A Possible Risk Gene for B-Cell Chronic Lymphocytic Leukemia: NRIP1

Yu Zheng[1], Xiaoyang Li[1], Qiusheng Chen[1], Soyeon Showman[2*], Lei Xu[2*]

[1] Department of Hematology, Ruijin Hospital affiliated to Shanghai Jiao Tong University, School of Medicine, Shanghai 200025, China;

[2] National Cancer Institute, 9000 Rockville Pike, Bethesda, Maryland 20892, USA.

Corresponding to: Lei Xu, Ph.D, National Cancer Institute, 9000 Rockville Pike, Bethesda, Maryland 20892, USA. Email: lei.xu2@nih.gov; Tel: 1-301-402-5535.

ABSTRACT

Background: The underlying mechanisms that cause B-cell chronic lymphocytic leukemia (B-CLL), the most common type of leukemia in adults, remain unclear. The aim of this study is to investigate the novel genetic risk of B-CLL through systematic literature review and meta-analysis.

Methods: A comprehensive search of electronic databases was completed using Illumina BioEngine. Twenty one B-CLL case/control bio-sets from four different studies were selected, including 195 B-CLL cases and 31 controls. The selected top B-CLL risk genes were further analyzed by integrating an online open source B-CLL genetic database. Pathway enrichment analysis (PEA) and network connectivity analysis (NCA) were conducted to identify the potential functional association between target genes and B-CLL.

Results: One novel gene (NRIP1) and two known genes (INPP5F and LEF1) were identified through the meta-analysis as top target genes for B-CLL. These genes play important roles within multiple B-CLL genetic pathways and are closely related to known B-CLL target genes. NCA results also revealed strong functional association between these genes and B-CLL.

Conclusion: This study identified known as well as novel B-CLL target genes and their functional pathways that involved in the B-CLL pathogenesis. Our results may provide new insights into the understanding of the genetic mechanisms of B-CLL.

Keywords: B-cell chronic lymphocytic leukemia, ResNet Database, Pathway enrichment analysis, Network connectivity analysis
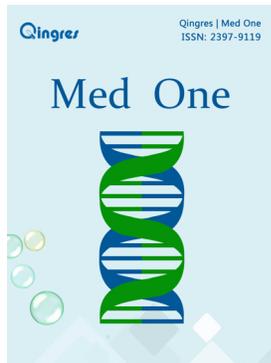
1 INTRODUCTION

Chronic lymphocytic leukemia (CLL) is the most frequent B cell leukemia in elderly patients [1], the average onset age of CLL are mostly over 50, with few occurances in children [2-4]. The cellular origin of CLL is still debated, although this information is critical to understanding its pathogenesis. It has been hypothesized that both environmental and gen etic factors play important roles in the development of CLL[5-7].

A large number of case-control studies and family-based genetic studies of B-CLL have been conducted to explore candidate genes for the disease [5, 6, 8-12]. Many studies have shown increased familial risk for CLL [13], and an ~8.5-fold increased relative risk in first-degree relatives[14]. In addition, genome-wide association (GWA) studies identified multiple CLL susceptibility loci [15] and novel genetic variants from familial CLL , however that was not seen in sporadic CLL[16, 17]. Furthermore, multiple modality genetic data from peripheral blood samples were employed to identify B-CLL genetic determinants [8-11]. These previously studies built a solid background for B-CLL genetic research, which could be leveraged for the discovery and evaluation of novel risk genes.

The risk estimates from individual studies often lack statistical power due to limited sample sizes and sample specificities in terms of phenotype characteristics. It is also difficult to come to a consistent conclusion as results are spread over a large number of independent studies. Therefore, a meta-analysis of multiple studies could provide a more effecitve assessment of the genetic risk factors of B-CLL.

A meta-analysis was conducted on four recent studies (2004-2012). Integrating a curated B-CLL genetic database (B-CLL_GD), the top genes from the study were further analyzed. The B-CLL_ GD database was constructed using a large scale literature knowledge database, Pathway Studio (PS) ResNet database. In recent years, the PS ResNet database has been widely used to study modeled relationships between proteins, genes, complexes, cells, tissues and diseases (http://pathwaystudio. gousinfo.com/Mendeley.html). Our study identified novel B-CLL genes and evaluated the effectiveness of integrating meta-analysis and PS ResNet database to identify and evaluate novel B-CLL risk genes.

2 METHODS AND MATERIALS

2.1 Genetic data selection

A systematic search of electronic databases was conducted using the Illumine BaseSpace Correlation Engine (http://www.illumina.com). Fig. 1 presents the diagram for the data selection. A search for the 'B-cell chronic lymphocytic leukemia' search result identified 28 B-CLL studies. Further filter criteria included: (1) whether the organism is Homo sapiens; (2) the data type is RNA expression; (3) the study is B-CLL case vs. healthy control study (or include case/control bio-sets). In total, 21 bio-sets (B-CLL case/control comparisons) from four studies satisfied the study selection criteria and were included in this systematic review and meta-analysis.

Fig. 1 Workflow diagram for meta-analysis data selection
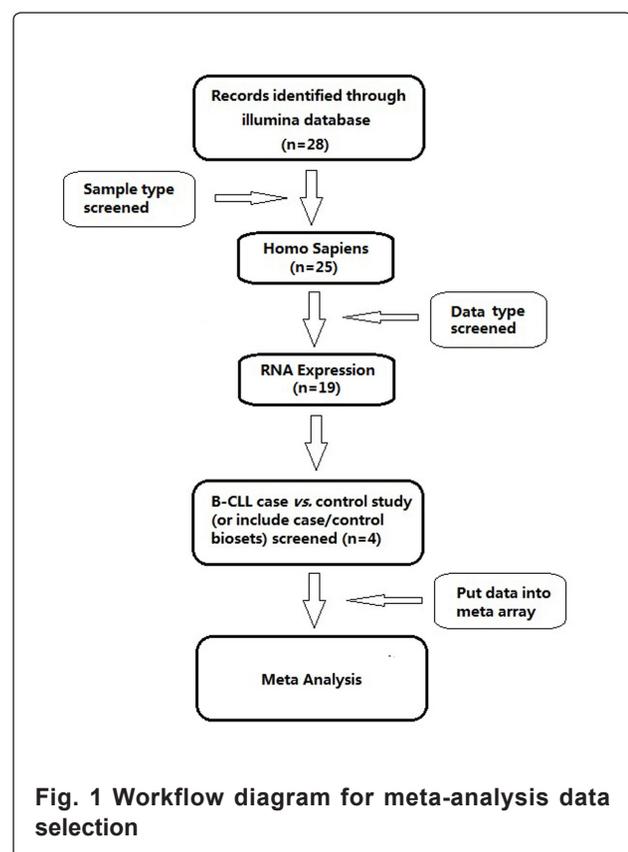
2.2 Genetic database B-CLL_GD

The B-CLL_GD is an online B-CLL targeted knowledge database available at 'Bioinformatics Database' (http://database.gousinfo.com/). The database is updated monthly or upon request. The current version of B-CLL_GD is composed of 753 B-CLL target genes (B-CLL_GD→Related Genes), 125 pathways (B-CLL_GD→Related Pathways), and 159 related diseases (B-CLL_GD→Related Diseases). The database also provides supporting

references for each B-CLL-Gene relation, including the titles and the sentences where the relation has been identified (B-CLL_GD→Ref for Related Genes). This information could be used to locate a detailed description of how a candidate gene/drug is related to B-CLL.

Using B-CLL_GD, further analysis of the B-CLL target genes from the meta-analysis were conducted, including identifying their related B-CLL pathways (B-CLL_GD →Related Pathways) and genes (B-CLL_GD →Related Genes). Here we defined two genes as functionally related if they play roles within same genetic pathway. Pathway enrichment analysis (PEA) was conducted using Pathway Studio to identify genetic pathways potentially linked to B-CLL [18]. The gene-disease relationships were identified using the network building module of Pathway Studio.

3 RESULTS

3.1 Selected Datasets

Using the previously identified selection criteria, 21 B-CLL case/control comparison bio-sets from four independent studies were retrieved and assessed (see in B-CLL_Meta→Selected Datasets). Only one of the four datasets contained the case/control study

(GSE19147) [8]. In this study, researchers analyzed T-cells isolated from CD3+ T-cells of patients with B-CLL, providing insights into the role of T-cells in B-CLL. The other three datasets contained separate case/control studies were available at NCBI GEO (ID: GSE2466, GSE26725 and GSE36907).

Datasets GSE26725 was designed to study the relationship between MYB (v-myb myeloblastosis viral oncogene homolog) and miR-155 host gene in B-CLL [9], It contains 4 separate case/control studies including: (1) B-CLL of diseased patients vs. normal CD19+ PBMCs (peripheral blood mononuclear cells of healthy subjects); (2) B-CLL of high-risk Rai stage 3-4 vs. normal CD19+ PBMCs; (3) B-CLL of intermediate-risk Rai stage 1-2 vs. normal CD19+ PBMCs; (4) B-CLL of low-risk Rai stage 0 vs. normal CD19+ PBMCs. The second dataset (GSE36907) contained two separate case/control studies to detect the cellular origin and pathogenesis of CLL [11], including: (1) B-CLL with mutated IgV vs. normal naive CD27- IdD+; (2) B-CLL with wildtype IgV vs. normal naive CD27- IdD+. The third identified datasets (GSE2466) contained 14 separate case/control studies, finding that a gene dosage effect may exert a pathogenic role in B-CLL, as well as genomic signature for the VH mutational status might be sex-related [10]. All the bio-sets studies are available at http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE2466 Statistics of the included bio-sets are presented in Table 1.

Table 1. Characteristics of the selected studies ordered by publication date

| First Author | Design | Cases | | Controls | | Country |
		N	Sources	N	Sources	
Haslinger et al., 2004	c-c	100	Boehringer Ingelheim Austria	11	Boehringer Ingelheim Austria	Austria
Eisele et al., 2009	c-c	25	University Hospital Essen	8	University Hospital Essen	Germany
Vargova et al., 2011	c-c	12	Charles University	5	Charles University	USA
Seifert et al., 2012	c-c	58	University of Duisburg-Essen	7	University of Duisburg-Essen	Germany

Note: c-c refers to case vs. control.

3.2 Meta-analysis results

The Meta-analysis results were deposited into the 'Bioinformatics Database' (http://database.gousinfo. com), under the title B-CLL_Meta. The top three genes (Score > 90) from the meta-analysis appear in Table 2., with more detailed statistics presented in B-CLL_Meta→Top 3 Genes. The full Meta-analysis results are presented in B-CLL_Meta→Full Gene List. A gene's score is defined by the meta-analysis Illumina BaseSpace Correlation Engine (http://www.illumina.com), which is based on the statistical significance and consistency of the gene across the queried bio-sets. The higher the score, the greater importance of the gene in the case/control comparison.

Table 2. The top 4 genes of the included in meta-analysis

Gene Correlation	EntrezGene ID	Score	Specificity	Associated Pathways	Gene Connectivity
INPP5F	22876	100	19/21	protein homodimerization activity (0042803).	78
NRIP1	8204	94	21/21	positive regulation of transcription, DNA-templated (0045941); receptor binding (0005102); negative regulation of transcription from RNA polymerase II promoter (0000122).	199
LEF1	51176	94	19/21	positive regulation of cell proliferation (0008284); negative regulation of apoptotic process (0006916); response to organic cyclic compound (0014070); negative regulation of transcription from RNA polymerase II promoter (0000122); positive regulation of transcription, DNA-templated (0045941); positive regulation of gene expression (0010628); positive regulation of cell migration (0030335); organ regeneration (0031100); transcription factor binding (0008134); negative regulation of transcription, DNA-templated (0061021); negative regulation of cysteine-type endopeptidase activity involved in apoptotic process (0043154); cellular response to cytokine stimulus (0071345)	342

Score: A gene's score is based on the statistical significance and consistency of the gene across the queried bio-sets. Specificity: A gene's specificity is the number of bio-sets in which the direction of a gene's regulation matches the selected filter. Associated Pathway: The known B-CLL related Pathways (B-CLL_Meta→Related Pathways) that contain the gene. GO ID is provided if any. Gene Connectivity: The number of known B-CLL related genes (B-CLL_GD→Related Genes) that connect with the target gene.

Out of the three genes listed in Table 2., only one gene was not included in the database B-CLL_GD (NRIP1), which suggested that it may be a novel B-CLL risk gene. Further study using the B-CLL_GD showed that this novel gene was enriched within multiple B-CLL target pathways and was connected to many other genes that were linked to B-CLL (Table 2., B-CLL_Meta → Related Pathways). Fig. 2 shows the 14 B-CLL pathways including these three genes. To note, two of these 14 pathways were among the top 10 B-CLL pathways (B-CLL_Meta → Related Pathways), including positive regulation of cell proliferation (0008284) and negative regulation of apoptotic process (0006916), as shown in Fig. 2.

Fig. 2 Fourteen B-CLL pathways where the three genes get enriched. The weight for a two-node edge is the number of shared genes by the two Pathways; The larger the size and brighter the color of a node, the larger the number of B-CLL candidate pathways including the gene.
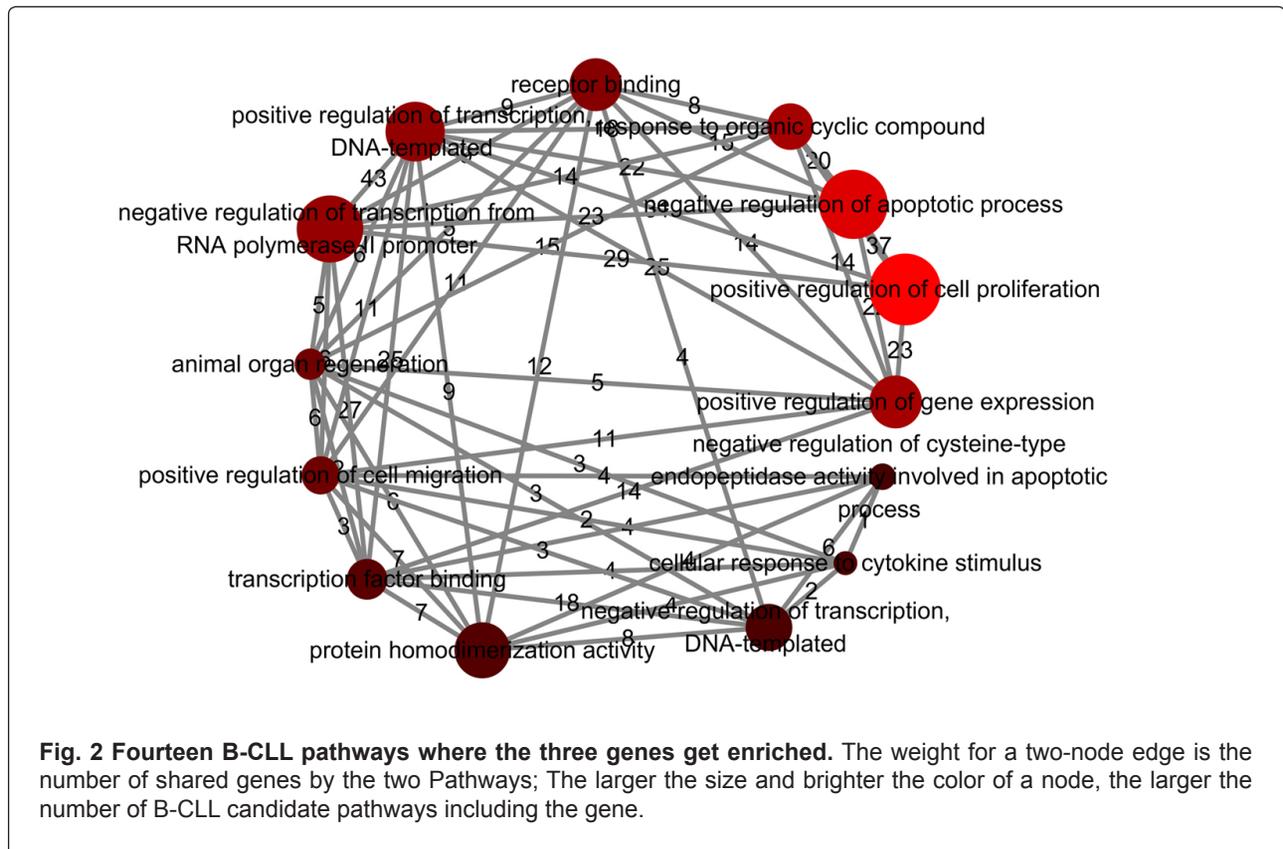
3.3 Network analysis

Additional functional network connectivity analysis (NCA) using PS showed that the novel gene from this meta-analysis (NRIP1) presents strong functional association with B-CLL. These genes influence the pathogenic development of B-CLL through multiple pathways (Fig. 3). Each relation (arrow) were supported by one or more references (see B-CLL_Meta → NRIP1).

Fig. 3 Network connectivity analyses between NRIP1 and B-CLL. The networks were generated using 'network building' module of Pathway Studio. For the definition of the entity types and relation types in the figure please refer to http://pathwaystudio.gousinfo.com/ResNetDatabase.html
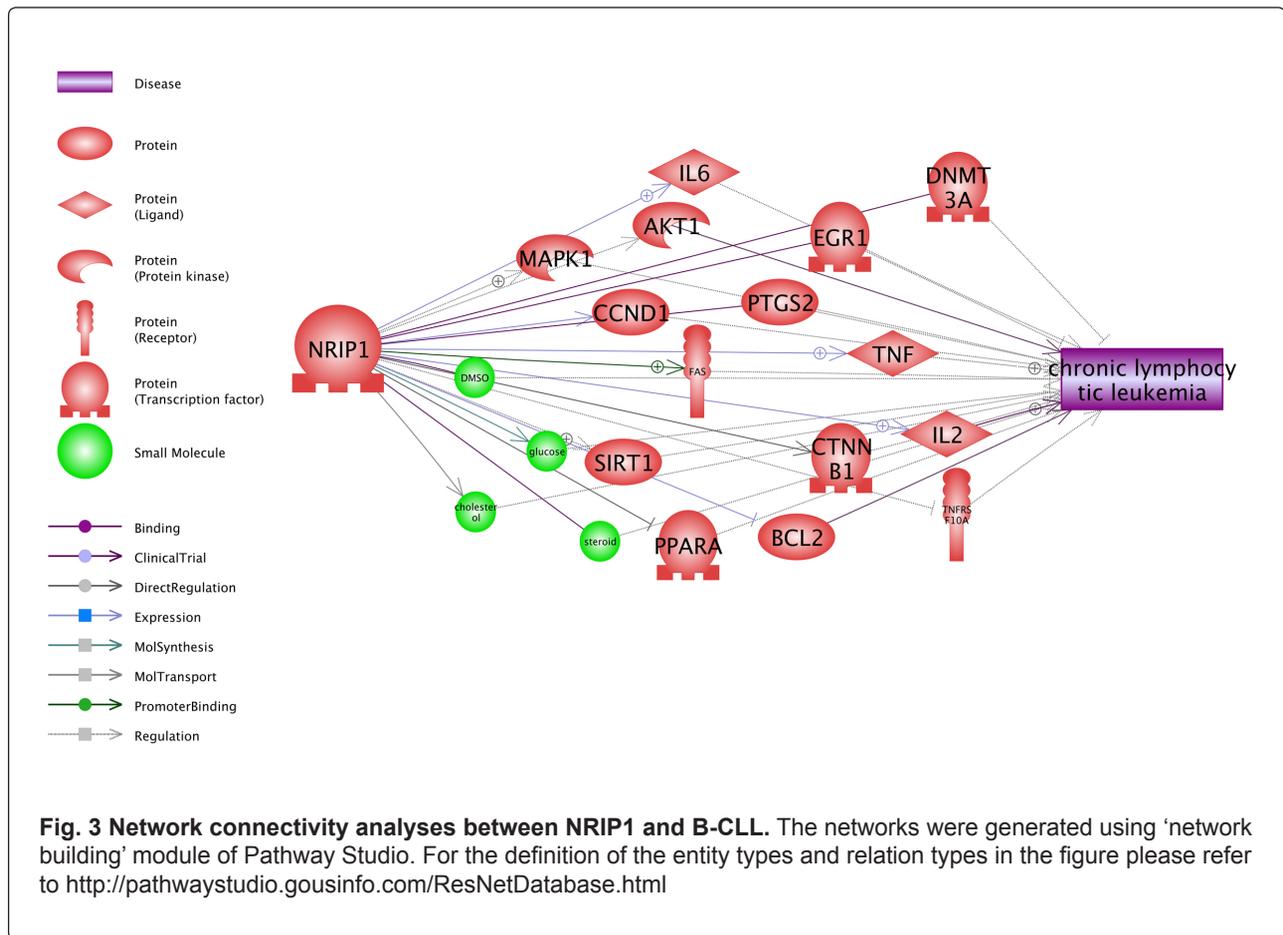
4 DISCUSSION

Although many previous genetic studies have been conducted to discovery genetic risk factors for B-CLL, combining the results from these separated studies by using meta-analysis could lead to a higher statistical power and a more robust point estimate for the disease. In this study, meta-analysis was performed on 21 B-CLL case/control bio-sets extracted from four recent studies. The B-CLL target genes from meta-analysis were sorted by gene score, which is based on the statistical significance and consistency of the gene across the queried bio-sets. Meta-analysis results suggested three top risk genes (INPP5F, NRIP1 and LEF1) for B-CLL (Score > 90), of which one is novel according to a recently updated database B-CLL_GD. Further analyses were conducted to study the possible correlation between B-CLL and these three genes, especially the novel gene.

Analysis using B-CLL_GD showed that the two known B-CLL target genes, INPP5F, and LEF1, are among the top B-CLL_GD genes , which is supported by (see B-CLL_GD → Related Genes).

Results from PEA showed that these three known genes and the one novel gene (NRIP1) are enriched within multiple B-CLL pathways (B-CLL_Meta → Related Pathways) and linked to hundreds of other B-CLL genes. These results support the relationship between the identified genes and B-CLL.

Additional network connectivity analysis (NCA) revealed multiple possible functional associations between B-CLL and the novel gene (Fig. 3). It has been shown that overexpression of NRIP1 could increase the mRNA levels of TNF-alpha [19], while TNF-alpha plays an important role in the progression of B-CLL [20]. TNF-alpha promotes the proliferation of malignant cell clones. Therefore inhibition of TNF may have therapeutic applications in CLL [21]. This suggests that NRIP1 may play a role in the development of B-CLL through a NRIP1 → TNF → B-CLL pathway.

Many studies indicated that INPP5F and LEF1 may have a role in the therapeutic strategies of CLL [22, 23]. More potential connections between these

genes and B-CLL could be identified from the B-CLL_ Meta database (see B-CLL_Meta → NRIP1, INPP5F and LEF1), which is available in the open source 'Bioinformatics Database' (http://database.gousinfo. com).

There are several limitations of this meta-analysis. The number of B-CLL patients and healthy controls were not well match (195 B-CLL cases and 31 controls). The unbalanced case/control comparison may influence the accuracy of the results. Additionally, due to the limitation of the space, we mainly focused on the most significant genes (Gene Score > 90). Genes with less significance from this meta-analysis may have potential linkages to B-CLL. A full list of the top 100 genesused in this meta-analysis is presented in the database B-CLL_Meta.

In summary, this meta-analysis supported the correlation between two genes (INPP5F and LEF1) and B-CLL, and revealed one novel potential risk gene (NRIP1). Network analysis supported the meta-analysis results and identified potential functional pathways and mechanisms, wherein which these genes play important roles on B-CLL. The findings in this study provide new insights into the current genetics research on B-CLL.

CONFLICT OF INTERESTS

Authors claim no conflict of interests.

REFERENCES

1. Zenz T, Mertens D, Küppers R, Döhner H, Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat Rev Cancer. 2010; 10: 37-50.

2. Mauro FR, Foa R, Giannarelli D, Cordone I, Crescenzi S, Pescarmona E, Sala R, Cerretti R, Mandelli F. Clinical characteristics and outcome of young chronic lymphocytic leukemia patients: a single institution study of 204 cases. Blood. 1999; 94(2): 448-454.

3. Montserrat E, Gomis F, Vallespi T, Rios A, Romero A, Soler J, Alcala A, Morey M, Ferran C, Diaz-Mediavilla J, Flores A, Woessner S, Batlle J, Gonzalez-Aza C, Rovira M, Reverter J, Rozman C. Presenting features and prognosis of chronic lymphocytic leukemia in younger adults. Blood. 1991; 78: 1545.

4. Demir HA, Bayhan T, Üner A, Kurtulan O, Karakuş E, Emir S, Ozyoruk D, Ceylaner S. Chronic lymphocytic leukemia in a child: a challenging diagnosis in pediatric oncology practice. Pediatr Blood Cancer. 2014; 61(5): 933-935.

5. Caporaso N, Marti GE, Goldin L. Perspectives on familial chronic lymphocytic leukemia: genes and the environment. Semin Hematol. 2004; 41(3): 201-206.

6. Goldin LR, Slager SL, Caporaso NE. Familial Chronic Lymphocytic Leukemia. Curr Opin Hematol. 2010; 17(4): 350-355.

7. Puiggros A, Blanco G, Espinet B. Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go. Biomed Res Int. 2014; 2014: 435983.

8. Eisele L, Prinz R, Klein-Hitpass L, Nückel H, Lowinski K, Thomale J, Moeller LC, Duhrsen U, Durig J. Combined PER2 and CRY1 expression predicts outcome in chronic lymphocytic leukemia. Eur J Haematol. 2009; 83(4): 320-327.

9. Vargova K, Curik N, Burda P, Basova P, Kulvait V, Pospisil V, Savvulidi F, Kokavec J, Necas E, Berkova A, Obrtlikova P, Karban J, Mraz M, Pospisilova S, Mayer J, Trneny M, Zavadil J, Stopka T. MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. Blood. 2011; 117(14): 3816-3825.

10. Haslinger C, Schweifer N, Stilgenbauer S, Döhner H, Lichter P, Kraut N, Stratowa C, Abseher R. Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. J Clin Oncol. 2004; 22(19): 3937-3949.

11. Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Dürig J, Ralf K. Cellular origin and pathophysiology of chronic lymphocytic leukemia. J Exp Med. 2012; 209(12): 2183-2198.

12. Kostopoulos IV, Tsakiridou AA, Pavlidis D, Megalakaki A, Papadhimitriou SI. Familial chronic lymphocytic leukemia in two siblings with ATM/13q14 deletion and a similar pattern of clonal evolution. Blood Cancer J. 2015; 5(7): e322.

13. Albright F, Teerlink C, Werner TL, Cannon-Albright LA. Significant evidence for a heritable contribution to cancer predisposition: a review of cancer familiality by site. BMC Cancer. 2012; 12: 138.

14. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. Haematologica. 2009; 94: 647-653.

15. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJ, Holroyd A, Wang Y, Sunter NJ, Mansouri L, Juliusson G, Smedby KE, Roos G, Jayne S, Majid A, Dearden C, Hall AG, Mainou-Fowler T, Jackson GH, Summerfield G, Harris RJ, Pettitt AR, Allsup DJ, Bailey JR, Pratt G, Pepper C, Fegan C, Rosenquist R, Catovsky D, Allan JM, Houlston RS. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2014; 46(1): 56-60.

16. Slager SL, Caporaso NE, De SS, Goldin LR. Genetic susceptibility to chronic lymphocytic leukemia. Semin Hematol. 2013; 50: 296-302.

17. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, Sullivan K, Vijayakrishnan J, Wang Y, Pittman AM, Sunter NJ, Hall AG, Dyer MJ, Matutes E, Dearden C, Mainou-Fowler T, Jackson GH, Summerfield G, Harris RJ, Pettitt AR, Hillmen P, Allsup DJ, Bailey JR, Pratt G, Pepper C, Fegan C, Allan JM, Catovsky D, Houlston RS. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet. 2008; 40(10): 1204-1210.

18. Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio--the analysis and navigation of molecular networks. Bioinformatics. 2003; 19(16): 2155-2157.

19. Zhang L, Chen Y, Yue Z, He Y, Zou J, Chen S, Liu M, Chen X, Liu Z, Liu X, Feng X, Li M, Liu P. The p65 subunit of NF-κB involves in RIP140-mediated inflammatory and metabolic dysregulation in cardiomyocytes. Arch Biochem Biophys. 2014; 554: 22-27.

20. Mainou-Fowler T, Miller S, Proctor SJ, Dickinson AM. The levels of TNF alpha, IL4 and IL10 production by T-cells in B-cell chronic lymphocytic leukaemia (B-CLL). Leuk Res. 2001; 25(2): 157-163.

21. Yan L, Anderson GM, DeWitte M, Nakada MT. Therapeutic potential of cytokine and chemokine antagonists in cancer therapy. Eur J Cancer. 2006; 42(6): 793-802.

22. Palermo G, Maisel D, Barrett M, Smith H, Duchateau-Nguyen G, Nguyen T, Yeh RF, Dufour A, Robak T, Dornan D, Weisser M. Gene expression of INPP5F as an independent prognostic marker in fludarabine-based therapy of chronic lymphocytic leukemia. Blood Cancer J. 2015; 5: e353.

23. Tandon B, Peterson L, Gao J, Nelson B, Ma S, Rosen S, Chen YH. Nuclear overexpression of lymphoid-enhancer-binding factor 1 identifies chronic lymphocytic leukemia/small lymphocytic lymphoma in small B-cell lymphomas. Mod Pathol. 2011; 24(11): 1433-1443.