UNIVERSITY OF CALIFORNIA, SAN DIEGO

SMAD2 and miR-181b roles in the transition from indolent to aggressive Chronic Lymphocytic Leukemia

A Thesis submitted in partial satisfaction of the requirements for the degree Masters of Science

in

Biology

by

Michelle Pablo Salcedo

Committee in charge:

Professor Thomas J. Kipps, Chair
Professor Steven Briggs, Co-Chair
Professor Trey Ideker
Professor Kees Murre

2012

The Thesis of Michelle Pablo Salcedo is approved and is acceptable in quality and form for publication and microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2012

DEDICATION

As my life long inspiration, she inspired me to reach for the stars and to keep shooting; I dedicate this work to my amazing mother, Majell, who has believed in me throughout my whole life and has made her sacrifices to make life for our whole family better.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Thomas Kipps for giving me the honor to work in his lab and be part of unparalleled work in the field of CLL. Although quite a challenge it has been throughout my journey here, he has made me into a stronger person and a better scientist. I am very grateful for his willingness and faith into my success. His mentorship has enabled me to become a better critical thinker and a harder worker.

I would also like to extend this gratitude to Dr. Emanuela Ghia for her monumental effort and mentorship throughout the duration of my thesis project. As a co-author of this work, her guidance, collaboration, and friendship have definitely impacted my success. I am undoubtedly appreciative of all the time she has bestowed upon my efforts to succeed and as well as open opportunities to be part of other interesting projects.

I would like to give recognition to Dr. Laura Rassenti, Carolina Bump, and Monica Cook for their support throughout my duration in the lab. Their support and effort to continue my studies and providing the materials needed have been nothing but positive and inspirational.

To Dr. Han-Yu (Brook) Chuang, I would like to express my thanks for the opportunity to work with her in the Subnetwork study and for contributing part of her work into my thesis. And to the co-authors of the Subnetwork study, Kate Licon, Dr. Alexander Kohlmann, Dr. Torsten Haferlach, Dr. Robin Foa, and Dr. Trey Ideker, I express my humble gratitude for their support on my thesis.

To my thesis committee, Dr. Steven Briggs, Dr. Trey Ideker, and Dr. Kees Murre, I would like to give my deepest thanks for their willingness to be part of my committee and for their support in my thesis and graduation.

Last, but not least, I would like to show my utter appreciation and thanks to all the members of the Kipps Lab, including all that have been mentioned here already and those who took leave. No matter how big or small, their helpfulness and advice has played an integral part in this milestone of my academic career. They are magnificent teachers and wonderful colleagues.

ABSTRACT OF THE THESIS

SMAD2 and miR-181b roles in the transition from indolent to aggressive Chronic Lymphocytic Leukemia

by

Michelle Pablo Salcedo

Master of Science in Biology

University of California, San Diego, 2012

Professor Thomas J. Kipps, Chair

The clinical course of Chronic Lymphocytic Leukemia (CLL) patients is heterogeneous: some patients progress relatively rapidly whereas others survive for many years without requiring therapy. Because static prognostic markers, such as *IGVH* mutation and ZAP-70 protein expression, lack indication of disease progression, alternative approaches for discriminating patients into their disparate risks for CLL

progression is needed. By integrating gene expression measurements over sets of genes that encode proteins known to interact within a protein subnetwork we found 22 of the 38 significant subnetworks had increased activity in defined high-risk patients (referred to as pro-onconets) whereas the other 16 had decreased activity (referred to as anti-onconets). MicroRNA 181b (miR-181b) was described to decrease over time in patients with progressive CLL and not in those who remained indolent. Interestingly, one of the targets of miR-181b is *SMAD2*, a gene upregulated in the pro-onconets. With that said, miR-181b may regulate some pro-onconet pathways. In this study, we asked whether protein expression of our pro-onconet genes increased in protein expression over time and also if miR-181b may target *SMAD2* transcript. Through western blot, quantitative Real-Time PCR, and transfection of miR-181b mimic into progressive CLL samples, we were able to answer these questions. This study shows that protein expression of these pro-onconet genes may increase over time and that miR-181b regulates *SMAD2* expression. These studies value the importance of future studies investigating the oncogenic role of *SMAD2*, the relative expressions of target transcripts for particular microRNAs, and epigenetic regulation in progressive CLL.

Introduction

Chronic Lymphocytic Leukemia (CLL) is a blood cancer consisting of neoplastic growth of mature monoclonal B-lymphocytes in the lymphoid tissue microenvironment (Burger, 2011) and high accumulation of these malignant cells, stuck in the G0-phase of the cell cycle, in the peripheral blood (Hamblin & Oscier, 1997). CLL is the most prevalent leukemia in the Western world. One of the main abnormalities of these CLL B-cells are their impaired function of programmed cell death (Danilov, Klein, Lee, Baez, & Huber, 2005) due to the halt at the G0-phase (in peripheral blood) and expression of many activation markers, a decrease in pro-apoptotic molecules, and upregulation or sustained expression of pro-survival molecules (Hamblin & Oscier, 1997).

Tumor evolution is comprised of a series of molecular events that direct the tumor cell to survive in anaerobic conditions, resist cell cycle control, and escape apoptosis. CLL is the most prevalent leukemia in the Western world. Powerful single-gene prognostic markers, such as *IGVH* mutation status and ZAP-70 protein expression, are markers that are static over time, thus are unlikely to associate with disease progression/tumor evolution in CLL. Included in this study, is a network-based analysis on gene expression data from a cohort of 130 CLL patients has revealed gene expression changes over time, of which such changes correspond to disease progression in CLL.

Several subnetworks discovered are pathways that are, but not limited to, involved in apoptosis, signal transduction, cell cycling, and cellular metabolism. Towards disease progression, this study shows that patients on this path will eventually have certain subnetworks that are highly active (referred to as pro-onconets) and other

subnetworks that decrease in activity (referred to as anti-onconets) when reached to an aggressive state of the disease and requiring treatment (Chuang et al., 2012). From Visone et al., there has been evaluation of microRNA expression over time (Visone et al., 2011). MicroRNA 181b (miR-181b) is seen to decrease over time in progressive patients. With this in mind, this microRNA may be a missing regulator to some of these subnetworks found in this study. From the subnetwork analysis, *SMAD2* was identified to be upregulated during disease progression. In association with the miR-181b story in CLL, *SMAD2* is predicted to be a target of miR-181b in silico. The aim of this study is to explore the relationship between expression levels of *SMAD2* and miR-181b in CLL cells of patients during disease progression. Together, miR-181b and *SMAD2* may be players in the same pathway and thus contribute to disease progression.

Prognostic Markers in CLL

The clinical course of CLL is heterogeneous. Patients with the indolent form of the disease are asymptomatic and have a longer time to treatment or may not require treatment at all, yet patients with the aggressive form have a shorter time to treatment and overall survival, and require immediate therapeutic action. Identifying the two clinical courses of CLL can be assessed by many factors.

Clinic-based prognostic systems, such as Rai (Rai et al., 1975) and Binet (Binet et al., 1981) staging systems, encompass physical examinations assessing for presence of lymphadenopathy, organomegaly, and cytopenias. Although these staging systems are beneficial for acknowledgment of systemic spread and disease severity, their prognostic power is not as robust as biology-based prognostic markers (Morabito et al., 2009).

Furthermore, patients with Rai Stage I-II (intermediate risk) are the hardest to predict if patients with indolent disease may become aggressive (Kufe et al., 2003).

Two of the most prevalent and robust biology-based prognostic markers used to stratify patients into the high risk group (aggressive CLL) and the low risk group (indolent CLL) are *Immunoglobulin Variable Region Heavy Chain* (*IGHV*) gene mutation status (Rosenwald et al., 2001) and Zeta-associated Protein 70 (ZAP-70) (Durig et al., 2003) protein expression. Mutation of the immunoglobulin variable region's heavy chain occurs on the hypervariable regions after antigen stimulation. In CLL, *IGHV* mutation status corresponds to the two disease courses in CLL, whereas the mutated form (<98% homology to germline sequence) is highly associated to the indolent course as the unmutated form (>=98% homology to germline sequence) is highly associated to the aggressive course (Damle et al., 1999).

Various studies have investigated protein marker associations with *IGHV* mutation status due to the long procedures to obtaining results from DNA sequencing. Such associations, or surrogate markers, to *IGHV* include ZAP-70 (Crespo et al., 2003) and CD38 (Hamblin et al., 2002). Flow cytometry is a faster technique and ZAP-70 protein expression was observed to be a better stand-alone prognostic marker to *IGHV* (Rassenti et al., 2004). In the Rassenti et al. study, of the patients in whom the rearranged *IGHV* gene had 98 % or greater homology with a known germ-line *IGHV* gene, 71 % were ZAP-70–positive; 83 % of the patients with a mutated IGHV gene were ZAP-70–negative (Rassenti et al., 2004). Despite these strong and statistically significant associations, there were also discordant associations between *IGHV* mutation status with ZAP-70 and CD38.

Although these biology-based prognostic markers do categorically define CLL patients into one risk group or the other to an extent, these markers cannot correspond their presence or absence to disease progression because they do not change expression over time (Rassenti et al., 2004). Although, there was some inference of changes in CD38 expression to occur over time (Hamblin et al., 2002), but actual proof has not yet surfaced. Furthermore, patients with discordant associations between IGHV mutation status and ZAP-70 expression cannot be determined on their clinical course. Last but not least, patients stratified to have good prognosis sometimes develop to the aggressive course, which then makes *IGVH* mutation status and ZAP-70 expression unreliable to identifying this discrepancy amongst patients with indolent CLL and their susceptibility to disease progression. Other means to determining time to treatment and disease progression early is necessary to overcome these flaws in biology-based single-gene prognostic markers.

Molecular Interaction Networks as Prognostic Profiles and Progression Indicators

An alternative means to understand, detect, and evaluate disease progression in CLL, one must realize that disease progression in any cancer involves a series of intracellular and molecular changes to achieve transformation from a normal cell into a cancer cell, which in other words, is the idea of cancer evolution. With this idea in mind, bioinformatic studies (Chen & Yuan, 2006; Trey Ideker, Ozier, Schwikowski, & Siegel, 2002) have utilized large data sets of human protein-protein and protein-DNA interactions, pooled from systematic yeast two-hybrid screens (Uetz et al., 2000) and co-immunopreciptation experiments (Gavin et al., 2002), and sought to investigate the

interaction of these molecules in order to define which ones interacted with other in a network via bioinformatic techniques. These defined networks can then reveal signaling and regulatory pathways that are significantly active and important. Significantly "active" networks are determined by meticulous statistical analyses that formulate a scoring system to identifying gene sets (genes encoding proteins that may interact with each other in a pathway) that may be differentially expressed significantly between two conditions, ie. *GAL80* knock out vs. Wild type in yeast (T Ideker et al., 2001; Trey Ideker et al., 2002). To do these analyses in CLL, we compiled gene expression data from DNA microarray experiments on 130 CLL patient samples to extrapolate molecular interaction networks significantly active in aggressive CLL using the same analysis as the study mentioned previously (Chuang et al., 2012). In our study, we compared two different time points of which the initial time point is closer to the diagnostic date as the latter time point is closer to the treatment date. Patients who were not treated had the same time interval between time point one and two as all the rest. With these sample criteria, we aimed to see the differences in gene expression over time in patients who were treated and untreated. From our results we have identified 1) gene expression profiles differentiate between stable and aggressive CLL patients, 2) stable CLL patients who have undergone disease progression eventually share the same gene expression patterns of patients who have aggressive CLL, and 3) we have identified 38 subnetworks known to be significantly active in aggressive CLL.

Of these 38 subnetworks, 22 had increased activity in the defined high-risk patients (referred to as pro-onconets) whereas the other 16 had decreased activity (referred to as anti-onconets). Many of these pro-onconet subnetworks involve pathways

that are not only perturbed in many cancers (ie. Cell cycle and apoptosis pathways) but are specifically perturbed in CLL (ie. WNT (Lu et al., 2004) and TGF-beta pathways (Laurence Lagneaux, Delforge, Bernier, Stryckmans, & Bron, 1998)). This work is submitted for publication and is included in this thesis with permission from the co-authors behalf.

TGF-beta Signaling and SMAD2

One of the aspects of CLL research includes specifically on the role of TGF-β1 in CLL. TGF-beta is a cytokine that binds to three TGF-beta receptors, referred to as TGF-Beta Receptor I, II, and III, to induce anti-proliferative and pro-apoptotic pathways, via Smad signaling, in normal B-lymphocytes (Sebestyén et al., 2005). Patients with CLL who are resistant to TGF-beta stimulation have low expression of TGF-beta receptors (Laurence Lagneaux et al., 1998), thus indicating a possible mechanism of how CLL B-cells have a growth advantage. Although CLL B-cells still produce TGF-beta cytokine, these cells produce more than normal B-cells, particularly because of the expansion of this CLL B-cell clone (Laurence Lagneaux et al., 1998).

TGF-beta resistance may also contribute an escape from apoptosis. Apoptotic response to TGF-beta treatment was shown to have a wide range of sensitivity in CLL patients to TGF-beta mediated apoptosis, but over all is resistant compared to normal B-lymphocytes (Douglas, Capocasale, Lamb, Nowell, & Moore, 1997). It was also shown that this resistance is not generally due to lack of TGF-beta receptor expression (Douglas et al., 1997), thus indicating that there could be deregulation of signal transduction pathways connected with apoptosis.

SMAD2 is a receptor-mediated SMAD protein involved in TGF-βR1 activated signal transduction (Itoh, Itoh, Goumans, & Ten Dijke, 2000). Upon activation, SMAD2 becomes phosphorylated, complexes with other SMAD proteins, and relocates to the nucleus, which in turn upregulate genes involved in cell proliferation, apoptosis, and differentiation (Lebman & Edmiston, 1999; Ten Dijke, Goumans, Itoh, & Itoh, 2002). Its role in progressive CLL is unknown, yet evidence of its upregulation in patients with aggressive CLL from our pro-onconets and its role in TGF-beta signaling, makes it a strong candidate for investigation in its functional significance in progressive CLL.

MiR-181b, Marker of Disease Progression in CLL

Aside from deregulation of pathways due to over expression of certain oncogenes and deletion of tumor suppressor genes, a new type of regulation has been discovered at the post-transcriptional level. This new mechanism of regulation at the mRNA level is due to the discovery of short, non-coding RNA sequences (~22 nt long), referred to as microRNAs (recognized as genes with "miR" as a prefix), that target specific mRNA transcripts via binding on semi-complementary "seed" regions located on the 3' UTR area of the target mRNA (Lai, 2002). In many studies in various cancers, microRNAs have been observed to have tumor suppressor or oncogenic roles.

A number of novel microRNAs have been discovered to be involved in CLL. A few tumor suppressor candidates include miR15-a/miR-16-1 cluster, miR-34a, miR-34b, miR-34c, miR-29, and miR-181 (Nana-Sinkam & Croce, 2010). The expression of these microRNAs has been proven to modulate the expression of anti-apoptotic or tumor suppressor genes such as *BCL2*, *TP53*, *TCL1* and *MCL1* as well as ZAP-70 tyrosine

kinase. In aggressive CLL, miR-29 and miR-181b targeting anti-apoptotic genes like *TCL1* and *MCL1*, are downregulated thus resulting to overexpression of those anti-apoptotic genes (Nana-Sinkam & Croce, 2010). In indolent CLL, miR15-a/miR-16-1 cluster located at the 13q14 region are deleted, those microRNAs are known to target not only anti-apoptotic proteins like BCL2 (Cimmino et al., 2005) and MCL1 (Calin et al., 2008) but also the tumor suppressor, TP53 (Fabbri et al., 2011). The dowregulation of such microRNAs is also associated with increased level of TP53 expression and transactivation of miR34a, miR-34b and miR-34c resulting to reduced level of ZAP-70 expression (Fabbri et al., 2011).

Of the many microRNAs (and counting) found to be associated with CLL, Visone et al. found that the single microRNA, namely miR-181b, was observed to decrease in serial samples of patients who developed progressive disease, but not in serial samples of patients who had CLL that remained indolent (Visone et al., 2011). As a biomarker for disease progression in CLL, it is highly likely that this microRNA may be a missing regulator of a few of the pro-onconets found in our study.

Methods

Patients and cell lines

All twenty-three patients chosen for this project are enrolled in the Chronic Lymphocytic Leukemia Research Consortium (CRC) in the Moores Cancer Center. All patients registered in the CRC Tissue Core have multiple sample collections at different dates. Whole blood sample collections of patients were processed in order to isolate total white blood cells (consisting of >90% CD5+CD19+, assessed by flow cytometry) and then stored in 10% DMSO in liquid nitrogen. Samples ready to be used for experimentation are then removed from liquid nitrogen and then thawed quickly for laboratory use.

Criteria of this cohort included that all samples used for analysis are untreated. This allows evaluation of the disease without any influence of any kind of treatment. Patients who were eventually treated during the course of their disease only had samples chosen prior to treatment, whereas patients with no treatment had samples chosen due to availability of total vials in CRC Tissue bank. For over time studies, a second time point was taken for each patient. Average years between sample collection 1/initial time point (SC1) and sample collection 2/second time point (SC2) was 2.7 years.

RNA extraction

Patient samples had their total RNA content extracted with standard TRIzol Reagent (Invitrogen) method. After extraction of RNA, samples were treated with TURBO DNAse enzyme for removal of any trace DNA remaining after extraction using

the TURBO DNA-free Kit (Ambion) protocol. Quantification of total RNA concentration was done by NanoDrop ND-1000 spectrophotometer. RNA quality was evaluated by Novex® 15% TBE-Urea Gels (Invitrogen) using Novex® 2X TBE Urea Sample Buffer (Invitrogen). Intact and useable RNA had distinctive 5.8S rRNA and 5S rRNA bands and as well as presence of small RNAs at the very bottom. Further validation of intact and useable RNA was done on the Agilent 2100 Bioanalyzer. Range of RNA Integrity Number (RIN) of all samples included in this study was from 7.8 to 10. The RIN is a quantitative grade on the degradation of RNA of which grades less than 7.0 are considered poor RNA quality. Patient samples with good and working RNA were then used to synthesize complementary DNA (cDNA) for further analysis on mRNA expression.

Reverse Transcriptase cDNA synthesis for random cDNA and gene-specific cDNA

Nine microliters (ul) of total RNA was reversely transcribed using 0.5 ug of Oligo(dT)12-18 Primers (Invitrogen), 1 uM of dNTPs, 20nM DTT, 5mM MgCl2, 1x of 10x PCR Buffer, and 200 units of Superscript II Reverse Transcriptase Enzyme (Invitrogen) in order to generate random cDNA. Standard thermocycler conditions consisted of 5 minutes at 65 degrees (after addition of only Oligo(dT) primers to total RNA aliquot) and then 42 degrees for 55 minutes and 70 degrees for 15 minutes after addition of mastermix consisting of the rest of the components mentioned previously. RNAse OUT™ RNAse Inhibitor enzyme was used (Invitrogen) by adding 40 units in each sample and incubated at 37 degrees for 20 minutes. Concentration of synthesized random cDNA was determined by a Nanodrop ND-1000 spectrophotometer. As random

cDNA was utilized for analysis on larger mRNA transcripts, microRNA cDNA required a different protocol. Like described in the methods of Visone et al. (Visone et al., 2011), gene-specific cDNA synthesis for miR-181b and RNU44 were generated with TaqMan® MicroRNA Assays (Applied Biosystems) that included reverse transcription primers specific for miR-181b and RNU44. RNU44 is a small nuclear RNA encoding in the C/D box 44 and was used as the endogenous control only for miR-181b expression in quantitative Real-Time polymerase chain reaction.

Quantitative Real-Time Polymerase Chain Reaction (qPCR)

Samples studied in this project were also evaluated on their mRNA expression of miR-181b and *SMAD2*. Expression of miR-181b via qPCR was conducted using the methods in TaqMan® MicroRNA Assay (mentioned in cDNA synthesis methods) specific for miR-181b and RNU44. TaqMan® Universal Master Mix II with no UNG (Applied Biosystems) was also used for miR-181b qPCR. Eight nanograms of random cDNA were used for detection of *SMAD2* and *GAPDH*. Each sample was done in triplicate on a standard 96-well plate with a non-template control for each gene. *GAPDH* was used as an internal control strictly for larger genes, such as *SMAD2*. QPCR reactions for *SMAD2* used Power® SYBR Green PCR Master Mix (Applied Biosystems). Primer sequences used were from our unpublished work from Chuang et al. in 2012 and are as follows: *SMAD2* forward, GCCTGTTGTATCCCACTGATC; *SMAD2* reverse, GCCATCACCACTCAAAACTGT. *GAPDH* primers were designed in this study and sequences are as follows: *GAPDH* forward, AGCCGAGCCACATCGCT; *GAPDH* reverse, TGGCAACAATATCCACRRRAC. All qPCR reactions were done in an

Applied Biosystems 7900HT Fast Real Time PCR system. Calculation of relative fold change to *GAPDH* used the 2^(-ΔCt) method on individual samples and the 2^(-ΔΔCt) method for relative fold changes over time between SC1 and SC2 (Schmittgen & Livak, 2008).

Transfection of miR-181b mimic in CLL cells

A HiPerFect ®Transfection Reagent (Qiagen) was utilized to introduce mimic miR-181b into CLL cells. For obtain physiological concentrations of miR-181b mimic in CLL cells, mimic negative control and mimic miR-181b (both from Ambion) in reactions had a final concentration of 50 nM and 0.05 nM, respectively. Amount of HiPerFect and RPMI without serum and PSG per reaction were 3 ul and 100 ul, respectively. A 24-well plate contained 2 x 10^6 cells per well. Aside from the mimic negative control, a non-stimulated (NS) control (only having growth media) was also added. Prior to transfection, CLL cells were thawed from storage, counted, and incubated for at least 24 hours over night in 5% CO2 and 37 degree conditions. On transfection day, cells were re-counted, aliquoted on plate, and then incubated in transfection complexes for one hour in 5% CO2 and 37 degree conditions. After the first hour, RPMI media with 10% Fetal Bovine Serum and 1% PSG was added. Five more incubation hours then followed where a total of six hours of incubation time passed. Twenty-four incubations with transfection complexes were incubated for approximately 23 hours more.

Protein lysis and Immunoblot

CLL patient samples and cell lines were lysed with RIPA Buffer (Thermo Scientific) containing 25 mM Tris•HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1 % SDS, and freshly added Halt™ Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific) and 5mM EDTA (Thermo Scientific) to get total cell lysates. Prior to addition of lysis buffer, cells were washed with cold PBS twice. Spinning conditions at 1100 rpm at 4 degrees for 6 minutes or 2000 rpm at 4 degrees for 3 minutes (depending on G-force of centrifuge) remained consistent throughout the entire lysis process, except for certain steps. To obtain more concentrated lysates, approximately 1 x 10^7 cells were treated with 10-70ul microliters of lysis buffer and then sonicated with 50% pulse for 5 seconds 5 times. Incubation in lysis buffer took place in 4 degrees for 15-30 minutes maximum. Lysates were then spun for 15 minutes at 14000 x g in 4 degrees to spin down cellular debris. Quantification of protein lysates were done using a BCA protein assay kit (BioRad).

Proteins on gel were transferred to a nitrocellulous membrane using an iBlot Dry Blot System from Invitrogen. Blocking step used 5% milk in 1X TBST either in the cold and overnight or 30 minutes in room temperature. Primary antibodies of SMAD2 and GAPDH (both from Cell Signaling #3122 and #2118) were diluted 1:2500 in 5% milk in 1X TBST and 1:1000 in 5% Bovine Serum Albumin in 1X TBST, respectively. Anti-CCT4 (Santa Cruz Biotechology) and anti-CSNK2A1 (Sigma-Aldrich) were diluted 1:400 and 1:1000 in 5% milk in 1X TBST as well. Secondary antibody dilution of Anti-Rabbit HRP (Cell Signaling #7074) was 1:10,000 in 5% milk in 1X TBST for one hour in room temperature. Chemiluminescence was detected using a SuperSignal West Femto

Chemiluminescent Substrate (Thermo Scientific) and Pierce ECL Plus Western Blotting Substrate (Thermo Scientific). After probing, membranes were stripped with Restore Plus Western Blot Stripping Buffer (Thermo Scientific) for next probes.

Chapter 1: Disease Phases in CLL have Distinct Transcription Activity

When assigning patients with CLL into clinical courses, one way of determining this is use of the Rai and Binet staging systems, which are basically physical assessments of the patient's lymphocyte counts and organ growth (lymph node and spleen size). These systems, although useful for determining the severity of the disease, are flawed in detecting disease progression at an earlier fashion since they are solely dependant on physical examinations. Furthermore, most studies in defining prognostic markers have investigated risk stratification in CLL based on time between diagnosis (DX) and treatment (TX), which ignores changes in the transcriptome during disease progression. To account these changes, gene expression changes of 130 CLL patients were evaluated at different disease phases of CLL. Figure 1A shows the distribution of patients based on amount of years between DX and sample collection (SC), referred to as DX→SC. Approximately 40 percent of our patient cohort had a gap of 1 year or less of DX→SC, indicating that most patients samples were collected at the early phase. The rest of the samples varied of DX→SC, thus reflecting the heterogeneity of DX→SC in our cohort.

To correspond disease phase to the point in time the samples were collected, we resorted to the amount of years between sample collection and time to treatment, referred to as SC→TX. Sample collections considered "early stage" of CLL had a SC→ TX greater than 4 years, whereas "intermediate stage" sample collections had a SC→ TX less than 4 years but no shorter than 1 year, and "late stage" had a SC→ TX less than one year. With this convention, the variability of disease phases in our cohort were quite widespread, as seen in Figure 1B (disease phases are the colored bars). Most sample collections from our cohort had a SC→TX about 6 months, thus were mainly taken from

the late phase of the disease. For further analysis we wanted to see the amount of patient samples that have a mutated *IGVH* status (Figure 1B, black bars), thus see whether any of the patients observed in this study were stratified with a good prognosis. Based on our results, we observed that a high percentage of patients using a mutated *IGVH* have sample collections at a late phase of the disease. This implicated that these patients were more likely patients who have eventually progressed and thus required eventual treatment. Observing this find also implicated that *IGVH* mutation status was an unreliable marker for progression.

Like single-gene expression studies from other laboratories utilizing microarray techniques to observe gene expression differences between *IGVH* mutation status, we aimed to evaluate the gene expression in our cohort the same way. In our analysis, we counted the number of differentially expressed genes between disease phases and *IGVH* mutation status subtypes. This approach showed that comparison between early and late phase gene expression had the largest number of differentially expressed genes suggesting two different transcriptional activities (Figure 1C, leftmost panel). By comparison between the two *IGVH* mutation subtypes, a large difference of transcriptional activity was also observed. Furthermore, we also compared the gene expression differences between disease phases within the *IGVH* mutated subtype groups (Figure 1C, middle panel). We found that the greater difference resided mainly in the early disease phase only. Interestingly, the number of genes differentially expressed drastically lowered in patients in intermediate phase, and even lower in patients in late phase (Middle panel; Figure 1C). Such a result suggests that transcription activity eventually becomes similar as the disease progresses. The heterogeneity of disease phases

found with either mutated or unmutated *IGVH* subtypes suggests that *IGVH* mutation is a static marker that cannot be associated with a changing transcription activity, thus irrelevant to associate with disease progression. This work has currently been submitted for publication and accredits co-authors of this submitted work listed under Chuang et al., 2012.

Chapter 2: Single-gene Prognostic Markers Cannot Distinguish Disease Phases

Other prognostic markers have been established in other studies, in addition to *IGVH* mutation status. With these other single-gene prognostic marker sets, we tested our cohort of 130 patients on their prognostic power on differentiating low- and high-risk groups in terms of DX→TX, a perspective most studies on prognosis are based on, and SC→TX, a perspective that represents treatment-free survival. As expected, *IGVH* mutation status was the most prevalent predictor with DX→TX, but failed in terms of SC→TX due to non-significant p-values of patients with a SC→TX less than one year (denoted by *) and with more than one year (denoted by #) (Figure 1D, leftmost panel). Again, such a result can be explained by the static nature of *IGVH* mutation status. Furthermore, other established prognostic marker sets failed to differentiate the risk groups in terms of SC→TX and sometimes with DX→TX. These results indicate that single-gene markers were not powerful enough to distinguishing between risk groups based on SC→TX, therefore unfit to predicting treatment-free survival. Evidence of this is seen in two key studies that took sample collection into consideration as seen at the rightmost panel in Figure 1D. This work has currently been submitted for publication and accredits co-authors of this submitted work listed under Chuang et al., 2012.

Chapter 3: The Subnetwork Profile of Aggressive CLL

Since single-gene prognostic markers failed to differentiate low- and high-risk groups in our cohort, we ventured towards a different approach by instead aggregating gene expression data as sets instead of individual gene markers. These sets are comprised of genes encoding proteins that may interact with each other through protein-protein or protein-DNA interactions, thus essentially comprising specific metabolic pathways, or subnetworks. Through bioinformatic analysis and scoring (as described in Chuang et al., 2012), we discovered thirty-eight subnetworks that are significantly active in aggressive CLL (Chuang et al., 2012) (Figure 2A). These thirty-eight subnetworks can be subdivided into two different groups. These two groups are: 1) genes in networks that are increased in aggressive CLL, referred to as the pro-onconets (A-O), and 2) genes in networks that decrease, referred to as the anti-onconets (P-T). Evidence of expression change of these subnetwork genes were mainly evaluated at the mRNA level via real-time qPCR. Interestingly, the pro-onconet subnetworks included pathways involved in cell cycle, apoptosis, and the WNT signaling pathway, all of which are known to be perturbed in CLL. Use of these subnetworks can give rise to potential studies of investigating key pathways and targets for therapy.

More so, we attempted to see the prognostic power of these subnetworks into stratifying our patient samples into their corresponding clinical courses. Based on a supervised clustering, these subnetworks were able to categorize the gene expression data of these patients into three distinct risk groups, which were a high risk group, low risk I, and low risk II (data not shown). Interestingly, the high risk group expression profile had sample collections with a short SC→TX, thus indicating a high association at the late

stage of disease, whereas the two low risk groups had longer SC→TX intervals, thus a high association to the early and intermediate disease phases. Over all, this cluster analysis utilizing subnetworks as prognostic markers represented the gene expression changes that may occur in the transition from indolent CLL to aggressive CLL. This work has currently been submitted for publication and accredits co-authors of this submitted work listed under Chuang et al., 2012.

Chapter 4: Change of Pro-onoconet Subnetwork Genes also Increase Over Time at the Protein Level

At this point, we have exemplified that aggressive CLL has certain subnetworks that increase over time (pro-onconets) and other that decrease over time (anti-onconets). Because of this change of expression, it was curious to see whether protein changes of members of these networks also occurred. To resolve this curiosity, protein expression analysis of some pro-onconet subnetwork genes were conducted via western blot. Such subnetwork members evaluated included, *CSNK2A1*, *SMAD2*, *CCT4, SUPT3H*, *SKP2*, *c-MYC*, and *DMD*. Optimization of antibodies against the encoded proteins of these genes included clean detection of the protein of interest (testing specificity of these antibodies), ensuring bands detected were at the correct molecular weight (also in correspondence to postive controls), and detection of expression in CLL samples. Of the seven antibodies tested, three were successfully optimized. These three antibodies, anti-CSNK2A1, anti-SMAD2, and anti-CCT4 were used to screen for protein expression of a few progressive patient samples (looking at both the early time point and last time point sample collections) also observed in our subnetwork analysis. As shown in Figure 2B, CCT4 and SMAD2 protein expression was observed to increase over time, just the same as it was observed at the mRNA level. CSNK2A1 had no distinguishable difference at the protein level (data not shown), which can be due to the limitations of the assay used. In addition, we also noted the *IGVH* mutation status of these samples that had increases of CCT4 and SMAD2 proteins, which included both *IGVH* mutation subtypes. Over all, we were able to conclude that the changes of expression of genes belonging in our subnetworks also

can happen at the protein level. This work has currently been submitted for publication and accredits co-authors of this submitted work listed under Chuang et al., 2012.

Chapter 5: *SMAD2* is a Predicted Target of miR-181b

In relation to the subnetworks found through the approach of aggregating gene expression as gene sets, this data implicates that there might be some dysregulation occurring during disease progression. Such dysregulation can be caused by different mechanisms, which includes regulation by microRNAs. MicroRNAs are regulators solely involved at the transcript level. As described in the Introduction section of this thesis, microRNAs are small RNA segments about 22 nucleotides long that target mRNA transcripts at the 3'UTR region, which then marks them for degradation or prevention from translation.

A global microRNA expression study observing the microRNA expression changes of two sequential time points in both progressive and indolent CLL patient samples was conducted by the Visone et al. group from Ohio State University (Visone et al., 2011). From this study, it was revealed that a single microRNA, namely miR-181b, significantly decreased over time, which also was verified via real-time qPCR (Visone et al., 2011). Furthermore, they displayed that progressive CLL patients samples generally have a lower expression of miR-181b in comparison to indolent CLL patients samples. Because of this find, we hypothesized that miR-181b may be a missing regulator in the pathways suggested by our subnetwork analyses.

Utilizing the DIANA microT 3.0 database (Maragkakis et al., 2009) we performed a search of all target genes of miR-181b using a "loose" threshold setting of 7.3. This threshold setting only shows targets with a total miTG score of 7.3 and above, of which a total miTG score is a prediction score that indicates the likelihood that the gene of interest is a possible target, thus the higher the total miTG score, the more likely

a possible target. Our search found 1705 targets in 735 genes for human miR-181b (data not shown). With this data, we attempted to find targets that belonged in the pro-onconet subnetworks. As listed in Table 1, three genes from the pro-onconet subnetworks were observed as predicted targets of miR-181b. Interestingly, one of these predicted targets, namely SMAD2, was a member part of a subnetwork that involves interaction with c-MYC (Figure 2A, subnetwork F), an important cell cycle regulator and a widely known oncogene observed to be over expressed in many cancers. Because of this target was centralized player of this particular subnetwork, further investigation of regulation of this gene was conducted.

Based on results generated from DIANA microT 3.0, miR-181b could bind at three different positions on the 3' UTR of *SMAD2* (Figure 3A) with a total miTG score of 9.88. Although the highest scoring predicted target (*TIMP3*) had a total miTG score of 11.34, *SMAD2* as a target was fairly possible due to other studies that have confirmed miR-181b directly targeted genes such as *AICDA* (De Yébenes et al., 2008) and *BCL2* (Zhu, Shan, Wang, Shu, & Liu, 2010), which in this public database have lower total miTG scores than *SMAD2* (data not shown). As *SMAD2* is proposed target of miR-181b, we hypothesized that miR-181b may be a regulator in pro-onconet pathway involving *SMAD2* (Figure 3B).

Chapter 6: Studies of MiR-181b and *SMAD2* Expression Over Time

To investigate the relationship between miR-181b and *SMAD2* during disease progression, we sought to observe messenger RNA (mRNA) expression of miR-181b and *SMAD2* via real-time qPCR in 18 CLL patients. Primer sets for *SMAD2* and *GAPDH* (reference gene) were first tested using a melting curve analysis with an evaluation of their products on an agarose gel (Figure S1). Sample collections at two different time points before the date of initial treatment were analyzed in every patient. As indicated in Tables 2 and 3, patient sample collections were calculated of the amount of years from sample collection to treatment (SC→TX). Samplings collected from treated patients had an initial time point collected from either the early phase and intermediate phase as all last time points collected were at a late phase of the disease. Thus, these sample collections were evaluated during disease progression. Patients who had an indolent disease course (not treated) were also included in our cohort for screening of miR-181b and *SMAD2* expression.

Because miR-181b was established as a marker for disease progression in CLL in the Visone et al. study, we investigated our cohort in the same terms as this paper. These terms, which were described as "properties", consisted that patients undergoing disease progression undergo a decrease equal or greater than 50% of miR-181b expression over time and/or have an initial miR-181b relative expression of 0.005 or less (Visone et al., 2011) .Within our cohort about 56% (10 out of 18) have these "properties" with most patients undergoing a 50% or more decrease in miR-181b relative expression over time and only two patients had an initial miR-181b relative expression of less than 0.005 (Figure 2S). Sample collections with large decreases also occurred in samples that had a

low miR-181b relative expression at the first initial time point, thus suggesting that decrease of miR-181b may be an affect of disease progression and not due to baseline expression. Notably, patients with "properties" were mostly treated (Tables 2 and 3). This shows that eventual loss of miR-181b expression contributes disease progression.

Along with miR-181b expression, we also evaluated the mRNA expression of *SMAD2*, a gene observed to be upregulated in aggressive CLL in our pro-onconet subnetworks and as well as a predicted target of miR-181b in silico. As represented in Figure 3S, majority of patients with a significant increase of *SMAD2* expression mainly resided in patients with "properties". A significant increase or decrease was considered to being at least 50% change from the initial time point's relative expression level. Those who experienced a significant increase of *SMAD2* over time, 83% of these patients belong to those who had "properties" whereas one patient belonged to those without. The frequent occurrence of a significant decrease (50% and above) of miR-181b and significant increase in *SMAD2* simultaneously suggests that perhaps that miR-181b may regulate *SMAD2*. All other cases that may not suggest this regulation are possibly due to the genetic heterogeneity amongst the different patients and as well as the possibility of over abundance of other transcripts for miR-181b. For the most part, these data suggest that increase of SMAD2 expression is an event that occurs during disease progression, which is what our subnetworks have shown.

Chapter 7: MiR-181b can Modulate *SMAD2* if Target is Abundantly Expressed

To confirm that miR-181b modulates *SMAD2* expression, we introduced miR-181b, via Hiperfect Transfection Reagent kit, in the cases that had a significant decrease in miR-181b and a significant increase in *SMAD2*. Viabilities were accounted before and after 24 hour incubation without treatment, which ranged from 98% to 87% throughout the whole duration of the experiment. Thus, viabilities after transfection were assumed to be just as good. To ensure that introduction of miR-181b mimic in CLL cells resulted in expressions at physiological levels detected in CLL, a test of optimal mimic concentrations were conducted (Figure 5A). Serial dilutions from starting point concentration of mimic, 20 nM, were made in 1:100 (0.2 nM), 1:1000 (20 picoM), and 1:10,000 (2 picoM) dilutions. After 6 hrs of introduction of mimic miR-181b, RNA was collected and then further processed for preparation on real time qPCR. Avoiding contamination of non-transfected mimic miR-181b and false positives, cells collected were washed with PBS before extracted for RNA. This test showed that a concentration of 20 picoM was the proper concentration to express high miR-181b expression and also stay within the normal physiological expression range seen in CLL.

As the ideal concentration was optimized, a time course of transfection in CLL cells was tested for the optimal time to introduce a maximum amount of miR-181b. Due to the common fact that CLL cells in culture tend to apoptose drastically after 48 hours, our time course only tested incubation durations of only 6 hours and 24 hours. Viabilities of samples were calculated before and after the initial 24 hour incubation in culture without treatment. Controls included in this experiment were a Non-Stimulated (NS) population, which only consisted of growth media, and a population introduced to a

mimic negative control (scrambled sequence), which allowed us to observe any effects by transfection alone and as well as show the baseline of miR-181b expression for comparison. According to our results from both 6-hour and 24-hour incubations, significant elevation of miR-181b expression was observed in all samples at both time points (Figure 5B). NS and mimic negative controls were very similar in expression, thus indicating that mimic negative controls reflected baseline expression of endogenous miR-181b. P-values resulted from comparing a difference between the negative control population and mimic transfected population were all very significant. Yet, observation of a maximum expression of miR-181b transfected occurred in 24-hour incubations (Figure 5B, two rightmost sets of columns), which, additionally, also had less variability in expression in comparison to 6 hours of incubation.

Along with successful transfection of miR-181b mimic in CLL cells at 24 hours, we also observed *SMAD2* expression to test if miR-181b modulates *SMAD2* mRNA. From our results, *SMAD2* mRNA expression, detected on real-time qPCR, was variable in our mimic miR-181b transfected samples (Figure 5C). Observed in this experiment was an inconsistency in regulating *SMAD2* transcript. Comparison between samples of CLL 4 and CLL 26 showed that regulation of *SMAD2* by miR-181b displayed no regulation and down regulation, respectively. Notably, the normal expression levels of *SMAD2* in these patient samplings (meaning that these samples were extracted of their RNA from the time of thawing) were relatively different (Figure 5D). CLL 4 had a *SMAD2* relative expression of 0.39 whereas CLL 26 had a *SMAD2* relative expression of 2.63. Such differential expression made it reasonable why *SMAD2* expression was possible to be down regulated by introduction of miR-181b in CLL 26 and not in CLL 4.

Due to these confounding results, this experiment suggests that miR-181b regulation of target transcripts may be determined by the abundance of target transcript. Between specific patient samplings, we concluded that variability of modulation of *SMAD2* is most likely due to the expression abundance of other miR-181b targets.

Discussion

Determining disease progression early in CLL has been a challenge due to the limitations of static markers, such as *IGVH* mutation status and ZAP-70 protein expression, and as well as late detection of progression by the clinic-based staging systems, such as the Rai and Binet systems. Furthermore, gene expression studies of the most accepted and robust prognostic marker, *IGVH* mutation status, has proven to be inconsistent on what genes are differentially expressed between the subtypes (Klein et al., 2001; Rosenwald et al., 2001). In our study, which was submitted for publication, we show that transcriptional activity towards a more aggressive stage of disease becomes similar, which then allowed us to use an alternative type of analysis described by Ideker et al. in 2002. This network-based analysis revealed key pathways that could be responsible for the pathogenesis of CLL (Figure 2A). In addition, we have shown that the encoded proteins of these pathways can also change over time, thus corresponding with the gene expression data (Figure 2B). Furthermore, we wanted to ask whether dysregulation of transcription regulation mechanisms could be a cause as to why we observe these changes of expression over time. Because miR-181b was observed to decrease significantly in progressive patients over time (Visone et al., 2011), we also monitored the expression behavior of this particular microRNA in our own cohort and correspond this microRNA to the genes found in our subnetworks. Interestingly, we found that miR-181b can regulate *SMAD2*, a member of the pro-onconet subnetworks, but depending on how much this target transcript is abundantly expressed (Figure 5C and 5D). Over all, these results open avenues for further investigation on SMAD2 function in

CLL, target favorability by microRNAs, and studies on other regulatory mechanisms at the transcription and gene level that may contribute to disease progression in CLL.

The fact that we were able to observe SMAD2 increase at the protein level suggests that there could be oncogenic function of SMAD2 in CLL. SMAD2 is a receptor-mediated SMAD of the TGF-Beta signaling pathway. Receptor-mediated SMADs are activated by phosphorylation via TGF-Beta and activin type I receptors, which then allow them to form complexes with cofactor-SMADs to act as a transcription factor complexes (Itoh et al., 2000); downstream targets depend on cell type, yet function of TGF-Beta generally for B-lymphocytes is cell cycle arrest and apoptosis (Lebman & Edmiston, 1999). Most studies have exemplified the importance of TGF-beta in mediating apoptosis, thus stating the tumor-suppressive functions of this pathway (Derynck, Akhurst, & Balmain, 2001; Douglas et al., 1997; Sebestyén et al., 2005). Despite these studies, there has only been one study to show otherwise (Baughn et al., 2009), which in our case, we have also shown adverse function of the TGF-Beta pathway at the signal transduction level. In CLL, there has been evidence of various degrees of apoptotic resistance against TGF-Beta stimulation due to variable expression of TGF-Beta receptors between individual CLL patients (L Lagneaux et al., 1997). Additionally, another study has described mutations of the TGF-Beta receptor that may contribute to this apoptotic resistance (Schiemann et al., 2004). Yet, reports have not confirmed any SMAD-signaling dysfunction in CLL. In this study, we have demonstrated that the downstream signal transduction mechanisms of TGF-Beta activation may also be a contributor to apoptotic resistance in aggressive CLL. A clue as to how SMAD2 can be anti-apoptotic may possibly be through targeting pro-survival genes. Evidence of CDK2

phosphorylation at the threonine 8 residue of SMAD2 protein, an abnormal site for SMAD2 post-translational modification, has been observed in bone marrow myeloma cells (Baughn et al., 2009). This modification may alter the binding of SMAD2 with its common binding partners and bind on other cofactors, thus resulting to transcription of different gene targets, which may possibly be pro-survival molecules. Future studies examining the phosphorylation of SMAD2 protein and its potential downstream targets in CLL can validate this insight.

As miR-181b was exemplified to regulate *SMAD2* transcript, it is highly possible that miR-181b may regulate other transcripts simultaneously expressed. MicroRNAs bind semi-complementary to their target transcript, thus suggesting up to hundreds of possible targets for each microRNA. Public databases, such as TargetScan, DIANA microT 3.0, miRbase.org, ncRNA.org, and etcetera, confirm this notion of numerous targets for a single microRNA. Although these sites give only predictions based on sequence homology of the microRNA and the 3' UTR region of target transcripts, they exemplify how complicated microRNA function can be. Our data suggests regulation by microRNAs may be dependent on the abundance of target transcript. Additionally, there has been evidence of this particular phenomenon in a study lead by Arvey et al. at the Memorial Sloan-Kettering Cancer Center in 2010. Multiple transcript targets for miR-181b in CLL are present and abundantly expressed. Such factors include *BCL-2* (Otake et al., 2007; Zhu et al., 2010), *TCL-1* (Pekarsky et al., 2006), and *AICDA* (De Yébenes et al., 2008; Palacios et al., 2010), all which are confirmed targets of miR-181b and over expressed in CLL. Conclusively, *SMAD2* transcript expression in CLL patients may vary across the board, thus miR-181b may not always regulate this transcript when others may

be highly expressed. Because of this situation, it is highly recommended to determine relative expression levels of genes targeted by miR-181b in CLL. With this data the true target(s) of miR-181b in CLL may come to light. Furthermore, other microRNAs may be variably expressed between different patients, thus other microRNAs may possibly regulate miR-181b target transcripts. Over all, the study of microRNAs in any disease is a complicated situation, thus research on the regulation dynamics of microRNAs is very important in order to know what microRNAs target which transcripts and which microRNAs solely regulate those transcripts.

We have exemplified that microRNAs, specifically miR-181b, can regulate genes in our subnetworks, thus in turn, translate as an overall regulatory effect on the entire subnetwork. Due to this evidence, it is plausible to hypothesize that other microRNAs may also regulate our subnetworks, thus aberrant expression of these microRNAs may result to the aberrant expression of the genes observed in aggressive CLL. The indication of microRNAs having tumor suppressor functions have been first described in CLL (Calin et al., 2008). Patients who have acquired 13q deletions over time have also lost miR-15a/16-1 expression, which are the microRNAs that are encoded in this region. Through searching possible targets of these microRNAs in public databases, pro-onconet members, *MAP2K1*, *IRAK2*, and *SMAD2*, are proposed targets. The loss of expression of miR-15a/16-1 thus result to the over expression of these factors in CLL, as observed in our subnetworks. Acquired loss of tumor suppressor microRNAs may occur during the coexistence of microRNAs functioning as oncogenes. In CLL, it has been observed that aggressive CLL patient samples have an over expression of miR-155 (Marton et al., 2008). Conversely, this microRNA targets a number of genes belonging in the anti-

onconet subnetworks (Figure 2A, subnetworks P-T). Such genes include, *MYB*, *CEBPA*, and *SP1*. Validation of these microRNAs with our subnetwork genes is necessary to relate this microRNA story in CLL with our subnetwork story in order to piece together the events responsible for disease progression.

In addition to other microRNAs, along with miR-181b, epigenetic regulation can also be responsible for the significant change of expression of our subnetwork. Epigenetics is way of regulation at the gene level where methylation and acetylation of histones (enzymes that spool DNA together) repress and activate gene transcription, respectively. These protein modifications can either compact the DNA to prevent transcription or "loosen" the DNA to allow transcription. In addition to the two types of modifications mentioned above, phosphorylation, ubiquination, and sumolyation are also the types of modifications to occur on histones. Epigenetic regulation is a genome-wide regulation, thus this form of regulation basically governs the function of the cell to its entirety. Each cell and tissue type has its own "epigenetic code"—its own combination of these modifications which then result to the differential expression of certain genes between cell and tissue types (Turner, 2007). In relation to CLL, there has been evidence of particular genes that are aberrantly methylated in comparison to normal B-cells (Kuang et al., 2008; Tong et al., 2010). Although these cited studies that have determined these epigenetic mishaps in CLL, studies determining whether these epigenetics are acquired over time have not been done. Studying the epigenetic changes during disease progression is a fruitful addition to answer why some subnetworks are aberrantly expressed in aggressive CLL.

Retrospectively, our research on *SMAD2* and miR-181b roles during the transition from indolent to aggressive CLL have opened avenues of research that are paramount understanding why progression happens. These future studies may lead us to discover key markers of disease progression in CLL, which in turn may allow better therapeutic advances and early detection of progression.

Appendix

Figure 1: Disease phase cannot be corresponded to *IGVH* mutation status and other established prognostic markers. A) Out of a cohort of 130 patients the amount of years between diagnostic date (DX) and sample collection (SC), referred to as DX→ SC, was calculated per patient sample and then the proportion of patients vs. DX→SC was graphed. Over 40% of our cohort have a DX→SC of a year or less. Median amount of years is 1.4 years. B) With the same patients, the amount of years between time of SC and treatment (TX), referred to as SC→ TX, was also calculated. Patients with an SC→TX greater than 4 years were considered "early stage" (E; green bars), SC→TX equal or less than 4 years but no shorter than 1 year were considered "intermediate stage" (I; yellow bars), and SC→TX less than 1 year were considered "late stage" (L; red bars). Majority of the patients in our cohort have sample collections taken at the late stage of CLL. Corresponding to that, we evaluated proportion of patients that have mutated *IGVH* (black bars). C) With gene expression data, we sought to calculate the number of genes differentially expressed between the disease phases and *IGVH* mutation status. Leftmost panel red bars show comparisons between early and late (E vs. L) disease stages and within the same phase (E vs. E and L vs. L). Comparisons of genes differentially expressed between *IGVH* mutation status was also evaluated (blue bar). Further analysis between disease phases based on categorization by *IGVH* status is seen in the middle panel (all blue bars). Rightmost panel are permutation tests performed to find false positives. D) Furthermore, analyses on treatment-free survival were performed. Graphically displayed are the p-values of the difference between low- and high-risk groups when using DX→TX (blue bars) and SC→TX (red bars) amount of years. Leftmost panel shows prognostic power of *IGVH* mutation status alone, whereas * denotes patients with an SC→TX of less than a year and # denotes an SC→TX more than a year. Middle panel shows prognostic power of markers established from other studies. Rightmost panel show the only two studies that have taken sample collection to account. This work has currently been submitted for publication and accredits co-authors of this submitted work listed under Chuang et al., 2012.

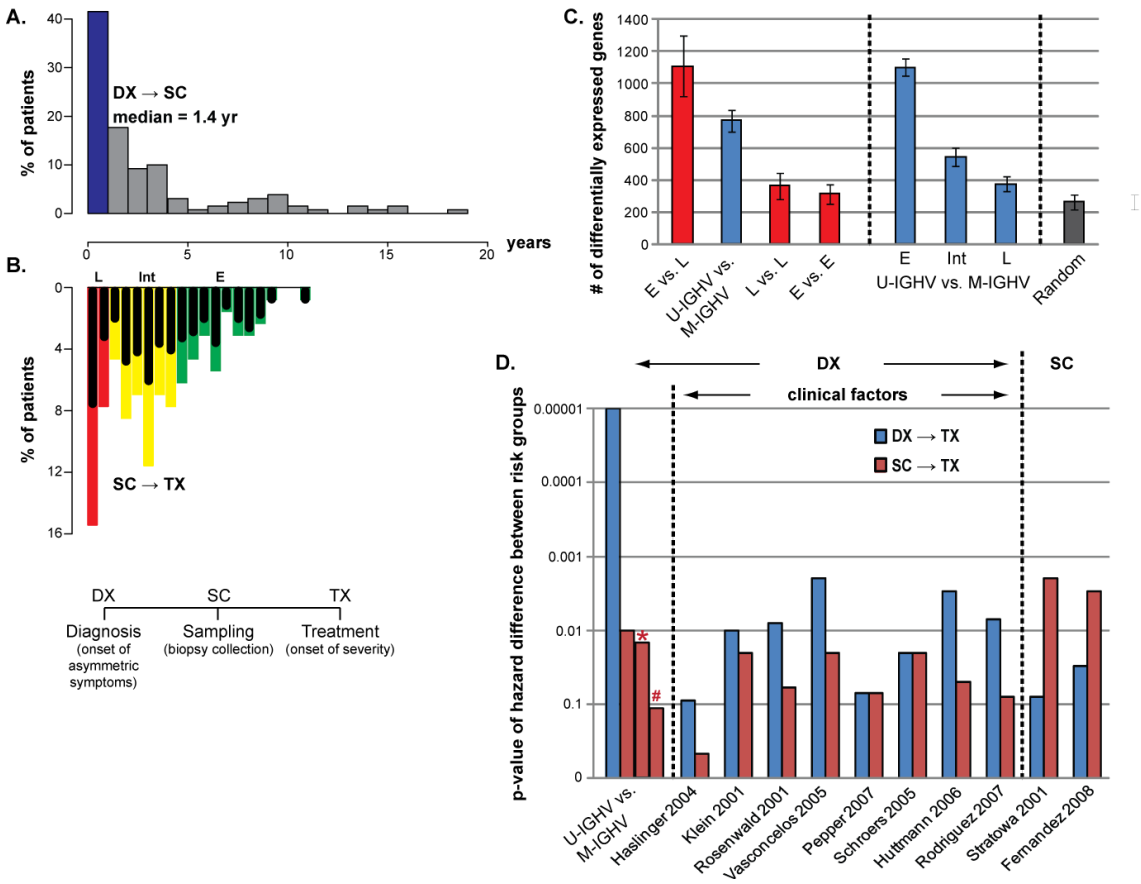
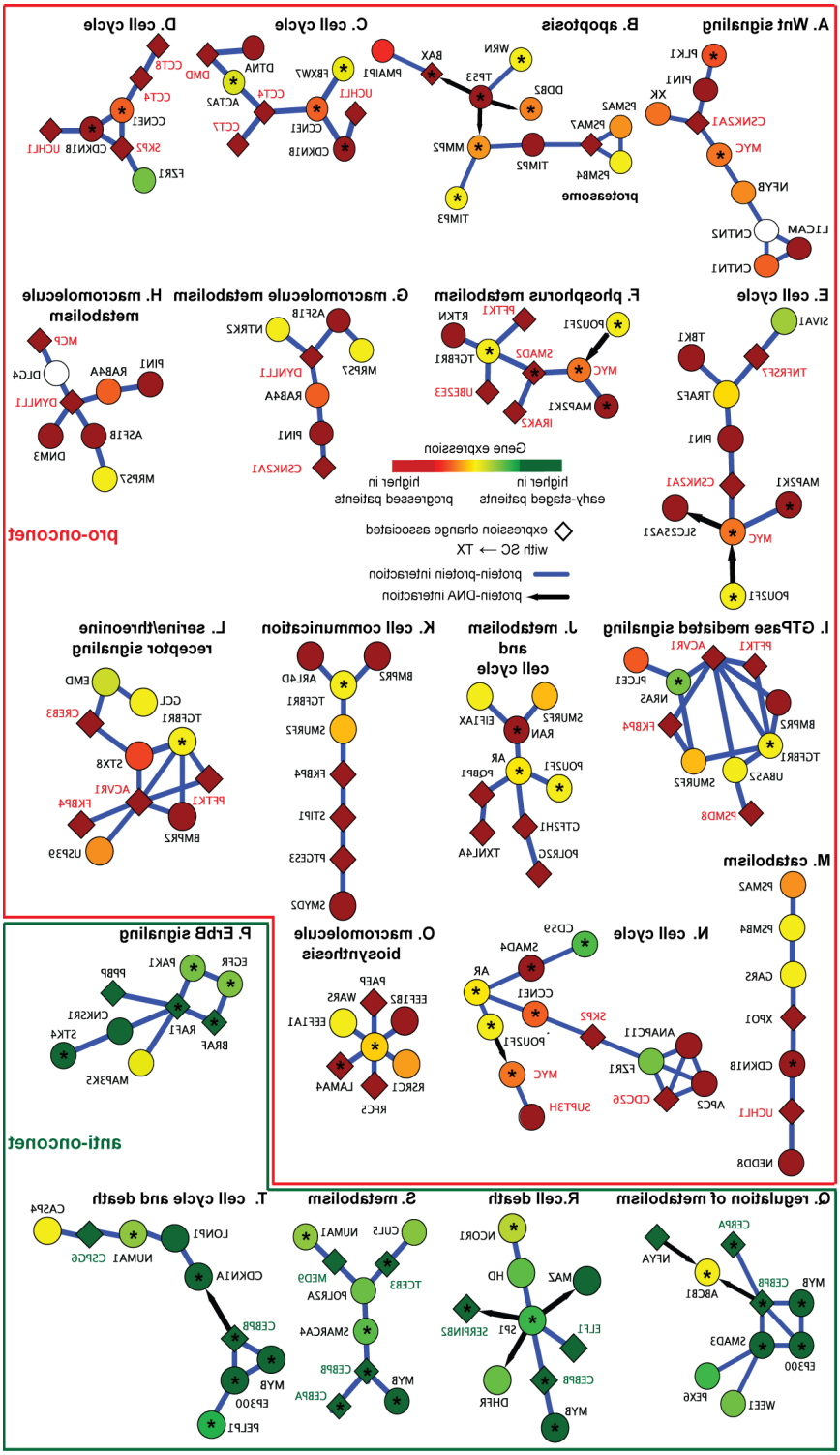
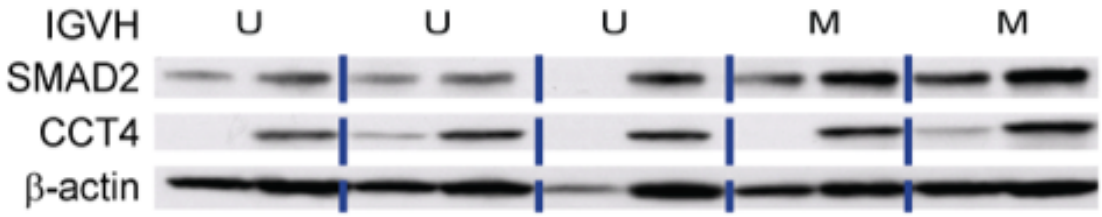

A.
% of patients
DX → SC
median = 1.4 yr
years
B.
L
Int
E
% of patients
SC → TX
DX
SC
TX
Diagnosis
(onset of asymmetric symptoms)
Sampling
(biopsy collection)
Treatment
(onset of severity)
C.
# of differentially expressed genes
E vs. L
U-IGHV vs. M-IGHV
L vs. L
E vs. E
E
Int
L
U-IGHV vs. M-IGHV
Random
D.
DX
SC
clinical factors
DX → TX
SC → TX
p-value of hazard difference between risk groups
U-IGHV vs. M-IGHV
Haslinger 2004
Klein 2001
Rosenwald 2001
Vasconcelos 2005
Pepper 2007
Schroers 2005
Huttmann 2006
Rodriguez 2007
Stratowa 2001
Fernandez 2008

Figure 2: Thrity-eight subnetworks were identified to be significantly active in aggressive CLL. A) Thirty-eight subnetworks found to be significantly active in aggressive CLL. Gene expression and subnetwork-based analysis on aggressive and indolent CLL patients revealed 22 subnetworks are significantly upregulated (referred to as pro-onconets, subnetworks A-O) and 16 other subnetworks significantly downregulated (referred to as anti-onconets, subnetworks P-T) in aggressive CLL. Protein-protein interactions are labeled as nodes as DNA-protein interactions have directed ends. B) To observe whether protein expression also changes over time of a few molecules belonging in the pro-onconet subnetworks, western blot analysis was done. SMAD2 and CCT4 protein expression are observed at two sequential time points of progressive patients. Different patient sample collections are divided by the dashed lines. Above, "U" denotes as patients using an unmutated IGVH as "M" denotes as patients using a mutated IGVH. Internal control used was Beta-actin. This work has currently been submitted for publication and accredits co-authors of this submitted work listed under Chuang et al., 2012.

A.


A. Wnt signaling
B. apoptosis
C. cell cycle
D. cell cycle
E. cell cycle
F. phosphorus metabolism
G. macromolecule metabolism
H. macromolecule metabolism
I. GTPase mediated signaling
J. metabolism and cell cycle
K. cell communication
L. serine/threonine receptor signaling
M. catabolism
N. cell cycle
O. macromolecule biosynthesis
P. ErbB signaling
Q. regulation of metabolism
R. cell death
S. metabolism
T. cell cycle and death
pro-oncoret
anti-oncoret
proteasome
Gene expression
higher in
early-stage patients
higher in
progressed patients
expression change associated
with SC → TX
protein-protein interaction
protein-DNA interaction

B.



(Figure 2 continued)

Figure 3: MicroT v3.0 Web Server (Maragkakis et al., 2009) predicts SMAD2 as a target of hsa-miR-181b. A) MiR-181b binds on SMAD2 3'UTR in 3 different ways in silico using the DIANA microT v3.0 Web Server. Binding type is the number of nucleotides that bind together within the first 9 nucleotides of the alignment and as well as presence of any G:U wobble pairs. The score for each alignment is based on their contribution of the total miTG score, which is a prediction score measuring the likelihood as possible target. Overall, SMAD2 as a target of miR-181b is pretty likely with a total miTG score of 9.88. Conservation value is the amount of species this sequence is conserved. B) Diagram of the SMAD2 subnetwork from the pro-onconets and miR-181b as a regulator of this pathway. This diagram was generated by the same program used in the submitted works of Chuang et al. in 2012.

A.

	Binding Type	UTR position	Score	Conservation	Also predicted by
⊟	8mer(pos 1)	8066 - 8094	2	2	

Position on Chromosome 18: 43614101 - 43614129
Conservation Info loxAfr1, monDom4

```
3'UTR:  5'  A    UUU   C          3'
            CAUU      C GUGAAUGUU
            ||..      | .|||||||||
            GUGG      G UACUUACAA
miRNA:  3' UGG    CUGUC U          5'
```

⊟	7mer(pos 2)	2708 - 2736	1	0	

Position on Chromosome 18: 43619459 - 43619487
Conservation Info Not Conserved

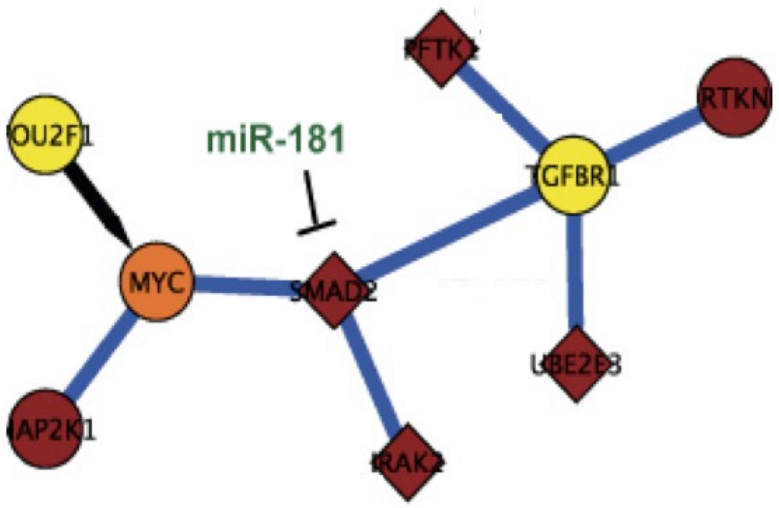
```
3'UTR:  5'          UU   AUUACAUUUGU        G 3'
            CUGG  GUA            GUGAAUGU
            |.|.  |.|            .|||||||
            GGCU  CGU            UACUUACA
miRNA:  3' UGGGU    GU                   A 5'
```

⊟	7mer(pos 1)	4551 - 4579	1	1	

Position on Chromosome 18: 43617616 - 43617644
Conservation Info canFam2

```
3'UTR:  5' U      AU   U          3'
           AUCUGC  AUA CA  GAAUGUU
           |.|..|  |.| ||  |||||||
           UGGGUG  UGU GU  CUUACAA
miRNA:  3'       GC   C  UA       5'
```

B.



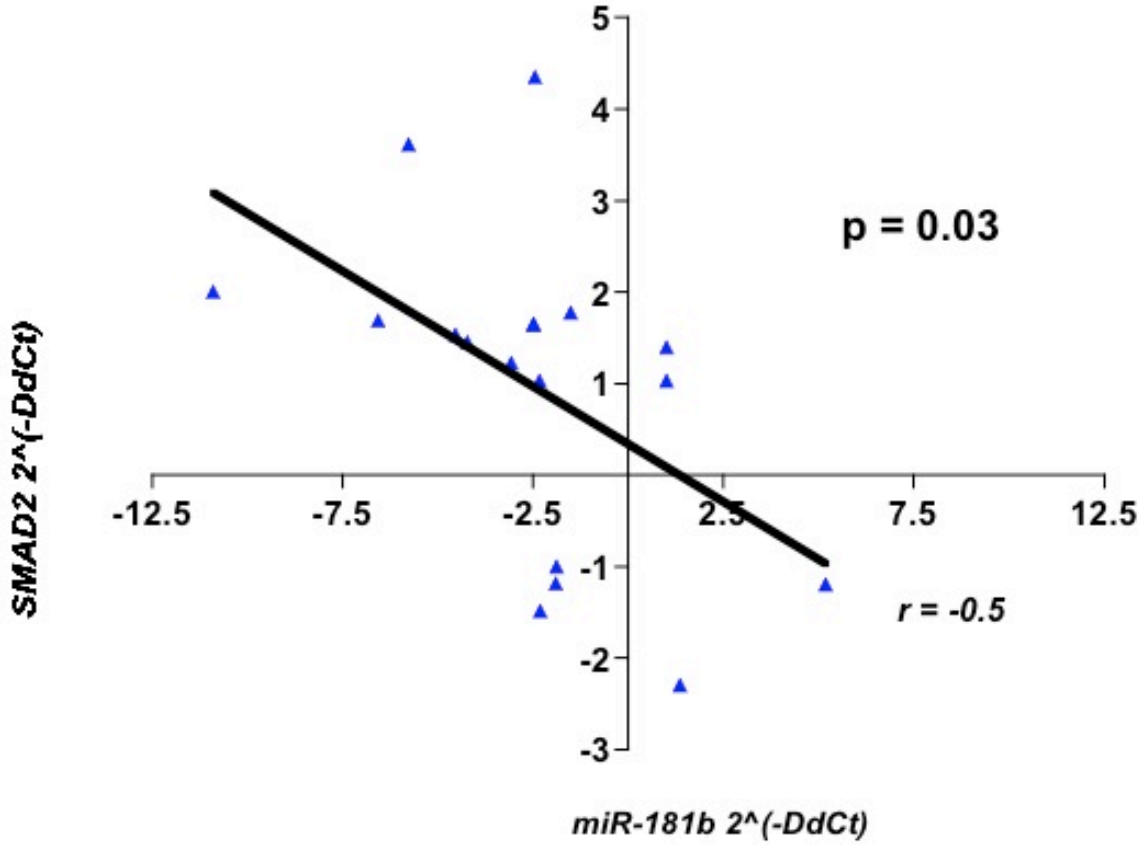
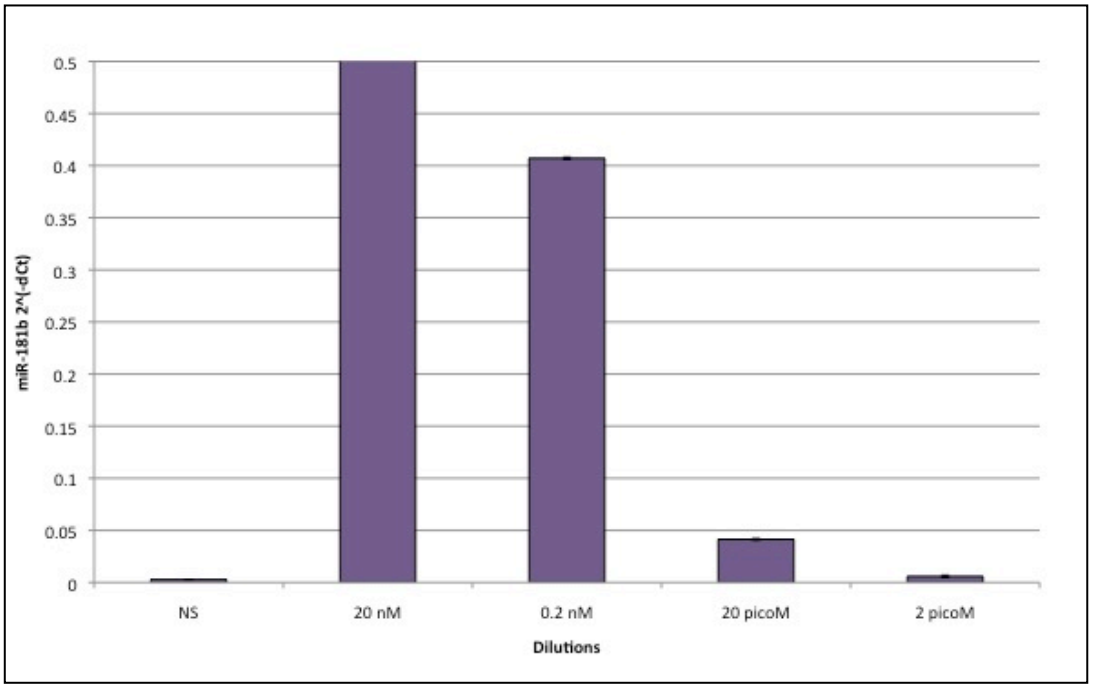


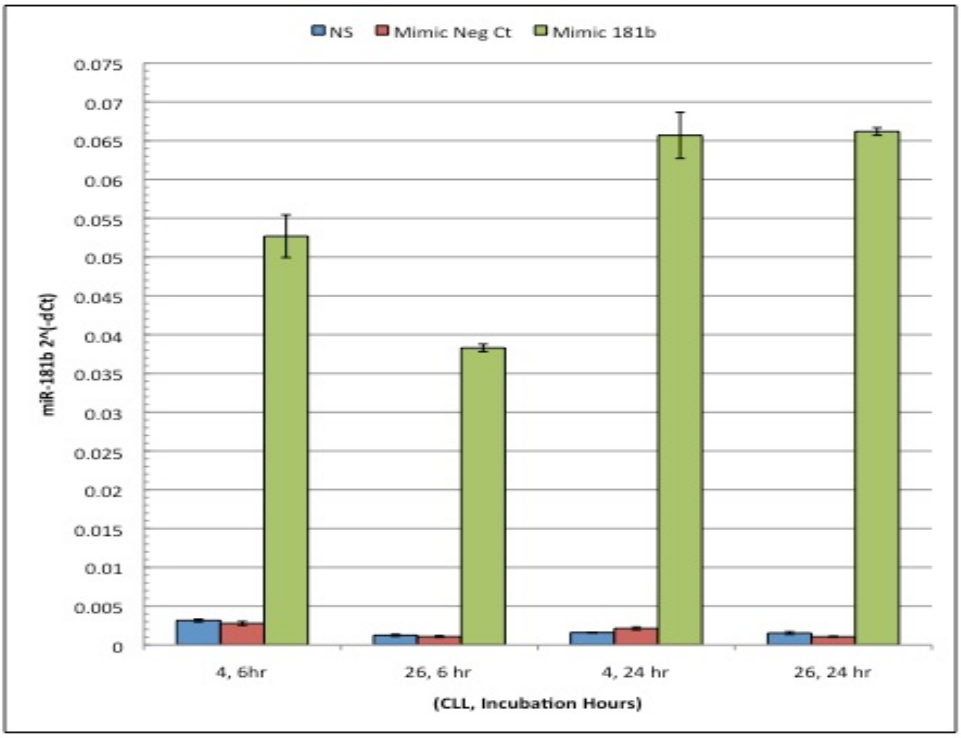
Figure 4: Inverse correlation of miR-181b and *SMAD2* mRNA change of expression over time. Each individual triangle represents different patients. Plotted are the 2-DdCt values of both miR-181b and SMAD2, which is a value that indicates the fold change of expression change over two sequential time points. Calculating these values refer to the methods described in Schmittgen and Livak, 2008. X-axis is for miR-181b change of relative expression over time as Y-axis is for *SMAD2* change of relative expression over time. Linear regression analysis displays an inverse correlation between changes of expression of these genes during disease progression, as verified by an R value of -0.5 and p-value of 0.03.

Figure 5: Progressive CLL cells introduced to miR-181b mimic can modulate *SMAD2* mRNA. A) To ensure appropriate concentrations of miR-181b mimic were transfected into CLL cells, we performed serial dilutions of miR-181b mimic starting from 20 nM. These dilutions were 1:100 (0.2 nM), 1:1000 (20 picoM), and 1:10,000 (2 picoM). Graphed are the relative expressions values of miR-181b expression for each condition. Relative expression values were calculated using the $2^{-\text{Delta Ct}}$ method (Schmittgen & Livak, 2008). B) Validating that 20 picoM concentration gave a high miR-181b expression at physiological levels in CLL, a time course of transfection was conducted at 6 hours and 48 hours. Relative miR-181b expression of Non-stimulated (NS) cells are the blue bars, cells treated with the scramble negative control mimic (Neg Ct) are the red bars, and cells introduced with miR-181b mimic (Mimic 181b) are the green bars. C) *SMAD2* relative expression was also monitored at the different treatment conditions after 24 hours of transfection. Colored bars represent the same conditions as the previous graphs. D) *SMAD2* relative expression of patient samples chosen for transfection. This data was from our initial screening of *SMAD2* expression via real-time qPCR. All error bars plotted are the standard deviations of relative expression ($2^{-\text{Delta Ct}}$) of 3 samples.

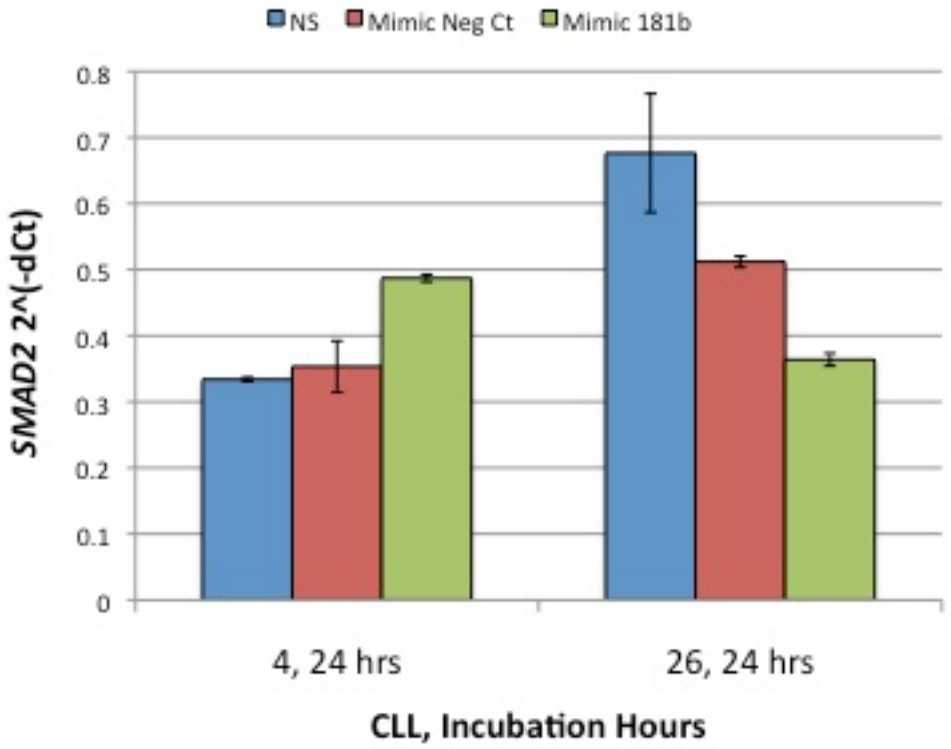
A.


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
miR-181b 2^(-dCt)
NS
20 nM
0.2 nM
20 picoM
2 picoM
Dilutions

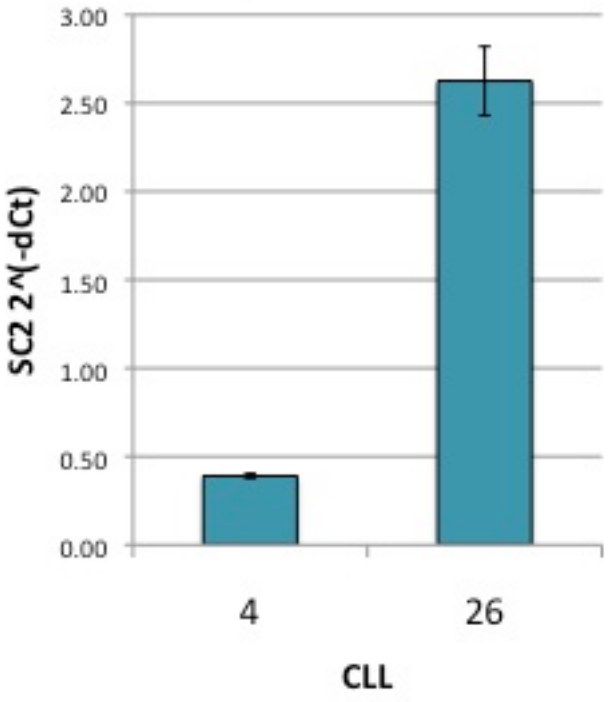
B.


NS
Mimic Neg Ct
Mimic 181b
0.075
0.07
0.065
0.06
0.055
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
miR-181b 2^(-dCt)
4, 6hr
26, 6 hr
4, 24 hr
26, 24 hr
(CLL, Incubation Hours)

C.



D.



(Figure 5 continued)

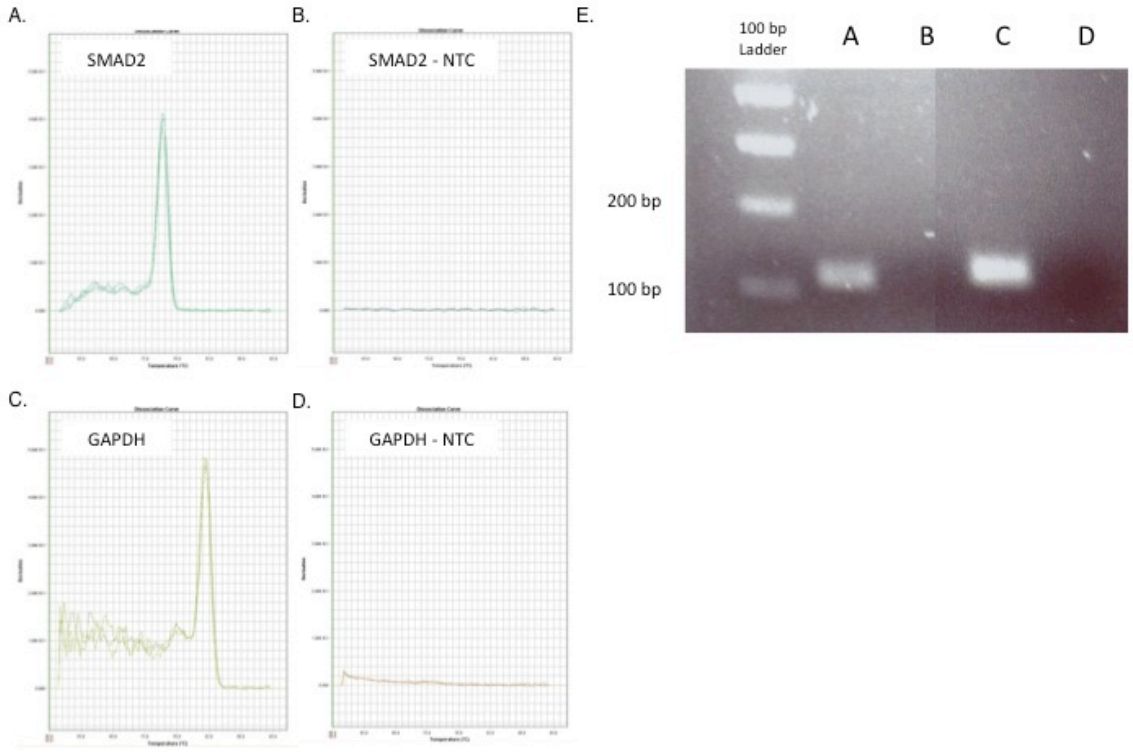


Figure 1S: Melting curve analysis for real-time qPCR of *SMAD2* and *GAPDH* primers. Each graph represents the negative first derivative of temperature vs. fluorescence disassociation curve, which indicates the melting temperature of the product synthesized. X-axis scales for temperature as Y-axis scales for the derivative. For both *SMAD2* and *GAPDH* primers, one peak from each primer set was observed at both at different temperatures, thus indicating different products (A and C). Non-template controls (NTC) for each primer set were tested per reaction plate for our experiment, which as expected, showed no sign of products (B and D). Products synthesized reactions were run in a 2% agarose gel for evaluation of amplicon size (E). Each lane, denoted as A, B, C, and D, represent the products made from the corresponding graphs in this figure. This verified that *SMAD2* and *GAPDH* products were the correct amplicons. As evaluated with Ensembl, our *SMAD2* and *GAPDH* primers would synthesize amplicons of 111 and 122 bp of length, respectfully. Our gel results verified these lengths.

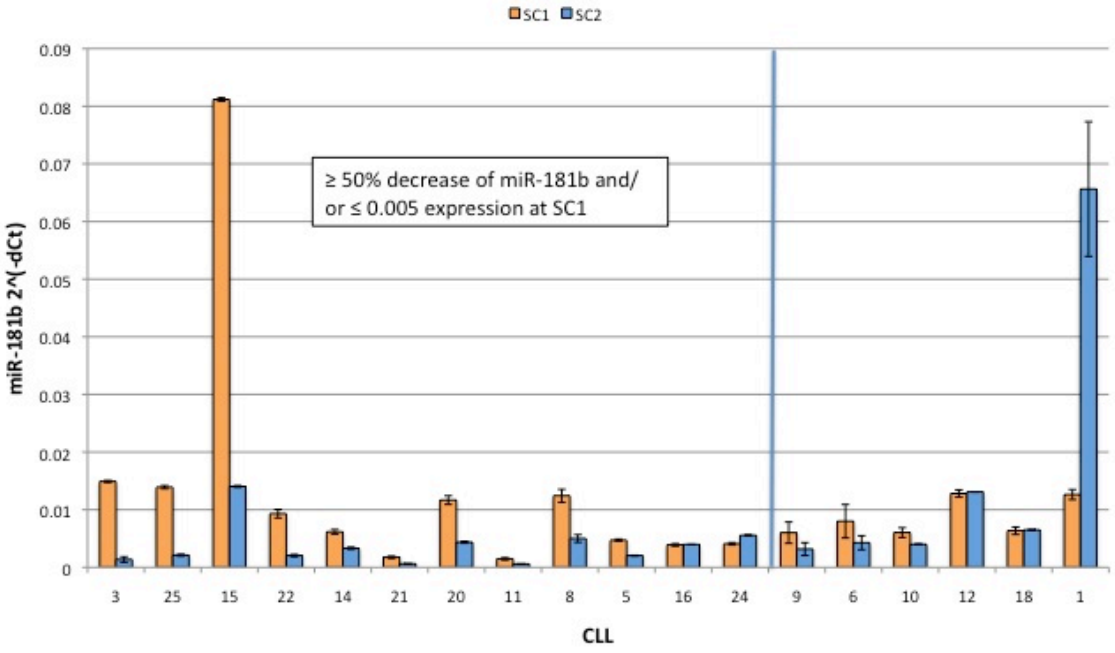


Figure S2: miR-181b relative expression over time in CLL. A cohort of 18 CLL patients were observed of their relative expression of miR-181b at two sequential time points before treatment (if given) via real-time qPCR. Orange bars and blue bars are expression levels of miR-181b at the first sample collection (SC1) and the last sample collection (SC2), respectively. Patients samples left of the blue line had at least a 50% decrease of miR-181b expression over time and/or had SC1 expression of less than or equal to 0.005. Right of this blue line are patient samples that have not undergone these characteristics. Relative expression values were calculated using the $2^{-deltaCt}$ method (Schmittgen & Livak, 2008).

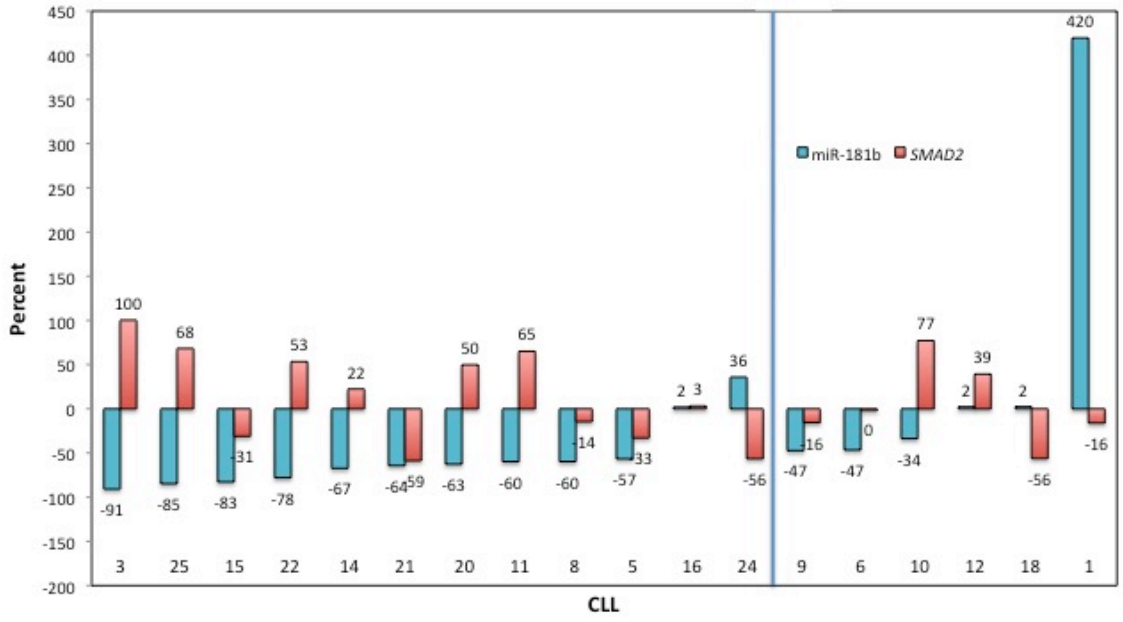


Figure S3: Percent change of miR-181b and *SMAD2* relative expression over time. Each patient (N = 18) was evaluated at two different sample collections (initial sample collection, SC1, and last sample collections, SC2) for their expression of miR-181b and SMAD2 via real-time qPCR. Percent change of the expression of each gene (blue represents miR-181b as red represents *SMAD2*) over time are graphically shown here whereas values above bars are the percent change. Like in Figure S2, blue line stands for the same reason described.

Table 1: Pro-onconet genes targeted by miR-181b. DIANA micro v3.0 was used for this search.

Targets of miR-181b	Binding sites	Precision	miTG score	Proven target
TIMP3	2	0.83	11.34	YES*
POU21	5	0.96	25.33	No
SMAD2	3	0.8	9.88	No

Precision value is the probability that target is correct.
miTG score is the final prediction score of the likely hood of being a real target. The higher the score the more likely.
* denotes that this validation was conducted from a published study (Wang et al., 2010)

Table 2: Summary data of patients with "properties". These sample collections have "properties", defined as a greater than or equal to a 50 percent decrease of miR-181b relative expression over time.

CLL ID	DX --> SC1 (years)	SC1-->SC2 (years)	SC1 --> TX (years)	SC2-->TX (years)	Disease Phase Transition	SC1 mir181b relative expression	SC2 miR-181b relative expression	miR-181b % change over time	SC1 *SMAD2* relative expression	SC2 *SMAD2* relative expression	*SMAD2* % change over time	TREATMENT STATUS
3*	4.5	2.2	3.1	0.9	E to L	0.015	0.001	-91	0.33	0.66	100	YES
11*	0.4	3.5	3.7	0.1	E to L	0.001	0.001	-60	0.34	0.57	65	YES
5	0.9	3.0	3.4	0.4	E to L	0.005	0.002	-57	0.33	0.22	-33	YES
24*	0.1	2.3	2.3	0.0	E to L	0.004	0.006	36	0.56	0.24	-56	YES
25*	2.1	1.3	2.0	0.6	I to L	0.014	0.002	-85	0.30	0.50	68	YES
22*	0.7	3.5	3.6	0.0	I to L	0.009	0.002	-78	0.08	0.13	53	YES
14	0.2	1.3	1.3	0.0	I to L	0.006	0.003	-67	0.33	0.40	22	YES
21*	1.1	1.9	2.1	0.2	I to L	0.002	0.001	-64	0.26	0.11	-59	YES
15	0.2	2.4	N/A	N/A	N/A	0.081	0.014	-83	0.42	0.29	-31	NO
20*	0.1	2.3	N/A	N/A	N/A	0.012	0.004	-63	0.08	0.11	50	YES
8	5.3	3.6	N/A	N/A	N/A	0.012	0.005	-60	0.46	0.40	-14	NO
16	0.4	2.4	N/A	N/A	N/A	0.004	0.004	2	0.08	0.08	3	NO
Average	**1.3**	**2.6**	**3.0**	**0.3**								

* Denotes CLL patients with 50% or more increase of *SMAD2* relative expression over time.
SC1 denotes sample collection 1 (first time point).
SC2 denotes sample collection 2 (last time point).
DX denotes time of diagnosis.
TX denotes time of treatment.
DX --> SC1 is the amount of years between DX and SC1.
SC1-->SC2 is amount of years between SC1 and SC2.
SC1-->TX is amount of years between SC1 and TX.
SC2-->TX is amount of years between SC2 and TX.
N/A stands for Not Applicable

Table 3: Summary data of patients without "properties".

CLL ID	DX --> SC1 (years)	SC1-->SC2 (years)	SC1 --> TX (years)	SC2-->TX (years)	Disease Phase Transition	SC1 mir181b Rel. Exp.	SC2 miR-181b Rel. Exp.	miR-181b % change ovt	SC1 *SMAD2* Rel. Exp.	SC2 *SMAD2* Rel. Exp.	*SMAD2* % change ovt	TREATMENT STATUS
10*	0.3	4.9	5.0	0.1	E to L	0.006	0.004	-34	0.21	0.37	77	YES
9	1.0	1.3	1.9	0.5	I to L	0.006	0.003	-47	0.40	0.34	-16	YES
12	0.0	1.1	1.2	0.1	I to L	0.013	0.013	2	0.18	0.25	39	YES
6	1.3	8.2	N/A	N/A	N/A	0.008	0.004	-47	0.35	0.35	0	NO
18	0.0	2.0	N/A	N/A	N/A	0.006	0.007	2	0.15	0.07	-56	NO
1	0.1	1.7	N/A	N/A	N/A	0.013	0.066	420	0.47	0.40	-16	NO
Average	**0.4**	**3.2**	**2.7**	**0.2**								

* Denotes CLL patients with 50% or more increase of *SMAD2* relative expression over time.
SC1 denotes sample collection 1 (first time point).
SC2 denotes sample collection 2 (last time point).
DX denotes time of diagnosis.
TX denotes time of treatment.
DX --> SC1 is the amount of years between DX and SC1.
SC1-->SC2 is amount of years between SC1 and SC2.
SC1-->TX is amount of years between SC1 and TX.
SC2-->TX is amount of years between SC2 and TX.
N/A stands for Not Applicable

References

Arvey, A., Larsson, E., Sander, C., Leslie, C., and Marks, D. "Target mRNA abundance dilutes microRNA and siRNA activity." *Molecular Systems Biology* 6.363 (2010): 363.

Baughn, L., Di Liberto, M., Niesvizky, R., Cho, H., Jayabalan, D., Lane, J., Liu, F., and Chen-Kiang, S. "CDK2 phosphorylation of Smad2 disrupts TGF-beta transcriptional regulation in resistant primary bone marrow myeloma cells." *The Journal of Immunology* 182.4 (2009): 1810-1817.

Binet, J., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F., Thomas, M., Tchernia, G., Jacquillat, C., Boivin, P., Lesty, C., Duault, M., Monconduit, M., Belabbes, S., and Gremy, F. "A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis." *Cancer* 48.1 (1981): 198-206.

Burger, J. "Nurture versus Nature: The Microenvironment in Chronic Lymphocytic Leukemia." *ASH Education Program Book* (2011): 96-103.

Calin, G., Cimmino A., Fabbri, M., Ferracin, M., Wohcik, S., Shimizu, M., Taccioli, C., Zanesi, N., Garzon, R., Aqeilan, R., Alder, H., Volinia, S., Rassenti, L., Liu, X., Liu, C., Kipps, T., Negrini, M., and Croce, C. "MiR-15a and miR-16-1 cluster functions in human leukemia ." *Proceedings of the National Academy of Sciences* 105 .13 (2008): 5166-5171.

Chen, Jingchun, and Bo Yuan. "Detecting functional modules in the yeast protein–protein interaction network ." *Bioinformatics* 22 .18 (2006): 2283-2290.

Chuang, H., Rassenti, L., Salcedo, M., Licon, K., Kohlmann, A., Haferlach, T., Foa, R., Ideker, T., and Kipps, T. "Subnetwork-Based Analysis of Chronic Lymphocytic Leukemia Identifies Pathways that Associate with Disease Progression." *Blood* (2012): Submitted for publication.

Cimmino, A., Calin, G., Fabbri, M., Iorio, M., Ferracin, M., Shimizu, M., Wojcik, S., Aqeilan, R., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C., Kipps, T., Negrini, M., Croce, M. "miR-15 and miR-16 induce apoptosis by targeting BCL2 ." *Proceedings of the National Academy of Sciences of the United States of America* 102 .39 (2005): 13944-13949.

Crespo, M., Bosch, F., Villamor, N., Bellosillo, B., Colomer, D., Rozman, M., Marce, S., Lopez-Guillermo, A., Campo, E., and Montserrat, E. "ZAP-70 Expression as a Surrogate for Immunoglobulin-Variable-Region Mutations in Chronic Lymphocytic Leukemia." *New England Journal of Medicine* 348.18 (2003): 1764-1775.

Damle, R., Wasil, T., Fais, F., Ghiotto, F., Valetto, A., Allen, S., Buchbinder, A., Budman, D., Dittmar, K., Kolitz, J., Lichtman, S., Schulman, P., Vinciguerra, V., Rai, K., Ferrarini, M., and Chiorazzi, N. "Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia ." *Blood* 94 .6 (1999): 1840-1847.

Danilov, A., Klein, A., Lee, H., Baez, D., and Huber, B. "Differential control of G0 programme in chronic lymphocytic leukaemia: a novel prognostic factor." *British Journal of Haematology* 128.4 (2005): 472-481.

De Yébenes, V., Belver, L., Pisano, D., Gonzalez, S., Villasante, A., Corce, C., He, L., and Ramiro, A. "miR-181b negatively regulates activation-induced cytidine deaminase in B cells." *The Journal of Experimental Medicine* 205.10 (2008): 2199-2206.

Derynck, R, R J Akhurst, and A Balmain. "TGF-beta signaling in tumor suppression and cancer progression." *Nature Genetics* 29.2 (2001): 117-129.

Dijke, P., Goumans, M., Itoh, F., and Itoh, S. "Regulation of cell proliferation by Smad proteins." *Journal of Cellular Physiology* 191.1 (2002): 1-16.

Douglas, R., Capocasale, R., Lamb, R., Nowell, P., and Moore, J. "Chronic Lymphocytic Leukemia B Cells Are Resistant to the Apoptotic Effects of Transforming Growth Factor-β." *Blood* 89 .3 (1997): 941-947.

Durig, J., Nuckel, H., Cremer, M., Fuhrer, A., Halfmeyer, K., Fandrey, J., Moroy, T., Klein-Hitpass, L., and Duhrsen, U. "ZAP-70 expression is a prognostic factor in chronic lymphocytic leukemia." *Leukemia* (2003): 2426-34.

Fabbri, M., Bottoni, A., Shimizu, M., Spizzo, R., Nicoloso, M., Rossi, S., Barbarotto, E., Cimmino, A., Adair, B., Wojcik, S., Valeri, N., Calore, F., Sampath, D., Fanini, F., Vannini, I., Musuraca, G., Dell'Aquila, M., Alder, H., Davuluri, R., Rassenti, L., Negrini, M., Nakamura, T., Amadori, D., Kay, N., Rai, K., Keating, M., Kipps, T., Calin, G., and Croce, C. "Association of a MicroRNA/TP53 Feedback Circuitry With Pathogenesis and Outcome of B-Cell Chronic Lymphocytic Leukemia ." *JAMA: The Journal of the American Medical Association* 305 .1 (2011): 59-67.

Gavin, A., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J., Michon, A., Cruciat, C., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M., Copley, R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. "Functional

organization of the yeast proteome by systematic analysis of protein complexes." *Nature* 415.6868 (2002): 141-147.

Hamblin, T. and Oscier, D. "Chronic lymphocytic leukaemia: the nature of the leukaemic cell." *Blood Reviews* 11.3 (1997): 119-128.

Hamblin, T., Orchard, J., Ibbotson, R., Davis, Z., Thomas, P., Stevenson, F., and Oscier, D. "CD38 expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease ." *Blood* 99 .3 (2002): 1023-1029.

Ideker, T., Thorrson, V., Ranish, J., Christmas, R., Buhler, J., Eng, J., Bumgarner, R., Goodlett, D., Aebersold, R., and Hood, L. "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network." *Science* 292.5518 (2001): 929-34.

Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A. "Discovering regulatory and signalling circuits in molecular interaction networks ." *Bioinformatics* 18 .suppl 1 (2002): S233-S240.

Itoh, S., Itoh, F., Goumans, M., and Dijke, P. "Signaling of transforming growth factor-β family members through Smad proteins." *European Journal of Biochemistry* 267.24 (2000): 6954-6967.

Klein, U., Tu, Y., Stolovitzky, G., Mattioli, M., Cattoretti, G., Husson, H., Freedman, A., Inghirami, G., Cro, L., Baldini, L., Neri, A., Califano, A., and Dalla-Favera, R. "Gene Expression Profiling of B Cell Chronic Lymphocytic Leukemia Reveals a Homogeneous Phenotype Related to Memory B Cells." *The Journal of Experimental Medicine* 194.11 (2001): 1625-1638.

Kuang, S., Tong, W., Yang, H., Lin, W., Lee, M., Fang, Z., Wei, Y., Jelinek, J., Issa, J., and Garcia-Manero, G. "Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia." *Leukemia official journal of the Leukemia Society of America Leukemia Research Fund UK* 22.8 (2008): 1529-1538.

Kufe, D., Pollock, R., Weichselbaum, R., Bast, R., Gansler, T., Holland, J., and Frei, E. "Clinical Staging and Other Prognostic Features." *Holland-Frei Cancer Medicine*. 6th ed. Ed. DW Kufe et al. BC Decker, 2003. Print.

Lagneaux, L., Delforge, A., Bron, D., Massy, M., Bernier, M., and Stryckmans, P. "Heterogenous response of B lymphocytes to transforming growth factor-beta in B-cell chronic lymphocytic leukaemia: correlation with the expression of TGF-beta receptors." *British Journal of Haematology* 97.3 (1997): 612-620.

Lagneaux, L., Delforge, A., Bernier, M., Stryckmans, P., and Bron, D. "TGF-β Activity and Expression of its Receptors in B-Cell Chronic Lymphocytic Leukemia." *Leukemia & Lymphoma* 31.1-2 (1998): 99-106.

Lai, Eric C. "Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation." *Nature Genetics* 30.4 (2002): 363-364.

Lebman, D. and Edmiston, J. "The role of TGF-beta in growth, differentiation, and maturation of B lymphocytes." *Microbes and infection Institut Pasteur* 1.15 (1999): 1297-1304.

Lu, D., Zhao, Y., Tawatao, R., Cottam, H., Sen, M., Leoni, L., Kipps, T., Corr, M., and Carson, D. "Activation of the Wnt signaling pathway in chronic lymphocytic leukemia." *Proceedings of the National Academy of Sciences of the United States of America* 101.9 (2004): 3118-3123.

Maragkakis, M., Reczko, M., Simossis, V., Alexiou, P., Papadopoulos, G., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Vergoulis, T., Koziris, N., Sellis, T., Tsanakas, P., and Hatzigeorgiou, A. "DIANA-microT web server: elucidating microRNA functions through target prediction ." *Nucleic Acids Research* 37 .suppl 2 (2009): W273-W276.

Marton, S., Garcia, M., Robello, C., Persson, H., Trajtenberg, F., Prtisch, O., Rovira, C., Naya, H., Dighiero, G., and Cayota, A. "Small RNAs analysis in CLL reveals a deregulation of miRNA expression and novel miRNA candidates of putative relevance in CLL pathogenesis." *Leukemia official journal of the Leukemia Society of America Leukemia Research Fund UK* 22.2 (2008): 330-338.

Morabito, F., Cutrona, G., Gentile, M., Matis, S., Todoerti, K., Colombo M., Sonaglio, C., Fabris, S., Reverberi, D., Megna, M., Spriano, M., Lucia, E., Rossi, E., Callea, V., Mazzone, C., Festini, G., Zupo, S., Molica, S., Neri, A., and Ferrarini, M. "Definition of progression risk based on combinations of cellular and molecular markers in patients with Binet stage A chronic lymphocytic leukaemia." *British Journal of Haematology* 146.1 (2009): 44-53.

Nana-Sinkam, S., and Croce, C. "MicroRNA in chronic lymphocytic leukemia: transitioning from laboratory-based investigation to clinical application." *Cancer Genetics and Cytogenetics* 203.2 (2010): 127-133.

Otake, Y., Soundararajan, S., Sengupta, T., Kio, E., Smith, J., Pineda-Roman, M., Stuart, R., Spicer, E., and Fernandes, D. "Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA." *Blood* 109.7 (2007): 3069-3075.

Pekarsky, Y., Santanam, U., Cimmino, A., Palamarchuk, A., Efanov, A., Maximov,V., Volinia, S., Alder, H., Liu, C., Rassenti, L., Calin, G., Hagan, J., Kipps, T., and Croce, M. "Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181." *Cancer Research* 66.24 (2006): 11590-11593.

Palacios, F., Moreno, P., Morande, P., Abreu, C., Correa, A., Porro, V., Landoni, A., Gabus, R., Giordano, M., Dighiero, G., Pritsch, O., and Oppezzo, P. "High expression of AID and active class switch recombination might account for a more aggressive disease in unmutated CLL patients: link with an activated microenvironment in CLL disease ." *Blood* 115 .22 (2010): 4488-4496.

Rai, K., Sawitsky, A., Cronkite E., Chanana, A., Levy, R., and Pasternack, B. "Clinical staging of chronic lymphocytic leukemia ." *Blood* 46 .2 (1975): 219-234.

Rassenti, L., Huynh, L., Toy, T., Chen, L., Keating, M., Gribben, J., Neuberg, D., Flinn, I., Rai, K., Byrd, J., Kay, N., Greaves, A., Weiss, A., and Kipps, T. "ZAP-70 Compared with Immunoglobulin Heavy-Chain Gene Mutation Status as a Predictor of Disease Progression in Chronic Lymphocytic Leukemia." *New England Journal of Medicine* 351.9 (2004): 893-901.

Rosenwald, A., Alizadeh, A., Widhopf, G., Simon, R., Davis, R., Yu, X., Yang, L., Pickeral, O., Rassenti, L., Powell, J., Botstein, D., Byrd, J., Grever, M., Cheson, B., Chiorazzi, N., Wilson, W., Kipps, T., Brown, P., and Staudt, M. "Relation of Gene Expression Phenotype to Immunoglobulin Mutation Genotype in B Cell Chronic Lymphocytic Leukemia ." *The Journal of Experimental Medicine* 194 .11 (2001): 1639-1648.

Schiemann, W., Rotzer, D., Pfeifer, W., Levi, E., Rai, K., Knaus, P., and Kadin, M. "Transforming growth factor-beta (TGF-beta)-resistant B cells from chronic lymphocytic leukemia patients contain recurrent mutations in the signal sequence of the type I TGF-beta receptor." *Cancer Detection and Prevention* 28.1 (2004): 57-64.

Schmittgen, T., and Livak, K. "Analyzing real-time PCR data by the comparative CT method." *Nature Protocols* 3.6 (2008): 1101-1108.

Sebestyén, A., Barna, G., Nagy, K., Janosi, J., Paku, S., Kohut, E., Berczi, L., Mihalik, R., and Kopper, L. "Smad signal and TGFbeta induced apoptosis in human lymphoma cells." *Cytokine* 30.5 (2005): 228-235. Print.

Tong, W., Wierda, W., Lin, E., Kuang, S., Bekele, B., Estrov, Z., Wei, Y., Yang, H., Keating, M., and Garcia-Manero, G. "Genome-wide DNA methylation profiling of chronic lymphocytic leukemia allows identification of epigenetically repressed molecular pathways with clinical impact." *Epigenetics* 5.6 (2010): 499-508.

Turner, Bryan M. "Defining an epigenetic code." *Nature Cell Biology* 9.1 (2007): 2-6.

Uetz, P., Glot, L., Cagney, G., Mansfield, T., Judson, R., Knight, J., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emill, Li, Y., Godwin, B., Conover, D., Kalbffeisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae." *Nature* 403.6770 (2000): 623-627.

Visone, R., Veronese, A., Rassenti, L., Balatti, V., Pearl, D., Acunzo, M., Stefano, V., Taccioli, Kipps, T., and Croce, C. "miR-181b is a biomarker of disease progression in chronic lymphocytic leukemia." *Blood* 118.11 (2011): 3072-3079.

Zhu, W., Shan, X., Wang, T., Shu, Y., and Liu, P. "miR-181b modulates multidrug resistance by targeting BCL2 in human cancer cell lines." *International Journal of Cancer* 127.11 (2010): 2520-2529.