UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOMEDICINA COMPARATA ED ALIMENTAZIONE

Corso di Laurea Magistrale in Biotecnologie per l'Alimentazione

UNIVERSITÀ DEGLI STUDI DI PADOVA

TESI DI LAUREA

RNA-seq data analysis of circular RNA expression relative to host-gene transcripts in Chronic Lymphocytic Leukemia with t(14;19)

Relatrice: **Prof.ssa Stefania Bortoluzzi**

Dipartimento di Medicina Molecolare

Correlatori:

Dott. Enrico Gaffo, Dipartimento di Medicina Molecolare
Prof.ssa Mery Giantin, Dipartimento di Biomedicina Comparata e Alimentazione

Laureanda: **Kimia Salek**

ANNO ACCADEMICO 2023/2024

Analisi mediante RNA-seq dell'espressione relativa di RNA circolari rispetto ai trascritti lineari del gene ospite nella Leucemia Linfocitica Cronica con t(14;19)

Riassunto

La leucemia più diffusa nei paesi Occidentali è la Leucemia Linfocitica Cronica (CLL), una malattia ematologica complessa ed eterogenea. Nonostante i miglioramenti nelle terapie, la CLL impatta sulla salute dei pazienti ed è ancora oggi una malattia sostanzialmente incurabile.

Gli RNA circolari (circRNA) sono emersi abbastanza recentemente quale nuova e interessante classe di RNA prevalentemente non codificanti ma con ruoli regolativi in svariati processi cellulari e nelle patologie, incluse quelle tumorali. Proprio la loro struttura circolare fa sì che questi RNA siano stabili e possano agire come regolatori dell'espressione genica nello sviluppo dei tumori. Evidenze recenti stanno mostrando che i circRNA svolgono la loro funzione mediante interazione con altri RNA oppure con proteine, mentre solo alcuni circRNA codificano peptidi funzionali. Gli RNA circolari sono abbondantemente espressi nel comparto ematopoietico, ma i ruoli della maggior parte di essi in diversi tipi e stadi di maturazione delle cellule del sangue, e anche nel contesto della CLL, sono ancora poco conosciuti.

Questa tesi mira all'esplorazione dettagliata dei profili di espressione differenziale degli RNA circolari nella CLL, ed in un sottotipo particolarmente aggressivo di questa patologia, recentemente associato alla traslocazione t(14;19)(q32;q13) causante sovraespressione del fattore di trascrizione BCL3, in confronto con la controparte normale rappresentata delle cellule B di donatori sani. Oltre a considerare l'espressione assoluta degli RNA circolari, le analisi si focalizzano particolarmente su quella relativa, ovvero sui cambiamenti del livello di espressione degli RNA circolari rispetto alla controparte lineare prodotta dallo stesso gene. Significativi cambiamenti nella "proporzione circolare su lineare" (circular to linear proportion, CLP) permettono di descrivere diversi stati cellulari. Inoltre, in questa tesi è stato preso in considerazione un tipo particolare di circRNA identificato solo molto recentemente: gli RNA circolari codificati dal genoma mitocondriale, andando a fornire una visione più ricca e multidimensionale delle alterazioni del trascrittoma della CLL.

Per raggiungere questi obiettivi, è stato usato il software CirComPara2 per analizzare i dati RNA-seq di una coorte di 47 pazienti con CLL, 22 con la traslocazione che coinvolge BCL3, 22 senza, e 9 di cellule B di controlli sani. Abbiamo identificato specifici circRNA con un CLP alterato nella CLL in generale oppure in associazione con la t(14;19)(q32;q13) e il cui ruolo in questa patologia, e la rilevanza come marcatori diagnostici e prognostici, dovrà essere studiato in futuro.

Studi esplorativi svolti in questa tesi indicano che alcuni dei circRNA identificati potrebbero agire legandosi a microRNA impedendo la funzione soppressoria di questi sull'espressione genica. Ulteriori indagini potranno essere condotte per indagare invece possibili interazioni tra circRNA e RNA messaggeri e il ruolo di queste nella CLL, con l'obiettivo di comprendere meglio gli aspetti molecolari di questa patologia nell'ottica di facilitare lo sviluppo di migliori e personalizzate terapie in futuro.

Abstract

Chronic lymphocytic leukemia (CLL) is a prevalent and complex hematological malignancy with a significant impact on patient health and despite numerous advancements in treatment modalities, it remains an incurable pathological condition. Circular RNAs (circRNAs) have recently emerged as a fascinating class of non-coding RNAs that possess diverse regulatory functions in various cellular processes and diseases including cancer. Accumulating evidence suggests that circRNAs can exert their biological functions by interacting with other RNA molecules and proteins, and even through their translation into functional peptides. Given their unique circular structure, circRNAs may play critical roles in the dysregulation of gene expression and the development of cancer. While circRNAs have been abundantly detected in the hematopoietic compartment, their specific roles and diversity among different blood cell types, particularly in the context of CLL, remain poorly understood.

This thesis aimed to comprehensively explore the differential expression patterns of circRNAs comparing Chronic lymphocytic leukemia (CLL), CLL cases with aggressive disease-bearing t(14;19)(q32;q13) rearrangement, and B-cells from healthy donors. However, instead of focusing solely on how the absolute read counts change in each condition, our emphasis is on understanding how the abundance of the circular form changes relative to the linear form. Concentrating on significant changes in circular to linear proportion (CLP) is crucial for unraveling the dynamic shifts across different states. Furthermore, an additional aspect involves investigating the presence and abundance of Mitochondrion-encoded circular RNA in these conditions. This exploration into Mitochondrion-encoded circular RNA adds a complementary dimension to the study, potentially revealing insights into their potential roles alongside the broader landscape of circRNAs in CLL pathogenesis.

To accomplish the objective, we utilized the CirComPara2 methodology to analyze RNA-sequencing data sourced from a cohort of 47 CLL patients, 25 CLL exhibiting and 22 without the BCL3 translocation, and a control group comprising 9 healthy donors. We identified specific circRNAs that show a meaningful difference in circular to linear proportions among conditions and could play a role in the processes associated with CLL pathogenesis and influenced by the t(14;19)(q32;q13) rearrangement, which can be further explored as diagnostic or prognostic biomarkers in the future.

Our findings indicate that certain circRNAs may contribute to leukemogenesis by interacting with microRNAs. A deeper understanding of the intricate relationships among circRNAs, mRNAs, miRNAs, and CLL pathophysiology holds promise for unveiling novel therapeutic strategies, potentially leading to personalized medicine approaches.

Table of Contents

List of Tables

List of Figures

Introduction

1. Chronic Lymphocytic Leukemia

1.1 Definition and Pathophysiology

Chronic lymphocytic leukemia (CLL) is a monoclonal lymphoproliferative disease, resulting in the proliferation and accumulation of morphologically mature B-cell lymphocytes with impaired immune function, commonly known as smudge cells. The primary disease sites affected by CLL include peripheral blood, spleen, lymph nodes, and bone marrow [1].

According to current knowledge, the pathophysiology of CLL involves a two-step process that results in the clonal replication of malignant B lymphocytes. Initially, various factors, such as antigenic stimulation, genetic mutations, and cytogenetic abnormalities, lead to the development of Monoclonal B-cell Lymphocytosis (MBL) cells. Subsequently, MBL progresses to CLL by encountering additional genetic abnormalities or changes in the bone marrow microenvironment. A pivotal step in CLL's pathogenesis is the expression of B-cell antigen receptors (BCRs), which triggers antigen-independent cell-autonomous signaling [2,3].

In CLL, CD5+ B cells are continually activated due to mutations leading to MBL. As genetic abnormalities accumulate in more mature B cells, the neoplastic B-cell undergoes clonal division primarily within the lymph nodes. Lymphadenopathy, detected in 50 to 90% of patients, is the most common physical examination finding in CLL, typically affecting lymph nodes in the cervical, supraclavicular, and axillary regions. As the number of B-cell lymphocytes increases, they eventually spill into the peripheral blood, leading to lymphocytosis observed on a Complete Blood Count [4–6].

These malignant B cells evade apoptosis and continue to divide within the lymph nodes over time. They eventually infiltrate the spleen and bone marrow, causing splenomegaly and hypercellular bone marrow on biopsy. This enlargement of the spleen results in increased sequestration of red blood cells (RBCs) and platelets, leading to anemia and thrombocytopenia. CLL patients are also more susceptible to autoimmune hemolytic anemia and autoimmune thrombocytopenia [7,8]. As the neoplastic B cells spread throughout the body, systemic symptoms like fever, night sweats, unintentional weight loss, fatigue, and early satiety may occur. The reduced functional B cells also lead to hypogammaglobulinemia, increasing the risk of infections [9,10].

1.2 Diagnosis

The initial step in diagnosing CLL involves conducting a peripheral blood smear, which reveals an absolute lymphocyte count exceeding 5000/mcL and the presence of characteristic "smudge cells," confirming the CLL diagnosis. Although the diagnostic criteria specify an absolute lymphocyte count of ≥5000/mcL B lymphocytes on the peripheral smear, a considerable number of patients exhibit an absolute lymphocyte count > 100,000/mcL [4,11].

Immunophenotypic analysis using peripheral blood flow cytometry aids in verifying the clonality of circulating B cells in CLL patients. This analysis can be performed on both peripheral blood and bone marrow aspirate, examining classical immunophenotypic markers of CLL. Characteristic features include low levels of immunoglobulins (mainly IgM immunoglobulin, and occasionally both IgM and IgD), expression of B-cell associated antigens (CD19, CD20, CD21, CD23, and CD24), and expression of CD5, a T-cell associated antigen. The most common immunophenotype expression in CLL is the coexpression of CD5, CD19, and CD23. However, some cases may show varying levels of expression of other immunophenotypic antigens. Measurement of serum immunoglobulins and free-light chains is conducted at baseline to assess immunodeficiency and during treatment to monitor

immune reconstitution, especially with newer generation B-cell receptor signaling drugs. Most importantly, only one type of light chain is observed, indicating the monoclonality of the lymphocytes, although rare cases may express both kappa and lambda light chains, referred to as "biclonal CLL" [4,12–17].

Fluorescent in situ hybridization (FISH) is a highly sensitive test employed to detect chromosomal abnormalities in CLL patients. While not necessary for diagnosis, bone marrow aspiration and biopsy are frequently performed as part of the diagnostic workup or prior to treatment. A bone marrow biopsy confirming greater than 30% lymphocytes of all nucleated cells in a normocellular/hypercellular aspirate confirms CLL diagnosis. Reduction of lymphocytic infiltration to less than 30% during treatment indicates a complete response [18–20].

Complications of CLL, such as autoimmune hemolytic anemia, are diagnosed based on a positive direct antiglobulin (Coombs) test, elevated reticulocyte count, increased serum lactate dehydrogenase (LDH), reduced haptoglobin, and elevated serum indirect bilirubin. Additional complications, such as pure red cell aplasia and thrombocytopenia, can be diagnosed through peripheral blood smear examination and bone marrow aspiration and biopsy. CLL may also present with hypogammaglobulinemia (observed in less than 15% of cases), elevated uric acid levels, and elevated hepatic enzymes. Lab studies, such as serum LDH and beta-2 microglobulin, are important for assessing disease activity, and creatinine levels are considered while interpreting beta-2 microglobulin levels [21].

1.3 Prognosis and Disease progression

The survival period in CLL patients varies from 2 to > 20 years, with a median survival of 10 years. Patients presenting with Rai stage 0-II may survive for 5 to 20 years without requiring treatment. The lymphocyte doubling time, which is the time it takes for the absolute lymphocyte count to double, is a prognostic factor. Patients with a lymphocyte doubling time of < 12 months typically have a more aggressive form of CLL. Certain factors, such as mutated Ig heavy chain variable region (potentially suggestive of an origin from more mature B-cells), 13q deletion, low ZAP-70 expression, and low CD38 levels, are associated with a favorable prognosis [22]. On the other hand, high-risk cytogenetic abnormalities like 17p deletion and 11q deletion are considered unfavorable prognostic factors. Patients with multiple-chain lymphadenopathy, hepatosplenomegaly, anemia, and thrombocytopenia generally have a worse prognosis. The International Prognostic Index for Chronic Lymphocytic Leukemia (CLL-IPI) is a commonly used tool to predict CLL-related outcomes [23].

CLL can have various consequences, including increased susceptibility to infections, particularly respiratory tract infections. There is also a risk of progression to diffuse large B-cell lymphoma, known as Richter syndrome. CLL patients also have an elevated risk of developing other malignancies, such as skin, lung, and gastrointestinal tract cancers. Additionally, some individuals may experience immune system issues, where the immune system attacks red blood cells or platelets, although this is rare [24].

1.4 Treatment and Management

1.4.1 Treatment Approaches

The treatment of CLL is based on the individual characteristics of the disease and the patient's overall health. Not all patients diagnosed with CLL require immediate treatment, as the disease is heterogeneous. Some patients may have survival rates similar to those of the general population and

may not need treatment, except in cases of allogeneic hematopoietic stem cell transplantation (HCT), which is the only curative option.

The treatment of CLL is divided into two categories: patients without "active disease" and patients with "active disease", For patients without "active disease," including early-stage asymptomatic CLL, the standard of care is observation. Regular clinical examinations and blood counts are performed, and a decision is made on whether aggressive treatment is required or if observation can continue. There are specific criteria, such as the International Prognostic Score for Early-stage CLL (IPS-E), which further divides early-stage CLL patients based on risk factors to determine the need for treatment at one and five years [25–27].

For patients with "active disease" the main goal of treatment is symptomatic relief, prolonged remission, and improved survival. However, CLL remains an incurable disease, and treatment choice is based on various factors such as genetic risk stratification, patient fitness, tumor characteristics, and therapy goals. The available treatment options include chemotherapy, immunotherapeutic agents like monoclonal antibodies (rituximab, ofatumumab, obinutuzumab) and CAR T-cell Therapy, targeted therapies including Bruton tyrosine kinase inhibitors (ibrutinib, acalabrutinib), BCL-2 inhibitor (venetoclax), purine analogs (fludarabine, pentostatin), and alkylating agents (cyclophosphamide, chlorambucil, bendamustine) [28].

For patients with relapsed/refractory CLL, treatment choices depend on the initial treatment received and the duration of response. Targeted therapies like ibrutinib, idelalisib, alemtuzumab, and venetoclax are commonly used, and allogeneic hematopoietic stem cell transplantation may be considered in specific cases. Moreover, emerging treatment options for CLL include novel chemotherapy-free triplet combinations, bispecific antibody-based therapies, and cell-based therapies [29].

1.5 Genomic Characteristics of CLL

1.5.1 Recurrent Genetic Lesions

The advent of genome-wide sequencing and copy-number analysis has ushered in a comprehensive understanding of the genomic landscape of CLL. Through whole-exome sequencing (WES) of over 1,000 CLL samples and whole-genome sequencing (WGS) of approximately 200 CLL patients, researchers have illuminated the genomic realm of CLL, revealing an average of around 0.9 mutations per megabase encompassing point mutations, copy number alterations, and rare chromosomal translocations. Each patient typically harbors 10–30 non-silent events, with Mutated Immunoglobulin Heavy Chain Variable Region (IGHV-M) CLL exhibiting a higher count (~2,800) than Unmutated IGHV(IGHV-UM) CLL (~2,000) [30–39].

Deletions of 13q14 (del13q14) emerged as the most frequent anomaly, affecting 50-60% of CLL patients, and are intricately linked to essential genetic elements governing CLL initiation and progression [35,40]. NOTCH1 mutations, found in 4-20% of cases [30,32–35], sustain continuous signaling and are associated with advanced disease stages and poorer survival [30,34,41,42]. Mutations in RNA processing genes like SF3B1, convergence on NF-κB activation, and the presence of deletions (del17p) encompassing TP53 underscore the intricate genetic landscape of CLL [33,43,44]. Additionally, deletions in the ATM gene (del11q) further contribute to disease complexity, particularly impacting the non-canonical NF-κB pathway [45]. Trisomy 12, another prevalent cytogenetic abnormality, brings atypical lymphocyte morphology into focus and assumes a role of importance, especially when co-occurring with NOTCH1 mutations, as it is linked to poor outcomes [46]. Furthermore, POT1 mutations disrupt telomere protection, contributing to chromosomal breaks and tumor progression, particularly in advanced disease stages and specific genetic contexts [47]. These multifaceted genetic

anomalies collectively shape the intricate landscape of CLL, influencing its clinical manifestations and prognosis.

1.5.2 Complex Karyotype

Recent advancements have highlighted the potential importance of complex karyotype (CK) in CLL. A CK is defined when three or more chromosomal aberrations, including structural and numerical anomalies, are assessed through chromosome-banding analysis (CBA).

A retrospective study involving 5290 patients with CBA data aimed to understand the clinical and biological associations of CK in CLL. The study found that patients with five or more abnormalities (high-CK) consistently had unfavorable clinical outcomes, regardless of disease stage, TP53 aberrations (deletions and mutations), and immunoglobulin gene mutation status. In contrast, cases with 3 or 4 aberrations (low-CK and intermediate-CK) showed aggressive disease progression, mainly with TP53 aberrations, whereas those with specific chromosomal changes, such as trisomy 12 (+12) and 19 (+19), alongside CK exhibited an indolent disease profile [48].

Baliakas et al. (2019) proposed a hierarchical model categorizing patients based on CK, TP53 aberrations, and immunoglobulin gene mutation status. High-CK cases had the worst prognosis (Figure 1), while mutated CLL without CK or TP53 aberrations, and CK with +12,+19 alterations, showed the longest survival. This model highlights the diverse clinical behaviors within CK and underscores that high-CK cases are significant adverse prognostic indicators, independently of other biomarkers (Figure 1). However, the authors stressed the need for prospective clinical validation before integrating high-CK into CLL risk stratification.

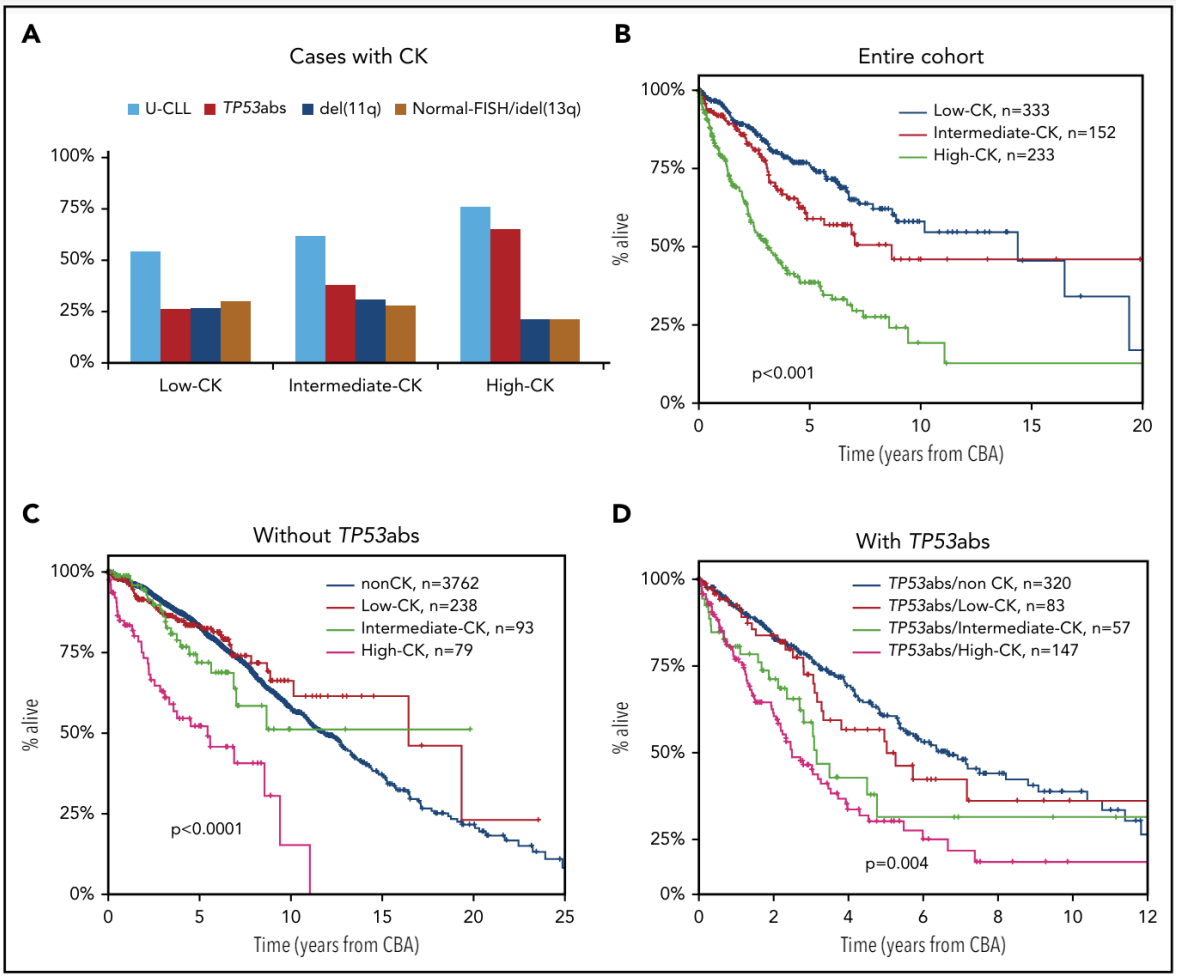

Figure 1: a hierarchical model categorizing patients based on CK, TP53 aberrations, and immunoglobulin gene mutation status. Baliakas et al. (2019) investigated patients with chronic lymphocytic leukemia (CLL) who had varying numbers of chromosomal abnormalities (CK) in their genetic profiles. The patients were divided into groups based on the number of aberrations: low-CK (3 aberrations), intermediate-CK (4 aberrations), and high-CK (≥5 aberrations). (A) The frequency of specific genetic markers was analyzed: U-CLL (unmutated IGHV genes), TP53abs (deletion of chromosome 17p and/or TP53 mutations), del(11q) (deletion of chromosome 11q), and normal-FISH/idel(13q) (normal FISH or isolated deletion of chromosome 13q based on Döhner hierarchical model). High-CK patients had higher occurrences of U-CLL and TP53abs compared to low-CK and intermediate-CK patients. However, patients with normal-FISH/idel(13q) were present in all CK groups. (B-D) Kaplan-Meier survival curves were used to assess overall survival (OS) for different patient groups. (B) For the entire cohort of CLL patients with CK, low-CK, intermediate-CK, and high-CK cases were represented by blue, red, and green lines, respectively. (C) Among patients without TP53abs, high-CK patients had the shortest OS (purple line), while there was no significant difference between low-CK (red line), intermediate-CK (green line), and the non-CK CLL group (blue line). (D) In patients with TP53abs, the number of aberrations worsened the clinical outcome, with high-CK patients showing the shortest OS (purple line) [48].

1.5.3 The t(14;19)(q32;q13) rearrangement marks an aggressive subtype of CLL

The t(14;19)(q32;q13) is an uncommon genetic abnormality found in less than 0.1% of B-cell neoplasms. Most cases of this abnormality are classified as "atypical" CLL. Despite similarities in clinical presentation with CLL, patients with this abnormality are usually younger, predominantly male, and tend to have a more aggressive disease course compared with the remaining CLL cases [49]. Conversely, the frequency and prognostic relevance of BCL3 translocation in typical CLL cases with a usual phenotype have not been extensively characterized. One study on 225 CLL patients with a

typical phenotype (phenotypic score>3) revealed that BCL3 translocation occurred in 1.8% of cases and was exclusively associated with the presence of +12 as the sole cytogenetic abnormality in the considered cohort [50–52].

The appearance of blood and bone marrow cells is "atypical", but these differences are not evident in bone marrow or lymph node tissue samples [49]. The immune profile shows positive CD5 and typically dim/negative CD23 expression, along with a low Matutes CLL score [53,54]. Confirmation of the translocation resulting in IGH/BCL3 fusion is achieved through fluorescence in situ hybridization because cytogenetic analysis may overlook the subtle reciprocal translocation between chromosomes 14 and 19 (Figure 2 and Figure 3), necessitating a more precise diagnostic approach [55].

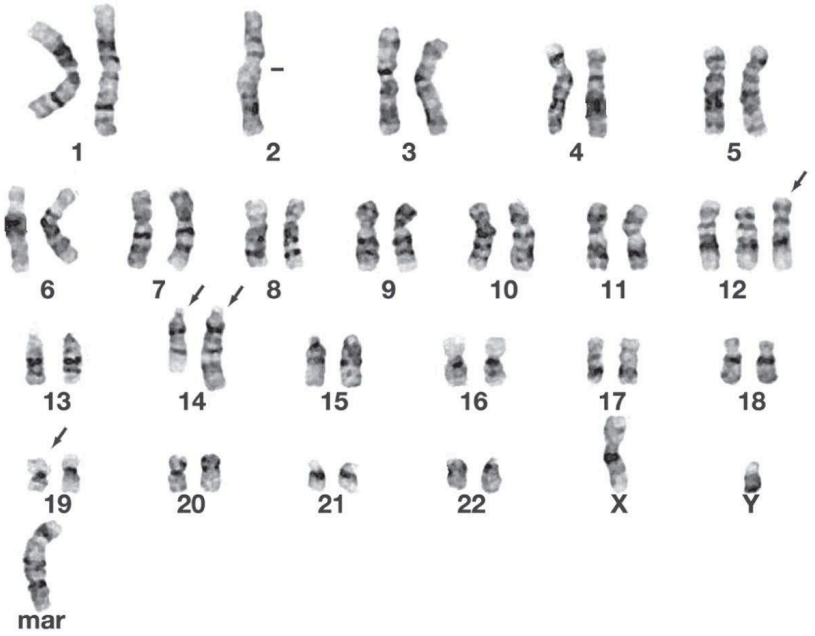

Figure 2 : BCL3 translocation in atypical CLL patient with trisomy 12.

BCL3 translocations in atypical CLL have been linked to the presence of +12 (Figure 2), unmutated IGHV genes, and preferential use of the IGHV4-39/IGHD6-13/IGHJ5 rearrangement. Clinically, these translocations in atypical CLL indicates an unfavorable outlook and poses an elevated risk of transforming into aggressive lymphoma [50,51,53]. Considering these findings, it has been proposed that B-cell lymphoproliferative disorders harboring the t(14;19)(q32.3;q13.2) translocation could constitute a distinct entity separate from CLL. This provisional designation is referred to as "t(14;19)-positive small B-CLL leukemia" [53].

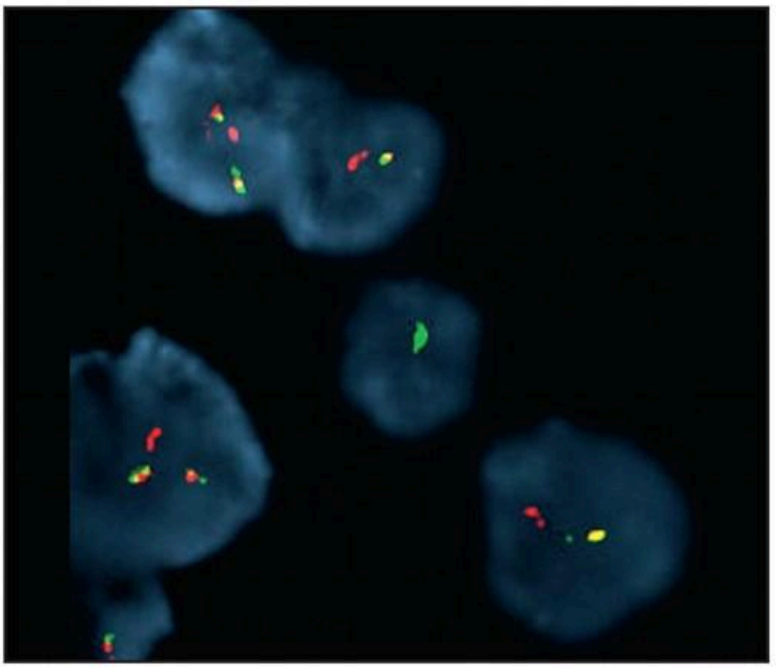

Figure 3: Fluorescence in situ hybridization (FISH) examination reveals the presence of the IGH-BCL3 fusion gene. This gene fusion is identified using specific fluorescent labels: the IGH region is marked with SpectrumGreen, while the BCL3 region is labeled with SpectrumOrange. The occurrence of the IGH-BCL3 fusion gene is indicated by a distinct yellow fusion signal, resulting from the overlap of green and orange signals, signifying the proximity and fusion of the IGH and BCL3 regions [54].

The t(14;19)(q32;q13) translocation involves the juxtaposition of the BCL3 gene located on chromosome 19q13 and the immunoglobulin heavy chain gene locus located on chromosome 14q32 (Figure 3), leading to the upregulation of BCL3 expression [53,54]. BCL3, primarily found in the nucleus, acts as an inhibitor of kappa B (IκB), interacting with nuclear factor kappa B (NFκB) p50 and p52 subunits to modify their DNA binding and gene expression [56–59]. BCL3's role includes blocking apoptosis induced by interleukin-4 deprivation and potentially inhibiting p53 activation [60,61]. Interestingly, BCL3 overexpression, as detected through immunohistochemistry, is not exclusive to tumors with the t(14;19)(q32;q13) translocation. It has been observed in cases of CLL lacking this translocation, as well as various B- and T-cell lymphomas and classical Hodgkin lymphoma. Despite its association with the t(14;19)(q32;q13) translocation, BCL3's functions and expression patterns extend to a broader spectrum of lymphoid malignancies [62–64].

Based on recent clinical data, as provided by Dr. Andrea Visentin and referring to a large patient cohort, individuals with t(14;19)-CLL exhibit several distinct characteristics in comparison to typical CLL patients [65,66]. Notably, they are diagnosed at a significantly younger age, with a median age of 60.5 years as opposed to 70.5 years for the typical CLL group. Additionally, these patients experience a shorter overall survival, averaging 7.4 years, and often require earlier treatment, as indicated by an average treatment-free survival of 2.0 years. This aggressive subtype remains poorly understood from a molecular perspective, with limited representation in the current literature, as evidenced by the paucity of relevant research papers. Moreover, the transcriptome characteristics of t(14;19)-CLL have received limited attention, with only a few cases studied, and the available data are fragmented, primarily focusing on linear RNA analysis.

2. The Emerging Role of Circular RNAs in Leukemias

2.1 Biogenesis and Regulation of circRNAs

As gene sequencing technology and transcriptome studies have progressed, it has become evident that a substantial portion of the human genome, ranging from 70% to 90%, is transcribed into RNAs. Surprisingly, only 2% of these RNAs are responsible for encoding proteins. This revelation underscores the potential significance of non-coding RNAs (ncRNAs) in driving essential biological processes. In recent years, the advent of advanced high-throughput sequencing techniques and transcriptome analysis has facilitated the discovery of circular RNAs (circRNAs), a class of transcripts that can participate in various biological mechanisms and influence the development of diseases [67,68].

CircRNAs have emerged as promising biomarker candidates, direct contributors of disease mechanisms, and therapeutic targets of different tumor types. Compelling evidence has demonstrated their involvement in hematological malignancies, notably leukemia. CircRNAs have shown aberrant features in cancer, exerting significant impacts on disease initiation, progression, metastasis, and patient prognosis. Therefore, circRNAs have garnered significant attention as a focal point for research and exploration within the context of hematological cancers.

Canonical splicing in eukaryotic cells is a fundamental process orchestrated by the spliceosome, responsible for removing introns and connecting exons to create a linear RNA transcript with 5′–3′ polarity [69]. Unlike this canonical process, most circRNAs are generated through backsplicing, a mechanism involving either the spliceosome or ribozymes like the group I and II ribozymes [70]. CircRNAs exhibit distinctive features due to their closed covalent bonds, lacking typical terminal structures like the 5′ cap or a polyadenylated tail [71].

Experimental evidence points to the involvement of the spliceosome in circRNA biogenesis. Inhibition of canonical splicing by a pre-mRNA splicing inhibitor, isoginkgetin, leads to decreased levels of circRNAs and spliced linear transcripts [72]. Interestingly, circRNA expression does not consistently align with the levels of their linear counterparts, suggesting a regulated circRNA expression, with the spliceosome differentiating between forward (canonical) and backward (backsplicing) splicing [73].

CircRNAs can stem from exons or introns (Figure 4), resulting in three types: exonic, intronic, and exon-intron circRNAs [74]. Exonic circRNAs form via pre-mRNA splicing when a 3′ splice donor attaches to a 5′ splice acceptor, either from a single exon or by linking the start of an upstream exon with the end of a downstream exon, creating multiple-exon circRNAs. Exon-intron circRNAs are produced if the intron between exons is retained [75]. Intronic circRNAs form from resistant intron lariats, featuring a unique 2′–5′ linkage, distinguished from exonic circRNAs [76]. Intronic circRNAs depend on specific GU-rich and C-rich sequences near the 5′ splice site and branch point, respectively. During backsplicing, the binding of two segments forms a circle, leading to the removal of exonic and intronic sequences, and leaving behind intronic circRNA [77].

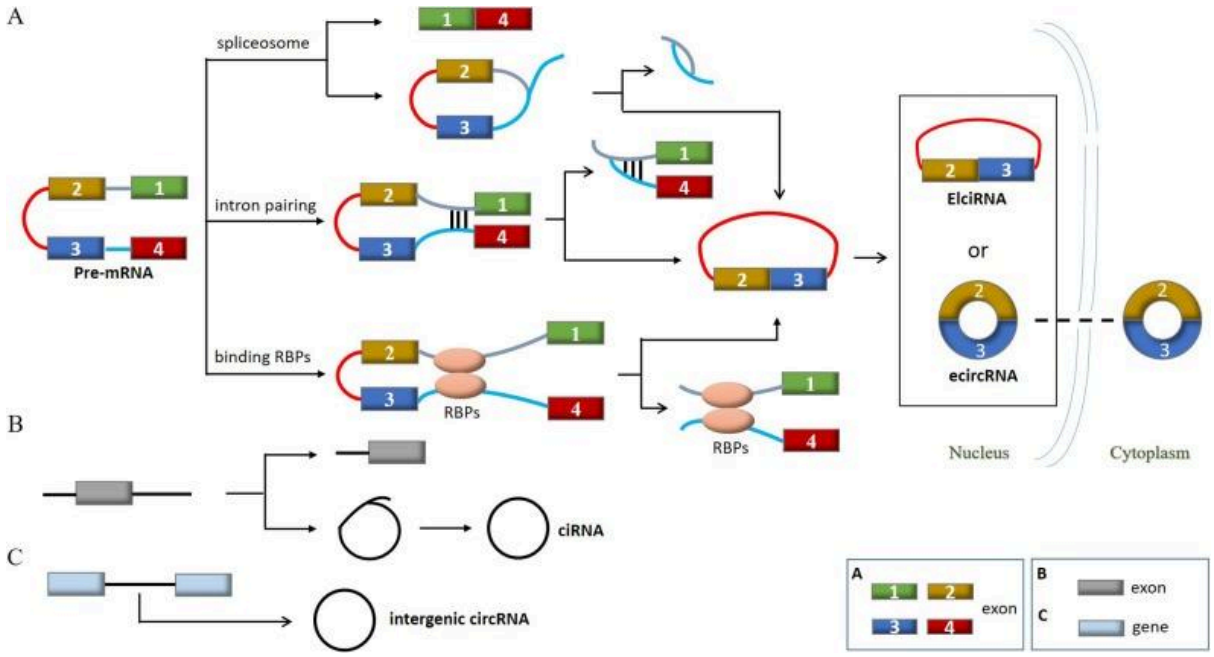

Figure 4: (A) The formation process of ecircRNAs and ElciRNAs. (B) The generation mechanism of ciRNAs. (C) The production process of intergenic circRNAs. circRNAs: Circular RNAs; ecircRNAs: Circular RNAs derived from exons; ElciRNAs: Circular RNAs spanning both exons and introns; ciRNAs: Circular intronic RNAs; RBPs: RNA-binding proteins [78].

In recent studies, the role of RNA binding proteins (RBPs) in regulating the formation of circRNAs has come to light, with RBPs acting as either enhancers or inhibitors under specific conditions [79]. For instance, in the fly brain, RBPs that govern the levels of muscle-blind protein (MBL) have been found to impact circRNA creation. Elevated RBP levels prompt binding to pre-mRNA, inducing splicing into circRNA instead of linear splicing, thus preventing translation into MBL protein [80,81]. This regulatory effect is further seen with the Quaking (QKI) protein and the adenosine deaminases acting on RNA (ADAR) protein, both being RNA binding proteins. QKI, linked to various diseases, including cancer, promotes circRNA biogenesis by binding to adjacent introns, enabling dimerization to form a looped structure that supports circularization. In contrast, ADAR, a regulatory RBP responsible for RNA editing and crucial for mammalian development, weakens and edits RNA duplexes, thus reducing the likelihood of circRNA circularization [82].

2.2 Biological Functions of circRNAs

2.2.1 MicroRNAs and Protein Sponges

Researchers are currently focused on understanding how circRNAs and miRNAs interact and regulate gene expression (Figure 5). CircRNAs contain sites where miRNAs can bind, acting as sponges to absorb and regulate their activity. This sponging effect is achieved through miRNA response elements (MREs) in circRNAs. CircRNAs can have multiple binding sites for miRNAs and play a significant role in controlling miRNA function. This interaction works as a competitive binding mechanism, indirectly influencing the expression of downstream target genes. When circRNAs are overexpressed, the corresponding target gene's expression increases, and inhibition of circRNAs leads to decreased target gene expression. This reveals the vital role circRNAs play in fine-tuning gene expression by acting as miRNA sponges [83].

CircRNAs' sequence or secondary structures can also be recognised and bound by proteins, making circRNAs able to serve as protein sponges or to positively regulate protein activity, for instance,

favoring protein interactions. CircRNA-protein interactions have been shown to impact protein expression, biogenesis, and pathophysiological progress [84,85]. Most circRNAs are derived from genes that encode proteins, and their frames can include exon sequences. CircRNAs binding to their target proteins will cause the functional suppression of these proteins [86].

2.2.2 Protein Translation

Some circRNAs can initiate protein translation, contrary to the traditional understanding that RNAs with circular structures cannot be translated. CircRNAs with internal ribosome entry site (IRES) elements have been found to generate proteins.

Although circRNAs are typically categorized as non-coding RNAs (ncRNAs), emerging research has indicated their potential for protein translation, particularly in cases where an internal ribosome entry site (IRES) is present [87–89]. Unlike traditional translation initiation, which depends on the 5′ cap structure and 3′ poly(A) tail recognition, IRES offers an alternative mechanism for translation initiation in eukaryotes [90]. Additionally, Adenosine methylation can enhance circRNA translation [91].

2.2.3 Direct Effects on Gene Expression and Translation

CircRNAs can have a direct impact on gene expression. They can interact with RNA polymerase II at gene promoters, affecting transcription.

Moreover, circularization and non-linear splicing of pre-mRNAs can hinder translation and lead to reduced production of proteins, termed the "mRNA trap" [92,93].

2.2.4 Regulation of mRNA Stability

Some circRNAs influence the stability of mRNA molecules. They can form double-stranded structures with target mRNAs, increasing their stability. This stabilization impacts gene expression, particularly in inflammatory signaling pathways [94,95]. Of interest, the regulation of oncogenic protein translation by circRNA can have a significant impact on cancer, as exemplified by the recently described role of circZNF609 in sarcomas, which affects tumor growth and drug response [96].

2.2.5 DNA Methylation Regulation

Certain circRNAs participate in gene regulation by affecting the DNA methylation of target genes. CircRNA ACR, for instance, prevents DNA Methyltransferase 3 Beta (Dnmt3B)-driven DNA methylation of the promoter region, activating gene expression [97].

2.2.6 Retrotransposon-Like Function

CircRNAs have been associated with retrotransposon-like functions, creating pseudogenes and potentially altering genome structure and gene expression regulation. They also play roles in various cellular processes, including RNA processing, genome rearrangement, and chromosome modification [98–100].

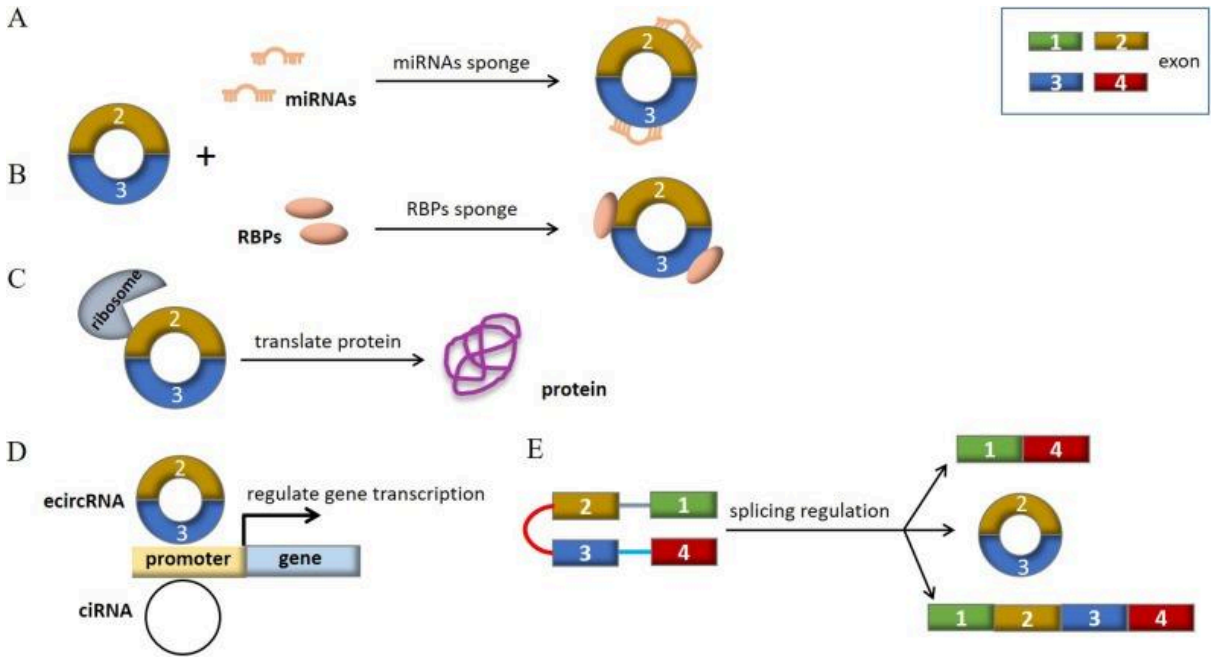

Figure 5: Main functions of circRNAs. (A) miRNAs sponge. (B) RBPs sponge. (C) Translate protein. (D) Regulate gene transcription. (E) Splicing regulation. circRNAs: circular RNAs; ecircRNAs: exonic circRNAs; ElciRNAs: exon-intron circRNAs; ciRNAs: circular intronic RNAs; RBPs: RNA-binding proteins [78].

2.3 Mitochondrion-Encoded Circular RNAs

2.3.1 Discovery

Cellular metabolic reprogramming, a recognized hallmark of cancer, holds promising avenues for therapeutic intervention. Mitochondria have gained increasing attention as a vital cellular compartment fueling the metabolic requirements of cancer cells. These organelles, as the primary source of ATP and essential metabolites, play a central role in fulfilling both bioenergetic and biosynthetic needs crucial for cancer cell survival and proliferation. Moreover, mitochondria are principal generators of reactive oxygen species (ROS), orchestrating cellular demise. In the context of CLL, there is mounting evidence of heightened significance in mitochondrial bioenergetics, biogenesis, ROS production, and adaptive responses to intrinsic oxidative stress [101].

Recent studies emphasize the potential of targeting mitochondrial metabolism as a viable strategy for cancer therapy [102]. Beyond their bioenergetic role, mitochondria are pivotal in various metabolic processes and cellular functions. While the mammalian mitochondrial genome contains only 37 genes, recent discoveries have unveiled a previously overlooked aspect: the presence of mecciRNAs, which are the circular RNAs originating from the mitochondrial genome. MecciRNAs have the potential to act as molecular chaperones, aiding in the proper folding of nuclear-encoded proteins and facilitating their entry into the mitochondria [103].

Despite these intriguing functions, the biogenesis of mecciRNAs remains enigmatic. This puzzle arises from the lack of essential components required for linear splicing, a process commonly observed in nuclear-encoded RNAs. This deficiency could be attributed to the absence of authentic introns in mitochondrial pre-RNA transcripts or linked to the scarcity of specific protein factors crucial for the splicing process within mitochondria [104]. The functional and characteristic attributes of mecciRNAs, as a novel category of circular RNAs, remain ripe for further investigation.

2.3.2 Role as Molecular Chaperones

The authors of a recent study [103] present compelling evidence, both *in vitro* and *in vivo*, demonstrating that mecciRNAs play a pivotal role in facilitating the translocation of nuclear-encoded proteins into mitochondria by acting as molecular chaperones during the folding process of imported proteins (Figure 6). Specifically, mecciND1 (chrM:3596-3739+) is shown to enhance the mitochondrial entry of RPA proteins, which are integral subunits of the replication protein A (RPA) complex. This complex is known for its involvement in various critical aspects of DNA metabolism, including DNA replication, repair, recombination, telomere maintenance, and the coordination of cellular responses to DNA damage through the activation of the ataxia telangiectasia and Rad3-related protein (ATR) kinase.

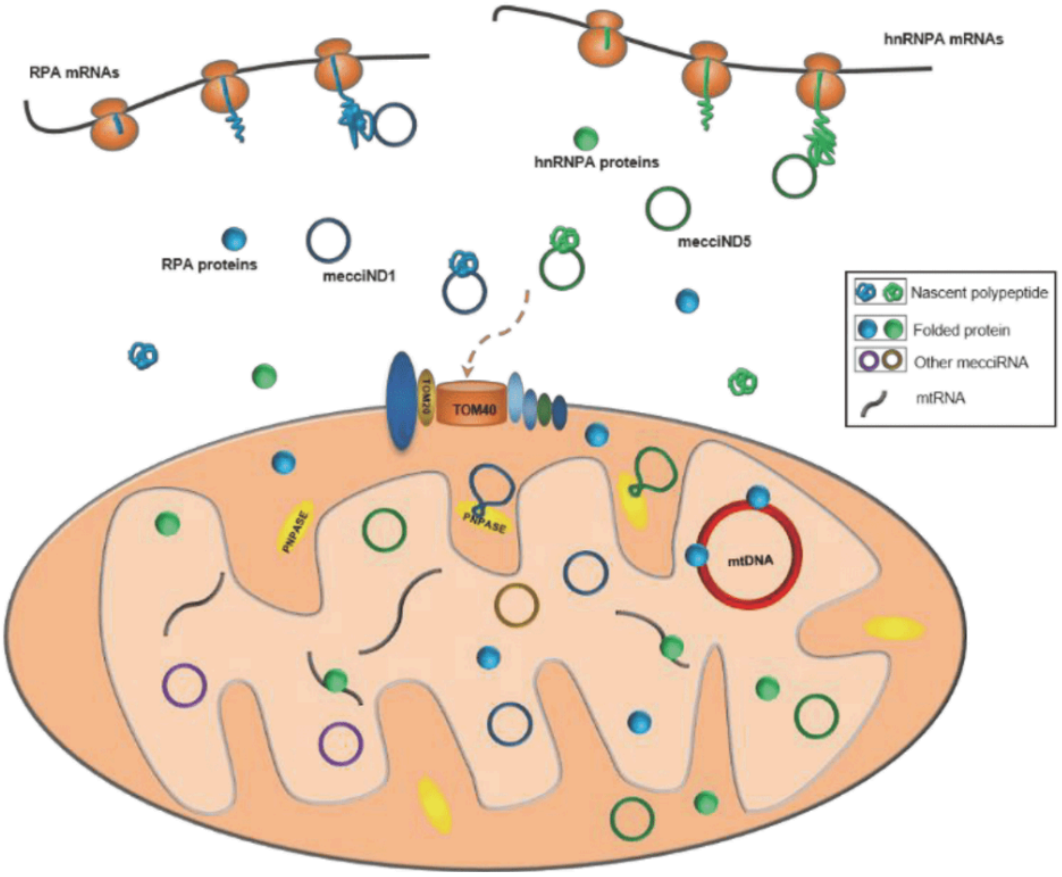

Figure 6: An operational framework elucidates the pivotal role played by mecciRNAs, exemplified by mecciND1 and mecciND5, in facilitating the intricate process of mitochondrial protein importation. These mecciRNAs initiate their functional journey by engaging nascent polypeptides within the cytosol. Subsequently, they assume the role of molecular chaperones, orchestrating and enhancing the efficient importation of proteins into the mitochondria via the TOM40 complex, encompassing vital players like RPA and hnRNPA proteins [103].

Moreover, mecciND5 (chrM:13844-13998+) is demonstrated to facilitate the entry of hnRNPA proteins into the mitochondria. These are RNA-binding proteins that interact with pre-mRNAs in the nucleus, exerting influence over pre-mRNA processing and other aspects of mRNA metabolism and transport. Notably, the protein encoded by hnRNPA1 is among the most abundant core proteins within hnRNP complexes, playing a central role in the regulation of alternative splicing [103].

The study further uncovers intriguing interactions between mecciRNAs and established components involved in mitochondrial protein and RNA importation, such as TOM40 and PNPASE. These

interactions appear to regulate the entry of proteins into the mitochondria; they also report an upregulation of mecciND1 and mecciND5 under stress stimuli and cancer (Figure 7) [103].

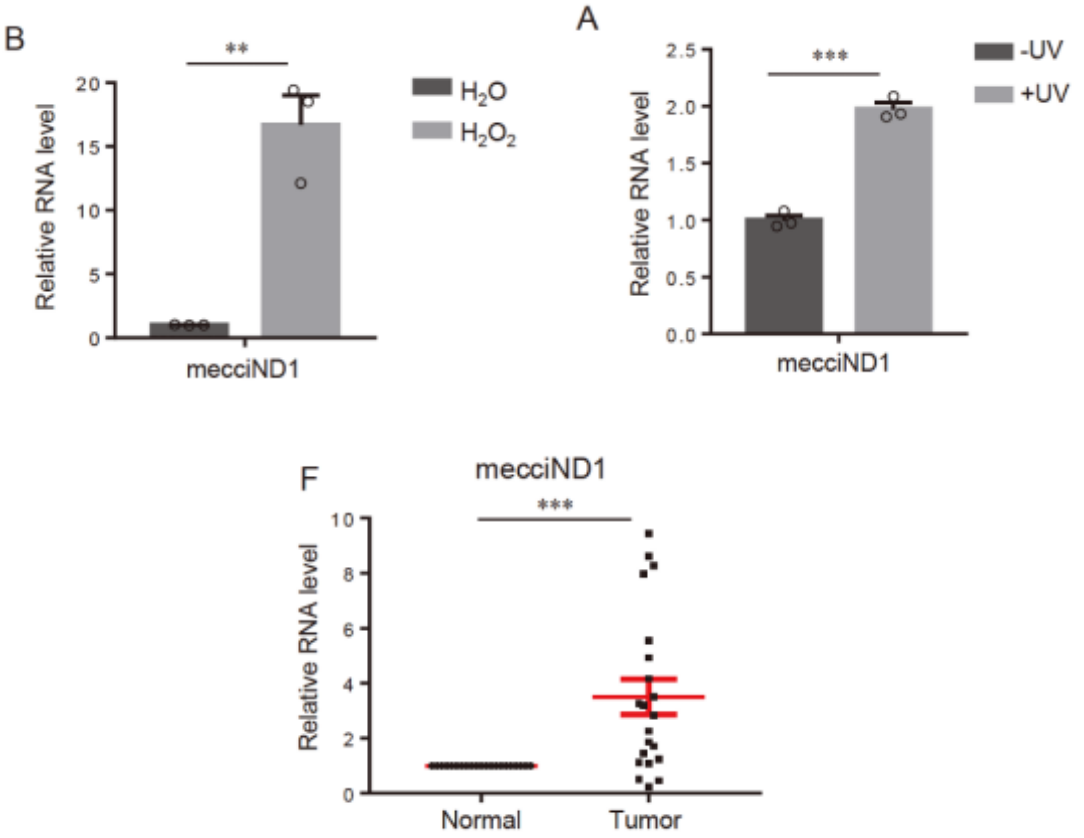

Figure 7: Upregulation of mecciND1 in response to stress. We can observe a notable upregulation of mecciND1 in response to stress stimuli induced by hydrogen peroxide and UV radiation, as well as within hepatocellular carcinoma (HCC) patients. [103].

The authors propose that in response to physiological stress, the upregulation of mecciND1 and mecciND5 leads to heightened levels of RPA proteins, specifically within the mitochondria. Importantly, this upregulation occurs without a simultaneous increase in overall cellular RPA levels (Figure 7, Figure 8). This intriguing observation suggests that mecciRNAs likely serve a pivotal role in facilitating the translocation of these proteins from the cytosol into the mitochondria rather than influencing their transcription or translation processes [103]. Another study (Figure 9) has corroborated the significance of overexpressed circMTND5 (positioned at chrM: 14068-14413+) in contributing to renal mitochondrial injury and kidney fibrosis. This circular RNA's role as a sponge for MIR6812 adds to its intrigue, making it an even more compelling subject of investigation [105].

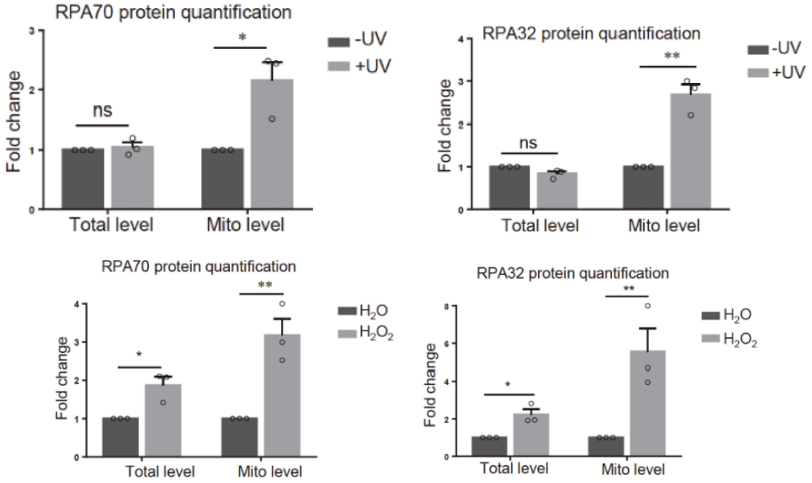

Figure 8: Upregulation of RPA proteins in response to stress. We can see a notable upregulation of RPA proteins in response to stress stimuli induced by hydrogen peroxide and UV radiation mainly in mitochondria. [103].

In the context of this thesis, an effort was made to detect these mecciRNAs within the conditions and investigate whether the RPA gene exhibits upregulation under the CLL condition.

2.3.3 Role in Chronic Lymphocytic Leukemia

A recent paper investigated the role of the mc-COX2 mecciRNA in CLL [106]. According to the authors, four circRNAs, namely chrM:8365-14148-, chrM:13417-13498+, chrM:13791-13854-, and chrM:7585-7981-, exhibit distinct origins, with the former three deriving from the MT-ND5 gene region and the latter from the MT-COX2 gene region, originating from mitochondrial genome; they demonstrated a significant upregulation in the exosome content of patients diagnosed with CLL [106]. The authors also reported that the endogenous reduction of mc-COX2 can have a substantial impact on mitochondrial functions, leading to the suppression of cell proliferation and the induction of cell apoptosis. Conversely, the upregulation of mc-COX2 has been found to exhibit a positive correlation with leukemogenesis and poorer survival outcomes among CLL patients. The functional analysis they performed unveiled intriguing characteristics of mc-COX2. Unlike conventional nuclear circRNAs, mc-COX2 exhibits lower stability and may function through innovative mechanisms beyond its role as a competing endogenous RNA. These unique features suggest that mc-COX2 may play a distinct and pivotal role in cellular processes.

2.4 Circular RNA Degradation

In contrast to the processes governing the formation of circular RNAs (circRNAs), the precise mechanisms through which cells ultimately break down circRNAs remain largely unexplored. Recently, a team of researchers led by Liu [107] has uncovered a previously unknown enzyme, RNase L, that is capable of degrading circRNAs on a global scale. CircRNAs lack free 5′ and 3′ ends, making endoribonucleolytic cleavage the sole viable approach for their degradation. Building upon the discovery by Kim et al. [108], it has been found that a subgroup of circRNAs containing m6A modifications interacts with YTHDF2 in a process influenced by HRSP12. Notably, HRSP12 is preferentially downregulated through RNase P/MRP. The presence of m6A appears to facilitate the degradation of both mRNAs and circRNAs, a revelation that has elevated the scientific community's awareness of this particular mechanism for circRNA breakdown [109].

2.5 CircRNA detection from RNA-seq data with CirComPara2

CirComPara2, as developed by Gaffo et al. in 2022, presents an automated computational pipeline that handles the identification, quantification, and annotation of circular RNAs (Figure 9). Its distinctive feature lies in its utilization of a combination of nine distinct detection tools to locate circular RNAs. It also assesses the levels of linear RNAs and gene expression, enabling comparisons and correlations between circRNA and gene/transcript expression levels [110].

The minimal input required includes (i) the RNA-seq reads in FASTQ format and (ii) a reference genome in FASTA format. Optionally, gene annotation in GTF format and various parameters for customized analysis can be provided. CirComPara2 offers an optional preliminary processing of input raw reads using Trimmomatic [111] to trim adapters or eliminate low-quality reads. The resulting reads are aligned co-linearly to the reference genome using HISAT2 [112]. These aligned reads are then used for the analysis of linear transcripts and for identifying reads that are not co-linearly aligned, which are essential for detecting back-splice events.

The circRNA analysis aligns these non-co-linearly aligned reads independently with five methods, allowing chimeric alignments through Bowtie2 [113], BWA-MEM [114], Segemehl [115], STAR [116], and TopHat-Fusion [117]. This computationally demanding chimeric alignment is performed once per aligner and shared across different circRNA detection tools to enhance the efficiency of CirComPara2.

The outcomes are then subjected to analysis by circRNA detection tools, forming the nine distinct circRNA detection sub-pipelines. The most reliable circRNA identifications are those detected by two or more tools, as indicated by Hansen et al. in 2016 and Hansen in 2018 [118,119], and experimentally demonstrated by Vromman and colleagues [120]. Such identifications exhibit high recall (0.98) and precision (0.99) values. While the soundness of circRNA identifications improves with an increase in the number of tools used, a corresponding rise in discarded circRNAs is observed. Consequently, the default parameters of CirComPara2 utilize seven circRNA detection methods concurrently. The tools used include CIRCexplorer2 [121] coupled with BWA, Segemehl , STAR , TopHat-Fusion aligners, in addition to CIRI2 [122], DCC [123], and Findcirc [71]. CirComPara2 then discards circRNAs not detected by at least two of these methods (Figure 9).

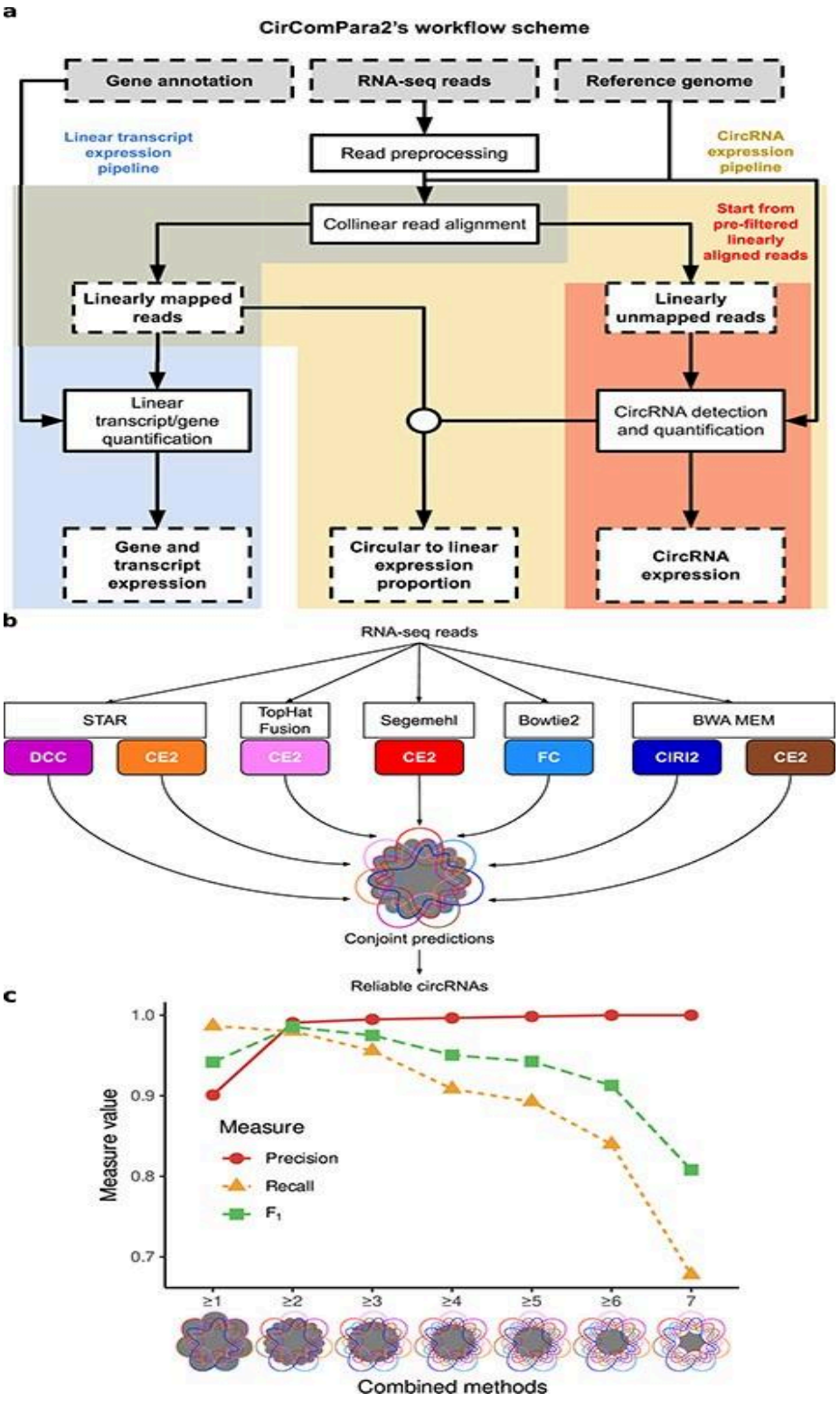

Figure 9: The CirComPara2 workflow. A: dashed lines indicate input and output data; solid lines indicate computing tasks; circular connector indicates merging data. Background colors highlight different pipeline branches (linear transcript analysis in blue, full circRNA analysis in yellow and strict circRNA analysis in red). B: a detail of the CirComPara2 strategy with the circRNA detection methods (coloured rounded corner boxes) and the respective chimeric read aligners (white boxes). The gray area of the Venn diagram represents circRNAs conjointly detected by two or more methods [110].

In a recent large benchmarking [120], CirComPara2 has been demonstrated to be superior to most tools for circRNA detection, with high sensitivity and precision.

CirComPara2 tallies the non-redundant backsplice junction reads (BJRs) by identifying BJR data from each method for every circRNA. This approach maintains information extracted from each individual method.

The analysis of circRNA expression typically entails post-detection data cleaning to eliminate circRNAs with low counts of backsplice junction reads. Small BJR counts often indicate false positives. Hence, filtering based on expression abundance can yield more dependable circRNA collection. By increasing the minimum BJR filter threshold, all methods achieve higher precision but with a corresponding decrease in recall. Despite this, CirComPara2 predictions exhibit robustness compared to methods with equivalent precision, even after low-count filtering. This suggests that circRNAs captured by CirComPara2 and missed by other methods hold relevance and should be regarded as significant molecules.

It is noteworthy to mention that recently, an international team of researchers included CirComPara2 in a comprehensive benchmarking investigation. The purpose was to assess the accuracy and effectiveness of nearly all existing tools for identifying and measuring circRNAs. Through rigorous experimentation involving three different methods and a double-blind procedure, CirComPara2 demonstrated its superiority over the majority of other tools. This led to the highest levels of sensitivity and accuracy in detecting circRNAs, with approximately 99% of circRNAs being successfully confirmed [120]. Consequently, there is a high level of assurance that the circRNAs identified and discussed in this thesis are overwhelmingly authentic.

2.6 Circular to Linear Transcript Expression Proportion

In the assessment of circRNA expression level, two perspectives emerge: firstly, the evaluation of absolute abundance, and secondly, the assessment of abundance relative to the linear transcripts originating from the circRNA host-gene. Notably, circRNA expression often demonstrates synchronization with its corresponding linear transcript. Consequently, the identification of circRNAs manifesting distinct expression patterns from their host-gene expression presents an intriguing avenue of exploration.

Circular RNAs exhibit distinctive expression profiles that set them apart from their linear RNA counterparts. Typically, circRNAs are expressed at considerably lower levels, accounting for approximately 2% to 10% of the total mRNA expression within cells. Conversely, it is not uncommon to observe specific circRNAs expressing at significantly higher levels, sometimes surpassing linear mRNA levels by a factor of ten or more. For instance, consider circRNA HIPK3, which demonstrates expression levels several times greater than linear HIPK3 RNA. Notably, circRNA ciRS-7 showcases an exceptional case, with overwhelming expression in the brain, reaching approximately five times the expression level of the commonly used housekeeping gene GAPDH. Another striking example is circRNA-SRY, which stands as the most abundantly expressed transcript in the mouse testis [124].

The concept of circular-to-linear transcript expression proportion (CLP) encompasses the correlation between circular and linear expressions, thereby reflecting how changes in CLP under different conditions convey the degree of independence between a circRNA and the linear expression of its host gene. CLP values span from 0 to 1: a value of 0 signifies the absence of circular expression, while $0 < CLP < 0.5$ indicates that circRNAs are expressed at lower levels than their corresponding linear isoforms. A CLP of 0.5 suggests an equal abundance of circular and linear transcripts, and values between 0.5 and 1 imply that circular isoforms are more abundant than their linear

counterparts. Notably, a CLP of 1 implies that linear expression relative to the circRNA is undetectable.

Circular ANRIL, unveiled in 2012, emerges as a remarkable offshoot of its linear counterpart, resulting from intricate back-splicing processes. Its fascinating attribute lies in its dual potential: depending on the degree of its circularity or linearity, it may either foster or safeguard against atherosclerosis, demonstrating the profound impact of its structural conformation on biological outcomes [125]. Drawing parallels with ANRIL's role in coronary heart disease, it is equally intriguing to contemplate the discovery of a similar biomarker specific to Chronic Lymphocytic Leukemia.

2.7 Aim of the thesis

The aim of this thesis is to address the existing knowledge gaps regarding t(14;19)-CLL by thoroughly analyzing the circRNAome – the complete landscape of circular RNAs. This largely unexplored area has the potential to provide new insights into the molecular complexities of t(14;19)-CLL. Specifically, this research investigates the dysregulation of circular RNAs in t(14;19)-CLL and compares these findings to both canonical CLL and normal conditions. By examining the circRNAome, we seek to gain a deeper understanding of the aggressive nature of this CLL subtype.

Additionally, this study aims to identify and analyze mitochondrion-derived circular RNAs, assessing their abundance and potential functional roles in the sampled conditions. This exploration is pivotal as mitochondrion-derived circular RNAs could offer further insights into the mitochondrial aspects of cancer biology, which may be crucial for understanding the unique metabolic demands and survival strategies of cancer cells.

This work not only enhances our comprehension of the molecular underpinnings of t(14;19)-CLL but also contributes to the broader field of circRNA research. It offers a more comprehensive understanding of circRNAs' role in cancer biology, potentially paving the way for personalized treatment approaches.

In conclusion, this thesis focuses on investigating independent variations in circRNA expression across three distinct conditions: Chronic Lymphocytic Leukemia (CLL), CLL cases influenced by the t(14;19)(q32;q13) rearrangement, and healthy B-cells. Through this exploration, the study aims to unravel the unique characteristics of circRNAs and their potential implications within these contexts, including an analysis of mitochondrion-derived circular RNAs.

Material and Methods

3.1 Study Design and Patients

This thesis project centers around an RNA-seq dataset of 64 samples, including 55 samples of leukemia cells with a purity level exceeding 97%, originating from a group of 50 patients diagnosed with CLL. Among them, 28 patients exhibited the t(14;19) translocation, and the remaining 22 showcased the typical canonical CLL profile. The collection of these samples was overseen by Prof. L. Trentin and A. Visentin from the Hematology and Clinical Immunology Unit within the Department of Medicine at the University of Padova. The study remarkably incorporates a significant number of instances displaying the rare t(14;19) translocation. These cases were procured from diverse countries, through an international collaboration involving multiple contributors to the ERIC (European Initiative for CLL) consortium, which operates under the broader European Leukemia Net (ELN).

Additionally, the research incorporates 9 samples of B-cells sorted from peripheral blood mononuclear cells (PBMC) of healthy donors. This subset comprises 5 samples collected and sequenced in tandem with the CLL samples, while another 4 samples were previously derived from a distinct project (GEO number ID: GSE110159, Gaffo et al. 2019).

RNA was extracted using the Qiaseq FastSelect kit following the manufacturer's instructions, which includes a step for the depletion of ribosomal RNA. RNA–seq libraries were prepared according to the Illumina Stranded mRNA Prep protocol and sequenced with an Illumina Novaseq 6000 to obtain 120 million reads 150bp paired–end reads per sample.

Each sample of CLL was meticulously linked to comprehensive clinical and molecular information. This encompassed essential parameters such as gender, the mutational status of IGHV and TP53 genes, the presence of trisomy 12, complex karyotype, and chromosomal deletions employed in the Dohner prognostic score. Furthermore, these samples were correlated with the stage of the disease at the time of sample collection, which included baseline cases (n=27), cases at progression (n=21), relapse before therapy (n=3), and relapse after therapy (n=4).

Among the 22 canonical CLL cases, 11 exhibited only a +12 mutation at FISH analysis, while the remaining 11 showed a normal FISH result. Within this group, 12 patients had a mutated IGHV, whereas the remaining 10 had an unmutated IGHV.

Among the 28 patients with t(14;19) translocations, 17 displayed a +12 mutation at FISH, while the other 11 had a normal FISH result. Additionally, 15 of these patients exhibited a complex karyotype, while 13 had a normal karyotype. Among these patients, 4 had a mutated IGHV, and 24 had an unmutated IGHV. Furthermore, 7 patients had mutated TP53, while 21 had unmutated TP53.

Moreover, 21 patients were not characterized by Dohner deletions, while 5 had the 17p deletion, 1 had the 13q deletion, and 1 had the 11q deletion.

3.2 SRA toolkit

SRA toolkit v3.0.1 was used to download the FASTQ files of the 4 B−cells from the Sequence Read Archive, as they were collected for another research project (Gaffo et al. 2019). These files can be found through the GEO number ID: GSE110159.

3.3 CirComPara2

CircRNAs were detected and quantified by CirComPara2 using 7 backsplice detection methods (Figure 9). The analysis was based on the Ensembl GRCh38 human genome and annotation v87.

CirComPara2 default parameters were set for the analyses, which are as follows: adaptors from the Trimmomaticv0.39 TruSeq3-PE-2.fa file; PREPROCESSOR = 'trimmomatic';PREPROCESSOR_PARAMS = 'MAXINFO:40:0.5 LEADING:20 TRAIL-ING:20 SLIDINGWINDOW:4:30 MINLEN:50 AVGQUAL:26'. For thePRJCA000751 data sets, the CROP:150 option was appendedto the parameter. STAR_PARAMS = '—runRNGseed 123 —outSJfilterOverhangMin 15 15 15 15 —alignSJoverhangMin 15 —alignSJDBoverhangMin 15 —seedSearchStartLmax 30 —outFilterScoreMin 1 —outFilterMatchNmin 1 —outFilterMismatchNmax2 —chimSegmentMin 15 —chimScoreMin 15 —chimScoreSeparation 10 —chimJunctionOverhangMin 15'. CIRCRNA_ METHODS = 'circexplorer2_bwa, circexplorer2_segemehl, circ-explorer2_star, circexplorer2_tophat, ciri, dcc, findcirc' ('cir-crna_finder' and 'testrealign' values were used in additionalruns to obtain CircRNA_finder and Segemehl predictions);CPUS = 12; BWA_PARAMS ='-T 19′; SEGEMEHL_PARAMS ='-D 0′ ;BOWTIE2_PARAMS ='—reorder —score-min = C,-15,0 -q —seed123′; DCC_EXTRA_PARAMS ='-fg -M -F -Nr 1 1 -N'; TESTRE-ALIGN_PARAMS = 'q median_1'; FINDCIRC_EXTRA_PARAMS='—best-qual 40 —filter-tags UNAMBIGUOUS_BP —filter-tagsANCHOR_UNIQUE' (this setting implements the optimization suggested by Hansen [119].) MIN_METHODS = 2; MIN_READS = 2;CIRC_MAPPING = "{'SE':['STAR','TOPHAT','BOWTIE2'],'PE':['BWA','SEGEMEHL']}"; HISAT2_PARAMS = '—seed 123'.

3.4 LiftOver

To facilitate the alignment of circular RNA genomic coordinates between different genome versions, the LiftOver tool has been employed. Specifically, this tool converted the genomic coordinates of select circular RNAs from the GRCh37 genome assembly to the GRCh38 assembly. This step was crucial to ensure the accurate and consistent comparison of circular RNA data across different reference genomes during the course of the research.

3.5 Data Analysis

The analysis was carried out with custom scripts in the R programming language (R version 4.3.0) running on a GNU/Linux Ubuntu 22.04LTS operating system. Graphics were generated with the ggplot2_3.4.2, ComplexHeatmap_2.16.0, VennDiagram_1.7.3, limma_3.56.2, sva_3.48.0 and data.table v1.14.8 R packages.

3.5.1 Data Transformation

Analyzing proportions through statistical methods can pose several challenges. The observations are inherently constrained to values between 0 and 1, including both endpoints. Additionally, the variability observed in proportions tends to change systematically with the average of the response. These characteristics likely contradict two key assumptions of conventional statistical approaches, which assume a normally distributed error term with constant variance. Distinctive traits of proportional data render traditional statistical techniques commonly employed by biologists (such as linear regression, ANOVA, and their extensions) generally unsuitable. The difficulties associated with analyzing proportional data have been acknowledged for a considerable time, leading to the

development of various strategies to address them. In cases where proportions stem from discrete counts, employing techniques like logistic or binomial regression is appropriate. These methodologies are well covered in most introductory biostatistics textbooks [126].

In this thesis, the logistic transformation, known as the logit transformation, has been employed to effectively handle the analysis of circular to linear proportional data. The logit transformation was chosen due to its inherent ability to map proportional data onto an unbounded interval, aligning well with the nature of our dataset. Mathematically they are expressed as logit(p)= ln(p/1-p), where p represents the original proportion value and ln denotes the natural logarithm. Furthermore, the symmetry of the logit transformation closely mirrors the behavior of probabilities as they approach their limits (0 and 1), enhancing its suitability for our data. The interpretability of the transformed values in terms of log-odds is advantageous when establishing relationships with other variables, providing valuable insights into direction and magnitude. Additionally, the logit transformation aids in stabilizing variance, making it suitable for statistical techniques that assume constant variance. By leveraging the logit transformation, we aimed to uncover linear relationships and derive meaningful insights from our circular to linear proportional data.

3.5.2 Surrogate Variable Analysis

In molecular biology, high-throughput data is now widely utilized for two main purposes: (i) identifying genomic characteristics linked to outcomes and (ii) constructing signatures for predictive purposes. However, these objectives become intricate due to the presence of latent variables or unwanted variations within the high-throughput data. Among these potential latent variables in experiments, batch effects are the most acknowledged. The impact of batch effects can be significant, potentially jeopardizing the reliability of biological findings [127]. Furthermore, batch effects are not the sole sources of latent variability that could compromise both the statistical and biological validity of a study [128].

To address these challenges, the sva Bioconductor package is introduced as a solution for detecting and mitigating batch effects and other undesirable sources of variation. The sva package encompasses techniques to eliminate artifacts through two approaches: (i) identifying and estimating surrogate variables to account for unidentified sources of variability in high-throughput experiments, and (ii) directly removing known batch effects using ComBat [129]. Research has shown that eliminating batch effects and utilizing surrogate variables can reduce dependency, stabilize error rate estimations, and enhance reproducibility [130].

3.6 Differential Relative Expression Analysis

Read counts spanning the back-splice junction are commonly utilized to approximate circRNA expression, while counts originating from exons or linear splice junctions are employed as an estimation of the expression of the parent gene.

Circular to linear expression proportion (CLP) for each circRNA was computed as such:

$$\mathrm{CLP} = \frac{circular\,reads}{circular\,reads + linear\,reads}$$

To assess variations in the expression of circular RNAs relative to their host gene transcripts across conditions, a Differential Expression Analysis (DEA) using the Limma-Voom method [131] was conducted. DE circRNAs were selected based on a Benjamini-Hochberg-adjusted P-value threshold of

0.01. The choice of employing the Limma-Voom method was informed by a systematic benchmarking study for evaluating the differential expression of circRNAs [132].

Limma, which stands for Linear Models for Microarray Data, is a powerful statistical approach designed for the identification of differential gene expression in high-throughput data. Voom, on the other hand, stands for Variance modeling at the observational level, and it complements Limma by providing a robust transformation of RNA-sequencing (RNA-seq) count data. The Limma-Voom method is particularly well-suited for RNA-seq data analysis, as it effectively accounts for the inherent variability and noise present in such datasets. It models the mean-variance relationship, allowing for precise statistical comparisons between experimental groups. Additionally, it generates robust and accurate estimates of differential expression, making it a reliable choice for identifying significant gene expression changes in our evaluations [131].

3.7 Functional Predictions

3.7.1 CircRNA-miRNA Interactions

To explore the potential interplay between circRNAs and miRNA target sites, we conducted a thorough analysis employing four distinct computational tools: TargetScan, miRanda, PITA, and IntaRNA. Additionally, we utilized mirBase and mirGeneDB as microRNA databases in this investigation.

In the process of identifying potential interactions between circRNAs and miRNA target sites, a rigorous selection process is employed. Firstly, a seed region consisting of 7 nucleotides is chosen for miRNA binding. Additionally, to ensure robust predictions, constraints on miRNA length are set, with a minimum of 16 nucleotides and a maximum of 28 nucleotides allowed. Subsequently, predictions are filtered based on their miranda scores, with only those scoring above the 25th percentile being retained, thereby discarding lower-scoring candidates. Furthermore, to enhance prediction reliability, interactions with binding energies exceeding 20 Kcal/Mol in miranda are excluded. Finally, in TargetScan, predictions containing 7 or more consecutive nucleotide matches (mers) are favored, providing a comprehensive yet highly selective approach to identifying potential circRNA-miRNA interactions [133–138].

3.7.2 OncomiRDB

To create a reliable and authoritative source for investigating the influence of miRNAs on target genes and cellular processes in cancer, researchers meticulously reviewed 2259 instances of miRNA regulations in cancer with direct experimental support extracted from approximately 9000 research abstracts. This comprehensive dataset encompasses over 300 miRNAs, 829 target genes, and spans across 25 different cancer tissues. They have also introduced a user-friendly web platform called oncomiRDB, equipped with both graphical and text-based interfaces, facilitating effortless navigation and retrieval of all these annotations. This valuable resource serves as a valuable tool for both computational analysis and experimental exploration of miRNA regulatory networks and their roles in cancer. For our analysis, we conducted a search within this database to identify the miRNAs that were predicted to have interactions with the circRNAs, with particular attention to those exhibiting either tumor-suppressing or oncogenic characteristics [139].

Results

4.1 Detection of Circular RNAs and quantification of circRNA and circRNA cognate linear expression

The employment of CirComPara2 on the RNA-seq data led to the identification of 71,818 circular RNAs (circRNAs) expressed in CLL samples and in the normal counterpart. Among these, the focus was directed toward the subset of 6237 (9%) circRNAs exhibiting substantial expression within the dataset—meaning they had at least 2 read counts in a minimum of 18 samples. This subset of circRNAs served as the basis for subsequent analysis.

The group of 6237 circRNAs could be attributed to a total of 2841 known genes, 6,112 were associated with exonic regions (Figure 10), while 81 were associated with intronic regions. Moreover, 44 were located in intergenic regions without any annotation, and were thus expressed by putative new human genes.

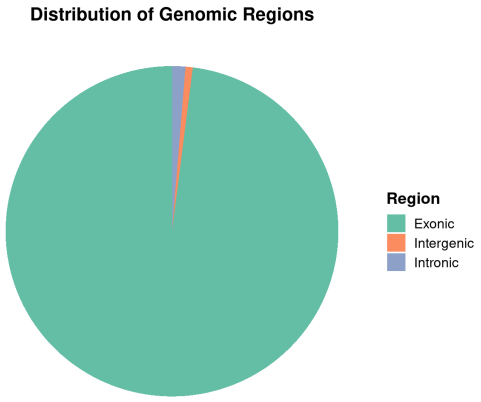

Figure 10: Biotype of circular RNAs. The majority (98%) of circRNAs, are exonic in nature.

Before delving into the Quality Control process, it's essential to understand the foundation upon which they are based. The Circular to Linear Proportion (CLP) is a crucial metric that quantifies the ratio of circular RNA to its cognate linear expression. It's important to note that CirComPara2 facilitated the quantification of circRNA cognate linear expression, forming the basis for calculating the CLP values.

4.2 Quality Control

Quality control (QC) procedures involved assessing sequencing data across all 64 samples, of which 57 were attributed to the baseline and progression-before-therapy stages—these being the primary stages of interest. Remarkably, a consistently good data quality was observed across the entire dataset. Samples 1224666 and 1224668 had substantial reduction in sequencing depth, a consequence of unfavorable shipping conditions exacerbated by a shortage of dry ice in Great Britain. Conversely, sample 1224677 displayed an unusually elevated sequencing depth, all three marking notable outliers in the dataset (Figure 11).

All the remaining samples demonstrated the expected sequencing depth or exceeded it. These samples proved suitable for transcriptomics analysis. In terms of sequence quality, an average of 27% of reads were eliminated during QC procedures. The remaining sequences of high quality effectively

underwent mapping to the reference genome at an 88% success rate, while the residual 12% were utilized for circRNA detection.

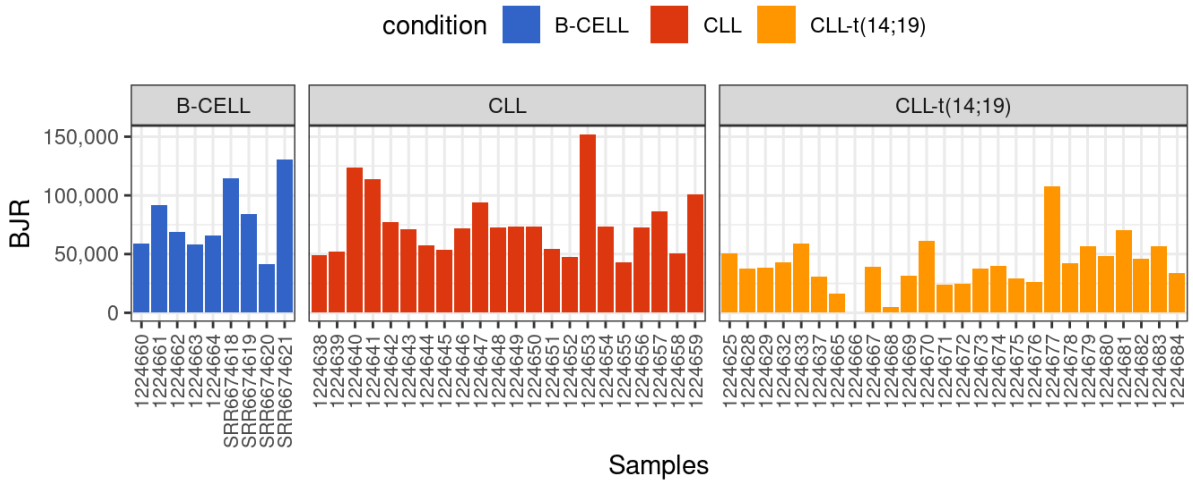

Figure 11: Barplot showing, for each sample, the total number of backsplice junction reads (total bar height).

4.2.1 Removal of Batch Effects and Sources of Noise

A subset of 3223 circRNAs (comprising 50% of the total dataset) characterized by the highest variability in terms of CLP were selected for comprehensive visualization. As vividly depicted through Principal Component Analysis (PCA) in Figure 12, the preeminent contribution to the observed variance rises from technical batch effect.

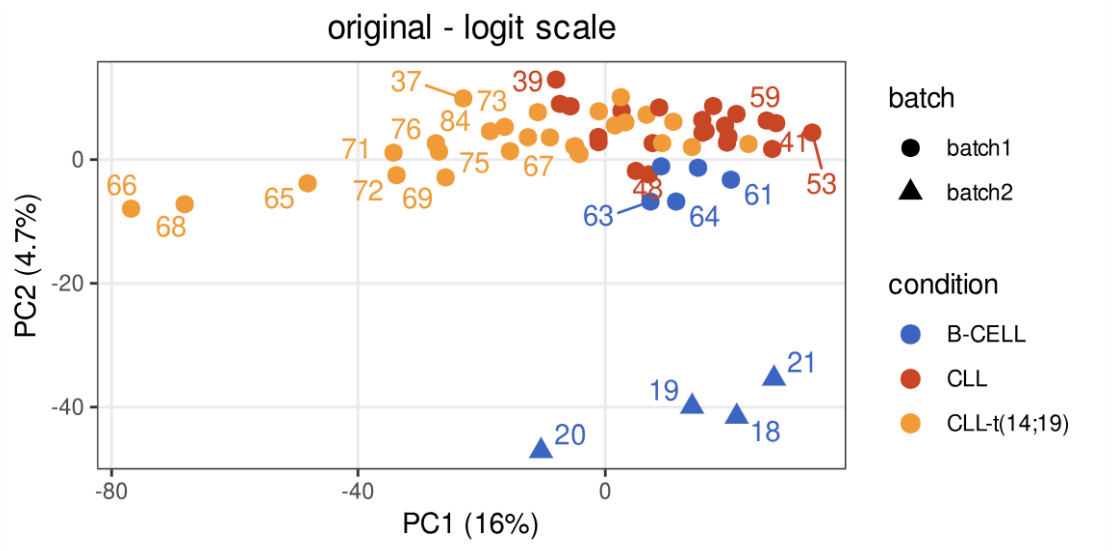

Figure 12: PCA plot based on circRNA expression profiles and with data logit-transformed. The plot underscores the presence of evident technical effects shaping the distribution of samples within the dataset.

Significantly, the observed effects can be traced back to the specific origin of the 4 B-CELL samples that come from an alternative laboratory setting and adhere to a distinct procedural protocol compared to the rest of the samples (Figure 12).

The *limma* package was used to implement batch effect removal. Our analysis revealed the successful mitigation of these effects. However, it is noteworthy that despite this correction, the samples

previously identified with sequencing depth issues persist as outliers. This phenomenon introduces an additional layer of technical variation into our results (Figure 13, Figure 14).

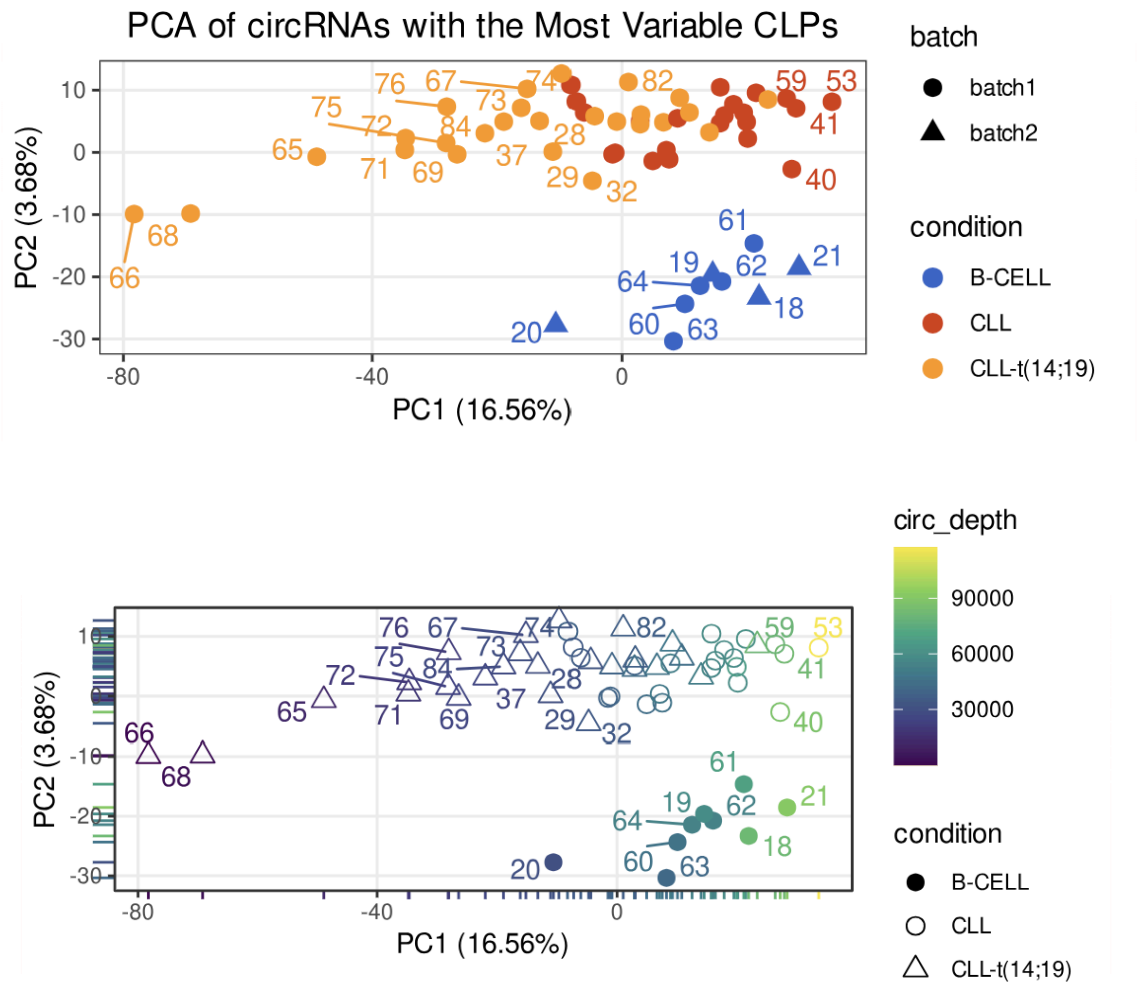

Figure 13: PCA showcasing logit-transformed data subsequent to batch effect mitigation via the limma package. Notably, the plot illustrates a noticeable stabilization and reduction in batch effects, compared to Figure 12. Nevertheless, discernible latent sources of variation persist, underscoring the ongoing necessity for further refinement.

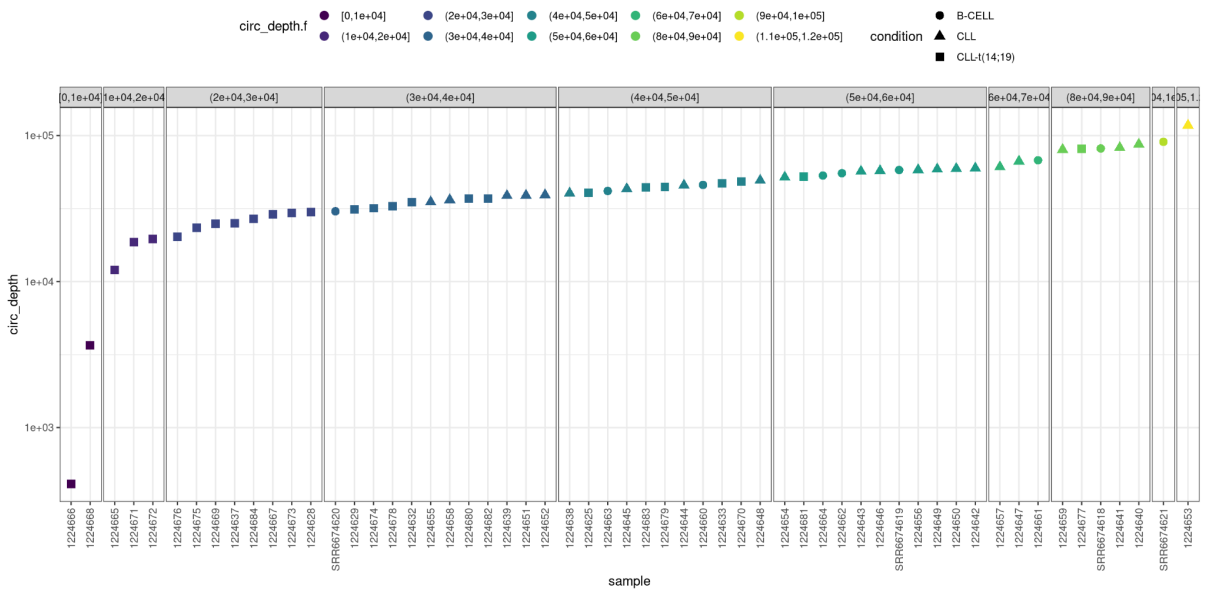

Figure 14: Variation in Sequencing Depths Across Samples. A diverse spectrum of sequencing depths is prominently evident across the examined samples, spanning a dynamic range from 10,000 to 120,000 reads. Notably, this inherent variation in sequencing depth stands out as a significant technical contributor to the observed variability.

In the context of addressing potential sources of bias in our analysis, Surrogate Variable Analysis (SVA) emerged as a pivotal tool. Hidden batch effects, often elusive yet capable of introducing substantial variability, were of particular concern in our high-dimensional dataset. To counteract this, we strategically employed SVA, a well-established statistical technique renowned for its ability to unveil and counterbalance latent sources of variation [140]. By assuming that not all relevant variability is overtly captured in the experimental design, SVA identifies these obscured factors—referred to as "surrogate variables"—through techniques like singular value decomposition (SVD). Incorporating these surrogate variables as covariates in subsequent analyses effectively disentangles their influence from the genuine biological variation. What sets SVA apart is its capability to rectify hidden batch effects while preserving authentic biological signals, thus ensuring that subsequent insights are rooted in genuine data patterns. By adopting SVA, we fortified the reliability and interpretability of our findings by accounting for unobserved sources of variability. Through this approach, our study underscores the importance of robust data preprocessing and analysis to derive credible biological conclusions in the presence of intricate hidden batch effects (Figure 15).

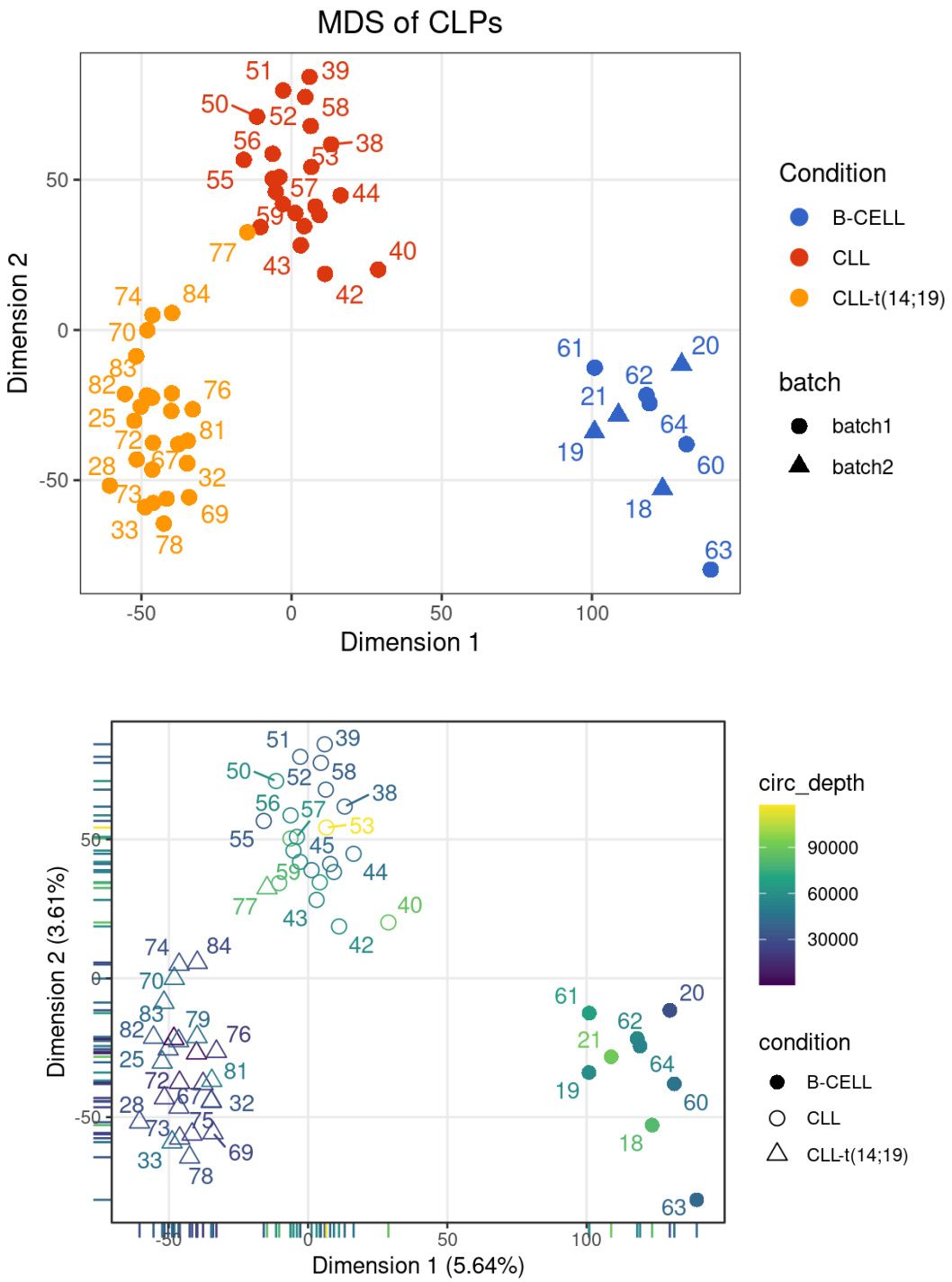

Figure 15: Multidimensional Scaling revealing the impact of batch effect removal through the application of the limma package on logit-transformed data. By concurrently considering confounding factors responsible for technical variation (SVA-correction), the plot elucidates the emergence of sample separation attributed to underlying biological factors.

In this comparative analysis, it becomes evident that by eliminating the three samples exhibiting problematic sequencing depth and subsequently applying batch effect correction to the logit-transformed, SVA-corrected data, the presence of unrelated variations has been significantly mitigated (Figure 16).

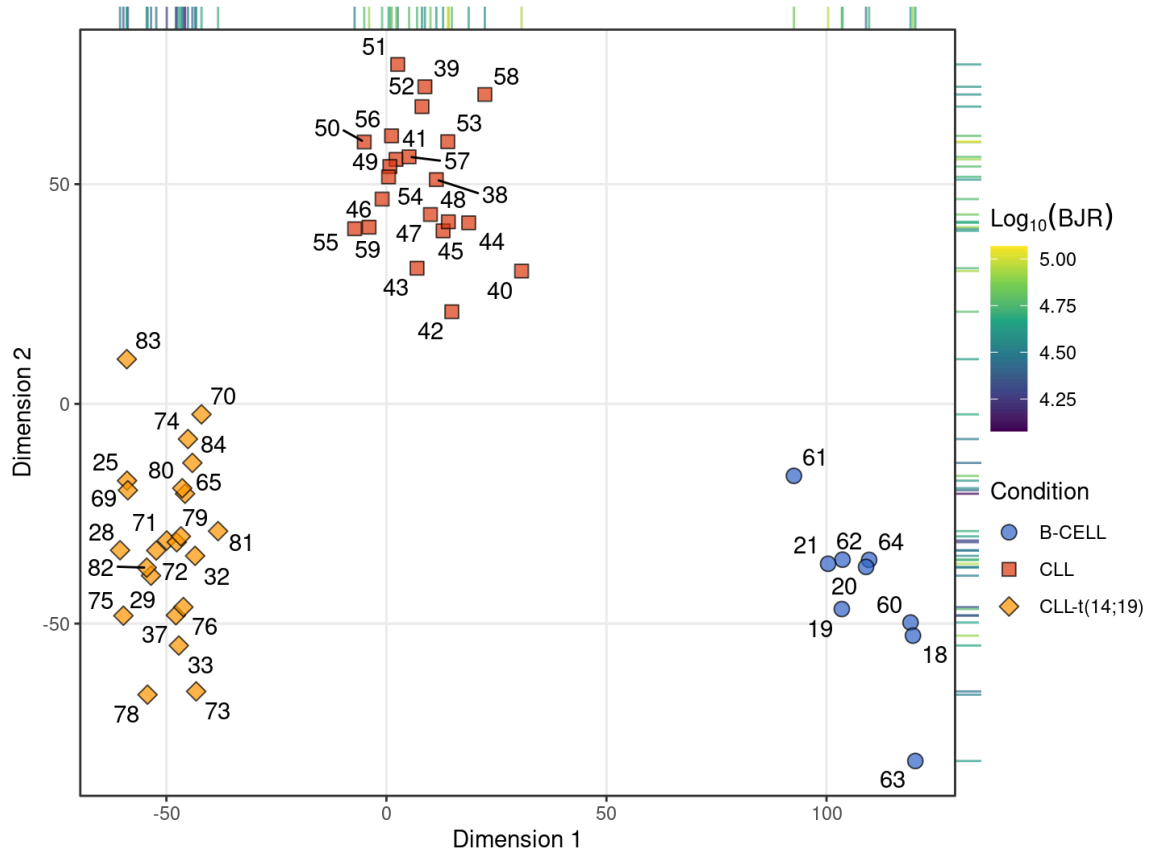

Figure 16: MDS of circRNAs after Removing problematic samples and SVA-correction.

4.3 Circular To Linear Proportion

After Preprocessing, the dataset consists of 54 samples in which a total of 6237 unique circRNAs were expressed. Out of which, the overarching majority (6203, 99.5%) were identified in all the sample groups, whereas 34 (0.5%) were only found in the CLL samples.

The CLP values, which quantify the proportion of circRNAs in each sample, exhibit a diverse range. The mean CLP across all samples is 0.04, with a median of 0.0068, indicating a positively skewed distribution where certain samples possess notably lower CLP values. The observed CLP values span from a very low value of almost zero to a maximum of 1, with circRNAs accounting for an almost negligible fraction of expression to others being the major if not the unique product of the gene region. The standard deviation of 0.1 highlights the variability in CLP values among samples, reflecting the heterogeneity in circRNA abundance. These statistics collectively provide insights into the composition of circRNAs within the dataset, aiding in the understanding of their prevalence and distribution across the samples (Figure 17).

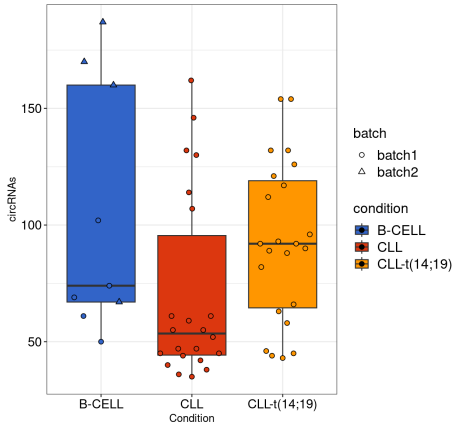

Figure 17: Boxplot of the count of circular RNAs that showed a minimum CLP of 0.50 in at least one sample. Notably, the median count of these circular RNAs is much higher in CLL cases with the translocation and under the B-CELL condition as compared to the typical CLL cases.

It is noteworthy that 41.78 percent of the examined circular RNAs display a Circular-to-Linear Proportion of 0.25 or higher in at least one sample (Figure 18). In contrast, only 2.4 percent exhibit an average CLP of 0.25 across all samples, all of which fall within the upper quantiles of absolute expression. Further analysis reveals that merely 0.35 percent of the circular RNAs display an average CLP of 0.50 or greater, indicating that they are equally or more abundant than their linear counterpart.

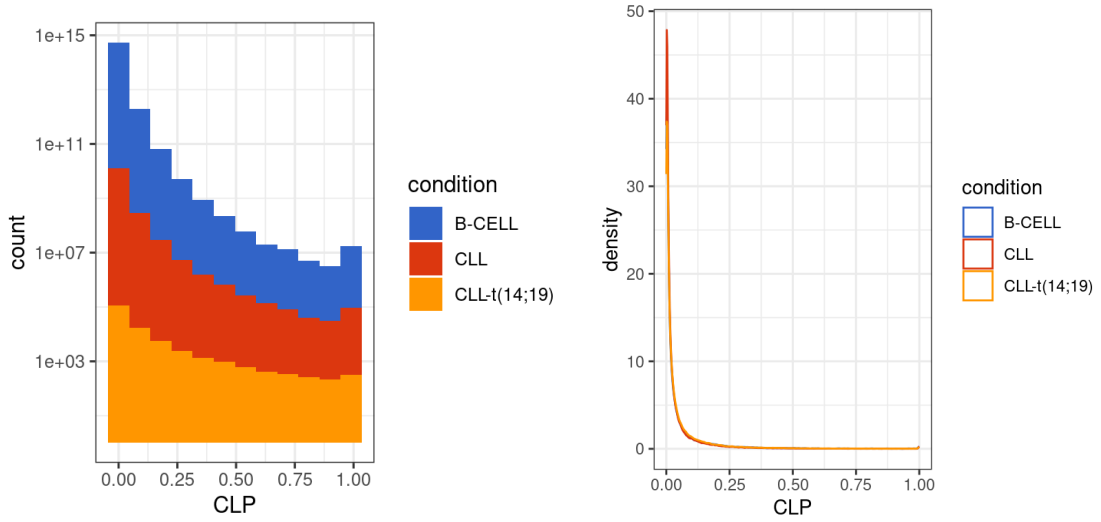

Figure 18: Histogram and Density Plot accounting only circRNAs expressed in each sample, show that there is an overlap in the CLP distribution among the three conditions, suggesting that the majority of CLP values fall within the range of 0 to 0.25.

A distinct subset of 555 circular RNAs (8.9%), boasts a CLP of 0.80 or more in at least one sample. This subgroup garners significant attention due to its notably higher circular abundance relative to the corresponding linear forms. Noteworthy is the fact that 74 of these circular RNAs are consistently present across all three investigated conditions (Figure 19 and Table 3). Remarkably, circNEIL3_4:177353307-177360677:+ exhibits an average CLP of 0.80, underscoring its exceptional status within this context.

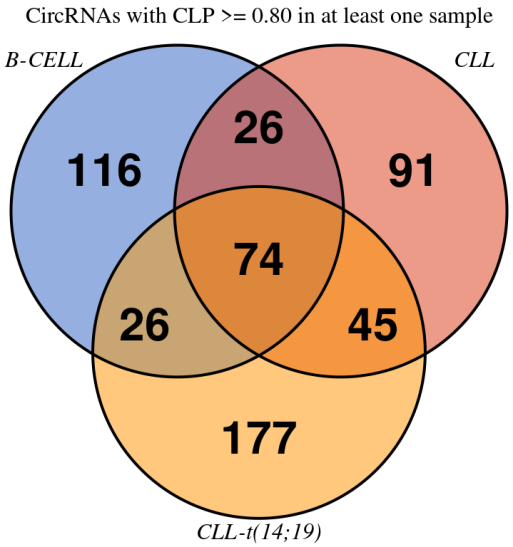

Figure 19: A Venn diagram of the circular RNAs with a CLP of 0.80 or more in at least one sample. Among the 555 circular RNAs displaying a CLP of 0.80 or more in at least one sample, a total of 74 circRNAs are found present across all three conditions. Furthermore, it is worth highlighting that 8 of these circRNAs are among the set of differentially expressed circRNAs.

Interestingly, a Spearman correlation of 0.51 has been identified between the average circular-to-linear proportion of circRNAs and the log10 of their average absolute expression levels. This finding suggests a relationship between circRNA circularization and expression (Figure 20).

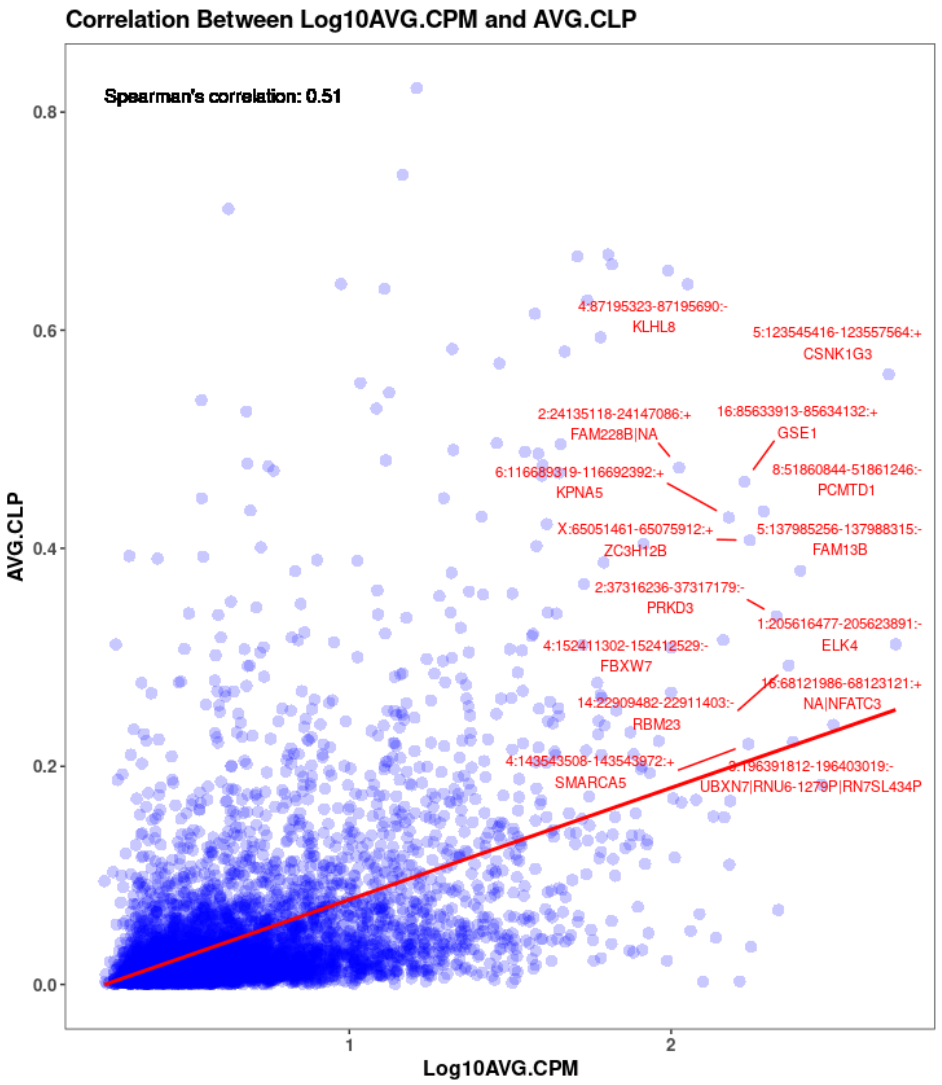

Figure 20: Exploring the relationship between Log_{10}(Average CPM) and Average CLP with a scatterplot. We observe the circRNAs with an average CLP higher or equal to 0.20 and an average CPM expression of more than 100, representing the circRNAs with high absolute count and high CLP.

4.4 Differential Relative Expression Assessment

When we encounter the term "differential expression analysis," it often refers to examining changes of circRNA counts, in other words, their absolute expression. However, in our case, we delve into a different dimension of analysis, focusing on the circular to linear proportion difference in sample groups, which we refer to as "differential relative expression", to discover possible imbalances arising in leukemia cells. This innovative approach goes beyond the conventional assessment of circRNA counts and offers a unique perspective on gene expression dynamics. It not only highlights the regulation of circRNA biogenesis, indicating shifts in splicing or processing machinery favoring circularization with an increase or reduced circRNA production with a decrease but it also offers functional implications. Changes in this ratio can signify alterations in the proportion of circRNAs relative to linear RNAs, which is crucial because circRNAs often play unique roles as microRNA sponges, protein sponges, or gene expression regulators. This metric can unveil disease-related trends, potentially identifying biomarkers or therapeutic targets. Thus, the CLP can provide a more

comprehensive understanding of circRNA biology and its significance in diverse biological processes and diseases.

The Differential Relative Expression Assessment (DREA) revealed significant differences in the relative expression of circRNAs in CLL patients compared to normal B-cells. Specifically, 387 circRNAs were identified as differentially expressed, with 334 showing increased expression and 53 showing decreased expression in CLL patients.

When comparing translocated CLL patients to normal B-cells, a total of 461 circRNAs were found to be differentially expressed, with 313 being upregulated and 148 downregulated. Thus, we observe an increase in CLP in both CLL conditions.

Furthermore, among CLL samples, 332 circRNAs exhibited varying expression patterns depending on the presence of the t(14;19) translocation. Of these, 84 were upregulated, and 248 were downregulated in the group of CLL patients bearing the t(14;19) translocation.

In summary, the analysis revealed distinct expression profiles of circRNAs in CLL patients compared to normal B-cells and further highlighted differences in circRNA expression associated with the t(14;19) translocation. Overall, 754 DE circRNAs were found (Figure 21). Among these, a subset of 10 circRNAs lacked both gene names and assigned gene IDs, while an additional 16 circRNAs were found to have no associated gene names but had gene IDs (Table 1).

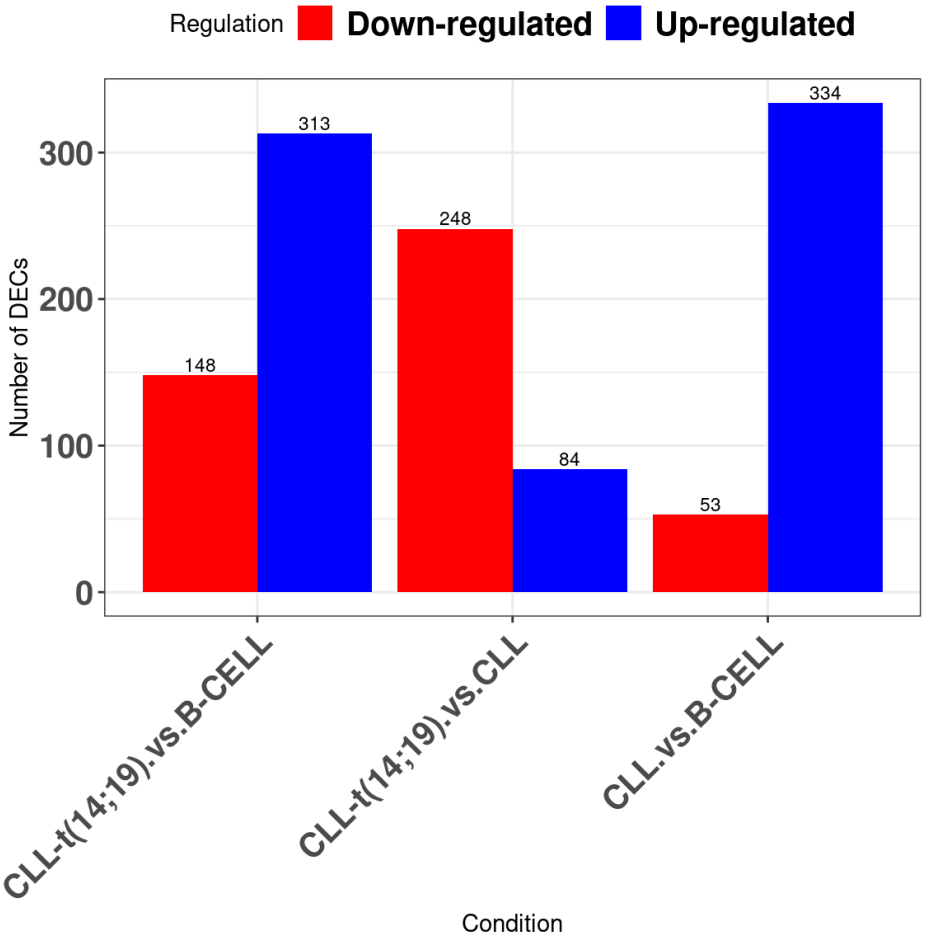

Figure 21: Barplot of DE circular RNAs.

Table 1: Table of 26 circRNAs Lacking Gene Names

circRNA	LFC CLL-t(14;19) vs B-CELL	LFC CLL-t(14;19) vs CLL	LFC CLL vs B-CELL	Padj CLL-t.(14;19) vs B-CELL	Padj CLL-t(14;19) vs CLL	Padj CLL vs B-CELL	Average CLP
17:3705080-3705645:-	-5.42	-5.71	0.29	0.1	0.01	9.50E-01	0.34
8:134812609-134813169:-	-3.34	-5.04	1.7	0.3	0.01	0.6	0.31
6:139598002-139606399:+	13.05	0.62	12.43	0.00E+00	2.10E-01	0	0.3
6:139598005-139606399:+	9.37	-3.33	12.7	0	0.03	0.00E+00	0.16
18:57731413-57732023:+	-0.73	0.08	-0.81	3.00E-02	0.8	0.00E+00	0.16
17:3705097-3705645:-	-9.88	-9	-0.88	0	0	0.73	0.15
13:80687235-80757286:+	8.04	-2.41	10.45	0.00E+00	1.60E-01	0	0.14
1:170954010-170960408:-	8.9	-1.7	10.6	0	0.03	0.00E+00	0.09
12:9584105-9607802:+	5.25	-3.58	8.83	0.14	0.12	0	0.08
4:11066835-11161821:+	11.73	9.92	1.81	0.00E+00	0	2.60E-01	0.06
13:80684355-80757286:+	3.45	-7.04	10.48	1.90E-01	0	0.00E+00	0.06
6:139429080-139429613:+	6.82	0.44	6.39	0	0.86	0	0.05
6:156230443-156294754:-	9.6	0.72	8.88	0.00E+00	2.00E-02	0	0.05
5:79256870-79257484:-	9.72	-0.03	9.74	0.00E+00	0.97	0.00E+00	0.04
6:139429080-139451669:+	6.03	-1.05	7.08	4.00E-02	0.69	0.00E+00	0.04
4:10859197-10871074:+	9.48	7.91	1.57	0.00E+00	0	3.60E-01	0.03
6:156252319-156294754:-	6.84	-1.35	8.19	0	0.19	0.00E+00	0.03
12:50129526-50130217:+	9.03	0.28	8.75	0	8.20E-01	0.00E+00	0.03
6:156236726-156294754:-	8.95	-0.24	9.19	0	0.76	0.00E+00	0.02
13:80687235-80823956:+	3.95	-4.9	8.85	1.00E-02	0	0.00E+00	0.02
6:156294668-156331021:-	1.66	-6.56	8.23	0.4	0	0	0.02
6:156252319-156331021:-	5.07	-3.76	8.83	0	0.00E+00	0	0.02

6:156252319-156370319:-	11.15	1.71	9.44	0	0.08	0	0.01
6:156370050-156372293:-	8.82	2.08	6.74	0	0.17	0	0.01
5:177619077-177641939:-	1.18	-6.55	7.72	0.7	0	0	0.01
17:7513537-7513726:-	0.93	4.39	-3.46	0.67	0	0.01	0

We observe that among the differentially expressed circRNAs, the majority exhibit upregulation in CLL conditions when compared to the B-CELL condition. However, in the canonical and non-canonical CLL comparison, the predominant trend is downregulation.

The clustering analysis of the 754 significantly differentially expressed circular RNAs (Figure 22) highlighted seven distinct clusters. Specifically, the first and fourth clusters exhibit upregulation in both CLL groups when compared to B-cells. The second cluster showcases upregulation in CLL-t(14;19) relative to CLL and B-cell groups. Conversely, the third and fifth clusters demonstrate downregulation in CLL and B-cell groups compared to CLL-t(14;19). Finally, the sixth and seventh clusters illustrate downregulation in CLL-t(14;19) when compared to the other two groups.

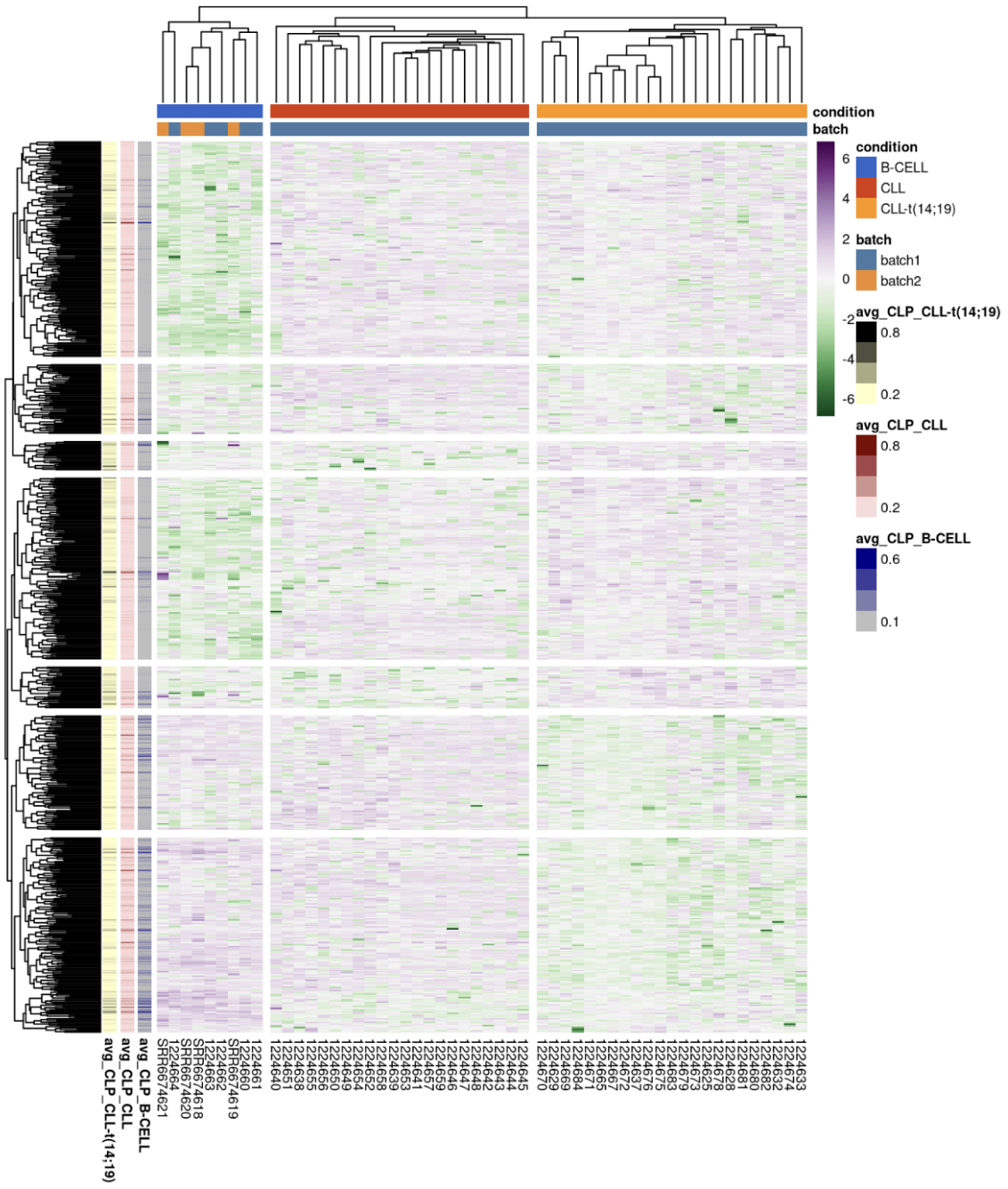

Figure 22: Heatmap of the 754 DE circRNAs in at least one comparison (t(14;19) vs. CLL, t(14;19) vs. B–cells or CLL vs. B–cells)

Notably, a considerable number of differentially expressed circRNAs (specifically, 272 circRNAs exhibiting significant differential expression between both CLL types and B-cells, in addition to 12 circRNAs displaying differential expression between the two CLL groups) exhibit distinct expression patterns when comparing the normal condition to the two CLL conditions, as illustrated in Figure 25.

These DE circRNAs exhibit a consistent pattern of dysregulation. In other words, when a circRNA is upregulated or downregulated in t(14;19) compared to B-cells, it also demonstrates a similar dysregulation pattern in canonical CLL compared to B-cells.

Among the 754 DE circular RNAs analyzed (Figure 22), 284 displayed differential expression in both the "CLL-t(14;19) vs. B-cell" and "CLL vs. B-cell" comparisons. Out of these, 258 circRNAs exhibited upregulation in both types of CLL compared to B-CELL, while 26 circRNAs demonstrated downregulation in both types of CLL compared to B-CELL.

Interestingly, within the group of circRNAs upregulated in both CLL types compared to B-CELL, 9 circRNAs were found to be downregulated in the CLL bearing the t(14;19) translocation when compared to the canonical CLL (circTNFRSF1A_12:6333068-6334244:-, circLINC00377_13:81018332-81020806:+, circCHD2_15:93000511-93009323:+, circCLNK_4:10501255-10513597:-, circCLNK_4:10513463-10542280:-, circCLNK_4:10525840-10542280:-, circARHGAP24_4:85923647-85942273:+, circGENE_6:156252319-156331021:-, circRAB11FIP1_8:37862736-37877551:-). On the other hand, one circRNA (circUGCG_9:111914604-111915836:+), which was downregulated in both CLL types compared to B-CELL, also exhibited downregulation in the CLL with the t(14;19) translocation compared to the canonical CLL. Additionally, two circRNAs (circSLC15A4_12:128799258-128815070:-, circC2orf76_2:119311621-119339971:-), upregulated in both CLL types compared to B-CELL, were also upregulated in the CLL with the t(14;19) translocation compared to the canonical CLL.

In both types of Chronic Lymphocytic Leukemia, we observed dysregulation in a set of 12 circRNAs. However, the extent of dysregulation varied between the two CLL types. Most of these circRNAs exhibited a more pronounced variation in typical CLL, except for two specific cases. First, circUGCGA 9:111914604−111915836 displayed a greater degree of downregulation in CLL with the translocation compared to its downregulation in typical CLL. Second, circC2orf76 2:119311621−119339971 exhibited a more substantial upregulation in CLL with the translocation compared to its upregulation in typical CLL (Figure 23 and Figure 24). These findings highlight the distinct molecular profiles and dysregulation patterns within these CLL subtypes.

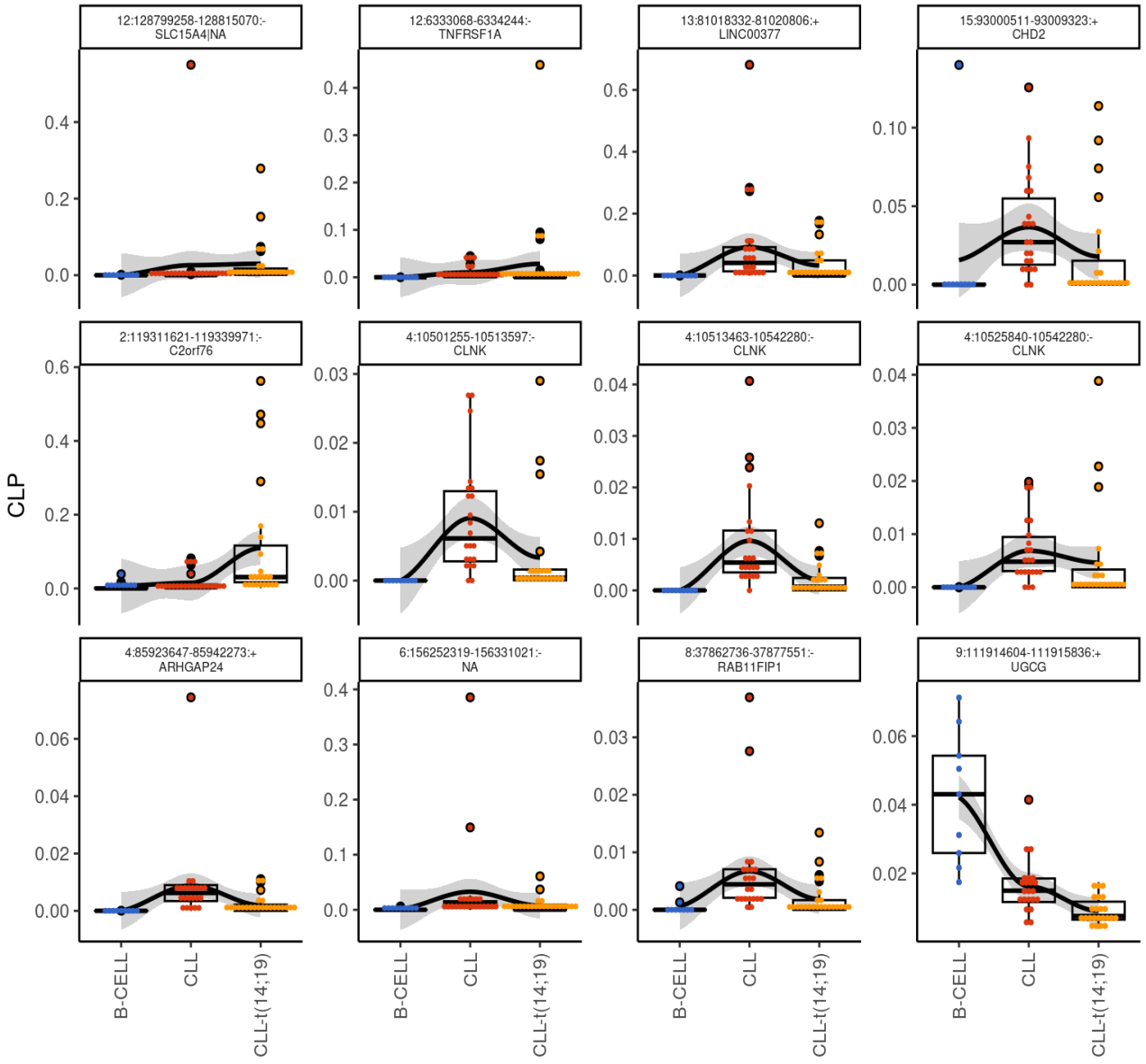

Figure 23: Boxplot illustrating the expression profiles of the twelve circRNAs found to be differentially expressed (DE) in all three comparisons (The threshold for p-value is 0.01).

Remarkably, we identified 100 circRNAs that exhibited differential expression (DE) in both the "CLL-t(14;19) vs. B-cell" and "CLL-t(14;19) vs. CLL" comparisons, without such differences in the "CLL vs. B-cell" comparison. This specific pattern of alteration was exclusive to the more aggressive t(14;19)-CLL subtype, implying that these circRNAs may serve as a distinct signature associated with circRNA relative expression variations linked to this genetic lesion (Figure 25).

Symmetrically, 30 circRNAs displayed DE in both the "CLL vs. B-cell" and "CLL-t(14;19) vs. CLL" comparisons, resulting in dysregulation only in canonical CLL and not in the aggressive form. It is intriguing to observe a reciprocal pattern in the expression of circular RNAs between chronic lymphocytic leukemia and CLL cases bearing the translocation. Specifically, those circRNAs that exhibit upregulation in CLL as compared to B cells display downregulation in CLL cases with the translocation compared to typical CLL, whereas the circRNAs showing downregulation in the former comparison show upregulation in the latter (Figure 25).

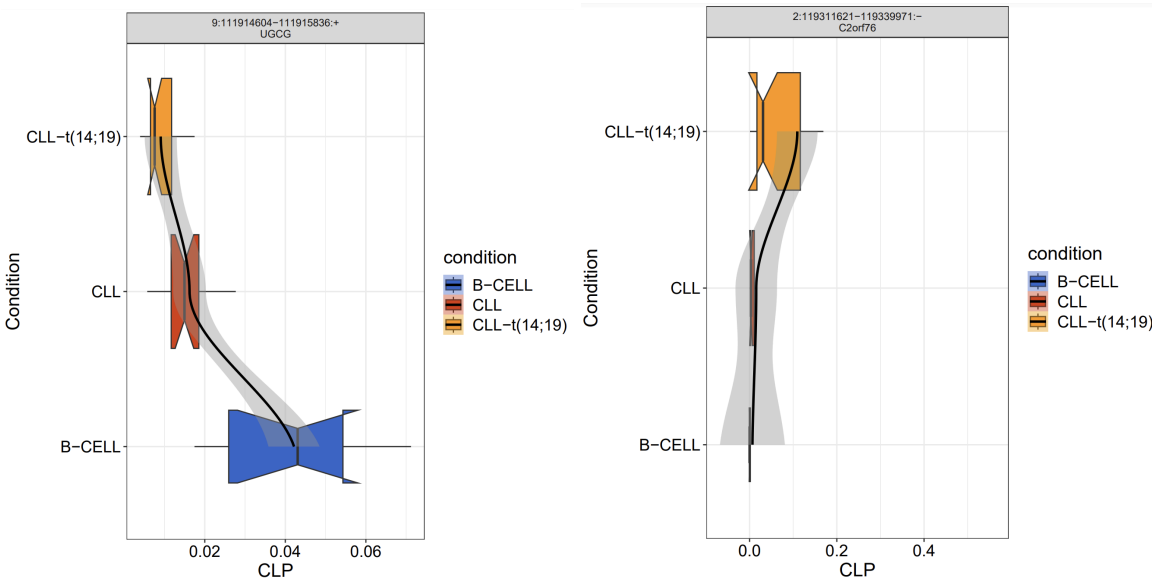

Figure 24: Boxplot illustrating the expression profiles of the two specific circular RNAs among the twelve circRNAs found to be differentially expressed (DE) in all three comparisons. These two circRNAs exhibit more aggressive dysregulation in CLL cases bearing the t(14;19) translocation compared to those with the typical CLL presentation. This observation makes the connection between circRNAs and the aggressiveness of CLL with the translocation even more intriguing, opening up new avenues for further investigation and potential therapeutic strategies.

Furthermore, there were 77 circRNAs that exclusively exhibited DE in the "CLL-t(14;19) vs. B-cell" comparison, 190 circRNAs specifically displayed DE in the "CLL-t(14;19) vs. CLL" comparison, and 73 circRNAs were solely DE in the "CLL vs. B-cell" comparison (Figure 25).

These findings indicate that there are specific circRNAs whose expression levels are associated with the different conditions. The overlapping circRNAs suggest potential shared regulatory mechanisms or biological significance across these different comparisons, while the circRNAs differentially expressed exclusively in one CLL group may play distinct roles in canonical and translocated CLL (Figure 25).

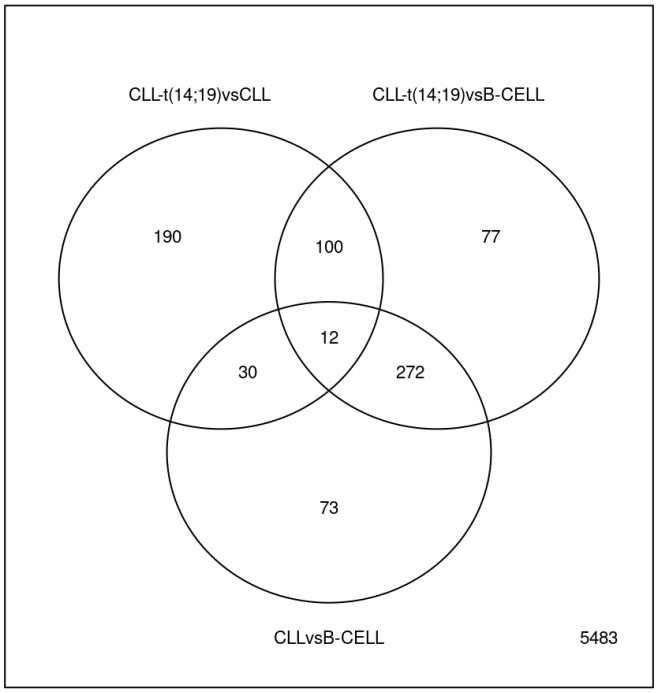

Figure 25: Overlap of circRNA with relative expression (CLP) altered in the different sample group comparisons.

4.5 Exploring Intriguing Circular RNAs

In the complex landscape of circRNAs, it is evident that many of these molecules undergo dysregulation. However, not all circRNAs functionally impact biological processes. To identify and prioritize the most captivating circRNAs, we focus on two distinct categories. We prioritize circRNAs that not only demonstrate significant differential relative expression across various conditions but also maintain an average circular-to-linear proportion of at least 0.20 in at least one of the experimental conditions characterized by a high degree of dysregulation and a substantial log-fold change (LFC). CircRNAs with established associations in the scientific literature with CLL also hold a place in our selection process due to their potential significance. These criteria collectively guide our quest to unravel the functional relevance of circRNAs in the intricate web of biological processes. It is worth noting that all of these circRNAs fall within the upper quantile of absolute expression.

We identified a set of 52 circular RNAs that displayed significant differential relative expression among conditions and also exhibited an average circular-to-linear proportion of at least 0.20 in at least one of the experimental conditions (Table 2).

Table 2: Fiftytwo circRNAs with Average CLP of at least 0.20 in at least one condition and significant CLP variation in at least one comparison. The table includes gene symbols and ENSEMBL gene IDs for the circRNAs, along with their average circular-to-linear proportion within each condition and across all conditions. Additionally, it provides the absolute expression percentile for each circRNA within individual conditions and their average across all conditions. Moreover, it presents the log fold change of circRNA dysregulation in each comparison, along with corresponding adjusted p-values; the p-value threshold is set at 0.01.

circRNA	Gene Symbol	LFC CLL-(14;19) vs B-CELL	LFC CLL-(14;19) vs CLL	LFC CLL vs B-CELL	Padj CLL-(14;19) vs B-CELL	Padj CLL-(14;19) vs CLL	Padj CLL vs B-CELL	ENSEMBL Gene ID	Percentile CLL-(14;19)	Percentile CLL	Percentile B-CELL.	Average Percentile	Average CLP B-CELL	Average CLP CLL	Average CLP CLL-t(14;19)	Average CLP
1:10347760-10348733:+	KIF1B	4.12	2.74	1.38	0	0	0.42	ENSG00000054523	0.91	0.81	0.36	0.87	0.1	0.05	0.31	0.15
1:203707008-203708104:+	ATP2B4	-0.51	0.59	-1.11	0.35	0.04	0	ENSG00000058668	0.97	0.97	0.89	0.96	0.23	0.06	0.17	0.15
1:235437321-235442911:+	TBCE|TBCE	-3.42	-1.69	-1.73	0	0.04	0.1	ENSG00000285053|ENSG00000284770	0.82	0.87	0.91	0.86	0.22	0.07	0.03	0.11
1:84865384-84866138:-	LPAR3	15.21	0.2	15.01	0.00E+00	0.94	0.00E+00	ENSG00000171517	0.99	0.97	0.32	0.98	0.26	0.83	0.85	0.65
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	-0.89	-0.41	-0.48	0	0.08	0.12	ENSG00000152078|ENSG00000271092	0.98	0.99	0.97	0.99	0.36	0.33	0.31	0.33
10:101667885-101676436:-	FBXW4	-0.53	0.03	-0.56	0.01	0.88	0	ENSG00000107829	0.99	0.99	0.99	0.99	0.21	0.13	0.14	0.16
10:103347237-103348999:-	PCGF6	-1.78	-1.64	-0.14	0.01	0	0.87	ENSG00000156374	0.88	0.92	0.91	0.91	0.31	0.23	0.07	0.2
10:17704430-17705741:+	STAM	-0.82	-0.6	-0.22	0.01	0	0.55	ENSG00000136738	0.99	1	0.99	0.99	0.22	0.2	0.12	0.18
12:25017133-25026866:+	IRAG2	9.99	0.3	9.69	0.00E+00	0.8	0.00E+00	ENSG00000118308	0.95	0.92	0.71	0.93	0.19	0.24	0.26	0.23

12:47104112-47108164:+	PCED1B	1.57	1.05	0.52	0.01	0	0.48	ENSG00000179715	0.97	0.94	0.81	0.96	0.06	0.07	0.29	0.14
13:20731840-20732121:-	EEF1AKMT1	1.9	2.15	-0.25	0.07	0	0.86	ENSG00000150456	0.95	0.89	0.87	0.93	0.07	0.1	0.47	0.21
13:40836596-40855815:-	NA|TPTE2P5	10.13	-0.87	10.99	0.00E+00	0.4	0.00E+00	ENSG00000290476|ENSG00000168852	0.88	0.88	0.36	0.86	0.01	0.27	0.13	0.14
13:40908158-40914105:-	NA|SUGT1P3|SUGT1P3	-0.9	-0.29	-0.6	0	0.14	0.01	ENSG00000290476|ENSG00000290464|ENSG00000239827	0.97	0.98	0.98	0.98	0.46	0.38	0.25	0.36
13:42917540-42970670:-	EPSTI1	-2.81	-1.98	-0.83	0	0	0.39	ENSG00000133106	0.86	0.88	0.9	0.88	0.22	0.11	0.04	0.12
13:75621762-75727098:+	NA|LMO7|NA	1.44	1.85	-0.41	0.07	0	0.67	ENSG00000261553|ENSG00000136153|ENSG00000228444	0.95	0.93	0.91	0.94	0.06	0.05	0.29	0.13
13:99238426-99244624:+	UBAC2	-0.54	-0.14	-0.4	0	0.34	0.016	ENSG00000134882	1	1	1	1	0.39	0.3	0.28	0.32
15:100330888-100334180:-	ADAMTS17	11.92	1.32	10.6	0	0.72	0	ENSG00000140470	0.7	0.84	0	0.74	0.11	0.21	0.3	0.21
15:58912562-58916999:-	SLTM	-0.69	-0.17	-0.51	0	0.34	0	ENSG00000137776	0.99	0.99	1	0.99	0.45	0.32	0.27	0.35
16:10428867-10440462:+	ATF7IP2	-0.94	-0.44	-0.5	0	0.06	0.13	ENSG00000166669	0.96	0.96	0.97	0.96	0.21	0.15	0.1	0.16
17:31155982-31169997:+	NF1	-8.86	-8.25	-0.6	0.00E+00	0.00E+00	0.75	ENSG00000196712	0.7	0.87	0.92	0.85	0.21	0.14	0.04	0.13
17:3705080-3705645:-	.	-5.42	-5.71	0.29	0.09	0	0.95	.	0.64	0.83	0.97	0.88	0.43	0.37	0.27	0.36

17:67945408-67975958:+	BPTF|NA|NA	0.48	0.13	0.35	0	0.34	0.02	ENSG00000171634|ENSG00000279573|ENSG00000279880	1	1	1	1	0.15	0.2	0.21	0.19
18:57731413-57732023:+	.	-0.73	0.08	-0.81	0.02	0.8	0	ENSG00000267787	0.99	1	1	1	0.26	0.13	0.14	0.18
19:23920690-23921010:+	ZNF726	-0.42	1.07	-1.49	0.6	0	0	ENSG00000213967	0.95	0.94	0.96	0.95	0.37	0.14	0.19	0.24
19:47362475-47362693:+	DHX34	-6.85	-8.1	1.25	0	0.00E+00	0.7	ENSG00000134815	0.71	0.83	0.84	0.8	0.21	0.29	0.04	0.18
19:52588531-52592228:+	ZNF137P|NA	-7.52	-6.31	-1.21	0	0	0.6	ENSG00000123870|ENSG00000290721	0.73	0.86	0.91	0.84	0.22	0.08	0.04	0.11
2:189719571-189720787:+	ANKAR	-0.93	0	-0.93	0	1	0	ENSG00000151687	0.96	0.96	0.97	0.96	0.23	0.11	0.1	0.15
2:210104103-210154611:-	KANSL1L	-4.28	-3.94	-0.35	0.04	0	0.9	ENSG00000144445	0.79	0.91	0.94	0.89	0.21	0.12	0.03	0.12
2:58221941-58232112:-	FANCL	-0.26	0.21	-0.47	0.21	0.12	0	ENSG00000115392	1	1	1	1	0.44	0.34	0.37	0.38
2:87793630-87794193:-	RGPD2	-8.9	-1.3	-7.6	0	0.52	0	ENSG00000185304	0.92	0.91	0.95	0.93	0.2	0.04	0.02	0.09
2:95148884-95153259:-	NA|ZNF514	0.79	1.4	-0.61	0.28	0	0.38	ENSG00000289685|ENSG00000144026	0.95	0.89	0.93	0.93	0.11	0.11	0.31	0.18
21:41226265-41257326:+	BACE2	-8.91	-7.1	-1.81	0.00E+00	0.00E+00	0.33	ENSG00000182240	0.82	0.95	0.98	0.95	0.38	0.14	0.02	0.18
3:53497151-53501720:+	CACNA1D	-8.03	-8.91	0.88	0	0	0.84	ENSG00000157388	0.56	0.68	0.98	0.91	0.18	0.4	0.03	0.2

4:10520790-10542280:-	CLNK	9.71	0.27	9.44	0.00E+00	0.55	0.00E+00	ENSG00000109684	0.96	0.97	0.67	0.96	0.22	0.05	0.02	0.1
5:123545416-123557564:+	CSNK1G3	-0.53	-0.07	-0.46	0	0.67	0	ENSG00000151292	1	1	1	1	0.66	0.55	0.53	0.58
5:161330882-161336769:-	GABRB2	-2.03	-6.71	4.67	0.7	0	0.22	ENSG00000145864	0.75	0.97	0.73	0.93	0.31	0.23	0.05	0.2
6:138943512-138944622:-	REPS1	-0.6	-0.3	-0.3	0	0.04	0.16	ENSG00000135597	1	1	1	1	0.32	0.26	0.21	0.27
6:139598002-139606399:+	.	13.05	0.62	12.43	0.00E+00	0.21	0.00E+00	ENSG00000226571	0.94	0.91	0.64	0.92	0.22	0.26	0.38	0.29
6:139598005-139606399:+	.	9.37	-3.33	12.7	0	0.03	0.00E+00	ENSG00000226571	0.85	0.87	0.5	0.84	0	0.3	0.08	0.13
6:145888019-145894977:-	SHPRH	-0.64	0.02	-0.66	0.01	0.94	0	ENSG00000146414	0.99	0.99	0.99	0.99	0.39	0.25	0.24	0.29
7:22976209-22991139:-	FAM126A	-1.12	-0.61	-0.51	0	0	0.06	ENSG00000122591	0.97	0.97	0.99	0.98	0.33	0.23	0.15	0.24
8:134812609-134813169:-	.	-3.34	-5.04	1.7	0.3	0	0.6	.	0.84	0.92	0.94	0.91	0.27	0.5	0.14	0.3
8:141253988-141254629:-	SLC45A4	0.18	-0.4	0.58	0.53	0	0	ENSG00000022567	0.99	0.99	0.99	0.99	0.51	0.64	0.53	0.56
8:47396375-47407961:+	SPIDR	-0.36	0.08	-0.44	0.04	0.59	0	ENSG00000164808	1	1	1	1	0.29	0.21	0.21	0.24
9:127339644-127342334:+	GARNL3	-5.26	-5.9	0.64	0	0.00E+00	0.73	ENSG00000136895	0.81	0.85	0.81	0.83	0.32	0.44	0.1	0.29

9:37126311-37147445:+	ZCCHC7	-0.46	-0.02	-0.44	0	0.9	0	ENSG00000147905	1	1	1	1	0.22	0.16	0.16	0.18
9:96458378-96458541:+	HABP4	11.75	-1.51	13.27	0.00E+00	0.34	0.00E+00	ENSG00000130956	0.82	0.86	0.13	0.82	0.11	0.31	0.15	0.19
9:99959916-99960155:+	STX17	-6.47	-7.19	0.72	0	0	0.82	ENSG00000136874	0.75	0.89	0.85	0.84	0.22	0.24	0.01	0.16
X:135545422-135556300:+	INTS6L	-0.32	-0.35	0.03	0.14	0	0.93	ENSG00000165359	0.99	0.99	0.99	0.99	0.24	0.23	0.19	0.22
X:22212903-22245409:+	PHEX	11	0.63	10.38	0.00E+00	0.07	0.00E+00	ENSG00000102174	0.99	0.97	0.65	0.98	0.21	0.17	0.33	0.24
X:65075473-65113813:+	ZC3H12B	0.25	0.47	-0.21	0.46	0	0.5	ENSG00000102053	1	1	0.99	1	0.25	0.23	0.3	0.26
X:85067126-85074391:+	APOOL	-0.98	-0.16	-0.82	0	0.62	0.01	ENSG00000155008	0.97	0.97	0.97	0.97	0.27	0.2	0.17	0.21

In total, 22 of these circRNAs exhibit differential relative expression in the CLL condition when compared to normal B cells. Additionally, 32 circRNAs demonstrate differential relative expression in CLL cases harboring the t(14;19) translocation relative to normal B cells. Moreover, 24 circRNAs exhibit differential relative expression in CLL cases carrying the t(14;19) translocation compared to those with the canonical CLL profile.

Furthermore, it is noteworthy that these circRNAs also demonstrated high absolute circRNA expression, consistently residing in the upper quantile of expression levels across our samples. This observation is particularly important because absolute circRNA expression levels provide insights into the number of molecules per cell. The high absolute expression suggests a substantial presence of these circRNAs within cells, which is crucial for their functional roles.

Achieving stoichiometric proportions in interactions with other RNAs or proteins can be critical for their proper functioning. Therefore, the combination of differential relative expression and high absolute expression highlights the significance of these circRNAs in potential interactions and functional roles within the cellular environment.

Notably, 50 out of these 52 circRNAs (96.15%) originate from exonic regions, while the remaining 2 (located at genomic positions 8:134812609-134813169 and 17:3705080-3705645) emerge from intergenic regions. Significantly, these intergenic regions have yet to be annotated, rendering them particularly intriguing subjects for prospective experimental investigations.

Out of these circRNAs, 46 exhibit an absolute Log Fold Change (LFC) of at least 0.5 in a minimum of one of the comparisons. This characteristic makes them particularly intriguing for functional annotation, owing to their high level of dysregulation (Figure 26).

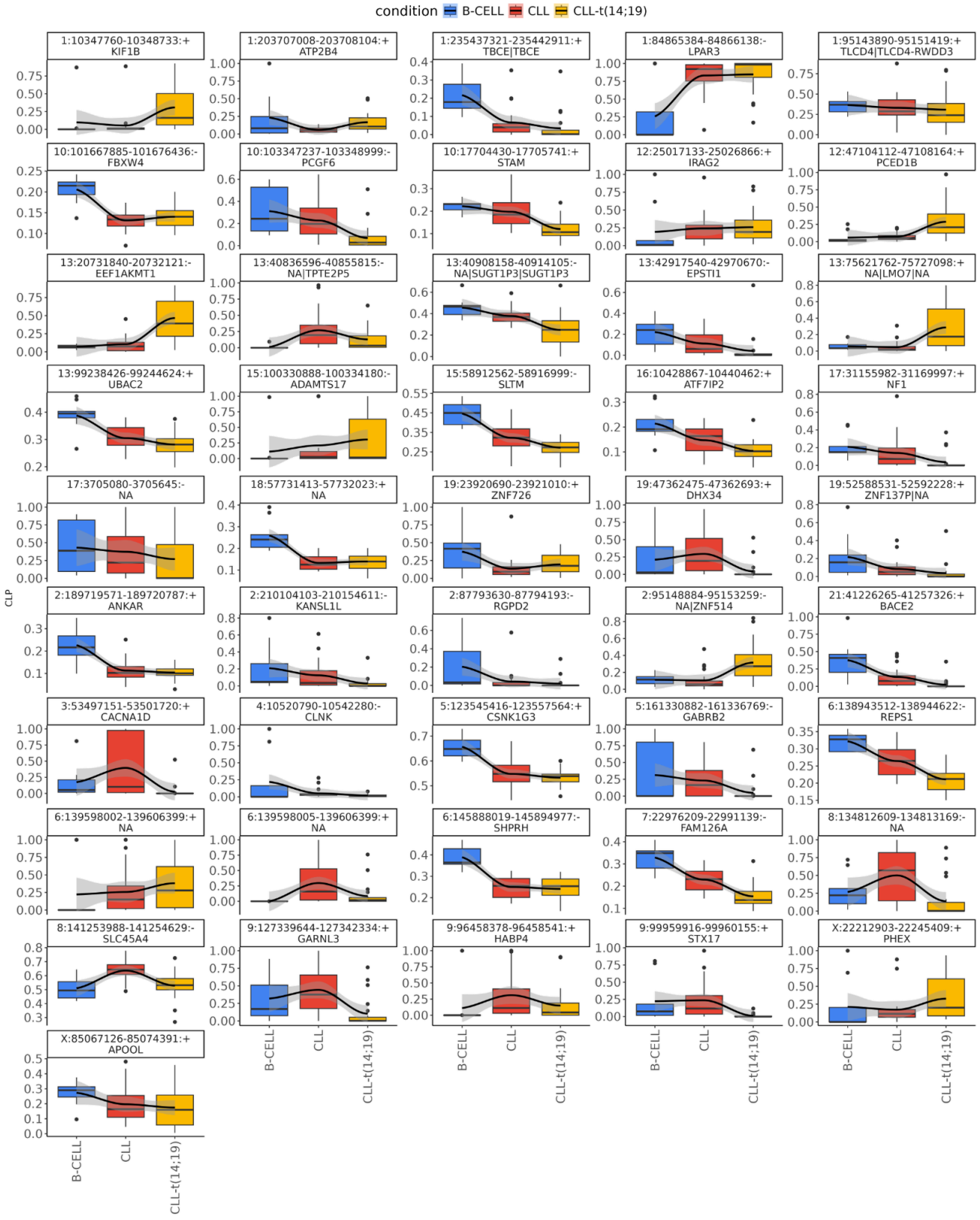

Figure 26: Relative Expression plots of 46 circRNAs with a minimum average CLP of 0.20 in at least one of the experimental conditions and a minimum absolute LFC of 0.5 in at least one of the comparisons.

Circular RNAs circIRAG2_12:25017133-25026866:+, circADAMTS17_15:100330888-100334180:-, circLPAR3_1:84865384-84866138:-, circGENE_6:139598002-139606399:+, circHABP4_9:96458378-96458541:+, and circPHEX_X:22212903-22245409:+ exhibit upregulation in both canonical CLL when compared to normal B cells and in CLL cases bearing the translocation compared to the normal condition (Table 3).

However, the circRNA at the locus 17:3705080-3705645:- shows downregulation specifically in CLL cases with the translocation compared to canonical CLL but shows no dysregulation in comparison to normal B cells (Table 3).

Conversely, circKIF1B_1:10347760-10348733:+ demonstrates upregulation in CLL cases with the translocation when compared to normal B cells and in comparison to canonical CLL (Table 3).

As previously mentioned, the circRNA located at 17:3705080-3705645 remains unannotated. Similarly, the circRNA at circGene_6:139598002-139606399, linked to gene ID ENSG00000226571 within an exonic region, lacks comprehensive annotations. Therefore, it would be intriguing to consider experimental validation to ascertain their existence, differential expression and functional significance.

Table 3: circRNAs with significant CLP variation among conditions exhibiting a CLP of at least 0.8 in at least one sample.

circRNA	**Gene Symbol**	**Ensembl Gene ID**	**Average Percentile**	**Average CLP**
1:10347760-10348733:+	KIF1B	ENSG00000054523	0.87	0.15
1:84865384-84866138:-	LPAR3	ENSG00000171517	0.98	0.65
12:25017133-25026866:+	IRAG2	ENSG00000118308	0.93	0.23
15:100330888-100334180:-	ADAMTS17	ENSG00000140470	0.74	0.21
17:3705080-3705645:-	.	.	0.88	0.36
6:139598002-139606399:+	.	ENSG00000226571	0.92	0.29
9:96458378-96458541:+	HABP4	ENSG00000130956	0.82	0.19
X:22212903-22245409:+	PHEX	ENSG00000102174	0.98	0.24

4.5.1 CircRNAs from genes previously associated with Chronic Lymphocytic Leukemia Pathogenesis

We have successfully pinpointed 21 circRNAs with significant CLP variation among conditions, which originate from genes previously linked to CLL in the scientific literature. Moreover, we have also calculated the average CLP and the percentile of absolute expression for these circRNAs (Table 4).

Table 4: Circular RNAs with differential relative expression derived from CLL-associated Genes

circRNA	Gene Symbol	LFC CLL-t(14;19) vs B-CELL	LFC CLL-t(14;19) vs CLL	LFC CLL vs B-CELL	Padj CLL-t(14;19) vs B-CELL	Padj CLL-t(14;19) vs CLL	Padj CLL vs B-CELL	Average CLP B-CELL	Average CLP CLL	Average CLP CLL-t(14;19)	Average CLP	Percentile CLL-t(14;19)	Percentile CLL	Percentile B-CELL	Average Percentile
11:108330213-108332037:+	ATM	-3.47	-5.91	2.44	0.17	0.00E+00	0.32	0.01	0.07	0.03	0.04	0.68	0.82	0.69	0.76
18:55254496-55279656:-	TCF4	-0.3	-4.2	3.91	0.93	0.00E+00	0.03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.57	0.82	0.8	0.75
18:55350358-55351000:-	TCF4	-0.14	-4.85	4.71	0.97	0.00E+00	0.01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.65	0.88	0.64	0.79
18:55350358-55403518:-	TCF4	2.52	-3.66	6.19	0.21	0.00E+00	0.00E+00	0.00E+00	0.00E+00		0.00E+00	0.74	0.89	0.53	0.82
20:36904156-36912552:-	SAMHD1	6.08	0.68	5.4	0.00E+00	0.75	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.8	0.82	0.25	0.78
2:61522610-61533903:-	XPO1	-0.71	0.03	-0.74	0.00E+00	0.91	0.00E+00	0.11	0.06	0.06	0.08	1	0.99	0.99	0.99
3:142560262-142568154:-	ATR	-3.52	-3.56	0.05	0.03	0.00E+00	0.99	0.02	0.00E+00	0.00E+00	0.00E+00	0.57	0.79	0.88	0.77
3:47046486-47067118:-	SETD2	6.67	-0.41	7.09	0.00E+00	0.78	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.83	0.8	0.78	0.82
4:10501255-10513597:-	CLNK	6.94	-1.67	8.61	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.85	0.91	0.54	0.87
4:10501255-10520831:-	CLNK	8.74	0.00E+00	8.73	0.00E+00	1	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.91	0.9	0.1	0.89
4:10501255-10542280:-	CLNK	8.04	0.03	8.01	0.00E+00	0.92	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.9	0.91	0.48	0.89
4:10507958-10542280:-	CLNK	8.39	-0.49	8.88	0.00E+00	0.51	0.00E+00	0.00E+00	0.02	0.00E+00	0.00E+00	0.81	0.89	0.31	0.84

4:10513463-10542280:-	CLNK	7.49	-1.68	9.17	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.00E-02	0.00E+00	0.00E+00	0.74	0.86	0.18	0.78
4:10520790-10542280:-	CLNK	9.71	0.27	9.44	0.00E+00	0.55	0.00E+00	0.22	0.05	0.02	0.1	0.96	0.97	0.67	0.96
4:10525840-10542280:-	CLNK	7.52	-1.12	8.64	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.83	0.89	0.26	0.84
4:10584926-10584955:-	CLNK	5.06	-2.49	7.54	0.12	0.29	0.00E+00	0.00E+00	0.03	0.00E+00	0.01	0.76	0.86	0.11	0.79
4:10584926-10667911:-	CLNK	7.13	-1.59	8.72	0.00E+00	0.29	0.00E+00	0.00E+00	0.02	0.00E+00	0.00E+00	0.83	0.91	0.33	0.85
4:108063622-108079614:-	LEF1	7.92	-0.18	8.1	0.00E+00	0.67	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.93	0.97	0.22	0.95
4:108072836-108079614:-	LEF1	7.82	-0.05	7.87	0.00E+00	0.95	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.89	0.93	0.21	0.9
4:108072850-108079614:-	LEF1	8.76	0.78	7.98	0.00E+00	0.08	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.9	0.91	0.46	0.89
5:158941693-158941979:-	EBF1	-7.44	1.46	-8.9	0.00E+00	0.15	0.00E+00	0.17	0.05	0.02	0.08	0.79	0.82	0.95	0.87

The CLP values exhibit a spectrum spanning from 0.001 to 0.096. Among them, circCLNK_4:10520790-10542280, circEBF1_5:158941693-158941979 and circXPO1_2:61522610-61533903 stand out with the highest CLP values compared to the rest. circXPO1_2:61522610-61533903 and circEBF1_5:158941693-158941979 demonstrate significant downregulation in both forms of CLL compared to the normal condition. Conversely, circCLNK_4:10520790-10542280 exhibits upregulation in both CLL forms compared to the normal condition. It is evident that they all exhibit a high level of absolute expression.

4.6 Functional Relevance

Two distinct groups of circular RNAs were selected for miRNA binding site predictions (Figure 27).

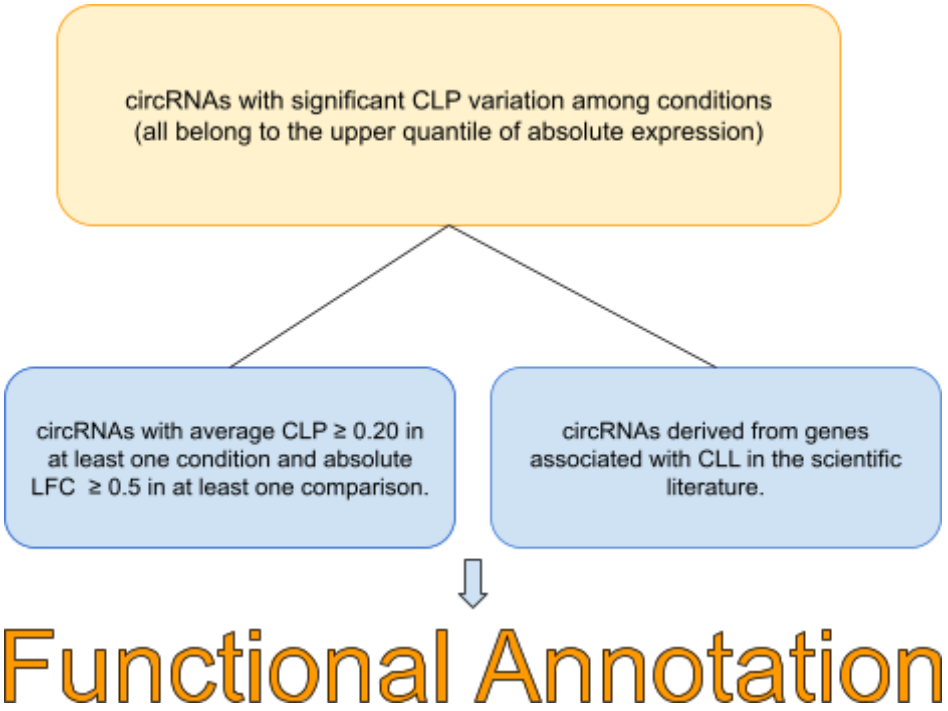

Figure 27: Diagram of Selected Circular RNAs for Functional Annotation. This diagram represents the selection and categorization process, providing a clear overview of the circular RNAs that were prioritized for further functional analysis based on their expression profiles and relevance to CLL.

The first group comprised of 46 circRNAs characterized by both high relative expression levels and significant dysregulation, with an average CLP of at least 0.20 in at least one condition and a Log Fold Change (LFC) exceeding 0.5 in at least one comparison. The second group consisted of 21 circRNAs that exhibited differential expression and were linked to CLL in existing scientific literature (regardless of their average CLP and LFC), thus highlighting their potential relevance to this specific medical context.

In the functional annotation of 66 unique circRNAs, potential binding sites were identified (Table 5), and 921 distinct miRNAs were predicted to interact with these circRNAs. Notably, some of these circRNAs were found to have multiple binding sites for the same miRNA (Table 8), suggesting a complex regulatory network where individual circRNAs can sequester and modulate the activity of specific miRNAs more effectively. Out of all these unique miRNAs, 90 have been identified in the OncomiRDB database as having roles in either tumor suppression or oncogenesis across 25 different tissues. Meanwhile, 31 of the predicted miRNAs are specifically associated with blood tissues [139] (Table 6 and Table 7).

Table 5: Number of potential binding sites for each circular RNA

Gene_circRNA	Binding Site Count
ADAMTS17_(15:100330888-100334180:-)	113
PCED1B_(12:47104112-47108164:+)	106
ZNF137P|NA_(19:52588531-52592228:+)	102
LEF1_(4:108072836-108079614:-)	98
LEF1_(4:108072850-108079614:-)	98
NA|ZNF514_(2:95148884-95153259:-)	63
SETD2_(3:47046486-47067118:-)	53
CLNK_(4:10501255-10542280:-)	48
TCF4_(18:55254496-55279656:-)	45
CLNK_(4:10507958-10542280:-)	43
FBXW4_(10:101667885-101676436:-)	40
NA_(18:57731413-57732023:+)	38
NA_(6:139598002-139606399:+)	37
BACE2_(21:41226265-41257326:+)	36
CLNK_(4:10513463-10542280:-)	36
ATF7IP2_(16:10428867-10440462:+)	35
ATR_(3:142560262-142568154:-)	31
DHX34_(19:47362475-47362693:+)	30
NA_(6:139598005-139606399:+)	30
CLNK_(4:10501255-10513597:-) CLNK_(4:10501255-10520831:-)	29
CACNA1D_(3:53497151-53501720:+)	28
KANSL1L_(2:210104103-210154611:-)	28
SLC45A4_(8:141253988-141254629:-)	28
LPAR3_(1:84865384-84866138:-)	26
NA_(17:3705080-3705645:-)	26
UBAC2_(13:99238426-99244624:+)	26
TBCE|TBCE_(1:235437321-235442911:+)	23
FAM126A_(7:22976209-22991139:-)	22
GABRB2_(5:161330882-161336769:-)	22
CLNK_(4:10520790-10542280:-)	21
EPSTI1_(13:42917540-42970670:-)	20
LEF1_(4:108063622-108079614:-)	20
NA_(8:134812609-134813169:-)	19
ATP2B4_(1:203707008-203708104:+)	18
CLNK_(4:10525840-10542280:-)	18
PHEX_(X:22212903-22245409:+)	17
SHPRH_(6:145888019-145894977:-)	17

ATM_(11:108330213-108332037:+)	16
HABP4_(9:96458378-96458541:+)	16
REPS1_(6:138943512-138944622:-)	16
TCF4_(18:55350358-55403518:-)	15
GARNL3_(9:127339644-127342334:+)	14
TCF4_(18:55350358-55351000:-)	13
NA|LMO7|NA_(13:75621762-75727098:+)	12
ZNF726_(19:23920690-23921010:+)	12
APOOL_(X:85067126-85074391:+)	11
TLCD4|TLCD4-RWDD3_(1:95143890-95151419:+)	11
IRAG2_(12:25017133-25026866:+)	9
EEF1AKMT1_(13:20731840-20732121:-)	8
KIF1B_(1:10347760-10348733:+)	8
SLTM_(15:58912562-58916999:-)	8
ANKAR_(2:189719571-189720787:+)	7
NF1_(17:31155982-31169997:+)	7
SAMHD1_(20:36904156-36912552:-)	7
XPO1_(2:61522610-61533903:-)	7
CLNK_(4:10584926-10667911:-)	6
EBF1_(5:158941693-158941979:-)	6
STAM_(10:17704430-17705741:+)	6
CSNK1G3_(5:123545416-123557564:+)	5
PCGF6_(10:103347237-103348999:-)	5
RGPD2_(2:87793630-87794193:-)	5
NA|TPTE2P5_(13:40836596-40855815:-)	4
STX17_(9:99959916-99960155:+)	4
CLNK_(4:10584926-10584955:-)	2
NA|SUGT1P3|SUGT1P3_(13:40908158-40914105:-)	2

Table 6: Thirty one predicted microRNAs associated with blood tissues sourced from OncomiRDB. The columns include the miRNA ID, predicted target genes, associated hematological disorders, regulatory mechanisms, functional roles, and any pertinent expression patterns or experimental evidence. This comprehensive overview provides insights into potential regulatory networks and biomarker candidates in hematological contexts.(NPA: Not Previously Annotated)

Pubmed	Tumor	Mirbase_r20	Direct target	Mirna regulator	Mirna function	Mirna expression	Clinical sample	Sum effect
23077663	acute myeloid leukemia	miR-370-3p	NF1	NPA	increase cell proliferation|increase colony formation	NPA	NPA	oncogenic
23059786	acute myeloid leukemia	miR-17-5p	CDKN1A|STAT3	HIF1A|MYC	promote HIF-1ŚÁ-induced differentiation	NPA	NPA	NPA
23059786	acute myeloid leukemia	miR-20a-5p	CDKN1A|STAT3	HIF1A|MYC	promote HIF-1ŚÁ-induced differentiation	NPA	NPA	NPA
22964640	acute myeloid leukemia	let-7c-5p	PBX2	NPA	promote granulocytic differentiation	down-regulated in PML/RARŚÁ-positive blasts from acute promyelocytic leukemia (APL) patients than normal promyelocytes	acute promyelocytic leukemia (APL) blasts	NPA
22723551	**chronic lymphocytic leukemia**	**miR-125b-5p**	**NPA**	**NPA**	**modulate glucose, glutathione, lipid, and glycerolipid metabolism**	**underexpressed in both aggressive and indolent CLL patients**	**NPA**	**NPA**
22610076	**chronic lymphocytic leukemia**	**miR-181b-5p**	**BCL2|MCL1|XIAP**	**NPA**	**increase apoptosis**	**frequently down-regulated in chronic lymphocytic leukemia cells|significantly lower in poor prognostic subgroups defined by unmutated immunoglobulin heavy chain variable status and p53 aberrations|negatively associated with shorter overall survival and treatment-free survival in CLL patients**	**six Chinese patients with chronic lymphocytic leukemia (CLL), and in peripheral B cells from pooled 30 healthy donors|primary CLL cells from 40 patients**	**tumor-suppressive**

22550173	diffuse large B-cell lymphoma	miR-125a-5p	TNFAIP3	NF-kB	promote NF-kB signaling	NPA	primary diffuse large B-cell lymphomas	NPA
22550173	diffuse large B-cell lymphoma	miR-125b-5p	TNFAIP3	NF-kB	promote NF-kB signaling	NPA	primary diffuse large B-cell lymphomas	NPA
22469780	B-cell acute lymphoblastic leukemia	miR-125b-5p	ARID3A	NPA	promote cell proliferation|inhibit apoptosis	overexpressed in B-ALL cases with the t(11;14)(q24;q32) translocation	NPA	oncogenic
22251480	cytogeneticallly abnormal acute myeloid leukemia	miR-181b-5p	NPA	NPA	promote apoptosis|reduce cell viability|inhibit cell proliferation|delay leukemogenesis	significantly associated with favorable overall survival	183 CA-AML patients|a validation set of 271 CA-AML patients	tumor-suppressive
22235305	cutaneous T-cell lymphoma	miR-122-5p	NPA	NPA	decrease chemotherapy-induced apoptosis|activate Akt signaling	NPA	NPA	NPA
22234685	**chronic lymphocytic leukemia**	**miR-650**	**CDK1|CDKN2A|EBF3**	**IGL(host gene)**	**influence cell proliferation**	**associated with a favorable CLL prognosis**	**chronic lymphocytic leukemia samples**	**NPA**
22139839	leukemic B-cells	miR-125b-5p	BCL2	CD40LG	increase cell proliferation	NPA	NPA	oncogenic
21976676	multiple myeloma	miR-34b-5p	NPA	promoter DNA hypermethylation	reduce cell proliferation|promote apoptosis	NPA	8 normal marrow controls, 8 MM cell lines, 95 diagnostic, and 23 relapsed/progressed MM	tumor-suppressive

							samples|12 MM patients with paired samples at diagnosis and relapse/progression	
21976676	multiple myeloma	miR-34c-5p	NPA	promoter DNA hypermethylation	reduce cell proliferation|promote apoptosis	NPA	8 normal marrow controls, 8 MM cell lines, 95 diagnostic, and 23 relapsed/progressed MM samples|12 MM patients with paired samples at diagnosis and relapse/progression	tumor-suppressive
21953646	Hodgkin's lymphoma	miR-17-5p	CDKN1A	NPA	promote cell cycle progression	NPA	NPA	oncogenic
21880154	pediatric acute promyelocytic leukemia	miR-125b-5p	BAK1	NPA	promote cell proliferation|inhibit apoptosis|increase drug resistance	highly expressed in pediatric APL compared with other subtypes of AML|correlated with treatment response, as well as relapse of pediatric APL|up-regulated in leukemic drug-resistant cells	169 pediatric acute myelogenous leukemia (AML) samples including 76 APL samples before therapy and 38 APL samples after therapy	oncogenic

21636858	**chronic lymphocytic leukemia**	**miR-181b-5p**	**MCL1**	**NPA**	**NPA**	**decreased in samples of patients with a progressive (P < .001, training and validation sets) but not in samples of patients with a stable disease (P = .3, training set; P = .2, validation set) over time**	**leukemic cells isolated from 358 sequential samples of 114 patients with either stable or progressive disease**	**NPA**
21569010	EVI1+ myeloid leukemia	miR-449a	NOTCH1|BCL2	EVI1	induce apoptosis|reduce cell viability	downregulated in MECOM rearranged leukemias	38 MECOM (MDS1 and EVI1 complex)-rearranged patient samples, normal bone marrow controls	tumor-suppressive
21460242	diffuse large B-cell lymphoma	miR-34a-5p	FOXP1	MYC	inhibit cell proliferation	NPA	NPA	tumor-suppressive
21399664	acute myeloid leukemia	miR-193a-3p	KIT	promoter DNA hypermethylation in cancer	inhibit cell growth|induce apoptosis|regulate granulocytic differentiation	epigenetically repressed by promoter hypermethylation in acute myeloid leukemia (AML) cell lines and primary AML blasts, but not in normal bone marrow cells|inversely correlated with KIT expression in leukemia cell lines and 27 primary AML samples	27 primary AML samples	tumor-suppressive
21297663	lymphoma	miR-34a-5p	MYC	NPA	induce apoptosis	NPA	NPA	tumor-suppressive
21205967	**B-cell chronic lymphocytic leukemia**	**miR-34b-5p**	**ZAP70**	**TP53**	**NPA**	**NPA**	**blood samples from untreated patients (n = 206)**	**NPA**

							diagnosed with B-cell CLL	
21205967	**B-cell chronic lymphocytic leukemia**	**miR-34c-5p**	**ZAP70**	**TP53**	**NPA**	**NPA**	**blood samples from untreated patients (n = 206) diagnosed with B-cell CLL**	**NPA**
21118985	B-cell acute lymphoblastic leukemia	miR-125b-5p	NPA	NPA	promote leukemia|promote cell proliferation	overexpressed in patients with myelodysplasia and acute myeloid leukemia up to 90-fold normal	patients with myelodysplasia and acute myeloid leukemia	oncogenic
20951946	multiple myeloma	miR-194-5p	MDM2|IGF1|IGF1R	TP53	reduce migration of plasma cells into bone marrow	NPA	NPA	tumor-suppressive
20889924	acute myeloid leukemia	miR-34a-5p	E2F3	CEBPA	inhibit cell proliferation|reprogram granulocytic differentiation	NPA	NPA	tumor-suppressive
20693279	chronic myelogenous leukemia	miR-181b-5p	MCL1	LYN	NPA	significantly reduced (?11- to 25-fold) in MYL-R cells	NPA	NPA
20651244	diffuse large B-cell lymphoma	let-7b-5p	PRDM1	NPA	NPA	overexpressed in DLBCL relative to normal GCB cells	a cohort of 25 primary DLBCL	NPA
20596961	acute myeloid leukemia	miR-181b-5p	MAP3K10	NPA	promote cell proliferation	highly expressed in acute myeloid leukemia (AML)	acute myeloid leukemia (AML) patients	oncogenic
20445018	myeloprolifera	miR-28-5p	MPL|E	NPA	inhibit terminal differentiation	overexpressed in platelets of a fraction of MPN patients and	MPN patients	oncogenic

	tive neoplasm		2F6			expressed at constant low levels in platelets from healthy subjects	and healthy subjects	
19826043	acute myeloid leukemia	miR-24-3p	DUSP16	RUNX1|AML1-ETO	promote cell growth|promote cell proliferation|block granulocytic differentiation	NPA	NPA	oncogenic
19692702	**chronic lymphocytic leukemia**	**miR-107**	**PLAG1**	**NPA**	**NPA**	**downregulated in primary cells of CLL patients compared with peripheral B cells of healthy donors**	**primary cells of 50 treatment-naive CLL patients and peripheral B cells of 14 healthy donors**	**tumor-suppressive**
19692702	**chronic lymphocytic leukemia**	**miR-181b-5p**	**PLAG1**	**NPA**	**NPA**	**downregulated in primary cells of CLL patients compared with peripheral B cells of healthy donors**	**primary cells of 50 treatment-naive CLL patients and peripheral B cells of 14 healthy donors**	**tumor-suppressive**
19258499	acute myeloid leukemia	miR-34b-5p	CREB1	NPA	inhibit cell cycle progression|reduce anchorage-independent cell growth	expressed significantly less in myeloid cell lines	a cohort of 78 pediatric patients at diagnosis of acute myeloid leukemia	tumor-suppressive
19151778	acute promyelocytic leukemia	miR-342-3p	NPA	SPI1|IRF1|IRF9	promote ATRA-induced cell differentiation	NPA	NPA	NPA
18936236	acute myeloid leukemia	miR-125b-5p	NPA	t(2;11)(p21;q23	inhibit terminal differentiation	NPA	NPA	NPA

				) translocation				
18728182	multiple myeloma	miR-181b-5p	KAT2B	NPA	reduce tumor growth	upregulated in MMs and MGUS compared to healthy plasma cells	CD138+ bone marrow PCs from subjects with MM (n = 16), monoclonal gammopathy of undetermined significance (MGUS) (n = 6), and normal donors (n = 6)	tumor-suppressive
18728182	multiple myeloma	miR-19b-3p	SOCS1	NPA	reduce tumor growth	upregulated in MMs but not in MGUS compared to healthy plasma cells	CD138+ bone marrow PCs from subjects with MM (n = 16), monoclonal gammopathy of undetermined significance (MGUS) (n = 6), and normal donors (n = 6)	tumor-suppressive
17942906	Burkitt lymphoma	let-7a-5p	NPA	MYC	inhibit cell proliferation	NPA	NPA	tumor-suppressive
17260024	acute promyelocytic leukemia	miR-107	NFIA	all-trans-retinoic acid	NPA	upregulated in acute promyelocytic leukemia patients	acute promyelocytic leukemia patients	NPA

Table 7: Twentysix circRNAs with potential interaction with the 31 miRNAs. The table includes gene symbols and IDs for the circRNAs, along with their average circular-to-linear proportion within each condition and across all conditions. Additionally, it provides the absolute expression percentile for each circRNA within individual conditions and their average across all conditions. Moreover, it presents the log fold change of circRNA dysregulation in each comparison, along with corresponding adjusted p-values, and identifies the blood-associated miRNAs with which they are predicted to interact.

CircRNA	Gene Symbol	ENSEMBL ID	Average CLP B-CELL	Average CLP CLL	Average CLP CLL-t(14;19)	Average CLP	Percentile CLL-t(14;19)	Percentile CLL	Percentile B-CELL	Average Percentile	LFC CLL-t(14;19) vs B-CELL	LFC CLL-t(14;19) vs CLL	LFC CLL vs B-CELL	Padj CLL-t(14;19) vs B-CELL	Padj CLL-t(14;19) vs CLL	Padj CLL vs B-CELL	miRNA
1:10347760-10348733:+	KIF1B	ENSG00000054523	0.1	0.05	0.31	0.15	0.91	0.81	0.36	0.87	4.12	2.74	1.38	5.30E-03	4.87E-03	4.20E-01	miR-370-3p
1:84865384-84866138:-	LPAR3	ENSG00000171517	0.26	0.83	0.85	0.65	0.99	0.97	0.32	0.98	15.21	0.2	15.01	1.02E-08	0.94	7.05E-10	miR-24-3p
1:84865384-84866138:-	LPAR3	ENSG00000171517	0.26	0.83	0.85	0.65	0.99	0.97	0.32	0.98	15.21	0.2	15.01	1.02E-08	0.94	7.05E-10	miR-370-3p
1:84865384-84866138:-	LPAR3	ENSG00000171517	0.26	0.83	0.85	0.65	0.99	0.97	0.32	0.98	15.21	0.2	15.01	1.02E-08	0.94	7.05E-10	miR-24-3p
1:84865384-84866138:-	LPAR3	ENSG00000171517	0.26	0.83	0.85	0.65	0.99	0.97	0.32	0.98	15.21	0.2	15.01	1.02E-08	9.40E-01	7.05E-10	miR-370-3p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7b-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7c-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7a-5p

1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7b-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7c-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7a-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7b-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	0.97	0.99	-0.89	-0.41	-0.48	6.17E-03	8.00E-02	0.12	let-7c-5p
1:95143890-95151419:+	TLCD4|TLCD4-RWDD3	ENSG00000152078|ENSG00000271092	0.36	0.33	0.31	0.33	0.98	0.99	9.70E-01	0.99	-0.89	-0.41	-0.48	6.17E-03	0.08	1.20E-01	let-7a-5p
11:108330213-108332037:+	ATM	ENSG00000149311	0.01	0.07	0.03	0.04	0.68	0.82	0.69	0.76	-3.47	-5.91	2.44	1.70E-01	4.06E-05	0.32	miR-145-5p
12:47104112-47108164:+	PCED1B	ENSG00000179715	0.06	0.07	0.29	0.14	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-650
12:47104112-47108164:+	PCED1B	ENSG00000179715	6.00E-02	7.00E-02	2.90E-01	1.40E-01	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-370-3p
12:47104112-47108164:+	PCED1B	ENSG00000179715	6.00E-02	7.00E-02	2.90E-01	1.40E-01	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-650

12:47104112-47108164:+	PCED1B	ENSG00000179715	6.00E-02	7.00E-02	2.90E-01	1.40E-01	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-650
12:47104112-47108164:+	PCED1B	ENSG00000179715	6.00E-02	7.00E-02	2.90E-01	1.40E-01	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-370-3p
12:47104112-47108164:+	PCED1B	ENSG00000179715	6.00E-02	7.00E-02	2.90E-01	1.40E-01	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-650
12:47104112-47108164:+	PCED1B	ENSG00000179715	6.00E-02	7.00E-02	2.90E-01	1.40E-01	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	0.48	miR-650
12:47104112-47108164:+	PCED1B	ENSG00000179715	0.06	0.07	0.29	0.14	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	4.80E-01	miR-370-3p
12:47104112-47108164:+	PCED1B	ENSG00000179715	0.06	0.07	0.29	0.14	0.97	0.94	0.81	0.96	1.57	1.05	0.52	0.01	3.72E-03	4.80E-01	miR-650
13:42917540-42970670:-	EPSTI1	ENSG00000133106	0.22	0.11	0.04	0.12	0.86	0.88	0.9	0.88	-2.81	-1.98	-0.83	2.71E-03	3.07E-03	0.39	miR-196b-5p
13:42917540-42970670:-	EPSTI1	ENSG00000133106	0.22	0.11	0.04	0.12	0.86	0.88	0.9	0.88	-2.81	-1.98	-0.83	2.71E-03	3.07E-03	0.39	miR-196a-5p
13:42917540-42970670:-	EPSTI1	ENSG00000133106	0.22	0.11	0.04	0.12	0.86	0.88	0.9	0.88	-2.81	-1.98	-0.83	2.71E-03	3.07E-03	0.39	miR-196b-5p
13:42917540-42970670:-	EPSTI1	ENSG00000133106	0.22	0.11	0.04	0.12	0.86	0.88	0.9	0.88	-2.81	-1.98	-0.83	2.71E-03	3.07E-03	0.39	miR-196a-5p
13:75621762-75727098:+	NA|LMO7|NA	ENSG00000261553|ENSG00000136153|ENSG00000228444	0.06	0.05	0.29	0.13	0.95	0.93	0.91	0.94	1.44	1.85	-0.41	0.08	4.06E-04	0.67	miR-125a-5p
13:75621762-75727098:+	NA|LMO7|NA	ENSG00000261553|ENSG00000136153|ENSG00000228444	0.06	0.05	0.29	0.13	0.95	0.93	0.91	0.94	1.44	1.85	-0.41	0.08	4.06E-04	0.67	miR-125b-5p

13:75621762-75727098:+	NA|LMO7|NA	ENSG00000261553|ENSG00000136153|ENSG00000228444	0.06	0.05	0.29	0.13	0.95	0.93	0.91	0.94	1.44	1.85	-0.41	0.08	4.06E-04	0.67	miR-125a-5p
13:75621762-75727098:+	NA|LMO7|NA	ENSG00000261553|ENSG00000136153|ENSG00000228444	0.06	0.05	0.29	0.13	0.95	0.93	0.91	0.94	1.44	1.85	-0.41	8.00E-02	4.06E-04	0.67	miR-125b-5p
13:99238426-99244624:+	UBAC2	ENSG00000134882	0.39	0.3	0.28	0.32	1	1	1	1	-0.54	-0.14	-0.4	3.37E-03	3.40E-01	0.02	miR-34a-5p
15:100330888-100334180:-	ADAMTS17	ENSG00000140470	0.11	0.21	0.3	0.21	0.7	0.84	2.10E-04	0.74	11.92	1.32	10.6	1.31E-03	7.30E-01	1.09E-03	miR-650
17:31155982-31169997:+	NF1	ENSG00000196712	0.21	1.40E-01	4.00E-02	1.30E-01	0.7	0.87	0.92	0.85	-8.86	-8.25	-0.6	9.81E-09	1.05E-11	0.75	miR-9-5p
17:3705080-3705645:-			0.43	3.70E-01	2.70E-01	3.60E-01	0.64	0.83	0.97	0.88	-5.42	-5.71	0.29	0.1	8.65E-03	9.50E-01	miR-342-3p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	9.30E-01	1.15E-03	3.00E-02	miR-34c-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	9.30E-01	1.15E-03	3.00E-02	miR-122-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	9.30E-01	1.15E-03	3.00E-02	miR-125b-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	9.30E-01	1.15E-03	3.00E-02	miR-28-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	9.30E-01	1.15E-03	3.00E-02	miR-125a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	3.00E-02	miR-34a-5p

18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	3.00E-02	miR-34c-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	9.30E-01	1.15E-03	0.03	miR-122-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	3.00E-02	miR-125b-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-28-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34c-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-122-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125b-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-28-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34a-5p

18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34c-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-122-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125b-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-28-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34c-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-122-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125b-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-28-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34a-5p

18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34c-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-122-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125b-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-28-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-125a-5p
18:55254496-55279656:-	TCF4	ENSG00000196628	3.12E-03	2.20E-03	3.74E-04	1.90E-03	0.57	0.82	0.8	0.75	-0.3	-4.2	3.91	0.93	1.15E-03	0.03	miR-34a-5p
18:57731413-57732023:+	NA	ENSG00000267787	0.26	0.13	0.14	0.18	0.99	1	1	1	-0.73	0.08	-0.81	0.03	0.8	3.76E-03	miR-196b-5p
18:57731413-57732023:+	NA	ENSG00000267787	0.26	0.13	0.14	0.18	0.99	1	1	1	-0.73	0.08	-0.81	0.03	0.8	3.76E-03	miR-196a-5p
18:57731413-57732023:+	NA	ENSG00000267787	0.26	0.13	0.14	0.18	0.99	1	1	1	-0.73	0.08	-0.81	0.03	0.8	3.76E-03	miR-196b-5p
18:57731413-57732023:+	NA	ENSG00000267787	0.26	0.13	0.14	0.18	0.99	1	1	1	-0.73	0.08	-0.81	0.03	0.8	3.76E-03	miR-196a-5p
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-29b-3p
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-29a-3p

19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-107
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-29b-3p
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-29a-3p
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-107
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-29b-3p
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-29a-3p
19:47362475-47362693:+	DHX34	ENSG00000134815	0.21	0.29	0.04	0.18	0.71	0.83	0.84	0.8	-6.85	-8.1	1.25	6.65E-03	4.65E-07	0.7	miR-107
19:52588531-52592228:+	ZNF137P|NA	ENSG00000123870|ENSG00000290721	0.22	0.08	0.04	0.11	0.73	0.86	0.91	0.84	-7.52	-6.31	-1.21	1.15E-04	4.70E-06	0.6	miR-128-3p
2:210104103-210154611:-	KANSL1L	ENSG00000144445	0.21	0.12	0.03	0.12	0.79	0.91	0.94	0.89	-4.28	-3.94	-0.35	0.05	6.46E-03	0.91	miR-9-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-181b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-199b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193a-3p

2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193b-3p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-181b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-199b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193a-3p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193b-3p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-181b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-199b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193a-3p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193b-3p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-181b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-199b-5p
2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193a-3p

2:95148884-95153259:-	NA|ZNF514	ENSG00000289685|ENSG00000144026	0.11	0.11	0.31	0.18	0.95	0.89	0.93	0.93	0.79	1.4	-0.61	0.29	7.31E-04	0.38	miR-193b-3p
21:41226265-41257326:+	BACE2	ENSG00000182240	0.38	0.14	0.02	0.18	0.82	0.95	0.98	0.95	-8.91	-7.1	-1.81	7.89E-07	3.72E-08	0.33	miR-199b-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-20a-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-17-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-193a-3p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-20a-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-17-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-193a-3p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-20a-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-17-5p
3:142560262-142568154:-	ATR	ENSG00000175054	0.02	7.57E-03	2.49E-03	8.91E-03	0.57	0.79	0.88	0.77	-3.52	-3.56	0.05	0.03	3.22E-04	0.99	miR-193a-3p
3:47046486-47067118:-	SETD2	ENSG00000181555	2.78E-03	6.41E-03	4.49E-03	4.56E-03	0.83	0.8	0.78	0.82	6.67	-0.41	7.09	1.48E-05	0.78	5.04E-07	miR-28-5p

4:108063622-108079614:-	LEF1	ENSG00000138795	1.00E-06	3.50E-03	2.50E-03	2.00E-03	0.93	0.97	0.22	0.95	7.92	-0.18	8.1	4.01E-23	0.67	2.80E-25	miR-34b-5p
5:123545416-123557564:+	CSNK1G3	ENSG00000151292	0.66	0.55	0.53	0.58	1	1	1	1	-0.53	-0.07	-0.46	2.99E-03	0.67	4.07E-03	miR-181b-5p
5:158941693-158941979:-	EBF1	ENSG00000164330	0.17	0.05	0.02	0.08	0.79	0.82	0.95	0.87	-7.44	1.46	-8.9	5.03E-08	0.15	1.80E-11	miR-370-3p
6:145888019-145894977:-	SHPRH	ENSG00000146414	0.39	0.25	0.24	0.29	0.99	0.99	0.99	0.99	-0.64	0.02	-0.66	0.02	0.94	4.25E-03	miR-15a-5p
6:145888019-145894977:-	SHPRH	ENSG00000146414	0.39	0.25	0.24	0.29	0.99	0.99	0.99	0.99	-0.64	0.02	-0.66	0.02	0.94	4.25E-03	miR-19b-3p
6:145888019-145894977:-	SHPRH	ENSG00000146414	0.39	0.25	0.24	0.29	0.99	0.99	0.99	0.99	-0.64	0.02	-0.66	0.02	0.94	4.25E-03	miR-15a-5p
6:145888019-145894977:-	SHPRH	ENSG00000146414	0.39	0.25	0.24	0.29	0.99	0.99	0.99	0.99	-0.64	0.02	-0.66	0.02	0.94	4.25E-03	miR-19b-3p
7:22976209-22991139:-	FAM126A	ENSG00000122591	0.33	0.23	0.15	0.24	0.97	0.97	0.99	0.98	-1.12	-0.61	-0.51	7.97E-05	1.26E-03	0.06	miR-107
8:141253988-141254629:-	SLC45A4	ENSG00000022567	0.51	0.64	0.53	0.56	0.99	0.99	0.99	0.99	0.18	-0.4	0.58	0.53	3.86E-03	2.28E-03	miR-194-5p

Table 8: Twentytwo circular RNAs that have more than one binding site for the same miRNA.

Gene-circRNA	miRNA	Binding Site Count
ZNF137P|NA_(19:52588531-52592228:+)	hsa-miR-12114	2
LEF1_(4:108072836-108079614:-)	hsa-miR-1229-3p	2
LEF1_(4:108072850-108079614:-)	hsa-miR-1229-3p	2
NA|ZNF514_(2:95148884-95153259:-)	hsa-miR-3085-3p	2
NA|ZNF514_(2:95148884-95153259:-)	hsa-miR-3925-3p	2
PCED1B_(12:47104112-47108164:+)	hsa-miR-412-3p	2
PCED1B_(12:47104112-47108164:+)	hsa-miR-4695-5p	2
ADAMTS17_(15:100330888-100334180:-)	hsa-miR-550a-5p	2
PCED1B_(12:47104112-47108164:+)	hsa-miR-650	2
DHX34_(19:47362475-47362693:+)	hsa-miR-6735-3p	2
LEF1_(4:108072836-108079614:-)	hsa-miR-6735-5p	2
LEF1_(4:108072850-108079614:-)	hsa-miR-6735-5p	2
CLNK_(4:10501255-10513597:-)	hsa-miR-6780a-3p	2
CLNK_(4:10501255-10520831:-)	hsa-miR-6780a-3p	2
CLNK_(4:10501255-10542280:-)	hsa-miR-6780a-3p	2
ZNF137P|NA_(19:52588531-52592228:+)	hsa-miR-6780a-5p	2
SETD2_(3:47046486-47067118:-)	hsa-miR-6781-5p	2
ZNF137P|NA_(19:52588531-52592228:+)	hsa-miR-6851-5p	2
ZNF137P|NA_(19:52588531-52592228:+)	hsa-miR-6880-5p	2
NA_(8:134812609-134813169:-)	hsa-miR-7157-3p	2
ZNF137P|NA_(19:52588531-52592228:+)	hsa-miR-7160-5p	2
ATF7IP2_(16:10428867-10440462:+)	hsa-miR-7703	2

4.7 Identification of Mitochondrial circular RNA (mecciRNAs)

Our analysis revealed the expression of nine mitochondrial-encoded circular RNAs (mecciRNAs) in all sample groups. Specifically, three mecciRNAs were identified in normal B-cell samples, while seven mecciRNAs were found in canonical CLL samples and seven in CLL samples bearing the translocation (Figure 28 and Table 9). This result attracted our attention due to recent breakthrough publications about mecciRNAs discovery and their role in cancer.

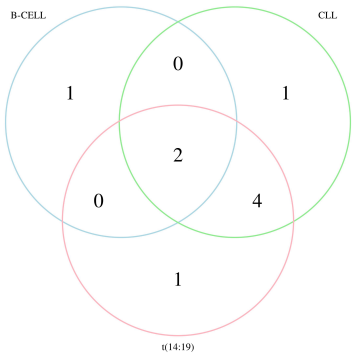

Figure 28: Venn Diagram Illustrating Common mecciRNAs Across Conditions.

Table 9: circRNAs derived from the mitochondrial genome in all conditions. We observe the average CPM per condition.

mecciRNA ID	Expression in BCELL Condition	Expression in CLL Condition	Expression in t(14;19)-CLL Condition	Gene Name	Ensembl Gene ID
MT:12548-13383:-	11.28	67.56	115.20	.	.
MT:13652-14054:-	0	0	6.37	.	.
MT:13677-14412:-	2.81	0	0	MT-ND6	ENSG00000198695
MT:13977-14420:-	0	2.41	13.82	MT-ND6	ENSG00000198695
MT:13998-14412:-	2.80	18.37	27.89	MT-ND6	ENSG00000198695
MT:15381-15539:-	0	2.40	5.33	.	.
MT:15381-15852:-	0	10.87	17.16	.	.
MT:3321-3828:-	0	1.20	12.83	.	.
MT:3566-3789:-	0	3.60	0	.	.

The mecciRNAs MT:12548-13383:-, MT:13652-14054:- are derived from the antisense strand (also known as the light strand) of the MT-ND5 gene. Similarly, MT:15381-15539:- and MT:15381-15852:- originate from the antisense strand of the MT-CYB gene. Additionally, MT:3321-3828:- and MT:3566-3789:- are derived from the antisense strand of the MT-ND1 gene, all belonging to intergenic regions of the mitochondrial genome. Meanwhile, MT:13977-14420:- and MT:13998-14412:- and MT:13677-14412:- are annotated to the MT-ND6 gene (refer to Figure 29 to view the coordinates) that encodes NADH dehydrogenase 6, a component of the large enzyme complex called complex I playing a crucial role in oxidative phosphorylation.

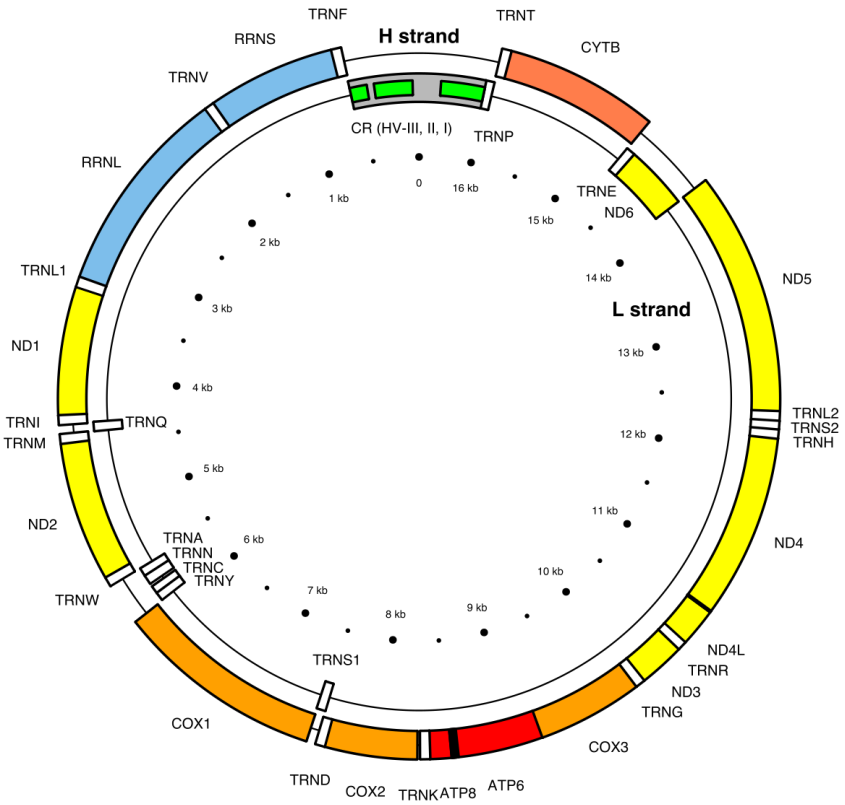

Figure 29: a comprehensive view of the mitochondrial genome. This diagram shows the genomic locations of the nine detected mecciRNAs in Table 9.

We have discovered the presence of two anti-sense MT-ND5, two anti-sense MT-ND1 and two anti-sense MT-CYB mecciRNAs (Table 9). The discovery of anti-sense mecciRNAs suggests that these regulatory molecules may be involved in modulating the expression of these genes, potentially affecting the synthesis of the corresponding proteins (NADH dehydrogenase 5 and NADH dehydrogenase 1) and thus impacting mitochondrial function and cellular energy metabolism.

Discussion

In this thesis, we employed Circompara2 to analyze RNA-seq data from 64 samples for the purpose of identification and quantification of circular RNAs. Subsequently, we conducted a quality control step to retain only those samples with sufficient quality, sequencing depth and those corresponding to the baseline and progression-before-therapy stages. Through this comparative analysis, we applied batch effect correction to the logit-transformed, SVA-corrected data, effectively reducing the influence of unrelated variations. As a result, we narrowed down our dataset to 54 samples and 6237 circular RNAs, each with a minimum of 2 read counts in at least 18 samples. This subset of circRNAs served as the foundation for our subsequent analyses, with the majority originating from exonic regions and a small number (44) deriving from intergenic regions.

Circular RNAs originating from unannotated genes have recently gained prominence for their enigmatic properties and emerging importance in biology. These circRNAs often arise from previously overlooked genomic regions, but advances in sequencing and bioinformatics have facilitated their discovery and characterization. Unannotated circRNAs exhibit diverse structures and are increasingly recognized for their regulatory roles, influencing processes like microRNA sequestration, RNA-binding protein interactions, and even epigenetic regulation. Their functional significance spans various biological processes, making them potential biomarkers and therapeutic targets. Conservation across species suggests fundamental roles. However, challenges in functional annotation persist, and future research will benefit from advanced techniques and expanding resources. Unannotated circRNAs hold promise in unraveling the intricacies of cellular biology, offering exciting prospects for future discoveries.

An essential aspect of our study involved calculating the circular-to-linear proportion of circRNAs. Intriguingly, we found that only 2.4 percent of the circRNAs exhibited an average CLP of 0.25 or higher across all samples. Furthermore, we identified a Spearman correlation of 0.51 between the average CLP of circRNAs and the log10 of their average absolute expression levels, suggesting a potential relationship between circRNA circularization and expression. One plausible explanation is that the circularization of circRNAs may play a pivotal role in regulating their stability and subcellular localization, thereby exerting influence over their expression levels [141]. This notion aligns with emerging research indicating that circRNAs can function as post-transcriptional regulators of gene expression [142].

Within the context of cancer, particularly in T-cell acute lymphoblastic leukemia (T-ALL), the dysregulation of CLP has become a subject of extensive investigation. Notably, one of the major factors contributing to these dysregulations is the aberrant splicing of RNA. The irregular CLP observed in cancer cells may not primarily result from a deliberate alteration in cellular behavior but rather stems from irregularities in the splicing process. Furthermore, numerous splicing factors, including QKI, are known to be dysregulated in cancer. By modulating the circular-to-linear ratio, cells might finely regulate the abundance of circRNAs to fulfill specific biological functions [143].

Then, we conducted a differential relative expression analysis with a specific focus on changes in the circular-to-linear proportion between different conditions. This approach allowed us to uncover potential imbalances present in leukemia cells, offering a novel perspective on the dynamics of gene expression.

Our analysis not only extended beyond the conventional evaluation of circRNA absolute counts but also provided a unique insight into the expression profiles of circRNAs. We observed distinctive relative expression patterns of circRNAs between the three conditions.

Intriguingly, our results revealed that 387 circRNAs exhibited differential relative expression among canonical CLL and normal conditions, while 332 circRNAs displayed such differences when

comparing CLL-t(14;19) to canonical CLL conditions. Additionally, we identified 461 circRNAs with altered relative expression in CLL-t(14;19) compared to normal conditions. These findings underscore the complexity and intricacies of circRNA expression in the context of leukemia, offering valuable insights into the molecular mechanisms at play.

Of note, 100 circRNAs were found to be DE in both the "CLL-t(14;19) vs. B-cell" and "CLL-t(14;19) vs. CLL" comparisons, and not in CLL vs. B-cell, resulting in specifically altered only in the aggressive t(14;19)-CLL and representing a potential signature of circRNA relative expression variation in link with this lesion.

The t(14;19) rearrangement instigates the overexpression of BCL3, a pivotal transcription factor, with possible implications for circRNA absolute expression. Notably, CLL patients harboring the translocation exhibit significant variations in the circular to linear proportion. These variations are ascribed predominantly to anomalies in RNA splicing, backsplicing regulation, and possibly controlled degradation of circRNAs, all intricately linked to the influence of BCL3 and its gene targets [144].

In the context of transcriptional regulation, BCL-3 functions as a transcriptional co-factor, facilitating interactions with other transcription factors to govern the expression of particular genes. Elevated levels of BCL3, characteristic of its overexpression, can enhance the transcription of specific genes, including those with roles in splicing. These genes often encode splicing factors or regulators that govern alternative splicing processes, potentially leading to alterations in the counts and splicing patterns of mRNA isoforms and circular RNAs [145].

BCL-3 is closely linked to the NF-κB signaling pathway, which, when activated due to BCL3 overexpression, can induce modifications in the expression of genes involved in splicing and RNA processing. These changes in gene expression can cascade down to affect the abundance and diversity of mRNA and circRNA species, thereby impacting the circRNA-to-linear RNA ratio [144,146].

Furthermore, BCL-3's interactions with RNA-binding proteins (RBPs) and other RNA processing factors can influence RNA stability and decay. CircRNAs are known for their enhanced stability compared to linear RNAs. BCL-3's influence on RNA stability might lead to differences in the degradation rates of circRNAs and linear RNAs, consequently altering their proportions over time [147].

Additionally, BCL-3 overexpression can affect the expression of miRNAs or other RNA molecules involved in regulating circRNAs and mRNAs. CircRNAs can act as miRNA sponges, sequestering miRNAs and thereby modulating mRNA expression. Changes in miRNA levels due to BCL-3 overexpression can disrupt these competing endogenous RNA (ceRNA) networks, influencing the circRNA-to-linear RNA ratio [144].

However, it is essential to consider that the impact of BCL-3 on circRNA-to-linear RNA ratios may also depend on the specific cellular context and the genes involved. Different cell types and conditions may respond differently to BCL-3 overexpression, leading to variable alterations in the proportion of circRNAs to linear RNAs. This intricate interplay highlights the multifaceted mechanisms through which BCL-3 can influence the dynamic balance between circRNAs and linear RNAs within the cell [144].

In summary, BCL-3's influence on transcriptional regulation, NF-κB signaling, and interactions with RNA processing factors can collectively impact the counts and splicing patterns of mRNA and circRNAs, adding a layer of complexity to the regulation of these genetic elements in response to BCL-3 overexpression [148,149].

It is important to note that within our analysis, certain findings may be grounded in speculations and hypotheses rather than established facts. Scientific research often involves a combination of

confirmed data and educated guesses, allowing for a more comprehensive understanding of complex phenomena.

There are several circRNAs that derive from the TCF4 gene. These circRNAs are noteworthy due to the TCF4 gene association with CLL. The TCF4 gene has been linked to CLL and is known to be overexpressed in CLL compared to normal conditions. What makes this transcription factor (TF) particularly intriguing is its role as a downstream target of the Wnt/TCF pathway. Furthermore, it is activated in human cancers characterized by β-catenin defects and plays a pivotal role in promoting neoplastic transformation [150].

Our analysis revealed the presence of three circular RNAs originating from the TCF4 gene, each exhibiting distinct expression patterns (circTCF4_18:55254496-55279656:-, circTCF4_18:55350358-55351000:- and circTCF4_18:55350358-55403518:-). Specifically, the circTCF4_18:55254496-55279656:- displayed downregulation in CLL-t(14;19) when compared to canonical CLL. On the other hand, circTCF4_18:55350358-55351000:- exhibited upregulation in canonical CLL in contrast to the normal condition, but it was downregulated in CLL-t(14;19) compared to canonical CLL. Lastly, circTCF4_18:55350358-55403518:- showed upregulation in canonical CLL compared to the normal condition, while it was downregulated in CLL-t(14;19) relative to canonical CLL. Hence, both circTCF4_18:55350358-55351000:- and circTCF4_18:55350358-55403518:- exhibit overexpression in canonical CLL.

These findings highlight that certain circRNAs have expression levels closely linked to the various conditions under investigation. The circRNAs common to multiple comparisons suggest potential shared regulatory mechanisms or biological significance across these diverse scenarios. On the other hand, the circRNAs that exhibit exclusive differential expression in one CLL group may fulfill distinct roles in the context of canonical CLL and the translocated CLL subtype.

Our next step involved the identification and prioritization of the most intriguing circRNAs. We focused on two distinct categories: firstly, circRNAs with a minimum average CLP of 0.20 in at least one of the experimental conditions and a minimum absolute LFC of 0.5 in at least one of the comparisons. Secondly, we also considered circRNAs deriving from genes that have well-established associations in the scientific literature with chronic lymphocytic leukemia. All these circRNAs also demonstrated high absolute circular expression, consistently residing in the upper quantile of expression levels across our samples. Eight of these circular RNAs showed a CLP of at least 0.8 in at least one sample (circKIF1B_1:10347760-10348733:+, circLPAR3_1:84865384-84866138:-, circIRAG2_12:25017133-25026866:+, circADAMTS17_15:100330888-100334180:-, circGene_17:3705080-3705645:-, circGene_6:139598002-139606399:+, circHABP4_9:96458378-96458541:+, circPHEX_X:22212903-22245409:+).

When circular RNA exhibits significantly higher abundance than its linear counterpart, it suggests that circRNAs may play pivotal roles in cellular processes, possibly owing to their remarkable stability, conferred by their closed-loop structures. Their abundance implies a regulatory significance in gene expression modulation, affecting transcription, splicing, and translation. Moreover, the tissue-specific nature of circRNA abundance underscores their potential for specialized functions in distinct physiological contexts. These molecules may arise from alternative splicing events, indicating their involvement in isoform regulation, potentially influencing protein diversity and function. Elevated circRNA levels might be linked to disease pathogenesis, offering potential biomarkers or therapeutic targets. Their high abundance likely underscores their fundamental roles in cellular homeostasis, potentially conserved across species, suggesting evolutionary importance. In essence, abundant circRNAs hold promise as key players in cellular regulation, meriting further investigation to elucidate their precise functions and implications in various biological systems.

To investigate the potential interaction between circRNAs and miRNA target sites, we conducted an extensive analysis employing four distinct computational tools: TargetScan, miRanda, PITA, and IntaRNA. Furthermore, we utilized mirBase and mirGeneDB as microRNA databases in this study. In the functional annotation of these 66 unique circRNAs, we predicted interactions with 921 unique miRNAs. Notably, 90 of these unique miRNAs have been previously identified in the OncomiRDB database for their roles in tumor suppression or oncogenesis across 25 different tissues. Additionally, 31 of the predicted miRNAs are specifically associated with blood tissues (Figure 30).

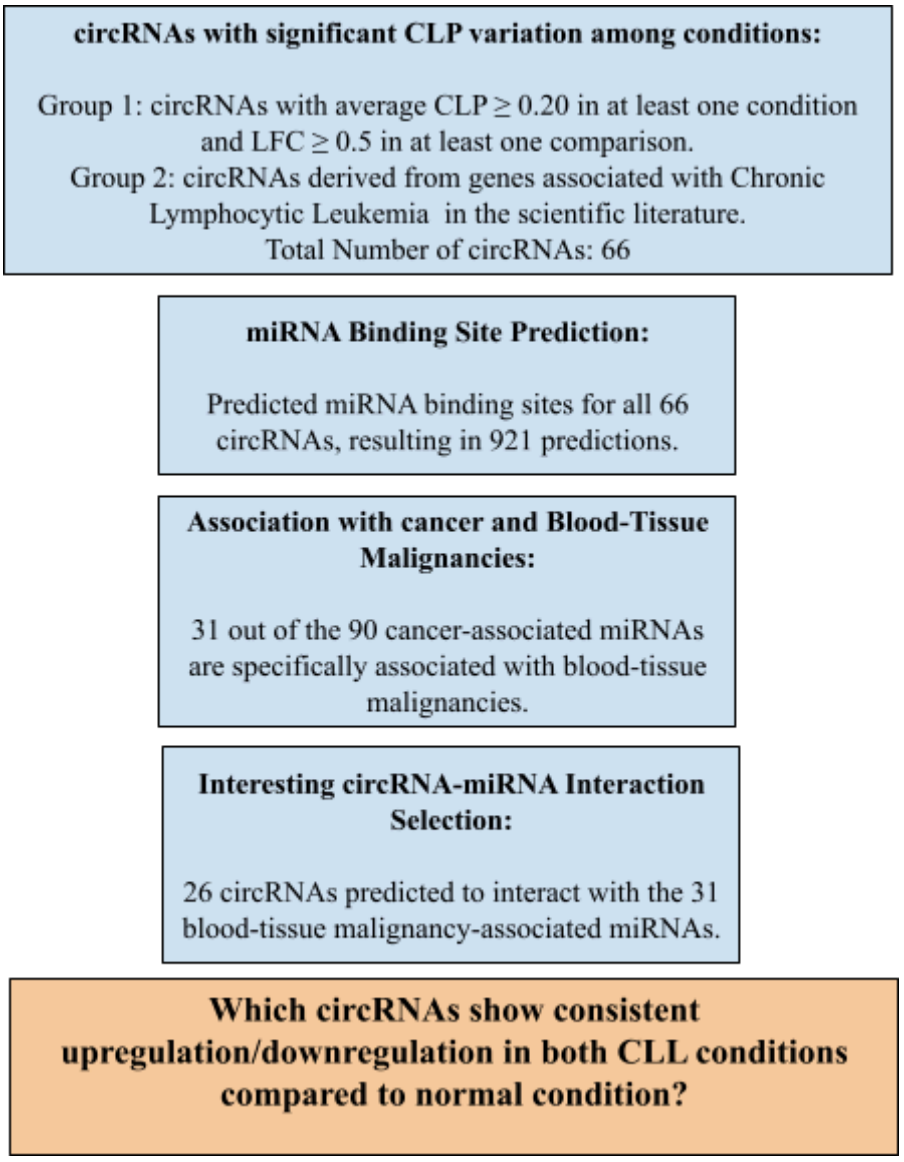

Figure 30: Identification of key circRNAs that could serve as biomarkers or therapeutic targets in blood-related cancers . The diagram visualizes the stepwise filtering and selection process, ultimately leading to a refined list of circRNAs with potential miRNA binding sites that are significantly involved in CLL and other blood malignancies.

The regulation of circRNAs, which are known to serve as miRNA sponges or competitive endogenous RNAs, plays a pivotal role in post-transcriptional gene expression control. The upregulation or downregulation of circRNAs can have biological implications by modulating the availability of miRNAs they interact with. In cases where circRNAs are downregulated, there is an increased risk of miRNA sequestration, thereby liberating miRNAs to target and suppress their cognate mRNA

transcripts. This can lead to a derepression of target genes that were previously under miRNA-mediated repression, potentially influencing critical cellular pathways and processes. Conversely, when circRNAs are upregulated, they can effectively absorb miRNAs, limiting their availability to interact with other mRNA targets. This can result in the stabilization of mRNAs that would have otherwise been subjected to miRNA-induced degradation, further emphasizing the regulatory impact of circRNAs on gene expression. Thus, the dysregulation of circRNAs can have a significant impact on the post-transcriptional control of gene expression and contribute to the development and progression of various biological and pathological conditions (Figure 30). Among the 26 circRNAs that were predicted to have interactions with miRNAs associated with blood tissues, the following circRNAs exhibit consistent downregulation in both forms of CLL when compared to the normal condition: circEBF1_5:158941693-158941979:-, circCSNK1G3_5:123545416-123557564:+. Conversely, circADAMTS17_15:100330888-100334180:-, circLPAR3_1:84865384-84866138:-,circSETD2_3:47046486-47067118:-, and circLEF1_4:108063622-108079614:- demonstrate significant upregulation in both forms of CLL compared to the normal condition.

Based on our predictions, circEBF1_5:158941693-158941979 (down-regulated in both forms of CLL compared to normal conditions) potentially interacts with six miRNAs, with miR-370-3p being one of them. According to a functional analysis conducted by Laura García-Ortí and her colleagues, miR-370-3p exhibits activity in blood tissue and is identified as a direct regulator of the NF1 gene. Notably, overexpressing miR-370-3p leads to effects similar to NF1 inactivation, resulting in increased proliferation and colony formation in acute myeloid leukemia (AML) cells. As a consequence of these findings, it is evident that miR-370-3p is associated with oncogenic effects in AML [151].

circGene_5:123545416-123557564:+ (down-regulated in both forms of CLL compared to normal conditions) potentially interacts with five miRNAs. Among these, miR-181b-5p stands out as particularly intriguing due to its recurrent associations in the scientific literature with various hematologic malignancies, including CLL, acute myeloid leukemia, multiple myeloma, and cytogenetically abnormal AML. MiR-181b-5p has been linked to the regulation of multiple genes, including BCL2, MCL1, XIAP, PLAG1, MAP3K10, and KAT2B. Notably, its expression profile varies in different contexts: in the context of multiple myeloma, it is upregulated in both multiple myeloma (MM) and monoclonal gammopathy of undetermined significance (MGUS) compared to healthy plasma cells. Conversely, in chronic lymphocytic leukemia, miR-181b-5p is frequently down-regulated, particularly in subgroups characterized by unmutated immunoglobulin heavy chain variable status and p53 aberrations. This downregulation is significantly associated with shorter overall survival and treatment-free survival in CLL patients (reported as a biomarker of disease progression in CLL by Visone and colleagues). Therefore, miR-181b-5p plays a multifaceted role in these hematologic malignancies, with its expression patterns and functions varying depending on the specific context [152–158].

circADAMTS17_15:100330888-100334180:- (up-regulated in both forms of CLL compared to normal condition) potentially interacts with 113 miRNAs, one of which is miR-650. Notably, miR-650 and its homologs share overlapping sequences with various variable (V) subgenes responsible for encoding lambda immunoglobulin (IgLλ). While the role of miR-650 has previously been explored in solid tumors, its specific function in CLL was the subject of a study conducted by Marek Mraz and colleagues. In this study, the authors reported that miR-650's expression is co-regulated with its host gene for IgLλ, revealing a unique and previously undiscovered mechanism of microRNA gene regulation. Moreover, higher levels of miR-650 expression are associated with a more favorable prognosis for CLL and exert influence over the proliferation of B cells. Furthermore, in B cells, miR-650 targets essential proteins involved in cell proliferation and survival, including

cyclin-dependent kinase 1 (CDK1), inhibitor of growth 4 (ING4), and early B-cell factor 3 (EBF3). This research underscores the substantial role played by miR-650 in both the biology of CLL and the normal physiology of B cells [159].

circLPAR3_1:84865384-84866138:- (up-regulated in both forms of CLL compared to normal condition) potentially interacts with 26 miRNAs, two of which are miR-24-3p and miR-370-3p. In a study conducted by Sayyed K Zaidi and colleagues, the authors reported that in acute myeloid leukemia, when the subnuclear localization of Runx1/AML1 is disrupted, it activates miR-24. MiR-24 is regulated by both Runx1 and the AML1-ETO protein encoded by a specific chromosomal translocation. MiR-24 downregulates MAPK phosphatase-7 and enhances the phosphorylation of key kinases. This, in turn, stimulates myeloid cell growth, reduces their dependence on interleukin-3, and blocks the differentiation process, showing the oncogenic nature of this miRNA in the context of leukemia [160].

circSETD2_3:47046486-47067118:- (up-regulated in both forms of CLL compared to normal condition) potentially interacts with 53 miRNAs, one of which is miR-28-5p. In a study conducted by Girardot and colleagues, it was found that miR-28-5p directly targets the MPL and E2F6 genes in the context of myeloproliferative neoplasms, resulting in the inhibition of terminal differentiation. Moreover, this miRNA was observed to be overexpressed in the platelets of some myeloproliferative neoplasm (MPN) patients while maintaining consistently low levels in platelets from healthy individuals, underscoring its oncogenic nature [161].

circLEF1_4:108063622-108079614:- (up-regulated in both forms of CLL compared to normal condition) potentially interacts with 20 miRNAs, including miR-34b-5p. In a study conducted by Kwan Yeung Wong and colleagues, it was revealed that miR-34b-5p, a component of the MIR34B/C cluster, plays a crucial role in myeloma. It is often inactivated by promoter hypermethylation in myeloma cells. When this inactivation is reversed, it can lead to reduced cell growth and increased cell death, making it a potential tumor suppressor. During the progression of myeloma, the inactivation of miR-34b-5p through hypermethylation becomes more common, suggesting that it may contribute to the development and worsening of myeloma [162].

Of particular interest is circGene_17:3705080-3705645:-, which represents an unannotated circRNA located within an intergenic region. It exhibits a high CLP and experiences profound downregulation in CLL bearing the translocation, compared to canonical CLL and normal conditions. This circRNA potentially interacts with 20 miRNAs, including miR-342-3p. In a study conducted by M L De Marchis and colleagues, it was reported that MiR-342 plays a crucial role in acute promyelocytic leukemia (APL) when treated with all-trans-retinoic acid (ATRA). ATRA induces the maturation and remission of leukemia in APL patients. MiR-342 is a microRNA that ATRA boosts during APL differentiation. It is regulated by key transcription factors like PU.1, interferon regulatory factor (IRF)-1, and IRF-9. While IRF-1 keeps miR-342 low, PU.1 and IRF-9 increase its expression during ATRA-induced differentiation. MiR-342's enforced expression in APL cells enhances their differentiation, contributing to the maturation and remission of leukemia in APL [163].

Furthermore, the last part of our analysis revealed the presence of nine mitochondrial-encoded circular RNAs across all sample groups. Specifically, we detected three distinct mecciRNAs in normal B-cell samples, while canonical CLL samples exhibited the presence of seven mecciRNAs, and CLL samples with the translocation also displayed seven such mecciRNAs.

It is worth mentioning that none of the circular RNAs referenced in Section 2.6 of the introduction were identified in our sample set. However, among the 9 mecciRNAs discovered, 3 were annotated to the MT-ND6 gene, while two were mapped to the anti-sense strand of the MT-ND5 gene, two to the anti-sense strand of the MT-ND1 gene, and two to the anti-sense strand of the MT-CYB gene.

The MT-ND5, MT-ND6 and MT-ND1 genes contain instructions for synthesizing proteins known as NADH dehydrogenase 5, NADH dehydrogenase 6 and NADH dehydrogenase 1, which are vital components of the large enzyme complex called complex I. Complex I operates within the mitochondria and plays a crucial role in oxidative phosphorylation, a process essential for ATP generation. The MT-CYB gene provides instructions for making a protein called cytochrome b, which is one of 11 components of a group of proteins called complex III. These enzyme complexes are housed within the intricately folded inner mitochondrial membrane. During oxidative phosphorylation, these mitochondrial enzyme complexes execute chemical reactions that lead to ATP production. More specifically, they facilitate the gradual transfer of negatively charged particles, or electrons, creating an uneven electrical charge on either side of the inner mitochondrial membrane. This disparity in electrical charge serves as the energy source for ATP synthesis.

We also observed an upregulation of circular RNAs originating from the RPA gene. This finding aligns with the results of the study by Liu et al. 2020 which reported the overexpression of the RPA gene in various cancer types [103].

Conclusion

In conclusion, this thesis has provided a comprehensive analysis of circular RNAs in normal and CLL samples, shedding light on their abundance, expression patterns, and potential significance in CLL pathogenesis. Our findings reveal a positive correlation between absolute and relative expression of circRNAs, highlighting their potential as primary gene products. Additionally, we have identified circRNAs with significantly different CLP in normal and CLL conditions, unveiling a distinct CLL-specific circRNA CLP profile. Moreover, our investigation into circRNA-miRNA interactions adds to our understanding of circRNA function.

Furthermore, the detection of mitochondrion-encoded circRNAs shows the necessity for innovative approaches in deciphering mitochondrial gene regulation in cancer. Moving forward, experimental validation of identified circRNAs and elucidation of their functional roles in CLL pathogenesis is imperative. Additionally, exploring the potential of validated circRNAs as biomarkers for CLL diagnosis, prognosis, and therapeutic response holds significant promise. Integrating a circRNA disease-related signature into diagnostic and prognostic algorithms could enhance CLL patient management, leading to more tailored treatment approaches.

Acknowledgements

I would like to extend my deepest gratitude to my esteemed professors, Dr. Stefania Bortoluzzi and Dr. Enrico Gaffo , for their unwavering support, guidance, and encouragement throughout this journey. The time I have spent in your lab has been an incredibly enriching experience, one that I will cherish for the rest of my life.

Your insightful advice and constructive feedback have been invaluable, guiding me through the complexities of my research. The knowledge I have gained under your tutelage extends far beyond academic learning. Through your mentorship, I have grown not only as a scholar but also as an individual. You have taught me the true meaning of dedication, perseverance, and intellectual curiosity.

Dr. Bortoluzzi, your passion for research and commitment to excellence have inspired me to strive for the highest standards in my work. Your patience and willingness to always provide support, no matter how busy your schedule, have been a source of great comfort and motivation.

Dr. Gaffo, your wisdom and thoughtful guidance have been instrumental in shaping my academic journey. You have shown me the importance of critical thinking and the value of exploring new perspectives. Your encouragement has given me the confidence to tackle challenges and embrace opportunities for growth.

The collaborative and stimulating environment you have fostered in the lab has been a perfect setting for both my academic and personal development. I have learned the importance of teamwork, the power of diverse ideas, and the joy of shared discovery.

Additionally, I would like to thank Dr. Merry Giantin for her presence and support. I appreciate her being part of this academic journey.

I am profoundly grateful for the opportunities you have provided me and the trust you have placed in my abilities. Your mentorship has not only guided my research but has also profoundly influenced my outlook on life and my future aspirations.

Thank you, Dr. Stefania Bortoluzzi and Dr. Enrico Gaffo, and Dr. Merry Giantin, for being exceptional mentors and for making this journey a transformative and fulfilling experience.

References

1. Mukkamalla, S. K. R., Taneja, A., Malipeddi, D. & Master, S. R. *Chronic Lymphocytic Leukemia*. (StatPearls Publishing, 2023).
2. Ghia, P. & Caligaris-Cappio, F. Monoclonal B-cell lymphocytosis: right track or red herring? *Blood* **119**, 4358–4362 (2012).
3. Dühren-von Minden, M. *et al.* Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. *Nature* **489**, 309–312 (2012).
4. Hallek, M. *et al.* iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood* **131**, 2745–2760 (2018).
5. Binet, J. L. *et al.* A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* **48**, 198–206 (1981).
6. Baccarani, M., Cavo, M., Gobbi, M., Lauria, F. & Tura, S. Staging of chronic lymphocytic leukemia. *Blood* **59**, 1191–1196 (1982).
7. Visco, C. *et al.* Impact of immune thrombocytopenia on the clinical course of chronic lymphocytic leukemia. *Blood* **111**, 1110–1116 (2008).
8. Dearden, C. *et al.* The prognostic significance of a positive direct antiglobulin test in chronic lymphocytic leukemia: a beneficial effect of the combination of fludarabine and cyclophosphamide on the incidence of hemolytic anemia. *Blood* **111**, 1820–1826 (2008).
9. Parikh, S. A. *et al.* Hypogammaglobulinemia in newly diagnosed chronic lymphocytic leukemia: Natural history, clinical correlates, and outcomes. *Cancer* **121**, 2883–2891 (2015).
10. Kyasa, M. J., Parrish, R. S., Schichman, S. A. & Zent, C. S. Autoimmune cytopenia does not predict poor prognosis in chronic lymphocytic leukemia/small lymphocytic lymphoma. *Am. J. Hematol.* **74**, 1–8 (2003).
11. Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375–2390 (2016).
12. Rawstron, A. C. *et al.* Reproducible diagnosis of chronic lymphocytic leukemia by flow

cytometry: An European Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project. *Cytometry B Clin. Cytom.* **94**, 121–128 (2018).

13. Rawstron, A. C. *et al.* International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. *Leukemia* **21**, 956–964 (2007).
14. Potter, K. N. *et al.* Structural and functional features of the B-cell receptor in IgG-positive chronic lymphocytic leukemia. *Clin. Cancer Res.* **12**, 1672–1679 (2006).
15. Geisler, C. H. *et al.* Prognostic importance of flow cytometric immunophenotyping of 540 consecutive patients with B-cell chronic lymphocytic leukemia. *Blood* **78**, 1795–1802 (1991).
16. Fournier, S., Delespesse, G., Rubio, M., Biron, G. & Sarfati, M. CD23 antigen regulation and signaling in chronic lymphocytic leukemia. *J. Clin. Invest.* **89**, 1312–1321 (1992).
17. Freedman, A. S. *et al.* Normal cellular counterparts of B cell chronic lymphocytic leukemia. *Blood* **70**, 418–427 (1987).
18. Lipshutz, M. D., Mir, R., Rai, K. R. & Sawitsky, A. Bone marrow biopsy and clinical staging in chronic lymphocytic leukemia. *Cancer* **46**, 1422–1427 (1980).
19. Pangalis, G. A., Boussiotis, V. A. & Kittas, C. B-chronic lymphocytic leukemia. Disease progression in 150 untreated stage A and B patients as predicted by bone marrow pattern. *Nouv. Rev. Fr. Hematol.* **30**, 373–375 (1988).
20. Rozman, C. *et al.* Bone marrow histologic pattern--the best single prognostic parameter in chronic lymphocytic leukemia: a multivariate survival analysis of 329 cases. *Blood* **64**, 642–648 (1984).
21. Borthakur, G. *et al.* Immune anaemias in patients with chronic lymphocytic leukaemia treated with fludarabine, cyclophosphamide and rituximab--incidence and predictors. *Br. J. Haematol.* **136**, 800–805 (2007).
22. Kipps, T. J. *et al.* Chronic lymphocytic leukaemia. *Nat Rev Dis Primers* **3**, 16096 (2017).
23. Molica, S. *et al.* The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: Independent validation in a prospective cohort of early stage patients. *Am. J. Hematol.* **91**, 1090–1095 (2016).
24. Iskierka-Jażdżewska, E. & Robak, T. Minimizing and managing treatment-associated

complications in patients with chronic lymphocytic leukemia. *Expert Rev. Hematol.* **13**, 39–53 (2020).

25. Chemotherapeutic options in chronic lymphocytic leukemia: a meta-analysis of the randomized trials. CLL Trialists' Collaborative Group. *J. Natl. Cancer Inst.* **91**, 861–868 (1999).
26. Hoechstetter, M. A. *et al.* Early, risk-adapted treatment with fludarabine in Binet stage A chronic lymphocytic leukemia patients: results of the CLL1 trial of the German CLL study group. *Leukemia* **31**, 2833–2837 (2017).
27. Condoluci, A. *et al.* International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. *Blood* **135**, 1859–1869 (2020).
28. Goede, V. *et al.* Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N. Engl. J. Med.* **370**, 1101–1110 (2014).
29. Woyach, J. A. Management of relapsed/refractory Chronic Lymphocytic Leukemia. *Am. J. Hematol.* **97 Suppl 2**, S11–S18 (2022).
30. Fabbri, G. *et al.* Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J. Exp. Med.* **208**, 1389–1401 (2011).
31. Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
32. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
33. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
34. Puente, X. S. *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
35. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
36. Ljungström, V. *et al.* Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations. *Blood* **127**, 1007–1016 (2016).
37. Kasar, S. *et al.* Whole-genome sequencing reveals activation-induced cytidine deaminase

signatures during indolent chronic lymphocytic leukaemia evolution. *Nat. Commun.* **6**, 8866 (2015).

38. Landau, D. A. *et al.* The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nat. Commun.* **8**, 2185 (2017).
39. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2011).
40. Döhner, H. *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343**, 1910–1916 (2000).
41. Fabbri, G. *et al.* Genetic lesions associated with chronic lymphocytic leukemia transformation to Richter syndrome. *J. Exp. Med.* **210**, 2273–2288 (2013).
42. Rossi, D. *et al.* The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* **117**, 3391–3401 (2011).
43. Mansouri, L. *et al.* Functional loss of IκBε leads to NF-κB deregulation in aggressive chronic lymphocytic leukemia. *J. Exp. Med.* **212**, 833–843 (2015).
44. Zenz, T. *et al.* TP53 mutation and survival in chronic lymphocytic leukemia. *J. Clin. Oncol.* **28**, 4473–4479 (2010).
45. Austen, B. *et al.* Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. *Blood* **106**, 3175–3182 (2005).
46. Balatti, V. *et al.* NOTCH1 mutations in CLL associated with trisomy 12. *Blood* **119**, 329–331 (2012).
47. Ramsay, A. J. *et al.* POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.* **45**, 526–530 (2013).
48. Baliakas, P. *et al.* Cytogenetic complexity in chronic lymphocytic leukemia: definitions, associations, and clinical impact. *Blood* **133**, 1205–1216 (2019).
49. Attygalle, A. D. t(14;19)(q32;q13)-associated B-cell neoplasms—a review. *J. Hematop.* **5**, 159–163 (2012).
50. Chapiro, E. *et al.* The most frequent t(14;19)(q32;q13)-positive B-cell malignancy corresponds to

an aggressive subgroup of atypical chronic lymphocytic leukemia. *Leukemia* **22**, 2123–2127 (2008).

51. Martín-Subero, J. I. *et al.* A comprehensive genetic and histopathologic analysis identifies two subgroups of B-cell malignancies carrying a t(14;19)(q32;q13) or variant BCL3-translocation. *Leukemia* **21**, 1532–1544 (2007).
52. Rossi, D. *et al.* BCL3 translocation in CLL with typical phenotype: assessment of frequency, association with cytogenetic subgroups, and prognostic significance. *Br. J. Haematol.* **150**, 702–704 (2010).
53. Huh, Y. O. *et al.* The t(14;19)(q32;q13)-positive small B-cell leukaemia: a clinicopathologic and cytogenetic study of seven cases. *Br. J. Haematol.* **136**, 220–228 (2007).
54. Huh, Y. O. *et al.* Chronic lymphocytic leukemia with t(14;19)(q32;q13) is characterized by atypical morphologic and immunophenotypic features and distinctive genetic features. *Am. J. Clin. Pathol.* **135**, 686–696 (2011).
55. Yamamoto, K. *et al.* Translocation (14;19)(q32;q13) detected by spectral karyotyping and lack of BCL3 rearrangement in CD5-positive B-cell lymphoma associated with hemophagocytic syndrome. *Cancer Genet. Cytogenet.* **130**, 38–41 (2001).
56. Fujita, T., Nolan, G. P., Liou, H. C., Scott, M. L. & Baltimore, D. The candidate proto-oncogene bcl-3 encodes a transcriptional coactivator that activates through NF-kappa B p50 homodimers. *Genes Dev.* **7**, 1354–1363 (1993).
57. Bundy, D. L. & McKeithan, T. W. Diverse effects of BCL3 phosphorylation on its modulation of NF-kappaB p52 homodimer binding to DNA. *J. Biol. Chem.* **272**, 33132–33139 (1997).
58. Wulczyn, F. G., Naumann, M. & Scheidereit, C. Candidate proto-oncogene bcl-3 encodes a subunit-specific inhibitor of transcription factor NF-kappa B. *Nature* **358**, 597–599 (1992).
59. Rayet, B. & Gélinas, C. Aberrant rel/nfkb genes and activity in human cancer. *Oncogene* **18**, 6938–6947 (1999).
60. Kashatus, D., Cogswell, P. & Baldwin, A. S. Expression of the Bcl-3 proto-oncogene suppresses p53 activation. *Genes Dev.* **20**, 225–235 (2006).
61. Rebollo, A. *et al.* Bcl-3 expression promotes cell survival following interleukin-4 deprivation and

is controlled by AP1 and AP1-like transcription factors. *Mol. Cell. Biol.* **20**, 3407–3416 (2000).

62. Schlette, E., Rassidakis, G. Z., Canoz, O. & Medeiros, L. J. Expression of bcl-3 in chronic lymphocytic leukemia correlates with trisomy 12 and abnormalities of chromosome 19. *Am. J. Clin. Pathol.* **123**, 465–471 (2005).
63. Ohno, H., Nishikori, M., Maesako, Y. & Haga, H. Reappraisal of BCL3 as a molecular marker of anaplastic large cell lymphoma. *Int. J. Hematol.* **82**, 397–405 (2005).
64. Canoz, O., Rassidakis, G. Z., Admirand, J. H. & Medeiros, L. J. Immunohistochemical detection of BCL-3 in lymphoid neoplasms: a survey of 353 cases. *Mod. Pathol.* **17**, 911–917 (2004).
65. Visentin, A. *et al.* Integrated CLL Scoring System, a New and Simple Index to Predict Time to Treatment and Overall Survival in Patients With Chronic Lymphocytic Leukemia. *Clin. Lymphoma Myeloma Leuk.* **15**, 612–20.e1–5 (2015).
66. Visentin, A. *et al.* Continuous treatment with Ibrutinib in 100 untreated patients with TP53 disrupted chronic lymphocytic leukemia: A real-life campus CLL study. *Am. J. Hematol.* **97**, E95–E99 (2022).
67. Wu, Z., Sun, H., Li, J. & Jin, H. Circular RNAs in leukemia. *Aging* **11**, 4757–4771 (2019).
68. Deng, W., Chao, R. & Zhu, S. Emerging roles of circRNAs in leukemia and the clinical prospects: An update. *Immun Inflamm Dis* **11**, e725 (2023).
69. Chen, X. *et al.* circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.* **6**, 34985 (2016).
70. Vicens, Q. & Westhof, E. Biogenesis of Circular RNAs. *Cell* vol. 159 13–14 (2014).
71. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
72. Starke, S. *et al.* Exon circularization requires canonical splice signals. *Cell Rep.* **10**, 103–111 (2015).
73. Chen, L.-L. & Yang, L. Regulation of circRNA biogenesis. *RNA Biol.* **12**, 381–388 (2015).
74. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
75. Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol.*

Biol. **22**, 256–264 (2015).

76. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* **32**, 453–461 (2014).
77. Zhang, Y. *et al.* Circular intronic long noncoding RNAs. *Mol. Cell* **51**, 792–806 (2013).
78. Wang, H., Feng, C., Wang, M., Yang, S. & Wei, F. Circular RNAs: Diversity of Functions and a Regulatory Nova in Oral Medicine: A Pilot Review. *Cell Transplant.* **28**, 819–830 (2019).
79. Conn, S. J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134 (2015).
80. Ashwal-Fluss, R. *et al.* circRNA biogenesis competes with pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014).
81. Meng, X. *et al.* Circular RNA: an emerging key player in RNA world. *Brief. Bioinform.* **18**, 547–557 (2017).
82. Rybak-Wolf, A. *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **58**, 870–885 (2015).
83. Panni, S., Lovering, R. C., Porras, P. & Orchard, S. Non-coding RNA regulatory networks. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194417 (2020).
84. Ulshöfer, C. J., Pfafenrot, C., Bindereif, A. & Schneider, T. Methods to study circRNA-protein interactions. *Methods* **196**, 36–46 (2021).
85. Aufiero, S., Reckman, Y. J., Pinto, Y. M. & Creemers, E. E. Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* **16**, 503–514 (2019).
86. Ho-Xuan, H. *et al.* Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Res.* **48**, 10368–10382 (2020).
87. Granados-Riveron, J. T. & Aquino-Jarquin, G. The complexity of the translation ability of circRNAs. *Biochim. Biophys. Acta* **1859**, 1245–1251 (2016).
88. Schneider, T. *et al.* CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs. *Sci. Rep.* **6**, 31313 (2016).
89. Pamudurti, N. R. *et al.* Translation of CircRNAs. *Mol. Cell* **66**, 9–21.e7 (2017).
90. Hernández, G., Vázquez-Pianzola, P., Sierra, J. M. & Rivera-Pomar, R. Internal ribosome entry

site drives cap-independent translation of reaper and heat shock protein 70 mRNAs in Drosophila embryos. *RNA* **10**, 1783–1797 (2004).

91. Di Timoteo, G. *et al.* Modulation of circRNA Metabolism by m6A Modification. *Cell Rep.* **31**, 107641 (2020).
92. Cortés-López, M. & Miura, P. Emerging Functions of Circular RNAs. *Yale J. Biol. Med.* **89**, 527–537 (2016).
93. Sinha, T., Panigrahi, C., Das, D. & Chandra Panda, A. Circular RNA translation, a path to hidden proteome. *Wiley Interdiscip. Rev. RNA* **13**, e1685 (2022).
94. Ng, W. L. *et al.* Inducible RasGEF1B circular RNA is a positive regulator of ICAM-1 in the TLR4/LPS pathway. *RNA Biol.* **13**, 861–871 (2016).
95. Garikipati, V. N. S. *et al.* Circular RNA CircFndc3b modulates cardiac repair after myocardial infarction via FUS/VEGF-A axis. *Nat. Commun.* **10**, 4317 (2019).
96. Rossi, F. *et al.* Circular RNA ZNF609/CKAP5 mRNA interaction regulates microtubule dynamics and tumorigenicity. *Mol. Cell* **82**, 75–89.e9 (2022).
97. Zhou, L.-Y. *et al.* The circular RNA ACR attenuates myocardial ischemia/reperfusion injury by suppressing autophagy via modulation of the Pink1/ FAM65B pathway. *Cell Death Differ.* **26**, 1299–1315 (2019).
98. An, Y., Furber, K. L. & Ji, S. Pseudogenes regulate parental gene expression via ceRNA network. *J. Cell. Mol. Med.* **21**, 185–192 (2017).
99. Akhter, R. Circular RNA and Alzheimer's Disease. *Adv. Exp. Med. Biol.* **1087**, 239–243 (2018).
100. Conn, V. M. *et al.* Circular RNAs drive oncogenic chromosomal translocations within the MLL recombinome in leukemia. *Cancer Cell* **41**, 1309–1326.e10 (2023).
101. Roy Chowdhury, S. & Banerji, V. Targeting Mitochondrial Bioenergetics as a Therapeutic Strategy for Chronic Lymphocytic Leukemia. *Oxid. Med. Cell. Longev.* **2018**, 2426712 (2018).
102. Jitschin, R. *et al.* Mitochondrial metabolism contributes to oxidative stress and reveals therapeutic targets in chronic lymphocytic leukemia. *Blood* **123**, 2663–2672 (2014).
103. Liu, X. *et al.* Identification of mecciRNAs and their roles in the mitochondrial entry of proteins. *Sci. China Life Sci.* **63**, 1429–1449 (2020).

104. Liu, X., Yang, Y. & Shan, G. Identification and detection of mecciRNAs. *Methods* **196**, 147–152 (2021).

105. Luan, J. *et al.* circMTND5 Participates in Renal Mitochondrial Injury and Fibrosis by Sponging MIR6812 in Lupus Nephritis. *Oxid. Med. Cell. Longev.* **2022**, 2769487 (2022).

106. Wu, Z. *et al.* Mitochondrial Genome-Derived circRNA mc-COX2 Functions as an Oncogene in Chronic Lymphocytic Leukemia. *Mol. Ther. Nucleic Acids* **20**, 801–811 (2020).

107. Liu, C.-X. *et al.* Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell* **177**, 865–880.e21 (2019).

108. Kim, S. M. *et al.* Cancer-derived exosomes as a delivery platform of CRISPR/Cas9 confer cancer cell tropism-dependent targeting. *J. Control. Release* **266**, 8–16 (2017).

109. Park, O. H. *et al.* Endoribonucleolytic Cleavage of m6A-Containing RNAs by RNase P/MRP Complex. *Mol. Cell* **74**, 494–507.e8 (2019).

110. Gaffo, E., Buratin, A., Dal Molin, A. & Bortoluzzi, S. Sensitive, reliable and robust circRNA detection from RNA-seq with CirComPara2. *Brief. Bioinform.* **23**, (2022).

111. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

112. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

113. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

114. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013) doi:10.48550/arXiv.1303.3997.

115. Hoffmann, S. *et al.* A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.* **15**, R34 (2014).

116. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

117. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12**, R72 (2011).

118. Hansen, T. B., Venø, M. T., Damgaard, C. K. & Kjems, J. Comparison of circular RNA

prediction tools. *Nucleic Acids Res.* **44**, e58–e58 (2015).

119. Hansen, T. B. Improved circRNA Identification by Combining Prediction Algorithms. *Front Cell Dev Biol* **6**, 20 (2018).
120. Vromman, M. *et al.* Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nat. Methods* **20**, 1159–1169 (2023).
121. Zhang, X.-O. *et al.* Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).
122. Gao, Y., Zhang, J. & Zhao, F. Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* **19**, 803–810 (2018).
123. Cheng, J., Metge, F. & Dieterich, C. Specific identification and quantification of circular RNAs from sequencing data. *Bioinformatics* **32**, 1094–1096 (2016).
124. Shi, Y. & Shang, J. Circular RNA Expression Profiling by Microarray-A Technical and Practical Perspective. *Biomolecules* **13**, (2023).
125. Razeghian-Jahromi, I., Zibaeenezhad, M. J., Karimi Akhormeh, A. & Dara, M. Expression ratio of circular to linear ANRIL in hypertensive patients with coronary artery disease. *Sci. Rep.* **12**, 1802 (2022).
126. Douma, J. C. & Weedon, J. T. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods Ecol. Evol.* doi:10.1111/2041-210X.13234.
127. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11**, 733–739 (2010).
128. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
129. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
130. Leek, J. T. & Storey, J. D. A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18718–18723 (2008).
131. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model

analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

132. Buratin, A., Bortoluzzi, S. & Gaffo, E. Systematic benchmarking of statistical methods to assess differential expression of circular RNAs. *Brief. Bioinform.* **24**, (2023).

133. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).

134. Marín, R. M. & Vanícek, J. Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res.* **39**, 19–29 (2011).

135. Chen, K., Maaskola, J., Siegal, M. L. & Rajewsky, N. Reexamining microRNA site accessibility in Drosophila: a population genomics study. *PLoS One* **4**, e5681 (2009).

136. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).

137. Fromm, B. *et al.* MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res.* **50**, D204–D210 (2022).

138. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).

139. Wang, D., Gu, J., Wang, T. & Ding, Z. OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics* **30**, 2237–2238 (2014).

140. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

141. Sun, M. & Yang, Y. Biological functions and applications of circRNAs—next generation of RNA-based therapy. *J. Mol. Cell Biol.* **15**, mjad031 (2023).

142. Liu, R., Ma, Y., Guo, T. & Li, G. Identification, biogenesis, function, and mechanism of action of circular RNAs in plants. *Plant Comm* **4**, (2023).

143. Buratin, A. *et al.* Large-scale circular RNA deregulation in T-ALL: unlocking unique ectopic expression of molecular subtypes. *Blood Adv* **4**, 5902–5914 (2020).

144. Liu, H., Zeng, L., Yang, Y., Guo, C. & Wang, H. Bcl-3: A Double-Edged Sword in Immune Cells and Inflammation. *Front. Immunol.* **13**, 847699 (2022).

145. Seaton, G., Smith, H., Brancale, A., Westwell, A. D. & Clarkson, R. Multifaceted roles for BCL3 in cancer: a proto-oncogene comes of age. *Mol. Cancer* **23**, 1–17 (2024).

146. Schuster, M., Annemann, M., Plaza-Sirvent, C. & Schmitz, I. Atypical IκB proteins - nuclear modulators of NF-κB signaling. *Cell Commun. Signal.* **11**, 23 (2013).

147. Babayev, M. & Silveyra, P. Role of circular RNAs in lung cancer. *Front. Genet.* **15**, 1346119 (2024).

148. Palmer, S. & Chen, Y. H. Bcl-3, a multifaceted modulator of NF-kappaB-mediated gene transcription. *Immunol. Res.* **42**, 210–218 (2008).

149. Oh, A. *et al.* NF-κB signaling in neoplastic transition from epithelial to mesenchymal phenotype. *Cell Commun. Signal.* **21**, 291 (2023).

150. Mallm, J.-P. *et al.* Linking aberrant chromatin features in chronic lymphocytic leukemia to transcription factor networks. *Mol. Syst. Biol.* **15**, e8339 (2019).

151. García-Ortí, L. *et al.* Integration of SNP and mRNA arrays with microRNA profiling reveals that MiR-370 is upregulated and targets NF1 in acute myeloid leukemia. *PLoS One* **7**, e47717 (2012).

152. Zhu, D.-X. *et al.* miR-181a/b significantly enhances drug sensitivity in chronic lymphocytic leukemia cells via targeting multiple anti-apoptosis genes. *Carcinogenesis* **33**, 1294–1301 (2012).

153. Li, Z. *et al.* Up-regulation of a HOXA-PBX3 homeobox-gene signature following down-regulation of miR-181 is associated with adverse prognosis in patients with cytogenetically abnormal AML. *Blood* **119**, 2314–2324 (2012).

154. Visone, R. *et al.* miR-181b is a biomarker of disease progression in chronic lymphocytic leukemia. *Blood* **118**, 3072–3079 (2011).

155. Zimmerman, E. I. *et al.* Lyn kinase-dependent regulation of miR181 and myeloid cell leukemia-1 expression: implications for drug resistance in myelogenous leukemia. *Mol. Pharmacol.* **78**, 811–817 (2010).

156. Chen, H., Chen, Q., Fang, M. & Mi, Y. microRNA-181b targets MLK2 in HL-60 cells. *Sci. China Life Sci.* **53**, 101–106 (2010).

157. Pallasch, C. P. *et al.* miRNA deregulation by epigenetic silencing disrupts suppression of the

oncogene PLAG1 in chronic lymphocytic leukemia. *Blood* **114**, 3255–3264 (2009).

158. Pichiorri, F. *et al.* MicroRNAs regulate critical genes associated with multiple myeloma pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 12885–12890 (2008).
159. Mraz, M. *et al.* MicroRNA-650 expression is influenced by immunoglobulin gene rearrangement and affects the biology of chronic lymphocytic leukemia. *Blood* **119**, 2110–2113 (2012).
160. Zaidi, S. K. *et al.* Altered Runx1 subnuclear targeting enhances myeloid cell proliferation and blocks differentiation by activating a miR-24/MKP-7/MAPK network. *Cancer Res.* **69**, 8249–8255 (2009).
161. Girardot, M. *et al.* miR-28 is a thrombopoietin receptor targeting microRNA detected in a fraction of myeloproliferative neoplasm patient platelets. *Blood* **116**, 437–445 (2010).
162. Wong, K. Y. *et al.* Epigenetic inactivation of the MIR34B/C in multiple myeloma. *Blood* **118**, 5901–5904 (2011).
163. De Marchis, M. L. *et al.* A new molecular network comprising PU.1, interferon regulatory factor proteins and miR-342 stimulates ATRA-mediated granulocytic differentiation of acute promyelocytic leukemia cells. *Leukemia* **23**, 856–862 (2009).