Published in final edited form as:

Cancer Genet. 2014 ; 207(0): 19–30. doi:10.1016/j.cancergen.2014.01.004.

Acquired Chromosomal Anomalies in Chronic Lymphocytic Leukemia (CLL) Patients Compared to >50,000 Quasi-normal Subjects

Cathy C. Laurie^{*a}, Cecelia A. Laurie^a, Stephanie A. Smoley^b, Erin E. Carlson^c, Ian Flinn^d, Brooke L. Fridley^{c,e}, Harvey A. Greisman^{f,g}, John G. Gribben^h, Diane F. Jelinek^{i,j}, Sarah C. Nelson^a, Elisabeth Paietta^k, Dan Schaid^c, Zhuoxin Sun^l, Martin S. Tallman^m, Richard Weinshilboumⁿ, Neil E. Kayⁱ, and Tait D. Shanafeltⁱ

^aBiostatistics Department, University of Washington, Seattle, WA, USA

^bDepartment of Laboratory Medicine and Pathology, Cytogenetics Laboratory, Mayo Clinic, Rochester, MN, USA

^cDepartment of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

^dSarah Cannon Research Institute, Nashville, TN, USA

^eUniversity of Kansas Medical Center, Kansas City, KS, USA

^fDepartment of Laboratory Medicine, University of Washington, Seattle, WA, USA

^gPhenoPath Laboratories, Seattle, WA, USA

^hBarts Cancer Institute, Queen Mary, University of London, London, UK

ⁱDivision of Hematology, Mayo Clinic, Rochester, MN, USA

^jDepartment of Immunology, Mayo Clinic, Rochester, MN, USA

^kThe North Division of Montefiore Medical Center, Bronx, NY, USA

^lDepartment of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA

^mMemorial Sloan-Kettering Cancer Center, New York, NY, USA

© 2014 Elsevier Inc. All rights reserved.

^{*}Corresponding Author: Cathy C. Laurie, Department of Biostatistics, 4373 Brooklyn Ave, Seattle, WA 98195, cclaurie@u.washington.edu, 781-640-2941 (voice).

Author Contributions: All authors reviewed, edited and approved the manuscript. C.C.L. analyzed and interpreted data and was the primary writer. T.D.S. and N.E.K. participated in study design, data collection, analysis, and interpretation. C.A.L. and S.A.S. analyzed and interpreted data. E.E.C., B.L.F. and D. S. prepared genotypic and phenotypic data for analysis. H.A.G., B.S.W. and R.W. contributed to data interpretation. J.G.G. analyzed and interpreted Ig rearrangement status. D.F.J. and I.F. collected and analyzed data. S.C.N. provided bioinformatics support for data interpretation. E.P. collected and characterized specimens. Z.S. helped with the initial proposal, and prepared clinical data. M.S.T. provided administrative support for conduct of the clinical trial and collection of data.

Conflict Of Interest Disclosures: J.G.G. declares honoraria from Roche/Genentech and Celgene. There are no other conflicts to disclose.

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

ⁿDivision of Clinical Pharmacology, Mayo Clinic, Rochester, MN, USA

Abstract

Pre-therapy CLL patients from US Intergroup trial E2997 were analyzed with single nucleotide polymorphism microarrays to detect acquired chromosomal anomalies. The four CLL-typical anomalies (11q-, +12, 13q- and 17p-) were found at expected frequencies. Acquired anomalies in other regions account for 70% of the total detected and their number per subject has a significant effect on progression-free survival after adjusting for the effects of 17p- (and other covariates). These results were compared with a previous study of >50,000 subjects from the GENEVA consortium of genome-wide association studies, which analyzed individuals with a variety of medical conditions and healthy controls. The percentage of individuals with acquired anomalies is vastly different between the two studies (GENEVA 0.8%; E2997 80%). The composition also differs, with GENEVA having a higher percentage of acquired uniparental disomies and a lower percentage of deletions. The four common CLL anomalies are among the most frequent in GENEVA subjects, some of whom may have CLL-precursor conditions or early stages of CLL. However, the patients from E2997 (and other studies of symptomatic CLL) have recurrent acquired anomalies that were not found in GENEVA subjects, thus identifying genomic changes that may be unique to symptomatic stages of CLL.

Keywords

Chromosomal aberration; Chromosomal mosaic; Chronic Lymphocytic Leukemia; Cancer precursor condition; Cytogenetics

Introduction

Chronic lymphocytic leukemia (CLL) is one of the most common hematological cancers in adults, with an incidence that increases with age[1]. CLL is defined by a total blood lymphocyte count greater than $5 \times 10^9/l$ and a specific immunophenotype[2]. The clinical course of disease is highly variable. There is strong evidence that CLL is nearly always preceded by monoclonal B cell lymphocytosis (MBL), defined by the presence of a low concentration of clonal B cells with an immunophenotype characteristic of CLL[3]. People with MBL are asymptomatic, but have an estimated 1-2% annual risk of progression to CLL[4]. The mechanisms of disease progression from MBL to CLL, and from indolent to aggressive CLL, are still largely unknown.

Many studies of CLL patients, using chromosome banding or interphase fluorescence *in situ* hybridization (FISH), have shown four or five common anomalies involving copy number variation (see reviews[5, 6]). For example, Döhner et al.[7] found 55% of patients with del(13q14), 18% with del(11q22), 16% with 12q trisomy, 7% with del(17p13) and 6% with del(6q21). The 17p and 11q deletions were associated with a poor prognosis, while the 13q deletions were associated with a better prognosis. Microarray technologies also have been used to detect acquired chromosomal anomalies in CLL patients (e.g.[8-13]). In addition to the five common abnormalities, these studies detected a variety of deletions, duplications and acquired uniparental disomies (aUPD). Genomic complexity, defined by the number or

length of anomalies per patient, has been associated with shorter progression-free survival[10] and time to treatment[14]. Genomic sequencing studies of CLL patients have shown that the recurrence of mutations in specific genes is low[15] and the total number of point mutations in coding sequences is lower than in many other cancers[16, 17]. While chromosome banding studies have found that balanced translocations in CLL are rare, sequencing studies have shown that some patients have multiple rearrangements associated with copy number variation in a phenomenon known as chromothripsis[18], which has been associated with a poor prognosis[19].

Here we describe genomic profiling, using single nucleotide polymorphism (SNP) microarrays, of 214 CLL patients enrolled in E2997, a randomized, phase III intergroup clinical trial led by the Eastern Cooperative Oncology Group (ECOG) [20]. This trial evaluated treatment with a combination of fludarabine and cyclophosphamide (FC) compared with fludarabine alone (FL). FC provides the backbone of a widely used chemotherapy for CLL, which now commonly includes rituximab (R) to constitute a regimen designated as FCR. Previously, the prognostic significance of a large panel of laboratory factors was evaluated in this trial, including *IgV_H* mutational status, *TP53* mutations and FISH assays of the five copy number alterations (CNAs) described above[21]. The occurrences of del(17p13) and del(11q22) were each associated with reduced progression free survival, while mutational status of *IgV_H* and *TP53* had no significant effect independent of del(17p13).

The blood cell chromosomal anomalies that we discuss in this paper were acquired during the lifetime of the individual. Ideally, acquired and inherited anomalies are distinguished using paired tumor-normal samples. However, in many studies (such as E2997), normal control tissues were not genotyped. An alternative method of identifying acquired anomalies is detection of a mosaic mixture of normal and abnormal cells in a single tissue, assuming that the abnormality was acquired during development. This approach works well using B Allele Frequency (BAF) and Log R Ratio (LRR) metrics from Illumina microarrays[22-24], provided that a single anomaly constitutes a significant fraction (> 5-10%) of the cells in the sample from which DNA is extracted. We define 'clonal mosaic anomalies' as a mixture of normal and abnormal cells derived from a single zygote, in which the abnormal cells contain a chromosomal anomaly presumed to be of clonal origin because it constitutes a large fraction of the total cell population. We also define a larger group of 'acquired' anomalies as clonal mosaics plus anomalies that are non-mosaic in the blood cells analyzed, but greater than 3 megabases (Mb) in length. The latter are considered to be acquired because inheritance of anomalies this large is very rare (see Supplementary Text). Figure 1 shows diagrams of inherited, mosaic and acquired anomalies to illustrate this terminology.

Previously, we developed a method for efficiently detecting chromosomal anomalies, including clonal mosaics, using Illumina SNP microarrays, and applied this method to over 50,000 subjects from genome-wide association studies in the GENEVA consortium[25]. The GENEVA studies involved subjects of all ages and addressed a variety of different medical conditions, such as glaucoma, prematurity and addiction. Using peripheral blood samples, we found that the frequency of clonal mosaicism is low (<0.5%) from birth until 50 years of age, after which it rapidly rises to 2-3% in the elderly. Many of the mosaic anomalies are

characteristic of those found in hematological cancers, although subjects with hematological cancer were not specifically recruited in GENEVA studies and, in fact, would generally have been excluded when their condition was known. Over all GENEVA studies, we estimated that only 3% of subjects with a mosaic anomaly had any prior record of hematological cancer. Thus, we consider the GENEVA subjects as ‘quasi-normal’ (with respect to hematological cancer). However, in four longitudinal cohort studies (a GENEVA subset), we found that subjects with acquired anomalies, but without a previous hematological cancer diagnosis, have an estimated tenfold higher risk of a subsequent hematological cancer than subjects without acquired anomalies (95% confidence interval = 6-18)[25].

Here we describe acquired chromosomal anomalies in the E2997 CLL patients, detected with the GENEVA methods, and examine their properties relative to other CLL studies and to the sample of quasi-normal GENEVA subjects.

Methods

In this section, we describe methods used to obtain and analyze data from the E2997 study. The E2997 genotypic and progression-free survival data are posted on dbGaP (accession phs000621.v1.p1). The GENEVA methods and results described in this paper were published previously[25]. The GENEVA data used here were published in Supplementary Table 3 of Laurie et al.[25]. Among the 50,222 GENEVA subjects analyzed, 43% were female and self-identified race was 66% white, 15% black and 19% other.

CLL patients

The E2997 study enrolled a total of 278 symptomatic CLL patients who were previously untreated but required therapy at the time of enrollment. Institutional review board approval was obtained at all institutions participating in this study and all patients provided written informed consent. Details of patient enrollment and characteristics were described previously[20].

Samples and laboratory analysis

For DNA analyses, peripheral blood was obtained before treatment, mononuclear cells were isolated using Ficoll density-gradient centrifugation and DNA was extracted using Gentra Puregene Cell Kit. DNA samples were genotyped on the Illumina HumanOmni-Quad-v1-0_B SNP microarray and quality control was performed as described previously[25, 26]. *IgV_H* mutational analysis was described previously[21]. Among the 278 enrolled patients, 215 were genotyped and 214 passed quality control for anomaly detection. Among the 214 analyzed patients, 31% were female and self-identified race was 78% white, 14% black and 8% other.

E2997 FISH results are based on whole blood samples drawn before treatment, analysis of 200 cells and probe sets described previously[27] to assay 11q23 (*ATM*), 12q15 (*MDM2*), 13q14 (*DIS319*), 17p13 (*TP53*) and 6q23.3 (*MYB*). Henceforth, we refer to anomalies detected with these probes as 11q-, +12, 13q- 17p- and 6q-, since most span multiple cytogenetic bands. Samples with a percent abnormal cells above the following thresholds

were considered to be positive for the anomaly: 6.5% (6q-), 5% (11q-), 1.5% (+12), 7% (13q- one copy), 1.5% (13q- two copies) and 8.5% (17p-).

Clinical phenotypes and statistical methods

Progression-free survival (PFS) was defined from the date of randomization to the date of progression or death (event) or to last follow-up (censored). PFS was analyzed, using Cox proportional hazards models, for association to chromosome anomalies along with several clinical covariates, including treatment arm, sex, age, hemoglobin concentration, white blood cell count, platelet count, lymphocyte count, and RAI stage at enrollment.

SNP microarray detection of chromosomal anomalies

Autosomal anomalies were detected as deviations from the normal biparental disomic state, using a method based on BAF and LRR, as described previously[25]. This method detects both mosaic and non-mosaic CNAs, as well as mosaic UPD, and was optimized to detect large anomalies (50 kilobases to whole chromosome). Smaller anomalies were not considered here because it is very difficult to distinguish acquired from non-acquired in the absence of paired normal control tissue. Non-mosaic UPDs were not identified because they are difficult to distinguish from regions of identity-by-descent inherited biparentally. Also, we do not report X chromosome anomalies here because of technical difficulties in the analysis of LRR for the X (see[25]). We distinguished mosaic from non-mosaic duplications using a bivariate normal, 95% prediction interval for BAF and LRR for polymorphic CNAs. All mosaic anomalies and all other anomalies greater than 2 Mb in length were confirmed by manual review of BAF/LRR plots. Immediately adjacent anomalies belong to a ‘compound’ set. Whole chromosome anomalies with gain of copy number (other than chromosome 21) are all classified as mosaic (see Supplementary Information). Reproducibility of acquired anomaly detection was 100% in a small sample of duplicates (see Supplementary Information). We note that detection and classification of anomalies was done independently in E2997 and GENEVA, except for the finding in GENEVA families that inheritance of anomalies > 3 Mb is very rare, which was used to classify non-mosaic E2997 anomalies > 3 Mb as acquired. See Supplementary Information for further discussion of these GENEVA family studies.

Results

Types of SNP-detected anomalies in E2997

Among 214 CLL patients, we found 1,112 anomalies, of which 628 were classified as acquired and 484 as non-acquired. Among the 628 acquired anomalies, 478 are clonal mosaics, 109 are non-mosaic (in blood) with length > 3 Mb (median 19.5 Mb) and 41 belong to compound sets. Figure 2 shows examples of the three types of clonal mosaic anomalies detected.

A global view of the lengths and types of acquired anomalies is shown as a “Circos” plot[28] in Figure 3, with more detail in Supplementary Figure 1 and Supplementary Table 1. Distributions of anomaly length are in Supplementary Figure 2 and the number of

anomalies per subject in Figure 4. Eight CLL patients (3.7%) showed evidence of chromothripsis (see Supplementary Information).

E2997 anomalies detected by FISH

Among 211 CLL patients, we detected CNAs by the standard FISH panel as follows: 16.6% with 11q-, 26.1% with +12, 52.6% with 13q-, 8.5% of 17p- and 3.8% with 6q-. These frequencies are similar to those from other CLL studies (Supplementary Table 2). The concordance between SNP- and FISH- detected anomalies is 96.7% overall and 98.9% when only the FISH anomalies with > 20% abnormal cells are considered as positive (see Supplementary Information for more detail).

E2997 anomalies not detected by FISH

Among the 628 acquired anomalies detected by SNP data, 53 (8.4%) were aUPD and thus could not have been detected by FISH. Among the 575 acquired CNAs, 403 (70.1%) do not overlap any of the FISH probes. Therefore, the SNP microarray detected many genomic alterations in CLL patients that were not detected by the standard FISH probes, as noted previously by others[9-14]. There was a highly significant positive association between 17p- status and the number of acquired anomalies that do not overlap FISH probes (i.e. 'non-target') anomalies ($p=7 \times 10^{-7}$). See Supplementary Information for other associations between FISH-targeted and non-target anomalies.

Acquired anomaly effects on progression-free survival in E2997

Analysis of PFS by multiple Cox regression in E2997 patients showed significant effects of treatment arm, 13q-, 17p- and non-target acquired anomalies (see Supplementary text for details). As in previous studies[7], 13q- has a positive effect on PFS, while 17p- has a negative effect. Non-target anomaly number has a significant negative effect on PFS that appears to be independent of 17p- status, as in Kay et al.[10].

Recurrent acquired anomalies in E2997

For each chromosome arm, we identified genomic intervals with maximal coverage by acquired anomalies of a given copy number type. Table 1 summarizes the properties of intervals covered by ≥ 4 anomalies in different E2997 patients (see also Supplementary Table 3). From this we note: (1) Intervals are often less than 2 Mb for deletions, but generally much larger for duplications and aUPD. (2) The fraction of the autosomal genome covered by these intervals is 2.7% for aUPD, 8.5% for gains and 2.8% for losses. (3) In most cases, anomalies spanning these intervals have been reported in other CLL studies (see references in Supplementary Table 3). (4) E2997 patients with a recurrent anomaly spanning a given interval generally also have other acquired anomalies. (5) The *IgV_H* mutational status varies within intervals.

Comparison of anomaly frequencies and types in E2997 and GENEVA

In comparisons of E2997 and GENEVA anomalies, we included all subjects in both studies despite some demographic differences, for the following reasons. (1) Previously, in GENEVA subjects, we found that neither genetic ancestry nor sex is significantly associated

with mosaic status (in total, or considering each copy number type separately)[25]. (2) Although E2997 includes only adults (all but two over age 40) and GENEVA includes some younger subjects, 88% of mosaic anomalies in GENEVA derive from subjects over age 40. (3) We were interested in comparing acquired anomalies in CLL patients with those in quasi-normal subjects (of any age) to determine what characteristics may be unique to the CLL condition.

The frequency of acquired anomalies was far higher in the E2997 patients than in GENEVA. In GENEVA, we detected 514 mosaic anomalies in 404 (0.8%) of the 50,222 subjects analyzed. In E2997, we detected 628 acquired anomalies in 171 (79.9%) patients and 478 mosaic anomalies in 146 (68.2%) of the 214 patients analyzed. This dramatic frequency difference clearly indicates that the vast majority of anomalies designated as acquired in E2997 were, in fact, acquired rather than inherited CNAs.

An overall comparison of the types and chromosomal distribution of acquired anomalies in GENEVA and E2997 is given in Figure 3, which shows all acquired anomalies in E2997 (N=628) and the mosaic anomalies in GENEVA (N=514). Supplementary Figure 3 shows only the mosaic anomalies in E2997 (N=478). More fine-scaled plots of anomaly coverage are given in Supplementary Figure 1 for E2997 and in Supplementary Figure 7 of Laurie et al.[25] for GENEVA.

GENEVA has a much higher proportion of anomalies with aUPD and a lower proportion of loss, while the proportion of gain is similar. The proportions of mosaic (acquired) anomalies in E2997 are 11.1% (8.4%) aUPD, 16.9% (20.7%) gain and 71.9% (70.9%) loss of copy number, while GENEVA has 34.4% aUPD, 15.6% gain and 50.4% loss. These differences in composition are highly significant ($p\text{-value} < 10^{-16}$). In both studies, aUPD and gains of copy number often cover a large fraction of a chromosome arm, the entire arm, or (for gains) the whole chromosome, whereas losses of copy number tend to be smaller (Supplementary Figure 4).

Comparison of recurrent anomalies in E2997 and GENEVA

Intervals with recurrent aUPD are rare in E2997 patients, except for 9q and 13q (which contain nested deletions that cover the 13q FISH probe). On the other hand, GENEVA subjects have recurrent aUPD on many chromosome arms, most notably 9p and 14q (but also including 9q and 13q).

The most frequent gain in copy number is +12 in both E2997 and GENEVA. Trisomy 12 dominates the landscape of acquired gains in E2997, while in GENEVA multi-megabase gains in other regions are nearly as common as +12. Chromosomes 3 and 7 have recurrent gains in E2997 (and other CLL studies), while no GENEVA subjects have gains in these regions (Table 1).

Comparisons of deletion coverage are given in Figure 5 and Supplementary Figure 5. In E2997, the intervals of highest coverage by deletions are those targeted by the FISH probes on 11q, 13q and 17p, containing the well-studied candidate genes *ATM*, *DLEU1* and *DLEU2*, and *TP53* [29]. The next highest coverage is a 1.3 Mb region on 6q, which does not

overlap the 6q-FISH probe. The other segments each have <10 deletions and most are large and/or contain many genes.

In GENEVA, the deletion intervals on 11q, 13q and 17p found in E2997 (and in other CLL studies) also have multiple deletions. On 13q, the interval of maximal coverage is slightly distal to that in E2997, but multiple deletions in both studies intersect key candidate genes[29] in this region. On 11q, the intervals with maximal deletion coverage overlap between studies and both include *ATM*. On 17p, the peaks of maximal coverage are broad in both studies and both include *TP53*.

Some regions have strikingly different deletion profiles in the two studies. In GENEVA, there are three intervals with high coverage that were not observed in the E2997 patients. These intervals are characteristic of myeloid malignancies (i.e. a large interval on 20q-[30], a small interval on 4q containing *TET2*[31], a small interval on 2p containing *DNMT3A*[32] and aUPD of 9p[33]). Correspondingly, several regions with recurrent deletions in E2997 (and other CLL studies) have no coverage in GENEVA (3p, 4p, 4q, 6p, 8p, 9p, 9q and 15q; Table 1).

The five CLL-typical anomalies targeted with FISH assays represent a larger fraction of the total anomalies in the CLL patients than in GENEVA subjects. They are 25.1% of mosaic (and 28.0% of all acquired) anomalies in E2997 patients, compared with 9.5% of mosaic anomalies in GENEVA subjects (Fisher's exact test p-value = 7×10^{-11}). The frequency of subjects with one or more of these five anomalies is vastly different between the two studies (176/214 for E2997 and 43/50222 for GENEVA). However, the relative frequencies of the five anomalies are not significantly different between studies (Chi-squared test of 2×5 table, p-value = 0.7). All GENEVA subjects with one of these CLL-typical anomalies were 45 years or older at the time of sampling.

Discussion

This study of acquired anomalies is unique in providing an extremely large baseline of quasi-normal GENEVA controls analyzed by the same methodology as E2997 CLL patients so that comparisons were very precise. As expected, the percentage of GENEVA subjects with acquired anomalies is much less than in CLL patients (0.8% versus 80%), but the total number of anomalies detected is similar (~500-600 each) and therefore provides an opportunity for comparison of anomaly characteristics. This comparison reveals both striking similarities and differences. Many of the GENEVA anomalies are typical of CLL: 17.3% span intervals with recurrent acquired anomalies in E2997 patients, but this is much less than the 72.9% in the E2997 patients themselves. E2997 patients, as well as CLL patients from other studies, have recurrent anomalies at multiple genomic locations that were not found in GENEVA subjects. Correspondingly, GENEVA subjects have recurrent deletions typical of myeloid malignancies, which were not seen in E2997.

As noted earlier, a prospective study of a subset of GENEVA subjects showed that subjects with mosaic anomalies but no prior history of hematological cancer have an estimated ten-fold higher risk of acquiring hematological cancer compared to those without such

anomalies[25]. Among the 15 incident cancers in subjects with mosaic anomalies, eight were lymphoid and six were myeloid malignancies, consistent with the occurrence of both lymphoid- and myeloid-characteristic anomalies. The mosaic anomalies in GENEVA occur primarily in elderly subjects and we previously suggested that they might represent precursor conditions such as MBL, undiagnosed indolent disease (as CLL often is) or under-reported types of malignancies such as myelodysplastic syndrome.

It is noteworthy that we did not detect acquired anomalies typical of myeloid malignancies in E2997 patients despite the observation of an increased incidence of myeloid malignancies after therapy[34]. These observations support the hypothesis that exposure to fludarabine and cyclophosphamide-based treatments may promote the development of myeloid malignancies, with the caveat that small clones of myeloid-typical anomalies (<5% of mononuclear cells) would have escaped detection with our methods.

The current study, which compares symptomatic CLL patients with GENEVA subjects, using the same methodology, supports the suggestion that some GENEVA subjects have unrecognized CLL or high B-cell count MBL. Subjects with MBL (both high and low count types) have CLL-typical anomalies at frequencies similar to CLL itself[35, 36]. Our method of anomaly detection, which requires at least 5-10% of abnormal cells, could identify anomalies in high count MBL and unrecognized cases of CLL, but most likely not the more prevalent low count MBL, which has clonal B-cells constituting less than 10% of total white blood cell count[4]. Unfortunately, white blood cell counts and other diagnostics necessary to determine the disease status of GENEVA subjects are not available.

Although some GENEVA subjects have the common CLL-typical anomalies and may have CLL-precursor conditions, the symptomatic CLL patients in E2997 (and other studies summarized in Table 1) have recurrent anomalies in several regions that were not found in any of the 50,222 GENEVA subjects analyzed. This finding indicates that some recurrent anomalies are characteristic of late stages in the development of CLL and do not occur in precursor conditions (or in the general population). At this point, however, the clinical significance of this finding is not clear, since the recurrently-covered regions are large and the level of recurrence is low (ranging from 4-9 of 214 patients affected).

The GENEVA studies represent a diverse collection of subjects in terms of age, race and health status, and this diversity may be considered a disadvantage relative to a more clearly defined set of matched normal controls in the comparison to CLL patients. However, the very large size of the GENEVA sample provided the opportunity to characterize a reasonably large set of acquired anomalies (514) that occur at very low frequency (0.8%). Recruiting matched normal controls specifically for this purpose would not be practical. The quasi-normal characteristic of the GENEVA sample may affect various aspects of the composition of anomalies detected, but does not compromise finding that CLL-typical anomalies are very rare in a diverse sample from the general population and that recurrent anomalies in several genomic regions are unique to CLL patients.

Our results encourage further work on acquired chromosomal anomalies in the general population of elderly subjects in order to clarify their relationship with hematological

cancers and related precursor conditions. It could be argued that such a study will begin to at least identify individuals at risk for clinical development of blood malignancies. Indeed, our analysis suggests that such work eventually may contribute to an understanding of early stages of malignant disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank all clinical investigators and patients who participated in the Eastern Cooperative Oncology Group clinical trial E2997. We also thank Gordon W. Dewald (deceased) for his work on the early stages of this study. We thank Jeffrey C. Murray, Terri H. Beaty, Mary L. Marazita and William L. Lowe for access to GENEVA data for the analysis of parent-offspring transmission of anomalies. This work was supported by the following grants: U01-HG 005157, P01-GM099568, HD52953, HD57192, U01-DE-018993, R01-DE016148, R01-DE014899, U01-DE018903, U10-CA02115, P01-CA81538, R01-CA136591, CA21115, CA114737, R01-CA138461, U19-GM61388, R01-GM28157, CA132780, CA95241, and U01-HG005137. T. D. Shanafelt acknowledges support as a Clinical Scholar of the Leukemia and Lymphoma Society.

References

1. SEER. US Estimated 33-Year L-D Prevalence Counts on 1/1/2008. Surveillance E, End Results (SEER) Program, National Cancer Institute, DCCPS, Surveillance Research Program, Statistical Research and Applications Branch, released April 2011 based on the November 2010 SEER data submission. 2011 editor.
2. Cheson BD, Bennett JM, Grever M, et al. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood*. 1996; 87:4990–7. [PubMed: 8652811]
3. Landgren O, Albitar M, Ma W, et al. B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med*. 2009; 360:659–67. [PubMed: 19213679]
4. Shanafelt TD, Ghia P, Lanasa MC, et al. Monoclonal B-cell lymphocytosis (MBL): biology, natural history and clinical management. *Leukemia*. 2010; 24:512–20. [PubMed: 20090778]
5. Oscier DG. Cytogenetic and molecular abnormalities in chronic lymphocytic leukaemia. *Blood Rev*. 1994; 8:88–97. [PubMed: 7950479]
6. Zenz T, Mertens D, Dohner H, et al. Importance of genetics in chronic lymphocytic leukemia. *Blood Rev*. 2011; 25:131–7. [PubMed: 21435757]
7. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*. 2000; 343:1910–6. [PubMed: 11136261]
8. Ouillette P, Collins R, Shakhan S, et al. The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. *Clin Cancer Res*. 2011; 17:6778–90. [PubMed: 21890456]
9. Gunnarsson R, Mansouri L, Isaksson A, et al. Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. *Haematologica*. 2011; 96:1161–9. [PubMed: 21546498]
10. Kay NE, Eckel-Passow JE, Braggio E, et al. Progressive but previously untreated CLL patients with greater array CGH complexity exhibit a less durable response to chemoimmunotherapy. *Cancer Genet Cytogenet*. 2010; 203:161–8. [PubMed: 21156228]
11. 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; 85:3610–8. [PubMed: 7780145]
12. Shao L, Kang SH, Li J, et al. Array comparative genomic hybridization detects chromosomal abnormalities in hematological cancers that are not detected by conventional cytogenetics. *J Mol Diagn*. 2010; 12:670–9. [PubMed: 20724749]

13. Hagenkord JM, Monzon FA, Kash SF, et al. Array-based karyotyping for prognostic assessment in chronic lymphocytic leukemia: performance comparison of Affymetrix 10K2.0, 250K Nsp, and SNP6.0 arrays. *J Mol Diagn*. 2010; 12:184–96. [PubMed: 20075210]
14. Ouillette P, Collins R, Shakhan S, et al. Acquired genomic copy number aberrations and survival in chronic lymphocytic leukemia. *Blood*. 2011; 118:3051–61. [PubMed: 21795749]
15. Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science*. 2013; 339:1546–58. [PubMed: 23539594]
16. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011; 475:101–5. [PubMed: 21642962]
17. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med*. 2011; 208:1389–401. [PubMed: 21670202]
18. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011; 144:27–40. [PubMed: 21215367]
19. Edelmann J, Holzmann K, Miller F, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood*. 2012; 120:4783–94. [PubMed: 23047824]
20. Flinn IW, Neuberg DS, Grever MR, et al. Phase III trial of fludarabine plus cyclophosphamide compared with fludarabine for patients with previously untreated chronic lymphocytic leukemia: US Intergroup Trial E2997. *J Clin Oncol*. 2007; 25:793–8. [PubMed: 17283364]
21. Grever MR, Lucas DM, Dewald GW, et al. Comprehensive assessment of genetic and molecular features predicting outcome in patients with chronic lymphocytic leukemia: results from the US Intergroup Phase III Trial E2997. *J Clin Oncol*. 2007; 25:799–804. [PubMed: 17283363]
22. Peiffer DA, Le JM, Steemers FJ, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res*. 2006; 16:1136–48. [PubMed: 16899659]
23. Conlin LK, Thiel BD, Bonnemann CG, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet*. 2010; 19:1263–75. [PubMed: 20053666]
24. Rodriguez-Santiago B, Malats N, Rothman N, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010; 87:129–38. [PubMed: 20598279]
25. Laurie CC, Laurie CA, Rice K, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012; 44:642–50. [PubMed: 22561516]
26. Laurie CC, Doheny KF, Mirel DB, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol*. 2010; 34:591–602. [PubMed: 20718045]
27. Dewald GW, Brockman SR, Paternoster SF, et al. Chromosome anomalies detected by interphase fluorescence in situ hybridization: correlation with significant biological features of B-cell chronic lymphocytic leukaemia. *Br J Haematol*. 2003; 121:287–95. [PubMed: 12694251]
28. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009; 19:1639–45. [PubMed: 19541911]
29. Klein U, Dalla-Favera R. New insights into the pathogenesis of chronic lymphocytic leukemia. *Semin Cancer Biol*. 2010; 20:377–83. [PubMed: 21029776]
30. Gurvich N, Perna F, Farina A, et al. L3MBTL1 polycomb protein, a candidate tumor suppressor in del(20q12) myeloid disorders, is essential for genome stability. *Proc Natl Acad Sci U S A*. 2010; 107:22552–7. [PubMed: 21149733]
31. Bejar R, Levine R, Ebert BL. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J Clin Oncol*. 2011; 29:504–15. [PubMed: 21220588]
32. Yan XJ, Xu J, Gu ZH, et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet*. 2011; 43:309–15. [PubMed: 21399634]
33. Klampfl T, Harutyunyan A, Berg T, et al. Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood*. 2011; 118:167–76. [PubMed: 21531982]
34. Smith MR, Neuberg D, Flinn IW, et al. Incidence of therapy-related myeloid neoplasia after initial therapy for chronic lymphocytic leukemia with fludarabine-cyclophosphamide versus fludarabine:

long-term follow-up of US Intergroup Study E2997. *Blood*. 2011; 118:3525–7. [PubMed: 21803850]

35. Rawstron AC, Bennett FL, O'Connor SJ, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med*. 2008; 359:575–83. [PubMed: 18687638]
36. Fazi C, Scarfo L, Pecciarini L, et al. General population low-count CLL-like MBL persists over time without clinical progression, although carrying the same cytogenetic abnormalities of CLL. *Blood*. 2011; 118:6618–25. [PubMed: 21876118]

Figure 1.

Classification of chromosome anomalies. (a) Normal individuals have biparental disomy in the zygote (fertilized egg) and in somatic cells such as blood. (b) Individuals with an inherited CNA have the abnormality in the zygote and somatic cells. (c) and (d) Individuals with a mixture of normal cells and abnormal cells in the soma are assumed to have developed from a normal zygote and, therefore, the abnormality is classified as acquired. In (c), the abnormality is a CNA and in (d) a segmental UPD. (e) Individuals with a large (> 3Mb) non-mosaic CNA in somatic cells are considered to have an acquired anomaly (derived from a normal zygote) because CNAs > 3 Mb are very rarely inherited (see text). Each cell is shown with a pair of homologous autosomes, one maternal (red) and one paternal (blue).

Figure 2.

Examples of four clonal mosaic anomalies detected in E2997 patients. A pair of BAF and LRR plots is given for each sample-chromosome combination and each point is a single SNP. Points in BAF plots are color-coded by genotype (red=AA, cyan=AB, purple-blue=BB, black=missing call). The vertical black lines indicate the breakpoint(s) of the anomaly. The vertical gray rectangle is the centromeric gap. Horizontal pink lines are drawn at 0, 1/3, 1/2, 2/3 and 1 in the BAF plots. The solid horizontal red line in each plot is the median value for non-anomalous regions of the autosomes. The horizontal dashed red line is

the median value within the anomaly. A red box on the ideogram indicates the segment displayed in the plots. (a) and (b) Mosaic deletions, which have reduced LRR and four bands on the BAF plot, indicating a mixture of monosomic and disomic cells. The BAF pattern differs from the normal disomy observed in neighboring regions (three bands at 0, $\frac{1}{2}$ and 1) and from non-mosaic deletions, which would have BAF bands only at 0 and 1. (c) Mosaic trisomy, in which there are four bands and an elevation of LRR. The pattern differs from non-mosaic trisomy because the two intermediate BAF bands are closer together than the expected positions of $\frac{1}{3}$ and $\frac{2}{3}$ (and outside a 95% prediction interval for non-mosaic trisomy). (d) Mosaic UPD, which also shows four BAF bands with no change in LRR. Non-mosaic UPD would have only two bands at 0 and 1. As in this example, mosaic UPD is typically found only on terminal segments, most likely because it is due to mitotic recombination followed by outgrowth of one of the daughter cells.

Figure 3.

Circos plots of anomalies. Each arc represents an anomaly, color coded by copy number type (red = loss of copy number, blue = gain of copy number and green = copy neutral aUPD). Anomalies in the same annulus generally are not from the same subject. (a) 628 acquired anomalies (478 mosaic and 150 non-mosaic) found in 214 E2997 patients. (b) 514 mosaic anomalies found in 50,222 GENEVA subjects.

Figure 4.

Distributions of the number of anomalies per patient in E2997. (a) Mosaic anomalies, (b) non-mosaic > 3 Mb, (c) acquired (mosaic plus non-mosaic > 3 Mb) and (d) non-acquired. Note that the number of acquired anomalies per subject has a higher mean and variance than that for the non-acquired anomalies. The mean(SD) numbers are 2.23 (3.6) for mosaic, 2.94 (4.1) for acquired (including mosaic), and 2.27 (1.64) for non-acquired.

Figure 5.

Deletion coverage. Coverage is the number of deletions spanning a genomic interval, provided separately for GENEVA and E2997 subjects. The gray rectangle labeled “cen” is the centromeric gap and the vertical blue lines are gene positions.

Table 1

Genomic intervals with maximal anomaly coverage in E2997 subjects.

Type ^a	Chr ^b	Position ^c	Cytoband ^d	Length (Mb) ^e	E2997 Number ^f (of 214)	GENEVA Number ^g (of 50,222)	Other Number ^h	IgV _H status ⁱ	Other Studies ^j	Gene Number ^k	Genes of Interest ^l
aUPD	9	125003271-139986510	q33.2-q34.3	14.98	5	7	4	1/3	0	298	NOTCH1; NRARP
aUPD	13	50805079-113656958	q14.3-q34	62.85	10	6	10	7/8	3	222	
gain	2	94979-65644593	p25.3-p14	65.55	8	1	8	2/8	7	381	XPO1; ACP1; ALK; BCL11A; MYCN; REL; ADAM17
gain	3	82595269-95019980	p12.2-q11.2	12.42	4	0	4	3/4	1	14	
gain	3	167732066-199049392	q26.1-q29	31.32	7	0	7	3/6	4	223	KLHL6; PIK3CA
gain	7	153066277-158790820	q36.2-q36.3	5.72	4	0	4	1/4	2	31	
gain	8	126172002-128225139	q24.13-q24.21	2.05	8	6	7	2/5	8	6	
gain	8	128225139-128299967	q24.21	0.07	8	6	7	2/5	8	0	
gain	8	128300750-128907151	q24.21	0.61	8	6	7	2/5	8	4	MYC
gain	12	497680-9291155	p13.33-p13.31	8.79	49	8	27	9/39	10	141	
gain	12	9299886-10345823	p13.31-p13.2	1.05	49	8	27	9/39	11	20	
gain	12	10345823-20897382	p13.2-p12.2	10.55	49	8	27	9/39	11	96	
gain	12	21307680-49030386	p12.1-q13.13	27.72	49	8	27	9/39	10	180	MLL2; KRAS
gain	12	49030386-71189568	q13.13-q21.1	22.16	50	9	28	10/40	11	302	DTX3
gain	18	39287884-43803277	q12.3-q21.1	4.52	4	1	4	4/4	2	22	
gain	19	10175448-13448926	p13.2-p13.13	3.27	6	3	6	5/6	5	115	
gain	19	15783724-47981160	p13.12-q13.31	32.20	4	3	4	4/4	5	461	DLL3; NUMBL
gain	19	48435900-63776118	q13.31-q13.43	15.34	4	1	4	4/4	5	659	
loss	1	232783129-233197149	q42.3	0.41	4	1	4	0/4	1	3	
loss	3	57010-561487	p26.3	0.50	4	0	4	0/2	1	1	
loss	3	9470385-9799651	p25.3	0.33	4	0	4	0/1	1	8	
loss	4	22663148-24305954	p15.31-p15.2	1.64	7	0	7	1/6	6	3	DHX15; PPARGC1A
loss	4	175186279-183245566	q34.1-q35.1	8.06	4	0	4	1/4	1	17	
loss	4	184392509-191152644	q35.1-q35.2	6.76	4	0	4	1/4	2	47	

Type ^a	Chr ^b	Position ^c	Cytoband ^d	Length (Mb) ^e	E2997 Number ^f (of 214)	GENEVA Numbers ^g (of 50,222)	Other Number ^h	IgV _H status ⁱ	Other Studies ^j	Gene Number ^k	Genes of Interest ^l
loss	6	43819437-44493419	p21.1	0.67	5	0	5	4/5	2	16	Laurie et al.
loss	6	99580131-112260893	q16.2-q21	12.68	13	2	9	4/12	5	65	PRDM1
loss	8	21310858-27470311	p21.3-p21.1	6.16	9	0	9	4/8	7	64	TNFRSF10A; TNFRSF10B
loss	9	9876403-14484388	p23-p22.3	4.61	5	0	5	3/4	0	7	
loss	9	25430150-28045900	p21.3-p21.2	2.62	5	0	5	2/4	1	13	
loss	9	28054791-28979497	p21.2-p21.1	0.92	5	0	5	2/4	0	3	
loss	9	70225758-77281262	q13-q21.13	7.06	4	0	4	2/4	0	35	
loss	9	77788425-78914594	q21.13	1.13	4	1	4	2/4	0	7	
loss	9	82752285-84950566	q21.31-q21.32	2.20	4	1	4	2/4	0	6	
loss	9	103659726-105431618	q31.1	1.77	4	0	4	2/4	0	1	
loss	11	106587051-108000065	q22.3	1.41	31	10	27	6/28	11	14	ATM
loss	13	49552639-49554583	q14.3	0.00	66	24	52	32/54	13	1	DLEU1
loss	13	49565213-49966576	q14.3	0.40	66	24	52	32/54	13	3	DLEU1;DLEU2
loss	13	50061996-50067538	q14.3	0.01	66	27	52	32/55	13	0	
loss	14	71246688-73305480	q24.2-q24.3	2.06	5	2	5	2/4	7	19	NUMB
loss	14	74043493-75290582	q24.3	1.25	5	2	5	2/4	8	21	MLH3
loss	14	87785125-88858120	q31.3	1.07	5	2	5	2/4	5	7	
loss	15	38083251-39191343	q15.1	1.11	8	0	7	3/8	5	34	DLL4
loss	15	39197997-40495135	q15.1	1.30	8	0	7	3/8	5	29	LTK; MGA
loss	17	113794-4101175	p13.3-p13.2	3.99	16	4	16	8/15	8	84	
loss	17	4109714-5222824	p13.2	1.11	16	4	16	8/15	8	38	
loss	17	5225280-7724723	p13.2-p13.1	2.50	16	4	16	8/15	8	95	TP53
loss	18	4753493-6392126	p11.31	1.64	6	2	6	2/6	5	10	
loss	20	3655715-6859631	p13-p12.3	3.20	5	4	4	0/4	3	36	
loss	20	10842564-12359403	p12.2-p12.1	1.52	5	4	4	0/4	3	2	
loss	22	20847479-20945493	q11.22	0.10	4	2	4	1/3	5	1	

Column descriptions:

- ^aType = copy number type;
- ^bChr = chromosome;
- ^cPosition = genomic interval in base pairs (genome build NCBI136/hg18);
- ^dCytoband = genomic interval in cytobands;
- ^eLength = interval length in Mb;
- ^fE2997 Number = number of E2997 patients (out of a total of 214) with an acquired anomaly of the relevant copy number type that span the interval;
- ^gGENEVA number = number of GENEVA subjects (out of a total of 50,222) with a mosaic anomaly of the relevant copy number type that span the interval;
- ^hOther number = number of E2997 subjects with an acquired anomaly spanning the interval that also have other acquired anomalies;
- ⁱIgVH status = the number of E2997 patients with positive mutation status divided by the total number with non-missing status;
- ^jOther studies = number of other studies that also report acquired anomalies that cover at least 90% of this interval using cytoband limits (see Supplementary Information for references);
- ^kGene number = number of RefSeq genes that intersect this interval;
- ^lGenes of Interest = Names of selected genes of interest that intersect the interval.