Journal of Molecular and Genetic Medicine

Fernández-Martínez et al., J Mol Genet Med 2017, 11:4
DOI: 10.4172/1747-0862.1000295

Commentary Open Access

The Effect of *NOP16* Mutation in Chronic Lymphocytic Leukemia

Fernández-Martínez JL*, DeAndrés-Galiana EJ and Cernea A*

Group of Inverse Problems, Optimization and Machine Learning, Department of Mathematics, C/ Federico García Lorca 8, 33007 Oviedo, University of Oviedo, Oviedo, Asturias, Spain

Abstract

NOP16 was the third most important mutation (6.84%) in a reduced-size cohort of 117 patients with Chronic Lymphocytic Leukemia (CLL) whose most recurrent mutations were *NOTCH1* (9.4%) and *SF3B1* (8.55%). In this paper we analyzed the effect of the *NOP16* mutation in gene expression. The *NOP16* mutation was predicted with 100% accuracy using a small-scale signature formed by the 26 genes with the highest Fisher's Ratio. *SLC39A4* (*ZIP4*) and *WARS* are the most discriminatory genes of this mutation providing a predictive accuracy of 97.4%. The Fold Change analysis also confirmed a very important role of the light (*IGKV3D-11*, *IGKC* and *IGLJ3*) and heavy (*IGHG1*) chain immunoglobulins, *SOX11*, *CCND1* and *CHL1*. This analysis also highlights the importance of several mechanisms such as the ZIP4-related apoptosis, the *SOX11-CCND1* over expression relationship observed in mantle cell-lymphoma, and CHL1 down regulation and over expression of the midkine-neurite growth-promoting factor that enhances the angiogenic and proliferative activities of cancer cells in different types of solid cancers. Besides, the holdout stability analysis has shown the importance of Signaling Events of B Cell Receptor (BCR), P53 signaling, Infectious disease, and TGF-beta Receptor Signaling. The integration of the *NOP16* mutation with the *IgHV*, *NOTCH1* and *SF3B1* mutations, that were previously analyzed, confirmed that these mutations only share two high discriminatory genes: *IGHG1* and *RGS13*. These genes are involved in different mechanisms concerning signaling and the immunological system. This analysis opens novel working hypothesis for CLL treatment and prognosis.

Keywords: Chronic lymphocytic leukemia; Machine learning; CLL mutations

Introduction

B-cell Chronic Lymphocytic Leukemia (CLL) is a heterogeneous disease from the molecular point of view, and the most common adult Leukemia in western countries. It is characterized by the accumulation of malignant B-cells in blood and lymphoid organs [1]. The diagnosis of CLL in clinical practice is based on the demonstration of an abnormal population of B lymphocytes in the blood, bone marrow, or tissues that display an unusual but characteristic pattern of molecules on the cell surface (*CD5* and *CD23* clusters of differentiation). Clinical staging systems to determine the extent of the disease, primarily based on a low platelet or red cell count, are currently used [2,3].

DNA analyses have served to distinguish two major types of CLL with different survival times [4]. This distinction is based on the maturity of the lymphocytes as discerned by the immunoglobulin variable-region heavy chain (*IgHV*) gene mutation status. High-risk patients have an immature cell pattern with few mutations in the DNA in the *IgHV* gene region, whereas low risk patients show considerable mutations of the DNA in the antibody gene region indicating mature lymphocytes [5]. Since the determination of the *IgHV* mutation status is very labor-intensive, other relevant markers have been investigated to better understand its progression. That way, *ZAP-70* and cell membrane expression of *CD38* became popular biomarkers of the *IgHV* mutational status [6,7].

Gene expression profiles were also used to understand the genesis and progression of CLL [8-11]. As result of these analyses different genetic signatures with high differential expression have been proposed, mainly to understand the *IgHV* mutational status. Whole genome sequencing also helped to gain more insights about CLL [12-19]. Using these techniques, four main genetic aberrations were recognized in CLL cells having a major impact on disease behavior, and are independent prognostic factors of the *IgHV* mutational status [12]. Recently, different recurrent mutations were identified in CLL patients [13]. *NOTCH1* and *SF3B1* turned to be the most frequently mutated genes that predicted poor prognosis. Disease progression has also been associated with a number of genetic alterations that include cytogenetic abnormalities and specific gene mutations [14-18]. Besides, aberrant methylation has been described for genes that are specifically deregulated in CLL [19].

Due to their low percentage of occurrence (no higher than 10%) in comparison to their mutational *IgHV* status (around 60/40%), these studies have shown that any of the individual mutations in CLL patients cannot explain alone the difference in prognosis observed in the CLL patients. Therefore, the main CLL mutations should share some common mechanisms that are responsible for CLL development and bad prognosis. A retrospective analysis to understand the effect the main mutations (*IgHV, NOTCH1* and *SF3B1*) in gene expression and the existing relationships between them have been recently performed using the same dataset [20].

Following the same methodology, we analyze the effect of the *NOP16* mutation in gene expression, highlighting the importance of different mechanisms involved in CLL, such as the ZIP4-related apoptosis, the *SOX11-CCND1* over expression relationship also observed in mantle cell-lymphoma, and the CHL1 down regulation and over expression of the midkine-neurite growth-promoting factor observed in different types of solid cancers. Besides, the holdout stability analysis has shown the importance of Signaling Events of B Cell Receptor (BCR), P53 signaling, Infectious disease, and TGF-beta Receptor Signaling.

***Corresponding author:** Dr. Juan Luis Fernández-Martínez, Group of Inverse Problems, Optimization and Machine Learning, Department of Mathematics, C/ Federico García Lorca 8, 33007 Oviedo, University of Oviedo, Oviedo, Asturias, Spain, Tel: 0034 985 103 199; E-mail: jlfm@uniovi.es

Received September 25, 2017; **Accepted** October 06, 2017; **Published** October 09, 2017

Citation: Fernández-Martínez JL, DeAndrés-Galiana EJ, Cernea A (2017) The Effect of *NOP16* Mutation in Chronic Lymphocytic Leukemia. J Mol Genet Med 11: 295 doi:10.4172/1747-0862.1000295

The *NOP16* mutation has not been studied before in CLL or in other hematologic malignancies. The current literature concerning *NOP16* expression is largely focused in breast cancer, since the expression of this gene induced by estrogens and Myc protein is a marker of poor patient survival [21].

Our expectation is that the knowledge issued from the analysis of the effect of the *NOP16* mutation in this reduced cohort where expression data and mutational status of the patients are available, will contribute in the near future to improve CLL treatment and prognosis, with the investigation of novel hypotheses and target therapies issued from this analysis. Particularly, this analysis has confirmed our previous findings concerning the effect of these mutations in the immunological system. We believe that a whole integration of the main mutations and genomic aberrations observed in CLL is needed to find a medical solution to this heterogeneous disease, and constitutes the subject of future research.

Materials and Methods

Data description

The microarray data includes a cohort of 117 patients and 48807 probes [21] and it is publically available at the European Bioinformatics Institute (EGAS00001000374). In this cohort only 8 patients presented the *NOP16* mutation. The exome sequencing data is described by Quesada et al. [15]. These authors identified 1246 mutations resulting in protein coding changes. Six genes appeared to be most frequently mutated in our restricted cohort (>5%): *NOTCH1, SF3B1, NOP16, CHD2, ATM* and *LRP1B.* This second data set we only use the *NOP16* mutational status of the different samples. Obviously, these statistics concerning these mutations might be different in larger cohorts, but independently of its frequency which is unknown, these mutations have been observed in CLL patients and constitute a perfect example to understand their effect in gene expression.

Concerning the degree of mutational overlap between patients, it can be said that most of the patients that have the *NOP16* mutation (7) have the *IgHV* mutated; none of these patients show the *NOTCH1* mutation and only one patient shows the *SF3B1* mutations. *NOP16* seems to be an independent factor for CLL development from *NOTCH1* and *SF3B1* mutations. Previous work has shown that these mutations (*NOTCH1* and *SF3B1*) are independent and jointly affecting the immunologic system (*IL-4*-mediated signaling events pathway) [20].

Methodology

The methodology used in this paper relates the *NOP16* mutation to the change in expression of the different genes. The phenotype consists in this case in the presence or absence of the *NOP16* mutation in the different samples. The aim is not to predict the *NOP16* mutation via gene expression, but understanding the mechanisms of action of the *NOP16* mutation and its impact in the expression of other genes in the transcriptome.

The methodology tries to determine the shortest lists of most discriminatory genes that predict the *NOP16* mutation and is described by Fernández-Martínez et al. [20] and De Andrés-Galiana et al. [22-24]. This classification problem is naturally unbalanced due to the low number of patients that show the *NOP16* mutation, and the classifier has to take this feature into account. In the case where the predictive accuracies of the small-scale signatures found are higher than those provided by the corresponding majority class classifiers, we could affirm that the classifier has really learnt the set of discriminatory genes for the *NOP16* mutation.

We have used the Fisher's Ratio and Fold Change to rank the genes according to their discriminatory power [20,22-24]. These gene-ranking methods and particularly Fisher's ratio turned to be very robust against different kind of noise [24]. Genes with the highest discriminatory power as described by these methods are expected to be involved in the genesis of the CLL. Finally, using a distance-based classifier we determine the shortest list of genes (small-scale signature) providing the highest predictive accuracy using a Leave One Out Cross Validation experiment (LOOCV). Besides, their stability has been confirmed using different random holds-outs (75% for learning and 25% for validation). This analysis has also provided the list of most-frequently sampled genes that are also used to explore the genetic mechanisms involved.

The correlation networks between the set of highly discriminatory genes are determined via the Pearson correlation coefficient [25]. This correlation network is a measure of the mutual dependence between gene expressions, and served to analyze inter-relationships between genes, which impacted both expression, and function. Finally, the pathways ontology is performed using Gene-Analytics™ to cover the altered and disease pathways [26]. This software uses the main ontological databases (Biosystems, Reactome, Qiagen, Kegg, Cell Signaling Technology and R&D Systems). Figure 1 shows the flowchart of the methodology that is followed in this paper.

Results

NOP16 is the third mutation by percentage of occurrence (6.84%) in our cohort. Other authors have identified *POT1* as the third most mutated gene using a more restricted dataset [18]. The most important feature using the present cohort is having disposal the gene expressions and the *NOP16* mutational status of these 117 patients, independently of its true statistical percentage in CLL. Besides, *NOP16* (*NOP16* nucleolar protein) is an interesting target, since it is transcriptionally regulated by c-Myc, a gene that plays an important role in cell cycle progression, apoptosis and cellular transformation. *NOP16* is up regulated in breast cancer and it is associated to poor patient survival [27].

Gene ranking

Table 1 shows the most discriminatory genes of the *NOP16* mutation provided by the Fisher's ratio. The highest LOOCV predictive

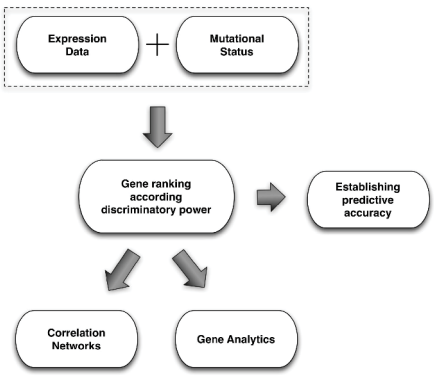

Figure 1: Flow diagram of the analytical procedure followed to perform this analysis. The dataset consists in the expression data and the mutational status of the patients. The genes are ranked according to their discriminatory power measured by their Fisher's ratio and absolute fold change values. Their predictive accuracy is established through leave-one-out-cross validation (LOOCV) and different hold-out analysis. Finally, the Correlation Networks and the Pathway Analysis using Gene Analytics are established and serve to find the biological processes impacted by the NOP16 mutation and how the most discriminatory genes are related.

accuracy reached was 100%, using the 26 most discriminatory genes. Interestingly, the predictive accuracy obtained with only the three first genes of this list (*SLC39A4, BRWD1* and *WARS*) is very high (94.9%). Table 2 shows the list of 32 most differentially expressed genes ordered by decreasing absolute FC (under or over expressed). This list provides a maximum accuracy of 94% using different list of genes containing the 32 most differentially expressed genes. *IGKV3D-11* (immunoglobulin kappa), transcription factor *SOX11* (Sex Determining Region Y-Box 11), *IGKC* (immunoglobulin kappa constant), *CCND1* (Cyclin D1), *CHL1* (cell adhesion molecule L1) and several probes of *IGLJ3* (Immunoglobulin Lambda Joining 3) appear in the list.

Correlation networks

Figure 2 shows the Pearson Correlation (PC) network of the most discriminatory genes (defined by FR) of *NOP16* mutational status. Figure 3 shows the same network constructed with the list of most differentially expressed genes provided by Fold Change. In both cases we built the minimum spanning tree using the maximum absolute value of the Pearson coefficient between the most discriminatory genes beginning by the header gene (the most discriminatory according to FR or to the absolute fold change). These correlation trees serve to understand the relationships among the discriminatory genes to regulate gene expression.

Pathways analysis

Table 3 shows the top canonical pathways using an expanded list of most discriminatory genes with Fisher's ratio greater than 0.6. Table 4 shows the top canonical pathways using an expanded list of most differentially expressed genes with absolute fold change value greater than 0.6. In both cases we try to use expanded lists with enough genes to improve the pathway matching in ontological databases.

Hold-outs experiments and PCA analysis

Figure 4 shows the cumulative probability distribution function (CDF) of the predictive accuracy for the small–scale signatures for

Genes	μ_1	σ_1	μ_2	σ_2	FR	Acc
SLC39A4	41	13	48	13	1.7	88.9
BRWD1	115	32	125	60	1.5	92.3
WARS	111	104	71	42	1.4	94.9
DSE	25	2	25	4	1.3	95.7
NONO	2265	257	2437	232	1.3	95.7
CORIN	15	1	16	1	1.3	95.7
DCX	16	2	15	1	1.3	95.7
OR51F1	16	1	16	2	1.2	96.6
SLC1A7	18	2	18	3	1.2	97.4
KLHL8	379	104	282	69	1.2	97.4
SIRT6	19	2	19	2	1.2	98.3
C9orf57	15	2	15	2	1.2	96.6
SNORA16B	20	4	21	5	1.1	97.4
COMMD9	81	22	86	18	1.1	98.3
UNC5B	24	5	25	4	1.1	98.3
OR1J4	16	1	16	2	1.1	99.1
TCOF1	155	38	158	40	1.0	98.3
ABHD2	24	6	22	4	1.0	97.4
GNA14	13	1	14	1	1.0	99.1
EDN3	17	1	16	1	1.0	99.1
SEMA6A	15	2	15	2	1.0	100.0

Table 1: *NOP16* mutational status prediction. List of the 26 most discriminatory genes ordered by decreasing Fisher's ratio. μ_1 and σ_1 refer respectively to the mean expression and standard deviation in class 1, (mutated NOP16), and μ_2 and σ_2 for the unmutated group. FR (log) stands for the logarithmic Fisher's ratio, and Acc is the LOOCV predictive accuracy.

Genes	μ_1	σ_1	μ_2	σ_2	*Fc*	Acc
IGKV3D-11	32	51	890	3367	-4.8	84.6
SOX11	448	1224	16	5	4.8	94.0
IGKC	24	22	613	2432	-4.7	94.0
CCND1	570	1506	35	16	4.0	94.0
CHL1	208	545	17	24	3.6	94.0
IGLJ3	1741	4866	156	1420	3.5	94.0
IGLJ3	1796	5023	168	1548	3.4	94.0
CCND1	192	484	20	5	3.3	94.0
PXDNL	18	2	164	590	-3.2	94.0
IGLJ3	1466	4096	165	1529	3.2	94.0
MDK	191	478	22	14	3.1	94.0
IGHG1	72	64	588	1461	-3.0	94.0
RASSF6	17	5	140	348	-3.0	94.0
SOX11	131	326	17	3	3.0	94.0
MDK	157	378	23	7	2.8	94.0
GTSF1	20	3	136	248	-2.8	94.0
LOC150568	102	242	16	2	2.7	94.0
CCND1	116	273	19	3	2.6	94.0
ITM2C	351	838	57	87	2.6	94.0
TUBB2B	133	308	22	5	2.6	94.0
MDK	146	342	24	7	2.6	94.0
MS4A7	315	675	54	76	2.5	94.0
CCND1	138	318	24	4	2.5	94.0
MAP1B	93	217	16	2	2.5	94.0
C5orf13	165	372	32	23	2.4	94.0
CSGALNACT1	550	1499	108	596	2.3	94.0
IFI44L	260	618	54	77	2.3	94.0
IFI44L	108	248	23	23	2.2	94.0
RGS13	551	715	119	406	2.2	94.0
MSR1	72	163	16	16	2.1	94.0
MS4A7	149	274	34	32	2.1	94.0
SEPT10	25	30	108	197	-2.1	94.0

Table 2: NOP16 mutational status prediction. List of the 32 most discriminatory genes, ordered by decreasing absolute Fold Change. The most differentially expressed genes are the first 4 genes. μ1 and σ1 refer respectively to the mean expression and standard deviation in class 1 (mutated *NOP16*), while μ2 and σ2 do for the unmutated group. *Fc* stands for the Fold Change. Acc is the LOOCV predictive accuracy. The shortest list with the maximum accuracy (94%) is obtained with the first 2 most differentially expressed genes (*IGKV3D-11* and *SOX11*). Adding other less differentially expressed genes does not improve the predictive accuracy that remains equal to 94 %.

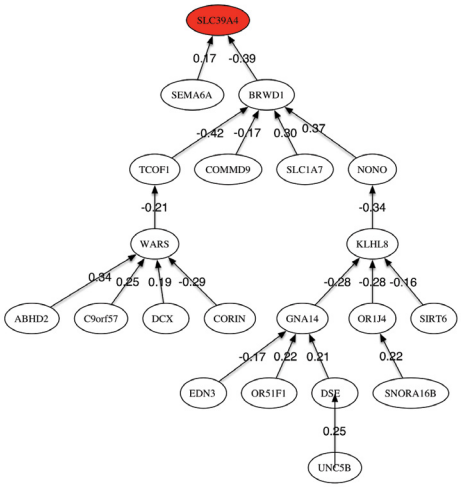

Figure 2: Correlation network of the most discriminatory genes (Fisher's ratio) for the NOP16 mutational status prediction using the Pearson correlation coefficient. The tree is restricted to the most discriminatory genes for the sake of visual clarity. The tree begins by the header gene (*SCLC39A* in this case) and the branches show the nodes (genes) that control expression and have the highest absolute Pearson correlation coefficient.

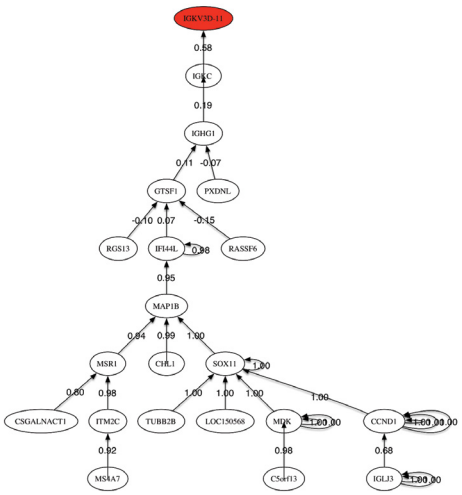

Figure 3: Correlation networks of the most discriminatory genes (Fold-change) for the *NOP16* mutational status prediction using the Pearson correlation coefficient. The tree is restricted to the most discriminatory genes for the sake of visual clarity.

Score	Name	# Genes	# Matched Genes
8.32	Apoptosis and Autophagy	178	7
7.95	Amyotrophic Lateral Sclerosis (ALS)	60	4
7.77	Oxidative Stress	31	3
7.47	Chondroitin Sulfate/dermatan Sulfate Metabolism	66	4
7.41	TGF-beta Receptor Signaling Activates SMADs	34	3
# Med score matches			

3a) Pathways

Score	Name	# Genes	# Matched Genes
10.22	Positive Regulation of Myoblast Fusion	17	3
10.17	Regulation of Renal Sodium Excretion	4	2
9.32	Neuromuscular Process Controlling Balance	46	4
9.02	Peptide Cross-linking Via Chondroitin 4-sulfate Glycosaminoglycan	6	2
8.80	Regulation of Vasoconstriction	24	3
# Med score matches			

3b) Biological processes

Table 3: Top pathways and biological processes impacted by the *NOP16* mutation using Fisher's ratio lists of most discriminatory genes (with FR greater than 0.6) (213 genes).

Score	Name	# Genes	# Matched Genes
12.10	Role of Phospholipids in Phagocytosis	45	4
11.64	Immuno-regulatory Interactions Between A Lymphoid and A Non-Lymphoid Cell	139	6
11.11	Binding and Uptake of Ligands by Scavenger Receptors	54	4
10.97	Regulation of Actin Dynamics for Phagocytic Cup Formation	99	5
10.40	Immune Response Fc Epsilon RI Pathway	164	6
9.16	C-MYB Transcription Factor Network	78	4
# High score matches			
# Med score matches			

4a) Pathways

Score	Name	# Genes	# Matched Genes
13.57	Negative Regulation of Long-term Synaptic Potentiation	2	2
13.48	Phagocytosis, Engulfment	35	4
13.11	Leukocyte Chemotaxis	14	3
12.41	Positive Regulation of Nucleotide-binding Oligomerization Domain Containing 2 Signaling Pathway	3	2
11.78	Regulation of Immune Response	193	7
# High score matches			
# Med score matches			

4b) Biological processes

Table 4: Top Pathways (high and medium score matches) biological processes impacted by the NOP16 mutation using the list of most differentially expressed genes found by fold change analysis (130 differentially expressed genes).

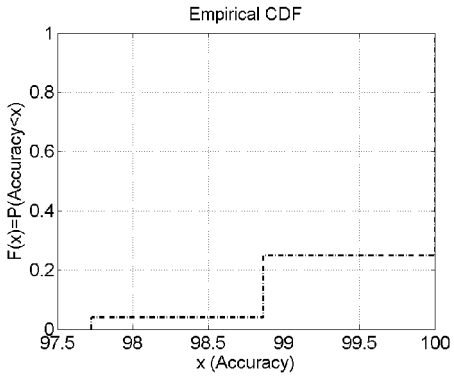

Figure 4: Hold-Out stability analysis. Cumulative probability functions for the predictive accuracy of the small-scale genetic signature of the *NOP16* mutation. The minimum scale signature is very stable since the accuracy varies from 97.5% to 100%.

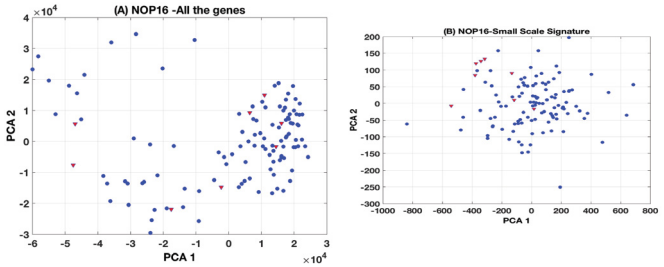

Figure 5: NOP16 mutational status. A) 2D PCA plot using all the genes. B) 2D PCA plot in the small-scale signature. This graph shows that restricting the expression data to the minimum scale signature improves the separability of two classes using unsupervised classification (Principal Component Analysis). PCA1 and PCA2 stand for the two first PCA coordinates of the different samples in the reduced PCA two-dimensional space.

the *NOP16*, obtained via 500 different random 75-25 hold-outs experiments. The CDF gives the probability that the predictive accuracy of the minimum-scale signature (found by FR) is smaller than a given cut-off (x value). The lower and upper percentiles and the median accuracy can be read directly in this plot that provides the stability in the predictive accuracy as a function of the variability in the training and validation sets. Table 5 shows the list of most frequently sampled genes in this hold out analysis, and Table 6 the pathway analysis made with this list of genes.

Besides, to show the discriminatory power of the small-scale signature found to explain the *NOP16* mutational status we have

Frequency	Gene-Name
0.69	*SLC39A4*
0.52	*CORIN*
0.48	*DSE*
0.46	*NONO*
0.46	*KLHL8*
0.45	*C9orf57*
0.45	*SNORA16B*
0.43	*BRWD1*
0.43	*UNC5B*
0.43	*DCX*
0.43	*WARS*
0.43	*SIRT6*
0.42	*SLC1A7*
0.41	*OR51F1*
0.40	*ABHD2*
0.38	*OR1J4*
0.36	*EDN3*
0.35	*NRIP1*
0.34	*COMMD9*
0.32	*TCOF1*
0.32	*TMEM14B*
0.29	*SYTL5*
0.28	*LUZP4*
0.28	*MECR*
0.27	*APBA2*
0.27	*SEMA6A*
0.27	*SNRPB2*
0.26	*TMEM14B*
0.25	*HES5*
0.25	*GNA14*
0.25	*ITGB3BP*
0.25	*KIAA0907*

Table 5: Holdout simulations and frequency analysis of the most frequently sampled discriminatory genes (with sampling frequency higher than 0.25) for the NOP16 mutation obtained after 500 random simulations (32 first genes).

Score	Name	# Genes	# Matched Genes
10.94	Downstream Signaling Events of B Cell Receptor (BCR)	390	11
10.27	P53 Signaling	128	6
10.16	Infectious Disease	773	16
8.66	P53 Pathway (Pathway Interaction Database)	66	4
8.46	TGF-beta Receptor Signaling Activates SMADs	33	3
# High score matches # Med score matches			

6a) Pathways

Score	Name	# Genes	# Matched Genes
13.02	Activation of Cysteine-type Endopeptidase Activity	11	3
12.82	Regulation of Apoptotic Process	278	10
12.82	Ovarian Follicle Rupture	2	2
11.83	Protein Ubiquitination	424	12
11.34	Apoptotic Process	644	15
# High score matches # Med score matches			

6b) Biological processes

Table 6: Top pathways (high ad medium score matches) and biological processes impacted by the NOP16 mutation using the most-frequently sampled genes (168 genes).

performed Principal Component Analysis (PCA) of this phenotype using all the genes available in the microarray (Figure 5A) and also the data set reduced to the small-scale signature (Figure 5B). It can be observed that reducing the dimension to the minimum scale signature improves greatly the separability of both classes.

Discussion

Fisher's analysis

With respect to the genes shown in Table 1, *SLC39A4* (also named *ZIP4*) is zinc transporter and it has been shown that *ZIP4* expression in Hepatocellular Carcinomas serves to repress apoptosis, enhancing cell cycle and increasing migration [28]. Also, Xu et al. have shown in hepatocellular carcinomas that suppression of *ZIP4* reduced cell migration and invasiveness, whereas *ZIP4* over expression caused increasing of cell migration and invasiveness [29]. Besides, over expression of *ZIP4* resulted in increased expression of pro-metastatic genes (*MMP-2*, *MMP-9*) and decreased expression of pro-apoptotic genes (caspase-3, caspase-9, Bax). *ZIP4* is also over expressed in pancreatic cancer, and regulates cell growth by activating the *IL-6/STAT3* pathway through Zinc Finger Transcription Factor *CREB* [30]. Similar conclusions were achieved by Cui et al. who analyzed the role of *ZIP4* in apoptosis resistance in pancreatic cancer, concluding that *ZIP4* confers resistance to zinc deficiency-induced apoptosis [31]. Also, *ZIP4* was found as a novel diagnostic and prognostic marker in human pancreatic cancer [32] and glioma [33]. Therefore, the result showed here gives an important role to *ZIP4* in CLL via the *NOP16* mutation.

BRWD1 encodes a member of the WD repeat protein family that are involved in a variety of cellular processes, including cell cycle progression, signal transduction, apoptosis, and gene regulation. Chromatin organization, Cytokine Signaling in Immune system, Innate Immune System and Interleukin-7 signaling are the pathways related to this gene.

WARS (Tryptophanyl-TRNA Synthetase) regulates different activation pathways that are associated with angiogenesis, cytoskeleton reorganization and shear stress-responsive gene expression. This gene is thought to be involved in the genesis of the Gulf War Syndrome (GWS).

The top canonical pathways show the importance of Apoptosis and Autophagy, Amyotrophic Lateral Sclerosis (ALS), Oxidative Stress, Chondroitin Sulfate/dermatan Sulfate Metabolism and TGF-beta receptor signaling. For instance, the disruption of TGF-beta signaling, either via mutational inactivation of components of the signaling pathway, or by down modulation of their expression, is known to play an important role in malignant transformation [34]. The main biological process involved is the Positive Regulation of Myoblast Fusion, highlighted in different diseases [35]. Other pathways involved are protein-protein interactions dopamine metabolism, *FOXA1*-transcription network or signaling by retinoic acid. The main related diseases are colorectal cancer, alzheimer, spastic paraplegia, neuroblastoma, prostate and breast cancer.

The correlation network (Figure 2) shows one main branch relating *SLC39A4* with *SEMA6A* and *BRWD1*. The main sub-tree develops under *BRWD1*. The correlation coefficients of all these genes are very low, therefore they should be considered as independent in explaining the phenotype. It has been shown that the Class 6 Semaphorin *SEMA6A* is induced by Interferon-γ in pathological situations [36].

Fold change analysis

The importance of immunoglobulin heavy and light chain gene

expression in CLL has been emphasized by Nakahashi et al. [37]. Besides, Lenormand et al. observed rearrangements of immunoglobulin light and heavy chain genes and correlation with phenotypic markers in B-cell CLL [38]. This analysis confirms the relation of the expression of the light chain immunoglobulins (kappa and lambda) to the *NOP16* mutation. The mean expression of *IGKV3D-11* and *IGKC* are 25-27 higher in the group with unmutated *NOP16*, and 8 times in the case of *IGHG1*. Conversely, the mean expression of *IGLJ3* in the mutated group is 10 times higher than in the case of unmutated *NOP16*.

SOX11 is a transcription factor involved in embryonic neurogenesis and tissue remodeling that has been found to be associated with mantle cell lymphoma, lymphoblastic neoplasms, Burkitt's lymphoma and B-cell lymphocytic proliferative diseases [39-41]. It has been shown that its expression correlates with *CCND1* and adverse prognosis [42]. In our cohort two different genetic probes of *CCND1* appears in 4th and 8th positions in Table 2. The protein encoded by this gene belongs to the cyclin family. *CCND1* protein has been shown to interact with tumor suppressor retinoblastoma protein (RB) and regulates the expression of this gene positively. Mutations, amplification and overexpression of this gene that alters cell cycle progression are frequently observed in a variety of tumors and may contribute to tumorigenesis [43-45]. The mean expressions of *SOX11* and *CCND1* in our cohort are 28 and 16 times higher in the group with mutated *NOP16*.

CHL1 differential expression has been highlighted for major type of cancers [46]: cancer profiling arrays revealed down-regulation/silencing of this gene in a majority of primary tumors and up-regulation associated with invasive/metastatic growth. The interpretation was that during the primary tumor growth *CHL1* could act as a putative tumor suppressor that is silenced to facilitate in situ tumor growth; and lately the re-expression of this gene on the edge of tumor mass might promote local invasive growth and enable further metastatic spread in certain types of cancers. In our cohort, *CHL1* is down-regulated in the group with no *NOP16* mutation.

MDK (Midkine-Neurite Growth-Promoting Factor 2) is a heparin-binding growth factor that is overexpressed in a number of solid cancers [47,48], and also in childhood B-precursor acute lymphoblastic leukemia [49]. *MDK* seems to enhance the angiogenic and proliferative activities of cancer cells. In some cases, these elevated levels of *MDK* also indicate poorer prognosis of the disease. In our cohort, three different probes of *MDK* appear in the list of 30 most differentially expressed genes, and their expression in the group of mutated *NOP16* is 6 to 9 times greater than in the group of unmutated *NOP16*.

The related top canonical pathways and biological processes involved show an important role of the immunological system, the role of phospholipids in phagocytosis, the regulation of Actin dynamics and C-MYB transcription factor network among others, whose importance has been highlighted in CLL [50]. These authors have shown that MYB is overexpressed in a subset of B-CLL patients. MYB physically associates with the promoter of miR-155 host gene, stimulating its transcription. Elevated levels of microRNA miR-155 represent a candidate pathogenic factor in chronic B-lymphocytic leukemia (B-CLL). The main biological processes are also related to Phagocytosis and the regulation of the immune system. Many of the discriminatory genes used in the pathway analysis are also related to Osteoarthritis, Prostate Cancer and Breast cancer. The correlation network (Figure 3) shows one main tree relating *IGKV3D-11* and IGKC. Nevertheless, the correlation coefficients of all these genes are very low, therefore they should be considered as independent in explaining the phenotype.

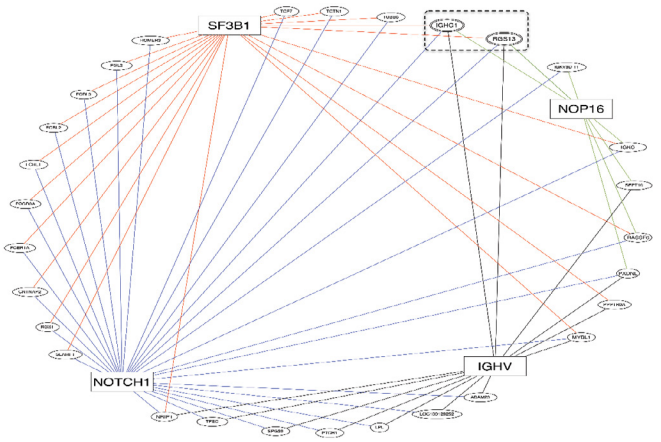

Figure 6: Cyclic diagram showing the intersection between the most discriminatory genes in each mutation: *NOTCH1, SF3B1, IGHV* and *NOP16*. The four mutations share only 2 genes (*IGHG1* and *RGS13*) within the intersection lists of discriminatory genes of each mutation.

Pathways analysis using the most frequently sampled genes

The pathway analysis using the most frequently sampled genes in the holdouts experiments complements the analysis that we did using the Fisher's ratio and the Fold Change. It is important to understand that the phenotype prediction problems admit different genetic signatures explaining equally well the phenotype, due to the reduced number of samples that are available compared to the number of genetic probes that are monitored [23]. Therefore, the pathway analysis is sensible to this high degree of under determinacy. For that reason, the analysis shown in Table 6, deduced with the most-frequently sampled genes would provide important clues about the genetic mechanisms that are involved. In this case the pathways are mainly related to Signaling Events of B Cell Receptor (BCR), P53 signaling, Infectious disease, and TGF-beta Receptor Signaling. The main biological processes involved are related to Cysteine Activation, Apoptotic processes and Protein Ubiquitination. Besides, many of these are also related to Colorectal Cancer, Alzheimer, Spastic Paraplegia, Neuroblastoma, Prostate cancer and Breast cancer.

Interaction of *NOP16* with the main mutations in CLL

Figure 6 shows the intersections between the lists of most discriminatory genes provided by the union of the Fisher's ratio and fold change lists.

The intersections are as follows:

1. The intersection between *NOP16* and *NOTCH1* contains 6 genes: *IGHG1, IGKC, IGKV3D-11, PXXDNL, RASF6* and *RSG13.*
2. The intersection between *NOP16* and *IgHV* contains 4 genes: *IGHG1, PXXDNL, SEPT10* and *RSG13.*
3. The intersection between *NOP16* and *SF3B1* contains 4 genes: *IGHG1, IGKC, RASF6* and *RSG13*

Therefore, the longest intersection of high discriminatory genes for *NOP16* is with the *NOTCH1* mutation. Besides, only two genes belong to the intersection of the 4 mutations: *IGHG1* and *RSG13.*

IGHG1 (Immunoglobulin Heavy Constant Gamma 1) has been already related to hypogammaglobulinemia and B-cell chronic lymphocytic leukemia. This gene also plays a major role in antigen

binding. *RGS13* (Regulator of G-protein signaling 13) is a protein-coding gene that is a member of the regulator of G protein signaling (*RGS*) family. Down-regulation of *RGS13* has been observed in mantle cell lymphoma [51]. In the present case *RGS13* is upregulated in the group with mutated *NOP16*. *RGS13* over expression inhibited *CXCL12*-evoked Ca (2+) mobilization, Akt phosphorylation and chemotaxis [52]. Also, it has been also shown that p53 negatively regulates *RGS13* protein expression in immune cells [53].

Other important genes appearing in these intersections are:

1. *SEPT10* (Septin) has been associated to CLL [54-57].
2. *PXDNL* (Peroxidasin) is involved in peroxidase activity and heme binding. Peroxidases serve diverse biological functions including well-characterized activities in host defense and hormone biosynthesis.
3. *RASSF6* is tumor suppressor protein regulates apoptosis and cell cycle via *MDM2* and p53 proteins [58]. Members of the *RAS* family form the core of the Salvador-Warts-Hippo (SWH) pathway. *SEPT10*, *RASSF6* are down regulated in the group of samples with *NOP16* mutation.

In a previous research [20] we have shown that 4 genes (*IGHG1, MYBL1, NRIP1* and *RGS13*) belong to the intersection of *IgHV, NOTCH1* and *SF3B1* mutations and that IL-4-mediated signaling events pathway seems to be involved as a common mechanism for disease progression. This analysis also highlights the importance of *IGHG1* and *RGS13* in the disease progression. Also, while the *IgHV* and *SF3B1* top genetic networks are related to cellular growth and proliferation and hematological diseases, the *NOP16* and *NOTCH1* networks are related to cancer and immunological response, although cellular apoptosis mechanisms are also important.

The main pathways described by the Fisher's ratio genes are more related to apoptosis processes and oxidative stress while the differentially expressed genes found by fold-change describe processes related to the regulation of the immunological system. Finally, the pathways found using the holdout simulations point to signaling events of B-cell receptor and to the P53 pathway well known in cancer. In conclusion these methods provide different mechanisms that might be involved in the disease progression.

Conclusion

Although recent analysis of non-coding recurrent mutations in CLL [59] have shown that the most frequently mutated genes in CLL are *NOTCH1* (12.6%), ATM (11%), *SF3B1* (8.6%), *BIRC3* (8.8%), *CHD2* (6%), TP53 (5.3%), *MYD88* (4%), and the *NOP16* mutation did not appear in this list, its presence in this reduced cohort served us to understand its effect in gene expression in CLL patients, concluding that *NOP16* mutation in CLL affects the expression of different genetic networks that are well-known in different types of solids cancers. Besides, the conclusions that were obtained confirm previous findings concerning the effect of the main mutations in gene expression [20]. It is important to understand that mutations are not only important for their frequency of occurrence, but also for their effect in the disease progression, since none of these mutations can explain alone the difference in prognosis observed in the CLL patients.

The main aim of this paper is to provide novel working hypothesis for CLL treatment based on the retrospective analysis of publically available data.

Acknowledgements

The authors declare that there are no conflicts of interest. They have also read and accepted the journal's authorship agreement. We would like to thank our colleagues from the Biochemistry Department (University of Oviedo, Spain) for having introduced us to this interesting problem and providing us with this dataset (gene expressions and landscape of mutations).

References

1. Rodríguez-Vicente AE, Diaz MG, Hernandez-Rivas JM (2013) Chronic lymphocytic leukemia: A clinical and molecular heterogenous disease. Cancer Genet 206: 49–62.
2. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, et al. (1975) Clinical staging of chronic lymphocytic leukemia. Blood 46: 219–234.
3. Binet J, Auquier A, Dighiero G, Chastang C, Piguet H, et al. (1981) A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer 48: 198–206.
4. Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, et al. (2001) Relation of gene expression phenotype to immunoglobulin mutation genotype in b cell chronic lymphocytic leukemia. J Exp Med 194: 1639– 1647.
5. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK (1999) Unmutated IG v(h) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94: 1848–1854.
6. Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, et al. (2003) Zap-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. New England Journal of Medicine 348: 1764–1775.
7. Cruse JM, Lewis RE, Webb RN, Sanders CM, Suggs JL (2007) Zap-70 and cd38 as predictors of IgHV mutation in cll. Exp Mol Pathol 83: 459–461.
8. Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, et al. (2001) Gene expression profiling of b cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory b cells. J Exp Med 94: 1625–1638.
9. Ferrer A, Ollila J, Tobin G, Nagy B, Thunberg U, et al. (2004) Different gene expression in immunoglobulin-mutated and immunoglobulin-unmutated forms of chronic lymphocytic leukemia. Cancer Genet Cyto- genet 153: 69–72.
10. Haslinger C, Schweifer N, Stilgenbauer S, Dohner H, Lichter P, et al. (2004) Microarray gene expression profiling of b-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and vh mutation status. J Clin Oncol 22: 3937–3949.
11. Vasconcelos Y, De Vos J, Vallat L, Reme T, Lalanne AI, et al. (2005) Gene expression profiling of chronic lymphocytic leukemia can discriminate cases with stable disease and mutated ig genes from those with progressive disease and unmutated ig genes. Leukemia 19: 2002–2005.
12. Döhner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, et al. (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. N Eng J Med 343: 1910–1916.
13. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, et al. (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukemia. Nature 475: 101–115.
14. Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabanian H, et al. (2011) Analysis of the chronic lymphocytic leukemia coding genome: role of notch1 mutational activation. J Exp Med 208: 1389–1401.
15. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, et al. (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44: 47–52.
16. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, et al. (2011) Sf3b1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 365: 2497–2506.
17. Oscier DG, Rose-Zerilli MJJ, Winkelmann N, Gonzalez De Castro D, Gomez B (2013) The clinical significance of notch1 and sf3b1 mutations in the uk lrf cll4 trial. Blood 121: 468–475.
18. Ramsay AJ, Quesada V, Foronda M, Conde L, Martinez-Trillos A, et al. (2013) Pot1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. Nat Genet 45: 526–530.
19. Rush LJ, Raval A, Funchain P, Johnson AJ, Smith L, et al. (2004) Epigenetic profiling in chronic lymphocytic leukemia reveals novel methylation targets. Cancer Res 64: 2424–2433.

20. Fernández-Martínez JL, De Andrés-Galiana EJ, Sonis S (2017) Genomic data integration in chronic lymphocytic leukemia. J Gene Med.

21. Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, et al. (2014) Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. Genome Res 24: 212–226.

22. Saligan L, Fernández-Martínez JL, De Andrés-Galiana EJ, Sonis S (2014) Supervised classification by filter methods and recursive feature elimination predicts risk of radiotherapy-related fatigue in patients with prostate cancer. Canc Inform 13: 141–152.

23. DeAndrés-Galiana EJ, Fernández-Martínez JL, Sonis S (2016) Design of biomedical robots for phenotype prediction problems. J Comput Biol.

24. deAndrés-Galiana EJ, Fernández-Martínez JL, Sonis S (2016) Sensitivity analysis of gene ranking methods in phenotype prediction. J Biomed Inform.

25. Witten IH, Eiben F, Hall MA (2011) Data mining: Practical machine learning tools and techniques. (3rd edn). Morgan Kaufmann, Burlington, Massachusetts, USA.

26. Stelzer G1, Inger A, Olender T, Iny-Stein T, Dalah I, et al. (2009) GeneDecks: Paralog hunting and gene-set distillation with Gene Cards annotation. Omics 13: 477-487.

27. Butt AJ, Sergio CM, Inman CK, Anderson LR, McNeil CM, et al. (2008) The estrogen and c-myc target gene hspc111 is over-expressed in breast cancer and associated with poor patient outcome. Breast Cancer Res 10: R28.

28. Weaver BP, Zhang Y, Hiscox S, Guo GL, Apte U, et al. (2010) Zip4 (slc39a4) expression is activated in hepatocellular carcinomas and functions to repress apoptosis, enhance cell cycle and increase migration. PLoS One 5: 1-2.

29. Xu X, Guo HJ, Xie HY, Li J, Zhuang RZ, et al. (2014) Zip4, a novel determinant of tumor invasion in hepatocellular carcinoma, contributes to tumor recurrence after liver transplantation. Int J Biol Sci 10(3): 245–256.

30. Li M, Zhang Y, Liu Z, Bharadwaj U, Wang H, et al. (2007) Aberrant expression of zinc transporter zip4 (slc39a4) significantly contributes to human pancreatic cancer pathogenesis and progression. Proc Natl Acad Sci USA 104: 18636–18641.

31. Cui X, Zhang Y, Yang J, Sun X, Hagan JP, et al. (2014) Zip4 confers resistance to zinc deficiency-induced apoptosis in pancreatic cancer. Cell Cycle 13: 1180–1186.

32. Xu C, Wallace MB, Yang J, Jiang L, Zhai Q, et al. (2014) Zip4 is a novel diagnostic and prognostic marker in human pancreatic cancer: a systemic comparison between eus-fna and surgical specimens. Curr Mol Med 14: 309–315.

33. Lin Y, Chen Y, Wang Y, Yang J, Zhu VF, et al. (2013) Zip4 is a novel molecular marker for glioma. Neuro Oncol 15: 1008–1016.

34. Kim SJ, Letterio J (2003) Transforming growth factor-beta signaling in normal and malignant hematopoiesis. Leukemia 17: 1731–1737.

35. Rochlin K, Yu S, Roy S, Baylies MK (2010) Myoblast fusión: When it takes more to make one. Develop Biol 341: 66-83.

36. Gautier G, De Saint-Vis B, Sénéchal B (2006) The class 6 Semaphorin SEMA6A is induced by interferon-γ and defines an activation status of langerhans cells observed in pathological situations. Am J Pathol 168: 453-465.

37. Nakahashi H, Tsukamoto N, Hashimoto Y, Koiso H, Yokohama A, et al. (2009) Characterization of immunoglobulin heavy and light chain gene expression in chronic lymphocytic leukemia and related disorders. Cancer Sci 100: 671–677.

38. Lenormand B, Ghanem N, Tilly H, Boussemart C, Monconduit M, et al. (1991) Rearrangements of immunoglobulin light and heavy chain genes and correlation with phenotypic markers in b-cell chronic lymphocytic leukemia. Leukemia 5: 928–936.

39. Chen YH, Gao J, Fan G, Peterson LC (2010) Nuclear expression of sox11 is highly associated with mantle cell lymphoma but is independent of t (11;14) (q13;q32) in non-mantle cell b-cell neoplasms. Mod Pathol 23: 105–112.

40. Dictor M, Ek S, Sundberg M, Warenholt J, Gyorgy C, et al. (2009) Strong lymphoid nuclear expression of sox11 transcription factor defines lymphoblastic neoplasms, mantle cell lymphoma and Burkitt's lymphoma. Haematologica 94: 1563–1568.

41. Cao X, Fan L, Fang C, Zhu DX, Dong HJ, et al. (2012) The ex- pression of sox11, cyclin d1, cyclin d2, and cyclin d3 in b-cell lymphocytic proliferative diseases. Med Oncol 29: 1190–1196.

42. Meggendorfer M, Kern W, Haferlach C, Haferlach T, Schnittger S (2013) Sox11 overexpression is a specific marker for mantle cell lymphoma and correlates with t (11;14) translocation, ccnd1 expression and an adverse prognosis. Leukemia 27: 2388–2391.

43. Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL (2011) Cyclin d as a therapeutic target in cancer. Nat Rev Cancer 11: 558–572.

44. Shan J, Zhao W, Gu W (2009) Suppression of cancer cell growth by promoting cyclin d1 degradation. Mol Cell 36: 469–476.

45. Matsuura I, Denissova NG, Wang G, He D, Long J, et al. (2004) Cyclin-dependent kinases regulate the antiproliferative function of smads. Nature 430: 226–231.

46. Senchenko VN, Krasnov GS, Dmitriev AA, Kudryavtseva AV, Anedchenko EA, et al. (2011) Differential expression of chl1 gene during development of major human cancers. PLoS One 6: e15612.

47. Ikematsu S, Yano A, Aridome K, Kikuchi M, Kumai H, et al. (2000) Serum midkine levels are increased in patients with various types of carcinomas. Br J Cancer 83: 701–716.

48. Kato M, Maeta H, Kato S, Shinozawa T, Terada T (2000) Immunohistochemical and in situ hybridization analyses of midkine expression in thyroid papillary carcinoma. Mod Pathol 13: 1060–1065.

49. Hidaka H, Yagasaki H, Takahashi Y, Hama A, Nishio N, et al. (2007) Increased midkine gene expression in childhood b-precursor acute lymphoblastic leukemia. Leuk Res 31: 1045–1051.

50. Curik N, Burda P, Basova P, Kulvait V, Pospisil V, et al. (2011) MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. Blood 117: 3816–3825.

51. Islam TC, Asplund AC, Lindvall JM, Nygren L, Liden J, et al. (2003) High level of cannabinoid receptor 1, absence of regulator of g protein signaling 13 and differential expression of cyclin d1 in mantle cell lymphoma. Leukemia 17: 1880–1890.

52. Bansal G, DiVietro JA, Kuehn HS, Rao S, Nocka KH, et al. (2008) Rgs13 controls g protein-coupled receptor-evoked responses of human mast cells. J Immunol 181: 7882–7890.

53. Iwaki S, Lu Y, Xie Z, Druey KM (2011) p53 negatively regulates rgs13 protein expression in immune cells. J Biol Chem 286: 22219–22226.

54. Oppezzo P, Vasconcelos Y, Settegrana C, Jeannel D, Vuillier F, et al. (2005) The lpl/adam29 expression ratio is a novel prognosis indicator in chronic lymphocytic leukemia. Blood 106: 650–657.

55. Maloum K, Settegrana C, Chapiro E, Cazin B, Lepretre S, et al. (2009) IGHV gene mutational status and lpl/adam29 gene expression as clinical outcome predictors in cll patients in remission following treatment with oral fludarabine plus cyclophosphamide. Ann Hematol 88: 1215– 1221.

56. Benedetti D, Bomben R, Dal-Bo M, Marconi D, Zucchetto A, et al. (2008) Are surrogates of IgHV gene mutational status useful in b-cell chronic lymphocytic leukemia? the example of septin-10. Leukemia 22: 224–226.

57. Travella A, Ripolles L, Aventin A, Rodriguez A, Bezares RF, et al. (2013) Structural alterations in chronic lymphocytic leukemia. Cytogenetic and fish analysis. Hematol Oncol 31: 79–87.

58. Iwasa H, Kudo T, Maimaiti S, Ikeda M, Maruyama J, et al. (2013) The RASSF6 tumor suppressor protein regulates apoptosis and the cell cycle via mdm2 protein and p53 protein. J Biol Chem 288: 30320–30329.

59. Puente XS (2015) Non-coding recurrent mutations in chronic lymphocytic leukaemia. Nature 526: 519–524.