![CUSJ - Columbia Undergraduate Science Journal]

Identification of LncRNAs as Therapeutic Targets in Chronic Lymphocytic Leukemia

Simay Dolaner[1,2], Harpreet Kaur[2], Mohit Mazumder[2], Julia Panov[2,3], Elia Brodsky[2]

[1]Department of Molecular Biology and Genetics, Bahcesehir University, Istanbul, Turkey
[2]Pine Biotech, New Orleans, USA
[3]Tauber Bioinformatics Research Center, Haifa, Israel

ABSTRACT: Chronic Lymphocytic Leukemia (CLL) is a type of blood cancer that has a very heterogeneous biological background and diverse treatment strategies. However, a small part of this malignancy may disappear without receiving any treatment, known as "spontaneous regression", which occurs as a result of a poorly investigated mechanism. Exposing the underlying causes of this condition can lead to a novel treatment approach for CLL. In this article, we applied in-silico analysis on total RNA expression data from 24 CLL samples to determine possible regulatory mechanisms of spontaneous regression in CLL. These were first selected by comparing spontaneous regression with progressive samples of CLL at the transcriptional level using two unsupervised machine learning algorithms, i.e., Principal Component Analysis (PCA) and Hierarchical Clustering. Subsequently, the DESeq2 algorithm was used to scrutinize only statistically significant (adjusted p-value < 0.01) RNA transcripts that can differentiate both conditions. Here, at first, we have elucidated 870 significantly differentially expressed protein-coding genes that were involved in the biogenesis and processing of RNA. Consequently, these findings led our study to investigate non-coding RNA, and 33 long non-coding RNAs (lncRNAs) were found to be significantly differentially expressed among two conditions based on differential gene expression analysis. Further, our analysis in the current study suggested lncRNAs, PTPN22-AS1, PCF11-AS1, SYNGAP1-AS1, PRRT3-AS1, and H1FX-AS1 as potential therapeutic targets to trigger spontaneous regression. Ultimately, the results presented here reveal new insights into spontaneous regression and its relationship with non-coding RNAs, particularly lncRNAs.

INTRODUCTION

According to the American Cancer Society statistics, leukemia is the second leading blood cancer, with roughly 60,000 new cases identified in 2020 [1]. Leukemia has different subtypes, which are classified in the context of the tumor origin [2]. The most common variety in adults is chronic lymphocytic leukemia (CLL), which is a lymphoid malignancy due to failed apoptosis and aggressive proliferation of mature B cells [3]. These cells circulate through the blood as non-proliferating cells or arrested cells in the G0/G1 phase of the cell cycle and may affect the function of normal cells in other organs [4].

CLL has a highly diverse biological and clinical background for each patient that determines the stage of the disease [5]. Although there are several stages of CLL classified according to their genetic background and B cell number, one of the most intriguing concepts is known as spontaneous regression, which is the disappearance of the tumor over time either without any treatment or with treatment that is categorized as insufficient to have an impact on the tumor [6].

Spontaneous regression can be seen in 1-2% of all CLL patients, and it is a phenomenon that is poorly understood [7]. In this process, cells that proliferate uncontrollably are transmitted to the quiescent state so that the tumor disappears partially or completely with time [6]. Spontaneous regression is not a common feature of cancer cells and is regulated by mechanisms that are not well-understood. If such mechanisms can be determined, target biological molecules that have a specific role in disease progression can be identified and manipulated in vitro.

Current strategies aim to inhibit BCR signaling, which is crucial for the survival of the B-cells, and chemokine signaling that creates survival signals and attracts leukemic cells to communicate with its microenvironment [5]. Moreover, activation of apoptosis pathways via blocking BCL2 activity, an anti-apoptotic protein, which is highly expressed in leukemic cells, is included in the aforementioned strategies [8]. Although these types of targeted therapies improve the outcome of the patients, the heterogeneity of the leukemia microenvironment reinforces the necessity of new targets. As the targeted therapies may have an impact on tumor surroundings and affect the other cells found in the tumor microenvironment, triggering spontaneous regression mechanisms may improve the strategies as well as patient outcome.

In CLL, the most frequent chromosomal abnormalities and somatic mutations on the protein-coding region of the genome have been distinguished as a result of genomic sequencing, and disrupted cellular pathways are identified using next-generation sequencing of mRNA expression [9]. In spite of this progress, nearly 20% of CLL cells do not show chromosomal abnormalities or genomic variation. Therefore, researchers recently shifted their focus to the non-coding region of the genome and regulatory RNA molecules, especially long non-coding RNAs, which are deregulated in many cancers [10].

Long non-coding RNAs (lncRNAs) are a subgroup of non-coding RNAs which are longer than 200 nucleotides and encompass thousands of diverse transcripts in humans [11]. There are approximately 100,000 known lncRNAs, and this quantity is expanding each year with the new studies [12].

LncRNAs play a significant role in gene regulation, controlling multiple cellular mechanisms involved in tumor progression. They are involved in epigenetic regulation of gene expression via histone modification, DNA methylation, or acetylation. Specifically, these epigenetic regulations may include recruitment of histone remodeling complexes, interaction with histone methyltransferases and demethylases to regulate DNA methylation, or histone acetyltransferases and deacetylases to modulate the acetylation [13]. Furthermore, lncRNAs may regulate gene expression at the transcriptional level by recruiting transcription factors [14]. Moreover, lncRNAs can produce hybrids or act as scaffolds through interaction with proteins to regulate expression at the post-translational level, including regulation of phosphorylation and ubiquitination [15].

In tumor development, lncRNAs can serve as either tumor suppressors, oncogenes, or even both at the same time for some cancer types [16]. Analysis of lncRNA expression patterns had led to the identification of putative biomarkers such as HOTAIR, H19, and DLEU1/2 [17]. Specifically, HOTAIR serves as an oncogene by inducing invasiveness and metastasis in several cancers via recruiting a demethylase [13]. Since this lncRNA is highly expressed in aggressive tumors, it is considered as a bio-

marker and a possible therapeutic target for many cancers [18]. DLEU1/2 epigenetically regulate the tumor suppressors and are deleted in CLL cells, which results in the progression of the CLL. This allows it to serve as a biomarker in the diagnosis [19]. On the other hand, H19 has a dual role in different cancers by stimulating distinct mechanisms through transcription factors, which makes it a target that needs to be studied separately for each cancer [13].

Furthermore, lncRNAs are known to play an important role in cell differentiation and tissue specificity [20]. However, characterization may be compelling because lncRNAs are transcribed in different loci and localized distinctly. More importantly, lncRNA expression is generally tissue-specific and can be detected under certain conditions [21]. As the lncRNAs are differentially expressed, their roles and activities can be identified in disease conditions.

Due to the diverse functions of lncRNAs, novel studies are concentrated on the identification of lncRNAs as therapeutic targets. As the expression of these molecules is tissue and disease-specific, this specificity makes them excellent targets compared to protein-coding genes [22].

In the scope of this project, the trigger mechanism behind the spontaneous regression process is investigated at the transcriptomic level to identify a pattern of lncRNA expression that would explain cell "decision" by comparing CLL tissue at the spontaneous regression and the progressive states.

METHODS

Dataset

The dataset of this project was generated by Kwok et al. (2020) and published as a BioProject on the NCBI with the accession number PRJNA535508 (Supplementary Notes 1) [6]. Transcriptome data contains raw reads of RNA sequencing from the Illumina Nextseq 550 platform by using paired-end sequencing. In this study, the authors compared multiple samples from multiple subtypes of chronic lymphocytic leukemia that can be seen in Table I. In our project, spontaneous regression and progressive states were chosen for further analysis to investigate expression variation specifically involved in the regression mechanism in CLL cells.

Table I. Dataset of the BioProject

Disease State	Number of Samples
Spontaneous Regression	16
Progressive State	8
Indolent State	16
Unmutated State	8
Healthy Condition	3

RNA-seq Raw Data Processing

Raw reads were used to construct an RNA sequencing pipeline that contains pre-processing using Trimmomatic, mapping of reads on reference transcriptome with Bowtie2-t, and quantification using RNA-Seq by Expectation-Maximization (RSEM) algorithms with T-Bioinfo server (Supplementary Notes 1) as represented in Figure 1. Briefly, in order to get gene expression levels, duplicated sequences which resulted in PCR amplification were removed by PCR clean considering best coverage. Then, Trimmomatic was used to remove adaptor sequences and poor-quality data at the end of the sequence. Mapping of these clean reads on transcriptome was performed by using the Bowtie2-t algorithm based on the reference genome (GRCh38). Bowtie2-t is an option of the Bowtie, which is a mapping algorithm and aligns short reads according to the seed approach [23]. Finally, for the quantification of gene expression, the RSEM algorithm was used with FPKM normalization to obtain the gene expression table.
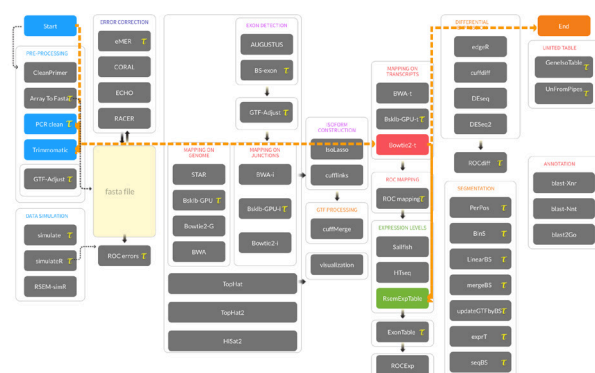
Figure 1: RNA-seq Data Processing Pipeline to generate an RNA expression table on T-Bioinfo Server

Exploratory Analysis

Exploratory analysis facilitated the examination of variation between all samples, including healthy and diseased patients, in order to select a comparison parameter for further analysis. Thus, the visual outputs were used to determine the patterns. Data was explored by using principal component analysis (PCA), which is a dimensionality reduction technique that discerns the variability between the samples [24]. PCA was performed twice, both for all the samples and for the spontaneous regression and the progressive samples to be able to observe the improvement on principle components. Hierarchical Clustering, which finds patterns among the samples using similarity measures [25], made it possible to understand the clustering aspects of spontaneous regression and progressive samples based on their gene expression, especially after the selection of statistically significant genes, which will be explained in the following section.

Differential Gene Expression Analysis

The Differential Gene Expression analysis was conducted by contrasting spontaneous regression with progressive CLL samples. Separate studies were performed for protein-coding and non-protein coding transcripts to understand the mechanisms involved in spontaneous regression.

The differential gene expression (DGE) pipeline, which includes pre-processing, mapping with HiSat2, RNA expression quantification using HTseq, and differential gene expression analysis with DESeq2 algorithm was constructed by using the T-Bioinfo server (Figure 2). Here, initially, PCR cleaning was performed by considering coverage to get rid of the duplicated sequences generated via PCR amplification. To eliminate the adaptors and bad quality sequences from the data, the Trimmomatic algorithm was used. Next, the HiSat2 was utilized for the mapping of sequence reads to the reference genome (GRCh38) by taking into account the splice junctions. Then, the HTseq algorithm quantified the gene expression through overlapping reads and generated a gene expression table in the form of count values. Finally, differential gene expression analysis was performed by using the DESeq2 algorithm which gives the expression differences between two groups using the shrinkage estimators [26]. DESeq2 provides results, which include p-value, log 2-fold change, and adjusted p-value. Then, an adjusted p-value or False Discovery Rate (FDR) that is a standard statistical value utilized for multiple testing correction, is calculated to eliminate the false-positive results [26, 27].

The T-Bioinfo server uses a one-step approach for DGE analysis and combines several methods. Although DESeq2 can be used for both normalization and statistical analysis, the T-Bioinfo server provides an additional approach, which includes gene set enrichment analysis (GSEA) to discover related pathways and processes [28].

Selection of Significant Genes

The significant genes were identified by their p-adjusted values and log2 fold change. The determined threshold for the adjusted p-value was 0.01 and the log2 fold change was ±1 in non-coding RNAs and ±1.5 in protein-coding genes.

Figure 2: Differential Gene Expression Analysis Pipeline Using the T-Bioinfo Server

Gene Ontology Analysis

The enrichment analysis was done via using the Enrichr platform for the protein-coding genes [29]. Respectively, the GO and KEGG pathways were considered for the observed upregulated and downregulated genes in spontaneous regression samples.

Data Visualization

The gene expression patterns were observed by heatmaps. Furthermore, PCA and H-Clustering were repeated with the selected significant RNA transcripts to make a comparison as before and after, and see an improvement on the basis of variance. Visualization was performed independently for protein-coding genes and non-protein coding genes.

RESULTS

Variation is Detected Between Spontaneous Regression and the Progressive Samples

The exploratory analysis using the PCA revealed that there was a variation between the spontaneous regression and the progressive state.

Based on Figure 3A, it can be seen that healthy samples and the diseased samples are well separated, which means there is significant variation between these groups. When the spontaneous regression and progressive samples were examined in a specif-

ic scatter plot, it was obvious that there was an improvement in principal components, and the two groups were clearly separated from each other (Figure 3B). After the observed variation between the two groups, the reason for this difference and its impacts could be investigated at the level of gene regulation.
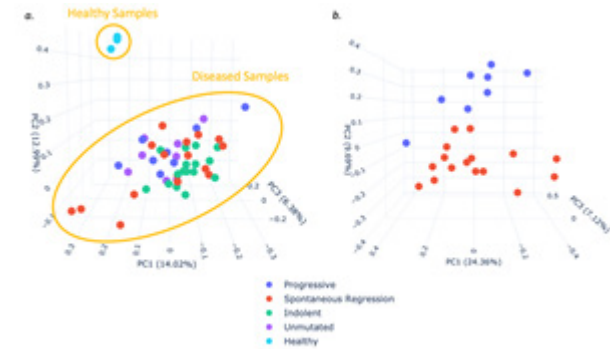
Figure 3. Principal Component Analysis,
(A) Healthy samples versus Diseased samples
(B) Spontaneous Regression (Red) versus Progressive Samples (Purple)

Revealed Pathways Involved in Biogenesis and Processing of RNA

Based on DESeq2 analysis, expression levels of 870 protein-coding genes were identified as significantly different between the two groups with p-adjusted values (<0.01) and log2 fold change (+/- 1.5) (Supplementary Table 1). The heatmap of the protein-coding genes shows the gene expression patterns between spontaneous regression and progressive samples (Figure 4A). It is obvious that most of the genes were upregulated in spontaneous regression while they downregulated in progressive ones. To understand the biological importance of pathways, gene ontology analysis was performed. Interestingly, although there were different pathways in upregulated genes, many represented biological pathways involved in biogenesis and processing of RNA, and mRNA translation (Figure 4B). In addition, identified pathways also included ribosome-related pathways, which imply translation of protein-coding RNAs is affected. The same concept could be detected with
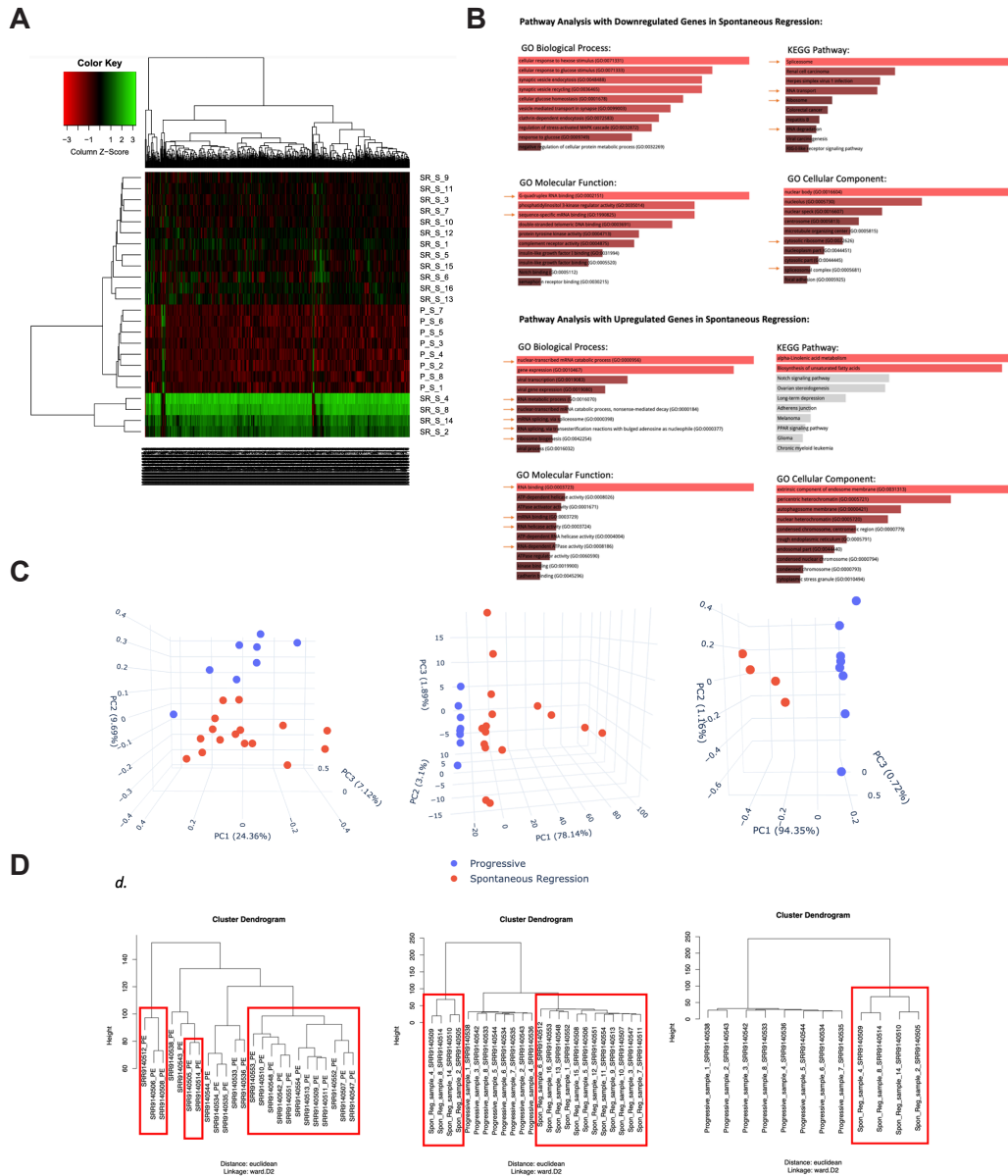
Figure 4. Data visualization of protein-coding genes
(A) Gene expression patterns of protein-coding genes. Heatmap exposes the downregulated and upregulated genes among the samples
(B) Gene ontology analysis of protein-coding genes. Pathway analysis shows the downregulated and upregulated pathways in the spontaneous regression samples, respectively. Arrows indicate the RNA related pathways
(C) Representation of principal component analysis with 3D scatter plots. First PCA shows the separation of spontaneous regression (SR) and progressive (P) samples before the selection of the protein-coding genes, second PCA represents the separation of SR and P samples after the selection of the protein-coding genes, and last PCA indicates the true variability source considering four outliers
(D) Hierarchical Clustering results as dendrograms. Each dendrogram point out the related scatter plots that are placed above and reveal the clustering perspective. Red boxes indicate the SR samples, and the other ones are the P samples

the downregulated genes as presented in Figure 4B. Consequently, these findings led this research to investigate the non-coding RNAs, which was the second branch of this study.

There was a Diversity Among the Spontaneous Regression Samples

According to the first principal component, spontaneous regression had a lot of variability within its own group (Figure 4C). This variation could be observed from the dendrograms in Figure 4D. Correspondingly, spontaneous regression samples had two diverse groups. The clinical data of these samples, therefore, needed to be examined to interpret them as outliers or true variability sources. However, no difference was detected between the four outliers and the other samples. This meant the difference had to be in their biological background. That is why these four diverse samples were selected in comparison with progressive samples to see the most significant difference between the two groups in the context of expression patterns.

As a result, it can be seen in Figure 4C that there was a significant improvement with the well-separated two groups after the selection of true variability source. Additionally, better separation could also be observed from the dendrogram in Figure 4D that four samples and the progressive samples have distinct branches compared to the remaining twelve samples.

Identified Non-Coding RNAs were Novel Transcripts

Based on differential gene expression analysis, 33 lncRNAs, which are significantly expressed, were identified with the specific parameters (Supplementary Table 2). Each lncRNA was investigated by their Ensembl ID and identified as novel transcripts. When the PCA was repeated with selected lncRNAs, clear separation could be seen among the spontaneous regression and the progressive samples (Figure 5A). Even though there was no dramatic improvement in principal components, hierarchical clustering results were highly distinctive. It can be seen in

Figure 5B that the two groups cluster separately in the dendrogram, which implies these significant genes might predict tumor characteristics.

To be able to observe the gene expression pattern of the lncRNAs, a heatmap was generated. As shown in Figure 5C, most of the genes were upregulated in progressive samples compared to spontaneous regression.

So far, pathway analysis of differentially expressed protein-coding genes in spontaneous regression samples has demonstrated that global gene expression in these samples might be modulated at the transcriptional and post-transcriptional level, and non-protein coding genes have shown that lncRNAs are differentially expressed in these samples. To make a biologically relevant interpretation, the analysis will regard lncRNAs identified as statistically most significant.

DISCUSSION

Although spontaneous tumor regression is rare, it is a phenomenon that can be observed in some types of cancer, such as neuroblastoma, renal cell carcinoma, lung cancer, lymphoma, and leukemia [30]. To be able to increase the occurrence rate of this mechanism, detailed studies should be conducted, and the related pathways should be examined.

In summary, we used RNA-seq datasets from CLL patients containing a total of 16 spontaneous regression and 8 progressive state samples. First of all, we looked for variation and pattern among these two groups by using PCA and H-Clustering based on gene expression. After the detection of variation, we hypothesized that the reason for this difference should be at the gene expression level. Therefore, we applied differential gene expression analysis by comparing spontaneous regression and progressive tumor states. We identified 870 protein-coding genes which were significantly differentially expressed and also were associated with RNA-related pathways based on gene ontology analysis. Depending on these findings, we also investigated non-coding RNAs and
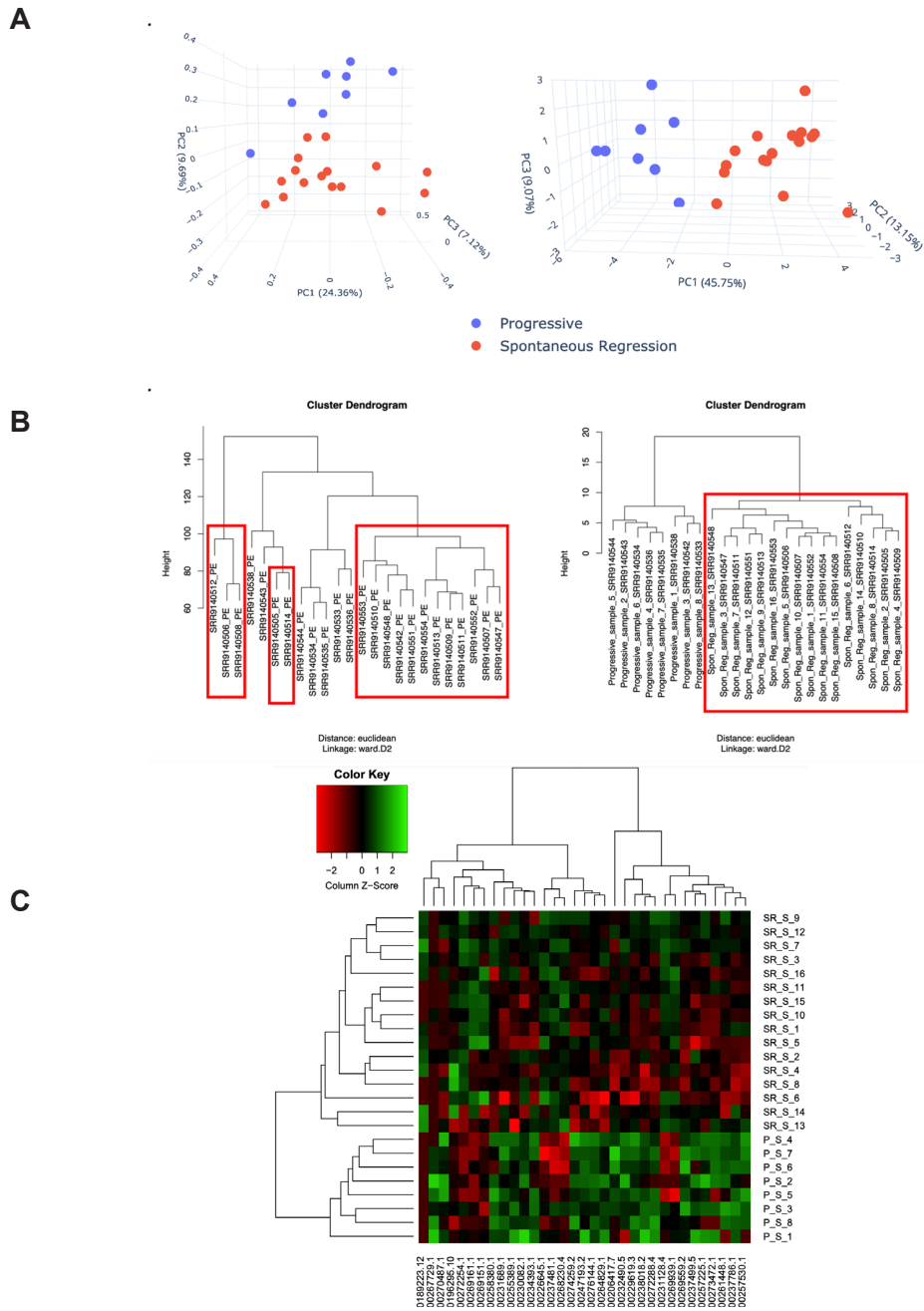
Figure 5. Data visualization of non-coding RNAs

(A) Representation of principal component analysis with 3D scatter plots. Initial PCA indicates the separation of spontaneous regression (SR) and progressive (P) samples before the selection of the non-coding RNAs and second PCA shows the separation of SR and P samples after the selection of the non-coding RNAs

(B) Hierarchical Clustering results as dendrograms. Each dendrogram specify the related scatter plots that are located above and expose the clustering perspective. Red boxes demonstrate the SR samples, and the other ones are the P samples

(C) Gene expression patterns of non-coding RNAs. Heatmap represents the downregulated and upregulated genes between the samples

identified significantly expressed 33 lncRNAs. This paper has highlighted that new therapeutic strategies may involve lncRNAs to trigger the spontaneous regression phenomena in cancer cells. Since the lncRNAs can regulate the important cellular mechanisms during cancer progression, identification of disease-specific lncRNAs may enlighten the way of new treatment strategies. Therefore, a detailed literature review for each lncRNA and their sense mRNAs had been done. Even though the detailed explanation can be found in Supplementary Table 2, identified lncRNAs are involved in several pathways including tumor growth, metastasis, cell survival, and regulation of the tumor microenvironment such as the immune system.

Previous studies showed that identified two lncRNAs, the PRRT3-AS1 and H1FX-AS1, were focused on cancer progression [31, 32, 33, 34]. Both lncRNAs were upregulated in the progressive state compared to spontaneous regression, which indicates that these lncRNAs were functioning as tumor enhancers in CLL cells. Li et al. (2020) showed that the lncRNA PRRT3-AS1 is upregulated in prostate cancer and targets the PPARγ gene by binding its 3' end, which leads to regulation of the Akt/mTOR signaling pathway [31]. The same study also revealed that the down-regulation of PRRT3-AS1 inhibits prostate cancer progression by regulating cell proliferation, migration, and apoptosis. Moreover, knocking down this lncRNA triggers autophagy via the mTOR pathway [31]. Another study suggested that PRRT3-AS1 is estimated as an immune-related lncRNA involved in PPAR signaling and can be used as a potential target in glioblastoma [32]. In light of these findings, it can be said that the PRRT3-AS1 functions as a tumor promoter gene in prostate cancer and glioblastoma. Since it is revealed that this lncRNA is highly expressed in the progressive state of CLL, further studies may clarify the related pathways and novel targets. Moreover, downregulation of this lncRNA might promote spontaneous regression by repressing cell proliferation and inducing apoptosis.

According to Shi et al. (2020), lncRNA H1FX-AS1 is downregulated in cervical cancer, and low expression is correlated with poor prognosis and linked to tumor size as well as metastasis. In silico analysis predicted that the possible target of the H1FX-AS1 was the miR-324-3p, and it was binding and regulating the DACT1 [33]. Furthermore, the same research included overexpression studies, which revealed that the high expression levels of this lncRNA significantly reduced the proliferation and invasiveness, and activated the apoptosis pathways. H1FX-AS1, therefore, is identified as a tumor suppressor gene in cervical cancer [33]. On the contrary, high expression levels of lncRNA H1FX-AS1 are associated with a poor prognosis in gastric cancer. It is predicted that this lncRNA was related to epithelial to mesenchymal transition (EMT) and metastasis pathways [34]. Additionally, in-silico prediction analysis indicated several targets of the H1FX-AS1 lncRNA, such as H1FX, COPG1, and MIR6826 that can be used as potential therapeutic targets in gastric cancer [34]. Since H1FX-AS1 is upregulated in the progressive state of CLL, the precise roles of this lncRNA should be examined for CLL cells. Besides, considering the studies on gastric cancer, downregulation of H1FX-AS1 may trigger the spontaneous regression through inactivation of metastasis.

As the other 31 lncRNAs were novel transcripts, several examples could be considered by examining the sense protein-coding genes of the lncRNAs to reveal the significance and possible functions of these lncRNAs. Thereby, the specified lncRNAs and their sense protein-coding genes can be targeted for further analysis and treatment strategies.

The first one is the lncRNA PTPN22-AS1 that was upregulated in the spontaneous regression samples. It is known that the B cell receptor signaling is vital for the growth and survival of the leukemic cells [35]. These signaling pathways can be stimulated with diverse tyrosine kinases as well as phosphatases. PTPN22 is a protein tyrosine phospha-

tase specifically expressed in immune cells. It can function as an enhancer or suppressor of BCR and TCR signaling by regulating phosphorylation status [36]. According to studies, the PTPN22 gene is upregulated in CLL cells, and this upregulation results in attenuation of the apoptosis signals produced by the BCR and stimulation of the AKT activity that generates a survival signal [35]. This regulation reveals that the cancer cells which express autoreactive BCRs can escape from apoptosis by upregulating the PTPN22 gene. Therefore, downregulation of this gene through the lncRNA PTPN22-AS1 at multiple levels may trigger the spontaneous regression in CLL by eliminating autoreactive B cells in the context of immune tolerance. So, if the upregulation of this lncRNA can be induced by an external factor, it could be used as a therapeutic target for CLL cells.

The next lncRNA would be PCF11-AS1, and it is also upregulated in spontaneous regression. Alternative polyadenylation (APA) leads to transcription of diverse isoforms at the RNA 3' end that affects the functioning of encoded proteins. This mechanism needs multicomponent protein complexes, and one of the protein complexes is called CFII [37]. This complex comprises the PCF11 gene that is a cleavage and polyadenylation factor subunit, and regulates the transcription termination and RNA 3' end maturation. Studies show that this gene particularly regulates the alternative polyadenylation of the genes associated with the WNT pathway as an oncogene in neuroblastoma cells [38]. Accordingly, the downregulation of this gene results in diminished cell growth as well as invasiveness. Interestingly, one of the studies indicates that spontaneously regressed neuroblastomas express the PCF11 gene at low levels compared to highly progressive neuroblastomas [39]. Basically, downregulation of this gene at the epigenetic or post-transcriptional level through the lncRNA PCF11-AS1 may induce the spontaneous regression in CLL via ceasing the cell growth. Thus, the upregulation of this lncRNA can be used as a therapeutic strategy.

Another lncRNA is the SYNGAP1-AS1 which is upregulated in progressive samples. Mutant RAS genes stimulate the GTP-bound state which constitutively activates RAS signaling in metastatic cells [40]. RasGAPs inactivate RAS signaling by converting active GTP-bound into its inactive state [41]. SYNGAP1, also known as RASA5, is a member of the RasGAP family and functions as a tumor suppressor gene [42, 43]. Studies proved that the RASA5 gene is epigenetically disrupted in multiple cancer types by promoter methylation, and gain of function assays exposed that RASA5 expression leads to a reduction in metastasis by regulating EMT and cell stemness [43]. As it is known that the lncRNAs can regulate gene expression at the epigenetic level through methylation, this epigenetic silencing may be due to the lncRNA SYNGAP1-AS1. So, the upregulation of the SYNGAP1-AS1 can knock down this gene at the epigenetic level by methylating the promoter region leading to aggressive tumor progression in CLL patients. Even though further studies are essential, targeting this lncRNA may also trigger spontaneous regression through metastasis and cell stemness pathways, and enhance our understanding of this mechanism. As a result, the deduction that can be made from these exemplary lncRNAs is that the detailed examination of the identified novel transcripts can be used to reveal the mechanisms that lead to spontaneous regression, as well as to identify these lncRNAs as targets for small molecule inhibitors or activators so that the spontaneous regression mechanism can be triggered in other cancer types.

CONCLUSION

These findings suggest that regulation through the lncRNAs might have a major role in cells' fate, and their detailed examination can enlighten the way of discovery of the possible therapeutic targets. In prospective studies, each lncRNA should be investigated to perceive their functional pathways by over-expression and suppression studies in various cancer types

in order to understand their specific roles. Additionally, the following study which considers these findings should combine the protein-coding genes and non-protein coding genes. Since these two groups revealed significant results independently, their combination can lead to future selection and determination of particular pathways which affected each other.

This study had certain limitations that need to be overcome to conduct more efficient studies. Spontaneous regression samples had diverse biological backgrounds, which increased the demand for the detailed investigation of these samples. Furthermore, it was problematic to identify the significant pathways through differential gene expression algorithms due to this variability. Finally, as the amount of the samples was limited, more comprehensive research is required to verify these findings and make a distinctive interpretation.

AUTHOR INFORMATION

Corresponding Author
*Simay Dolaner (College of Engineering and Natural Sciences, Department of Molecular Biology and Genetics, Bahcesehir University, Istanbul, Turkey)
simay.dolaner@bahcesehir.edu.tr

Author Contributions
Simay Dolaner performed the data analysis and produced the figures and manuscript. Dr. Harpreet Kaur provided technical assistance with analysis and pipelines. Elia Brodsky and Dr. Mohit Mazumder added project direction and provided guidance. Dr. Julia Panov provided expert feedback for the project.

Competing Interests
The authors declare no competing financial and non-financial interests.

SUPPLEMENTARY DOCUMENTS
Supplementary Notes 1, which includes external sources and related links mentioned in the Methods section, can be found in the following link. https://www.dropbox.com/s/mlo17y1yjdxfekm

The list of the significantly differentially expressed protein-coding genes can be found in Supplementary Table 1 through the following link. https://www.dropbox.com/s/zc1oaryr3zyhxlf

Supplementary Table 2, which contains the name, Ensembl ID, and the possible functions of the 33 significant lncRNAs, can be found in the following link. https://www.dropbox.com/s/4p67jf4d54if2b8

ACKNOWLEDGMENTS

ABBREVIATIONS
CLL: Chronic Lymphocytic Leukemia
NCBI: National Center for Biotechnology Information
BCR: B Cell Receptor
TCR: T Cell Receptor
PCA: Principal Component Analysis
H-Cluster: Hierarchical Clustering
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genomes
lncRNA: Long Non-Coding RNA
PRRT3-AS1: Antisense to PRRT3
H1FX-AS1: Antisense to H1FX
PTPN22-AS1: Antisense to PTPN22
PCF11-AS1: Antisense to PCF11
SYNGAP1-AS1: Antisense to SYNGAP1
EMT: Epithelial to Mesenchymal Transition

REFERENCES
[1] American Cancer Society. Cancer Statistics Center. https://cancerstatisticscenter.cancer.

org. 2020

[2] A. A. Ferrando, C. López-Otín, Clonal evolution in leukemia. Nature Medicine, 23(10), 1135–1145. doi:10.1038/nm.4410 (2017).

[3] M. V. Haselager, A. P. Kater, E. Eldering, Proliferative Signals in Chronic Lymphocytic Leukemia; What Are We Missing? Frontiers in Oncology, 10(), 592205–. doi:10.3389/fonc.2020.592205 (2020).

[4] G. Fabbri, R. Dalla-Favera, The molecular pathogenesis of chronic lymphocytic leukaemia. Nature Reviews Cancer, 16(3), 145–162. doi:10.1038/nrc.2016.8 (2016).

[5] D. L. Longo, J. A. Burger, Treatment of Chronic Lymphocytic Leukemia. New England Journal of Medicine, 383(5), 460–473. doi:10.1056/NEJMra1908213 (2020).

[6] M. Kwok et al., Integrative analysis of spontaneous CLL regression highlights genetic and microenvironmental interdependency in CLL. Blood, 135 (6): 411–428. doi: https://doi.org/10.1182/blood.2019001262 (2020).

[7] J. Delgado, F. Nadeu, D. Colomer, E. Campo, Chronic lymphocytic leukemia: from molecular pathogenesis to novel therapeutic strategies. Haematologica, 105:xxx doi:10.3324/haematol.2019.236000 (2020).

[8] T. J. Kipps, M. Y. Choi, Targeted Therapy in Chronic Lymphocytic Leukemia. Cancer journal (Sudbury, Mass.), 25(6), 378–385. https://doi.org/10.1097/PPO.0000000000000416 (2019).

[9] F. Bosch, R. Dalla-Favera, Chronic lym-phocytic leukaemia: from genetics to treatment. Nat Rev Clin Oncol, 16, 684–701 https://doi.org/10.1038/s41571-019-0239-8 (2019).

[10] L. Fabris, J. Juracek, G. Calin, Non-Cod-ing RNAs as Cancer Hallmarks in Chronic Lymphocytic Leukemia. Interna-tional Journal of Molecular Sciences, 21(18), 6720–. doi:10.3390/ijms21186720 (2020).

[11] J. Fernandes, et al., Long Non-Coding RNAs in the Regulation of Gene Expres-sion: Physiology and Disease. Non-coding RNA, 5(1), 17. https://doi.org/10.3390/ncrna5010017 (2019).

[12] L. Nobili, D. Ronchetti, E. Taiana, A. Neri, Long non-coding RNAs in B-cell malig-nancies: a comprehensive overview. On-cotarget, 8(36), –. doi:10.18632/oncotarget.17303 (2017).

[13] M. Aprile, V. Katopodi, E. Leucci, V. Cos-ta, LncRNAs in Cancer: From garbage to Junk. Cancers, 12(11), 3220. https://doi.org/10.3390/cancers12113220 (2020).

[14] M. Gourvest, P. Brousset, M. Bousquet, Long Noncoding RNAs in Acute Myeloid Leukemia: Functional Characterization and Clinical Relevance. Cancers, 11(11), 1638. https://doi.org/10.3390/cancers11111638 (2019).

[15] X. Zhang, et al., Mechanisms and Functions of Long Non-Coding RNAs at Multi-ple Regulatory Levels. International jour-nal of molecular sciences, 20(22), 5573. https://doi.org/10.3390/ijms20225573 (2019).

[16] M. A. Diamantopoulos, P. Tsiakanikas, A. Scorilas, Non-coding RNAs: the riddle of the transcriptome and their perspectives in cancer. Annals of translational medi-cine, 6(12), 241. https://doi.org/10.21037/atm.2018.06.10 (2018).

[17] J. Gao, F. Wang, P. Wu, Y. Chen, Y. Jia, Aberrant LncRNA Expression in Leuke-mia. Journal of Cancer, 11(14), 4284–4296. https://doi.org/10.7150/jca.42093 (2020).

[18] Q. Tang, S. S. Hann, HOTAIR: An On-co-genic Long Non-Coding RNA in Human Cancer. Cellular Physiology and Biochem-istry, (), 893–913. doi:10.1159/000490131 (2018).

[19] Y. Chi, D. Wang, J. Wang, W. Yu, J. Yang, Long Non-Coding RNA in the Pathogene-sis of Cancers. Cells, 8(9), 1015. https://doi.org/10.3390/cells8091015 (2019).

[20] J. Iwakiri, G. Terai, M. Hamada, Compu-ta-tional prediction of lncRNA-mRNA inter-ac-tions by integrating tissue specificity in human transcriptome. Biology Direct 12, 15 https://doi.org/10.1186/s13062-017-0183-4 (2017).

[21] G. M. Cruz-Miranda, et al., Long Non-Coding RNA and Acute Leuke-mia. Inter-national journal of molecular sciences, 20(3), 735. https://doi.org/10.3390/ijms20030735 (2019).

[22] G. Arun, S. D. Diermeier, D. L. Spector, Therapeutic Targeting of Long Non-Coding RNAs in Cancer. Trends in molecu-lar medi-cine, 24(3), 257–277. https://doi.org/10.1016/j.molmed.2018.01.001 (2018).

[23] B. Langmead, S. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat Meth-ods 9, 357–359 (2012). https://doi.org/10.1038/nmeth.1923

[24] J. Lever, M. Krzywinski, N. Altman, Points of Significance: Principal component analy-sis. Nature Methods, 14(7), 641–642 (2017). doi:10.1038/nmeth.4346

[25] N. Altman, M. Krzywinski, Points of Sig-nificance: Clustering. Nature Meth-ods, 14(6), 545–546 (2017). doi:10.1038/nmeth.4299

[26] M. I. Love, W. Huber, S. Anders, Mod-erat-ed estimation of fold change and dis-per-sion for RNA-seq data with DESeq2.

Ge-nome Biol 15, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

[27] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Pow-erful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Meth-odological), 57: 289-300 (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

[28] A. Subramanian, et al., Gene set en-rich-ment analysis: A knowledge-based ap-proach for interpreting genome-wide ex-pres-sion profiles. Proceedings of the Na-tional Academy of Sciences, 102(43), 15545–15550. doi:10.1073/pnas.0506580102 (2005).

[29] M. V. Kuleshov, et al., Enrichr: a com-pre-hensive gene set enrichment analysis web server 2016 update. Nucleic acids re-search, 44(W1), W90–W97. https://doi.org/10.1093/nar/gkw377 (2016).

[30] P. Ghatalia, C. J. Morgan, G. Sonpavde, Meta-analysis of regression of advanced solid tumors in patients receiving placebo or no anti-cancer therapy in prospective trials. Crit-ical Reviews in Oncolo-gy/Hematology, 98(), 122–136. doi:10.1016/j.critrevonc.2015.10.018 (2016).

[31] L. Fan, H. Li, W. Wang, Long noncoding RNA PRRT3-AS1 silencing inhibits pros-tate cancer cell proliferation and promotes apopto-sis and autophagy. Experimental Physiology, (), EP088011–. doi:10.1113/ep088011 (2020).

[32] X. Li, M. Yutong, Immune-Related lncRNA Risk Signatures Predict Survival of IDH Wild-Type and MGMT Promoter Unmethylated Glioblastoma, BioMed Re-search Internation-al, vol. 2020, Article ID 1971284, 11 pages, https://doi.org/10.1155/2020/1971284 (2020).

[33] X. Shi, et al. A newly identified lncRNA H1FX-AS1 targets DACT1 to inhibit cervi-cal

cancer via sponging miR-324-3p. Can-cer Cell Int 20, 358 https://doi.org/10.1186/s12935-020-01385-7 (2020).

[34] E. Delshad, et al. In-silico identification of novel lncRNAs with a potential role in di-agnosis of gastric cancer, Journal of Bio-molecular Structure and Dynamics, (), 1–9. DOI: 10.1080/07391102.2019.1624615 (2019).

[35] R. Negro, et al., Overexpression of the au-toimmunity-associated phosphatase PTPN22 promotes survival of antigen-stimulated CLL cells by selectively activat-ing AKT. Blood, 119(26), 6278–6287. https://doi.org/10.1182/blood-2012-01-403162 (2012).

[36] R. Cubas, et al., Autoimmunity linked protein phosphatase PTPN22 as a target for cancer immunotherapy. Journal for Im-munoTherapy of Cancer, 8(2), e001439–. doi:10.1136/jitc-2020-001439 (2020).

[37] S. W. Yang, et al., A Cancer-Specific Ubiq-uitin Ligase Drives mRNA Alternative Pol-yadenylation by Ubiquitinating the mRNA 30 End Processing Complex, Molecular Cell, 77(6), 1206–1221.e7. https://doi.org/10.1016/j.molcel.2019.12.022 (2020).

[38] J. Nourse, S. Spada, S. Danckwardt, Emerging Roles of RNA 3′-end Cleavage and Polyadenylation in Pathogenesis, Di-agnosis and Therapy of Human Disorders. Biomole-cules, 10(6), 915. doi:10.3390/biom10060915 (2020).

[39] A. Ogorodnikov, et al., Transcriptome 3′end organization by PCF11 links alterna-tive polyadenylation to formation and neu-ronal dif-ferentiation of neuroblasto-ma. Nature commu-nications, 9(1), 5331. https://doi.org/10.1038/s41467-018-07580-5 (2018).

[40] M. H. Berryer, et al., Decrease of SYN-GAP1 in GABAergic cells impairs in-hibitory

synapse connectivity, synaptic in-hibition and cognitive function. Nature communications, 7, 13340. https://doi.org/10.1038/ncomms13340 (2016).

[41] O. Maertens, K. Cichowski, An expand-ing role for RAS GTPase activating proteins (RAS GAPs) in cancer. Advances in Bio-logical Regulation, 55(), 1–14. doi:10.1016/j.jbior.2014.04.002 (2014).

[42] M. Berryer, et al. Decrease of SYNGAP1 in GABAergic cells impairs inhibitory syn-apse connectivity, synaptic inhibition and cognitive function. Nature Communica-tions 7, 13340 https://doi.org/10.1038/ncomms13340 (2016).

[43] L. Li, et al., Tumor Suppression of Ras GTPase-Activating Protein RASA5 through Antagonizing Ras Signaling Perturbation in Carcinomas. iScience, 21, 1–18. doi: 10.1016/j.isci.2019.10.007 (2019).