



## OPEN ACCESS

## EDITED BY

Festenstein,  
Richard, Imperial College London,  
United Kingdom

## REVIEWED BY

Wenan Chen,  
St. Jude Children's Research Hospital,  
United States  
Richa Batra,  
Weill Cornell Medicine, United States

## \*CORRESPONDENCE

Ying Gao,  
✉ [yinggao@sinh.ac.cn](mailto:yinggao@sinh.ac.cn)

## SPECIALTY SECTION

This article was submitted to  
Epigenomics and Epigenetics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 28 September 2022

ACCEPTED 14 December 2022

PUBLISHED 11 January 2023

## CITATION

Zhang Q, Gao Y, Lin S, Goldin LR,  
Wang Y, Stevenson H, Edelman DC,  
Killian K, Marti G, Meltzer PS, Xiang S and  
Caporaso NE (2023), Genome-wide  
DNA methylation profiling in chronic  
lymphocytic leukaemia.  
*Front. Genet.* 13:1056043.  
doi: 10.3389/fgene.2022.1056043

## COPYRIGHT

© 2023 Zhang, Gao, Lin, Goldin, Wang,  
Stevenson, Edelman, Killian, Marti,  
Meltzer, Xiang and Caporaso. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Genome-wide DNA methylation profiling in chronic lymphocytic leukaemia

Qiuyi Zhang<sup>1</sup>, Ying Gao<sup>1,2\*</sup>, Shuchun Lin<sup>1</sup>, Lynn R. Goldin<sup>2</sup>,  
Yonghong Wang<sup>3</sup>, Holly Stevenson<sup>3</sup>, Daniel C. Edelman<sup>3</sup>,  
Keith Killian<sup>3</sup>, Gerald Marti<sup>4</sup>, Paul S. Meltzer<sup>3</sup>, Song Xiang<sup>5</sup> and  
Neil E. Caporaso<sup>2</sup>

<sup>1</sup>CAS Key Laboratory of Nutrition, Metabolism and Food Safety, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, <sup>2</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States, <sup>3</sup>Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States, <sup>4</sup>Lymphoid Malignancies Section, Hematology Branch, NHLBI, National Institutes of Health, Bethesda, MD, United States, <sup>5</sup>Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

**Background:** DNA methylation aberrations are widespread among the malignant B lymphocytes of patients with chronic lymphocytic leukaemia (CLL), suggesting that DNA methylation might contribute to the pathogenesis of CLL.

**Aim:** We aimed to explore the differentially methylated positions (DMPs) associated with CLL and screen the differentially methylated and expressed genes (DMEGs) by combining public databases. We aimed to observe the direction of each DMEG in CLL based on the DMPs in the promoter and the body region respectively to narrow down DMEGs. We also aimed to explore the methylation heterogeneity of CLL subgroups and the effect of B cells maturation on CLL.

**Methods:** In this population-based case control study, we reported a genome-wide DNA methylation association study using the Infinium HumanMethylation450 BeadChip, profiling the DNA methylation of CD19<sup>+</sup> B Cells from 48 CLL cases and 28 healthy controls. By integrating methylation data and expression data from public databases, gene sets were jointly screened, and then the relationship between methylation sites in promoter and body region and expression of each gene was explored. In addition, support vector machine (SVM) classification algorithm was used to identify subgroups of CLL cases based on methylation pattern, and the effect of B-cell differentiation related methylation sites on CLL-related sites was observed.

**Abbreviations:** CLL, chronic lymphocytic leukemia; IGHV, immunoglobulin heavy chain variable region; TSGs, tumor suppressor genes; MBL, monoclonal B-cell lymphocytosis; DMPs, differentially methylated positions; DMRs, differentially methylated regions; CNAs, copy number aberrations; PCA, principal component analysis; DEGs, differentially expressed genes; DMEGs, differentially methylated and expressed genes; SVM, support vector machine.

**Results:** We identified 34,797 DMPs related to CLL across the genome, most of which were hypomethylated; the majority were located in gene body regions. By combining these DMPs with published DNA methylation and RNA sequencing data, we detected 26,244 replicated DMPs associated with 1,130 genes whose expression were significantly different in CLL cases. Among these DMEGs, nine low expressed DMEGs were selected with hypermethylated in promoter and hypomethylated in body region, and 83 high expressed DMEGs were selected with both hypomethylated in promoter and body region. The 48 CLL cases were divided into 3 subgroups based on methylation site by SVM algorithm. Over 92% of CpGs associated with B cell subtypes were found in CLL-related DMPs.

**Conclusion:** The DNA methylation pattern was altered across the genome in CLL patients. The methylation of *ZAP70*, *FMOD*, and *ADAMTS17* was significantly different between CLL cases and controls. Further studies are warranted to confirm our findings and identify the underlying mechanisms through which these methylation markers are associated with CLL.

#### KEYWORDS

DNA methylation, chronic lymphocytic leukaemia (CLL), illumina 450 K beadchip methylation array, CD19+ B cells, differentially methylated and expressed genes (DMEGs)

## Introduction

Chronic lymphocytic leukaemia (CLL) is characterized by the accumulation of B lymphocytes in the peripheral blood, bone marrow, and secondary lymphoid tissues, and it is the most common adult leukaemia in Western countries (Mansouri et al., 2018). It was estimated that 20,720 Americans would be diagnosed with CLL and 3,930 patients would die of CLL in 2019 (Siegel et al., 2019). At present, there is no efficient early treatment available for this haematologic malignancy, and in most cases, treatments start only when clinical symptoms develop. A clear understanding of the disease pathogenesis could facilitate the prevention and better treatment of CLL.

CLL is a clinically heterogeneous disease, and CLL subtypes differ in terms of B Cell activation, maturation, and cellular subgroup (Chiorazzi et al., 2005). Biological and genetic characteristics such as *ZAP70* expression, immunoglobulin heavy chain variable region (*IGHV*) mutation, and cytogenetic abnormality (del17p) can be linked to CLL prognosis (Parikh and Shanafelt, 2016). A meta-analysis of genome-wide association studies has defined genetic components of CLL aetiology (Berndt et al., 2016). However, the genetic variations identified for CLL in high-risk families can only partially explain CLL risk.

DNA methylation perturbations are common in cancer genomes and even in precancerous normal tissue, playing a

pivotal role in the early tumorigenic process (Feinberg, 2018). DNA methylation serves as an intermediate between gene and environmental interactions, study of DNA methylation might explain epigenetic mechanism of CLL pathogenesis. The methylation pattern in CLL patients of peripheral blood and lymph node samples is relatively stable over time, suggesting that aberrant methylation might be involved in the early phase of leukaemogenesis (Cahill et al., 2013). Studies have suggested that DNA methylation changes not only distinguished CLL and healthy control but also exhibited large interpatient differences in CLL, which could be used to distinguish CLL subtypes with different *IGHV* statuses (Supplementary Table S1). It has been reported that the transcriptional start sites of many tumour suppressor genes (TSGs) are located in CpG islands (CGIs), which are usually hypermethylated in tumours (Herman and Baylin, 2003). Besides the gene silence by hypermethylation of promoter region, altered methylation observed in the gene bodies could also influence gene expression (Kulis et al., 2012; Yang et al., 2014). However, the relationship between DNA methylation and gene expression level in CLL was not consistent, more epigenome-wide association studies, integrating expression data, are needed to explore the role of DNA methylation in the CLL pathogenesis. In addition, CLL heterogeneity and the cell types caused by B Cell differentiation are considered to bias the pathogenesis of CLL. In this study, we also explored the influence of CLL subtypes and the methylation changes related to B Cell differentiation on the methylation change in the case-control screening.

## Materials and methods

### Study population

Participants were recruited from the NCI B-CLL Registry, which has been described previously (Ishibe et al., 2001). Demographic information was collected from surveys and questionnaires. Clinical information was obtained from private physicians, patient interviews, hospital records, death certificates, and NIH clinic visits.

### Sample preparation

Fifty CLL patients, two MBL patients, and 29 unrelated healthy controls were included in the current study. The peripheral blood samples collected closest to diagnosis were used for the assay. B lymphocytes were isolated from cryopreserved peripheral blood lymphocytes by magnetic cell separation (ZenBio Advanced Cell-based Solutions and Services, North Carolina) using a CD19<sup>+</sup> antibody (Miltenyi Biotec, Bergisch Gladbach, Germany). The cell purity of the isolated samples was evaluated by flow cytometry using propidium iodide for live/dead cell discrimination with CD45/CD19 antigens. Sample aliquots of CD19<sup>+</sup> B Cells with purity greater than 90% were selected for DNA extraction.

### Microarray-based DNA methylation assay

DNA from the isolated CD19<sup>+</sup> B Cells was extracted with a DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). For each sample, 500 ng of genomic DNA was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA), and 125 ng bisulfite-converted DNA was used to evaluate methylation levels with Infinium HumanMethylation450K BeadChip (Illumina, Inc, San Diego, CA) according to the manufacturer's protocol.

### 450k DNA methylation data analysis

Raw 450K BeadChip methylation data were extracted from *GenomeStudio* (Illumina, Inc.). Two samples were excluded due to the high average detection *p*-values (over 1% CpG sites with detection *p* > 0.05). One sample was excluded due to the misclassification of sex by multidimensional scaling (MDS) of CpG sites on the X chromosome. CpG sites were excluded if 1) over 1% of samples had a detection *p* > 0.05; 2) over 5% of samples had bead counts <3; 3) they were located on X or Y chromosomes; or 4) they were annotated with SNPs and cross-reactive probes (Zhou et al., 2017). Data normalization was performed using *dasen* and *beta-mixture* quantile

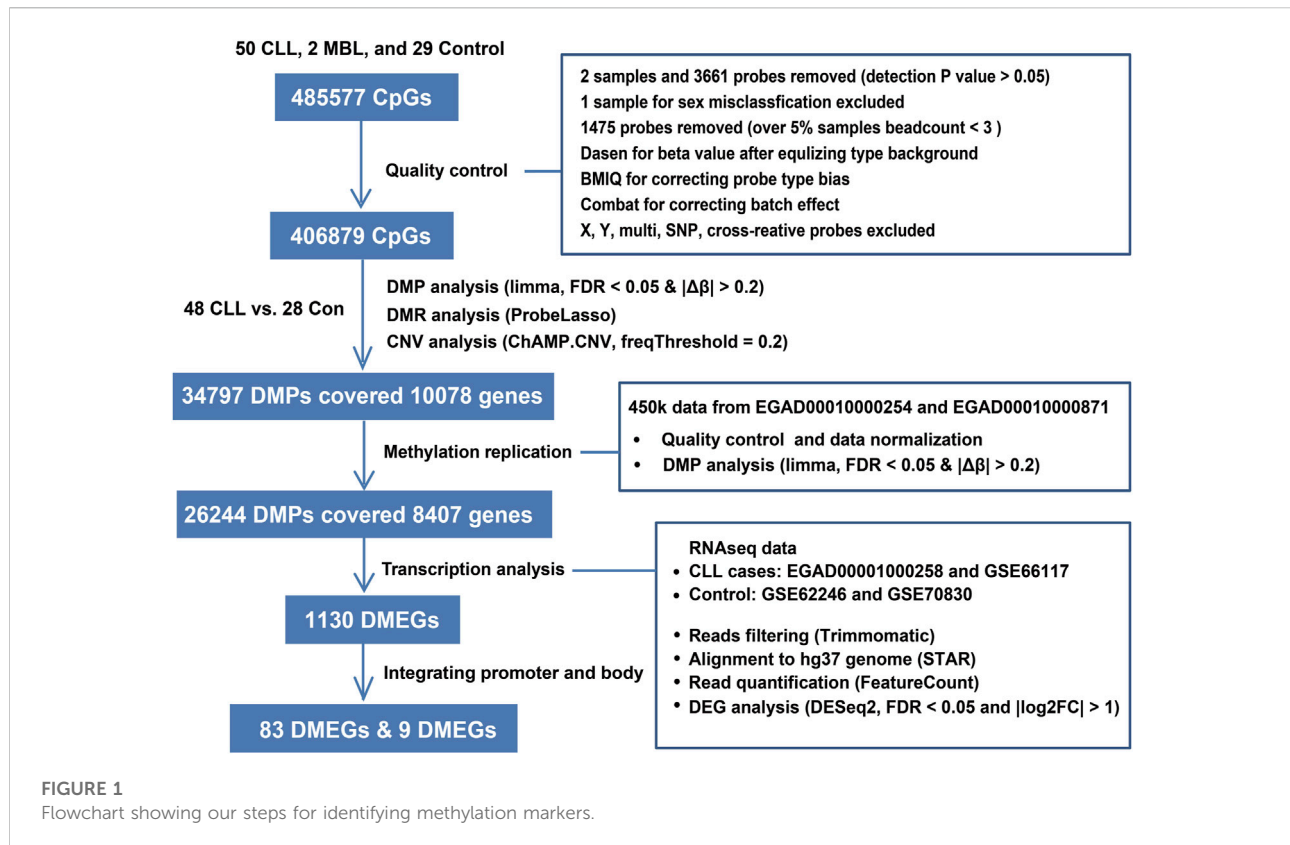
normalization. Batch effects were corrected by *ComBat* (Johnson et al., 2007). Altogether, 406,879 autosomal probes from 76 samples (48 CLL cases and 28 controls) were included in the final analysis. Two MBL samples were used for preliminary exploration.

Linear regression adjusted for age and sex comparing CLL cases and controls was performed using the R package *limma*. Beta values were transformed to M-values for the linear regression model. *p* values were corrected for multiple testing by Benjamini–Hochberg FDR (q-value). Probes with methylation differences over 0.2 ( $|\Delta\beta| > 0.2$ ) between cases and controls and FDRs less than 0.05 (q-value < 0.05) were identified as differentially methylated positions (DMPs). The chi-square goodness of fit test was used to compare the distribution of DMPs with genomic background CpGs. Differentially methylated regions (DMRs) containing more than seven DMPs with lengths over 50 bp were defined by *ProbeLasso* (Butcher and Beck, 2015). The minimum distance between two adjacent DMRs was 1,000 bp.

Copy number aberrations (CNAs) were profiled using the total signal intensity of probes by the Infinium array (Feber et al., 2014). The copy gain or loss on each chromosome was defined using a threshold of 0.2.

### Integration with public DNA methylation and transcription data

The publicly available datasets of DNA methylation and RNA transcription used in this study are shown (Supplementary Table S2A). DNA methylation data of Illumina 450K methylation array from the European Genome-Phenome Archive (EGA), EGAD00010000254 and EGAD00010000871 were requested. The CD19<sup>+</sup> B Cell samples of 329 CLL cases (139 from EGAD00010000254, 190 from EGAD00010000871) and 21 healthy controls (14 from EGAD00010000254, seven from EGAD00010000871) were finally selected from the dataset of EGAD00010000871 after quality control (Supplementary Table S2B), and their Methylation data were extracted and analysed with the procedure described above for replication. DMPs identified with methylation changes in the same direction as our data were defined as replicated DMPs. RNA sequencing data, including 145 CLL samples (98 from EGAD00010000254, 47 from GSE66117) and seven control samples (2 from GSE62246, five from GSE70830) from the Gene Expression Omnibus, were requested (Supplementary Table S2C). RNA sequencing data were evaluated and merged by *FastQC* and *multiQC*. *Trimmomatic* (Bolger et al., 2014) was used for read filtering. RNA reads were aligned to the human reference genome hg19 (UCSC) using *STAR* (Dobin et al., 2013) and were assembled for quantification using *featureCounts* to generate counts (Liao et al., 2014). Altogether, RNA expression data of



CD19<sup>+</sup> B Cells from 144 CLL and seven controls were included for downstream analysis after filtering one case due to the total counts of genes less than 10 000. Genes with expression differences larger than 2-fold ( $|\log_2\text{fold change}| > 1$ ) and Benjamini–Hochberg adjusted  $p$  values less than 0.05 ( $q\text{-value} < 0.05$ ) were defined as differentially expressed genes (DEGs) using DESeq2 (Love et al., 2014). DEGs covered by replicated DMPs were defined as differentially methylated and expressed genes (DMEGs) by combining methylation data in our study and DEG lists analysed with public datasets. Gene Ontology (GO) enrichment analysis and KEGG pathway analysis were performed on DMEGs using *clusterProfiler* (Yu et al., 2012) for enriched biological process (BP), molecular function (MF), and cell component (CC) terms and KEGG pathways. The significantly enriched GO terms and KEGG pathways were identified as those with  $q$ -values less than 0.05.

Since DMPs in promoter and gene body region could be associated with gene expression (Kulis et al., 2012), the relationship of DMPs in different regions (promoter and body region) and expression was inconsistent. We conducted integrative analysis by combining DNA methylation data in our study and expression data from public datasets. We defined the promoter region as the combination of TSS1500, TSS200, and the first exon (Jiao et al., 2014). To narrow down the DMEG gene list, we selected the DMPs in the promoter region

had a negative direction with expression, and the DMPs in the gene body region had a positive or negative direction with expression; and the overlap genes of these two types of DMPs might be the potential methylation regulated DEGs.

A flow chart describing these analysis steps is shown in Figure 1.

## Methylation of CLL subtypes

Based on IGHV status, CLL can be classified as M-CLL (IGHV mutated) or U-CLL (IGHV unmutated). CLL can also be classified as naive B-cell-like (n-CLL), intermediate (i-CLL), and memory B-cell-like CLL (m-CLL) based on methylation level, where n-CLL and m-CLL shared high similarity with U-CLL and M-CLL, respectively. We applied the support vector machine (SVM) model ( $C = 10$ ,  $\gamma = 0.01$ ) reported to classify the 48 CLL patients in our study into subgroups of m-CLL, n-CLL, and i-CLL based on the five CpGs (cg00869668, cg03462096, cg09637172, cg11472422, cg17014214) selected with 133 CLL cases in the ICGC database reported by Queiros et al. (2015). Linear regression adjusted for sex and age was fitted to find the DMPs between n-CLL and m-CLL. Subtype-related DMPs were compared to CLL-related DMPs to evaluate the impact of CLL methylation heterogeneity on DMP identification.

**TABLE 1 Characteristics of the 76 subjects in this study.<sup>a</sup>**

Characteristics	CLL (N = 48)	Control (N = 28)	<i>p</i> <sup>b</sup>
Sex			
Female	29 (60%)	19 (68%)	0.69
Male	19 (40%)	9 (32%)	
Age <sup>c</sup>	61 (55–71.5)	57 (42–66)	0.015
WBC (10 <sup>9</sup> /L)	29.2 (22.1–53.7)	7.4 (6.1–8.5)	1.2e-7
ALC (10 <sup>9</sup> /L)	21.0 (13.9–36.4)	2.0 (1.6–2.4)	3.0e-5
Rai stage <sup>d</sup>			
Low (0)	21 (46%)		
Intermediate (I/II)	22 (48%)		
High (III/IV)	3 (6%)		

<sup>a</sup>Two MBL cases were used to explore the tumor progression related issue assuming MBL, as intermediate point between healthy to CLL.

<sup>b</sup>Continuous variables were expressed as medians (IQRs), and categorical variables are expressed as frequencies (%). *p* values were calculated by using a wilcoxon rank sum test for continuous variables and chi-square test for categorical variables.

<sup>c</sup>Age was recorded when blood samples were collected. Age of three subjects in the control group was not available.

<sup>d</sup>Rai stage for two subjects in the CLL group were not available.

Abbreviations: WBC, white blood cells; ALC, absolute lymphocyte count.

## Methylation of normal B Cell subtypes

To explore the impact of normal B Cell differentiation on CLL-related DMPs detection, DNA methylation data of 14 samples of CD19<sup>+</sup> B Cell mixture and three samples of each purified B Cell subtypes, including naive B Cells (NBCs), CD5<sup>+</sup> naive B Cells (CD5<sup>+</sup> NBCs), class-switched memory B Cells (csMBCs), and non-class-switched memory B Cells (ncsMBCs) were requested from EGAD00010000254 dataset (Kulis et al., 2012). A linear regression was fitted to identify DMPs between CD19<sup>+</sup> B Cells (N = 14) and these four subtypes (N = 3 each). DMPs identified in these four comparisons were combined as normal B Cell differentiation related DMPs. After removing these DMPs from the comparison of normal B Cells to the total CpG sites after quality control, a linear model was fitted on CLL patients and healthy controls to find CLL-related DMPs and DMEGs, which should be independent from B Cell maturation. Gene enrichment analysis was performed on these genes to identify relevant biological processes other than B Cell differentiation.

## Results

### Characteristics of 48 CLL patients and 28 healthy controls

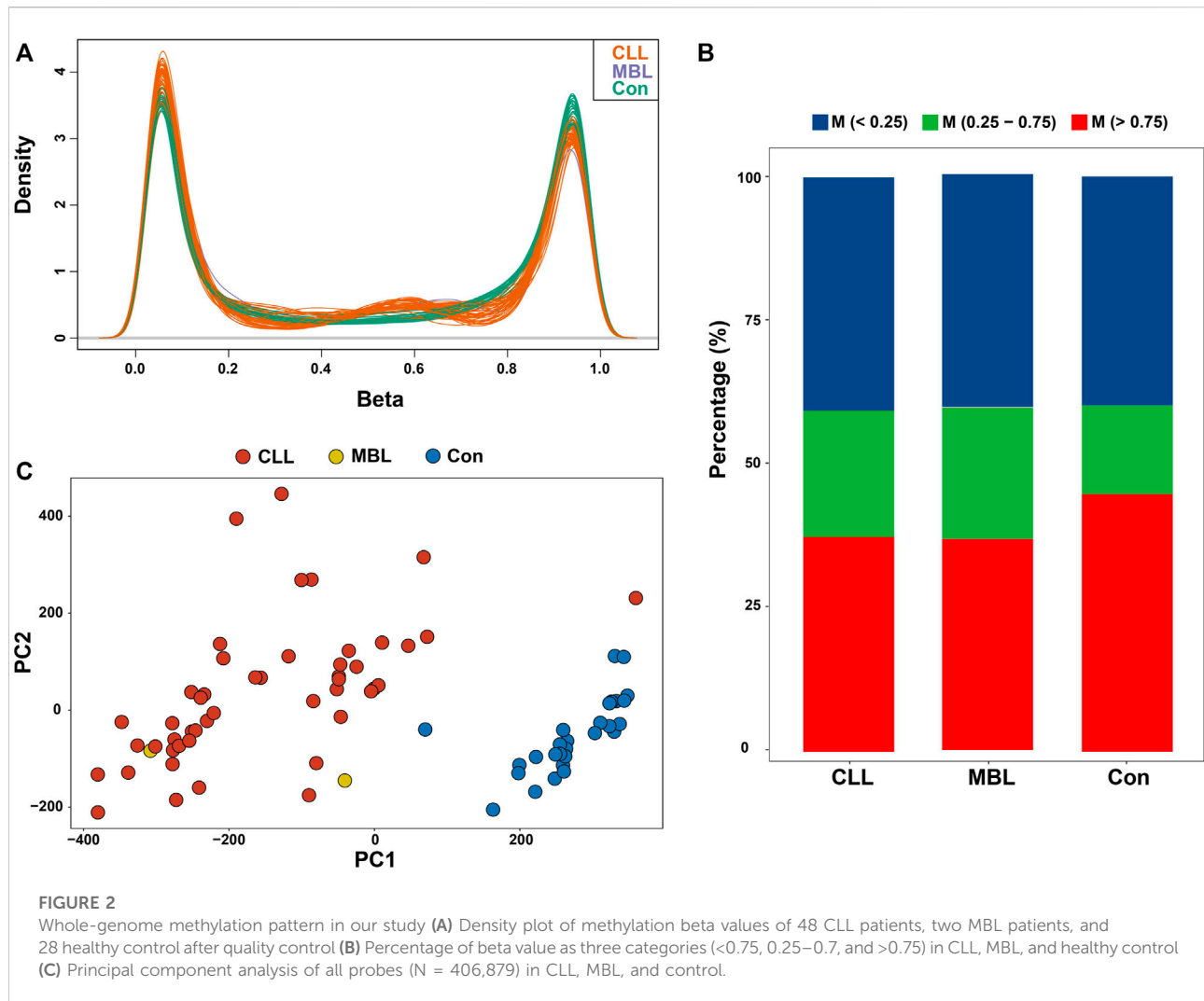
Genome-wide DNA methylation status was evaluated in 48 CLL cases and 28 healthy controls (Table 1). CLL patients were on average approximately 4 years older than controls (*p* =

0.015). Both white blood cell counts (WBCs) and absolute lymphocyte counts (ALCs) in CLL cases were significantly higher than those of controls (*p* = 1.2e-7; *P* = 3e-5). Among CLL cases, 21 were at Rai stage 0, 22 were at Rai stage I/II, and three were at Rai stage III/IV. Two patients with monoclonal B Cell lymphocytosis (MBL), one male and one female, were also included for preliminary exploration.

### CLL-related differentially methylated positions (DMPs)

The density plot of beta values showed that genomic CpG sites followed a bimodal distribution in cases and healthy controls (Figure 2A). Compared to controls, both CLL and MBL cases contained more sites with decreased methylation (Figure 2B). Principal component analysis (PCA) based on all probes showed that the overall methylation pattern of CLL cases was significantly different from that of controls, and MBL patients were grouped into CLL cases (Figure 2C). With the criteria of  $|\Delta\beta| > 0.2$  and *q*-value < 0.05, we identified 34,797 DMPs between CLL cases and controls (Supplementary Table S3). Unsupervised clustering analysis showed that most DMPs were hypomethylated in CLL cases compared to controls (Figure 3A). The DMPs were annotated with genomic features and CGI features, and the distribution of DMPs on 22 autosomes is shown in Figure 3B. Over 57.7% of DMPs in CpG islands were hypermethylated (Figure 3C). Nearly 93.9% of DMPs in gene body regions were hypomethylated (Figure 3D). A total of 477 differentially methylated regions (DMRs) located in





232 genes were identified (Supplementary Table S4). Chromosome six contained the highest number of DMRs, with 77 (Supplementary Figure S1).

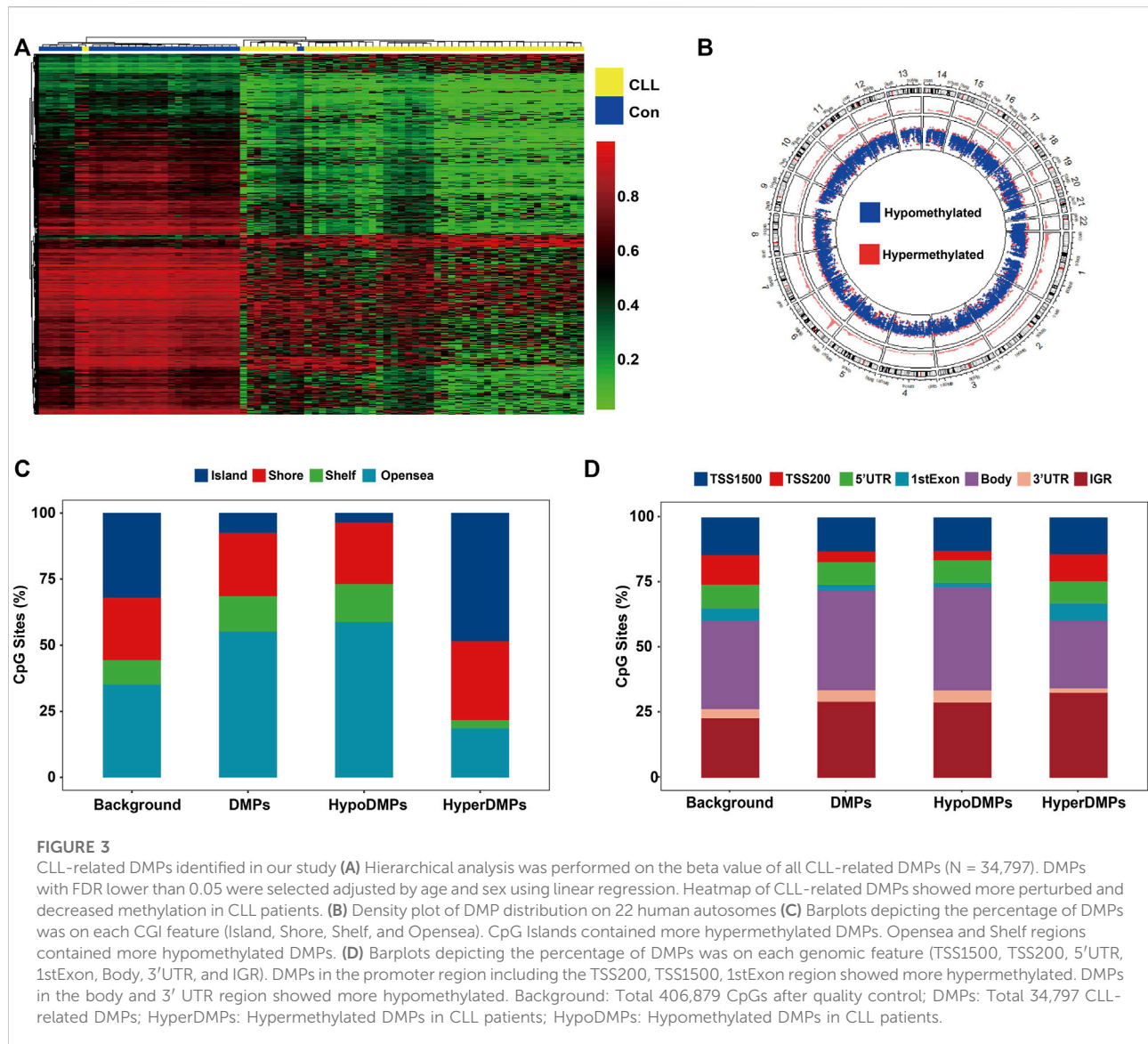
### CLL-related copy number aberrations (CNAs)

Moreover, 82 CNA regions were detected in CLL cases compared to controls (Supplementary Figure S2; Supplementary Table S5). The copy number loss mapped to chr13, chr17 and chr19 were top three chromosomes. Only three of the 82 CNAs were copy number gain regions, including two regions across *HOXC8* and *GALNT9* on chromosome 12 and one region across *GSTTP1* on chromosome 22. In contrast, 79 copy number loss CNAs were identified; these included two regions across *GSTT1* and *LOC391322* on chromosome 22 that exhibited both copy number gains and losses. The other 77 copy number

losses were located on chromosomes 1, 5, 6, 7, 11, 12, 13, 14, 17, and 19.

### Differentially methylated and expressed genes

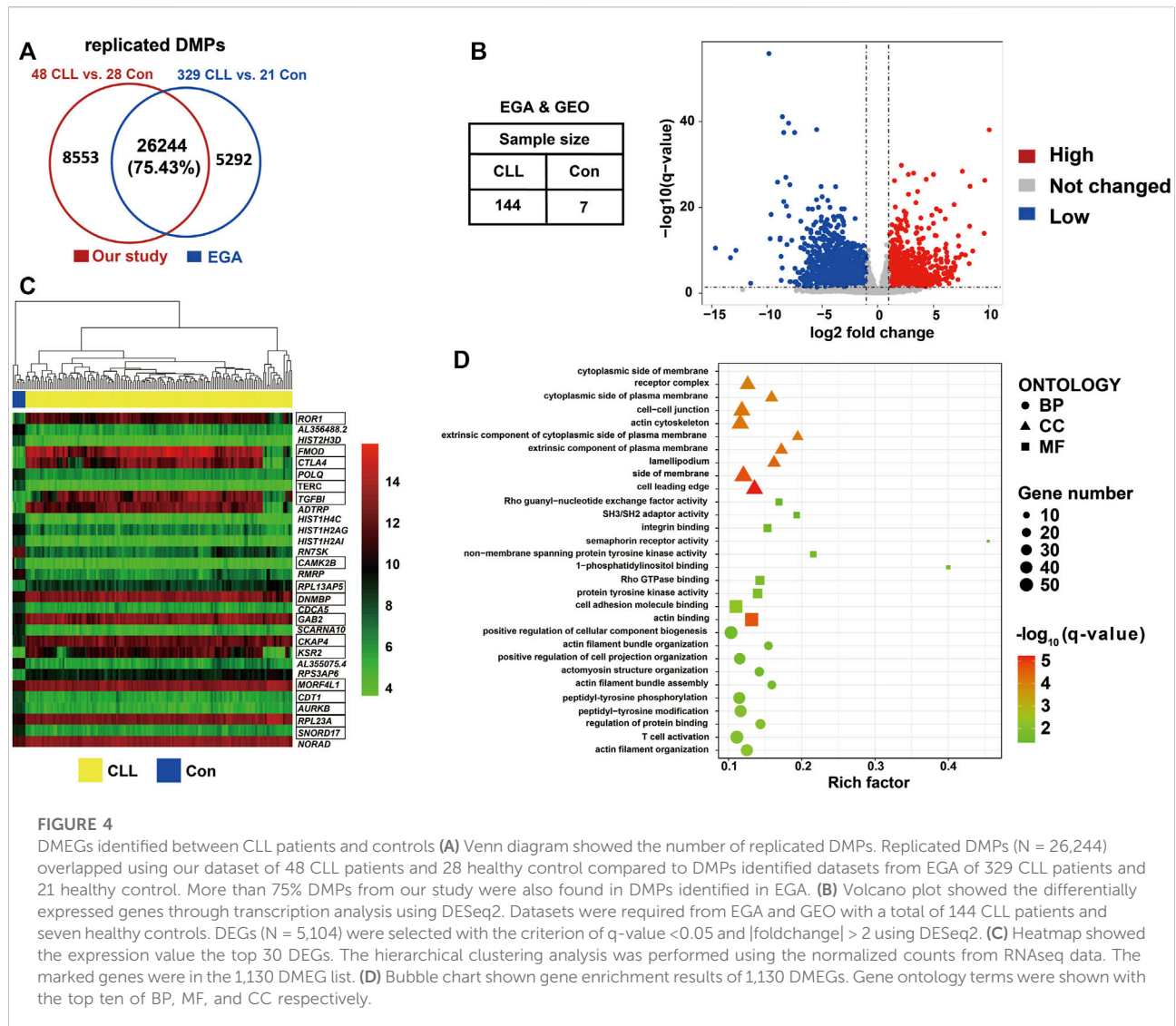
A comparison of 450k methylation data from 329 CLL cases and 21 controls from the EGA revealed 31,536 DMPs. Over 75% of the DMPs found in our study ( $n = 26,244$ ) could be replicated with the same direction as in the EGA data (Figure 4A; Supplementary Table S6A). We constructed a volcano plot showing the 5,104 differentially expressed genes (DEGs) that were detected by comparing the RNA sequencing data of 144 CLL cases and seven controls from the EGA and GEO datasets (Figure 4B; Supplementary Table S6B); the heatmap of the top 30 DEGs according to normalized counts using *vst* method was shown in Figure 4C. By combining the DEGs



and the replicated DMPs, 1,130 DMEGs were identified (Supplementary Table S6C). We found that *ROR1* and *ZAP70* was hypomethylated in the promoter region and highly expressed in CLL patients (Table S6C). The top ten BP, MF, and CC Gene Ontology terms identified in enrichment analysis of these DMEGs (Supplementary Table S7) are shown in Figure 4D. The biological process of actin filament organization and T cell activation were involved in CLL progression.

The DMEGs were classified as three classes: 1,335 DMEGs with DMPs located in gene body regions, and the methylation levels of all these DMPs and gene expression had the same direction in CLL (Supplementary Table S8A). 2,326 DMEGs with DMPs located in body regions, and the methylation levels of all these DMPs and gene expression were in the inverse direction

in CLL (Supplementary Table S8B). 3,226 DMEGs with DMPs located in promoter regions, and the methylation levels of all these DMPs and the expression were in the inverse direction (Supplementary Table S8C). Then we combined the DMPs of promoter and body for each DMEG. There were 83 DMEGs whose DMPs were both in body and promoter regions by combining 335 DMEGs and 226 DMEGs. These 83 DMEGs were all highly expressed in CLL with hypomethylated DMPs in the two regions, including *ROR1* and *ZAP70* (Supplementary Table S8D). There were nine DMEGs by combining 326 DMEGs and 226 DMEGs, including *MACROD2*, *ADAMTS17*, *TJP1*, *MET*, *OSBPLIA*, *SYN2*, *KCNQ2*, *AGBL4*, and *ME3*, were lower expressed in CLL cases with hypermethylated DMPs in promoter regions and hypomethylated DMP in body regions (Supplementary Table S8E).



## Methylation heterogeneity in CLL

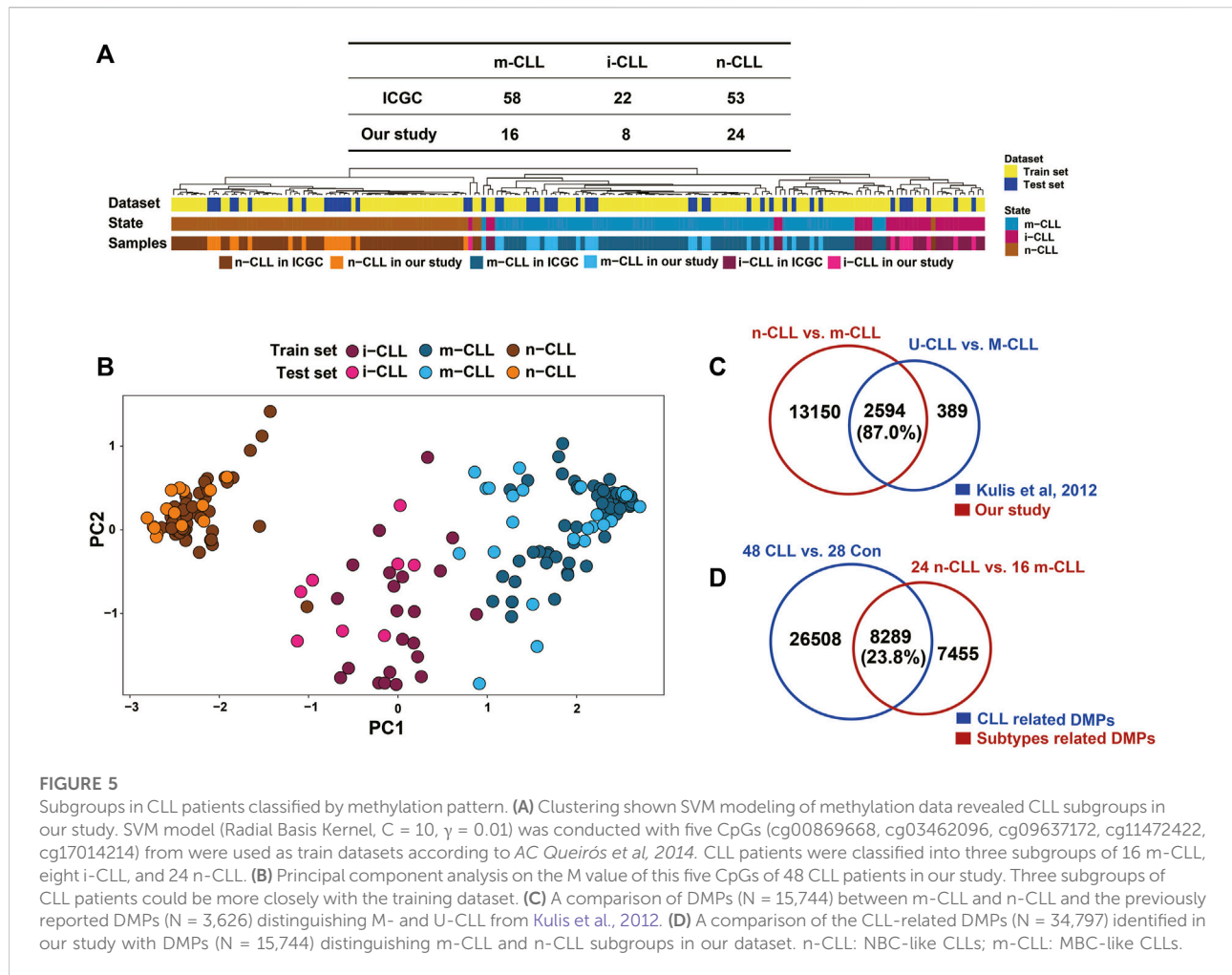
The 48 CLL patients in our study were classified into subgroups composed of 24 n-CLL, 16 m-CLL, and eight i-CLL patients using the SVM model of five CpG sites reported by Queiros et al. (2015). The hierarchical clustering of the five CpG sites reported in Queiros et al., defined as in ICGC and in our study is shown in Figure 5A. PCA also showed that the subgroups were distinct from each other (Figure 5B). By fitting the linear model adjusted for age and sex, 15,744 DMPs covering 5,906 genes were identified by comparing the n-CLL and m-CLL subgroups. It was previously reported that 3,266 DMPs covering 1,519 genes could distinguish the U-CLL and M-CLL subgroups (Kulis et al., 2012); 87% of the 3,266 DMPs covering 1,427 genes also differentiated the n-CLL and m-CLL subtypes in our study (Figure 5C). In addition,

compared to healthy controls, only 23.8% of DMPs were overlapped distinguishing the n-CLL and m-CLL subtypes overlapped with the CLL-related DMPs (Figure 5D).

## B Cell differentiation-related methylation changes

Studies have shown that dynamic methylation changes during normal B Cell differentiation may play a role in B-CLL aetiology (Kulis et al., 2015; Oakes et al., 2016). To test this hypothesis, we retrieved a 450k DNA methylation dataset for CLL (EGAD00010000254) and compared the DNA methylation patterns of four main CD19<sup>+</sup> B Cell subtypes (CD5<sup>+</sup> NBCs, NBCs, ncsMBCs, csMBCs) compared to those of the CD19<sup>+</sup> B Cell mixture. A total of 84,273 CpGs related to





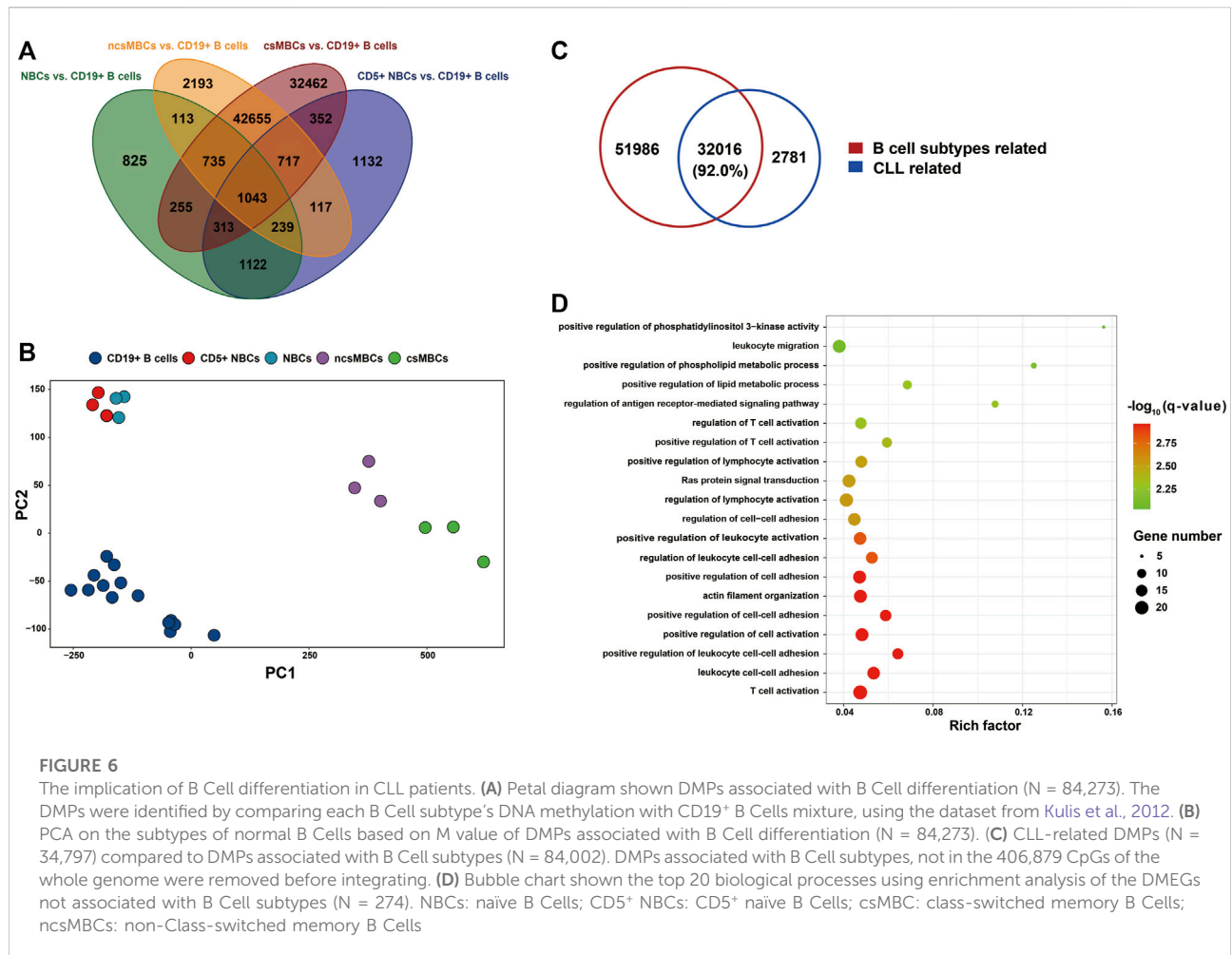
B Cell differentiation subtypes were identified (Figure 6A). The top two principal components of the M-value of these CpGs could also differentiate the subtypes from the CD19<sup>+</sup> B Cell mixture (Figure 6B). Comparing the B Cell differentiation-related DMPs to the CLL-related DMPs showed that more than 92% of CLL-related DMPs were normal B Cell subtype-related (Figure 6C). After removing the 84,273 CpGs related to B-cell differentiation, 2,781 CpG sites covered by 274 genes out of 34,797 CpG sites with expression changes were detected. The top 20 most enriched biological processes identified by gene enrichment analysis among those 274 genes are shown in Figure 6D. And the enriched terms included T cell activation, cell-cell adhesion, and lipid metabolic process.

## Discussion

In this study, we evaluated 406,879 CpGs and identified 34,797 DMPs related to CLL. More than 90% of the DMPs in

CLL patients were hypomethylated and mainly in gene body regions. Combining our study with public datasets of 450k methylation data and RNAseq data, we identified 1,130 DMEGs with methylation and gene expression both altered in CLL patients. Nine low expressed DMEGs were selected with hypermethylated in promoter and hypomethylated in body, such as *ADAMTS17*, and 83 high expressed DMEGs were selected with both hypomethylated in promoter and body, such as *ZAP70*, *FMOD*, *PAX9*. 48 CLL cases were characterized into three subgroups based on SVM model. Methylation sites associated with B Cell subtypes account for large methylation alteration in CLL. Apart from B Cell differentiation, methylation changes in DMEGs involved in T cell activation and lipid metabolism might also contribute to CLL.

Large-scale methylation alterations may contribute to chromosomal instability (*Landau et al., 2014; Oakes et al., 2014*) and subsequently promote CLL pathogenesis and progression. Losses of 13q and 17p were reported as common cytogenetic aberrations in CLL patients and are associated with



the prognosis of CLL (Ouillette et al., 2011; Puiggros et al., 2014). We also observed that the loss of chromosome 13q was the most frequent CNA in CLL patients. The deletion of the 13q14.3 region spans several TSGs, including *FAM10A4* (Sossey-Alaoui et al., 2002), *DLEU7* (Palamarchuk et al., 2010), *KCNRG* (Bierdinc et al., 2010), and *RBI* (Liu et al., 1993), which have been reported to be related to CLL aetiology. A region on chromosome 17q including *MIR21* (17q23.1, 3' UTR *TMEM49*), one of the microRNA fingerprints in CLL patients (Rossi et al., 2010), was lost in more than 50% of the CLL cases in our study.

There were 34,797 DMPs related to CLL, of which more than 90% were hypomethylated in CLL patients, and mainly located in gene body regions. The gene body regions contained a large number of hypomethylated DMPs in CLL, consistent with a recent study exploring the global distribution of DNA hydroxymethylation and methylation in CLL (Wernig-Zorc et al., 2019).

Combining our study with public datasets of 450k methylation data and RNAseq data, we identified

1,130 DMEGs associated with CLL. Among them, *PAX9* (Rani et al., 2017), *FMOD* (Beekman et al., 2018), and *LEF1* (Beekman et al., 2018) were reported previously. *FMOD* was the top DMEGs with expression level increased more than a thousand folds in CLL patients, which was consistent with report of Beekman et al. (Beekman et al., 2018). *FMOD* belongs to the family of small interstitial proteoglycans, and its silencing could induce apoptosis in CLL cell lines (Choudhury et al., 2010). *SNORD17* the DMEG with the lowest q-value in our study, was reported to play a role in the biogenesis of small nuclear RNAs (snRNAs). *SNORD17* can be regulated by the non-coding splicing factor *SF3B1*, which is frequently mutated in CLL (Shuai et al., 2019). The methylation status of *ZAP70* has been reported to be associated with its expression and has been recognized as a strong prognostic marker for CLL (Corcoran et al., 2005). We found that *ZAP70* was both hypomethylated in the promoter and body region, and was highly expressed in CLL patients. Similarly, the CD5 antigen has been reported to be highly expressed in CLL patients (Bashford-Rogers et al., 2017), and we also found that it was

hypomethylated in promoter and body regions and highly expressed in CLL cases. The tyrosine kinase receptor *ROR1* gene, which is involved in the peptidyl-tyrosine modification pathway, was reported to be upregulated in CLL patients (Hojjat-Farsangi et al., 2015). Consistent with this report, we observed hypomethylation of *ROR1* both in promoter and body region and higher *ROR1* expression with 32-fold in CLL patients. *CTLA-4*, a gene that serves as an immune checkpoint to regulate T cell function (Palma et al., 2017), was hypomethylated in the first exon region and body region in our CLL cases and had 128-fold higher expression in CLL patients than in healthy controls according to publicly available transcription data.

Human MHC class II molecules (HLA-DMB, HLA-DMA, HLA-DRA, HLA-DOA) in the top 30 DMRs, which could attract inflammatory tumour-specific CD4<sup>+</sup> T cells and dampen CD8<sup>+</sup> T cell antitumour reactivity (Donia et al., 2015), were found to be hypomethylated in all CLL patients in our study. Considering that T cell activation was among the top two enriched biological processes in CLL samples, these results suggest that T cell-related pathways may play an important role in CLL pathogenesis. The cytoskeleton plays a key role during several stages of B Cell activation (Kuokkanen et al., 2015). We observed three actin filament processes among the top ten enriched biological processes in our gene enrichment analysis. One study in 1986 had reported that cytoskeletal organization was aberrantly rearranged with adhesive properties in CLL (Caligaris-Cappio et al., 1986).

There were nine DMEGs were lower expressed in CLL patients, with DNA hypermethylated in the promoter region and hypomethylated in gene body regions, including *MACROD2*, *ADAMTS17*, *TJPI1*, *MET*, *OSBPL1A*, *SYN2*, *KCNQ2*, *AGBL4*, and *ME3*. Among these genes, *ADAMTS17* was also reported to be hypermethylated and transcriptionally silenced in CLL (Barrow et al., 2021). Although the other eight genes have not been reported in CLL, aberrant methylation, transcription, or genetic alteration of these genes has been observed in various cancers, including breast cancer (Mohseni et al., 2014), colorectal cancer (Thorsen et al., 2011; Lin et al., 2015; Kordowski et al., 2018), prostate cancer (Devaney et al., 2013), malignant glioma (Milinkovic et al., 2013), and some other epithelial cancers (Martin and Jiang, 2009; Kordowski et al., 2018).

By comparing DMPs between m-CLL and n-CLL (classified by SVM model) with CLL-related DMPs, only 23.8% of the CLL-related DMPs we identified were associated with CLL subtypes. This might suggest that the heterogeneity of CLL had a limited impact on the detection of CLL-related DMPs. Using the DNA methylation of B Cell subtypes reported by Kulis et al. (2012), we found that 92% of CLL-related DMPs were involved in the B Cell differentiation process, which suggests that B Cell maturation may play a prominent role in CLL pathogenesis. There were 274 DMEGs identified by comparing CLL and controls after removing the B Cell maturation-related DMPs from background

CpG sites. Gene enrichment analysis of the 274 DMEGs suggested that T cell activation and cell adhesion related processes might be involved in the pathogenesis of CLL. Ras protein signal transduction, which has been recognized as an attractive therapeutic target for myeloid leukaemia (Willman, 2001), was also observed as an enriched pathway. Interestingly, the lipid metabolism pathway was also one of the top 20 enriched biological processes. We also found that the gene expression of lipoprotein lipase (LPL) and CD36 were significantly upregulated in CLL patients. These two proteins could activate the uptake of fatty acids and were highly expressed in a lipogenic cancer model (Zaidi et al., 2013). LPL expression was reported to be closely correlated with IGHV mutational status and may be involved in *BCR* and *NOTCH1*-dependent signalling pathways (Bilous et al., 2019). Further studies are needed to decipher the role of lipid metabolism in CLL aetiology. We thought these findings might shed light on finding the new biological process for CLL carcinogenesis.

A previous study showed that methylation status was stable during CLL progression for 6.8 years (Cahill et al., 2013), suggesting that methylation is altered in a very early stage of neoplastic transformation. MBL is defined as the presence of CLL phenotypic cells in the peripheral blood in the absence of other features of CLL. MBL can be categorized as either low-count or high-count based on whether the B Cell count is above or below  $0.5 \times 10^9/L$ . High-count MBL progresses to CLL requiring therapy at a rate of 1%–2% per year (Strati and Shanafelt, 2015). Therefore, we examined two high-count MBL cases for preliminary exploration. The methylation patterns of these two MBL cases were more similar to that of CLL cases than that of healthy controls, supporting the hypothesis that methylation alteration occurs in a very early stage of CLL development. Since the sample size of MBL was limited in our study, further complementary longitudinal studies of MBL are warranted to validate our findings.

Our study presents several strengths. First, 94% of CLL cases in our study were in the early stage (0, I, II). Thus, the influences of CLL progression and medical treatment on methylation status were limited in these early stage patients compared to late-stage (III and IV) patients. Second, we integrated our data with publicly available methylation and expression data, which provided independent validation of the methylation results and provided with transcription data. There are also limitations in our study. We could not directly validate the DNA methylation and RNA expression for the top DMEGs with our population due to the lack of biospecimens. However, we acquired only 450k DNA methylation data from public databases to validate our findings. The larger population from the public data may provide extra information for methylation pattern and also made up the sample size limited by our own biospecimen.

## Conclusion

In summary, our study revealed that the DNA methylation pattern was altered across the genome in CLL patients. Differentially expressed genes with changes in methylation status, were detected by combining publicly available DNA methylation and RNA expression data. CLL-related DMEGs reported previously, including *ZAP70*, *FMOD*, and *ADAMTS17*, were also detected in our study. Our study also found that the methylation heterogeneity of CLL, and abnormal methylation of B Cell differentiation should be an important factor in the pathogenesis of CLL. Apart from B Cell differentiation, we detected DMEGs involving T cell activation, cell adhesion, Ras signalling, and lipid metabolism were associated with CLL, suggesting that these pathways might contribute to CLL pathogenesis. Further longitudinal large-scale population studies are warranted to replicate our results. Subsequent studies are also needed to decipher the mechanism of these DMEGs in CLL pathogenesis.

## Data availability statement

The 450K methylation datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.biosino.org/node/, OEP000173>.

## Ethics statement

The studies involving human participants were reviewed and approved by the National Institutes of Health Office of Human Subjects Research. All participants provided their written informed consent to participate in this study. The information of population was also described by our previously published research (Lin et al., 2019).

## Author contributions

YG and NC initiated and designed the study. LG and GM participated in sampling the subjects in our study. HS, DE, PM, and KK conducted the experiments. QZ, SL, and YW performed the data analysis. QZ, YG, SL, and SX wrote and revised the manuscript. All authors read, gave comments, and approved the final manuscript.

## Funding

This study was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant NO.

XDA26040304 to YG), Shanghai Science and Technology Committee (STCSM-20ZR1466400), and grants from the Intramural Research Award of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, United States of America.

## Acknowledgments

We thank Laura Fontaine, at Westat in Rockville, MD for assistance in requesting the clinical and demographic data. We thank Tao Huang for polishing description of our manuscript. We thank Keyu Jia and Kaiqi Hou for the assistance in formatting figures needed for the scientific manuscript. We are also very indebted to the participants involved in our study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1056043/full#supplementary-material>

### SUPPLEMENTARY FIGURE S1

Chromosome six contains the highest number of DMRs.

### SUPPLEMENTARY FIGURE S2

Composite CNV analysis of 48 CLL cases compared to 28 reference controls based on 450k microarray data.

### SUPPLEMENTARY TABLE S1

Summary of previously reported genome-wide DNA methylation associated studies on CLL.

### SUPPLEMENTARY TABLE S2

(A) Public datasets used for methylation replication and transcription analysis in this study. (B) Clinical characteristics of samples used for methylation replication (EGAD00010000871 and EGAD00010000254). (C) RNASeq data of CLL cases (EGAD0001000258 and GSE66117), and healthy control (GSE62246 and GSE70830).



**SUPPLEMENTARY TABLE S3**

CLL-related DMPs (N = 34,797).

**SUPPLEMENTARY TABLE S4**

CLL-related DMRs (N = 477).

**SUPPLEMENTARY TABLE S5**

CLL-related CNVs (N = 82).

**SUPPLEMENTARY TABLE S6****(A)** Replicated DMPs (N = 26,244). **(B)** CLL-related DEGs (N = 5,104). **(C)** CLL-related DMEGs (N = 1,130).**SUPPLEMENTARY TABLE S7**

GO enrichment results of 1,130 DMEGs in BP, MF, and CC.

**SUPPLEMENTARY TABLE S8****(A)** DMEGs with DMPs in body regions all positive with expression (N = 335). **(B)** DMEGs with DMPs in body regions all negative with expression (N = 326). **(C)** DMEGs with DMPs in promoter regions all negative with expression (N = 225). **(D)** DMEGs with DMPs both in promoter and body regions all negative with expression (N = 83). **(E)** DMEGs with DMPs in promoter negative and body regions positive with expression (N = 9).

## References

- Barrow, T. M., Nakjang, S., Lafta, F., Bilotkach, K., Woodhouse, L., Junge, G., et al. (2021). Epigenome-wide analysis reveals functional modulators of drug sensitivity and post-treatment survival in chronic lymphocytic leukaemia. *Br. J. Cancer* 124 (2), 474–483. doi:10.1038/s41416-020-01117-8
- Bashford-Rogers, R. J., Palser, A. L., Hodgkinson, C., Baxter, J., Follows, G. A., Vassiliou, G. S., et al. (2017). Dynamic variation of CD5 surface expression levels within individual chronic lymphocytic leukemia clones. *Exp. Hematol.* 50, 31–37. doi:10.1016/j.exphem.2016.09.010
- Beekman, R., Chapaprieta, V., Russinol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, N., Martens, J. H. A., et al. (2018). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* 24 (6), 868–880. doi:10.1038/s41591-018-0028-4
- Berndt, S. I., Camp, N. J., Skibola, C. F., Vijai, J., Wang, Z., Gu, J., et al. (2016). Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat. Commun.* 7, 10933. doi:10.1038/ncomms10933
- Bilous, N., Abramenko, I., Chumak, A., Dyagil, I., and Martina, Z. (2019). Analysis of LPL gene expression in patients with chronic lymphocytic leukemia. *Exp. Oncol.* 41 (1), 39–45.
- Birerdinc, A., Nohelty, E., Marakhonov, A., Manyam, G., Panov, I., Coon, S., et al. (2010). Pro-apoptotic and antiproliferative activity of human KCNKG, a putative tumor suppressor in 13q14 region. *Tumour Biol. J. Int. Soc. Oncodev. Biol. Med.* 31 (1), 33–45. doi:10.1007/s13277-009-0005-0
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170
- Butcher, L. M., and Beck, S. (2015). Probe lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. *Methods (San Diego, Calif.)* 72, 21–28. doi:10.1016/j.jymeth.2014.10.036
- Cahill, N., Bergh, A. C., Kanduri, M., Goransson-Kultima, H., Mansouri, L., Isaksson, A., et al. (2013). 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. *Leukemia* 27 (1), 150–158. doi:10.1038/leu.2012.245
- Caligaris-Cappio, F., Bergui, L., Tesio, L., Corbascio, G., Tousou, F., and Marchisio, P. C. (1986). Cytoskeleton organization is aberrantly rearranged in the cells of B chronic lymphocytic leukemia and hairy cell leukemia. *Blood* 67 (1), 233–239. doi:10.1182/blood.v67.1.233.233
- Chiorazzi, N., Rai, K. R., and Ferrarini, M. (2005). Chronic lymphocytic leukemia. *N. Engl. J. Med.* 352 (8), 804–815. doi:10.1056/NEJMra041720
- Choudhury, A., Derkow, K., Daneshmanesh, A. H., Mikaelsson, E., Kiaii, S., Kokhaei, P., et al. (2010). Silencing of ROR1 and FMOD with siRNA results in apoptosis of CLL cells. *Br. J. Haematol.* 151 (4), 327–335. doi:10.1111/j.1365-2141.2010.08362.x
- Corcoran, M., Parker, A., Orchard, J., Davis, Z., Wirtz, M., Schmitz, O. J., et al. (2005). ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia. *Haematologica* 90 (8), 1078–1088.
- Devaney, J. M., Wang, S., Funda, S., Long, J., Taghipour, D. J., Tbaishat, R., et al. (2013). Identification of novel DNA-methylated genes that correlate with human prostate cancer and high-grade prostatic intraepithelial neoplasia. *Prostate Cancer Prostatic Dis.* 16 (4), 292–300. doi:10.1038/pcan.2013.21
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635
- Donia, M., Andersen, R., Kjeldsen, J. W., Fagone, P., Munir, S., Nicoletti, F., et al. (2015). Aberrant expression of MHC class II in melanoma attracts inflammatory tumor-specific CD4+ T- cells, which dampen CD8+ T-cell antitumor reactivity. *Cancer Res.* 75 (18), 3747–3759. doi:10.1158/0008-5472.CAN-14-2956
- Feber, A., Guilhamon, P., Lechner, M., Fenton, T., Wilson, G. A., Thirlwell, C., et al. (2014). Using high-density DNA methylation arrays to profile copy number alterations. *Genome Biol.* 15 (2), R30. doi:10.1186/gb-2014-15-2-r30
- Feinberg, A. P. (2018). The key role of epigenetics in human disease prevention and mitigation. *N. Engl. J. Med.* 378 (14), 1323–1334. doi:10.1056/NEJMra1402513
- Herman, J. G., and Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* 349 (21), 2042–2054. doi:10.1056/NEJMra023075
- Hojjat-Farsangi, M., Jeddi-Tehrani, M., Daneshmanesh, A. H., Mozaffari, F., Moshfegh, A., Hansson, L., et al. (2015). Spontaneous immunity against the receptor tyrosine kinase ROR1 in patients with chronic lymphocytic leukemia. *PLoS one* 10 (11), e0142310. doi:10.1371/journal.pone.0142310
- Ishibe, N., Sgambati, M. T., Fontaine, L., Goldin, L. R., Jain, N., Weissman, N., et al. (2001). Clinical characteristics of familial B-CLL in the national cancer Institute familial Registry. *Leuk. Lymphoma* 42 (1-2), 99–108. doi:10.3109/10428190109097681
- Jiao, Y., Widschwendter, M., and Teschendorff, A. E. (2014). A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinforma. Oxf. Engl.* 30 (16), 2360–2366. doi:10.1093/bioinformatics/btu316
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostat. Oxf. Engl.* 8 (1), 118–127. doi:10.1093/biostatistics/kxj037
- Kordowski, F., Kolarova, J., Schafmayer, C., Buch, S., Goldmann, T., Marwitz, S., et al. (2018). Aberrant DNA methylation of ADAMTS16 in colorectal and other epithelial cancers. *BMC Cancer* 18 (1), 796. doi:10.1186/s12885-018-4701-2
- Kulis, M., Heath, S., Bibikova, M., Queiros, A. C., Navarro, A., Clot, G., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* 44 (11), 1236–1242. doi:10.1038/ng.2443
- Kulis, M., Merkel, A., Heath, S., Queiros, A. C., Schuyler, R. P., Castellano, G., et al. (2015). Whole-genome fingerprint of the DNA methylation during human B cell differentiation. *Nat. Genet.* 47 (7), 746–756. doi:10.1038/ng.3291
- Kuokkanen, E., Šuštar, V., and Mattila, P. K. (2015). Molecular control of B cell activation and immunological synapse formation. *Traffic* 16 (4), 311–326. doi:10.1111/tra.12257
- Landau, D. A., Clement, K., Ziller, M. J., Boyle, P., Fan, J., Gu, H., et al. (2014). Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 26 (6), 813–825. doi:10.1016/j.ccell.2014.10.012
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* 30 (7), 923–930. doi:10.1093/bioinformatics/btt656
- Lin, P. C., Lin, J. K., Lin, C. H., Lin, H. H., Yang, S. H., Jiang, J. K., et al. (2015). Clinical relevance of plasma DNA methylation in colorectal cancer patients identified by using a genome-wide high-resolution array. *Ann. Surg. Oncol.* 22, S1419–S1427. doi:10.1245/s10434-014-4277-2
- Lin, S., Liu, Y., Goldin, L. R., Lyu, C., Kong, X., Zhang, Y., et al. (2019). Sex-related DNA methylation differences in B cell chronic lymphocytic leukemia. *Biol. Sex Differ.* 10 (1), 2. doi:10.1186/s13293-018-0213-7



- Liu, Y., Szekely, L., Grander, D., Soderhall, S., Juliusson, G., Gahrton, G., et al. (1993). Chronic lymphocytic leukemia cells with allelic deletions at 13q14 commonly have one intact RB1 gene: Evidence for a role of an adjacent locus. *Proc. Natl. Acad. Sci. U. S. A.* 90 (18), 8697–8701. doi:10.1073/pnas.90.18.8697
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Mansouri, L., Wierzbinska, J. A., Plass, C., and Rosenquist, R. (2018). Epigenetic deregulation in chronic lymphocytic leukemia: Clinical and biological impact. *Seminars Cancer Biol.* 51, 1–11. doi:10.1016/j.semcancer.2018.02.001
- Martin, T. A., and Jiang, W. G. (2009). Loss of tight junction barrier function and its role in cancer metastasis. *Biochim. Biophys. Acta* 1788 (4), 872–891. doi:10.1016/j.bbmem.2008.11.005
- Milinkovic, V., Bankovic, J., Rakic, M., Stankovic, T., Skender-Gazibara, M., Ruzdijic, S., et al. (2013). Identification of novel genetic alterations in samples of malignant glioma patients. *PLoS One* 8 (12), e82108. doi:10.1371/journal.pone.0082108
- Mohseni, M., Cidado, J., Croessmann, S., Cravero, K., Cimino-Mathews, A., Wong, H. Y., et al. (2014). MACROD2 overexpression mediates estrogen independent growth and tamoxifen resistance in breast cancers. *Proc. Natl. Acad. Sci. U. S. A.* 111 (49), 17606–17611. doi:10.1073/pnas.1408650111
- Oakes, C. C., Claus, R., Gu, L., Assenov, Y., Hüllein, J., Zucknick, M., et al. (2014). Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov.* 4 (3), 348–361. doi:10.1158/2159-8290.CD-13-0349
- Oakes, C. C., Seifert, M., Assenov, Y., Gu, L., Przekopowicz, M., Ruppert, A. S., et al. (2016). DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.* 48 (3), 253–264. doi:10.1038/ng.3488
- Ouillette, P., Collins, R., Shakhani, S., Li, J., Li, C., Shedden, K., et al. (2011). The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. *Clin. Cancer Res. Official J. Am. Assoc. Cancer Res.* 17 (21), 6778–6790. doi:10.1158/1078-0432.CCR-11-0785
- Palamarchuk, A., Efanov, A., Nazaryan, N., Santanam, U., Alder, H., Rassenti, L., et al. (2010). 13q14 deletions in CLL involve cooperating tumor suppressors. *Blood* 115 (19), 3916–3922. doi:10.1182/blood-2009-10-249367
- Palma, M., Gentilcore, G., Heimersson, K., Mozaffari, F., Nasman-Glaser, B., Young, E., et al. (2017). T cells in chronic lymphocytic leukemia display dysregulated expression of immune checkpoints and activation markers. *Haematologica* 102 (3), 562–572. doi:10.3324/haematol.2016.151100
- Parikh, S. A., and Shanafelt, T. D. (2016). Prognostic factors and risk stratification in chronic lymphocytic leukemia. *Seminars Oncol.* 43 (2), 233–240. doi:10.1053/j.seminoncol.2016.02.009
- Puiggros, A., Blanco, G., and Espinet, B. (2014). Genetic abnormalities in chronic lymphocytic leukemia: Where we are and where we go. *BioMed Res. Int.* 2014, 435983. doi:10.1155/2014/435983
- Queiros, A. C., Villamor, N., Clot, G., Martinez-Trillos, A., Kulis, M., Navarro, A., et al. (2015). A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* 29 (3), 598–605. doi:10.1038/leu.2014.252
- Rani, L., Mathur, N., Gupta, R., Gogia, A., Kaur, G., Dhanjal, J. K., et al. (2017). Genome-wide DNA methylation profiling integrated with gene expression profiling identifies PAX9 as a novel prognostic marker in chronic lymphocytic leukemia. *Clin. Epigenet.* 9, 57. doi:10.1186/s13148-017-0356-0
- Rossi, S., Shimizu, M., Barbarotto, E., Nicoloso, M. S., Dimitri, F., Sampath, D., et al. (2010). microRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival. *Blood* 116 (6), 945–952. doi:10.1182/blood-2010-01-263889
- Shuai, S., Suzuki, H., Diaz-Navarro, A., Nadeu, F., Kumar, S. A., Gutierrez-Fernandez, A., et al. (2019). The U1 spliceosomal RNA is recurrently mutated in multiple cancers. *Nature* 574, 712–716. doi:10.1038/s41586-019-1651-z
- Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics. *CA a cancer J. Clin.* 69 (1), 7–34. doi:10.3322/caac.21551
- Sossey-Alaoui, K., Kitamura, E., Head, K., and Cowell, J. K. (2002). Characterization of FAM10A4, a member of the ST13 tumor suppressor gene family that maps to the 13q14.3 region associated with B-Cell leukemia, multiple myeloma, and prostate cancer. *Genomics* 80 (1), 5–7. doi:10.1006/geno.2002.6792
- Strati, P., and Shanafelt, T. D. (2015). Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: Diagnosis, natural history, and risk stratification. *Blood* 126 (4), 454–462. doi:10.1182/blood-2015-02-585059
- Thorsen, K., Schepeler, T., Oster, B., Rasmussen, M. H., Vang, S., Wang, K., et al. (2011). Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis. *BMC Genomics* 12, 505. doi:10.1186/1471-2164-12-505
- Wernig-Zorc, S., Yadav, M. P., Kopparapu, P. K., Bemark, M., Kristjansdottir, H. L., Andersson, P. O., et al. (2019). Global distribution of DNA hydroxymethylation and DNA methylation in chronic lymphocytic leukemia. *Epigenet. Chromatin* 12 (1), 4. doi:10.1186/s13072-018-0252-7
- Willman, C. L. (2001). Targeted AML therapy: New biologic paradigms and therapeutic opportunities. *Leukemia* 15 (4), 690–694. doi:10.1038/sj.leu.2402087
- Yang, X., Han, H., De Carvalho, D. D., Lay, F. D., Jones, P. A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26 (4), 577–590. doi:10.1016/j.ccr.2014.07.028
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics a J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118
- Zaidi, N., Lupien, L., Kuemmerle, N. B., Kinlaw, W. B., Swinnen, J. V., and Smans, K. (2013). Lipogenesis and lipolysis: The pathways exploited by the cancer cells to acquire fatty acids. *Prog. Lipid Res.* 52 (4), 585–589. doi:10.1016/j.plipres.2013.08.005
- Zhou, W., Laird, P. W., and Shen, H. (2017). Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic acids Res.* 45 (4), e22. doi:10.1093/nar/gkw967