each segment and its neighbors are validated for significance by the Kolmogorov-Smirnov (K-S) algorithm. If the p-value of compared segments is less than 10^{-5}, then said segment is accepted as real. If not, the segments are merged. The segmenter also reports statistics such as mean, standard deviation, and median for each segment. We viewed graphs of all ratio data and the algorithmically derived segmentation patterns for each sample (Figures 1, 2, 3 and 4 for examples) to assure data quality.

Segmented data were further annotated with a script we developed to query a local UCSC hg18 database mirror to annotate segments for genome objects (i.e. genes, RNA genes, pathway information, gene ontology (GO) terms, etc), within, as well as spanning, breakpoints (not shown) to facilitate further data analysis.

Frequency plots were computed on the segmented data from 58 CLL samples hybridized to 390K arrays, using 1.1 and 0.9 (1/1.1) as upper and lower cut-off values (Figure 5). A copy number variant (CNV) database based on our ROMA platforms was used to ensure the lesions observed in our dataset are not in fact CNVs. CNV frequencies were determined from a set of 500 profiles of cancer-free genomes, hybridized on the NimbleGen 85K array platform.

Results

Strategic Approach

CLL cells and neutrophils (PMNs) were prepared from peripheral blood samples as described in Patients and Methods. We compared the respective DNAs, using patient PMNs as the normal genome. By comparing the CLL genome to the normal genome from the same patient, as opposed to an unrelated normal, we intrinsically avoid detecting copy number mutations that are frequent occurrences in the human gene pool, and that might otherwise be mistaken for recurrent genomic lesions in CLL.

Although we compare CLL to normal cells from the same person, it is still possible to confuse a copy number variation with a leukemic lesion because loss-of-heterozygosity (LOH) in the leukemia, such as arising by gene conversion, could unmask a heterozygous copy number variant present in the patient's germ-line. To guard against this, we also compared the PMN genome to an unrelated normal, by which method we can detect most germ-line copy number variants that could be unmasked by LOH. Comparing PMN DNA to an unrelated normal also rules out a genomic abnormality arising in the PMN lineage (see, for example, Figure S1). Even this expedient, meant to avoid confusing copy number variation and leukemic lesions, fails when both patient, normal DNA and the DNA from an unrelated, normal control share the same copy number polymorphism. To further safeguard against this last source of error in interpretation, we further compared our results to a database of copy number variants derived from 500 normal humans (see, for example, Table S1 and Figure 5).

Most of our samples were analyzed on two platforms, 85K and 390K ROMA arrays. This has given us independent validation of variation seen at the 85K resolution

(addressing false positives), and enabled us to reliably assess the value of increased resolution (addressing false negatives). For example, Figure 2 illustrates a deletion at *CDKN2A* (*p16*$^{-INK4}$) in CLL 189 that is detected as a segment in the 390K data, but missed as a segment in the 85K data, although it is plainly present by visual inspection. Our estimate of the false positive rate for the 85K array data is 3% (1/34) based on a discrepancy between events detected with the 85K array and not detected with the 390K array on identical samples. We do not have an independent estimate of the false positive rate for the 390K array data, but we have no reason to believe it is greater than that of the 85K array.

Detailed Summary of ROMA Data

The number of lesions detected in samples is highly variable (for example, see Figure 1). The leukemia sample CLL 334 of Figure 1 did not display any lesions at 390K (except of course for rearrangements at loci encoding immunoglobulins). However, even that sample displayed genomic lesions when analyzed on the 2.1 million high resolution array (see Figure S2 HD2).

Supplementary Table S2 (segmented data summary) and Table S1 (frequency plot summary) contain our summary of findings from hybridizations performed on 390K microarrays (58 CLL samples), including the boundaries of all leukemic events. We defined the minimal regions of overlap for all the recurrent lesions, determined their frequency and the number of genes therein, and compared them to the frequencies of known CNVs. We excluded from Table S1 rearrangements at the immunoglobulin loci but not for alpha (13q14.2) and beta (7q34) T-cell receptor (TCR) loci. TCR is known to recombine in malignant B-cells, though the mechanism permitting this is yet

unknown.[25,26] Figure 5 is a graphical representation of all the segmented data from the 58 profiles, including the immunoglobulin loci. Thus, the height at each locus reflects the number of times an event had been observed there. This figure also contains frequency plots of CNVs derived from a study of 500 normal humans. There was no significant overlap between the set of known CNVs and the genomic changes we observe in CLL.

With the exception of the deletion at 6q21, we observed all the previously reported major cytogenetic imbalances, and in many cases to a higher resolution than found in the literature. The majority of lesions (315/419) are deletions and not amplifications (Figure 5), which is typical of CLL.

The common lesions include 11q, 13q, 17p deletions as well as trisomy 12. Previously published reports observed the deletion at 6q21 in ~ 1.5-8% of samples. [12,27-31] Since we have sufficient probe coverage (roughly 1000 probes in the 9.1Mb region of 6q21), if this abnormality were present in our sample dataset, we would have observed it. The lack of observance of this lesion is possibly due to sampling error.

There are eighteen (18) distinct regions where we have observed recurrent copy number mutations (duplications or deletions), two of them novel. Table 1 (an abridged version of Table S1) depicts our new information at recurrent loci. It includes the two newly identified regions, as well as recurrent regions, which are narrower due to our higher resolution. [32-36] Novel regions, highlighted in yellow, are a 3.6 Mb deletion at 8p21.2-p12 and a 587 kb deletion at 2q37.1 including genes *TRIM35* and *SP100/110/140*, respectively. Of the refined regions, a 249kb region at 9p21.3 spanning the *CDKN2A* (*p16*$^{-INK4}$) and a 156kb region at 18q23 containing *NFATC1* are particularly interesting. In the case of *NFATC1*, the minimal region of overlap spans that single gene.

Since breakpoints of both deletions and amplifications can disrupt the structure of a gene, we have analyzed all genes that span the breakpoints of deletions as well as amplifications. Genes were ranked according to how frequently breakpoints occur within them in all 58 samples (data not shown). Although no breakpoints were found within genes of known clinical significance (i.e. *ATM, TP53, miR-15a/16-1*), they were found to occur within genes flanking them. Furthermore, breakpoints were also frequently found in or near areas of segmental duplications.

Comparing ROMA to Classical Cytogenetics

The power of ROMA is further illustrated by its overall ability to detect lesions. Only 1.7% (1/58) of our sample set did not have observable lesions, when analyzed with the 85K and the 390K arrays, as compared to about 20% using FISH, 17% using chromosomal G-banding, and 15% utilizing other CGH platforms. [7,10-12,33-35] The increased resolution of ROMA allows the observation of lesions too fine to be identified with cytogenetics/FISH technologies or lower resolution CGH technologies. While the median size of lesions we observe is 933 kb, the minimal lesion size observed is just 20 kb (Row 17, Table S2). Previous claims of smallest observed lesions using CGH on CLL were 18 kb on a 644 probe BAC/PAC array and 70 kb on a 44K oligonucleotide array.[33,37] These groups however, resorted to two (2) probe confidence intervals to make these claims, whereas we use four (4), giving us much higher confidence in our calls.

As might be expected, using a still-higher resolution platform could reveal additional lesions, too fine to be observed with current generation technologies. To explore this, we hybridized our most stable CLL genome (see Figure 1, CLL334) to a high-density, 2.1 million probe, prototype array (HD2). CLL334 exhibited no discernible

lesions on either the 85K or the 390K array, apart from *IGKC* and *IGH* rearrangements, at 2p11.2 and 14q32.33 respectively. Hybridizing this sample to the HD2 array reveals multiple lesions, some of which occur within larger regions in our dataset. One example is an 8.3kb event at 2q37.2 spanning 8 probes, too fine to be observed even by the 390K array on CLL334 (Figure S2).

Clonal cell population analysis (CD38)

Elevated numbers of $CD38^+$ cells within a CLL patient's B-cell population have been associated with poor prognosis. [5] It is as yet unclear whether the $CD38^+$ cells arise from genetically distinct subpopulations of the CLL clone. To test this, we analyzed separated $CD38^+$ and $CD38^-$ fractions of four CLL samples. Copy number differences were detected in 3/4 samples (Figure 3 and Table 2) at various loci throughout the genome, some of which are of clinical relevance (i.e. *ATM* and *TP53*). Since we assayed $CD38^{+/-}$ fractions in a small number of samples, conclusions cannot yet be drawn on the role of certain loci in generating this diversity. However, we have clearly provided evidence of continuing genetic evolution in the CLL clones of some patients, and such continuing evolution may be related to disease outcome.

Discussion

In recent years, CGH has emerged as a powerful tool for detecting chromosomal duplications and deletions at a greater resolution than cytogenetics. Classical cytogenetics has identified gross regions of genomic instability in CLL, e.g. the common lesions, del 13q14.3, trisomy 12, del 11q22.3-23.2, and del 17p13.1. Enhanced cytogenetic techniques, such as refined G-banding, led to the narrowing of these

lesions.[11,14] CGH can detect novel lesions, ascertain the frequencies of gains and losses with greater accuracy, and pinpoint candidate genes associated with the disease within known regions of recurrent abnormality. [33-35,37-39] Therefore, based on previous experiences, [14,40,41] it was reasonable to expect that increased resolution would yield more accurate delineation of previously described lesions, as well as identification of new ones.

We used a high-resolution CGH to study CLL in order to consolidate existing knowledge of its genetics and to offer new insights into the nature of the disease. However, there is a danger to the application of high resolution CGH techniques. Previous comparative studies, performed at lower resolution, have ignored the issue of normal variation in the human genome. This is dangerous because at the scale we used to scan CLL the human genome is teeming with copy number variations. [24,42] We took several precautions to guard against this. First and foremost we compared the CLL genome to the normal genome from the same patient. Additional steps to guard against mistaking a genome copy number change in CLL with a copy number polymorphism are described in "Strategic Approach" and the discussion of Figure 5 in the "Results" section.

Our results with high resolution arrays validated all but one of the previous set of known CLL genomic lesions. We confirmed that even at high resolution deletions are more abundant than amplifications. Previous studies report lesions in about 80% of cases,[33-35,39] but we observed genomic lesions in all CLL samples. We saw lesions at most known loci at higher resolution than before, for example further delineating the complex epicenter of the highly recurrent deletion on 13q and shortening the list of candidate genes at other loci (see Table 1). Although, the smallest lesion we observe at the 13q region is 60kb in size, the minimal region of overlap from the frequency plot is

just 26kb in size and spans *miR-15a/16-1* (Row 278, Table S1). Additionally, we have observed multiple, discrete, genomic alterations in the 13q region including *miR-15a/16-1*, *Rb* and others (Figure 4). This observation suggests even greater complexity of lesions in the 13q region.

Additionally, we saw two recurrent lesions at new loci 2q37.1 and 8p21.2-p12, and many more genomic lesions at loci that were not recurrent. Our results suggest that the diversity of genomic aberrations in CLL is much greater than previously appreciated.

We used both a 390K array and an 85K array on most of our samples. By and large, the data sets agreed, but the 390K data showed somewhat more events (Figure 2). Still higher levels of resolution are now possible, and so we hybridized a limited number of patient samples (5) utilizing our HD2 prototype array with 2.1 million probes, seeking to ascertain whether additional lesions could be observed. One of the patients studied, CLL334, exhibited no discernible lesions by either 85K or 390K analyses, other than rearrangements at IGKC and IGH (Figure 1). However using the HD2 array, vastly more detail was observed, even at previously reported loci (see Figure S2 for one such example). We envision future studies utilizing HD2 will aid in narrowing down lesion breakpoints as well as uncovering many novel lesions that we did not observe with the 85K or 390K CGH platforms.

The amplitude of copy number changes (as observed in our figures) in CLL is often small, suggesting intraclonal heterogeneity. We estimate from doping experiments that we can observe lesions present in a minimum 30% of the total cell population. To find clearer evidence of intraclonal heterogeneity within patients, we looked for genomic differences between $CD38^+$ and $CD38^-$ populations in the same patient. We chose the

CD38 activation marker as CLL can differ in the proportion of cells expressing CD38, and patients with ≥30% $CD38^+$ cells have an unfavorable prognosis. [5,7]

If subclones within a patient harbor different genetic lesions and they have different proportions of cells expressing the CD38 marker at any given point in time, then we expect to observe these genomic differences by comparing $CD38^+$ and $CD38^-$ fractions. Indeed, we observed copy number differences between $CD38^+$ and $CD38^-$ fractions in three out of four cases (Figure 3). Since the CD38 marker may be transiently expressed, [43,44] our observation suggests that subclones of CLL spend differing amounts of time in the activated $CD38^+$ state, compared to other clonal members that may be $CD38^+$ or $CD38^-$.

This type of analysis enabled us to "time" the occurrence of events, as events not shared between two populations must occur subsequently to their divergence. In one case, this involved a loss of the p53 locus in the $CD38^+$ fraction, a marker that was not observed in the parallel $CD38^-$ fraction (Table 2). More generally, our observations point to the possibility of monitoring an aspect of the evolution of the disease that might have profound clinical significance. Within an overall apparently constant leukemic burden, the outgrowth of a subclone with additional genomic lesions might signal the start of a new phase of the disease. Additional studies, combining data on fractionated subpopulations with clinical outcomes, are needed to test this hypothesis.

In summary, we have demonstrated that ROMA is a highly sensitive CGH method to examine genomic changes in CLL. We have detected novel lesions, ascertained the frequencies of gains and losses with greater accuracy, and pinpointed candidate genes. The apparent continuing evolution of clones of CLL within a patient

may lead to improved understanding of the disease and the ability to identify patients at-risk. Overall, the capabilities we have demonstrated here offer opportunities for selective patient treatment and the identification of new therapeutic targets.

Acknowledgements

This work was supported by grants to M.W. from the Karches Family Foundation and the Simons Foundation. M.W. is an American Cancer Society Research Professor. N.C. and K.R.R. received support from the Karches Family Foundation, the Prince Foundation, and the Marks Foundation. We wish to thank Linda Rodgers, Patty Bird, Michael Riggs, Christopher Algieri, and Anthony Leotta (CSHL) and Herb Borerro and Stella Stefanova (FIMR) for their technical support. All authors declare no competing financial interests.

Specific Author Contributions:

V.G.: Designed and performed research, analyzed data, contributed data analysis tools, wrote the paper
A.K.: Contributed data analysis tools, analyzed data
J.T.: Performed research
J.M.: Prepared the patient samples, performed research
B.L.: Contributed data analysis tools
J.K.: Contributed data analysis tools
B.Y.: Contributed data analysis tools
G.A.: Performed research
D.P.: Performed research
N.N.: Performed research
L.H.: Performed research
Y.L: Contributed data analysis tools
K.C.: Performed research
S.A.: Collected patient samples
K.R.: Collected patient samples
R.N.D: Performed research
C.C.: Performed research
N.C.: Designed research, prepared patient samples, wrote the paper
M.W.: Designed research, wrote the paper
D.E.: Designed and performed research, analyzed data, wrote the paper

References

1. Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. N Engl J Med. 2005;352:804-815.
2. American Cancer Society. Cancer facts and figures and CLL detailed guide. Atlanta, Ga; 2008.
3. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. Blood. 1975;46:219-234.
4. Binet JL, Auquier A, Dighiero G, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer. 1981;48:198-206.
5. Damle RN, Wasil T, Fais F, et al. Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia. Blood. 1999;94:1840-1847.
6. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94:1848-1854.
7. Krober A, Seiler T, Benner A, et al. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. Blood. 2002;100:1410-1416.
8. Rassenti LZ, Huynh L, Toy TL, et al. ZAP-70 Compared with Immunoglobulin Heavy-Chain Gene Mutation Status as a Predictor of Disease Progression in Chronic Lymphocytic Leukemia. N Engl J Med. 2004;351:893-901.
9. Gahrton G, Robert KH, Friberg K, Zech L, Bird AG. Nonrandom chromosomal aberrations in chronic lymphocytic leukemia revealed by polyclonal B-cell-mitogen stimulation. Blood. 1980;56:640-647.
10. Dicker F, Schnittger S, Haferlach T, Kern W, Schoch C. Immunostimulatory oligonucleotide-induced metaphase cytogenetics detect chromosomal aberrations in 80% of CLL patients: A study of 132 CLL cases with correlation to FISH, IgVH status, and CD38 expression. Blood. 2006;108:3152-3160.
11. Haferlach C, Dicker F, Schnittger S, Kern W, Haferlach T. Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. Leukemia. 2007;21:2442-2451.

12. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med. 2000;343:1910-1916.
13. Xu W, Li JY, Pan JL, et al. Interphase fluorescence in situ hybridization detection of cytogenetic abnormalities in B-cell chronic lymphocytic leukemia. Int J Hematol. 2007;85:430-436.
14. 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;85:3610-3618.
15. Han T, Ozer H, Sadamori N, et al. Prognostic importance of cytogenetic abnormalities in patients with chronic lymphocytic leukemia. N Engl J Med. 1984;310:288-292.
16. Juliusson G, Robert KH, Ost A, et al. Prognostic information from cytogenetic analysis in chronic B-lymphocytic leukemia and leukemic immunocytoma. Blood. 1985;65:134-141.
17. Lucito R, West J, Reiner A, et al. Detecting Gene Copy Number Fluctuations in Tumor Cells by Microarray Analysis of Genomic Representations. Genome Res. 2000;10:1726-1736.
18. Lucito R, Healy J, Alexander J, et al. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res. 2003;13:2291-2305.
19. F Fais FG, S Hashimoto, B Sellars, A Valetto, S L Allen, P Schulman, V P Vinciguerra, K Rai, L Z Rassenti, T J Kipps, G Dighiero, H W Schroeder, Jr, M Ferrarini, and N Chiorazzi. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest. 1998;102:1515-1525.
20. Hicks J, Krasnitz A, Lakshmi B, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome Res. 2006;16:1465-1479.
21. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921.
22. Yang YH, Speed T. Design issues for cDNA microarray experiments. Nat Rev Genet. 2002;3:579-588.
23. Tseng GC, Oh M-K, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucl Acids Res. 2001;29:2549-2557.
24. Sebat J, Lakshmi B, Troge J, et al. Large-Scale Copy Number Polymorphism in the Human Genome. Science. 2004;305:525-528.

25. Szczepanski T, Beishuizen A, Pongers-Willemse MJ, et al. Cross-lineage T cell receptor gene rearrangements occur in more than ninety percent of childhood precursor-B acute lymphoblastic leukemias: alternative PCR targets for detection of minimal residual disease. Leukemia. 1999;13:196-205.
26. LeBien TW, Elstrom RL, Moseley M, Kersey JH, Griesinger F. Analysis of immunoglobulin and T-cell receptor gene rearrangements in human fetal bone marrow B lineage cells. Blood. 1990;76:1196-1200.
27. Juliusson G, Oscier D, Juliusson G, et al. Cytogenetic Findings and Survival in B-cell Chronic Lymphocytic Leukemia. Second IWCCLL Compilation of Data on 662 Patients. Leukemia and Lymphoma. 1991;5:21 - 25.
28. Stilgenbauer S, Bullinger L, Benner A, et al. Incidence and clinical significance of 6q deletions in B cell chronic lymphocytic leukemia. Leukemia. 1999;13:1331-1334.
29. Cuneo A, Rigolin GM, Bigoni R, et al. Chronic lymphocytic leukemia with 6q- shows distinct hematological features and intermediate prognosis. Leukemia. 2004;18:476-483.
30. Athanasiadou A, Stamatopoulos K, Tsezou A, et al. Report of novel chromosomal abnormalities in a series of 130 chronic lymphocytic leukemia patients studied by classic cytogenetic analysis. Leukemia and Lymphoma. 2006;47:2084 - 2087.
31. Nelson BP, Gupta R, Dewald GW, Paternoster SF, Rosen ST, Peterson LC. Chronic Lymphocytic Leukemia FISH Panel: Impact on Diagnosis. American Journal of Clinical Pathology 2007;128:323 - 332.
32. Bea S, Lopez-Guillermo A, Ribas M, et al. Genetic imbalances in progressed B-cell chronic lymphocytic leukemia and transformed large-cell lymphoma (Richter's syndrome). Am J Pathol. 2002;161:957-968.
33. 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;101:1039-1044.
34. 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;109:1202-1210.
35. Tyybakinoja A, Vilpo J, Knuutila S. High-resolution oligonucleotide array-CGH pinpoints genes involved in cryptic losses in chronic lymphocytic leukemia. Cytogenet Genome Res. 2007;118:8-12.

36. Fernandez V, Jares P, Salaverria I, et al. Gene expression profile and genomic changes in disease progression of early-stage chronic lymphocytic leukemia. Haematologica. 2008;93:132-136.
37. Bernheim A, Dessen P, Lazar V, et al. Cryptic del(13q14.2) and physiological deletions of immunoglobulin genes detected by high-resolution array comparative genomic hybridization in a patient with indolent chronic lymphocytic leukemia. Cancer Genet Cytogenet. 2007;176:89-91.
38. Patel A, Kang SH, Lennon PA, et al. Validation of a targeted DNA microarray for clinical evaluation of recurrent abnormalities in chronic lymphocytic leukemia. Am J Hematol. 2007.
39. 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 Res. 2008;68:1012-1021.
40. Odero MD, Soto JL, Matutes E, et al. Comparative genomic hybridization and amplotyping by arbitrarily primed PCR in stage A B-CLL. Cancer Genet Cytogenet. 2001;130:8-13.
41. Tsukasaki K, Lohr D, Sugahara K, et al. Comparative genomic hybridization analysis of Japanese B-cell chronic lymphocytic leukemia: correlation with clinical course. Leuk Lymphoma. 2006;47:261-266.
42. Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36:949-951.
43. Damle RN, Temburni S, Calissano C, et al. CD38 expression labels an activated subset within chronic lymphocytic leukemia clones enriched in proliferating B cells. Blood. 2007;110:3352-3359.
44. Patten PEM, Buggins AGS, Richards J, et al. CD38 expression in chronic lymphocytic leukemia is regulated by the tumor microenvironment. Blood. 2008;111:5173-5181.

Table 1: Novel and narrowed down regions and their frequencies

*Note: Table 1 is an abridged version of Table S1.

Regions not in the literature are highlighted in yellow. The total number of patients analyzed is 58. The positions and size are in base pairs. Frequency.up/down – frequency of amplification/deletion in a given region respectively; No.patients.up/down – the number of patients out of the total (58) where amplification/deletion were observed; Row.in.TableS1 – corresponding row number in Table S1; No.of.Probes – number of probes within a given lesion on the 390K array.

Chromosome	Cytoband	Start.Position	End.Position	Size	Frequency.Up	No.patients.up	Frequency.Down	No.patients.down	Row.in.TableS1	No.of.Probes
2.00	p16.3+	48084669	48252448	167780	0.03	2	0.00	0	26	18
2.00	p16.1-p15+	60246950	61709066	1462117	0.03	2	0.00	0	28	187
2.00	q37.1	230772768	231360013	587246	0.00	0	-0.03	2	63	79
8.00	p23.3-p23.2	16012	4660930	4644919	0.00	0	-0.03	2	140	523
8.00	p22	12910916	14630827	1719912	0.00	0	-0.05	3	149	237
8.00	p22	15020444	15914166	893723	0.00	0	-0.05	3	151	140
8.00	p22	16643703	17640961	997259	0.00	0	-0.05	3	153	130
8.00	p21.2-p12	27044904	30667738	3622835	0.00	0	-0.03	2	156	459
8.00	q23.3+	113960835	114091941	131107	0.03	2	0.00	0	174	15
8.00	q24.22+	135955922	136072387	116466	0.03	2	0.00	0	178	16
8.00	q24.3+	142693357	142713614	20258	0.03	2	0.00	0	180	7
9.00	p21.3	21883391	22132407	249017	0.00	0	-0.03	2	184	33
12.00	p13.31-q11+	8602964	36143885	27540922	0.16	9	0.00	0	239	3525
12.00	q11-q13.13+	36267342	48565240	12297899	0.16	9	0.00	0	241	1480
12.00	q13.13-q13.2+	48589017	53195850	4606834	0.17	10	0.00	0	242	581
12.00	q14.2+	61905334	61957668	52335	0.17	10	0.00	0	244	8
13.00	q14.3	49498772	49524994	26223	0.00	0	-0.40	23	279	4
13.00	q33.1	102118193	103445915	1327723	0.00	0	-0.03	2	300	182
14.00	q24.1	68334407	69105518	771112	0.00	0	-0.03	2	310	93
14.00	q24.1-q31.3	69112027	87740939	18628913	0.00	0	-0.05	3	311	2361
14.00	q31.3-q32.13	87971637	94858088	6886452	0.00	0	-0.05	3	313	843
18.00	p11.32	300920	2612639	2311720	0.00	0	-0.03	2	372	290
18.00	p11.32-p11.31	2615771	3673994	1058224	0.00	0	-0.05	3	373	141
18.00	p11.31-p11.21	3678333	13937187	10258855	0.00	0	-0.03	2	374	1315
18.00	q12.2	31085600	32130507	1044908	0.00	0	-0.03	2	380	137
18.00	q21.2+	49011586	49188996	177411	0.03	2	0.00	0	385	27
18.00	q22.1+	63409427	63997439	588013	0.03	2	0.00	0	387	77
18.00	q23	75240717	75397269	156553	0.00	0	-0.03	2	390	11

Table 2: Genomic Alterations in $CD38^+$ vs. $CD38^-$

Three CLL samples exhibit clonality with respect to CD38. All three samples exhibit differences between $CD38^+$ and $CD38^-$ fractions.

Sample	355			625			931		
CD38 Status (%)		25%			54%			63%	
	38+ vs. -	38+ vs. PMN	38- vs. PMN	38+ vs.-	38 + vs. PMN	38- vs. PMN	38+ vs. -	38 + vs. PMN	38- vs. PMN
Locus									
1p22.1							loss	loss	neutral
4q13.3-q22.1				loss	loss	neutral	loss	loss	loss
4q32.1							loss	loss	neutral
5q13.1-q13.2							loss	loss	loss
5q13.3-q14.1							loss	loss	loss
5q14.3-q15							loss	loss	loss
5q21.1-q21.3							loss	loss	loss
5q21.3							loss	loss	loss
8p23.3-p12	loss	loss	neutral						
8p11.22-p11.21	loss	loss	neutral						
11q22.3-q23.3	gain	gain	loss				loss	loss	neutral
11q23.3-q24.1							loss	loss	neutral
17p13.3-q21.31	loss	loss	neutral						
17q24.2-q25.3	loss	loss	neutral						
18p11.32-p11.21				gain	loss	loss	loss	loss	loss
18q22.3-q23				gain	neutral	loss			
21q22.11-q22.3				loss	gain	gain			

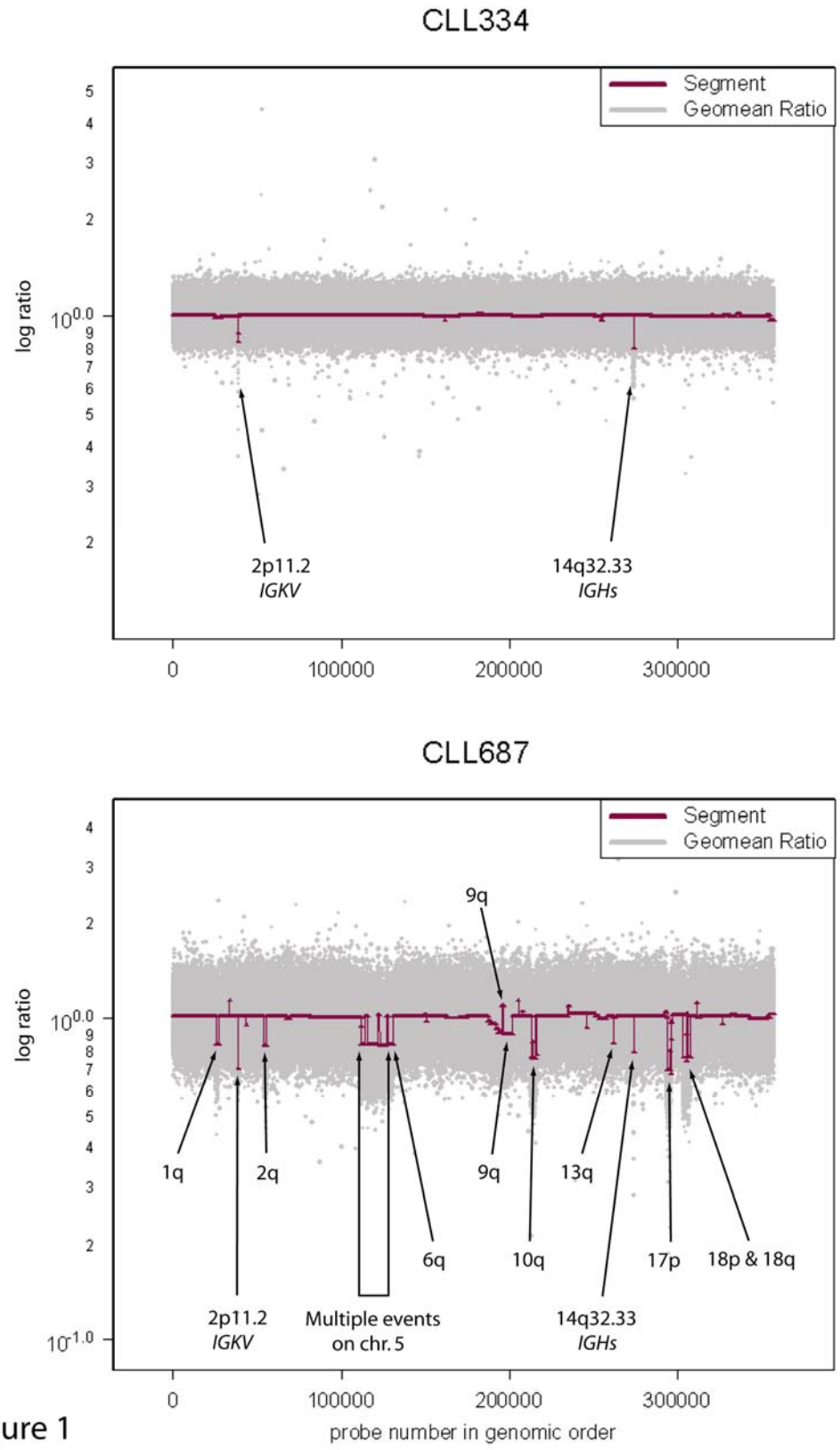


Figure 1

Figure 1: Variation in genome stability of CLL

Top: Sample CLL334 has a stable genome with immunoglobulin recombinations at 2p11.2 and 14q32.33 as the only "lesions". Raw geometric mean data ("Geomean Ratio") are shown in light gray and segmenter output in maroon.

Bottom: Sample CLL687 has a very unstable genome exhibiting multiple lesions in addition to the immunoglobulin recombinations described in sample CLL334 (top).

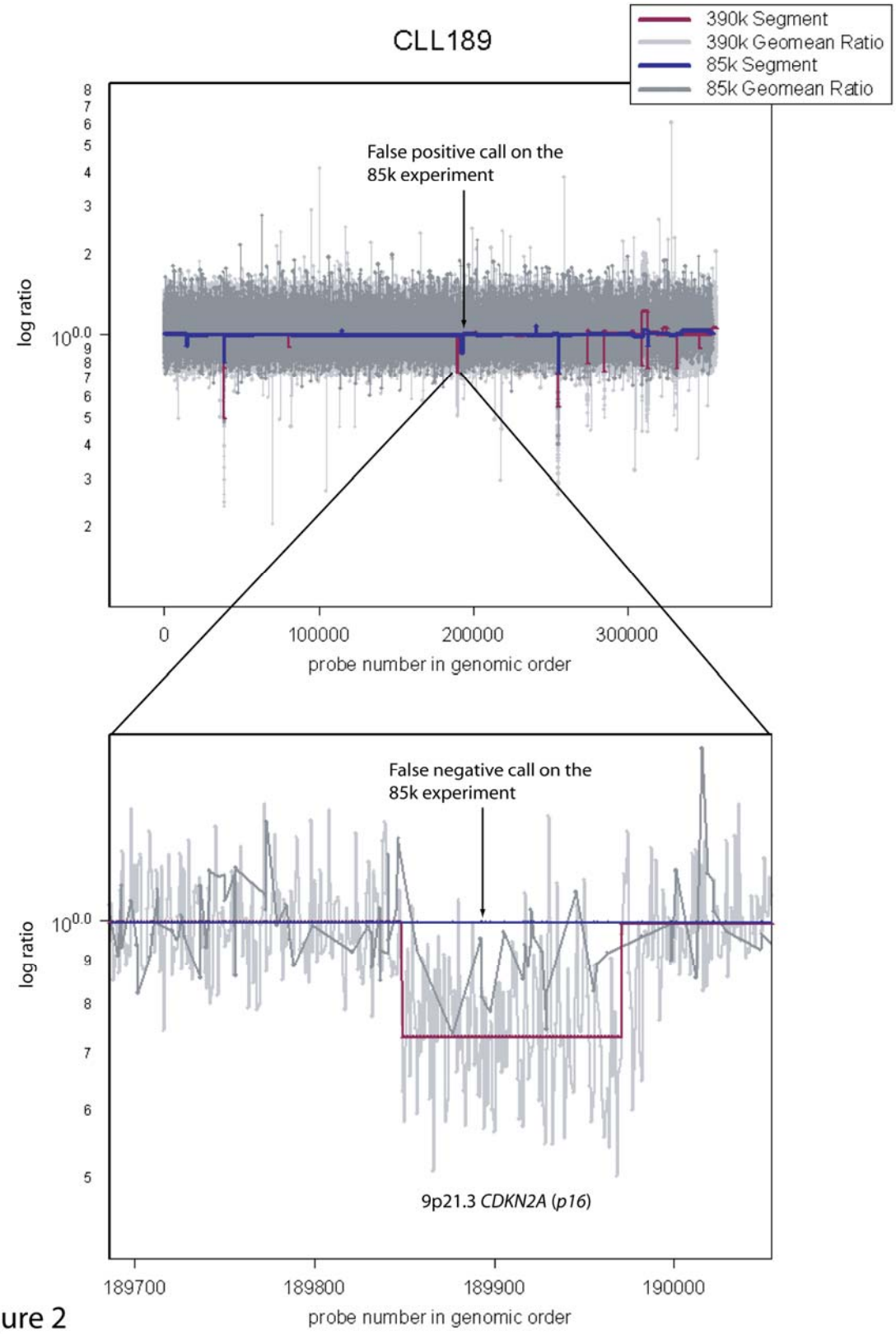


Figure 2

Figure 2: Comparison between the 85K array and the 390K array

Top: Sample CLL189 was hybridized both on 85K and 390K arrays. Arrow points to a false positive call on the 85K array data by the segmenter.

Bottom: Demonstration of a false negative call by the segmenter on the 85K data (dark blue). It spans *CDKN2A* (*p16*$^{-INK4}$). Since the 85K array did not have sufficient coverage in this region, the segmenter could not make a reliable lesion call, even though it appears obvious upon inspection of the 85K raw data (dark gray).

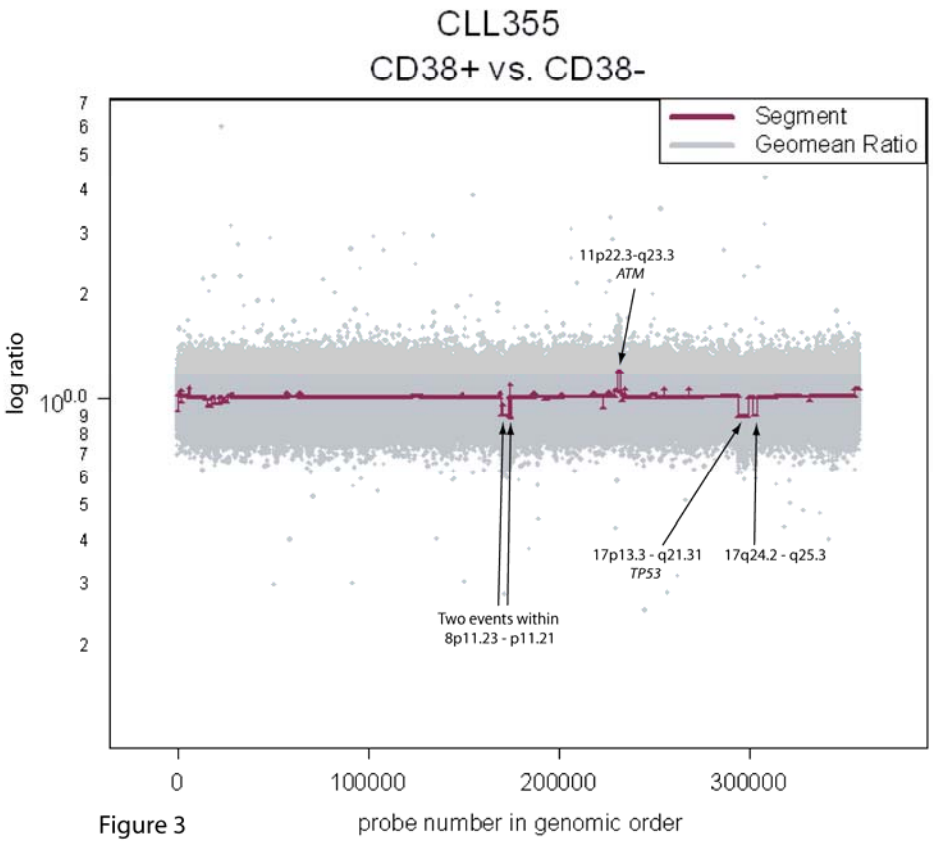
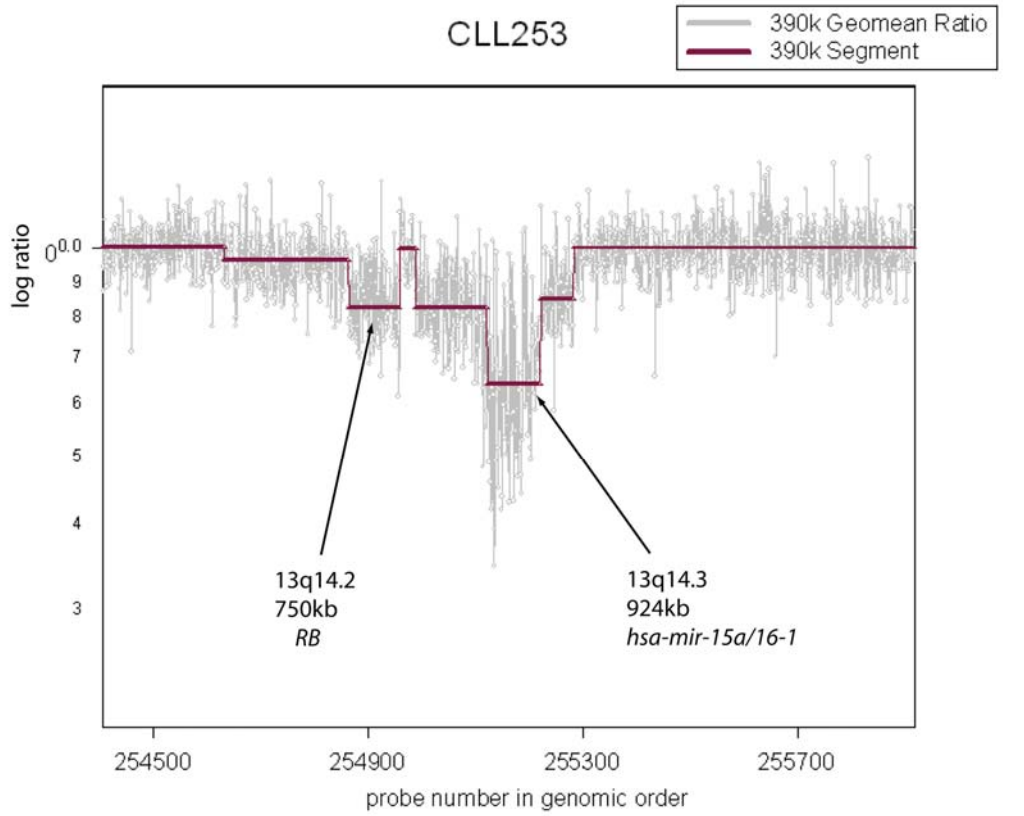


Figure 3

Figure 3: Clonality between the CD38^{+} and CD38^{-} fractions in CLL355

Copy number aberrations are clearly visible and span known loci of clinical significance.



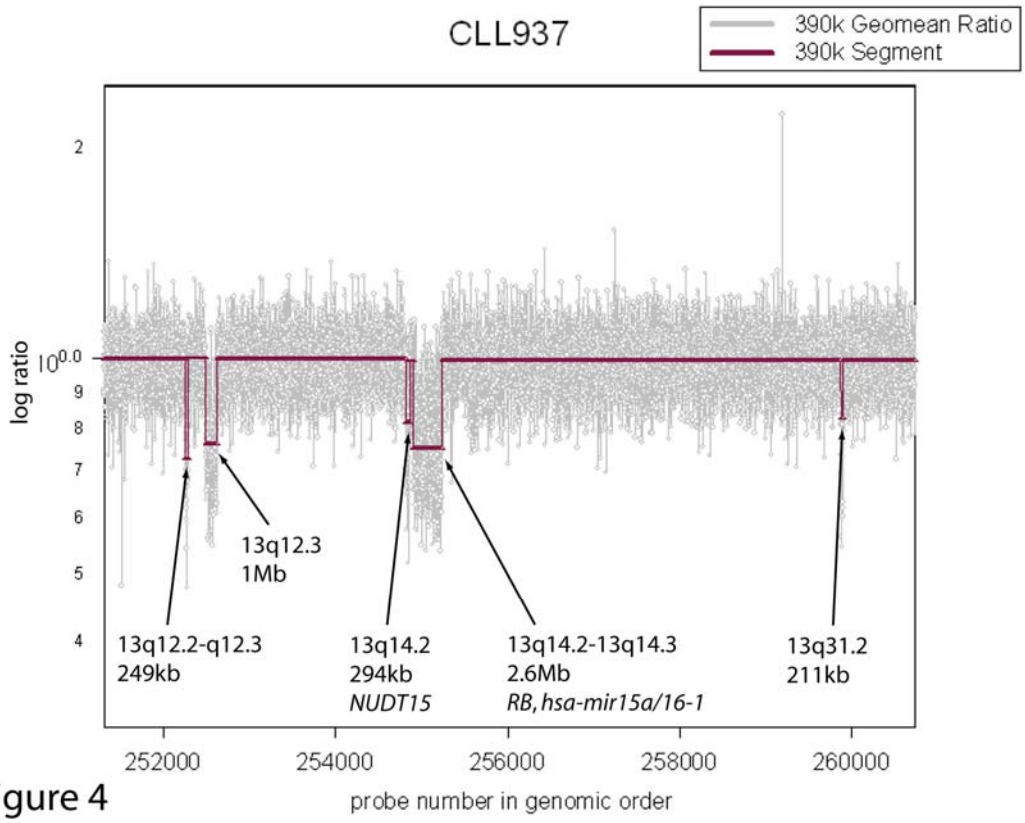


Figure 4

Figure 4: Multiple lesions observed in the 13q region

Top: CLL cells from patient CLL253 exhibit two distinct lesions on 13q. *RB* is hemizygously deleted and the *hsa-mir-15a/16-1* cluster is homozygously lost.

Bottom: Sample CLL937 exhibits multiple, distinct lesions on 13q.

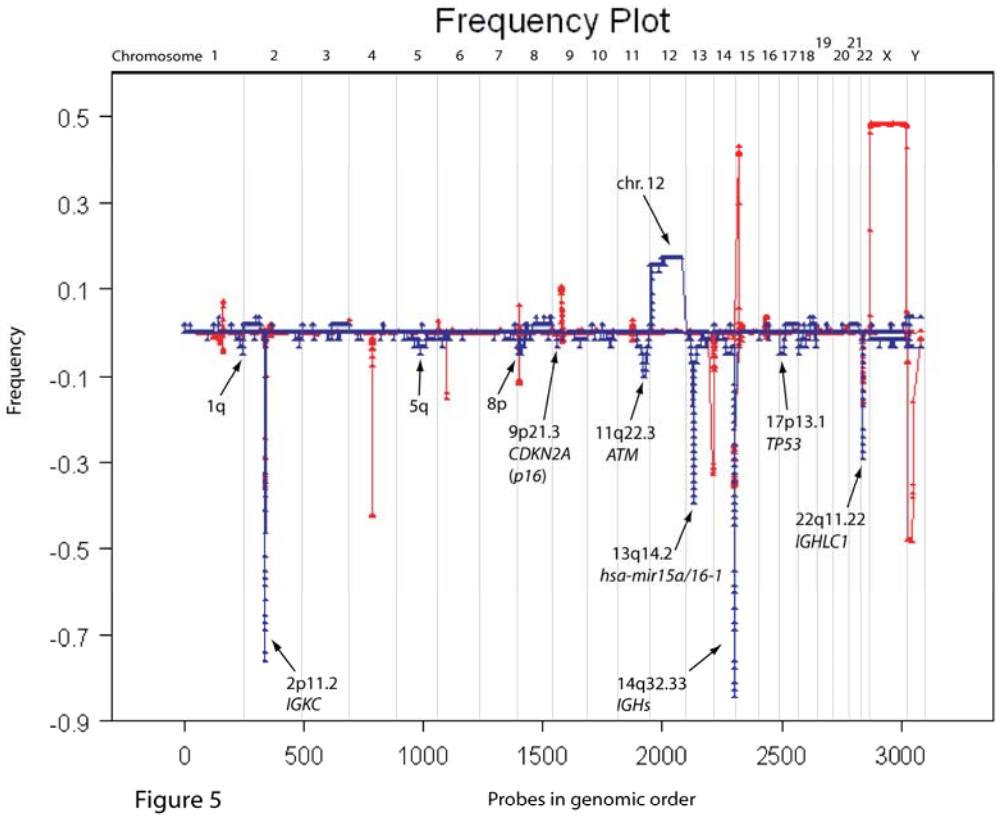


Figure 5

Figure 5: Frequency plot all CLL samples vs. normal copy number variant (CNV) database

The frequency plot for 58 CLL samples is shown in blue and that of 500 "normal" (i.e. disease free) individuals is in red. The red frequency plot shows CNVs in the human genome and was used to eliminate any CNVs in our data exposed through copy-neutral LOH (i.e. regions where CNVs overlap with lesions observed in CLL). Most lesions observed in CLL are deletions and trisomy 12. Regions of interest are marked with arrows.