8-2011

LDOC1, A NOVEL BIOMARKER OF PROGNOSIS IN CHRONIC LYMPHOCYTIC LEUKEMIA

Hatice Duzkale

LDOC1, A NOVEL BIOMARKER OF PROGNOSIS IN

CHRONIC LYMPHOCYTIC LEUKEMIA

by

Hatice Duzkale, M.D., M.P.H.

APPROVED:

_____
Lynne V. Abruzzo, M.D., Ph.D.
Supervisory Professor

_____
Michelle C. Barton, Ph.D.

_____
Gilbert J. Cote, Ph.D.

_____
Emil J Freireich, M.D., D. Sci. (Hon)

_____
Craig D. Logsdon, Ph.D.

_____
David J. McConkey, Ph.D.

APPROVED:

_____
George M. Stancel, Ph.D.
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

LDOC1, A NOVEL BIOMARKER OF PROGNOSIS IN

CHRONIC LYMPHOCYTIC LEUKEMIA

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Hatice Duzkale, M.D, M.P.H

Houston, Texas

August, 2011

DEDICATION*

Bana hep destek olan aileme, ozellikle de benden dualarini hic esirgemeyen

emektar anneme, sevgili kizkardesime ve biricik yegenime ithaf ediyorum.

* Dedicated to my family who has always supported me, especially to my mother, my sister and my nephew.

ACKNOWLEDGMENTS

LDOC1, A NOVEL BIOMARKER OF PROGNOSIS IN
CHRONIC LYMPHOCYTIC LEUKEMIA

Publication No._____

Hatice Duzkale, M.D., M.P.H

Supervisory Professor: Lynne V. Abruzzo, M.D., Ph.D.

In chronic lymphocytic leukemia (CLL), one of the best predictors of outcome is the somatic mutation status of the immunoglobulin heavy chain variable region (*IGHV*) genes. Patients whose CLL cells have unmutated *IGHV* genes have a median survival of 8 years; those with mutated *IGHV* genes have a median survival of 25 years. To identify new prognostic biomarkers and molecular targets for therapy in untreated CLL patients, we reanalyzed the raw data from four published gene expression profiling microarray studies. Of 88 candidate biomarkers associated with *IGHV* somatic mutation status, we identified *LDOC1* (Leucine Zipper, Down-regulated in Cancer 1), as one of the most significantly differentially expressed genes that distinguished mutated from unmutated CLL cases.

LDOC1 is a putative transcription factor of unknown function in B-cell development and CLL pathophysiology. Using a highly sensitive quantitative RT-PCR (QRT-PCR) assay, we confirmed that *LDOC1* mRNA was dramatically down-regulated in mutated compared to unmutated CLL cases. Expression of *LDOC1* mRNA was also

strongly associated with other markers of poor prognosis, including ZAP70 protein and cytogenetic abnormalities of poor prognosis (deletions of chromosomes 6q21, 11q23, and 17p13.1, and trisomy 12). CLL cases positive for *LDOC1* mRNA had significantly shorter overall survival than negative cases. Moreover, in a multivariate model, *LDOC1* mRNA expression predicted overall survival better than *IGHV* mutation status or ZAP70 protein, among the best markers of prognosis in CLL. We also discovered *LDOC1S*, a new *LDOC1* splice variant. Using isoform-specific QRT-PCR assays that we developed, we found that both isoforms were expressed in normal B cells (naïve > memory), unmutated CLL cells, and in B-cell non-Hodgkin lymphomas with unmutated *IGHV* genes.

To investigate pathways in which LDOC1 is involved, we knocked down LDOC1 in HeLa cells and performed global gene expression profiling. *GFI1* (Growth Factor-Independent 1) emerged as a significantly up-regulated gene in both HeLa cells and CLL cells that expressed high levels of *LDOC1*. GFI1 oncoprotein is implicated in hematopoietic stem cell maintenance, lymphocyte development, and lymphomagenesis.

Our findings indicate that *LDOC1* mRNA is an excellent biomarker of overall survival in CLL, and may contribute to B-cell differentiation and malignant transformation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1.　　　　INTRODUCTION

1.1　　Chronic Lymphocytic Leukemia (CLL)

Epidemiology

Chronic lymphocytic leukemia (CLL) is the most common leukemia in the Western hemisphere, and it accounts for one third of all leukemias in the United States. According to the American Cancer Society 2010 estimates (American Cancer Society, 2010;http://www.cancer.org/Cancer/LeukemiaChronicLymphocyticCLL/OverviewGuide/index), each year there are about 14,990 new cases and about 4,390 deaths due to CLL. The lifetime risk of developing CLL for an average person is about 1 in 200. CLL mainly affects adults of advanced age; the average age at diagnosis is 72 years. It is uncommon in individuals younger than 40 years, and very rare in children. It is also more common in men than women, for unknown reasons. CLL is more common in individuals of European than Asian ancestry. There are no proven environmental risk factors for CLL. Family history influences CLL risk, with first degree relatives of CLL patients having a 2 to 4-fold increased risk to develop CLL.

Diagnosis

In most patients (70-80%), the disease is identified incidentally during a routine blood test. The remaining patients present with enlarged lymph nodes or systemic symptoms, such as malaise, night sweats and weight loss. The diagnosis is established by evaluation of peripheral blood lymphocyte count combined with the characteristic immunophenotype (1). In advanced stages of the disease patients show signs of impaired function of normal bone marrow elements, such as anemia and susceptibility to infections.

1

Prognosis

The clinical course of CLL is variable and difficult to predict. Some patients live for many years without treatment, while others have a short survival even with treatment. As a result, a variety of clinical and laboratory parameters have been developed to predict prognosis. Clinical staging systems developed by Binet, et al. (1981) (2) and Rai et al. (1975) (3) remain among the most useful means to determine prognosis. The Rai staging system, used more commonly in North America than the Binet system, measures peripheral blood lymphocyte count, disease spread to lymph nodes, spleen or liver, and hemoglobin levels and platelet counts. Rai stages advance from Stage 0 (increased lymphocyte number only) to stage I (lymph node involvement), or stage II (enlargement of spleen or liver and/or lymph nodes), stage III (anemia), and stage IV (reduced platelet count). Physicians wait to initiate treatment until patients reach advanced stages and develop symptomatic disease. While these staging systems are useful to make therapeutic decisions, they are unable to predict long-term survival with high precision, especially for patients with early stage disease. A variety of serum markers have also been used to predict prognosis in early stage disease, and include β2 microglobulin, lactate dehydrogenase, soluble CD23, and serum thymidine kinase.

One of the best predictors of outcome is the somatic mutation status of the immunoglobulin heavy chain variable region (*IGHV*) genes. Patients whose CLL cells have unmutated *IGHV* genes, about 40% of patients, have a median survival of 8 years; patients whose CLL cells have mutated *IGHV* genes, about 60% of patients, have a median survival of 25 years (4). Although the somatic mutation status is highly associated with prognosis, it has been shown recently that the use of particular heavy chain variable region genes, such as VH3-21, is associated with a poor prognosis,

regardless of somatic mutation status (5, 6). Thus, the relationship between somatic mutation status and prognosis is not absolute.

Chromosomal abnormalities, predominantly gains and deletions (del), are strong independent predictors of prognosis in CLL. The most common abnormality is del(13)(q14.3), followed by del(11)(q22.3), trisomy 12, del(6)(6q21-q23), and del(17)(p13.1). In clinical practice these abnormalities are usually assessed using a panel of fluorescence in situ hybridization (FISH) probes (7-9). As the sole abnormality, del(13)(q14.3) is associated with a good prognosis. In contrast, del(6)(q21-q23), del(11)(q22.3), del(17)(p13.1), del(13)(q14.3) with other abnormalities, and trisomy 12, are associated with more rapid disease progression and inferior survival. The abnormalities del(17)(p13.1), the site of the *TP53* gene, and del(11)(q22.3), the site of the *ATM* gene, are the most important independent cytogenetic markers of poor prognosis. Deletion of (17)(p13.1) is associated with resistance to therapy with purine analogs, such as fludarabine, and short survival (10-12).

Gene expression profiling studies have demonstrated that the majority of unmutated CLL cases express *ZAP70* mRNA (13). Subsequently, others showed that expression of ZAP70 protein correlates with mutation status and clinical outcome (14, 15). Recent studies suggest that ZAP70 protein expression, usually measured by flow cytometry, may be a better predictor of time-to-treatment than somatic mutation status (16, 17). However, standardization of ZAP70 protein measurement has remained challenging for many clinical laboratories, which has limited its use as a routine diagnostic test.

A robust biomarker of prognosis that is easily standardized between laboratories would have a major clinical impact for CLL patients.

Treatment

Currently the most effective therapy for CLL is combination chemoimmunotherapy with fludarabine (a purine analog that inhibits repair of DNA damage), cyclophosphamide (a DNA damaging agent), and Rituximab (a monoclonal antibody directed against CD20, a protein on the surface of mature B cells), i.e., FCR (**F**ludarabine-**C**yclophosphamide-**R**ituximab). These agents act synergistically to enhance apoptosis of the CLL cells (18). The FCR regimen is the most successful combination therapy developed to date, and provides a 6 year survival rate of 77% (18). Despite this promising initial response, about one-half of patients treated with frontline FCR develop relapse within 3 years of treatment, which is often resistant to further FCR therapy. Thus, alternative specific and effective therapy options are needed to treat CLL patients.

1.2   Emergence of LDOC1 as a potential biomarker

To identify new prognostic biomarkers and molecular targets for therapy in untreated patients with CLL, we began this study by reanalyzing the raw data from four published gene expression profiling microarray studies (14, 19-21). Of 88 candidate biomarkers of *IGHV* somatic mutation status, we were able to confirm expression of 37 using a highly sensitive quantitative real-time polymerase chain reaction assay performed on microfluidics cards (MF-QRT-PCR) (22). Of these candidate biomarkers, the gene *LDOC1* (Leucine zipper Down-regulated in Cancer) was one of the most

4

significantly differentially expressed gene that distinguished mutated from unmutated CLL cases. ("Mutated" or "unmutated" CLL refers to patients whose CLL cells have somatically mutated or unmutated *IGHV* genes, respectively.) Since unmutated CLL is strongly associated with poor prognosis and *LDOC1* mRNA is highly upregulated in this group of patients, we hypothesized that LDOC1 might not only serve as a promising biomarker of prognosis, but might also contribute to CLL pathogenesis and serve as a novel candidate molecule for targeted therapies in cancers where it is expressed abundantly.

Schema of selection of LDOC1 as a candidate biomarker

Screening of CLL literature for microarray studies for genes distinguishing unmutated from mutated CLL

⬇

Evaluation of 88 genes in 49 untreated CLL patients (MF-QRT-PCR)

⬇

36 genes significant

⬇

LDOC1 most significant high in unmutated CLL, low in mutated CLL
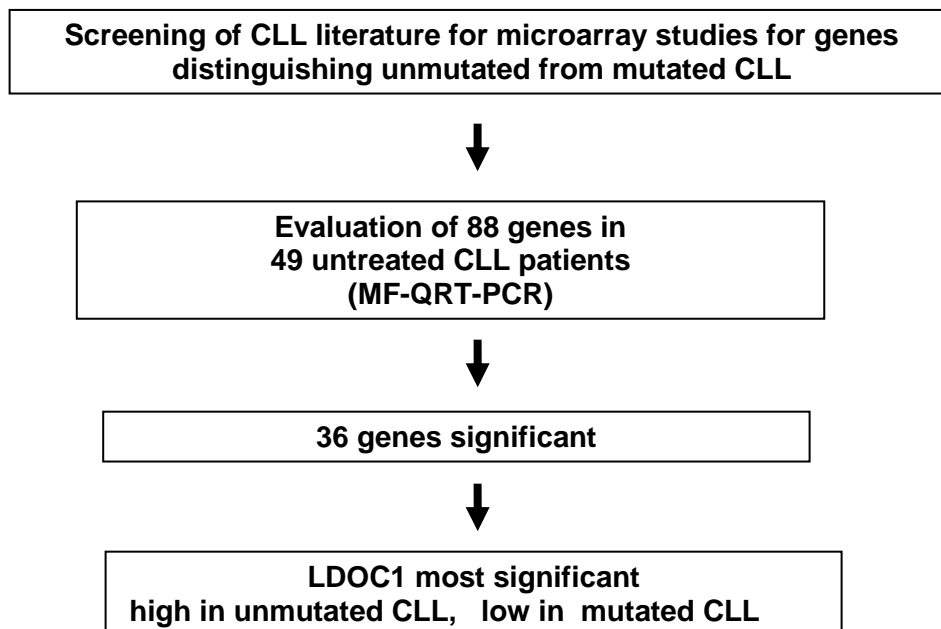
1.4 LDOC1

The *LDOC1* gene, located on chromosome Xq27, encodes a 17 kDa protein about which very little is known. A leucine zipper motif in the N-terminal region is followed by a short proline-rich region, which contains an SH3-binding consensus sequence, and then an acidic region in the C-terminus (23). Because leucine zipper and SH3-binding motifs mediate protein-protein interactions, LDOC1 protein may regulate transcription by homodimerization or heterodimerization with other transcription factors through its leucine zipper domain. *LDOC1* also may participate in cell signaling by providing a binding surface for signaling cascade proteins within its SH3 domain. Others have assessed *LDOC1* mRNA expression in a wide range of normal tissues and in carcinoma cell lines (23). *LDOC1* mRNA is expressed ubiquitously in normal tissues, although at relatively low levels in leukocytes, liver, and placenta compared with other tissues. In tumor cell lines, *LDOC1* mRNA is expressed in most breast cancer cell lines, but rarely in pancreatic or gastric carcinoma cell lines. Because *LDOC1* mRNA is expressed in many normal tissues, but is not expressed in most pancreatic and gastric carcinoma cell lines, Nagasaki and colleagues have hypothesized that *LDOC1* is a tumor suppressor gene. Ectopic *LDOC1* expression is reported to inhibit NF-kB activation in cell lines (24). Others have reported that *LDOC1* induces apoptosis in Jurkat lymphoma and K562 leukemia cell lines, but not in HeLa cervical carcinoma cells (25). Based on these findings, it has been suggested that *LDOC1* may have pro-apoptotic and anti-proliferative effects. The biologic functions of *LDOC1* in normal B-cell development and the pathophysiology of CLL are unknown.

1.5   Overview of Dissertation

LDOC1 is a novel gene and a putative transcription factor. It has emerged as one of the most significantly differentially expressed genes in CLL prognostic subtypes: *LDOC1* mRNA expression is dramatically reduced in CLL cases with somatically mutated *IGHV* genes (associated with a good prognosis) compared to CLL cases with somatically unmutated *IGHV* genes (associated with a poor prognosis) (22). Because robust biomarkers of prognosis in CLL are needed, we extended our study to determine the value of *LDOC1* mRNA expression as a novel clinical biomarker of prognosis. Because it is highly expressed in unmutated CLL cases, it also has the potential to serve as a candidate molecule for targeted therapies.

The studies presented in this dissertation had four goals. Our first goal was to determine if *LDOC1* mRNA expression could serve as a clinically useful biomarker of prognosis. For this, we evaluated *LDOC1* mRNA expression in a large cohort of CLL patients to assess its differential expression and correlation with overall survival. As part of this evaluation, we also correlated *LDOC1* mRNA expression with other known prognostic parameters in CLL including *IGVH* somatic mutation status, cytogenetic abnormalities, and ZAP70 expression. Our data suggest that *LDOC1* mRNA expression has prognostic significance in CLL (26). Our second goal was to investigate changes in *LDOC1* mRNA levels in normal peripheral blood B cells at different stages of development and differentiation, as well as in primary malignant B cells from a variety of different lymphoma subtypes. This might provide insights into the possible contributions of *LDOC1* in physiologic and pathologic changes in B cells. Our findings suggest that its dysregulation may contribute to the pathophysiology of CLL and other B-cell malignancies (26). It may also play a role in normal B cell development. Our third goal

was to gain insight into the biologic pathways in which LDOC1 is involved. To achieve this goal, we used siRNAs to reduce LDOC1 protein in HeLa cervical carcinoma cells, which abundantly express the protein compared to other solid tumors and hematologic malignancies that we have screened. We acquired the RNA and evaluated it for changes in 47,000 transcripts in LDOC1-reduced cells relative to the cells containing unaltered LDOC1 protein. We found that LDOC1 protein is involved in key signaling pathways in HeLa cells including cellular function and maintenance, cell cycle, cellular growth and differentiation, cell death, DNA replication and cancer. Our fourth and final goal was to validate key genes in these pathways in CLL patient samples using a highly sensitive QRT-PCR method. We looked for genes whose expression was concordant with respect to *LDOC1* expression in transfected HeLa cells and CLL samples, suggesting that they might be involved in overlapping biologic pathways in the different cell types. We identified the gene, *GFI1*, whose expression was decreased in HeLa cells in which LDOC1 protein had been knocked down, and also in mutated CLL cells, which express little or no *LDOC1*. GFI1 has been implicated in the development and function of hematopoietic stem cells, normal lymphoid development and differentiation, and lymphomagenesis.

In summary, we have shown that *LDOC1* mRNA expression is a promising novel biomarker of survival in untreated CLL patients. Correlative gene expression profiling studies in transfected HeLa cells and CLL samples suggest that LDOC1 might participate in a common pathway with GFI1, and contribute to CLL pathogenesis.

CHAPTER 2.   MATERIALS AND METHODS

2.1   Collection of patient and healthy donor samples

Peripheral blood samples were collected from 131 previously untreated CLL patients and six healthy volunteers at The University of Texas M.D. Anderson Cancer Center after informed consent was obtained. The study was approved by the Institutional Review Board and conducted according to the principles expressed in the Declaration of Helsinki. Normal CD19+ peripheral blood B cells (NBC) and CLL cells were purified by negative selection using immunomagnetic beads, and the samples processed as described previously (27). In addition, we enriched NBC from healthy donors for naïve or memory B cells using a CD27 antibody column, according to the manufacturer's instructions (Miltenyi Biotec, Inc., Auburn, CA). In peripheral blood, antigen-naïve B cells (CD27−) constitute about 60% of the B cells and antigen-experienced memory B cells (CD27+) about 40% (28). We confirmed the purity of all cell preparations by flow cytometry.

Primary lymphoma samples were obtained from either lymph node biopsy specimens from patients with follicular lymphoma (two patients) or from the peripheral blood of patients in leukemic phase of lymphomas (one patient with follicular lymphoma, four patients with mantle cell lymphoma, two patients with splenic marginal zone lymphoma, and one patient with marginal zone lymphoma). CD19+ B cells were purified as described above (27). We ensured, by flow cytometry, that each sample contained at least 95% CD19+ B cells.

9

2.2   Cell lines

The Epstein-Barr virus-negative Burkitt lymphoma cell line, GA-10, and the T-cell lymphoblastic lymphoma cell line, Jurkat, were maintained in RPMI medium (Mediatech, Inc., Manassas, VA); the breast carcinoma cell line, MCF-7, and the cervical carcinoma cell line, HeLa, were maintained in DMEM (Mediatech). Both media were supplemented with 10% fetal bovine serum (Sigma, St. Louis, MO) and 2 mM L-glutamine (Mediatech). We harvested the lymphoma cell lines during the exponential phase of growth; the confluence of the MCF-7 and HeLa cells at the time of harvest is indicated in the text.

2.3   Nucleic Acid Preparation

Total RNA was extracted using guanidine isothiocyanate/phenol-chloroform extraction with TRIzol reagent (Life Technologies, Inc., Gaithersburg, MD) according to the manufacturer's instructions. The quality of the RNA was assessed by agarose gel electrophoresis. For all subsequent PCR assays, total RNA was reverse transcribed using random hexamers and a First-Strand cDNA synthesis kit (GE Healthcare, Piscataway, NJ). We used a final concentration of 10 ng/µL of reverse-transcribed product. For microarray gene expression profiling experiments, total RNA was further purified with RNeasy MinElute Cleanup Kit (Qiagen, Valencia, CA). DNA was extracted using a Qiagen DNeasy kit (Qiagen Sciences, Maryland) according to the manufacturer's instructions. The quality of the RNA used for gene expression profiling microarray studies was assessed by Bioanalyzer analysis in the Genomics Core Facility and met their quality control standards.

2.4   Evaluation of the *IGHV* somatic mutation status

For our studies, we obtained the somatic mutation status of the *IGHV* genes for a total of 130 CLL and 10 primary B-cell lymphoma samples. We assessed the somatic mutation status of the *IGHV* genes in 66 CLL and 10 primary B-cell lymphoma samples as described previously, with minor modifications (27). Patient DNA sequences were aligned to the germline DNA sequences in VBASE II and the degree of *IGHV* somatic mutation was determined (29). For 54 patients, analysis of the somatic mutation status was performed in the clinical Molecular Diagnostics Laboratory at our institution and obtained by review of the medical records. For 10 patients, the somatic mutation status was performed by the Chronic Lymphocytic Leukemia Research Consortium laboratory, and the results were obtained from the CLL Database of the Protocol Data Management System (CLL-PDMS), which is maintained by our collaborator, Dr. Michael Keating. The *IGHV* somatic mutation status was designated as "unmutated" if there were fewer than 2%, or as "mutated" if there were 2% or more mutations compared to the germline sequences (30).

2.5   Assessment of *ZAP70* protein expression

Expression of ZAP70 protein was assessed by either immunohistochemistry or flow cytometry. Immunohistochemical staining was performed using routinely fixed and processed paraffin-embedded tissue sections of bone marrow core biopsy and/or clot specimens and a specific monoclonal antibody (Upstate Cell Signaling Systems, Lake Placid, NY), as described previously (17, 31). Immunohistochemical stains for ZAP70 protein were scored as either positive or negative by hematopathologists, and the results obtained from the patients' medical records. The flow cytometry assay for

ZAP70 protein expression was performed by the Chronic Lymphocytic Leukemia Research Consortium laboratory, as described previously (16). The results were obtained from the CLL Database of the Protocol Data Management System (CLL-PDMS), which is maintained by our collaborator, Dr. Michael Keating.

2.6 *LDOC1* mRNA expression by reverse-transcriptase polymerase chain reaction assay

We used two primer pairs to amplify either the entire coding region (primer pair AB) or the mRNA (primer pair AC) of the *LDOC1* gene in a reverse-transcriptase polymerase chain reaction (RT-PCR) assay. Primer pair AB, designed to amplify the entire *LDOC1* coding region, yields a product of 464 bp, as reported by others (23). We designed a second reverse primer, C (5'-AGCAGGTAACTGGAGCGCTA-3'), which binds within the 3' untranslated region (3' UTR). Primer pair AC was expected to yield a product of 649 bp. The cDNA (80 ng) was amplified in the presence of primers, reaction buffer, deoxynucleotide triphosphates (2.5 μM each), and HotStar Taq DNA polymerase. Following incubation at 95°C for 10 minutes, the cDNA was amplified for 35 cycles of at 95°C for 15 seconds, 55°C for 30 seconds, and 72°C for 45 seconds, followed by a final extension at 72°C for 7 min. The amplified products were separated by agarose gel electrophoresis, extracted from the gel, and purified. The sequence of the PCR products was determined directly using the forward and reverse PCR primers, and an ABI 3700 or 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). The sequences were aligned to the *LDOC1* reference sequence (GenBank RefSeq NM_012317) using NCBI SPLIGN algorithm and LaserGene v7.2 software (DNASTAR). The cDNA amount and integrity were ensured by amplifying the housekeeping gene,

beta-actin, using forward (5'-GATCATGTTTGAGACCTTCAAC-3') and reverse (5'-TCTTTGCGGATGTCCACGTC-3') primers (27). The RT-PCR conditions were identical to those used for the *LDOC1* RT-PCR assay; and the reactions were run simultaneously with the reactions amplifying *LDOC1*.

2.7 Total *LDOC1* mRNA expression determined by quantitative real-time polymerase chain reaction assays

We used two different quantitative real-time polymerase chain reaction assays (QRT-PCR) to detect total *LDOC1* mRNA expression: 1) a high-throughput assay using microfluidics cards (MF-QRT-PCR) and 2) a standard QRT-PCR assay. The *LDOC1* TaqMan probe and primer sets were identical in both assays. The primers and probe bind to 3' UTR sequences that are present both in wild-type *LDOC1* and its splice variant, *LDOC1S* (TaqMan Assay, Hs00273392_s1, Applied Biosystems). In the MF-QRT-PCR assay, custom microfluidics cards were printed with primers and probes corresponding to 88 candidate mRNA biomarkers of *IGHV* somatic mutation status, including *LDOC1* and *ZAP70*. We used five endogenous control genes (*18S rRNA*, *GAPD*, *PGK1*, *GUSB*, and *ECE-1*), and performed the MF-QRT-PCR assays, as described previously (22). The expression of each gene was assessed in duplicate.

In the standard QRT-PCR assay, the PCR reactions for *LDOC1* mRNA were carried out in 25 μL reaction volumes that contained 5 μL cDNA at a concentration of 10ng/μL. In addition, 1X TaqMan Universal PCR Master Mix without AmpErase UNG, unlabeled *LDOC1*-specific PCR primers, and a 6-carboxy fluorescein (FAM)–labeled TaqMan minor groove binder (MGB) probe were added. In all experiments, we amplified

of 18S ribosomal RNA (rRNA) as an internal control to normalize the *LDOC1* values. The probe for 18S rRNA is labeled with the VIC reporter dye. The PCR reaction conditions were as follows. After incubation at 95°C for 10 minutes, the cDNA was amplified for 45 cycles of denaturation at 95°C for 15 seconds and combined annealing/extension at 60°C for 1 minute. Each sample was analyzed in triplicate in a PRISM 7000 Sequence Detector (Applied Biosystems). We used the 7500 Fast System version 1.4.0 software (Applied Biosystems) to analyze the fluorescence emission data following QRT-PCR. The threshold cycle (Ct) values of each sample were exported to Microsoft Excel for further analysis. The Ct value represents the cycle number at which fluorescence originating from each sample passes a predetermined single threshold. The ΔCt for *LDOC1* mRNA was obtained by subtracting the Ct value of 18S rRNA from the Ct value of *LDOC1* mRNA for each sample. The *LDOC1* mRNA expression levels in test samples are presented as relative quantities (RQ), computed as: $2^{-\Delta\Delta Ct} = 2^{-(\Delta Ct\ test\ -\ \Delta Ct\ calibrator)}$, where the calibrator represents an equal mixture of cDNA obtained from GA-10 and Jurkat cells.

2.8 Expression of *LDOC1* mRNA isoforms by isoform-specific QRT-PCR assays

In order to assess the contribution of each of the *LDOC1* mRNA isoforms to the total *LDOC1* mRNA expression, we designed two specific TaqMan assays that distinguish between the isoforms.We used Primer Express software (Applied Biosystems) to search for and select the primer and probe combinations from a ranked list generated by the software. In the TaqMan assay that specifically recognizes the wild-type *LDOC1* sequences (Custom LDwt1, part number 4331348), the 5' primer anneals to sequences 5'–TGGTGCCCTACATCGAGATG-3', the 3' primer anneals to

sequences 5'-CGAGGAAGGCCCGGTAA-3', and the TaqMan probe anneals to sequences 5'–ATAGCCCCATCCTAGGTG-3'. In the TaqMan Assay that specifically detects the splice variant *LDOC1S* transcript, the TaqMan probe (Custom LDsv1, part number 4331348) targets the junction sequence located between nucleotides 183 to 233 and 718 to 785. The 5' primer anneals to sequences 5'-TTCCAAGCACTTCCGAGTGA-3', the 3' primer anneals to sequences 5'-ATGGAACAGCTGCGGCTG-3', and the TaqMan probe anneals to sequences 5'-CTATTCCTGGCGCAGCAG-3'. Assays were ordered on the Applied Biosystems Custom TaqMan Assay Design Tool (https://www5.appliedbiosystems.com/tools/cadt/).

2.9 Detection of genomic gains and losses by single nucleotide polymorphism (SNP) genotyping

Genomic DNA was extracted from purified CLL cells, as described above. Genotypic analysis was performed on DNA obtained from 100 CLL samples using the Illumina HumanHap610 chip, according to the manufacturer's instructions, in collaboration with Dr. Bogdan Czerniak, Department of Anatomic Pathology. Background subtraction and normalization were performed using the default settings in the Illumina BeadStudio software. Log R ratios (LRR), B allele frequencies (BAF), and genotype calls were exported from BeadStudio for analysis in the R statistical programming environment (version 2.8.1). Segments of constant copy number in the LRR data were identified by applying the circular binary segmentation (CBS) algorithm, as implemented in the DNAcopy package (version 1.16.0) (32). Segments with mean $LRR < -0.15$ and two bands in the BAF plot were called "deleted"; segments with mean $LRR > 0.15$ and four bands in the BAF plot were called "gained". The data were evaluated to detect common abnormalities associated with CLL, i.e., deletions of 6q21,

11q22.3, 13q14.3, and 17p13.1, and trisomy 12; we also evaluated the DNA for gains or deletions of *LDOC1*. The data were analyzed by our collaborator, Dr. Kevin Coombes, Department of Bioinformatics and Computational Biology. Following analysis, the genomic changes were visually inspected to ensure the accuracy of the calls made by the computer algorithm were correct.

2.10   Statistical analysis

Comparisons of *LDOC1* positive and negative patient groups were performed using two different tests (Table 1). For discrete variables (all parameters except age), we used the two-sided Fisher's exact test to compare the means of the two groups. For the continuous variable (age), we used an unpaired two-sample t-test assuming unequal variances to compare the mean. Time-to-event (survival) analysis was performed using Cox proportional hazards models. Significance was assessed using the log-rank (score) test. To assess multivariate models, we used a forward-backward stepwise algorithm to eliminate redundant factors and optimize the Akaike Information Criterion (AIC) (33). All computations were performed using either the survival package (version 2.35-8) in the R statistical programming environment (version 2.11.1) (performed by Dr. Kevin Coombes) or STATA Statistics/Data Analysis software version 10.0 (performed by Hatice Duzkale). Multivariate survival analysis and analysis of the 88 candidate biomarkers of prognosis by MF-QRT-PCR were performed in collaboration with Dr. Kevin Coombes.

2.11 LDOC1 protein knock-down in HeLa cells

HeLa cells were obtained from ATCC, and were at passage 2 since their receipt from the ATCC at the beginning of the experiments. The cells were thawed and passaged twice before seeding for the experiment. When they reached approximately 90% confluence they were harvested by trypsinization and counted; the cell viability, assessed by trypan blue dye exclusion, was greater than 90%. The cells were seeded at 1.5×10^5 cells per well in 2 mL complete DMEM medium (Mediatech), into 6-well plates.

Twenty four hours after seeding, the cultures were inspected using an inverted light microscope to ensure that the cells had reached 40-50% confluence and were evenly distributed. For transfection, the complete medium containing 10% FBS was replaced with DMEM medium containing 1% FBS, according to the manufacturer's instructions (Lipofectamine 2000, Invitrogen, Carlsbad, CA). Knock-down of *LDOC1* and *LDOC1S* mRNAs was performed using siRNA pools that contain four different siRNAs, all of which hybridize to the amino acid coding region (ON-TARGETplus SMARTpool siRNA pool; Thermo Scientific, Lafayette, CO). One of these siRNAs also recognizes the splice variant *LDOC1S*. For the negative controls, we used either a single non-targeting siRNA (ON-TARGETplus Non-targeting siRNA #1 for the time-course optimization experiments) or a pool of four non-targeting siRNAs (ON-TARGETplus Non-targeting Pool for the global gene expression profiling microarray experiments). The non-targeting siRNAs do not hybridize to any region in the human genome, and control for sequence-nonspecific silencing effects during siRNA transfection experiment. Water was used for the "mock" transfection control. The siRNAs were added to each culture at a final concentration of 100 nM. The cells were incubated with the siRNAs in

the presence of 1% FBS for 24 hours at 37°C with 5% CO_2. After this incubation, FBS was added to achieve a final concentration of 10% FBS. For the optimization experiments, the cells were continuously incubated with the siRNAs until they were harvested. For the gene expression profiling studies, the cells were incubated with the siRNAs for 56 hours, the medium replaced with complete medium containing 10% FBS, and then harvested at 93 hours post-transfection.

In the same experiment we also assessed the transfection efficiency using fluorescent-labeled control siRNAs (siGLO Transfection indicator-Red; Thermo Scientific). The transfection efficiency was determined using fluorescence microscopy by counting three different fields at 40x magnification, and calculated as: Efficiency = mean number of the red-fluorescing cells / (mean number of the total fluorescent + non-fluorescent cells).

2. 12   Cell cycle analysis

After harvesting, cells were washed with cold PBS and fixed with cold ethanol (70%). They were incubated at 4°C at least for 24 hours, and stained with propidium iodide (PI). A flow cytometer was used to measure DNA content of the cells. Cell cycle analysis was done using software MultiCycle (Phoenix Flow Systems, Inc., San Diego, CA).

2. 13   Evaluation of LDOC1 protein expression K\knock-down by Western blot

Whole cell lysates were prepared from freshly isolated cultured HeLa cells. After removing the medium, the cells were washed, detached with trypsin, transferred to cold 15 mL culture tubes, and spun at 1500 rpm for 5 minutes at 4°C in a refrigerated centrifuge. After removing the supernatant, the cell pellets were stored at -80°C for a minimum of 24 hours to denature the cellular proteins. The pelleted cells were lysed on ice in lysis buffer (25 mM Hepes, pH 7.7, 400 mM sodium chloride, 1.5 mM magnesium chloride, 2 mm ethylenedinitrilotetraacetic acid (EDTA), 0.5% Triton X-100). Immediately before use, the reducing agent 3 mM dithiothreitol (DTT), 20 mM beta-glycerophosphate, 2 mm sodium orthovanadate (Na$_3$VO$_3$), 5 mM sodium fluoride, 25 mM para-nitrophenyl-phosphate disodium (PNPP), 0.1 mM phenyl methyl sulfonyl fluoride (PMSF), and complete protease inhibitor cocktail (Roche, Indianapolis, IN) were added to the lysis buffer. The lysates were homogenized by sonication, placed on ice for 30 minutes, and spun at 4°C in a refrigerated centrifuge to remove cellular debris. The protein concentration was measured by Bradford assay. Total protein (100 µg) was heat denatured for 3 minutes in sample buffer (50 mM Tris-HCl (pH 6.8), 10% glycerol, 2% SDS, 0.025% bromophenol blue, 2.5% ß-mercaptoethanol), separated by 15% Tris-HCl SDS polyacrylamide gel electrophoresis, and transferred to a polyvinylidene fluoride (PVDF) membrane. The membrane was blocked with 3% non-fat milk in PBS-T buffer (PBS with 0.05% Tween 20) and incubated overnight at 4°C with either affinity-purified polyclonal mouse anti-human LDOC1 antibody (dilution 1:500; Abnova, Taiwan) or monoclonal mouse anti-human beta-actin antibody conjugated to horse-radish-peroxidase (HRP) (dilution 1:25,000; Sigma, St. Louis Missouri). All antibodies were diluted in 3% milk with PBS-T. The membranes were washed twice with PBS-T for 10

min, once with PBS for 10 min, and then once with ddH$_2$O for 30 min. To detect LDOC1, the membranes were incubated for a minimum of 60 minutes at room temperature with anti-mouse secondary antibody conjugated to HRP (GE Healthcare, United Kingdom) and washed, as described above. After exposure to chemiluminescent substrate for 1 minute, the peroxidase reaction was detected using the Enhanced Chemiluminescent-plus (ECL-plus, GE Healthcare) system for LDOC1 and the ECL system (GE Healthcare) for beta-actin.

2.14   Global gene expression profiling

We used GeneChip Human Genome U133 Plus 2.0 arrays (Affymetrix, Santa Clara, CA) for the 3' in vitro transcription (IVT) expression analysis. This chip contains more than 54,000 probe sets that interrogate 47,000 transcripts from approximately 38,500 well-characterized human genes. The hybridizations were performed in triplicate according to the manufacturer's instructions. The experiment was performed using six chips, three chips for each of two conditions: LDOC1 siRNA-treated HeLa cells (LD group) and non-targeting siRNA-treated HeLa cells (NT group). Briefly, total RNA (100 ng per chip) was reverse transcribed using a T7 oligo(dT) primer to synthesize cDNA that contains the T7 promoter sequence. This first strand cDNA was then converted to double-stranded DNA by DNA polymerase in the presence of RNase H, which simultaneously degrades the RNA during the reaction. The in vitro transcription reaction was performed in the presence of T7 RNA polymerase and biotinylated ribonucleotide analogues. Subsequently, the amplified RNA (cRNA) was purified to remove unincorporated nucleotide triphosphates (NTP), enzymes, salts and inorganic phosphate to improve the stability of the biotin-modified cRNA. The expected size of the

cRNA is between 250-5500 nucleotides (nt), with a peak size of 600-1200 nt. The cRNA was then fragmented to obtain 35-200 nt fragments, with a peak size ranging from 100-120 nt. The fragments were then hybridized to the chips at 45°C for 16 hours. The chips were then scanned with a GeneArray Scanner (Hewlett-Packard, Palo Alto, CA). The entire hybridization procedure was performed in the institutional Core Facility.

2.15   Statistical analysis of gene expression profiling microarray data

All data pre-processing and statistical analyses were performed in R, a free software environment for statistical computing and graphics (http://www.r-project.org/). As part of standard quality control (QC) analysis, the .CEL files were quantified using the MAS5 algorithm. The data quality was examined by preparing RNA degradation plots, Bland-Altman (M-versus-A) pairwise plots, density plots, and box plots. Then, the expression levels were quantified using the method "Robust Multiarray Analysis (RMA)". A two-sample t-test was performed to identify differentially expressed genes between LD and NT groups. To account for multiple testing, a beta-uniform mixture (BUM) model was applied to estimate false discovery rate (FDR).

In order to identify genes that were coordinately expressed in HeLa and CLL cells, we compared the data obtained from the HeLa knock-down experiments with gene expression profiling data previously acquired on 30 CLL samples in our laboratory (34). These CLL samples had been hybridized to the Affymetrix U133A 2.0 gene expression microarrays, which contain 22,283 probe sets. The array assesses the expression level of 18,400 transcripts and variants, including 14,500 well-characterized human genes. Data were processed using RMA, and two group comparisons between

21

LDOC1-high and LDOC1-low samples were performed using two-sample t-tests. The microarray gene expression analyses were performed in collaboration with Dr. Kevin Coombes.

2.16   Pathway analysis

Pathway analysis of differentially expressed genes was performed using the Ingenuity Pathway Analysis (IPA), a web-based algorithm (Ingenuity® Systems, IPA version 9.0, www.ingenuity.com) (35). For this analysis, the mean gene expression values were first base-two log transformed. To determine fold changes, the ratio of the LD/NT groups were calculated using Excel. Based on their p values, the ratios of the most significantly differentially expressed 107 genes were uploaded into the IPA algorithm. A core analysis was performed using a 0.5 fold change in ratio as the cut-off, querying only the direct interactions. The IPA algorithm uses literature-curated and database-originated a priori knowledge (Ingenuity Knowledge Base) and builds networks using those interactions as scaffolds, filling in with the gene list uploaded by the user.

CHAPTER 3. LDOC1 IS DIFFERENTIALLY EXPRESSED IN CLL PROGNOSTIC SUBTYPES AND PREDICTS OVERALL SURVIVAL IN UNTREATED PATIENTS

3.1   Introduction

In an earlier study performed to identify candidate biomarkers associated with *IGHV* somatic mutation status in 49 previously untreated CLL patients, our group identified *LDOC1* as one of the genes that was differentially expressed with strong statistical significance between mutated and unmutated cases (22). We have expanded this analysis to a total of 131 samples obtained from previously untreated CLL patients and evaluated the potential of *LDOC1* mRNA as biomarker of prognosis. Our analyses included correlations with known standard clinical and laboratory markers of prognosis, and univariate and multivariate survival analyses.

3.2   Results

3.2.1 *LDOC1* mRNA expression is strongly associated with known markers of poor prognosis

We expanded our previous analysis of 49 samples (22) to a total of 131 samples obtained from untreated CLL patients. We found that the distribution of *LDOC1* expression was bimodal, with no patients exhibiting ΔC_T values between 7 and 8 cycles **Figure 1**). Thus, we defined samples to be *LDOC1*-positive if $\Delta C_T \leq 7.5$ and *LDOC1*-negative if $\Delta C_T > 7.5$.
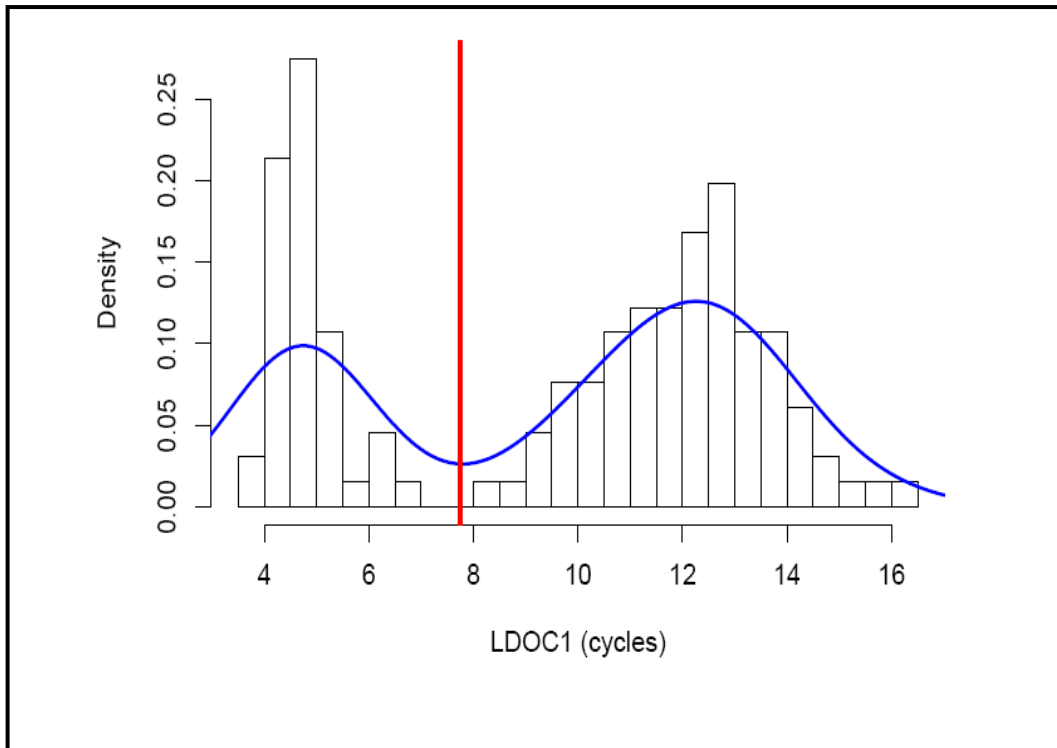
Figure 1. Histogram of the observed LDOC1 values in normalized cycles demonstrates a clear separation of CLL samples. Normalized Ct values of 131 CLL samples obtained by MF-QRT-PCR assay are shown on the X-axis. The red vertical line indicates the cut-off point to allocate *LDOC1* mRNA expression into positive (to the left of the red line) or negative (to the right of the red line) groups. The figure contains continuous (blue line) and discrete (histogram) approximations of the "probability density" function. The area of each bar represents the percentage of patient samples that have values in the interval represented by the x-axis values.

The patient characteristics in both groups are summarized in **Table 1**. With respect to age, gender, Rai stage, white blood cell count, or serum beta-2 microglobulin levels at the time the sample was obtained, there were no statistically significant differences between patients whose cells expressed *LDOC1* mRNA and those whose cells were negative for *LDOC1* mRNA. Since we originally identified *LDOC1* mRNA as a biomarker of *IGHV* somatic mutation status (22), its expression strongly correlated with the somatic mutation status, as expected (Fisher's exact test; p=2.20 x 10^{-16}). For each case, the data for the *IGHV* somatic mutation status, *IGHV* family, percent homology with the germline sequence, and *LDOC1* mRNA expression, as well as ZAP70 protein expression and cytogenetic findings, are listed in **Supplementary Table 1**.

Out of a total of 131 cases, 65 out of 67 mutated cases (97%) were negative for *LDOC1* mRNA, and 43 out of 63 unmutated cases (68%) were positive for *LDOC1* mRNA; the *IGHV* somatic mutation status was unavailable for one case (**Figure 2A, Supplementary Table 1**). Equivalently, 43 out of 45 (96%) *LDOC1*-positive cases were unmutated and 65 out of 85 (76%) *LDOC1*-negative cases were mutated. Thus, 22 cases (17%) showed discordance between *LDOC1* mRNA expression and *IGHV* somatic mutation status; two mutated cases were *LDOC1*-positive and 20 unmutated cases were *LDOC1*-negative.

Table 1. Clinical and Laboratory Features[†, **]

	All Patients (n = 131)	*LDOC1* positive (n = 46)	*LDOC1* negative (n = 85)	p value*
Age in years				
Median	59	60	59	0.6796
(range)	(27 – 82)	(27 – 82)	(27 – 81)	
Gender				
Male, n (%)	81 (62%)	27 (59%)	54 (64%)	0.7066
Female, n (%)	50 (38%)	19 (41%)	31 (36%)	
Rai stage	(n = 131)	(n = 46)	(n = 85)	
0-2, n (%)	102 (78%)	35 (76%)	67 (79%)	0.826
3-4, n (%)	29 (22%)	11 (24%)	18 (21%)	
WBC count	(n = 131)	(n = 46)	(n = 85)	
$\leq$ 150x10^9/L, n (%)	118 (90%)	38 (83%)	80 (94%)	0.0621
> 150x10^9/L, n (%)	13 (10%)	8 (17%)	5 (6%)	
Serum β2 microglobulin	(n = 130)	(n = 46)	(n = 84)	
< 4, n (%)	98 (75%)	31 (67%)	67 (80%)	0.1384
$\geq$ 4, n (%)	32 (25%)	15 (33%)	17 (20%)	
IGHV somatic mutation status	(n = 130)	(n = 45)	(n = 85)	
Mutated, n (%)	67 (52%)	2 (4%)	65 (76%)	2.20 x 10^{-16}
Unmutated, n (%)	63 (48%)	43 (96%)	20 (24%)	
ZAP70 protein status	(n = 113)	(n = 39)	(n = 74)	
Positive, n (%)	51 (45%)	30 (77%)	21 (28%)	1.06 x 10^{-6}
Negative, n (%)	62 (55%)	9 (23%)	53 (72%)	
Cytogenetic changes	(n = 100)	(n = 39)	(n = 61)	
None, n (%)	27 (27%)	14 (36%)	14 (23%)	
isolated del(13q), n (%)	36 (36%)	5 (13%)	30 (49%)	0.0005834
del(6q), del(11q), del(17p), +12, or del(13q) with other abnormalities, n (%)	37 (37%)	20 (51%)	17 (28%)	

† Age, Rai stage, WBC count, and serum β2 microglobulin values are reported for the time the sample was obtained for *LDOC1* mRNA expression; *IGHV* somatic mutation status, ZAP70 protein status, and cytogenetic changes were determined on samples obtained before treatment.

* All p values were calculated using the two sided Fisher's exact test except for age in years, which was calculated using the two-sided t-test.

** For serum β2 microglobulin and *IGHV* somatic mutation status, one value was unavailable; for ZAP70 protein, 18 values were unavailable, for genomic abnormalities, 31 values were unavailable.

Figure 2
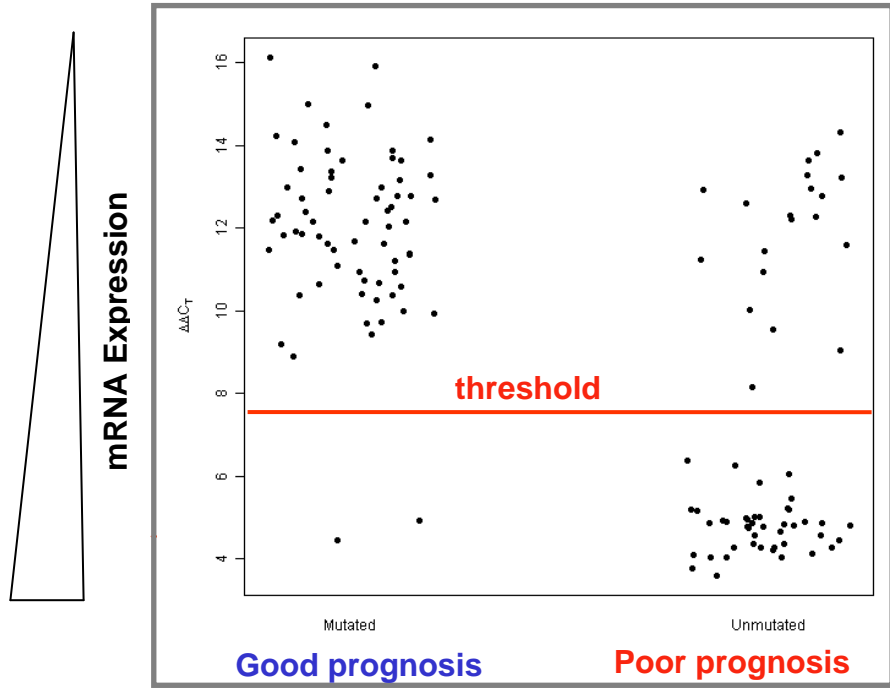
A. LDOC1
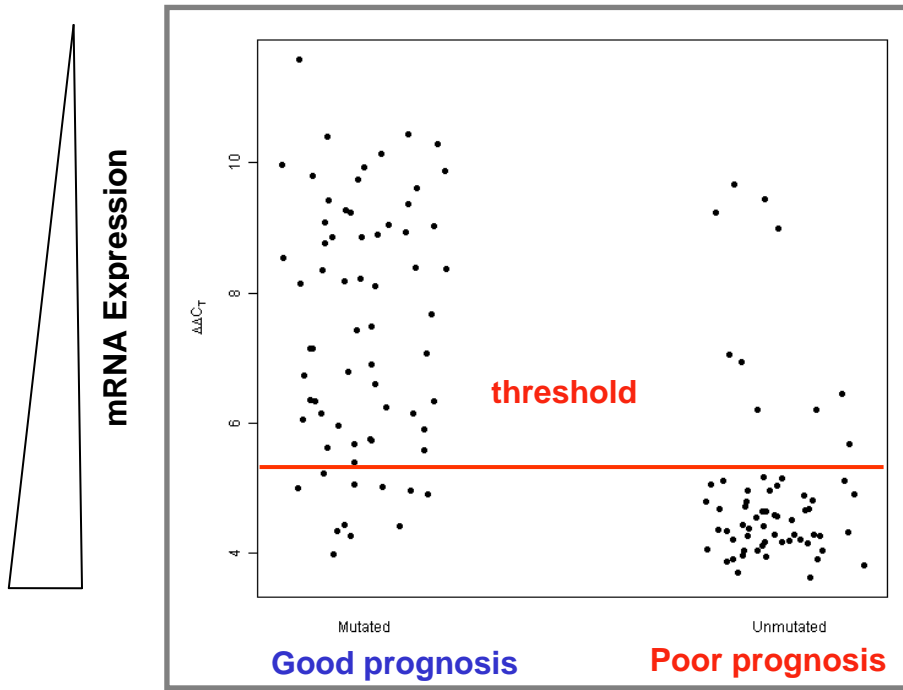
B. ZAP70

Figure 2: **Expression of *LDOC1* and *ZAP70* mRNAs measured by MF-QRT-PCR assay distinguishes between mutated and unmutated cases of CLL. (A) Dot-plot for *LDOC1* mRNA expression**. For the mutated cases, 65 out of 67 mutated cases were unambiguously negative for *LDOC1* mRNA (higher $\Delta\Delta$CT values) and 2 were positive (lower $\Delta\Delta$CT values). For the unmutated cases, 43 out of 63 unmutated cases were positive for *LDOC1* mRNA and 20 were negative. **(B) Dot-plot for *ZAP70* mRNA expression**. For the unmutated cases, 53 out of 63 unmutated cases were positive for *ZAP70* mRNA and 10 were negative. Eleven out of 67 mutated cases were positive for *ZAP70* mRNA and 56 were negative. The $\Delta\Delta$CT value considered for threshold = 5.3

3.2.2 *LDOC1* **mRNA expression more strongly predicts** *IGHV* **somatic mutation status than does ZAP70 protein expression**

Since evaluation of ZAP70 protein expression has been considered a surrogate biomarker for *IGHV* somatic mutation status (36, 37), we evaluated the association between *IGHV* somatic mutation status and ZAP70 protein expression, measured by immunohistochemistry or flow cytometry (IHC/Flow). We detected a positive association between ZAP70 protein positivity by IHC/Flow and *IGHV* somatic mutation status (Fisher's exact test; $p=1.42 \times 10^{-9}$). Of the 112 cases for which ZAP70 protein expression data were available, 24 (21%) showed discordance between *IGHV* somatic mutation status and ZAP70 protein expression **(Table 2)**, consistent with the results of previous studies (14, 16, 17, 31, 36, 37). Eleven out of 59 mutated cases (19%) were ZAP70 positive; 13 out of 53 unmutated cases (25%) were ZAP70 negative. Thus, *LDOC1* mRNA expression was more strongly associated with *IGHV* somatic mutation status than was ZAP70 protein expression.

ZAP70 protein	UM	M	Total
Positive	40	11	51
Negative	13	48	61
Total	53	59	112

Table 2. Correlation of ZAP70 protein expression and *IGHV* mutation status. Only in 112 patients data were available for both ZAP70 protein expression and *IGHV* mutation status. UM, *IGHV*-unmutated; M, *IGHV*-mutated

3.2.3 *LDOC1* **mRNA expression is associated with cytogenetic markers of prognosis**

We found that expression of *LDOC1* mRNA correlated with cytogenetic markers of prognosis (**Table 1,** Fisher's exact test; p= 0.0005834; **Supplementary Table 1**). Cases that were negative for *LDOC1* mRNA were more likely to harbor isolated deletions in chromosome 13q14.3, a marker of good prognosis, compared to samples that were positive for *LDOC1* mRNA. In contrast, cases that were positive for *LDOC1* mRNA were more likely to harbor genomic abnormalities associated with poor prognosis, i.e., del(6)(q21), del(11)(q22.3), del(17)(p13.1), +12, or with del(13)(q14.3), another cytogenetic marker of poor prognosis, than cases that were negative for *LDOC1* mRNA expression.

3.2.4 **Total *LDOC1* mRNA expression is a better predictor of overall survival than either *IGHV* somatic mutation status or ZAP70 protein expression**

Since *LDOC1* mRNA expression was associated with *IGHV* somatic mutation status and ZAP70 protein expression, both strong predictors of prognosis in CLL patients, we sought to determine if *LDOC1* mRNA expression could also serve as a biomarker of prognosis. Thus, we analyzed the relationship between *LDOC1* mRNA expression and overall survival. We found that patients whose cells were negative for *LDOC1* mRNA had a significantly longer median survival than patients whose cells were positive, regardless of whether overall survival was measured from the time of diagnosis (**Figure 3**; logrank test, p = 0.009581) or from the time the sample was obtained (logrank test, p = 0.02294; data not shown). The median survival for the

LDOC1 mRNA negative patients was not reached, whereas the median survival for *LDOC1* mRNA positive patients was 164 months.

Further, we applied the Akaike Information Criterion (AIC) to a multivariate model that incorporated *LDOC1* mRNA expression, ZAP70 protein expression, and *IGHV* somatic mutation status. The optimal model retained only *LDOC1* mRNA expression (AIC = 167.93), eliminating ZAP70 protein expression (AIC = 169.2) and mutation status (AIC=171.02). Smaller values of AIC provide better models (33). Thus, in this sample set *LDOC1* mRNA expression was a better predictor of overall survival than either *IGHV* somatic mutation status or ZAP70 protein expression. (The multivariate model was constructed by our collaborator, Dr. Kevin Coombes).

3.3 Summary

We have shown that *LDOC1* mRNA expression positively correlates with important known biomarkers of poor prognosis, i.e., unmutated *IGHV* somatic mutation status, ZAP70 protein expression, and cytogenetic markers of poor prognosis. Further, *LDOC1* mRNA expression is associated with shorter overall survival in a univariate analysis. Finally, *LDOC1* mRNA is an excellent biomarker of prognosis in untreated CLL patients, and is a better predictor of overall survival than either *IGHV* somatic mutation status or ZAP70 protein expression, the current gold standard biomarkers of prognosis in CLL.
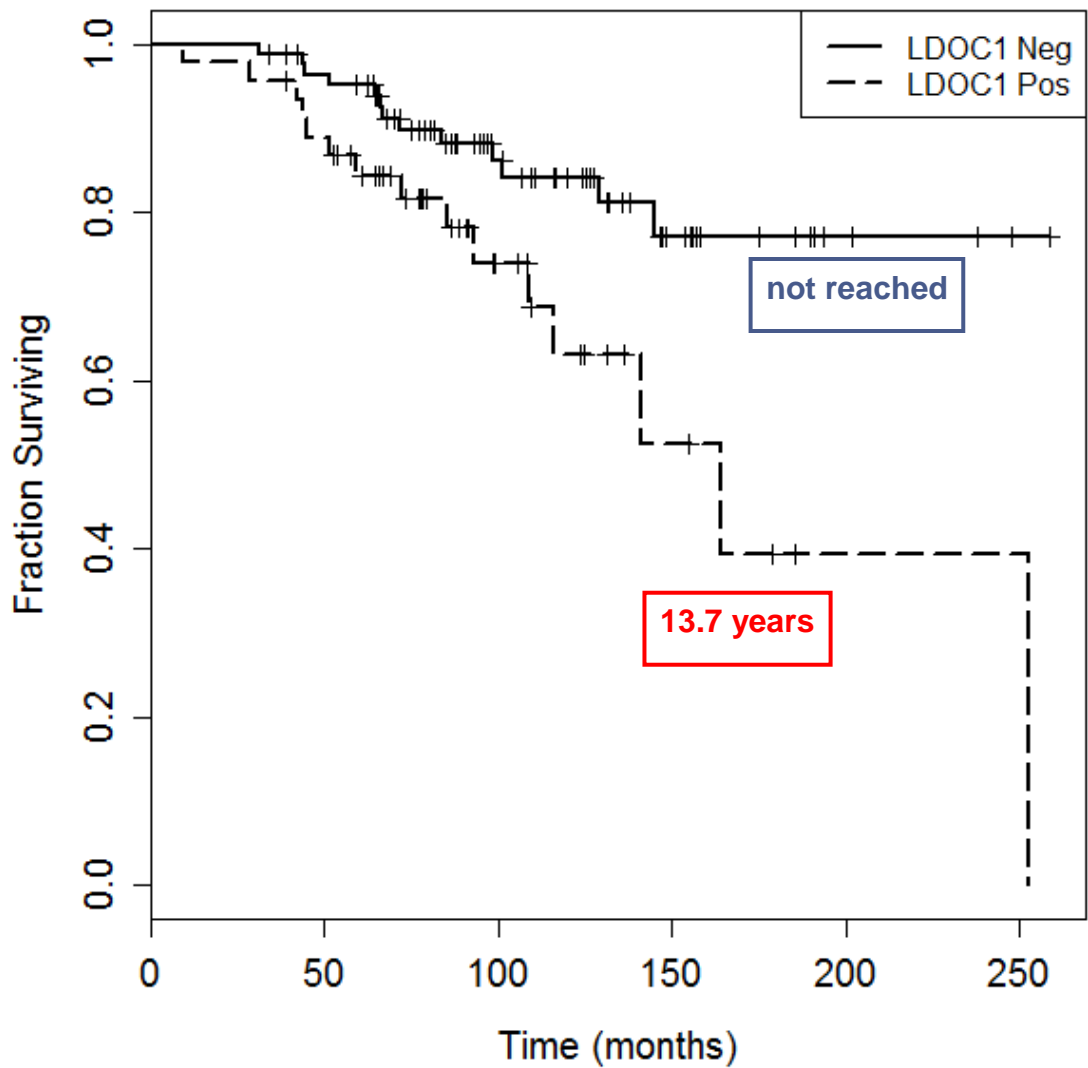
Figure 3. *LDOC1* **mRNA expression predicts overall survival.** Overall survival was measured from the time of CLL diagnosis. The median survival for *LDOC1* mRNA positive patients was 164 months (13.7 years); the median survival for the *LDOC1* mRNA negative patients was not reached.

CHAPTER 4. LDOC1 IS DIFFERENTIALLY EXPRESSED IN NORMAL PERIPHERAL BLOOD B CELL SUBSETS AND IN SUBTYPES OF PRIMARY NON-HODGKIN LYMPHOMAS

4.1 Introduction

Little is known about the expression of *LDOC1* mRNA in normal B cells (NBC) or in lymphoid malignancies other than CLL. To begin to understand its biological role in B cell development, differentiation, activation, and transformation we screened a variety of normal and malignant B-cell subsets for LDOC1 mRNA expression. We used the RT-PCR to identify novel mRNA variants and the QRT-PCR to measure the relative amounts of the isoforms. During the course of our studies we discovered a new splice variant, *LDOC1S*.

4.2 Results

4.2.1 RT-PCR analysis of LDOC1 mRNA reveals a novel splice variant, *LDOC1S*

We assessed *LDOC1* mRNA expression in NBC, a variety of primary B-cell lymphoma samples, and carcinoma cell lines, using two primer pairs, AB and AC (**Figure 4**). Primer pair AB was designed to amplify the entire *LDOC1* coding region and yield a product of 464 bp (23). The reverse primer, C, in primer pair AC was designed to bind within the 3' UTR and yield a product of 649 bp. Using primer pair AB we detected strong expression of the expected wild-type 464 bp product in seven unmutated CLL cases (CLL 42, 46, 49, 51, 53, 54, 60), two NBC samples (NBC 4, 6), and the MCF-7 breast carcinoma cell line, which has been shown to express high levels of *LDOC1* (23)

(**Figure 5A and B**). We detected faint or no RT-PCR products in seven mutated CLL cases (CLL 58, 62, 12, 37, 61, 67, 99), or in the GA-10 Burkitt lymphoma and Jurkat T-cell lymphoblastic lymphoma cell lines. However, primer pair AC yielded two distinct bands in all positive samples, the expected 649 bp product and an unexpected 165 bp product. Three out of four additional NBC samples, as well as normal peripheral blood T cells and HeLa cervical carcinoma cells were also positive with both primer pairs (data not shown).

In a subset of CLL cases that expressed *LDOC1* mRNA (CLL 46, 49, 53), we determined the sequence of the 464, 165, and 649 bp products. We also determined the sequence of the 464 and 649 base pair products in a subset of NBC cases (NBC 1, 3, 5). The sequences from the 464 and 649 bp products were identical to the published *LDOC1* reference sequence (GenBank RefSeq NM_012317). Sequence analysis of the 165 bp product revealed that it was a splice variant (**Figure 4B**). The *LDOC1* gene is an intronless gene that encodes an mRNA of 1376 bp (23). The *LDOC1* and *LDOC1S* mRNAs have identical translation start sites, with a sequence similar to that described by Kozak (38). The splice variant contains canonical splice donor (AG|GTACGT at nucleotide 232) and acceptor sequences (TGTCTTTGTTCCAG|G at nucleotide 704), as well as a branch sequence (TTCAT at nucleotide 685) (Alex's Splice Site Finder, version 0.5; NNSplice, version 0.9). Thus, in the splice variant, approximately the first third of the amino acid coding region is joined with the 3' UTR at nucleotide 718. After one codon (GAA, glutamic acid), the coding sequence is terminated by a stop codon (TAG), followed by a 3'UTR that is identical to the wild-type sequence. If translated, the 165 bp splice variant would produce a truncated protein of 44 amino acids that contains the leucine zipper region of the wild-type protein; the proline rich-region (amino acids 46-65) and the remainder of the coding region would be absent (**Figure 4C**).

Figure 4

C
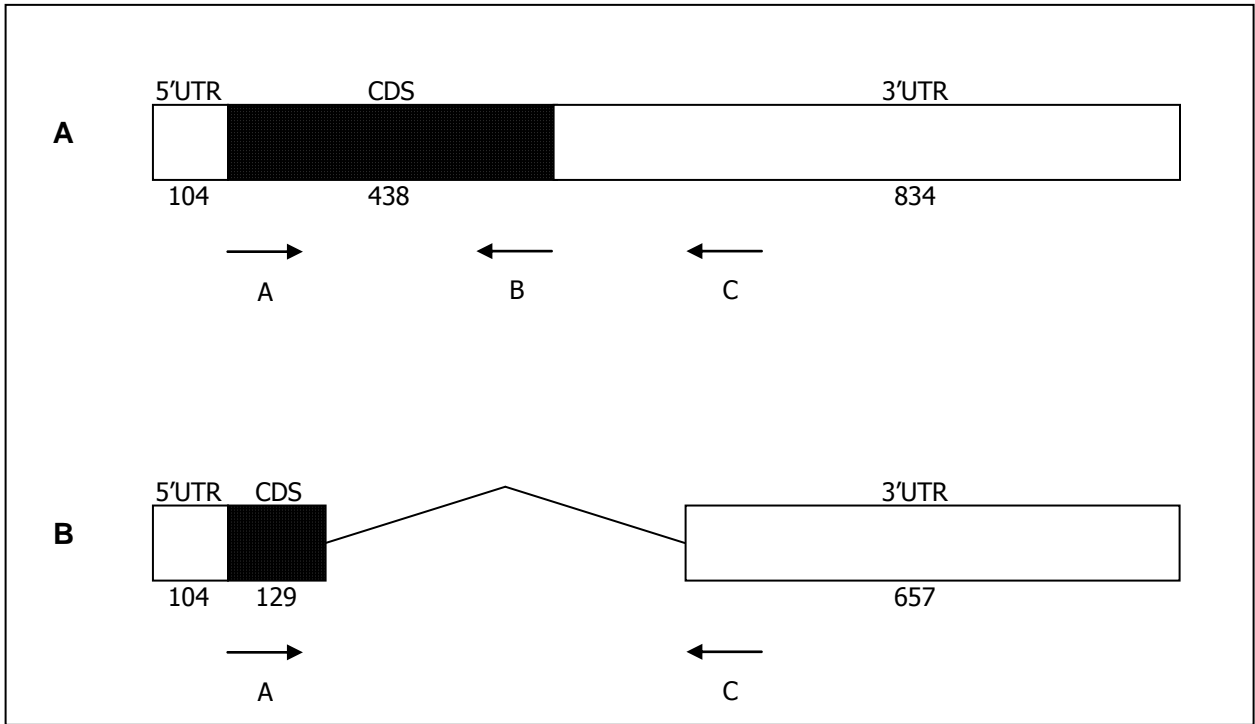
```
LDOC1   MVDELVLLLHALLMRHRALSIENSQLMEQLRLLVCERASLLRQVRPPSCPVPFPETFNGE 60
LDOC1S  *****************************************-------------------

LDOC1   SSRLPEFIVQTASYMLVNENRFCNDAMKVAFLISLLTGEAEEWVVPYIEMDSPILGDYRA 120
LDOC1S  ------------------------------------------------------------

LDOC1   FLDEMKQCFGWDDDEDDDDEEEEDDY 146
LDOC1S  ----------------------------------------------------------- 180

LDOC1S  ----------------------------------------------------------- 240

LDOC1S  ----------------------------------------------------------- 300

LDOC1S  --------------------E
```

Figure 4: Structure of the *LDOC1* wild-type and splice variant mRNAs, and translated proteins. **(A and B) Structure of the *LDOC1* and *LDOC1S* mRNAs, respectively.** The LDOC1 gene is an intronless gene that spans 1376 bp. Open boxes represent 5' and 3' untranslated regions (UTR) and shaded boxes represent the coding sequences (CDS); the number of nucleotides is indicated below. **(C) Alignment of the amino acid sequences of wild-type LDOC1 protein and the putative splice variant protein.** The wild-type mRNA encodes a protein composed of 146 amino acids. The splice variant mRNA, if translated, would yield a truncated protein of 44 amino acids that corresponds mainly to leucine zipper region of the wild-type LDOC1 protein. Identical residues are indicated by asterisks; the dashes indicate nucleotides that have been removed from the *LDOC1S* mRNA by alternative splicing. The leucine zipper domain (amino acids 5-40) in the wild-type protein, indicated by an open blue box, would be preserved in the splice variant.
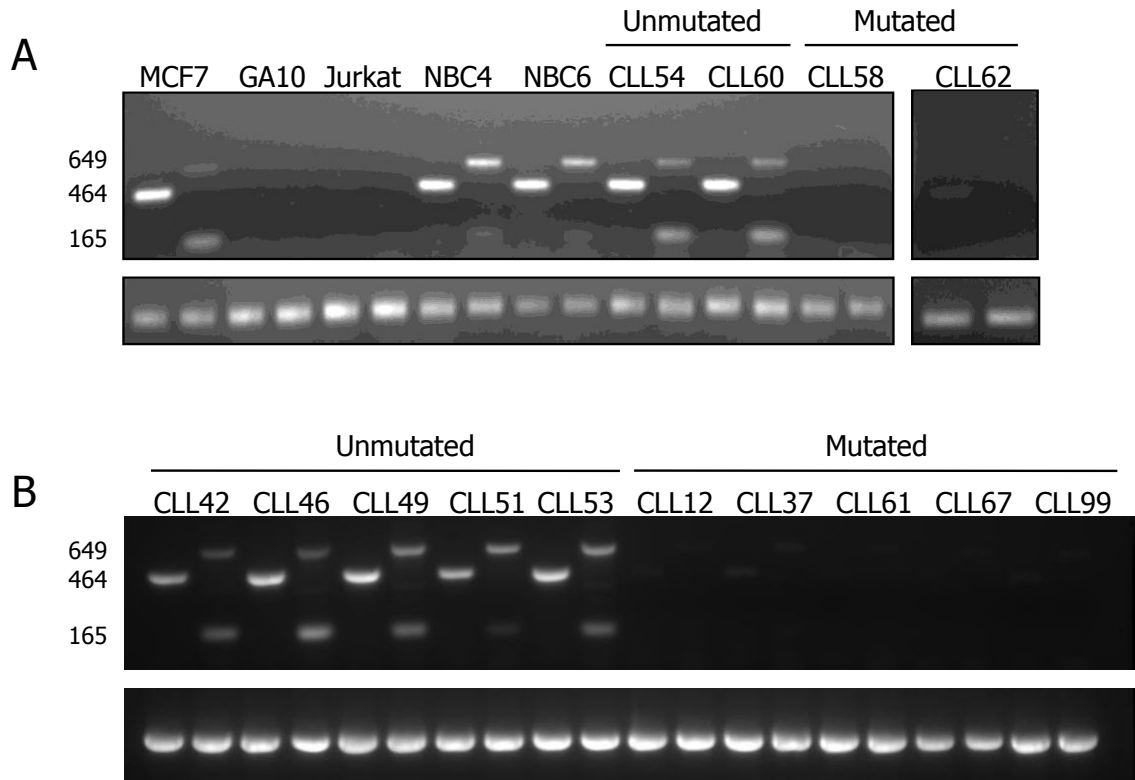
Figure 5: **Expression of *LDOC1* and *LDOC1S* assessed by RT-PCR. (A) Expression of *LDOC1* and *LDOC1S* in cell lines, normal peripheral blood B cells, and CLL cells**. Wild type *LDOC1* (464 and 649 bp) and *LDOC1S* (165 bp) were detected in the MCF7 breast cancer cell line, two normal peripheral blood B cell (NBC) samples, and two unmutated CLL samples. Little or no *LDOC1* or *LDOC1S* were detected in the Burkitt lymphoma cell line, GA10, the T-cell lymphoblastic lymphoma cell line, Jurkat, or two mutated CLL samples. **(B) Expression of *LDOC1* and *LDOC1S* in CLL cells**. Wild type *LDOC1* (464 and 649 bp) and *LDOC1S* (165 bp) were detected in the five additional unmutated CLL samples. Little or no *LDOC1* or *LDOC1S* were detected in five additional mutated CLL samples. The amount of cDNA amplified for each sample was comparable, as shown by the beta-actin signal.

37

4.2.2 Total *LDOC1* mRNA expression varies in NBC and B-cell subsets, CLL and primary B-cell lymphoma samples, and cell lines by QRT-PCR assay

We assessed the expression of *LDOC1* mRNA in unfractionated NBC samples and in NBC samples that had been enriched for naïve (CD27−) or memory (CD27+) B cells. We also assessed its expression in 10 primary B-cell non-Hodgkin's lymphoma samples, for which we had determined the *IGHV* somatic mutation status. The primary lymphoma samples included three follicular lymphoma samples (FL1, FL2, FL3, all mutated), four mantle cell lymphoma samples (MCL1, MCL2, and MCL4, mutated; MCL3 unmutated), one extranodal marginal zone B-cell lymphoma sample (MZL, unmutated), and two splenic marginal zone lymphoma samples (SMZL1, unmutated; SMZL2, mutated). For these experiments we used the commercially available QRT-PCR assay that detects total *LDOC1* mRNA; it does not distinguish between the isoforms. The results for unfractionated NBC and CLL samples, and lymphoma and carcinoma cell lines were consistent with the results obtained by the RT-PCR assay (**Figure 6**). For the fractionated NBC samples, the fraction enriched for naïve B cells expressed higher levels of *LDOC1* mRNA than the fraction enriched for memory B cells. We found that *LDOC1* mRNA was also expressed in primary B-cell lymphoma samples. Although the sample size is insufficiently large for a statistical analysis of individual lymphoma subtypes, there was a trend for the unmutated lymphoma samples, regardless of subtype, to express higher levels of *LDOC1* mRNA than mutated samples.
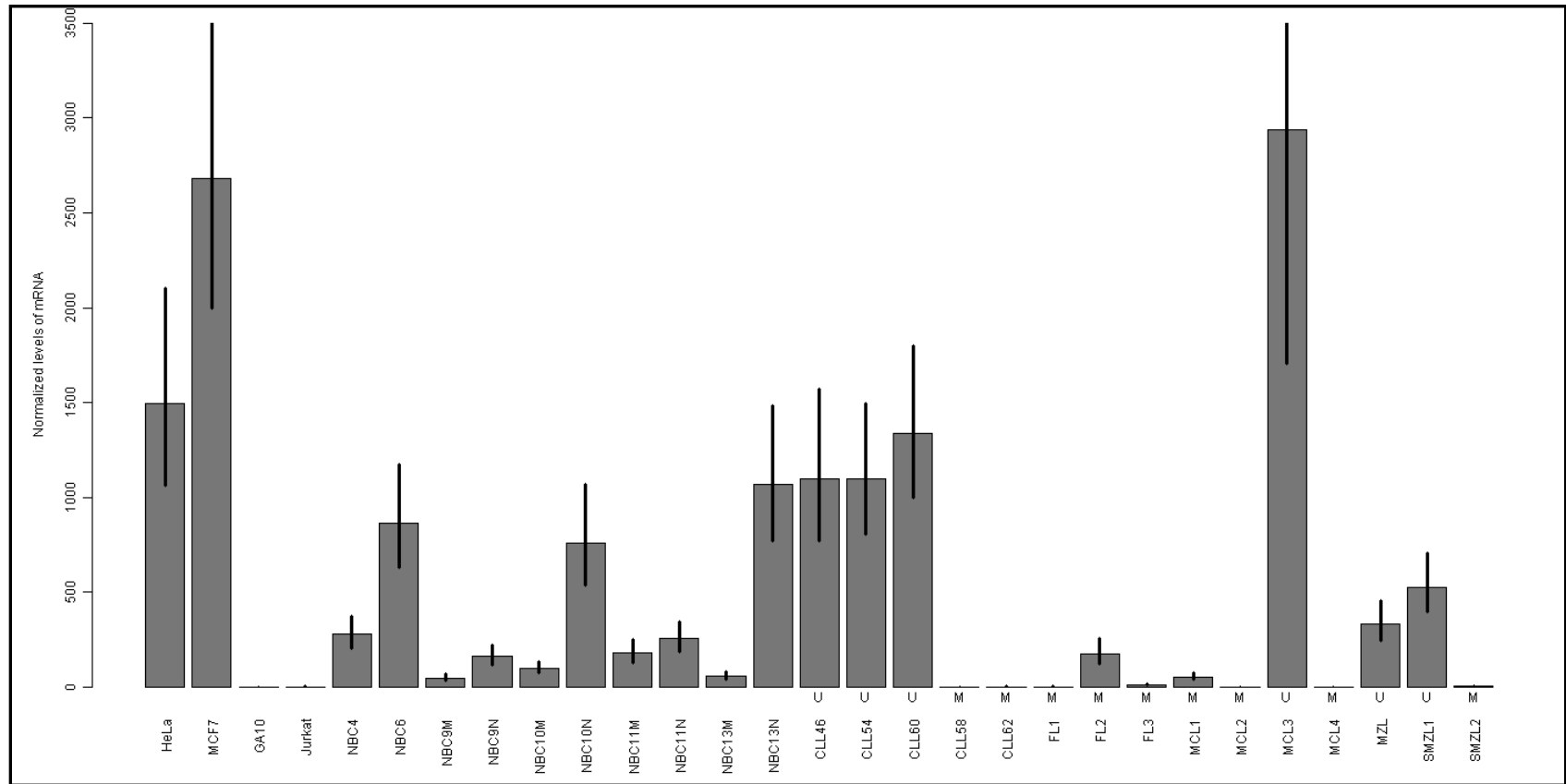
Figure 6

Figure 6: Expression of total *LDOC1* mRNA measured by QRT-PCR assay. LDOC1 levels are measured using a commercially-available TaqMan assay that does not distinguish between the *LDOC1* and *LDOC1S* mRNA isoforms; mRNA quantity shown on the y-axis refers to total *LDOC1* mRNA. *LDOC1* expression was measured in carcinoma (HeLa, MCF7) and lymphoma (GA10, Jurkat) cell lines, unfractionated normal peripheral blood B cells (NBC4, NBC6), normal peripheral blood B cells enriched for memory B cells (NBC9M, NBC10M, NBC11M, NBC13M), normal peripheral blood B cells enriched for naïve B cells (NBC9N, NBC10N, NBC11N, NBC13N), and unmutated (U) or mutated (M) CLL and primary B-cell lymphoma samples, including follicular (FL), mantle cell (MCL), marginal zone (MZL), or splenic marginal zone (SMZL) lymphoma samples. Error bars represent the standard error of the ΔΔCt values.

4.2.3 LDOC1 and LDOC1S mRNA isoform expression by QRT-PCR assay

Because the commercially available QRT-PCR assay fails to distinguish between the *LDOC1* and *LDOC1S* mRNA isoforms, we were unable to assess the contribution of each isoform to the total *LDOC1* mRNA levels that we observed using the RT-PCR and commercially available QRT-PCR assays. Thus, we designed QRT-PCR assays that distinguish between the isoforms. (The primer and TaqMan probe components of each assay are described in Chapter 2.)

In order to determine the assay specificity, that is, whether the TaqMan assay designed against the sequences of either *LDOC1* or *LDOC1S* mRNA recognize their own sequence specifically, and discriminate efficiently between the two, we prepared synthetic templates that consist of the amplicons between the forward and reverse primer sequences, inclusive: the wild-type template is 5'-TGGTGCCCTACATCGAGAT GATAGCCCCATCCTAGGTGTTACCGGGCCTTCCTCG-3', and the splice variant template is 5'-TTCCAAGCACTTCCGAGTGACTATTCCTGGCGCAGCAGCAGCCGCA GCTGTTCCAT-3'. We used 33,000 molecules of each primer as the cDNA template. Unlike cDNA obtained from samples, the primer templates allowed us to calculate more precisely the amount of input template in PCR reactions. We determined that the assay for the wild-type *LDOC1* mRNA isoform recognizes the wild-type template with 14×10^6 fold specificity compared to the splice variant template, as determined by RQ of $2^{-\Delta Ct} = 2^{-(Ct\ target\ -\ Ct\ homolog)}$. The assay for *LDOC1S* mRNA isoform recognizes the splice variant template with 43×10^3 fold specificity compared to the wild-type template **(Figure 7)**. Since a specificity of more than 30×10^3 fold is the standard used for the commercially available TaqMan assays (Dr. John Pfeifer, Life Sciences, personal communication), the assays we designed are highly specific. To account for the different assay

efficiencies, we used the data from all samples to fit a model expressing total *LDOC1* mRNA as a weighted linear combination of wild-type (WT) and splice variant (SV) contributions. For normalized mRNA levels, we found the optimal model to be Total = (0.67 x WT) + (0.24 x SV) (developed by Dr. Kevin R. Coombes).

We re-assessed the previously-tested samples described above using the isoform-specific TaqMan assays (**Figure 8**). In general, for both benign and malignant cells, cells that expressed the wild-type *LDOC1* mRNA also expressed the splice variant *LDOC1S* mRNA, but the wild-type isoform was predominant.
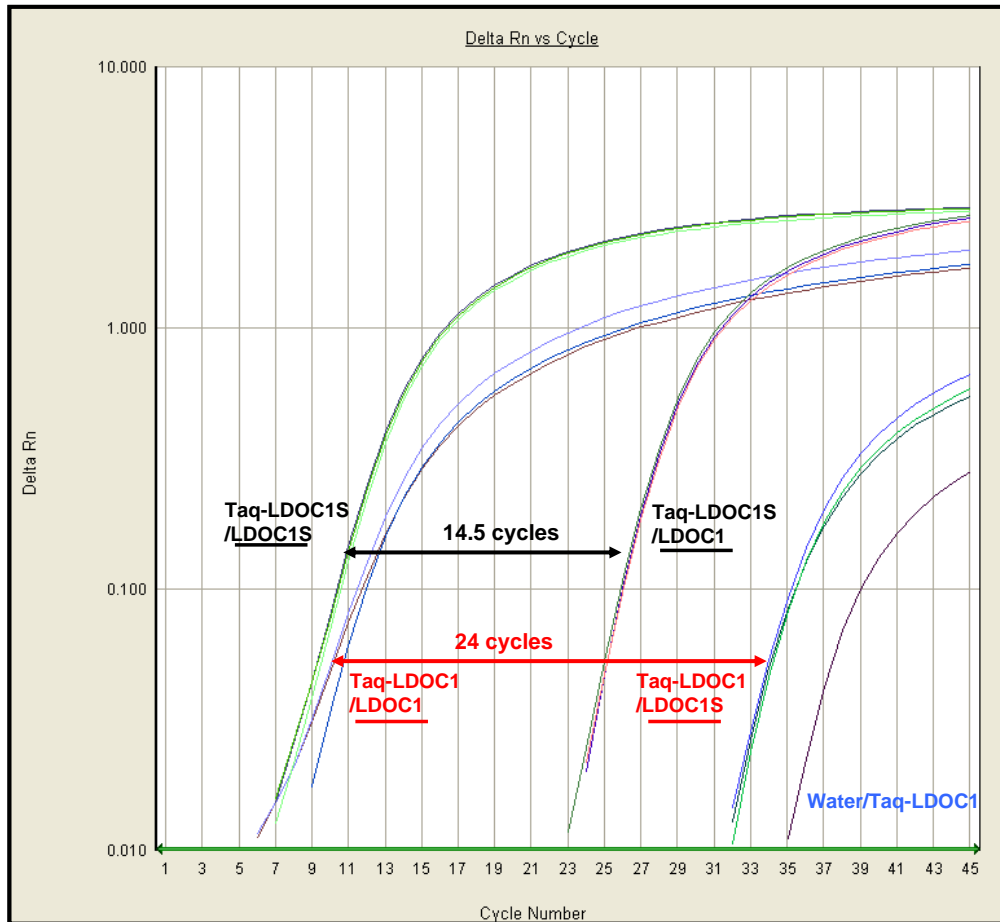
Figure 7: Isoform-specific TaqMan assays are highly specific for their target transcript. Synthetic templates consist of the amplicons between the forward and reverse primer sequences were synthesized for LDOC1 and LDOC1S, corresponding to the sequences uniquely found in each isoform. TaqMan assays designed to specifically recognize their own templates were tested for specificity against each other's template. Each arrow marks the threshold set for calculating the ΔCt values for the corresponding TaqMan assay against two different templates. The underlines indicate the templates, LDOC1 or LDOC1S. The specificities of the TaqMan assays are calculated as follows:

ΔCt TaqLD1sv (Ct target $_{\text{LD1svTemplate}}$-Ct homolog $_{\text{LD1wtTemplate}}$)= $2^{-(11.012-26.415)}$ = 43,327

ΔCt TaqLD1wt (Ct target $_{\text{LD1wtTemplate}}$-Ct homolog $_{\text{LD1svTemplate}}$)= $2^{-(10.358-34.104)}$ = 14,068,839

TaqLD1: TaqMan assay for corresponding LDOC1 isoform (sv or wt); LD1wt: wild type LDOC1; LD1sv: splice variant of LDOC1
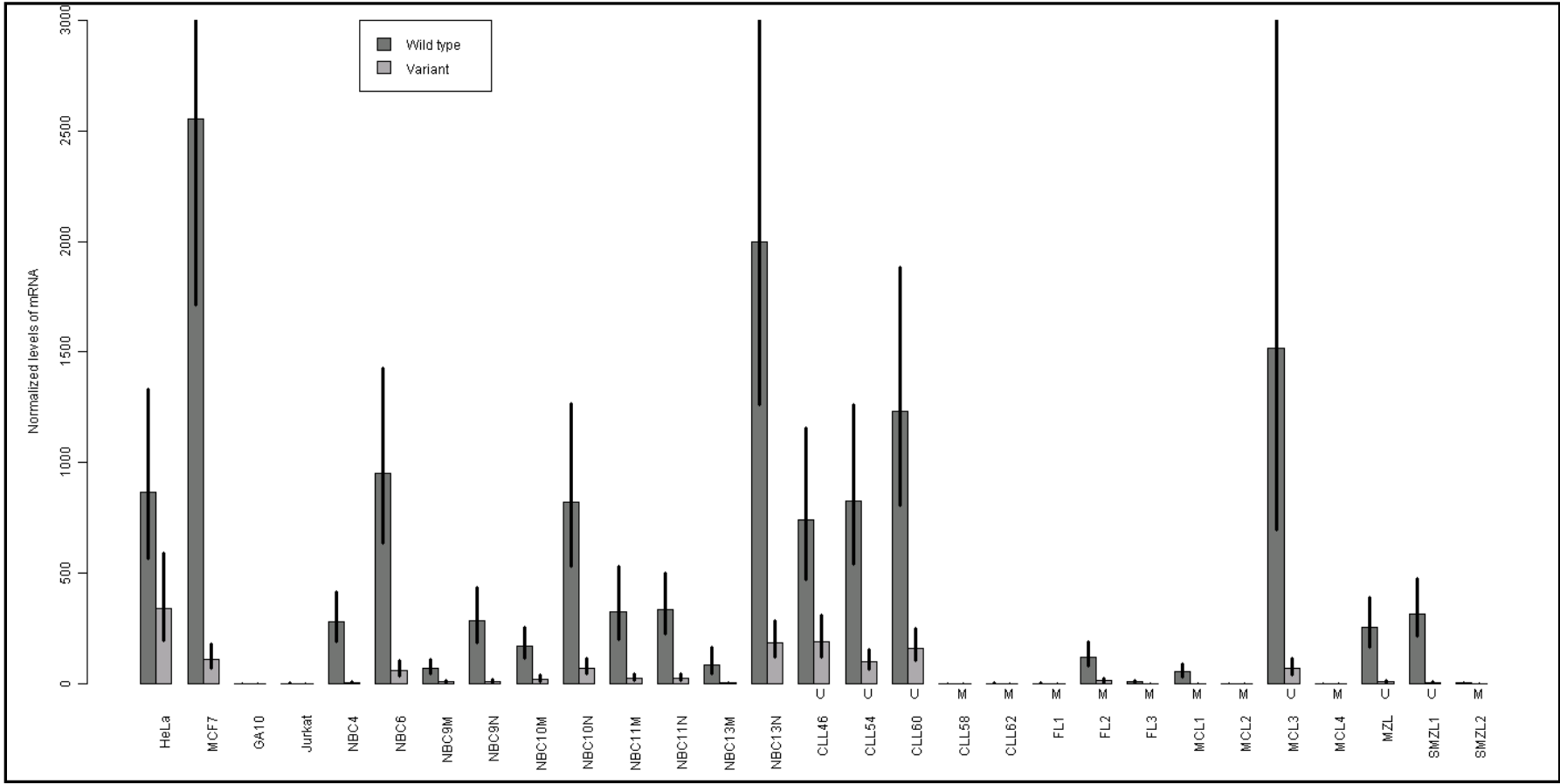
Figure 8

Figure 8: Expression of wild-type *LDOC1* and *LDOC1S* mRNA isoforms measured by QRT-PCR assay. We designed an assay that distinguishes between the isoforms. Gene expression was measured in carcinoma (HeLa, MCF7) and lymphoma (GA10, Jurkat) cell lines, unfractionated normal peripheral blood B cells (NBC4, NBC6), normal peripheral blood B cells enriched for memory B cells (NBC9M, NBC10M, NBC11M, NBC13M), normal peripheral blood B cells enriched for naïve B cells (NBC9N, NBC10N, NBC11N, NBC13N), and unmutated (U) or mutated (M) CLL and primary B-cell lymphoma samples, including follicular lymphoma (FL), mantle cell lymphoma (MCL), marginal zone lymphoma (MZL), or splenic marginal zone lymphoma (SMZL) samples. Error bars represent the standard error of the ΔΔCt values.

3.3 Summary

We evaluated *LDOC1* mRNA expression in subsets of normal and malignant B cells, and in solid tumor cell lines. We discovered a shorter form of *LDOC1*, a novel splice variant that we called *LDOC1S*. If translated into protein, the splice variant would consist almost entirely of the leucine zipper region of the wild type protein, raising the possibility that it might interfere with the function of the wild type protein through dimerization between leucine zipper regions. The *LDOC1S* mRNA is co-expressed with wild type *LDOC1* although its amount might vary between the cell types. However, we have not screened a sufficiently large number of samples to derive a statistically meaningful conclusion about the biological significance of *LDOC1/LDOC1S* ratio.

During our initial screen of 131 CLL samples for differential *LDOC1* mRNA expression and subsequent survival analyses we used an MF-QRT-PCR assay. This assay uses an inventoried TaqMan assay that recognizes both *LDOC1* and *LDOC1S* mRNAs. After we discovered *LDOC1S,* we developed and tested isoform-specific TaqMan assays to determine the contribution of *LDOC1S* to the total *LDOC1* mRNA measured by the MF-QRT-PCR assay. We found that the contribution of *LDOC1S* was very small. Had the expression of *LDOC1S* been significantly higher than that of wild type *LDOC1* we would have screened a larger cohort of CLL patients to measure *LDOC1S* mRNA to evaluate its value as a biomarker of overall survival by itself.

CHAPTER 5. GENE EXPRESSION PROFILING FOLLOWING LDOC1 KNOCK-DOWN IN HeLa CELLS IDENTIFIES POTENTIAL COORDINATELY-REGULATED BIOLOGIC PATHWAYS

5.1   Introduction

LDOC1 is a novel gene of unknown function. We hypothesized that global gene expression profiling would illuminate the biologic processes in which *LDOC1* is involved. We chose HeLa cervical carcinoma cells, which highly express endogenous LDOC1 protein, in order to elucidate changes in the transcriptome induced by LDOC1 protein knock-down.

5.2   Results

5.2.1   Optimization of LDOC1 protein knock-down

There are no previously published studies that demonstrate expression of endogenous LDOC1 protein using Western blot analysis. In order to demonstrate LDOC1 protein expression, we screened polyclonal and monoclonal LDOC1 antibodies from six commercial sources. Despite numerous experiments to optimize the conditions of the assay, we identified only one polyclonal antibody that gave reproducible results (although with some background) in a Western blot assay, which we used for our studies.

Before performing gene expression profiling, we first performed experiments to optimize HeLa cell transfection. As described in Materials and Methods, we transfected

HeLa cells with a siRNA pool that contains siRNAs directed against four different regions of the *LDOC1* mRNA, including sequences within the *LDOC1S* mRNA. The transfected cells were harvested every 24 hours following transfection for 4 days, and LDOC1 protein expression assessed by Western blot analysis using a polyclonal antibody to LDOC1. The results of a representative experiment are shown in **Figure 9**. The cell confluence at 24 (data not shown), 48 (data not shown), 72, and 96 hours was 60%, 95%, 100%, and >100%, respectively. In untransfected cells (lanes 1, 5, and 9) LDOC1 protein was not observed until the cells reached 100% confluence, 72 hours after initiating the cultures, and was maximally expressed when the cells reached >100% confluence, 96 hours after initiating the cultures. By 72 hours after transfection, LDOC1 protein expression was substantially reduced in cells transfected with LDOC1 siRNA, compared to untransfected cells, cells transfected with non-targeting siRNAs, or mock transfected cells. Western blot analysis performed using an antibody to beta-actin demonstrates that equivalent amounts of protein were added to each well.

In the same experiment, we also assessed the transfection efficiency by fluorescence microscopy 24 hours after transfection with fluorescently-labeled siRNA, siGLO Red. The efficiency was calculated by dividing the number of fluorescent cells by the total number of cells (mean of three different fields counted at 40X magnification), and was 92% **(Figure 10).**
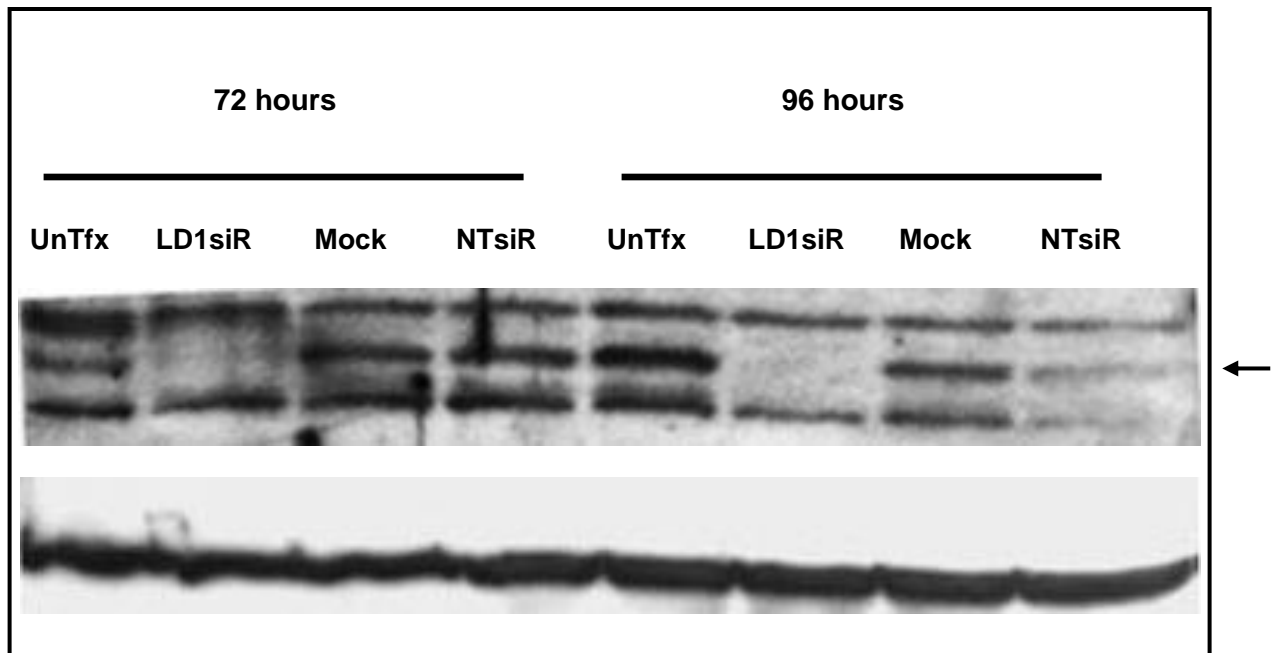
Figure 9. LDOC1 protein is knocked-down by specific LDOC1 siRNAs. Cells were harvested at 72 or 96 hours after transfection and protein lysates (50 μg) were loaded into each well. Western blot analysis was performed using a polyclonal antibody to LDOC1 (upper panel) or a monoclonal antibody to beta actin (lower panel). The LDOC1 protein is 17 kD (band is shown by arrow on upper panel). Abbreviations: UnTfx, untransfected; LD1siR, LDOC1 siRNA pool; Mock, mock transfection (all transfection reagents other than siRNAs); NTsiR, non-targeting siRNA #1.
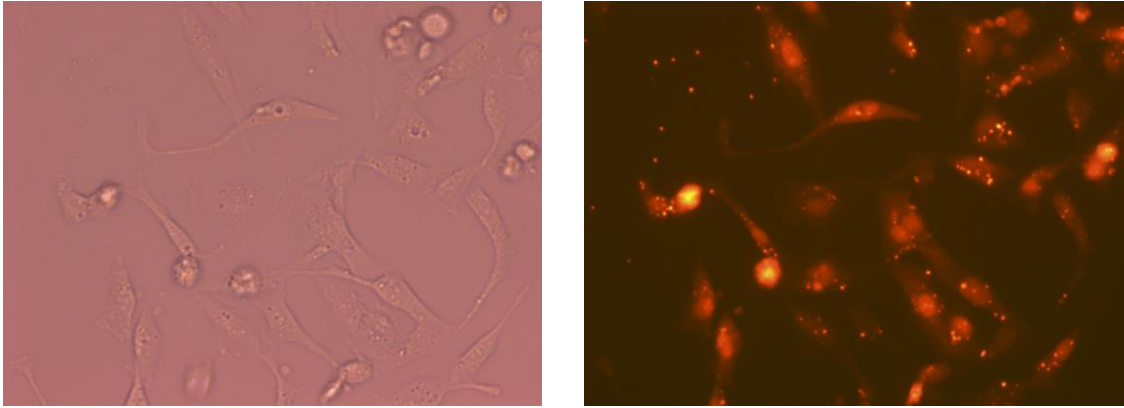
Figure 10. HeLa cells demonstrate high transfection efficiency by siRNA.

HeLa cells transfected by fluorescently labeled siRNA, siGLO Red, were visualized 24

hours after transfection under inverted light microscope (left) or fluorescent microscope

(right) (20X magnification).

5.2.2. LDOC1 protein knock-down for microarray experiment

The results of our optimization experiments indicated that LDOC1 protein was maximally knocked down 96 hours after transfection. Thus we chose to harvest cells at approximately this time point for subsequent gene expression profiling. Knock-down experiments were conducted in triplicate for each condition (untransfected, LDOC1 siRNA transfected, non-targeting siRNA transfected, and mock transfected). At 93 hours after transfection the cells had reached > 100% confluence. They were harvested and evaluated for cell number and viability (trypan blue exclusion), and cell cycle (propidium iodide staining followed by flow cytometry). In addition, protein and RNA were extracted to perform Western blot analysis (to confirm knock-down) and gene expression profiling. With respect to cell number, viability (data not shown), or cell cycle phase we found no difference between HeLa cells transfected with LDOC1 siRNAs and non-targeting siRNAs (**Figure 11 A and B**). We confirmed the high efficiency of LDOC1 protein expression knock-down by Western blot analysis **(Figure 12)**.

A
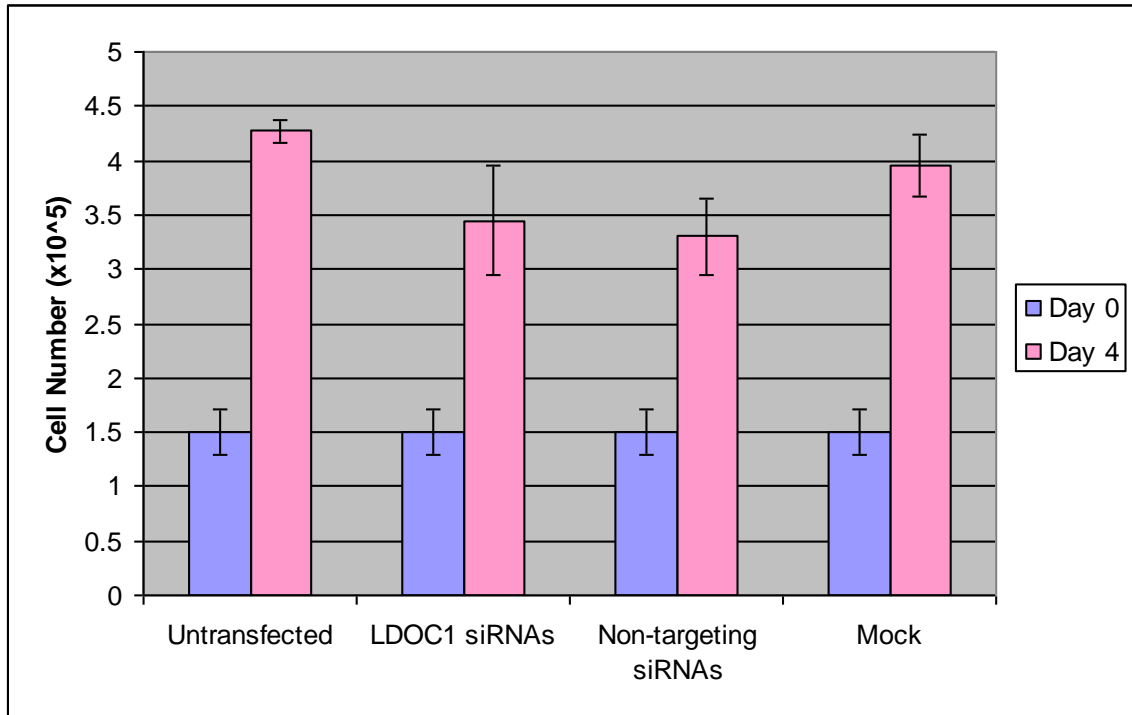
B
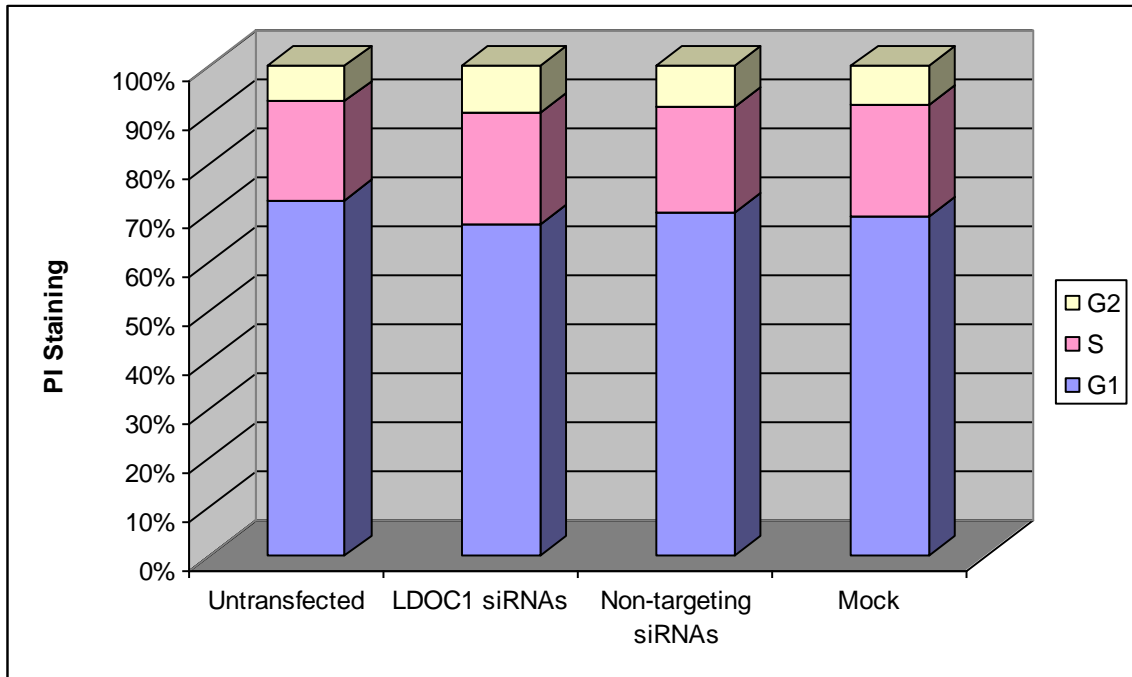
Figure 11. Transfection with LDOC1 siRNA has no significant effect on cell number or cell cycle phase. (A) The cell number was evaluated by trypan blue exclusion and (B) cell cycle phases were evaluated by propidium iodide staining.
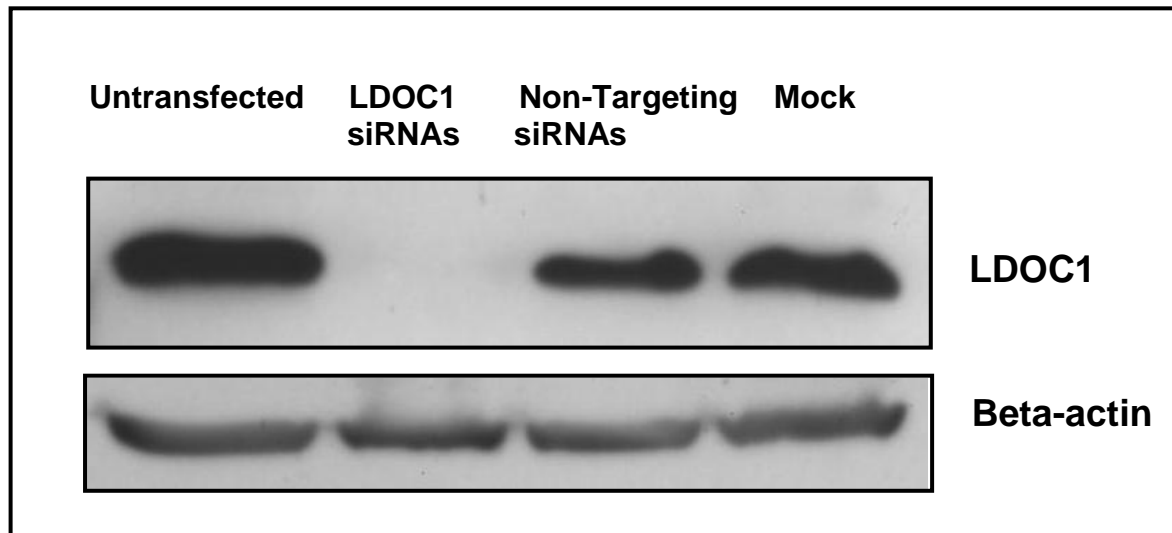
Figure 12. LDOC1 protein knock-down demonstrated by Western blot analysis. Protein was obtained 93 hours after transfection with siRNAs and 100 µg protein lysate was loaded into each well. LDOC1 protein was efficiently knocked down by specific *LDOC1* siRNAs, but was unaffected by non-targeting siRNAs (upper panel). Antibody to beta-actin demonstrates that equivalent amounts of protein (100 µg per lane) were added to each well (lower panel).

5.2.3 Differentially expressed genes in siRNA-transfected HeLa cells

We compared the gene expression profiles on RNA obtained from HeLa cells treated with LDOC1 siRNAs (LD) or non-targeting (NT) siRNAs. The array quality control analysis indicated that there was no significant evidence of outliers or abnormal hybridization patterns, and the overall array quality was good. The percentage of genes present for all the arrays ranged between 47-50% (expected between 30-60%), and the average, minimum, and maximum backgrounds were similar across the arrays.

In order to identify differentially expressed genes between the LD and NT groups, we performed two-sample t-tests. Using false discovery rates (FDR) of 0.05, 0.1, and 0.2, we identified 1, 4, and 107 differentially expressed genes, respectively. The fold value changes and p values of the 107 differentially expressed genes are presented in **Table 3**. In this table, fold changes indicate the ratio LD / NT. A positive fold change value indicates an increase in gene expression in LD compared to NT cells; a negative value indicates a decrease in gene expression in LD compared to NT cells.

Table 3. List of Differentially Expressed Genes in siRNA-transfected HeLa Cells

Probe ID	Gene Symbol	Fold Change	P value
204454_at	LDOC1	-5.24	0.000000144
1556551_s_at	SLC39A6	1.39	0.0000154
214352_s_at	KRAS	1.35	0.0000199
205809_s_at	WASL	1.28	0.0000209
223266_at	STRADB	1.07	0.0000385
211949_s_at	NOLC1	1.27	0.0000494
221539_at	EIF4EBP1	-1.48	0.0000536
210537_s_at	TADA2L	1.22	0.0000723
212113_at	LOC552889	1.23	0.0000877
211212_s_at	ORC5L	1.12	0.000088
209169_at	GPM6B	1.17	0.0000883
212010_s_at	CDV3	-1.27	0.0000927
226264_at	SUSD1	1.43	0.0000968
202132_at	WWTR1	1.20	0.000100038
211503_s_at	RAB14	1.20	0.000109765
226020_s_at	DAB1 /// OMA1	1.11	0.000139154
219529_at	CLIC3	1.09	0.000139573
221059_s_at	COTL1	-1.04	0.000141816
207940_x_at	CNR1	-1.10	0.000156992
228801_at	ORMDL1	1.18	0.000172399
1561403_at	SOHLH1	-1.24	0.000207897
243349_at	KIAA1324	-1.19	0.000210587
1558378_a_at	AHNAK2	1.43	0.000216582
212775_at	OBSL1	-1.07	0.000224804
227639_at	PIGK	1.09	0.000238186
223592_s_at	RNF135	1.13	0.000242704
202412_s_at	USP1	1.16	0.000252094
205315_s_at	SNTB2	1.20	0.000259242
209920_at	BMPR2	1.39	0.000261638
243145_at	---	1.13	0.000273762
226897_s_at	ZC3H7A	1.22	0.000276911
212808_at	NFATC2IP	1.19	0.000278652
206383_s_at	G3BP2	1.13	0.00028099
220144_s_at	ANKRD5	-1.06	0.000288035
217457_s_at	RAP1GDS1	-1.07	0.00029406
204078_at	SC65	1.09	0.000300573
215071_s_at	HIST1H2AC	1.14	0.000302292
224404_s_at	FCRL5	-1.16	0.000305902
241715_x_at	ACPT	-1.19	0.000307205

Probe ID	Gene Symbol	Fold Change	P value
1570130_at	SPATS2	1.22	0.000310473
232503_at	---	-1.16	0.00031898
221734_at	PRRC1	1.09	0.000320222
234724_x_at	PCDHB18	-1.16	0.000322211
1568640_at	ASPRV1	-1.17	0.000324145
236053_at	LOC100128653	-1.12	0.000336741
218576_s_at	DUSP12	-1.08	0.00034521
235266_at	ATAD2	1.34	0.000351744
212906_at	GRAMD1B	-1.10	0.000361663
203491_s_at	CEP57	1.20	0.000364253
237837_at	---	-1.10	0.000367065
56256_at	SIDT2	1.13	0.000369489
1552307_a_at	TTC39C	1.15	0.000370903
218751_s_at	FBXW7	1.14	0.000383996
212451_at	SECISBP2L	1.34	0.000385468
232349_x_at	DCAF6	1.18	0.000386059
221520_s_at	CDCA8	-1.09	0.00040314
224471_s_at	BTRC	1.04	0.000404038
221396_at	TAS2R7	-1.15	0.000412271
226297_at	---	1.15	0.000412343
219066_at	PPCDC	1.14	0.000426938
205226_at	PDGFRL	-1.39	0.000431297
230155_x_at	MSL1	-1.29	0.000436594
1557896_at	---	-1.22	0.00045447
208018_s_at	HCK	-1.14	0.000459992
223692_at	NMNAT1	1.29	0.000462599
239055_at	FLJ43663	-1.22	0.00046488
1556127_at	DIP2A	-1.15	0.00046601
237740_at	---	-1.13	0.000477778
224190_x_at	NOD1	1.14	0.000492446
211587_x_at	CHRNA3	-1.12	0.000511636
241273_at	---	1.13	0.000514578
208378_x_at	FGF5	-1.24	0.000515624
228936_at	---	-1.03	0.000522271
208297_s_at	EVI5	1.28	0.000527051
201959_s_at	MYCBP2	1.12	0.00052857
211801_x_at	MFN1	1.15	0.000529138
1562783_at	LOC100128840	-1.28	0.000533645
201562_s_at	SORD	1.21	0.000534678

Probe ID	Gene Symbol	Fold Change	P value
229274_at	GNAS	1.35	0.00053949
214869_x_at	GAPVD1	1.09	0.000546859
227780_s_at	ECSCR	-1.31	0.00054713
222975_s_at	CSDE1	1.13	0.000548207
211993_at	WNK1	1.38	0.000549098
1555057_at	NDUFS4	1.17	0.000562023
1569846_at	---	-1.20	0.000573165
239788_at	---	1.36	0.000587081
226413_at	LOC400027	1.10	0.000592997
202658_at	PEX11B	-1.23	0.000599795
200598_s_at	HSP90B1	1.29	0.00061511
229795_at	---	1.16	0.000629294
203056_s_at	PRDM2	1.40	0.000646625
233261_at	EBF1	1.32	0.000655267
217988_at	CCNB1IP1	1.11	0.00065986
200841_s_at	EPRS	1.44	0.00066462
219658_at	PTCD2	1.20	0.000665816
200975_at	PPT1	1.09	0.000671499
222142_at	CYLD	1.08	0.000681179
1554372_at	---	-1.19	0.000691453
209254_at	KLHDC10	1.25	0.000700331
1558747_at	SMCHD1	1.46	0.000718725
1568957_x_at	SRGAP2P1	1.11	0.000721806
220653_at	ZIM2	-1.25	0.000734443
1568931_at	---	-1.17	0.000743993
212520_s_at	SMARCA4	1.16	0.000753233
229909_at	B4GALNT3	1.08	0.000758928
201479_at	DKC1	-1.06	0.000761875
226568_at	FAM102B	-1.06	0.000762703
233827_s_at	SUPT16H	1.36	0.000766658
1556336_at	LOC100131735 /// LOC100291994 /// RBMX /// RBMXL1	1.19	0.000766725
235201_at	---	1.09	0.000784229
239009_at	KIAA0754	-1.26	0.000787619
1570165_at	---	-1.06	0.000789026
212220_at	PSME4	1.23	0.000801126
235059_at	RAB12	1.28	0.000809335
217753_s_at	RPS26	-1.06	0.000814191
1553542_at	CCDC125	1.26	0.000814995
213286_at	ZFR	1.31	0.000822981
208209_s_at	C4BPB	-1.12	0.000837838

Probe ID	Gene Symbol	Fold Change	P value
202197_at	MTMR3	1.14	0.000855254
227087_at	INPP4A	1.14	0.000856684
236806_at	---	1.12	0.000868555

5.2.4 Networks constructed from gene expression signature

The differentially expressed genes listed in **Table 3** belong to a variety of functional categories, including enzymes (*BMPR2*, *BTRC*, *GNAS*, *KRAS*, and *RAB14*), kinases (*HCK*, *PDGFRL*, and *STRADB*), and transcriptional regulators (*FBXW7*, *PRDM2*, and *SMARCA4*). To better understand the potential function of these genes and their relationships, we performed a network analysis using Ingenuity Pathways. A Core Analysis was performed using a 0.5 fold change value as the cut-off, querying only direct interactions between the 107 most significantly expressed genes. As shown in **Figure 13**, LDOC1 protein knock-down induced alterations in expression of a variety of genes that participate in wide range of biological processes, including regulation of gene expression, cellular function and maintenance, cancer, cell cycle, cellular growth and proliferation, cellular assembly and organization, cell death, and DNA replication, recombination and repair. To illustrate the construction of networks, Network # 2 is shown in **Figure 14**.

ID	Molecules in Network	Score	Focus Molecules	Top Functions
1	ASGR2, ATG3, ↑BTRC, CLPX, DBI, ↑DCAF6, DCAF8, DDX10, DLG4, ↓DUSP12, FNTB, ↑G3BP2, GIT2, HNF1A, HNF4A, KCNA5, KCNJ2, KCNJ10, KCNJ12, KRR1, MTTP, ↑NDUFS4, ↑NOD1, NRD1, ↓OBSL1, ↑ORMDL1, ↓PEX11B, ↑PPCDC, PPCS, PRPF38B, ↑SLC39A6, ↑SNTB2, SUPT5H, ↑WASL, ↑ZFR	25	14	Cardiac Arrythmia, Cardiovascular Disease, Gene Expression
2	26s Proteasome, ARRB1, ↑BMPR2, C2orf44, CBL, ↓DKC1, ↑EBF1, EGFR, ↑EPRS, ↑EVI5, ↑GAPVD1, ↑GPM6B, ↓HCK, ↑KRAS, KRT18, LIMK1, LRP1B, MGEA5, ↑NOLC1, NOVA1, OC90, PAK1, PLK1, PRPF4B, PSMA1, ↑PSME4, PSMF1, PTPN2, QARS, ↓RAP1GDS1, RIN1, RIN2, TSR1, YWHAZ, ↑ZC3H7A	22	13	Cellular Function and Maintenance, Cancer, Gastrointestinal Disease
3	ARL6IP1, ↑ATAD2, CANX, ↑CCNB1IP1, CDK1, CEBPE, ↑CEP57, EXOSC1, EXOSC7, EXOSC8, FAM107A, ↑FBXW7, FOSL1, Gcn5l, ↑INPP4A, ↓LDOC1, MAP4, ↓MSL1, ↑MTMR3, MYC, ↑MYCBP2, Npm1, NRIP1, ↑PIGK, RNA polymerase I, SMAD1, SNIP1, SUPT3H, ↑TADA2A, TAF12, TAF5L, TAF6L, TOPBP1, TP53BP1, ZBTB16	18	11	Cell Cycle, Gene Expression, Cellular Growth and Proliferation
4	BAK1, BAX, CCR4, CDC25C, CLOCK, ↓CNR1, ↑CSDE1, E2F1, E2F3, ↑HIST1H2AC, HIST3H3, HLTF, ING1, KAT2B, ↑MFN1, MFN2, MYCN, ↑NMNAT1, PARP1, PCNA, PDPK1, poly(ADP-ribose), ↑PPT1, ↓RPS26, SIRT1, SLC9A3R2, ↑SORD, STRAP, TEAD1, TEAD3, TNRC6A, ↑USP1, ↑WNK1, ↑WWTR1, ZYX	18	11	Cellular Assembly and Organization, Cardiac Necrosis/Cell Death, Cell Death
5	↑AHNAK2, BMP2, ↓C4BPB, CAB39, CALCR, CCNE2, CTGF, ↑CYLD, ↓DIP2A, ↓EIF4EBP1, Estrogen Receptor, Fgf, ↓FGF5, Fgfr, FST, FSTL1, GNAL, ↑GNAS, heparin, HIC1, ↑HSP90B1, IFN Beta, INHBA, LRP1, LRPAP1, PP2A, PPP2R2A, ↑PRDM2, ↑SMARCA4, STK11, ↑STRADB, TG, TGFB1, TNFSF13, VWF	18	11	Gene Expression, Cell Death, Tissue Development
6	CDC7, CDC45, ↓CDCA8, ↓CHRNA3, CHRNA5, ↓COTL1, DBF4, H3F3A/H3F3B, HIST2H2BE (includes others), Histone h3, Histone h4, ↑KLHDC10, MCM3, MCM4, MCM5, MT1F, ↑NFATC2IP, NUDCD3, ORC3, ORC4, ↑ORC5, ORC6, poly(ADP-ribose), ↑PTCD2, ↑RAB14, RCC2, RNF4, SOX10, SSRP1, ↑SUPT16H, TAGLN, TFIIE, TRAF4, UBE2I, VHL	14	9	DNA Replication, Recombination, and Repair, Cardiovascular Disease, Cancer

Figure 13. Ingenuity Pathways (IPA) Core Analysis reveals functional categories of differentially expressed genes. Of 16 networks, only the most significant 6 networks are shown from the IPA analysis. Genes in bold-face type are focus molecules that we identified as differentially expressed **(Table 3)** and were uploaded into the Core Analysis by the investigator. Genes in non-bold-face type are identified by the Ingenuity Knowledge Base as potential interaction partners and incorporated into networks. Red arrows indicate up-regulated molecules and green arrows indicate down-regulated molecules.

Figure 14. IPA Network 2 demonstrates potential interactions between genes that participate in cellular function and maintenance, cancer, and gastrointestinal disease. The shapes of gene symbols indicate their functional group. Solid lines indicate the direct interactions. Numerical values under the genes indicate mRNA fold changes in LDOC1 knocked-down cells compared to the NT control cells. Red indicates up-regulation and green indicates down-regulation of the mRNA of the respective gene. Genes are placed according to their location in the cell, i.e., extracellular space, plasma membrane, cytoplasm, or nucleus; genes for which the location is unknown are located on the right ("unknown").

Figure 14
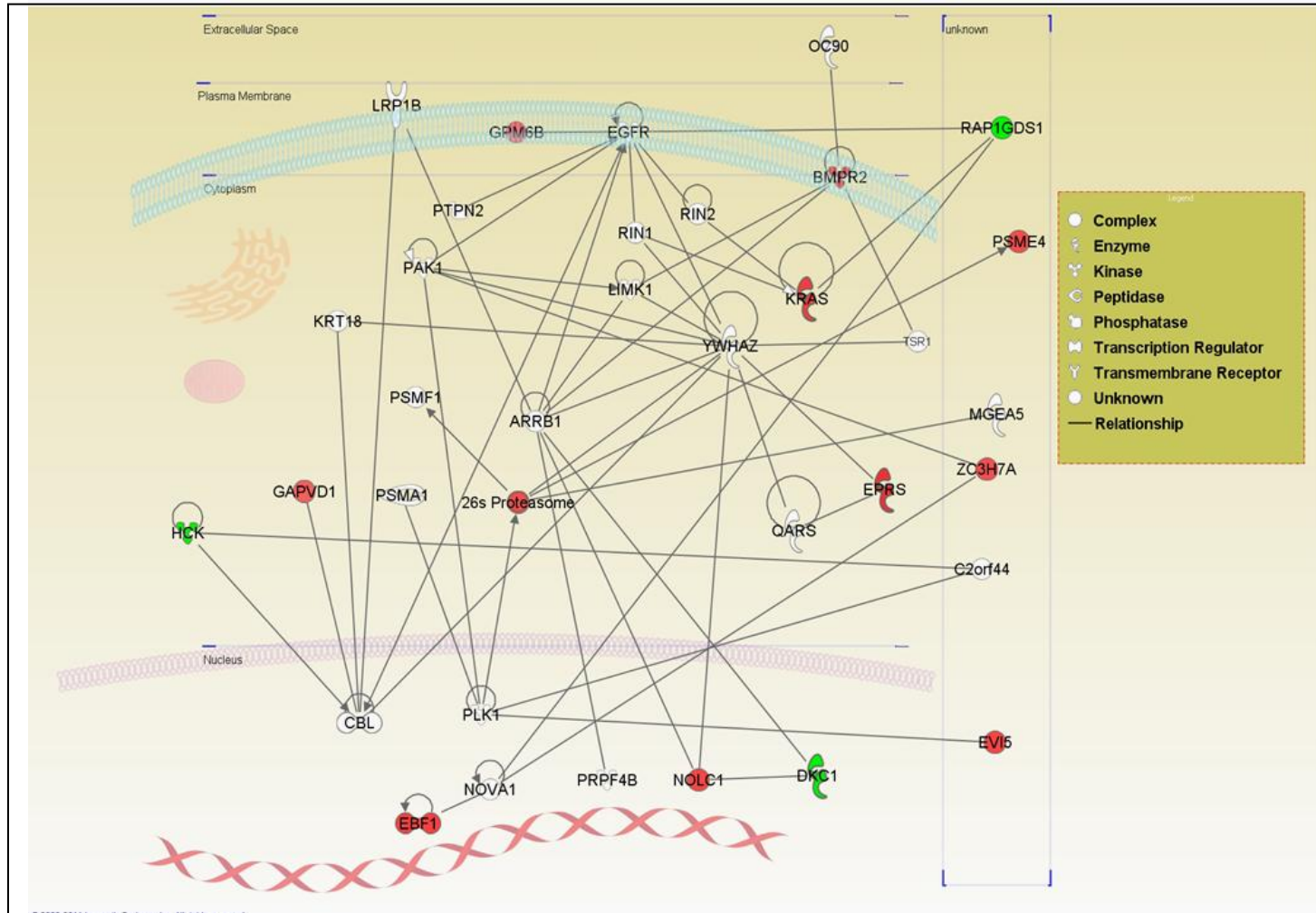
5.3　Summary

We have successfully knocked-down LDOC1 protein in HeLa cells and performed gene expression profiling microarray analysis. This analysis revealed 107 genes that are significantly differentially expressed between the cells transfected with *LDOC1* siRNAs and non-targeting control siRNAs. These genes belong to a wide variety of functional groups. The list of differentially expressed genes were subjected to further analysis by performing a Core Analysis in Ingenuity Pathways, which allowed us to assign the differentially expressed genes into potential networks and visualize interactions between the differentially expressed genes within functional categories. The IPA analysis also revealed other potential interaction partners extracted from the Ingenuity Knowledge Base, in addition to the 107 genes identified by gene expression profiling experiment. Upon inspection of the networks, *LDOC1* was found only in network 3. Relatively little is known about the function of LDOC1, and none of the molecules indicated as interaction partners of LDOC1 in this network have been shown experimentally to interact with LDOC1 in a mammalian system. Ingenuity Pathways constructs networks using all available published information, including bioinformatic analyses as well as experimental data, from the literature and databases. Thus, the results require validation in different tissue types and cellular contexts.

CHAPTER 6. INTERSECTION AND SUBSEQUENT VALIDATION OF COORDINATELY DIFFERENTIALLY EXPRESSED GENES IN HeLa CELLS AND CLL SAMPLES

6.1 Introduction

Although gene expression profiling using microarrays allows the expression of a large number of genes to be assessed in a single experiment, the dynamic range of microarrays is relatively small (39, 40). Thus, genes identified as differentially expressed using microarrays should be validated using a more sensitive method, such as a QRT-PCR assay. We chose to validate genes that we identified as "coordinately regulated" in LDOC1 knock-down experiments in HeLa cells and in CLL samples. We approached this problem in two ways.

First, we compared two lists of differentially expressed genes generated by gene expression profiling microarray studies using Affymetrix arrays: (1) the list of genes differentially expressed between HeLa cells transfected with LDOC1 siRNAs (to knock-down LDOC1 expression) and non-targeting siRNAs, and (2) the list of genes differentially expressed with respect to *LDOC1* mRNA expression in 30 CLL samples, obtained from a previous gene expression profiling microarray (34). Genes that were concordant for LDOC1 expression in both groups (HeLa and CLL) and whose function suggests that they might have relevance to the pathophysiology of CLL were to be evaluated by QRT-PCR assay.

Second, we compared the two lists of differentially expressed genes generated by a MF-QRT-PCR assay for 43 candidate biomarkers of prognosis and 5 endogenous

control genes; 23 potential biomarkers have been reported previously (22) and 20 have been identified subsequently (Abruzzo LV, et al., manuscript submitted). We had previously assessed 76 CLL samples for expression of these 48 genes (including LDOC1). We performed the same MF-QRT-PCR assay on RNA obtained from transfected HeLa cells, as described above.

6.2   Results

In the gene expression profiling experiments of HeLa cells using Affymetrix microarrays, the groups with "high" and "low" LDOC1 expression corresponded to cells that had been transfected with non-targeting siRNAs or LDOC1 SiRNAs, respectively. For the CLL samples, we reanalyzed the Affymetrix microarray raw data to allocate samples into groups with "high" and "low" *LDOC1* mRNA expression. To achieve this, we first plotted *LDOC1* mRNA expression levels for each sample to determine if there was a clear separation between samples with respect to *LDOC1* mRNA expression (data not shown). Since this analysis failed to separate the samples into two distinct groups, we then plotted the Affymetrix expression data against the MF-QRT-PCR assay data collected previously in our laboratory (**Figure 15**). Based on this analysis, we concluded that the expression value of 7.6 on the log2 scale defined the best cut-off to separate the groups with *LDOC1*-high from *LDOC1*-low expression for Affymetrix data (shown as the black vertical line in the figure). Differentially expressed genes between the *LDOC1*-high and *LDOC1*-low groups were determined using two sample t-tests.
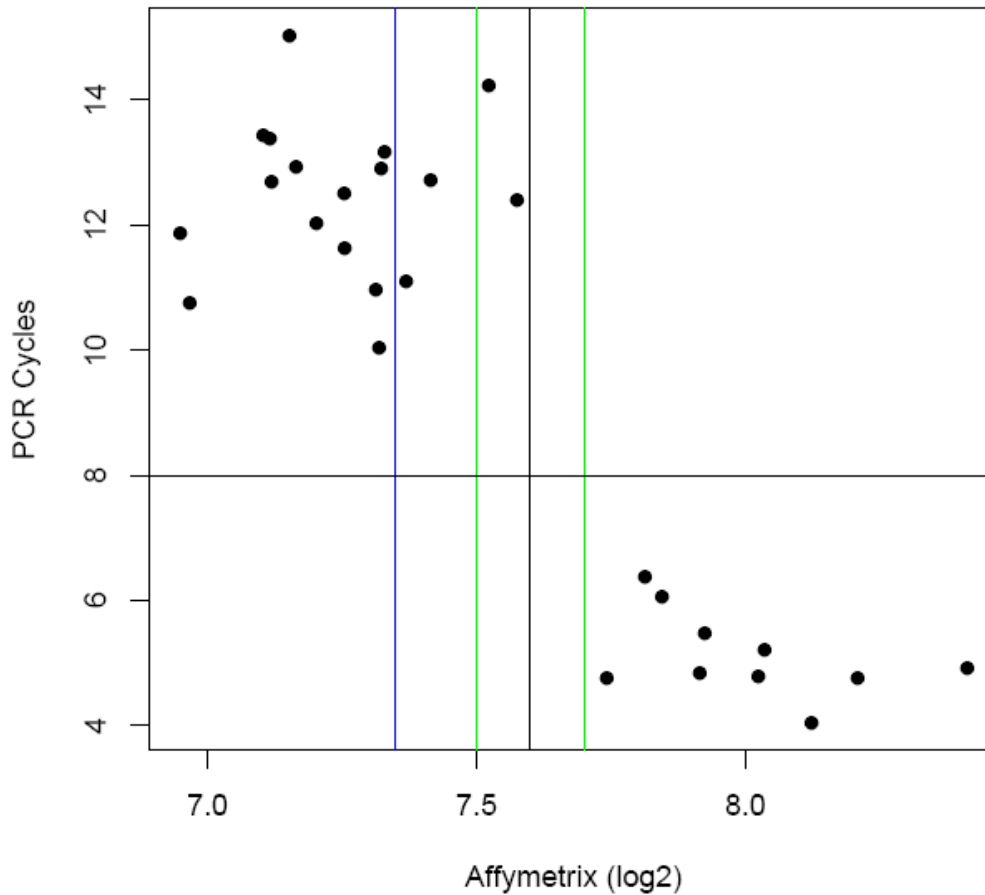
Figure 15. Scatter plot comparing Affymetrix expression estimates against QRT-PCR cycles defines the cut-off value for *LDOC1* mRNA expression on Affymetrix arrays. Vertical lines indicate plausible cut-offs to separate *LDOC1* mRNA expression into groups with high (right of the line) or low expression (left of the line), based on gene expression profiling Affymetrix data. The horizontal line separates *LDOC1* mRNA expression into groups with high (below the line) or low expression (above the line) based on the MF-QRT-PCR data. The expression value of 7.6 on the log2 scale defined the best cut-off to separate the groups with *LDOC1*-high from *LDOC1*-low expression for Affymetrix data (shown as the black vertical line in the figure). PCR cycles = Mean ΔCt value based on the MF-QRT-PCR assay. The colored vertical lines indicate

alternative cutoff values to set the threshold for LDOC1 mRNA positivity measured by Affymetrix microarrays.

Next we compared the list of differentially expressed genes in CLL groups and HeLa cell groups with "high" and "low" LDOC1 expression. To narrow down the number of genes for validation, we selected concordantly expressed genes for which the p value in CLL and HeLa cells was < 0.05. The overlapping, differentially expressed genes are listed in **Table 4**.

Table 4. Intersection of Differentially Expressed Genes Between CLL Samples and HeLa Cells. The CLL and HeLa p values refer to the significance of differential expression for that gene between LDOC1-high and LDOC1-low samples. The p combined value refers to the significance of gene expression change in the CLL and HeLa data when they are assessed simultaneously (Fisher's exact test). For CLL cells the fold change values indicate the gene expression ratios of cases with high LDOC1 / low LDOC1; for HeLa cells the fold change values indicate the gene expression ratios of NT (high LDOC1) / LD (low LDOC1) cells. A positive value indicates an increase in gene expression; a negative value indicates a decrease in gene expression. Genes that are concordant show the same direction of change in CLL and HeLa samples.

AffyProbeID	CLL P value	CLL Fold Change	HeLa P value	HeLa Fold Change	P Combined	Concordant
204454_at	7.17E-11	1.653	1.44E-07	5.253	4.44E-16	Yes
217753_s_at	0.041280928	1.692	0.000814191	1.059	0.000379822	Yes
219024_at	0.011023732	0.745	0.003040962	0.736	0.000378917	Yes
201676_x_at	0.045090524	1.095	0.003547553	1.065	0.001558113	Yes
216222_s_at	0.019039041	0.881	0.003716954	0.548	0.000747027	Yes
208794_s_at	0.007330058	1.278	0.003920538	1.039	0.000329257	Yes
202944_at	0.041408589	0.875	0.004546794	0.883	0.001803235	Yes
48659_at	0.039342604	1.088	0.004699016	1.062	0.001773999	Yes
214452_at	3.93E-06	1.246	0.005504185	1.285	4.04E-07	Yes
205780_at	0.025449019	1.404	0.00559713	1.294	0.001403986	Yes
202678_at	0.005916584	1.192	0.005796159	1.080	0.000386849	Yes
200881_s_at	0.038323819	1.173	0.005981868	1.043	0.002150508	Yes
202600_s_at	0.019171443	0.649	0.00651801	0.803	0.001248037	Yes
205882_x_at	0.031204126	0.600	0.007339724	0.960	0.002148678	Yes
207031_at	0.044814279	0.926	0.007569084	0.878	0.003049066	Yes
213017_at	0.028034783	1.148	0.007682522	1.112	0.002033838	Yes
201524_x_at	0.041995482	1.149	0.010959026	1.066	0.003996535	Yes
214474_at	0.040485963	0.900	0.011108058	0.799	0.003915664	Yes
205469_s_at	0.00720127	1.260	0.013697712	1.193	0.001008507	Yes
200834_s_at	0.001163252	0.891	0.013924734	0.984	0.000194872	Yes
219411_at	0.027896294	1.107	0.014487817	1.241	0.003562117	Yes
216037_x_at	0.035630998	0.851	0.015468312	0.966	0.004686716	Yes
214136_at	0.003592161	0.899	0.01557867	0.835	0.000603868	Yes
213475_s_at	0.02024539	1.260	0.016225922	1.177	0.002963391	Yes
221579_s_at	0.001444114	1.166	0.017597489	1.075	0.000294286	Yes
204476_s_at	0.028243382	0.887	0.019006562	0.920	0.004578915	Yes
202350_s_at	0.001019723	0.924	0.019681207	0.871	0.000237146	Yes
212288_at	0.037095652	0.817	0.020512879	0.850	0.006225207	Yes
205280_at	0.038571401	0.949	0.02161636	0.863	0.00674485	Yes
216576_x_at	0.02880998	2.351	0.022686296	1.191	0.005446398	Yes
217835_x_at	0.041520332	1.118	0.023418633	1.034	0.007716366	Yes
200629_at	0.048204048	1.376	0.023496573	1.151	0.0088155	Yes
201798_s_at	0.006273149	0.902	0.023795307	0.930	0.001464315	Yes
201969_at	0.002178054	0.831	0.024250526	0.942	0.000573014	Yes
203767_s_at	0.019546553	0.954	0.024951662	0.924	0.004206952	Yes
204520_x_at	0.008232359	1.212	0.025267505	1.070	0.001971513	Yes
212744_at	0.049204322	0.930	0.025816847	0.843	0.009741301	Yes
201460_at	0.017569059	1.240	0.026763556	1.090	0.004073119	Yes
217418_x_at	0.040932625	0.707	0.030087019	0.936	0.009482234	Yes
203988_s_at	0.001676991	0.706	0.031232095	0.944	0.000568649	Yes
219128_at	0.014004784	0.843	0.032302933	0.959	0.003936273	Yes
212162_at	0.027990074	0.874	0.035029663	0.888	0.007772744	Yes
202669_s_at	0.025205352	0.909	0.035307156	0.829	0.007141121	Yes
200002_at	0.047204876	0.918	0.036583127	0.994	0.012712462	Yes
218571_s_at	0.023344055	1.136	0.04066558	1.035	0.007556221	Yes
218145_at	0.006111609	1.261	0.04164203	1.148	0.002360794	Yes
41858_at	0.017332907	0.930	0.041647164	0.975	0.00594361	Yes
78383_at	0.039295836	0.933	0.042754313	0.950	0.01241388	Yes
214949_at	0.034035352	1.280	0.044100829	1.089	0.011259854	Yes
205149_s_at	0.011042873	1.138	0.044118806	1.160	0.004202982	Yes
207149_at	0.030557123	0.930	0.04499689	0.843	0.010435128	Yes
216401_x_at	0.030799181	2.376	0.045939624	1.066	0.010697648	Yes
212350_at	0.026389523	1.256	0.047558577	1.052	0.0096395	Yes
210087_s_at	0.014594328	0.897	0.048712568	0.967	0.005864396	Yes
216699_s_at	0.030924111	1.268	0.049976695	1.105	0.011548508	Yes

Of the 51 statistically significantly differentially expressed genes between CLL and HeLa groups we performed a review of the literature to identify genes whose reported function suggested that they might contribute to the pathogenesis of CLL. We selected six for validation by QRT-PCR assay: *NRIP1* (Hs00942766_s1, Applied Biosystems), *DNAJA1* (Hs00266011_m1, Applied Biosystems), *UBE2N* (Hs00854751_s1, Applied Biosystems), *ITGAL* (Hs00158218_m1, Applied Biosystems), *MAPKAPK2* (Hs00358962_m1, Applied Biosystems), and *CLCN4* (Hs00156541_m1, Applied Biosystems). None of these genes were significantly differentially expressed between HeLa groups. Therefore, we did not proceed with validation in CLL samples.

For the second approach, we assessed the HeLa groups (LD and NT) for the expression of 43 genes previously identified as potential biomarkers of prognosis in CLL (22). The format of the MF-QRT-PCR assay (microfluidics card) is presented in **Figure 16**. The cDNA prepared from the same RNA used for the gene expression profiling microarray experiments was assessed. The microfluidics card contained eight ports; each port contains TaqMan assays that recognize the 48 different genes. The card was loaded with cDNA from triplicate cultures (three LD and three NT cultures, one sample from each culture in duplicate). The analysis identified three differentially expressed genes with a p value of less than 0.05: *LDOC1*, *GFI1*, and *FOXO1* (**Table 4**).

Figure 16. Topographic view of the QRT-PCR assays for the genes printed on microfluidics card. The card includes 8 ports for loading 8 samples (on the left vertical side). Each sample is evaluated by QRT-PCR assay for 48 genes, extended to two rows. The genes *18S rRNA*, *GAPD*, *PGK1*, *GUSB*, and *ECE-1* served as endogenous controls.

Gene	Fold Change	T stats	P value
LDOC1	6.14	24.75	0.00002
GFI1	2.79	8.37	0.00111
FOXO1	1.21	4.62	0.00986
TRIB2	1.34	2.42	0.07292
AURKA	1.11	2.32	0.08075
FGL2	0.28	-1.88	0.13397
NUDC	0.86	-1.85	0.13782
GLI1	0.72	-1.78	0.14972
AICDA	0.18	-1.73	0.15958
18S	0.77	-1.68	0.16744
LPL	1.07	1.64	0.17612
ATRX	1.98	1.57	0.19128
NT5C2	0.66	-1.50	0.20780
LASS6	1.30	1.46	0.21713
EGR3	1.75	1.37	0.24332
SIRT1	0.92	-1.33	0.25569
P2RX1	1.90	1.32	0.25755
CCL5	0.90	-1.31	0.25888
SLAMF1	0.15	-1.31	0.26063
ATF4	1.08	1.19	0.30134
GAPDH	1.14	1.11	0.33028
OAS3	1.33	1.03	0.36025
SEPT10	0.94	-0.98	0.38411
CD14	1.26	0.96	0.39318
PGK1	1.10	0.93	0.40309
RIOK2	0.82	-0.93	0.40489
SKI	0.85	-0.89	0.42531
ZAP70	1.07	0.75	0.49723
TNFRSF8	2.94	0.70	0.52318
CD86	2.44	0.61	0.57745
PRAME	1.06	0.59	0.58430
NRIP1	1.05	0.55	0.61246
MLXIP	0.81	-0.54	0.62055
CRY1	1.05	0.45	0.67826
BCL7A	0.91	-0.44	0.67990
GUSB	1.02	0.44	0.68507
CHEK1	1.05	0.40	0.71118
ANXA2	0.98	-0.36	0.73460
FLNB	1.02	0.33	0.75607
COBLL1	0.93	-0.32	0.76576
WSB2	1.02	0.26	0.80527
BANK1	0.98	-0.25	0.81418
GZMK	1.18	0.21	0.84117
TPST2	0.99	-0.15	0.88745
ECE1	1.01	0.14	0.89667
ZBTB20	1.01	0.10	0.92558
FGFR1	1.00	0.08	0.93820

Table 5. The expression fold change values and significance of the genes in HeLa samples that are run on microfluidics cards. The T statistics show the direction of the expression change relative to *LDOC1*. A positive value indicates an increase in gene expression; a negative value indicates a decrease in gene expression. Genes that are concordant show the same direction of change in CLL and HeLa samples. Internal controls are *18S rRNA*, *GAPD*, *PGK1*, *GUSB*, and *ECE-1* genes.

Next, the CLL data that were previously collected using this card from 76 patients were grouped into LDOC1-high (27 patients) and LDOC1-low (49 patients) categories, and a two-group t-test was applied to determine the genes that were concordantly differentially expressed. We identified 30 differentially-expressed genes between the LDOC1-high and LDOC1-low CLL groups. When we compared this list of differentially expressed genes with the list obtained from our MF-QRT-PCR analysis of the HeLa samples, only *LDOC1* and *GFI1* showed concordant differential expression.

6.3   Summary

In this section, we used two approaches to identify and subsequently validate genes with concordant LDOC1 expression in transfected HeLa cells and CLL samples. First, we compared the gene expression profiling data obtained using Affymetrix microarrays from the current HeLa experiment and a previously-performed CLL study in our laboratory (34). This approach identified 51 genes in common. We attempted to validate, by QRT-PCR assay, 6 genes from this list whose function suggested that they might contribute to oncogenesis: *NRIP1*, *DNAJA1*, *UBE2N*, *ITGAL*, *MAPKAPK2*, and *CLCN4.*   However, we were unable to validate any of these genes as differentially

expressed in HeLa cells that are transfected with *LDOC1* siRNAs or non-targeting siRNAs. Second, we compared the gene expression data for 43 candidate biomarkers of prognosis obtained by the MF-QRT-PCR assay from 76 CLL samples with the transfected HeLa cells. This analysis identified *GFI1* (Growth Factor-Independence 1) as concordantly differentially expressed in both transfected HeLa cells and CLL samples.

CHAPTER 7. DISCUSSION

We previously identified *LDOC1* as one of the most significantly differentially expressed genes in untreated CLL patients (22). This dissertation extends that study and addresses two major questions with respect to *LDOC1*: 1) Does *LDOC1* mRNA expression correlate with other clinical parameters of prognosis and patient outcome, thus serve as a novel biomarker of survival in untreated CLL patients?, and 2) How does *LDOC1* mRNA upregulation in unmutated CLL contribute to disease pathogenesis?

To address the first question, we have expanded our knowledge of *LDOC1* mRNA expression in neoplastic and benign B cells (26). First, we have confirmed that *LDOC1* mRNA is dramatically down-regulated in mutated CLL cases compared with unmutated cases in a larger patient cohort. It is possible that differences in *LDOC1* mRNA expression may be related to the enhanced ability of unmutated CLL cases to respond to proliferative stimuli and resist apoptosis compared to mutated CLL cases (41). Second, we show that high levels *LDOC1* mRNA expression correlate with cytogenetic markers of poor prognosis and with high ZAP70 expression. Further, we demonstrate that *LDOC1* mRNA expression is an excellent predictor of overall survival in previously-untreated CLL patients. Third, although the sample size is small, we find that *LDOC1* mRNA expression is associated with unmutated *IGHV* somatic mutation status in other primary small B-cell lymphomas. Finally, we show that *LDOC1* mRNA is expressed in normal peripheral blood B cells, and that its expression is higher in the naïve B cell than in the memory B cell fraction. A recent study that evaluated the gene expression profiles of human cord blood subpopulations identified *LDOC1* mRNA as upregulated in the CD34+/CD133+ subpopulation, which contains hematopoietic stem cells and progenitor cells, compared to the more mature CD34−/CD133− subpopulation

(42). During the course of our study, we also discovered a novel splice variant, *LDOC1S* that is co-expressed with *LDOC1* mRNA in normal B cells, CLL and primary lymphoma cells that we examined. These findings suggest that during the course of normal B-cell development *LDOC1* mRNA levels may vary with the maturational stage and state of activation. An assessment of B cells obtained from different compartments and subjected to a variety of different stimuli is required to address this question.

The presence or absence of somatic mutations in the *IGHV* genes separates CLL patients into two prognostic subsets (4, 43). Patients with unmutated *IGHV* genes have a median survival of 8 years compared to 25 years in patients with mutated *IGHV* genes. Gene expression profiling studies using microarrays demonstrated that the majority of unmutated CLL cases express ZAP70 mRNA (13). Subsequently, others showed that expression of ZAP70 protein correlates with mutational status and clinical outcome (14, 15). Recent studies suggest that ZAP70 protein expression may be a better predictor of time to treatment than *IGHV* somatic mutation status (16, 17). In our study, *LDOC1* mRNA expression, *IGHV* somatic mutation status, and ZAP70 protein expression all predicted time to treatment in univariate analyses. In multivariate analyses, *IGHV* somatic mutation status performed marginally better than ZAP70 protein expression, which performed marginally better than LDOC1 mRNA expression (data not shown). However, our data indicate that expression of *LDOC1* mRNA may predict overall survival better than either *IGHV* somatic mutation status or ZAP70 protein expression in previously untreated CLL patients.

Chromosomal abnormalities, predominantly gains and losses, are strong independent predictors of prognosis in CLL. The most common abnormality is del(13)(q14), followed by del(11)(q22.3), the site of the *ATM* gene, trisomy 12,

74

del(6)(6q21-q23), and del(17)(p13), the site of the *TP53* gene (7-9). In the clinical setting these abnormalities are usually assessed using a panel of fluorescence in situ hybridization (FISH) probes. As the sole abnormality, del(13)(q14) is associated with a good prognosis. In contrast, del(6)(q21-q23), del(11)(q22.3), del(17)(p13), del(13)(q14) with other abnormalities, and trisomy 12, are associated with more rapid disease progression and inferior survival. The abnormalities del(17)(p13) and del(11)(q22.3) are the most important independent cytogenetic markers of poor prognosis. Deletions in 17p13, the site of the *TP53* gene, are associated with resistance to therapy with purine analogs, such as fludarabine, and short survival (10-12). The *IGHV* somatic mutation status is associated with cytogenetic abnormalities detected by FISH analysis (44-46). For example, del(13)(q14) is found more often in mutated cases. Our results indicate that *LDOC1* mRNA expression was associated with cytogenetic markers of prognosis. Cases that were *LDOC1* mRNA negative were more likely to contain isolated del(13)(q14), while *LDOC1* mRNA positive cases were more likely to demonstrate cytogenetic markers of poor prognosis. By SNP genotyping, no case showed loss of the *LDOC1* gene; three mutated cases showed a gain in the *LDOC1* copy number, but were negative for *LDOC1* mRNA expression (data not shown). Further, we found no mutations in *LDOC1* mRNA in the subset of CLL cases that we subjected to sequence analysis. Thus, the differences that we observed in levels of *LDOC1* mRNA expression in mutated compared to unmutated CLL cases appear to result neither from copy number variation in the gene, nor mutations in the coding regions.

If LDOC1 functions as a transcription factor, as hypothesized, then small alterations in its level of expression could profoundly affect other genes that it regulates, and could promote or inhibit tumor formation, depending upon the context. We are the first to report *LDOC1S*, a new splice variant of *LDOC1*. In cancer cells, altered

expression of mRNA splice variants may result either from the generation of a new mRNA variant or from a change in the tissue-specific ratio of normal mRNA isoforms. An example of the former is *Ikaros*, a zinc finger DNA binding protein that is critical for normal lymphocyte development network. Alternative splicing of *Ikaros* pre-mRNA yields eight different isoforms, each with a different DNA binding capacity and differential expression in normal and neoplastic lymphocytes (47). Alterations in the tissue-specific ratio of normal mRNA isoforms may also be associated with tumorigenesis. For example, Bcl-X, a member of the Bcl-2 family, undergoes alternative splicing to yield a long mRNA, Bcl-XL, which has anti-apoptotic activity, and a short form, Bcl-XS, which has pro-apoptotic activity. While Bcl-XL is primarily found in long-lived post-mitotic adult tissues such as brain, Bcl-XS is expressed abundantly in cells with high turnover, such as human lymphocytes. Loss of Bcl-XS expression is associated with shorter relapse-free and overall survival in patients with acute myeloid leukemia (AML) (48), and a high ratio of Bcl-XL to Bcl-XS is associated with a poor prognosis in AML (49). Similarly, the interferon regulatory factor-1 (IRF-1) is a transcriptional activator that may function as a tumor suppressor gene. IRF-1 has five splice variants that lack various combinations of exons 7, 8, and/or 9. Although the variants are expressed in both normal and malignant cervical cells, they are found more abundantly in malignant cells. Most of the variants have been shown to inhibit the transcriptional activity of the wild type IRF-1 (50).

If the *LDOC1S* mRNA splice variant were translated into protein, it would contain the leucine zipper motif of the wild-type protein, but would lack the proline-rich region, which contains an SH3-binding consensus sequence, and the acidic region in the C-terminus. Mizutani and co-workers (51) constructed LDOC1 deletion mutants, and studied their localization and protein interactions following transfection into MDCK (Madine-Darby canine kidney) cells. They found that full-length LDOC1 localized

predominantly to the nucleus, whereas the N-terminal mutant protein (the leucine zipper region) localized to both the nucleus and cytoplasm. They also identified WAVE3, a predominantly cytoplasmic protein, as a binding partner of LDOC1. Co-expression of full-length LDOC1 and WAVE3 proteins shifted the localization of LDOC1 from the nucleus to the cytoplasm, and was associated with decreased apoptosis. They concluded, therefore, that WAVE3 may inhibit the pro-apoptotic activity of LDOC1, either by sequestering it in the cytoplasm or by shuttling it from the nucleus to cytoplasm. Similarly, it is conceivable that LDOC1S binds to LDOC1 to form non-functional dimers that inhibit the pro-apoptotic and anti-proliferative activities of LDOC1, possibly by sequestering LDOC1 in the cytoplasm. Alternatively, within the nucleus LDOC1S could form nonfunctional dimers with LDOC1; these could compete with functional LDOC1 dimers for DNA binding sites and inhibit its transcriptional regulatory activity. This scenario would be similar to the inhibitory actions displayed by members of the Id protein family, which contain helix-loop-helix dimerization domains but lack DNA binding domains (52). Id proteins act in a dominant negative fashion and control critical events in cell differentiation, proliferation and tumorigenesis. It is also possible that LDOC1S may form non-functional heterodimers with other pro-apoptotic and/or anti-proliferative proteins. Either mechanism might contribute to the aggressive behavior of some tumor types that aberrantly express LDOC1 isoforms.

The relatively high expression of *LDOC1* mRNA isoforms in unmutated CLL cases compared to mutated cases, and their expression in a variety of tumor cell lines suggest that *LDOC1* may contribute to aggressive clinical behavior. Because we are a tertiary care center, many of our patients received the diagnosis of CLL months to years before seeking care at our hospital. Thus, we do not know the *LDOC1* mRNA expression status of their CLL cells at the time of initial diagnosis. We also do not know if LDOC1

protein expression predicts overall survival in previously treated CLL patients. We have been unable to identify a sensitive and specific commercially-available antibody for use in either a Western blot or flow cytometry-based assay. However, QRT-PCR assays for a variety of different mRNA transcripts, such as BCR/ABL1 fusion transcripts, are now performed routinely in clinical molecular diagnostics laboratories (53). Thus, the lack of a robust antibody does not preclude the use of *LDOC1* mRNA expression as a clinically relevant biomarker of prognosis. Whether *LDOC1* mRNA expression is stable in cases that have undergone clonal evolution over the disease course or following therapeutic interventions, and if it is truly a better predictor of overall survival than either *IGHV* somatic mutation status or ZAP70 protein expression can only be answered by a larger longitudinal study.

After identifying LDOC1 as a novel biomarker to predict CLL prognosis, we next asked whether the differential mRNA expression had biologic relevance to CLL pathogenesis. Since very little is known about *LDOC1*, we performed global gene expression profiling to gain an understanding of the biologic processes in which *LDOC1* participates in general, and in B cell-related pathways, in particular. Gene expression profiling of HeLa cells revealed 107 genes that are significantly differentially expressed following LDOC1 protein knock-down. We further processed the 107 genes using Ingenuity Pathways to identify potential interaction networks. We found that the differentially expressed genes could be placed into a variety of pathways including regulation of gene expression, cellular function and maintenance, cancer, cell cycle, cellular growth and proliferation, cellular assembly and organization, cell death, and DNA replication, recombination and repair. When these networks were inspected, we identified several genes that participate in biologic processes in hematopoietic cells, such as *EBF* (Early B cell Factor), which is essential for initiation of early B cell

78

development from the lymphoid progenitor towards the pro-B cell stage (54, 55), and HCK (Hematopoietic Cell Kinase), a protein tyrosine kinase that is expressed predominantly in cells of lymphoid and myeloid lineage (56). However, the majority of these 107 differentially expressed genes are of unknown function in hematopoiesis or lymphomagenesis.

To identify and subsequently validate genes with concordant LDOC1 expression in transfected HeLa cells and CLL samples, we used two approaches. First, we compared the gene expression profiling data obtained using Affymetrix microarrays from the current HeLa experiment and a previously-performed CLL study in our laboratory (34). This approach identified 51 genes in common. We attempted to validate, by QRT-PCR assay, 6 genes from this list whose function suggested that they might contribute to oncogenesis: *NRIP1*, *DNAJA1*, *UBE2N*, *ITGAL*, *MAPKAPK2*, and *CLCN4.* However, we were unable to validate any of these genes as differentially expressed in HeLa cells that are transfected with *LDOC1* siRNAs or non-targeting siRNAs. Second, we compared the gene expression data for 43 candidate biomarkers of prognosis obtained by the MF-QRT-PCR assay from 76 CLL samples with the transfected HeLa cells. This analysis identified *GFI1* as concordantly differentially expressed in both transfected HeLa cells and CLL samples. We failed to detect *GFI1* as differentially expressed in Affymetrix data acquired from HeLa and CLL cells grouped into *LDOC1*-high and *LDOC1*-low categories, possibly due to the smaller dynamic range of Affymetrix arrays compared to the QRT-PCR assay (three orders of magnitude of expression compared to seven, respectively) (20).

The gene Growth Factor-Independence 1 (*GFI1*) is a zinc finger transcription factor that was first identified in T-cell lymphoma cell lines as the target of Moloney

murine leukemia virus insertion, which resulted in enhanced cell growth. Upon activation by viral insertion into its promoter, GFI1 conferred interleukin-2 (IL-2)-independent proliferation to a IL-2-dependent T-cell lymphoma cell line (57). Subsequently, it was shown that GFI1 is a transcription factor that contains a Snail/Gfi-1 (SNAG) domain, and regulates gene expression by repressing transcription through its repressor domain SNAG (58). The same study showed that GFI-1 overexpression inhibited cell death induced by IL-2 withdrawal in IL-2 dependent T-cell lymphoma cell line. This escape from cell death was strictly dependent on intact SNAG repression function (58). GFI1 inhibits cell death of IL-2-dependent T cell lymphoma cell-lines in IL-2-deficient media by directly binding and repressing the promoter of Bax, a pro-apoptotic member of Bcl-2 family. The data also suggest that GFI1 represses the expression of another apoptosis promoting protein, Bak (59).

Evidence for the oncogenic potential of *GFI1* comes from *in vivo* studies using transgenic mice. While a minority of mice transgenic for *GFI1* alone developed T-cell lymphoma, the vast majority of mice transgenic for *GFI1* in combination with *PIM-1* or *L-MYC* developed T-cell lymphomas (60). *In vitro* studies support the hypothesis that overexpression of GFI1 mediates lymphomagenesis. GFI1 has been shown to repress cell cycle inhibitory genes, *CDKN1A* and *CDKN2b*, which encode the proteins p21^{Cip1} and p15^{INK4B}, respectively (35, 61). However, the mechanism of this repression is independent of DNA-binding by GFI1. GFI1 and c-Myc are recruited to the promoters of these target genes by Miz-1, a transcription factor and interaction partner of c-Myc, to form a complex. This protein complex allows GFI1 and c-Myc to cooperate to repress transcription, without directly binding to the promoters of the target genes, and to contribute to proliferation.

GFI1 is expressed by hematopoietic stem cells (HSC), common lymphoid progenitors, and granulocytic and monocytic progenitors (62). *In vivo* experiments using *GFI1* knock-out mice have shown that GFI1 maintains the self-renewal capacity of the HSC by restricting their proliferative capacity (62, 63). Further, *GFI1* also protects HSCs from stress-induced apoptosis (64). In the hematopoietic system the gene expression patterns of *LDOC1* and *GFI1* are similar: both are highly expressed in HSCs and progenitor cells (42) and show higher expression in naïve peripheral blood B cells than memory B cells according to our (26). Because both are highly expressed in HSCs, one can speculate that LDOC1 may cooperate with GFI1 in maintaining HSC renewal; this may have implications in maintenance of the leukemic stem cell as well.

Antigenic stimulation rapidly induces GFI1 expression in mature peripheral T cells, which suggests that GFI1 participates in T cell activation (65). Constitutive GFI1 expression in Jurkat T lymphoma cells inhibits phorbol ester-induced G1 arrest by blocking the negative cell cycle regulator, p21, and also inhibits activation-induced cell death. It is not known if GFI1 expression is induced in benign or malignant B cells following activation with antigen or other stimuli, such as IL-4, CpG oligonucleotides, or IgM. Longo and co-workers have demonstrated that cells obtained from patients with an aggressive disease course (usually unmutated CLL cases) respond to *in vitro* stimulation by CpG oligonucleotides by proliferating. In contrast, CLL cells from patients with an indolent disease course (usually mutated CLL cases) respond to stimulation by undergoing apoptosis (41). Further, responsiveness to *in vitro* stimulation by CpG oligonucleotides may be a better predictor of prognosis than *IGHV* somatic mutation status (66). The findings that *LDOC1* and *GFI1* are expressed at significantly higher levels in unmutated CLL cases (22), which are more responsive to antigenic stimuli, than

in mutated cases, raise the possibility that both *LDOC1* and *GFI1* participate in signaling pathways that promote proliferation in CLL cells and contribute to its pathogenesis.

A recent report demonstrates that the microRNA, miR155, is upregulated in normal B cells that have been stimulated with IgM/CD40 ligand or CpG oligonucleotides and in primary CLL cells, compared to unstimulated normal B cells. In human embryonic kidney 293 cells miR155 has been shown to down-regulate *LDOC1* mRNA expression (67). Although we can find no published reports of GFI1 regulation by miR155, a bioinformatic query of the TargetScanHuman database for microRNA binding sites (www.targetscan.org, Release 5.1) for miR155 indicates that GFI1 has a putative binding site for this microRNA in its 3'UTR. Thus, LDOC1 and GFI1 expression may be coordinately regulated through miR155.

Our data raise the possibility that a potential interaction between *GFI1* and *LDOC1* contributes to CLL pathophysiology, possibly by enhancing responsiveness to antigenic stimulation. We have shown previously that high levels of mRNA expression of both *GFI1* (22) and *LDOC1* (22, 26) are associated with unmutated *IGHV* somatic mutation status, a marker of poor prognosis. In the current study, LDOC1 knock-down in HeLa cells resulted in down-regulation of *GFI1* mRNA. It also resulted in decreased mRNA expression of two known GFI1 transcription targets, the pro-apoptotic gene *BAX* (59) and the cell cycle repressor gene *CDKN1A* (35) (p values, 0.065 and 0.008, respectively). Thus, it is possible that LDOC1 and GFI1 cooperate to inhibit apoptosis and promote cell proliferation in aggressive CLL. Interestingly, a recent report has shown that downregulation of Bax protein correlates with several parameters of poor prognosis in CLL patients, including higher Binet stage, shorter lymphocyte doubling time, higher

CD38 protein expression, and presence of chromosomal abnormalities in 11q and 17p and/or a mutation of the *ATM* or *TP53* genes (68).

It is conceivable that LDOC1 may directly regulate *GFI1* gene expression by binding to its promoter and activating its transcription. Alternatively, LDOC1 may interact indirectly with GFI1 as part of the protein complex formed by Miz-1, GFI1 and c-Myc, as described above, to regulate transcription of GFI1 target genes, such as *BAX* and *CDKN1A* (35). Recent data suggest that c-Myc interacts with LDOC1 (69). Additional studies, such as chromatin immunoprecipitation and reporter assays, could be performed to determine if LDOC1 binds directly to the *GFI1* promoter and controls its transcription. Similar studies could also be performed to determine if LDOC1 interacts with other proteins, for example, c-Myc, that regulate expression of GFI1 target genes and contribute to malignant behavior.

It has been shown that following stimulation with CpG oligonucleotides *in vitro*, CLL cells obtained from patients with an aggressive clinical course proliferate, whereas CLL cells obtained from patients with an indolent course undergo apoptosis (41). The same investigators also demonstrated that this proliferative response was generally associated with unmutated *IGHV* status, and predicted shorter time-to-treatment, progression free survival, and overall survival in a cohort of CLL patients. However, mutated cases that proliferated under these conditions had a poorer prognosis than mutated cases that underwent apoptosis (66). The finding that expression of both *LDOC1* and *GFI1* are upregulated in unmutated CLL cases raises the possibility that they interact to regulate the expression of the genes that enhance proliferation and/or inhibit apoptosis in CLL cells. Since selection of antigens through the B cell receptor has been implicated in the pathogenesis and progression of CLL (70-77), an *in vitro* study of

LDOC1 and GFI1 expression following stimulation, as well as established downstream pathways in CLL survival and proliferation, for example, Akt signaling (41), would be highly illuminating.

In summary, we have shown that *LDOC1* mRNA expression is an excellent biomarker of overall survival in untreated CLL patients. Longitudinal studies in a larger cohort of CLL patients are warranted to determine the value of *LDOC1* mRNA measurement as a clinical prognostic test. Our findings from gene expression profiling studies suggest that *GFI1* might be one of the key genes with which LDOC1 interacts to contribute to CLL pathophysiology.

Supplemental Table 1. *IGHV* Somatic Mutation Status and Family, Expression of *LDOC1* mRNA and ZAP70 Protein, and Genomic Abnormalities of the CLL cases

CLL	Mutation	IGHV Family	% Homology	LDOC1*	ZAP70†	Genomic Abnormality
1	M	2-05	95	NEG	NEG	del(13q)
2	U	5-51	100	POS	NA	del(13q), +12
3	M	4-34	92	NEG	NEG	None
4	M	1-69	97.6	NEG	NA	del(13q)
5	M	3-23	97.7	NEG	POS	del(13q)
6	U	1-69	100	POS	POS	None
7	M	3-33	91	NEG	NEG	del(13q), del(6q)
8	U	3-15	98.7	NEG	NA	del(13q)
10	M	3-43	95.3	POS	NA	None
11	M	3-11	94.1	NEG	NEG	del(13q)
12	M	3-23	88	NEG	NEG	del(13q)x2‡
13	M	4-34	95.5	NEG	POS	NA
14	M	4-30	88.4	NEG	NEG	NA
15	M	4-34	93	NEG	NEG	None
16	U	3-21	100	POS	POS	NA
17	M	3-48	97	NEG	NEG	del(13q)
20	U	3-73	100	NEG	POS	NA
21	M	4-34	96.6	NEG	NEG	del(13q)x2
22	M	4-39	89	NEG	NEG	del(13q)
23	M	1-02	92	NEG	NEG	NA
24	M	3-30	97.5	POS	NEG	NA
25	M	5-a	96.3	NEG	NA	NA
26	M	1-02	93.3	NEG	NEG	del(13q)
27	M	3-30	96	NEG	NEG	NA
28	M	3-34	93	NEG	NEG	NA
29	M	3-15	93	NEG	NEG	NA
32	M	3-09	95	NEG	NEG	NA
33	M	3-21	97.7	NEG	POS	del(13q), del(11q)
35	M	1-18	90	NEG	NEG	del(13q)x2
37	M	4-04	95	NEG	NEG	None
38	M	3-15	97.3	NEG	POS	NA
39	M	3-15	92	NEG	NEG	NA
40	M	3-23	95	NEG	NEG	NA
41	U	4-39	99.6	POS	POS	NA
42	U	4-39	100	POS	NA	None
43	U	3-30	99.7	POS	NEG	del(13q), del(11q)
44	U	4-39	100	NEG	POS	NA
45	U	3-30	100	NEG	NEG	NA
46	U	4-34	100	POS	POS	None
47	U	2-05	98.3	NEG	NEG	NA

CLL	Mutation	IGHV Family	% Homology	LDOC1*	ZAP70†	Genomic Abnormality
48	U	4-04	100	POS	POS	del(13q)
49	U	4-39	100	POS	POS	None
50	U	3-30	100	POS	NEG	NA
51	U	7-04	100	POS	NA	+12
52	M	3-30	91.9	NEG	NEG	NA
53	U	1-69	100	POS	POS	+12
54	U	3-48	100	POS	NEG	del(13q), del(6q)
55	M	3-74	89.9	NEG	NA	None
56	U	3-53	99	NEG	POS	NA
57	U	4-39	100	NEG	NA	del(13q)
58	M	4-34	95.9	NEG	NEG	del(13q)
59	U	4-39	100	POS	NEG	del(13q), del(6q), del(17p)
60	U	1-69	99	POS	POS	del(13q), del(17p)
61	M	3-23	97	NEG	POS	NA
62	M	3-07	93	NEG	NEG	None
63	M	3-23	92.5	NEG	NEG	None
66	U	1-69	100	POS	POS	del(13q), del(17p)
67	M	4-30	97	NEG	NEG	del(13q)
68	U	1-08	99.6	NEG	POS	del(13q)
69	M	3-23	91.5	NEG	NEG	del(13q)
70	U	4-34	100	NEG	NA	None
71	M	3-07	92.2	NEG	NA	del(13q)
72	M	3-23	97.3	NEG	NA	NA
73	U	1-69	100	POS	POS	+12
75	M	6-01	98	NEG	NEG	None
76	M	3-30	97.6	NEG	NEG	None
78	M	3-30	90.5	NEG	NEG	+12
80	U	3-33	99.7	POS	POS	None
81	M	4-04	91.5	NEG	NEG	del(13q)
83	M	3-11	94.6	NEG	NEG	del(13q), +12
84	U	5-a	99	POS	NA	+12
87	M	3-22	87	NEG	POS	None
88	U	3-33	100	POS	POS	del(11q)
90	U	4-04	100	POS	NA	None
92	U	1-02	100	POS	POS	None
93	M	4-54	95	NEG	NEG	del(13q), del(17p)
95	M	4-34	97.9	NEG	NEG	del(13q)
97	U	3-09	100	POS	POS	None
99	M	4-34	93.5	NEG	NEG	del(13q)
100	M	4-34	94.6	NEG	NEG	del(13q)
102	U	1-02	100	NEG	NEG	del(11q)
103	M	3-09	94.2	NEG	NEG	del(13q), del(17p)
104	M	3-07	95.6	NEG	POS	+12
105	U	1-03	100	NEG	POS	del(11q)
107	U	1-69	100	NEG	NEG	del(13q), del(11q)
109	U	4-39	100	POS	NA	+12
110	M	3-23	96.6	NEG	POS	del(13q)
113	U	3-09	100	POS	POS	NA

CLL	Mutation	IGHV Family	% Homology	LDOC1*	ZAP70†	Genomic Abnormality
116	M	3-30	92.5	NEG	NEG	+12
118	M	3-23	97.6	NEG	NEG	del(13q)
119	M	3-21	98	NEG	POS	del(13q)
121	M	4-34	96.2	NEG	NEG	del(13q), +12
124	U	3-13	99.3	POS	POS	+12
135	U	3-21	100	POS	POS	del(13q), +12
143	U	4-59	99.3	POS	POS	+12
144	U	3-23	99.3	NEG	POS	None
145	U	4-b	99.7	POS	POS	del(13q)
158	U	4-30	100	POS	POS	None
159	M	3-07	91.2	NEG	NEG	del(13q)
161	U	3-30	99.6	NEG	POS	del(13q)
162	NA	NA	NA	POS	NEG	+12
167	M	3-09	94.7	NEG	NEG	del(13q)x2
168	M	1-03	91.8	NEG	NA	NA
170	U	7-04	99.1	NEG	POS	del(13q)
171	U	3-23	99.1	NEG	NEG	del(13q)
174	M	1-08	92.2	NEG	NEG	None
176	U	1-69	100	POS	POS	+12
178	U	1-69	100	POS	NEG	None
179	U	4-31	100	NEG	POS	NA
180	M	3-23	95.7	NEG	POS	NA
181	M	4-31	88.4	NEG	NEG	del(13q), del(17p)
182	U	2-70	98.2	POS	POS	NA
184	U	3-48	100	POS	POS	del(13q)
186	M	3-07	90.8	NEG	NA	+12
188	U	4-31	99.7	POS	POS	+12
189	U	4-34	99.6	POS	POS	None
191	M	4-34	94	NEG	NEG	None
192	U	3-30	99.6	POS	POS	None
202	U	1-18	100	POS	POS	NA
203	M	3-23	96.5	NEG	NEG	None
204	U	5-51	99	NEG	NA	del(11q)
205	M	1-02	96.9	NEG	NEG	del(13q)
206	U	1-69	100	POS	NEG	del(13q)
207	U	4-34	100	POS	NEG	+12
209	U	3-11	98.6	NEG	POS	NA
212	M	6-01	95.7	NEG	POS	del(13q), del(11q), del(17p)
213	U	4-39	100	NEG	NEG	del(6q), del(17p)
216	U	1-18	99.3	POS	POS	del(13q)
217	U	3-33	100	POS	POS	del(11q)
219	U	3-20	99.6	POS	POS	del(13q), del(6q)
222	M	4-34	92.8	NEG	NEG	NA

Abbreviations: M, mutated; U, unmutated; POS, positive; NEG, negative; NA, not available.
* Total *LDOC1* mRNA expression measured by MF-QRT-PCR assay
† ZAP70 protein expression measured by immunohistochemistry or flow cytometry

‡ Biallelic loss

CHAPTER 7. BIBLIOGRAPHY

1.   Oscier, D., C. Fegan, P. Hillmen, T. Illidge, S. Johnson, P. Maguire, E. Matutes, and D. Milligan. 2004. Guidelines on the diagnosis and management of chronic lymphocytic leukaemia. Br J Haematol 125:294-317.

2.   Binet, J. L., A. Auquier, G. Dighiero, C. Chastang, H. Piguet, J. Goasguen, G. Vaugier, G. Potron, P. Colona, F. Oberling, M. Thomas, G. Tchernia, C. Jacquillat, P. Boivin, C. Lesty, M. T. Duault, M. Monconduit, S. Belabbes, and F. Gremy. 1981. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer 48:198-206.

3.   Rai, K. R., A. Sawitsky, E. P. Cronkite, A. D. Chanana, R. N. Levy, and B. S. Pasternack. 1975. Clinical staging of chronic lymphocytic leukemia. Blood 46:219-234.

4.   Hamblin, T. J., Z. Davis, A. Gardiner, D. G. Oscier, and F. K. Stevenson. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94:1848-1854.

5.   Thorselius, M., A. Krober, F. Murray, U. Thunberg, G. Tobin, A. Buhler, D. Kienle, E. Albesiano, R. Maffei, L. P. Dao-Ung, J. Wiley, J. Vilpo, A. Laurell, M. Merup, G. Roos, K. Karlsson, N. Chiorazzi, R. Marasca, H. Dohner, S. Stilgenbauer, and R. Rosenquist. 2006. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. Blood 107:2889-2894.

6.   Tobin, G., U. Thunberg, A. Johnson, I. Eriksson, O. Soderberg, K. Karlsson, M. Merup, G. Juliusson, J. Vilpo, G. Enblad, C. Sundstrom, G. Roos, and R. Rosenquist. 2003. Chronic lymphocytic leukemias utilizing the VH3-21 gene

display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood 101:4952-4957.

7. Zenz, T., H. Dohner, and S. Stilgenbauer. 2007. Genetics and risk-stratified approach to therapy in chronic lymphocytic leukemia. Best Pract Res Clin Haematol 20:439-453.

8. Juliusson, G., D. G. Oscier, M. Fitchett, F. M. Ross, G. Stockdill, M. J. Mackie, A. C. Parker, G. L. Castoldi, A. Guneo, S. Knuutila, and et al. 1990. Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. N Engl J Med 323:720-724.

9. Dohner, H., S. Stilgenbauer, A. Benner, E. Leupolt, A. Krober, L. Bullinger, K. Dohner, M. Bentz, and P. Lichter. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med 343:1910-1916.

10. Dohner, H., S. Stilgenbauer, K. Dohner, M. Bentz, and P. Lichter. 1999. Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis. J Mol Med 77:266-281.

11. Dohner, H., K. Fischer, M. Bentz, K. Hansen, A. Benner, G. Cabot, D. Diehl, R. Schlenk, J. Coy, S. Stilgenbauer, and et al. 1995. p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. Blood 85:1580-1589.

12. Cordone, I., S. Masi, F. R. Mauro, S. Soddu, O. Morsilli, T. Valentini, M. L. Vegna, C. Guglielmi, F. Mancini, S. Giuliacci, A. Sacchi, F. Mandelli, and R. Foa. 1998. p53 expression in B-cell chronic lymphocytic leukemia: a marker of disease progression and poor prognosis. Blood 91:4342-4349.

13. Haslinger, C., N. Schweifer, S. Stilgenbauer, H. Dohner, P. Lichter, N. Kraut, C. Stratowa, and R. Abseher. 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.

14.   Wiestner, A., A. Rosenwald, T. S. Barry, G. Wright, R. E. Davis, S. E. Henrickson, H. Zhao, R. E. Ibbotson, J. A. Orchard, Z. Davis, M. Stetler-Stevenson, M. Raffeld, D. C. Arthur, G. E. Marti, W. H. Wilson, T. J. Hamblin, D. G. Oscier, and L. M. Staudt. 2003. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood 101:4944-4951.

15.   Orchard, J. A., R. E. Ibbotson, Z. Davis, A. Wiestner, A. Rosenwald, P. W. Thomas, T. J. Hamblin, L. M. Staudt, and D. G. Oscier. 2004. ZAP-70 expression and prognosis in chronic lymphocytic leukaemia. Lancet 363:105-111.

16.   Rassenti, L. Z., L. Huynh, T. L. Toy, L. Chen, M. J. Keating, J. G. Gribben, D. S. Neuberg, I. W. Flinn, K. R. Rai, J. C. Byrd, N. E. Kay, A. Greaves, A. Weiss, and T. J. Kipps. 2004. ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. N Engl J Med 351:893-901.

17.   Admirand, J. H., R. J. Knoblock, K. R. Coombes, C. Tam, E. J. Schlette, W. G. Wierda, A. Ferrajoli, S. O'Brien, M. J. Keating, R. Luthra, L. J. Medeiros, and L. V. Abruzzo. 2010. Immunohistochemical detection of ZAP70 in chronic lymphocytic leukemia predicts immunoglobulin heavy chain gene mutation status and time to progression. Mod Pathol 23:1518-1523.

18.   Tam, C. S., and M. J. Keating. Chemoimmunotherapy of chronic lymphocytic leukemia. Nat Rev Clin Oncol 7:521-532.

19.   Rosenwald, A., A. A. Alizadeh, G. Widhopf, R. Simon, R. E. Davis, X. Yu, L. Yang, O. K. Pickeral, L. Z. Rassenti, J. Powell, D. Botstein, J. C. Byrd, M. R. Grever, B. D. Cheson, N. Chiorazzi, W. H. Wilson, T. J. Kipps, P. O. Brown, and

L. M. Staudt. 2001. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. J Exp Med 194:1639-1647.

20.   Abruzzo, L. V., K. Y. Lee, A. Fuller, A. Silverman, M. J. Keating, L. J. Medeiros, and K. R. Coombes. 2005. Validation of oligonucleotide microarray data using microfluidic low-density arrays: a new statistical method to normalize real-time RT-PCR data. Biotechniques 38:785-792.

21.   Klein, U., Y. Tu, G. A. Stolovitzky, M. Mattioli, G. Cattoretti, H. Husson, A. Freedman, G. Inghirami, L. Cro, L. Baldini, A. Neri, A. Califano, and R. Dalla-Favera. 2001. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J Exp Med 194:1625-1638.

22.   Abruzzo, L. V., L. L. Barron, K. Anderson, R. J. Newman, W. G. Wierda, S. O'Brien, A. Ferrajoli, M. Luthra, S. Talwalkar, R. Luthra, D. Jones, M. J. Keating, and K. R. Coombes. 2007. Identification and validation of biomarkers of IgV(H) mutation status in chronic lymphocytic leukemia using microfluidics quantitative real-time polymerase chain reaction technology. J Mol Diagn 9:546-555.

23.   Nagasaki, K., T. Manabe, H. Hanzawa, N. Maass, T. Tsukada, and K. Yamaguchi. 1999. Identification of a novel gene, LDOC1, down-regulated in cancer cell lines. Cancer Lett 140:227-234.

24.   Nagasaki, K., C. Schem, C. von Kaisenberg, M. Biallek, F. Rosel, W. Jonat, and N. Maass. 2003. Leucine-zipper protein, LDOC1, inhibits NF-kappaB activation and sensitizes pancreatic cancer cells to apoptosis. International journal of cancer 105:454-458.

25. Inoue, M., K. Takahashi, O. Niide, M. Shibata, M. Fukuzawa, and C. Ra. 2005. LDOC1, a novel MZF-1-interacting protein, induces apoptosis. FEBS letters 579:604-608.

26. Duzkale, H., C. D. Schweighofer, K. R. Coombes, L. L. Barron, A. Ferrajoli, S. O'Brien, W. G. Wierda, J. Pfeifer, T. Majewski, B. A. Czerniak, J. L. Jorgensen, L. J. Medeiros, E. J. Freireich, M. J. Keating, and L. V. Abruzzo. LDOC1 mRNA is differentially expressed in chronic lymphocytic leukemia and predicts overall survival in untreated patients. Blood 117:4076-4084.

27. McCarthy, H., W. G. Wierda, L. L. Barron, C. C. Cromwell, J. Wang, K. R. Coombes, R. Rangel, K. S. Elenitoba-Johnson, M. J. Keating, and L. V. Abruzzo. 2003. High expression of activation-induced cytidine deaminase (AID) and splice variants is a distinctive feature of poor-prognosis chronic lymphocytic leukemia. Blood 101:4903-4908.

28. Klein, U., K. Rajewsky, and R. Kuppers. 1998. Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. J Exp Med 188:1679-1689.

29. Retter, I., H. H. Althaus, R. Munch, and W. Muller. 2005. VBASE2, an integrative V gene database. Nucleic Acids Res 33:D671-674.

30. Fais, F., F. Ghiotto, S. Hashimoto, B. Sellars, A. Valetto, S. L. Allen, P. Schulman, V. P. Vinciguerra, K. Rai, L. Z. Rassenti, T. J. Kipps, G. Dighiero, H. W. Schroeder, Jr., M. Ferrarini, and N. Chiorazzi. 1998. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest 102:1515-1525.

31. Admirand, J. H., G. Z. Rassidakis, L. V. Abruzzo, J. R. Valbuena, D. Jones, and L. J. Medeiros. 2004. Immunohistochemical detection of ZAP-70 in 341 cases of non-Hodgkin and Hodgkin lymphoma. Mod Pathol 17:954-961.

32. Olshen, A. B., E. S. Venkatraman, R. Lucito, and M. Wigler. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5:557-572.

33. Burnham, K. P., and , and D. R. Anderson, editors. 2002. Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach. Springer-Verlag.

34. Wang, J., K. R. Coombes, W. E. Highsmith, M. J. Keating, and L. V. Abruzzo. 2004. Differences in gene expression between B-cell chronic lymphocytic leukemia and normal B cells: a meta-analysis of three microarray studies. Bioinformatics 20:3166-3178.

35. Liu, Q., S. Basu, Y. Qiu, F. Tang, and F. Dong. A role of Miz-1 in Gfi-1-mediated transcriptional repression of CDKN1A. Oncogene 29:2843-2852.

36. Crespo, M., N. Villamor, E. Gine, A. Muntanola, D. Colomer, T. Marafioti, M. Jones, M. Camos, E. Campo, E. Montserrat, and F. Bosch. 2006. ZAP-70 expression in normal pro/pre B cells, mature B cells, and in B-cell acute lymphoblastic leukemia. Clin Cancer Res 12:726-734.

37. Orchard, J., R. Ibbotson, G. Best, A. Parker, and D. Oscier. 2005. ZAP-70 in B cell malignancies. Leuk Lymphoma 46:1689-1698.

38. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125-8148.

39. Irizarry, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed. 2003. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31:e15.

40.   Zhang, L., M. F. Miles, and K. D. Aldape. 2003. A model of molecular interactions on short oligonucleotide microarrays. Nat Biotechnol 21:818-821.

41.   Longo, P. G., L. Laurenti, S. Gobessi, A. Petlickovski, M. Pelosi, P. Chiusolo, S. Sica, G. Leone, and D. G. Efremov. 2007. The Akt signaling pathway determines the different proliferative capacity of chronic lymphocytic leukemia B-cells from patients with progressive and stable disease. Leukemia 21:110-120.

42.   Hemmoranta, H., S. Hautaniemi, J. Niemi, D. Nicorici, J. Laine, O. Yli-Harja, J. Partanen, and T. Jaatinen. 2006. Transcriptional profiling reflects shared and unique characters for CD34+ and CD133+ cells. Stem Cells Dev 15:839-851.

43.   Damle, R. N., T. Wasil, F. Fais, F. Ghiotto, A. Valetto, S. L. Allen, A. Buchbinder, D. Budman, K. Dittmar, J. Kolitz, S. M. Lichtman, P. Schulman, V. P. Vinciguerra, K. R. Rai, M. Ferrarini, and N. Chiorazzi. 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 94:1840-1847.

44.   Krober, A., T. Seiler, A. Benner, L. Bullinger, E. Bruckle, P. Lichter, H. Dohner, and S. Stilgenbauer. 2002. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. Blood 100:1410-1416.

45.   Stilgenbauer, S., L. Bullinger, P. Lichter, and H. Dohner. 2002. Genetics of chronic lymphocytic leukemia: genomic aberrations and V(H) gene mutation status in pathogenesis and clinical course. Leukemia 16:993-1007.

46.   Oscier, D. G., A. C. Gardiner, S. J. Mould, S. Glide, Z. A. Davis, R. E. Ibbotson, M. M. Corcoran, R. M. Chapman, P. W. Thomas, J. A. Copplestone, J. A. Orchard, and T. J. Hamblin. 2002. Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. Blood 100:1177-1184.

47. Yagi, T., S. Hibi, M. Takanashi, G. Kano, Y. Tabata, T. Imamura, T. Inaba, A. Morimoto, S. Todo, and S. Imashuku. 2002. High frequency of Ikaros isoform 6 expression in acute myelomonocytic and monocytic leukemias: implications for up-regulation of the antiapoptotic protein Bcl-XL in leukemogenesis. Blood 99:1350-1355.

48. Yamaguchi, H., K. Inokuchi, and K. Dan. 2002. The study for loss of bcl-xs expression as a prognostic factor in acute myeloid leukemia. Leuk Res 26:1119-1123.

49. Deng, G., C. Lane, S. Kornblau, A. Goodacre, V. Snell, M. Andreeff, and A. B. Deisseroth. 1998. Ratio of bcl-xshort to bcl-xlong is different in good- and poor-prognosis subsets of acute myeloid leukemia. Mol Med 4:158-164.

50. Lee, E. J., M. Jo, J. Park, W. Zhang, and J. H. Lee. 2006. Alternative splicing variants of IRF-1 lacking exons 7, 8, and 9 in cervical cancer. Biochem Biophys Res Commun 347:882-888.

51. Mizutani, K., D. Koike, S. Suetsugu, and T. Takenawa. 2005. WAVE3 functions as a negative regulator of LDOC1. J Biochem 138:639-646.

52. Perk, J., A. Iavarone, and R. Benezra. 2005. Id family of helix-loop-helix proteins in cancer. Nat Rev Cancer 5:603-614.

53. Luthra, R., and L. J. Medeiros. 2006. TaqMan reverse transcriptase-polymerase chain reaction coupled with capillary electrophoresis for quantification and identification of bcr-abl transcript type. Methods Mol Biol 335:135-145.

54. Medina, K. L., J. M. Pongubala, K. L. Reddy, D. W. Lancki, R. Dekoter, M. Kieslinger, R. Grosschedl, and H. Singh. 2004. Assembling a gene regulatory network for specification of the B cell fate. Dev Cell 7:607-617.

55. Zhuang, Y., A. Jackson, L. Pan, K. Shen, and M. Dai. 2004. Regulation of E2A gene expression in B-lymphocyte development. Mol Immunol 40:1165-1177.

56.   Quintrell, N., R. Lebo, H. Varmus, J. M. Bishop, M. J. Pettenati, M. M. Le Beau, M. O. Diaz, and J. D. Rowley. 1987. Identification of a human gene (HCK) that encodes a protein-tyrosine kinase and is expressed in hemopoietic cells. Mol Cell Biol 7:2267-2275.

57.   Gilks, C. B., S. E. Bear, H. L. Grimes, and P. N. Tsichlis. 1993. Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma lines to IL-2-independent growth following activation of a gene (Gfi-1) encoding a novel zinc finger protein. Mol Cell Biol 13:1759-1768.

58.   Grimes, H. L., T. O. Chan, P. A. Zweidler-McKay, B. Tong, and P. N. Tsichlis. 1996. The Gfi-1 proto-oncoprotein contains a novel transcriptional repressor domain, SNAG, and inhibits G1 arrest induced by interleukin-2 withdrawal. Mol Cell Biol 16:6263-6272.

59.   Grimes, H. L., C. B. Gilks, T. O. Chan, S. Porter, and P. N. Tsichlis. 1996. The Gfi-1 protooncoprotein represses Bax expression and inhibits T-cell death. Proc Natl Acad Sci U S A 93:14569-14573.

60.   Schmidt, T., H. Karsunky, E. Gau, B. Zevnik, H. P. Elsasser, and T. Moroy. 1998. Zinc finger protein GFI-1 has low oncogenic potential but cooperates strongly with pim and myc genes in T-cell lymphomagenesis. Oncogene 17:2661-2667.

61.   Basu, S., Q. Liu, Y. Qiu, and F. Dong. 2009. Gfi-1 represses CDKN2B encoding p15INK4B through interaction with Miz-1. Proc Natl Acad Sci U S A 106:1433-1438.

62.   Zeng, H., R. Yucel, C. Kosan, L. Klein-Hitpass, and T. Moroy. 2004. Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. EMBO J 23:4116-4125.

63. Hock, H., M. J. Hamblen, H. M. Rooke, J. W. Schindler, S. Saleque, Y. Fujiwara, and S. H. Orkin. 2004. Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. Nature 431:1002-1007.

64. Khandanpour, C., C. Kosan, M. C. Gaudreau, U. Duhrsen, J. Hebert, H. Zeng, and T. Moroy. Growth Factor Independence 1 (Gfi1) Protects Hematopoietic Stem Cells Against Apoptosis But Also Prevents the Development of a Myeloproliferative-Like Disease. Stem Cells.

65. Karsunky, H., I. Mende, T. Schmidt, and T. Moroy. 2002. High levels of the onco-protein Gfi-1 accelerate T-cell proliferation and inhibit activation induced T-cell death in Jurkat T-cells. Oncogene 21:1571-1579.

66. Tarnani, M., L. Laurenti, P. G. Longo, N. Piccirillo, S. Gobessi, A. Mannocci, S. Marietti, S. Sica, G. Leone, and D. G. Efremov. 2010. The proliferative response to CpG-ODN stimulation predicts PFS, TTT and OS in patients with chronic lymphocytic leukemia. Leuk Res 34:1189-1194.

67. Skalsky, R. L., M. A. Samols, K. B. Plaisance, I. W. Boss, A. Riva, M. C. Lopez, H. V. Baker, and R. Renne. 2007. Kaposi's sarcoma-associated herpesvirus encodes an ortholog of miR-155. J Virol 81:12836-12845.

68. Pepper, C., T. T. Lin, G. Pratt, S. Hewamana, P. Brennan, L. Hiller, R. Hills, R. Ward, J. Starczynski, B. Austen, L. Hooper, T. Stankovic, and C. Fegan. 2008. Mcl-1 expression has in vitro and in vivo significance in chronic lymphocytic leukemia and is associated with other poor prognostic markers. Blood 112:3807-3817.

69. Miyamoto-Sato, E., S. Fujimori, M. Ishizaka, N. Hirai, K. Masuoka, R. Saito, Y. Ozawa, K. Hino, T. Washio, M. Tomita, T. Yamashita, T. Oshikubo, H. Akasaka, J. Sugiyama, Y. Matsumoto, and H. Yanagawa. A comprehensive resource of

interacting protein regions for refining human transcription factor networks. PLoS One 5:e9289.

70.   Chiorazzi, N., and M. Ferrarini. 2003. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. Annu Rev Immunol 21:841-894.

71.   Stevenson, F. K., and F. Caligaris-Cappio. 2004. Chronic lymphocytic leukemia: revelations from the B-cell receptor. Blood 103:4389-4395.

72.   Tobin, G. 2007. The immunoglobulin genes: structure and specificity in chronic lymphocytic leukemia. Leuk Lymphoma 48:1081-1086.

73.   Stamatopoulos, K., C. Belessi, C. Moreno, M. Boudjograh, G. Guida, T. Smilevska, L. Belhoul, S. Stella, N. Stavroyianni, M. Crespo, A. Hadzidimitriou, L. Sutton, F. Bosch, N. Laoutaris, A. Anagnostopoulos, E. Montserrat, A. Fassas, G. Dighiero, F. Caligaris-Cappio, H. Merle-Beral, P. Ghia, and F. Davi. 2007. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood 109:259-270.

74.   Messmer, B. T., E. Albesiano, D. G. Efremov, F. Ghiotto, S. L. Allen, J. Kolitz, R. Foa, R. N. Damle, F. Fais, D. Messmer, K. R. Rai, M. Ferrarini, and N. Chiorazzi. 2004. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. J Exp Med 200:519-525.

75.   Tobin, G., A. Rosen, and R. Rosenquist. 2006. What is the current evidence for antigen involvement in the development of chronic lymphocytic leukemia? Hematol Oncol 24:7-13.

76.   Refaeli, Y., R. M. Young, B. C. Turner, J. Duda, K. A. Field, and J. M. Bishop. 2008. The B cell antigen receptor and overexpression of MYC can cooperate in the genesis of B cell lymphomas. PLoS Biol 6:e152.

77.   Binder, M., F. Muller, A. Jackst, B. Lechenne, M. Pantic, U. Bacher, C. Zu
Eulenburg, H. Veelken, R. Mertelsmann, R. Pasqualini, W. Arap, and M. Trepel.
B-cell receptor epitope recognition correlates with the clinical course of chronic
lymphocytic leukemia. Cancer 117:1891-1900.

VITA

Hatice Duzkale was born in Ankara, Turkey, the daughter of Fatma Duzkale and Kemal Duzkale. After completing her high school education at Ankara High School, Ankara, Turkey in 1991 she entered Ankara University School of Medicine. She received the degree of medical doctor in April 1998, and came to the United States to obtain research training. She worked as visiting scientist in the Department of Experimental Therapeutics at UT M.D. Anderson Cancer Center for one year. In August 1999, Dr. Duzkale entered The University of Texas School of Public Health and received the degree of Master of Public Health in Community Health Practice in December 2003. She entered the Ph.D. program in the Graduate School of Biomedical Sciences in December 2003. She has been accepted into the Clinical Molecular Genetics Fellowship training program at Harvard Medical School beginning in July 2011.