Integrative genomics analyses reveal molecularly distinct subgroups of B-cell chronic lymphocytic leukemia patients with 13q14 deletion

Laura Mosca,1,* Sonia Fabris,1,* Marta Lionetti1, Katia Todoerti1, Luca Agnelli,1 Fortunato Morabito,2 Giovanna Cutrona,3 Adrian Andronache,1 Serena Matis,3 Francesco Ferrari,4 Massimo Gentile,2 Mauro Spriano,5 Vincenzo Callea,6 Gianluca Festini,7 Stefano Molica,8 Giorgio Lambertenghi Deliliers,1 Silvio Bicciato,4 Manlio Ferrarini,3,9 Antonino Neri1

1Centro per lo Studio delle Leucemie, Dipartimento di Scienze Mediche, Università di Milano, U.O. Ematologia 1, Fondazione IRCCS Policlinico, Milan, Italy; 2U.O.C. di Ematologia, Azienda Ospedaliera di Cosenza, Italy; 3Divisione di Oncologia Medica C, Istituto Nazionale per la Ricerca sul Cancro, IST, Genoa, Italy; 4Dipartimento di Scienze Biomediche, Università di Modena e Reggio Emilia, Modena, Italy; 5Dipartimento di Ematologia, Azienda Ospedaliera S. Martino, Genoa, Italy; 6Divisione di Ematologia, Azienda Ospedaliera, Reggio Calabria, Italy: 7Centro di Riferimento Ematologico-Seconda Medicina Azienda Ospedaliero-Universitaria, Ospedali Riuniti, Trieste, Italy; 8U.O.C. di Oncologia, Azienda Ospedaliera “Pugliese-Ciaccio”, Catanzaro, Italy and 9Dipartimento di Oncologia, Biologia e Genetica, Università degli Studi di Genova, Genoa, Italy

*These authors contributed equally to this work.

Correspondence: Prof. Antonino Neri, Department of Medical Sciences, University of Milan, Via F. Sforza 35, 20122 Milano, Italy. Tel. +39.02.55033328; Fax: +39.02.50320403
E-mail: antonino.neri@unimi.it

Running title: Integrative genomics analysis of 13q14 deleted CLL
Translational Relevance

The 13q14 deletion represents the most common genomic aberration in CLL (50%). Although it is associated with a favorable prognosis when present as the sole abnormality, the imprecise knowledge of the genes implicated and its genetic heterogeneity have limited our understanding on the pathogenetic mechanisms contributing to the disease. Microarray technology has improved the possibility of combining genome-wide DNA with transcriptomic profiles to identify potential candidate tumor genes related to underlying chromosomal aberrations. Based on SNP-array, our study shows the presence of two distinct molecular groups of patients with del(13q14) based on the deletion size and the presence of biallelic deletions. Notably, global gene expression profiling identified a significant transcriptional deregulation specifically associated with the two groups. Our data highlight the presence of two distinct molecular types of 13q14 deletion that may be of clinical relevance for the biology of as well as for the prognostic stratification of the disease.

Abstract

Purpose: Chromosome 13q14 deletion occurs in a substantial number of chronic lymphocytic leukemia (CLL) patients and it is believed to play a pathogenetic role. The exact mechanisms involved in this lesion have not yet been fully elucidated because of its heterogeneity and the imprecise knowledge of the implicated genes. This study was addressed to further contribute to the molecular definition of this lesion in CLL.

Experimental Design: We applied SNP-array technology and gene expression profiling data to investigate the 13q14 deletion occurring in a panel of 100 untreated, early-stage (Binet A) patients representative of the major genetics, molecular and biological features of the disease.

Results: Concordantly with FISH analysis, SNP-arrays identified 44 patients with
del(13)(q14) including 11 cases with a biallelic deletion. The shorter monoallelic deletion was 635 kb long. The loss of the miR-15a/16-1 cluster occurred in all del(13)(q14) cases except in two patients with a monoallelic deletion who retained both copies. MiR-15a/16 expression was significantly down-regulated only in patients with the biallelic loss of the miRNA cluster compared to 13q normal cases. Finally, the natural grouping of SNP profiles by means of non-negative matrix factorization algorithm, showed that patients could be classified into two separate clusters mainly characterized by short/biallelic vs. wide/monoallelic 13q14 deletions. Supervised analyses of expression data demonstrated that specific transcriptional profiles are correlated with these two genomic subgroups.

**Conclusions:** Overall, our data highlight the presence of two distinct molecular types of 13q14 deletions which may be of clinical relevance in CLL.

Key words: CLL; integrative genomics; 13q deletion; miR-15a/miR-16-1
Introduction

B-cell chronic lymphocytic leukemia (CLL) is a lymphoproliferative disorder characterized by a variable clinical course: some patients progress rapidly toward more advanced stages whereas others survive for a long time without disease progression (1). Recently, considerable efforts have been addressed to the identification of genomic aberrations which could explain the pathogenetic mechanisms and the clinical heterogeneity of the disease (2-5). Deletions of 13q14, 11q22.3 and 17p13, and trisomy 12 are common in CLL and may play a role in pathogenesis and disease progression (4). 13q14 deletion occurs in approximately 50% of CLL and is associated with a favorable prognosis when present as the sole abnormality (6). Notably, two microRNA genes, miR-15a and miR-16-1, located at 13q14, have been reported to be down-regulated in del(13)(q14) patients (7), and strongly suggest their potential role in the disease. However, the 13q14 deletion is not always accompanied by defects in miR-15a/16-1 expression, suggesting a more complex heterogeneity of the deletion itself (8, 9).

The recent introduction of microarray technology has improved the possibility of combining genome-wide DNA analyses with transcriptomic profiles, thus allowing the identification of potential candidate tumor genes related to underlying chromosomal alterations. In order to further elucidate the genomic complexity of the 13q14 deletion in CLL, we applied SNP-array technology to a panel of 100 untreated patients with Binet stage A disease including 44 patients with 13q deletion as assessed by FISH. We then investigated the expression levels of the miR-15a/16 cluster by means of quantitative real-time PCR (Q-RT-PCR) and global gene expression profiling in a representative panel of cases. The integration between genomic and expression data allowed the identification of two distinct molecular groups of patients with 13q14 deletions. Our data may have implications for the biology and the prognostic stratification of CLL.

Materials and Methods
Patients

The study included samples from 100 untreated CLL patients in Binet stage A: 60 from a retrospective collaborative Italian study (10) and 40 enrolled in an Italian prospective multicentre study (GISL O-CLL1). Eligibility required a diagnosis of a typical CLL phenotype: CD5/CD19+ and CD23+, weak surface immunoglobulin [sIg] and the monotypical expression of k or λ light chains by neoplastic cells. Patients were selected in order to provide a good numerical representation of cytogenetic lesions, as assessed by fluorescence in situ hybridization (FISH), and IgVH mutation status. All of the patients gave their informed consent in accordance with our institutional guidelines.

Sample preparation, immunophenotype and prognostic markers

Peripheral blood mononuclear cells were isolated by means of Ficoll-Hypaque gradient centrifugation (Seromed, Biochrom KG, Berlin, Germany), and CD38 and ZAP-70 expression were determined by flow-cytometry (11). IgVH gene usage and mutational status were established as previously described (12), with a 2% cut-off value being used to define mutated and unmutated patients. For the microarray analyses, the CLL cells were enriched by negative selection when less than 90% (11).

Fluorescence in situ hybridization (FISH)

The most common genomic aberrations, del(17)(p13), del(11)(q23), del(6)(q23), del(13)(q14) and trisomy 12, were investigated by means of interphase FISH hybridization. All of the probes are commercially available (Vysis, Downers Grove, IL) (13).

High-density SNP-arrays and data analysis

Two hundred and fifty nanograms of total genomic DNA were processed in accordance with the manufacturer’s protocol (Affymetrix), hybridized using Affymetrix GeneChip® Human
Mapping 250K NspI microarrays, and subsequently scanned using a GeneChip® Scanner 3000 7G. The images were acquired using the Affymetrix GeneChip® Operating System (GCOS 1.4). The accuracy of the SNP-array data were confirmed by the mean and median call rates of respectively 95.75% and 96.12% (the quality control specification for 250K arrays is a call rate greater than 93%).

The entire procedure for the CN estimation has been fully described previously (14). Briefly, the raw data relating to the individual SNPs were extracted from CEL files and converted into signal intensities using GTYPE 4.1 and Affymetrix Copy Number Analysis Tool (CNAT 4.0.1) software. Each sample was compared with a set of 48 normal Caucasian HapMap references available on the Affymetrix web site ¹ and the genomic smoothing window of the Hidden Markov Model algorithm was set at 0. After pre-processing, piecewise constant estimates of the underlying local DNA copy number alterations were calculated using the DNAcopy Bioconductor package, which looks for optimal breakpoints on the basis of circular binary segmentation (CBS) (15), and the median values of the estimated profiles were scaled back to a nominal multiplicity of two. After scaling, we determined all the clusters appearing on the frequency distribution (histogram) of the SNP values for all the SNP probes using the $k$-means algorithm on the cumulative profile of all of the array data. The transition from one cluster to the other, namely the threshold from one CN to another, was then estimated as the meeting point of two Gaussian curves and approximating the distributions of mean $\mu$ and variance $\sigma^2$ of the two neighboring clusters of SNPs identified in the final clustering. Thus, the thresholds result as follows: inferred CNs of more than 2.1 and 2.5 corresponded to gain and amplification, whereas CNs of less than 1.9 and 1.34 corresponded to loss and biallelic deletion. After segmentation, the SNP dataset was compressed to 1346 probes by eliminating redundant probes to better balance less represented regions with those showing a large

¹ http://www.affymetrix.com/support/technical/sample_data/500k_data.affx
number of probes, as previously described (14). A probe was defined as redundant when it showed copy number values identical to those of the most contiguous upstream probe in all samples. Then a non-negative matrix factorization (NMF) procedure adapted to R from the original MATLAB package (16) was used to evaluate the meaningful clusters across the whole dataset. For each factor level from two to ten, NMF was repeated 100 times to build a consensus matrix, and the samples were assigned to clusters on the basis of the consensus results.

Sequence copy number determination by means of quantitative real-time PCR (Q-RT-PCR)

Real-time PCR was performed according to a published protocol (17) using the ABI Prism® 7900 sequence detection system. Singleplex amplification reactions were carried out in triplicate using 40 ng of template DNA, 1X TaqMan® Universal Master Mix, no AmpErase UNG, and a 1X Primer-Probe mix (Applied Biosystems, Foster City, CA) containing sequence-specific primers and a fluorogenic probe. The TaqMan® RNase P Detection Reagents kit and a Custom TaqMan® Gene Expression Assay (forward primer: 5’-GCAATGTCAGCAGTGCCTTAG-3’; reverse primer: 5’-CAGCAGCAGTATCTGGAGAT-3’; probe: 5’FAM-CAGCACGTAAATATTG-3’)) were used to amplify the ribonuclease P RNA component H1 (RPPH1) gene mapped within 14q11.2 (present in two copies in all of the subjects and thus used as an internal standard) and the miR-15a/16-1 cluster. The comparative ΔCt method was used for quantification purposes. DNA samples from ten control individuals with two copies of the commonly deleted region were selected for PCR calibration. The estimated haploid gene CN was given by the formula $2^{-\Delta\Delta C_t}$, and the predicted CN was calculated as the closest integer number to the estimated CN (Determining Gene Copy Number using TaqMan® Real-Time PCR Assays on the 7900 HT-Quick Reference Card–Applied Biosystems (18).
Quantification of specific gene expression by means of Q-RT-PCR

Q-RT-PCR of specific genes (TPT1, WBP4, PEA15 and LGALS1) and mature miRNAs (hsa-miR-15a and hsa-miR-16) was performed using commercial TaqMan® assays (Applied Biosystems) as previously described (19). The relative gene and miRNA expression levels were calculated using the $2^{-\Delta Ct}$ method (Applied Biosystems User Bulletin No. 2) as previously described (19). Pearson’s correlation coefficient was calculated to test correlation between gene expression and Q-RT-PCR data.

Gene expression profiling (GEP)

Twenty-two CLL samples with 13q deletion included in the retrospective database underwent GEP analysis. Total RNA was extracted using the TRIZOL reagent (Invitrogen, Carlsbad, CA, USA) and purified using the RNeasy total RNA Isolation Kit (Qiagen, Valencia, CA, USA). The biotin-labeled complementary RNA was prepared and hybridized with GeneChip® Human Genome U133A Arrays (Affymetrix Inc., Santa Clara, CA, USA), which were scanned (GeneChip® Scanner 3000 7G; Affymetrix Inc.) in accordance with manufacturer’s protocols. The probe data were converted to expression values using the Bioconductor function for the robust multiarray average procedure, as described previously (20). Supervised analysis was made using Significant Analysis of Microarrays software, version 3.0.2 ²(21). The cut-off point for significance at a median false discovery rate (FDR) ≤5% (i.e. $q$-value<0.05) was determined by tuning the $\Delta$ parameter on the FDR and controlling the $q$-value of the selected probes. dChip software (22) was used to represent the selected probe lists; NetAffx ³ was used for the functional annotation study of the list. The genotyping and gene expression data are available at the NCBI’s Gene Expression Omnibus through GEO Series Accession Nos. GSE15643 and GSE16746.

³ https://www.affymetrix.com/analysis/netaffx/
Statistical analysis

The data were statistically analyzed using conventional procedures in R software (Kruskal-Wallis tests, Kendall’s τ correlations, Fisher’s exact tests, and q-value calculations).

Results

Molecular characteristics of the CLL patients

The main characteristics of the 100 CLL patients included in the study are reported in Supplementary Tables 1 and 2. Based on FISH analyses, 44 cases carried the 13q14 deletion, 34 of which as the sole abnormality, while the remaining cases showed 11q22.3 (6 pts), 17p13.1 (2 pts), 11q22.3 and 17p13.1 deletions (1 pt), or 6q23 deletion (1pt). Biallelic 13q14 deletions were identified in 11 cases, 5 of which (CS3, CZ42, GE110, CS0100 and LD0062) showed the presence of subclones characterized by the biallelic deletion (ranging from 68% to 83.5%). Trisomy 12 was present in 21 patients as a sole abnormality. Overall, the 11q22.3, 17p13.1 and 6q23 deletions were present in 15, 7 and 2 patients respectively. Fifty-five cases had unmutated IgVH genes; ZAP-70 and CD38 were expressed in 42 and 46 cases, respectively. 13q deletion was present in 25/45 (55.5%) patients with mutated IgVH genes, 15/42 (35.7 %) of the ZAP-70 positive and 15/46 (32.6 %) of the CD38 positive patients.

SNP-array data were concordant with FISH results (Supplementary Tables 2 and 3). In addition, SNP-arrays detected short deletions involving the second 13q allele in three del(13)(q14) patients (TS12, VB0013 and PS0044); deletions were respectively 467, 468 and 291 kb long and located centromerically to the LSI D13S25 FISH probe (Figure 1A-B). Furthermore, SNP-arrays revealed a 6q deletion of approximately 39 Mb in length and located upstream of the LSI MYB FISH probe in a single patients (CZ47) (data not shown). These aberrations were confirmed by specific FISH probes in all cases (data not shown).
Characterization of the 13q deletion by SNP-arrays

The SNP-arrays showed that the 13q deletions varied considerably in size, ranging from a minimum of 291 kb (PS0044) to a maximum of 56 Mb (CZ36) (Figure 1A-B). The minimal monoallelic deletion (MMD) was 635 kb long spanning from SNP_A-2003314 to SNP_A-2003318 (physical position 49,635,024 bp to 50,270,550 bp) (Figure 1C); notably, the miR-15a/16-1 cluster is located upstream (~87 kb) of the centromeric SNP_A-2003314. With regard to the miR-15a/16-1 cluster, SNP-array analysis depicted the following scenario: (i) retention of two copies of the cluster in 56 patients with a normal 13q and in two cases (CD0018 and CS95) showing a 13q monoallelic deletion telomeric to the cluster; (ii) retention of one copy of the cluster in 29 patients with a 13q monoallelic deletion and in two cases (CZ42 and PS0044) showing a deletion telomeric to the cluster in the second allele; (iii) loss of both copies of the cluster in the remaining 11 patients showing biallelic deletions (Figure 1A-B and Supplementary Table 4).

The status of the miR-15a/16-1 cluster was investigated further using a custom Q-RT-PCR assay on DNA from 32 patients, including 10 with biallelic deletions, ten with monoallelic deletions, and 12 normal at 13q14 based on SNP-array data. As shown in Figure 2A and Supplementary Table 4, the estimated CN obtained with this approach were concordant with the SNP-array data (P<0.0001) with three exceptions: two patients (CS3 and GE110) classified as CN=1 by RT-PCR and CN=0 by SNP-array, who showed two distinct cell populations by FISH characterized by either mono- or biallelic losses, and one patient (RC21) classified as normal by RT-PCR and CN=1 by SNP-array who showed a monoallelic loss in a fraction of interphase nuclei (61%) (Figure 2A). These findings suggest that clonal heterogeneity may account for the discrepancies between the real-time PCR and SNP-array data.

Furthermore, we compared miR-15a and miR-16 expression in 78/100 cases including 11 with biallelic deletions, 28 with monoallelic deletions, and 39 normal cases based on SNP-
array CN values. Cases with biallelic deletions had a miRNA expression significantly lower than cases retaining one or two copies ($P<0.0001$ for $miR-15a$, $P=0.0001$ for $miR-16$) whereas no statistically significant differences in miRNA expression levels could be observed between normal cases and those retaining a single copy of the cluster (Figure 2B and C and Supplementary Table 5).

Finally, SNP-array documented a loss of the $RB1$ gene in 28/44 cases with 13q14 deletion (63.6%). Two samples (CS0100 and TS12) showed a biallelic loss of the $RB1$ locus which was associated with a homozygous deletion of the 13q14 FISH probe (case CS0100) or was the result of a small deletion involving the $RB1$ locus on the second allele (case TS12) (Figure 1B). All these findings were validated by FISH using the $RB1$ specific clone RP11-305D15 (data not shown).

**Identification of two genetically distinct patient subgroups with del(13)(q14)**

In order to identify the most significant natural grouping of genome profiles in our panel, we used the NMF algorithm. This analysis led to the identification of four major groups (correlation coefficient=0.95) characterized by 13q deletion (group I and II, 27 and 13 cases, respectively), and by trisomy 12 (group III, 21 cases), whereas no specific alteration could be associated with group IV (39 cases) (Figure 3 and Table 1). Patients from group I and II differed in the deletion size and for the occurrence of mono- or biallelic deletions: group I included samples with relatively smaller losses, ten of which showed biallelic deletions, whereas group II included samples with larger losses, all but one (GE110) showing monoallelic deletions. The presence of four samples with 13q14 deletions in group IV was in all likelihood related to the fact that the inferred CN values (ranging from 1.74 to 1.85) were very close to the threshold (1.9) between monoallelic deletion and the retention of two copies. These “marginal” CN data are consistent with clonal heterogeneity and are in accordance with the small percentage of interphase nuclei detected by FISH showing 13q14 deletion (no more
than 21% in all cases) which may affect the clustering analysis. *RB1* deletion occurred in 14/27 cases in group I (52%), 12/13 in group II (92%), and 2/4 in group IV. Only 10 of the 44 patients with del(13q14) carried additional known genetic lesions, namely 11q22.3 (6 pts), 17p13.1 (2 pts), 11q22.3 plus 17p13.1 (1 pt), 6q23 deletions (1 pt); however no correlation was found with the two specific subgroups I and II. Finally, there was no significant association between cases within NMF group I and II and CD38 or ZAP-70 expression or *IgVH* mutational status (data not shown).

*Gene expression patterns associated with distinct groups of patients with 13q deletion*

Fifteen cases from group I and 7 from group II, for whom RNA material was available, were profiled on GeneChip® HG-U133A arrays. In order to verify whether the two groups could be divided based upon the differences of their expression profiles, we made a conventional unsupervised analysis using hierarchical agglomerative clustering at different levels on: i) all of the probes in the array; ii) all of the probes mapped to the chromosome 13; or iii) all of the probes within the 13q14 region. Notably, a significant grouping (i.e. group I and group II patients in two separate branches) was only detected when the 89 probes (65 genes) mapped on the 13q14 region were used (*P*<0.0001) (Figure 4A). This finding indicates that, albeit differences exist between the two groups, these are not sufficient to drive the clustering when the global transcriptional profiling was considered (i.e. the whole matrix). Thus, we carried out a supervised analysis to characterize the specific transcriptional profiles distinguishing NMF group I and II. We identified 76 differentially expressed probe sets, specific to 63 well-characterized genes (Figure 4B), ten of which were down-regulated and fifty-three up-regulated in group II compared to group I. Six of the ten down-regulated genes mapped to 13q13-q14 and exhibited a monoallelic deletion in most or all cases within group II (3/7=43% for *FOXO1*; 4/7=57% for *EXOSC1* and *WBP4*; 6/7=86% for *TPT1* and *NUFIP1*; 7/7=100% for *ESD*) compared to cases in group I (1/15=6.7% for *EXOSC1*, *TPT1*, *ESD*, *FOXO1* and...
NUFIP1; 2/15 = 13.3% for WBP4). The whole list of differentially expressed genes is reported in Table 2.

We selected two down-regulated (TPT1 and WBP4) and two up-regulated (PEA15 and LGALS1) genes in group II vs. group I for Q-RT-PCR validation of the microarray data. The Q-RT-PCR analyses were made in a subset of 20/22 patients for whom RNA material was available (fourteen belonging to group I, six to group II). The correspondence between the microarray and Q-RT-PCR data were evaluated by assessing the correlation coefficients of the expression levels determined by the two analyses: the coefficients were 0.71 for TPT1; 0.79 for WBP4; 0.60 for PEA15 and 0.78 for LGALS1 probe, thus indicating a very good concordance for all of the tested genes (Supplementary Figure 2).

Discussion

The present study focused on the molecular characterization of the 13q14 deletion based on SNP-arrays and gene expression profiling analyses. Several aspects distinguish this from previous reports (9, 23) including a homogeneous cohort of untreated early-stage patients (Binet A), the use of highly purified CLL samples and a higher-resolution SNP-array (250K NspI), the application of stringent statistics to define the genetic groups, and the integration of genomic and gene expression data in a significant number of cases. The main findings relate to the description of the topography of 13q14 deletions, the re-assessment of the fate of miR-15a/16-1 cluster in relation to the genomic losses and, perhaps more importantly, the definition of two major genomic groups of patients with 13q14 deletions also characterized by distinct transcriptional patterns.

The natural grouping of genome profiles by SNP-array analysis showed that the complex scenario of CN alterations affecting CLL is mainly driven by the presence of the 13q14 deletion and trisomy 12. Our SNP-arrays approach extends previous limited evidence showing that del(13)(q14) patients are characterized by differently sized deletions. Ouillette et
al. documented a significant correlation between a wider deletion encompassing the \textit{RB1} gene and a higher Rai clinical stage at diagnosis or in previously treated patients (9). In preliminary observations on our cohort of Binet stage A patients, we did not find any significant differences in the time to treatment between the patients with shorter/biallelic (group I) or wider/monoallelic losses (group II) (data not shown), although this finding awaits confirmation by ongoing prospective studies.

Our data support the notion that loss of the \textit{miR-15a/16-1} locus occurs in virtually all the CLL cases with del(13)(q14): in fact, among 44 cases with 13q deletion in our study, both copies of miRNA cluster were retained in only two patients with a monoallelic 13q14 deletion. Furthermore, we found that reduced \textit{miR-15a} and \textit{miR-16} expression levels in patients with del(13)(q14) significantly correlated only with the presence of a biallelic loss, which is in agreement with some recent data (8, 9) but not with older findings (7). Overall, these findings suggest a need to redefine the pathogenetic role of \textit{miR-15a} and \textit{miR-16-1} in the context of the molecular subtypes of 13q14 deleted patients.

The availability and integration of GEP data allowed a more comprehensive overview of the genetic complexity of the 13q14 deletion. Indeed, this approach led to the determination of distinct transcriptional signatures associated with different groups of 13q14 deleted patients (short/biallelic \textit{vs.} long/monoallelic lesion) identified by the NMF algorithm. In particular, a significant gene dosage effect has been observed involving the down-regulation of genes in group II (long/monoallelic deletion), six of which localized within the 13q14 region. Among the down-regulated genes, we should note \textit{TPT1/TCTP} (translationally controlled tumor protein) encoding for a highly conserved multifunctional protein acting as a pro-survival and growth stimulating factor (24) which inhibits BAX-induced apoptosis (25).

Most of the up-regulated genes in group II are involved in cell motility and adhesion, regulation of cell proliferation, tumor cell migration, metastasis, angiogenesis and apoptosis and some of these may contribute to lymphomagenesis. This is the case of the autocrine
motility factor (AMF)/glucose phosphate isomerase (GPI) gene which is upregulated in several human cancers and encodes for a housekeeping cytosolic enzyme involved in both glycolysis and gluconeogenesis (26). The basigin gene (BSG) encodes a cell surface glycoprotein of the Ig superfamily expressed at the surface of tumor cells metastasizing in bone marrow (27) and is believed to induce matrix metalloproteinases (MMP) production (28). The Galectin-1 (alias LGALS1) gene encoding a 14 kDa lectin, is overexpressed in numerous tumors including lymphomas (29) as well as CLL and is implicated in abnormal mechanisms of cell adhesion, induction of apoptosis and tumor angiogenesis (30, 31). PAK2 belongs to the p21-activated kinase family, which are well-known regulators of cytoskeletal remodelling, cell motility, proliferation and apoptosis (32). Finally, two other upregulated genes, β-parvin (PARVB) and vimentin (VIM), have been reported to play a critical role in transducing signals from integrins to the actin cytoskeleton and intracellular signaling proteins (33). Low levels of PARVB correlated with low adhesion to collagen (34), while increased levels reduce the activating phosphorylation of AKT (35), causing propensity to apoptosis. Up-regulation of VIM is thought to provide a selective advantage to tumor cells following signaling cues from mesenchymal and epithelial extracellular matrixes (36). Notably, a modulation of genes thought to act as regulators of tumor invasion and integrin-mediated cell motility and adhesion has been recently described in CLL, in particular during disease progression (37, 38).

In conclusion, our data may represent a valid contribution to the definition of the genomic profile of CLL. In particular, we provide evidence of two clearly-distinguishable molecular subtypes among CLL patients with 13q14 deletion that may contribute towards the better understanding of the pathogenetic and clinical relevance of this lesion in CLL.

Conflicts of interest. The authors declare that they have no conflict of interest.
Acknowledgements. This study was supported by grants from AIRC to AN (IG 4659) and FM (RG 6432 co-financed by AIRC, CARICAL, Fondazione ‘Amelia Scorza’ and Provincia di Cosenza); AIL Sezione Milano; Fondazione ‘Amelia Scorza’ ONLUS, Cosenza; Progetti Strategici - Ricerca Finalizzata Ministero Italiano della Salute “RFPS20063339960” (to GC) and “RFPS2006340196” (to FM and MF); and FIRB (Grant No RBIP06LCA9 to MF). The Progetto Ordinario Ricerca Finalizzata Ministero Italiano della Salute-2007 (to GC), Progetto Compagnia San Paolo (to GC), and the Fondazione Internazionale Ricerche Medicina Sperimentale (FIRMS) provided financial and administrative assistance. LA and SM were supported by fellowships from the Fondazione Italiana Ricerca sul Cancro (FIRC).
References


FIGURE LEGENDS

**Figure 1.** Chromosome 13 deletion pattern in 44 CLLs. A. Monoallelic (grey lines) and biallelic losses (black lines) in the 44 deleted patients. B. The enlarged subregion spanning 13q14.2-q14.3 between physical positions 44.50 and 52.00 Mb, including the MMD. C. The localization of SNPs and the FISH probe encompassing the minimally altered region (grey bar). Gene locations and transcriptional orientation are indicated by the horizontal arrows at the bottom.

**Figure 2.** Sequence copy number and expression quantification of miR-15a/16 as assessed by Q-RT-PCR. A. Vertical axis: $2^{-\Delta\Delta Ct}$ values for the 32 investigated patients. The inferred CN values obtained on the basis of the criteria described in Material and Methods are shown for each patients as white bars (CN=0), striped bars (CN=1) and black bars (CN=2). Horizontal axis: patient distribution into three classes on the basis of the number of alleles (0, 1 and 2) detected by SNP-array. B. and C. Expression of miR-15a and miR-16 as calculated using the $2^{-\Delta Ct}$ method and ranked according to the corresponding inferred CN value (CN 0 = 11 cases; CN 1 = 28 cases, including PS0044 patient; CN 2 =39 cases, including CD0018 and CS95 patients). P-values of the comparison of between the biallelic and monoallelic groups.

**Figure 3.** Heat-map of significantly altered DNA regions in 100 CLLs, as assessed by means of GeneChip® Human Mapping 250K Nsp. The genomic profiles of the CLL samples (horizontal lines) are clustered into four groups in accordance with the NMF method. Horizontal axis: chromosome localization. The dashed lines represent the centromeres. Light-green = biallelic deletion; dark-green = loss; white = normal copy number; red = gain.

**Figure 4:** Identification of gene signatures characterizing 13q14 classes. A. Dendrogram of the 22 CLL samples clustered according to the expression profiles of the genes located at...
13q14. B. Expression profiles of the NMF CLL group I vs II for the 76 probesets selected by means of a SAM two-class analysis. Information on chromosome 11, 13 and 17 deletions, chromosome 12 trisomy (+ = positive, - = negative, /= biallelic deletion), CD38 (+ = ≥30%, - = <30%), ZAP-70 expression (+ = ≥30%, - = <30%) and IgVH mutational status (+ = mutated, - = unmutated) is included alongside the patient ID. The color scale bar represents the relative changes in gene expression normalized by the standard deviation, and the color changes in each row represent gene expression in relation to the mean across the samples.
Figure 1.
Figure 2.

A.

B.

C.
Table 1. Molecular characteristics of the four genomic groups

<table>
<thead>
<tr>
<th>NMF * cluster</th>
<th>No. of cases</th>
<th>del(11q22.3)</th>
<th>Monoallelic del(13q14)</th>
<th>Biallelic del(13q14)</th>
<th>del(17p13.1)</th>
<th>12</th>
<th>CD38+</th>
<th>ZAP70+</th>
<th>Unmutated IgVH</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>27</td>
<td>(11.1%)</td>
<td>(63.0%)</td>
<td>(37.0%)</td>
<td>(3.7%)</td>
<td>0</td>
<td>8</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>(30.8%)</td>
<td>(92.3%)</td>
<td>(7.7%)</td>
<td>(15.4%)</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>II</td>
<td>13</td>
<td>(30.8%)</td>
<td>(92.3%)</td>
<td>(7.7%)</td>
<td>(15.4%)</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>III</td>
<td>21</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>21 (100%)</td>
<td></td>
<td>17</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>IV</td>
<td>39</td>
<td>(20.5%)</td>
<td>(10.3%)</td>
<td>0%</td>
<td>(10.3%)</td>
<td>0</td>
<td>17</td>
<td>17</td>
<td>23</td>
</tr>
</tbody>
</table>

* NMF = non-negative matrix factorization
<table>
<thead>
<tr>
<th>Probe Set ID</th>
<th>Gene Symbol</th>
<th>Gene Title</th>
<th>Score(d)</th>
<th>Fold Change</th>
<th>q-value(%)</th>
<th>Cytoband</th>
</tr>
</thead>
<tbody>
<tr>
<td>212809_s_at</td>
<td>TPT1</td>
<td>tumor protein, translationally-controlled 1</td>
<td>-6.3308</td>
<td>0.79</td>
<td>0.13q12-q14</td>
<td></td>
</tr>
<tr>
<td>209090_s_at</td>
<td>ESD</td>
<td>esterase D/formylglycine hydratase</td>
<td>-5.0130</td>
<td>0.56</td>
<td>0.13q14-q14.2</td>
<td></td>
</tr>
<tr>
<td>215136_s_at</td>
<td>EXOSC8</td>
<td>exosome component B</td>
<td>-4.5888</td>
<td>0.61</td>
<td>2.13q13</td>
<td></td>
</tr>
<tr>
<td>221183_s_at</td>
<td>CNOT2</td>
<td>CCR4-5, NOT transcription complex, subunit 2</td>
<td>-4.4811</td>
<td>0.69</td>
<td>2.19q12-q15</td>
<td></td>
</tr>
<tr>
<td>202723_s_at</td>
<td>FOXD1</td>
<td>forkhead box O1</td>
<td>-3.9060</td>
<td>0.67</td>
<td>3.26q14</td>
<td></td>
</tr>
<tr>
<td>205134_s_at</td>
<td>NUP1P1</td>
<td>nuclear fragile X mental retardation protein interacting protein 1</td>
<td>-3.2572</td>
<td>0.66</td>
<td>3.26q14</td>
<td></td>
</tr>
<tr>
<td>210346_s_at</td>
<td>CLCA4</td>
<td>CDC-like kinase 4</td>
<td>-3.3146</td>
<td>0.65</td>
<td>3.26q35</td>
<td></td>
</tr>
<tr>
<td>232526_s_at</td>
<td>MARCHE3</td>
<td>membrane-associated ring finger (C3HC4) 3</td>
<td>-4.1060</td>
<td>0.51</td>
<td>4.20q5p2.2</td>
<td></td>
</tr>
<tr>
<td>203597_s_at</td>
<td>WBP4</td>
<td>WW domain binding protein 4 (formin binding protein 21)</td>
<td>-4.1258</td>
<td>0.66</td>
<td>4.20q14</td>
<td></td>
</tr>
<tr>
<td>204155_s_at</td>
<td>QSK</td>
<td>serine/threonine-protein kinase QSK</td>
<td>-4.0622</td>
<td>0.72</td>
<td>4.72q11</td>
<td></td>
</tr>
<tr>
<td>213011_s_at</td>
<td>TP12</td>
<td>triosephosphate isomerase 1</td>
<td>-3.9355</td>
<td>1.60</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>201013_s_at</td>
<td>PAICS</td>
<td>phosphatidylinositol phosphate phosphatase, type 5</td>
<td>-3.9407</td>
<td>1.73</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>204042_s_at</td>
<td>PHACTR3</td>
<td>phosphatase and actin regulator 3</td>
<td>-3.9428</td>
<td>1.33</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>201888_s_at</td>
<td>LGALS1</td>
<td>lectin, galactoside-binding, soluble, 1</td>
<td>-3.9627</td>
<td>2.74</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>204155_s_at</td>
<td>QSK</td>
<td>serine/threonine-protein kinase QSK</td>
<td>-4.0622</td>
<td>0.72</td>
<td>4.72q11</td>
<td></td>
</tr>
<tr>
<td>213011_s_at</td>
<td>TP12</td>
<td>triosephosphate isomerase 1</td>
<td>-3.9355</td>
<td>1.60</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>201013_s_at</td>
<td>PAICS</td>
<td>phosphatidylinositol phosphate phosphatase, type 5</td>
<td>-3.9407</td>
<td>1.73</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>204042_s_at</td>
<td>PHACTR3</td>
<td>phosphatase and actin regulator 3</td>
<td>-3.9428</td>
<td>1.33</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>201888_s_at</td>
<td>LGALS1</td>
<td>lectin, galactoside-binding, soluble, 1</td>
<td>-3.9627</td>
<td>2.74</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>204155_s_at</td>
<td>QSK</td>
<td>serine/threonine-protein kinase QSK</td>
<td>-4.0622</td>
<td>0.72</td>
<td>4.72q11</td>
<td></td>
</tr>
<tr>
<td>213011_s_at</td>
<td>TP12</td>
<td>triosephosphate isomerase 1</td>
<td>-3.9355</td>
<td>1.60</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>201013_s_at</td>
<td>PAICS</td>
<td>phosphatidylinositol phosphate phosphatase, type 5</td>
<td>-3.9407</td>
<td>1.73</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>204042_s_at</td>
<td>PHACTR3</td>
<td>phosphatase and actin regulator 3</td>
<td>-3.9428</td>
<td>1.33</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>201888_s_at</td>
<td>LGALS1</td>
<td>lectin, galactoside-binding, soluble, 1</td>
<td>-3.9627</td>
<td>2.74</td>
<td>4.72q12</td>
<td></td>
</tr>
<tr>
<td>204155_s_at</td>
<td>QSK</td>
<td>serine/threonine-protein kinase QSK</td>
<td>-4.0622</td>
<td>0.72</td>
<td>4.72q11</td>
<td></td>
</tr>
</tbody>
</table>
Integrative genomics analyses reveal molecularly distinct subgroups of B-cell chronic lymphocytic leukemia patients with 13q14 deletion

Laura Mosca, Sonia Fabris, Marta Lionetti, et al.

Clin Cancer Res  Published OnlineFirst October 14, 2010.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-10-0151

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2010/12/08/1078-0432.CCR-10-0151.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.