SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Check for updates

DNA methylation of chronic lymphocytic leukemia with differential response to chemotherapy

Deyan Yordanov Yosifov [1,2], Johannes Bloehdorn[1], Hartmut Döhner[1], Peter Lichter[3], Stephan Stilgenbauer[1] & Daniel Mertens [1,2] ✉

Acquired resistance to chemotherapy is an important clinical problem and can also occur without detectable cytogenetic aberrations or gene mutations. Chronic lymphocytic leukemia (CLL) is molecularly well characterized and has been elemental for establishing central paradigms in oncology. This prompted us to check whether specific epigenetic changes at the level of DNA methylation might underlie development of treatment resistance. We used Illumina Infinium HumanMethylation450 BeadChips to obtain DNA methylation profiles of 71 CLL patients with differential responses. Thirty-six patients were categorized as relapsed/refractory after treatment with fludarabine or bendamustine and 21 of them had genetic aberrations of *TP53*. The other 35 patients were untreated at the time of sampling and 15 of them had genetic aberration of *TP53*. Although we could not correlate chemoresistance with epigenetic changes, the patients were comprehensively characterized regarding relevant prognostic and molecular markers (e.g. IGHV mutation status, chromosome aberrations, *TP53* mutation status, clinical parameters), which makes our dataset a unique and valuable resource that can be used by researchers to test alternative hypotheses.

Background & Summary

Chronic lymphocytic leukemia (CLL) is the most common leukemia in the Western world and mainly affects elderly patients[1]. Its incidence rate was 8.3 cases per 100 000 men and 5.8 cases per 100 000 women in Germany in 2014[2]. CLL is characterized by accumulation of small B lymphocytes with a mature appearance in blood, bone marrow, lymph nodes and other lymphoid tissues[3]. The clinical course of CLL differs depending on the biological characteristics of the disease (hypermutation status of the immunoglobulin heavy-chain genes (IGHV), presence of specific genomic aberrations and/or recurrent mutations in oncogenes and tumor suppressor genes)[4–6]. Some of these genetic features are associated with distinct epigenetic profiles, e.g. CLL tumours with high level of IGHV somatic hypermutation (M-CLL) have distinct DNA methylation patterns compared to CLL tumours with a low or absent IGHV mutational load (U-CLL)[7].

Chemoimmunotherapeutic regimens like fludarabine, cyclophosphamide and rituximab (FCR) or bendamustine and rituximab (BR) achieve durable remissions in the majority of treatment-naïve CLL patients[8–11]. Although novel targeted and effective treatments for CLL were introduced in the past five years, FCR is not inferior to them as first-line therapy in the subgroup of young and fit patients with M-CLL without 17p deletion and/or *TP53* mutation (del(17p)/*TP53*mut)[12,13]. Additionally, the high cost of novel targeted drugs limits their use in developing countries where conventional cytotoxic chemotherapy is still a viable option[14,15]. Thus, drugs like fludarabine and bendamustine will continue to be used in the future for treatment of CLL and development of resistance to these classical chemotherapeutics remains an important problem to study.

Chemorefractoriness of CLL is most often caused by functional impairment of the ATM-p53 DNA damage response pathway, mostly as a result of cytogenetic aberrations or mutations[16,17]. Del(17p) is found in 5% to 10% of patients at diagnosis but in up to 40% of patients relapsing after fludarabine-based treatment regimens[18].

[1]Department of Internal Medicine III, Ulm University, Ulm, Germany. [2]Cooperation Unit "Mechanisms of Leukemogenesis", German Cancer Research Center (DKFZ), Heidelberg, Germany. [3]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ✉e-mail: d.mertens@dkfz-heidelberg.de
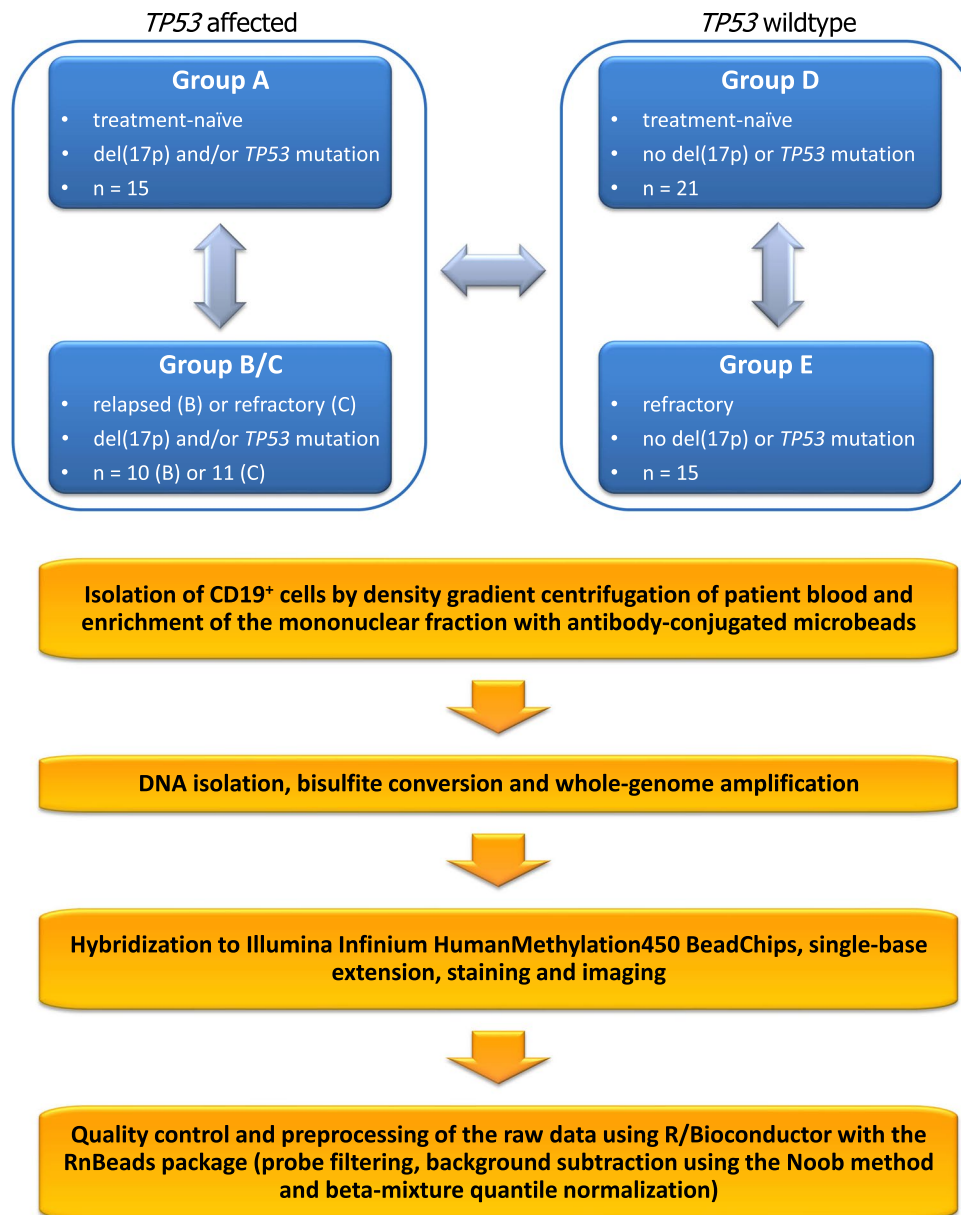
TP53 affected

TP53 wildtype

Group A
- treatment-naïve
- del(17p) and/or *TP53* mutation
- n = 15

Group D
- treatment-naïve
- no del(17p) or *TP53* mutation
- n = 21

Group B/C
- relapsed (B) or refractory (C)
- del(17p) and/or *TP53* mutation
- n = 10 (B) or 11 (C)

Group E
- refractory
- no del(17p) or *TP53* mutation
- n = 15

Isolation of CD19⁺ cells by density gradient centrifugation of patient blood and enrichment of the mononuclear fraction with antibody-conjugated microbeads

DNA isolation, bisulfite conversion and whole-genome amplification

Hybridization to Illumina Infinium HumanMethylation450 BeadChips, single-base extension, staining and imaging

Quality control and preprocessing of the raw data using R/Bioconductor with the RnBeads package (probe filtering, background subtraction using the Noob method and beta-mixture quantile normalization)

Fig. 1 Schematic overview of the study design and experimental procedure.

Del(17p) causes loss of one allele of the tumour suppressor *TP53* but in about 80% of the cases the other allele is also inactivated by somatic mutation[6,18]. Nevertheless, even monoallelic aberrations of *TP53* confer poor prognosis. Interestingly, some cases of chemorefractory CLL show dysfunction of the ATM-p53 pathway without respective genetic lesions[16,17]. Additional genes and pathways have been implicated in development of resistance to fludarabine, although also in these cases mutations are not always detectable[17,19,20]. These observations leave the possibility that chemoresistance in CLL can also be driven by epigenetic mechanisms. In order to find epigenetic changes associated with chemoresistance, we selected samples from patients that were relapsed/refractory after treatment with fludarabine or bendamustine and/or had del(17p)/*TP53*mut, as well as samples from CLL patients without del(17p)/*TP53*mut who had treatment-naïve disease or who achieved prolonged remission after treatment with fludarabine- or bendamustine-based regimens. The grouping of the samples is shown in Fig. 1. This selection of samples allows comparing relapsed/refractory patients to untreated patients after stratification for the presence or absence of aberrations affecting the *TP53* locus. In our opinion, this stratification is important because presence of *TP53* aberrations could obscure the effect of epimutations, as *TP53* aberrations themselves are a strong determinant of chemoresistance[8,21,22]. On the other hand, the chosen design of the study could allow to detect epimutations that additionally occur in the subgroup of *TP53*-disrupted CLL tumours to further reduce their sensitivity to chemotherapy. Genome-wide DNA methylation in all selected samples (N = 72) was quantified using Illumina Infinium HumanMethylation450 BeadChips. The resulting raw signal data and a normalized data matrix are provided here as a resource for studying relationships between epigenetics and chemoresistance in

Characteristic	All patients (N = 71)	A (del(17p), untreated, N = 15)	B (del(17p), relapsed, N = 10)	C (del(17p), refractory, N = 11)	D (no del(17p), chemosensitive, N = 20)	E (no del(17p), refractory, N = 15)	P Value*
Age at sampling, years							0.27
Median	64	66	62	67	62	63	
Range	38–84	49–77	53–68	54–76	38–84	38–71	
IGHV mutational status, number (%)							0.45
Mutated (<98% homology)	13 (18.3)	2 (13.3)	0 (0)	3 (27.3)	5 (25)	3 (20)	
Unmutated (≥98% homology)	58 (81.7)	13 (86.7)	10 (100)	8 (72.7)	15 (75)	12 (80)	
FISH analysis, number (%)							P < 0.0001
Normal karyotype	12 (16.9)	0 (0)	0 (0)	0 (0)	10 (50)	2 (13.3)	
Del(13q14)	46 (64.8)	9 (60)	6 (60)	8 (72.7)	10 (50)	13 (86.7)	
Trisomy 12	7 (9.9)	3 (20)	1 (10)	3 (27.3)	0 (0)	0 (0)	
Del(11q22)	13 (18.3)	3 (20)	3 (33.3)	3 (27.3)	0 (0)	4 (26.7)	
Del(17p13)	35 (49.3)	15 (100)	10 (100)	10 (90.9)	0 (0)	0 (0)	
TP53 mutational status, number (%)							P < 0.0001
Mutated	30 (42.3%)	10 (66.7)	9 (90)	11 (100)	0 (0)	0 (0)	
Unmutated	41 (57.7%)	5 (33.3)	1 (10)	0 (0)	20 (100)	15 (100)	

Table 1. Biological and molecular features of CLL patients included in the study. *All P values are for comparisons across all five groups and are two-sided. P values for numerical variables were calculated with the use of the Kruskal–Wallis test, and P values for categorical variables were calculated with the use of the chi-square test or Fisher's exact test. FISH profiles were summarized according to the hierarchical risk model[4] before performing the test.

CLL. Basic clustering and principal component analyses did not intuitively show grouping of samples according to chemoresistance status. However, we cannot exclude that more sophisticated analyses will be able to extract relevant differences and correlations. Notably, the dataset is unique with the high proportion of patients with del17p and/or mutated *TP53*. This dataset thus allows comparison of epigenetic profiles of CLL patients with negative prognostic markers to profiles of patients with chemosensitive CLL and CLL not harbouring *TP53* defects.

Methods

Patient sample selection and molecular characterization. The biological and molecular characteristics of the 71 CLL patients included in the study are listed in Table 1 and Online-only Table 1. Fifty-one of the patients were subjects of the multi-centre CLL2O clinical trial (clinicaltrials.gov: NCT01392079) and were subdivided here into 4 subgroups depending on their del(17p)/*TP53*mut and treatment/response statuses as follows: groups A (N = 15), B (N = 10) and C (N = 11) consisted of patients with del(17p) and/or *TP53* mutation and group E (N = 15) consisted of patients without del(17p) or *TP53* mutation. Patients in group A were not treated previously but required treatment, patients in group B had relapsed after treatment with fludarabine- or bendamustine-containing regimens and patients in groups C and E were refractory to fludarabine or bendamustine. Additional 20 cases (group D) were patients whose tumours did not harbour del(17p) or *TP53* mutation and who were not previously treated but some of whom required treatment and responded to subsequent therapy with fludarabine- or bendamustine-containing regimens (N = 6, Online-only Table 1). All patients had a confirmed diagnosis of CLL by flow cytometry; their IGHV mutational status and cytogenetics were also determined during the diagnostic workup. Unmutated IGHV gene (≥98% homology to germline) was detected by sequencing in 58 patients (81.7%). Fluorescent *in-situ* hybridization (FISH) analysis revealed the presence of del(17p) in 35 of a total 36 patients in groups A, B and C, as well as the absence of such an aberration in all patients from groups D and E. Mutated *TP53* was detected in 30 of total 36 patients in groups A, B and C, and in none of the patients in groups D and E. One of the patients in group D had two consecutive samples taken with a time difference of 40 months (Online-only Table 1). All patients provided informed consent to subsequent analysis and research in accordance with the Declaration of Helsinki and under a protocol approved by the ethical committee of the University of Ulm.

Sample preparation. Blood samples from CLL patients were subjected to density gradient centrifugation (Pancoll human, #P04-60500, PAN-Biotech, Germany) to isolate peripheral blood mononuclear cells (PBMCs), which were then enriched for CD19+ B cells using CD19 MicroBeads (#130-050-301, Miltenyi Biotec, Germany) and LS columns (#130-042-401, Miltenyi Biotec). The purity of the enriched cell fractions was confirmed using a FACSCalibur flow cytometer (Becton Dickinson & Co.) and a monoclonal mouse anti-human CD19 antibody (clone HD37, DakoCytomation, Denmark). Purified cell samples were flash frozen and stored as dry cell pellets at −80 °C for further analysis.

DNA extraction, bisulfite conversion and methylation level quantification. The 72 frozen cell pellets were processed in 6 batches, taking care that samples from each of the 5 subgroups (A-E) were approximately

equally divided among the 6 batches to mitigate possible batch effects. DNA was extracted from the cell pellets by the Qiagen AllPrep kit (#80204) and quantification and quality control were performed using a NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Scientific, USA). One and a half micrograms of DNA from each sample were sent to the Genomics and Proteomics Core Facility of the German Cancer Research Center (DKFZ) for bisulfite conversion and hybridization to Illumina Infinium HumanMethylation450 BeadChips, according to the manufacturer's instructions. The bisulfite conversion was performed using the EZ DNA Methylation Kit (Zymo Research) and then the converted DNA was whole-genome amplified and fragmented. The processed samples were distributed randomly among 6 Illumina Infinium HumanMethylation450 BeadChips. The core facility was blinded regarding the identity of the samples and the experimental groups to which they belonged. After hybridization, single-base extension and staining, BeadChips were scanned using an Illumina iScan reader, and the fluorescence intensity raw data for each sample was recorded as two IDAT files, one for the green (Cy3) and one for the red (Cy5) channel[23]. Quality control of the whole procedure was performed using the Methylation Module of Illumina's GenomeStudio software.

Data processing and statistics. After acquiring the raw data, we performed quality control, preprocessing and basic analysis using R/Bioconductor with the RnBeads package[24]. Illumina probes known to be cross-reactive or overlapping known SNPs[25] were excluded from analysis. This was also done for probes giving unreliable measurements as determined by the Greedycut algorithm implemented in RnBeads. The data from the remaining probes were subjected to background subtraction using the Noob method[26] and beta-mixture quantile normalization (BMIQ)[27]. In a subsequent step, probes of non-CpG context, probes binding to sequences on sex chromosomes and probes with low standard deviation were filtered out. CpG sites on the sex chromosomes were excluded to avoid gender-specific methylation bias, as groups within our study did not contain equal numbers of males and females. CpG sites with low standard deviation are generally not informative and removing them from the analysis is a common approach to increase power for detection of differentially methylated CpGs and to improve sensitivity of clustering[28,29]. The data obtained by the remaining probes[23] were used in downstream analyses. Methylation levels of CpG sites were calculated as β-values (β = intensity of the methylated allele (M)/ [intensity of the unmethylated allele (U) + intensity of the methylated allele (M) + 100].

Both multidimensional scaling (MDS) and principal-component analysis (PCA) were used as dimension reduction techniques. Hierarchical clustering was carried out using the Manhattan distance metric and complete linkage criteria.

Data records

The complete DNA methylation microarray dataset has been deposited in the NCBI Gene Expression Omnibus (GEO) database and consists of the raw data in the form of 72 pairs (red/green fluorescence) of raw Intensity Data files (.idat), the processed data matrix and a metadata table describing the samples and their groups[23]. For convenience, Online-only Table 1 lists all patients and samples with their characteristics, as well as experimental and analytical procedures and output data file names.

Technical Validation

Quality control of genomic DNA. Genomic DNA 260 nm/280 nm absorbance ratios were determined using a NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Scientific, USA). All samples had ratios in the range 1.8–2.0, as expected for DNA of high purity (Online-only Table 2).

Quality control of bisulfite conversion and Infinium 450k data. Quality control of bisulfite conversion and of data obtained by the Illumina Infinium HumanMethylation450 BeadChips was performed independently by the team of the core facility using the Methylation Module of Illumina's GenomeStudio software (Supplementary File 1 and Online-only Table 3) and by us using the rnb.run.qc command of the RnBeads package (Fig. 2). Both analyses ascertained the correct execution of the separate steps of the whole experimental procedure: bisulfite conversion, hybridization, single-base extension and stripping. Figure 2a,b demonstrate bisulfite conversion efficiency as reported by control probes of Infinium I or II design, respectively. Overall hybridization performance was assessed using synthetic reference targets that are present in the hybridization buffer at three concentrations (low, medium and high) and that resulted in signals with well separable intensity intervals, as expected (Fig. 2c). The extension controls showed high efficiency of extension with any of the 4 nucleotides (Fig. 2d) and the staining controls demonstrated high efficiency and sensitivity of the staining step (Fig. 2e). The overall performance of the assay from amplification to detection is summarized by the signal from probes that query non-polymorphic bases in the genome – one probe for each nucleotide (Fig. 2f).

The Infinium 450k BeadChip contains 65 genotyping probes that are useful for identification of sample mix-ups. These probes produced highly similar signal patterns in two of our samples, which was expected as these two samples (07PB1887 and 10PB6041) came from the same patient (Fig. 3).

Quality control of normalization procedure. Preprocessing of the raw data was performed using R/ Bioconductor with the RnBeads package[24]. Probes known to be cross-reactive (43 230) or overlapping known single-nucleotide polymorphisms (SNPs; 8 704)[25], as well as probes giving unreliable measurements (884 as determined by the Greedycut algorithm) were excluded from analysis. The data from the remaining 432 759 probes were subjected to background subtraction using the Noob method[26] and beta-mixture quantile normalization (BMIQ)[27]. This normalization strategy successfully mitigated the inherent bias in β-value distributions between the two different types of probes (Infinium I and II) that are present on the HumanMethylation450 BeadChip[30], as shown in Fig. 4. In a subsequent step, probes of non-CpG context (1 251), probes binding to sequences on sex chromosomes (9 917) and probes with standard deviation <0.005 (69 867) were filtered out. Thus, the data obtained by the remaining 351 724 probes qualified for downstream analysis.
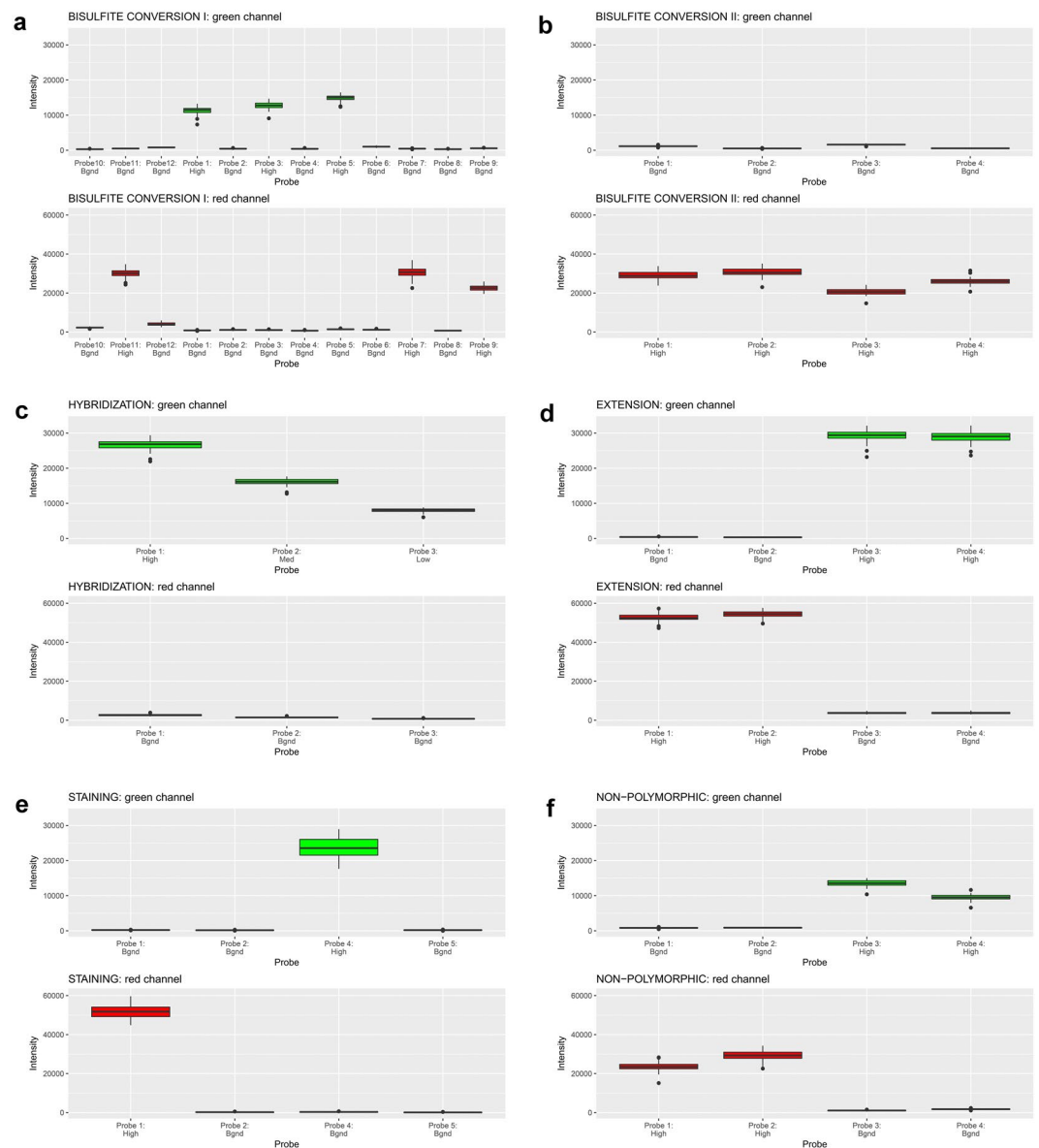
Fig. 2 Distribution (median and range) of signal intensity for quality control probes on Illumina Infinium HumanMethylation450 arrays across all samples and in each of the colour channels (green/red). **(a,b)** Bisulfite conversion efficiency as reported by control probes of Infinium design I **(a)** or II **(b)**. **(c)** Hybridization performance using synthetic reference targets present in the hybridization buffer at three concentrations. **(d)** Efficiency of extension of A, T, C and G nucleotides from hairpin probes (sample-independent). Probe 1 is specific for A, probe 2 for T, probe 3 for C and probe 4 for G. **(e)** Efficiency and sensitivity of the staining step (independent of the hybridization and extension steps). **(f)** Overall efficiency of the procedure estimated by querying non-polymorphic bases in the genome – one probe for each nucleotide. In all plots, labels give the expected intensity level: high, medium (Med), low or background (Bgnd).

Check for batch effects. Dimension reduction techniques are a powerful way of visualizing associations between different variables and global trends in DNA methylation data[24]. Applying PCA on the 10000 most variable CpGs in our data did not result in visible grouping of samples according to the BeadChip that they were applied on (Fig. 5). The lack of batch effects was further verified by Kruskal-Wallis one-way analysis of variance taking into account the first 8 primary components (Table 2).

Biological validation of the DNA methylation data. A technically sound dataset would allow confirmation of known facts. Using our dataset, we could replicate the finding that M-CLL and U-CLL are associated with distinct DNA methylation profiles[7]. In addition to the PCA in Fig. 5 and Table 2, we performed unsupervised hierarchical clustering using the 5000 most variable CpG sites. The resulting dendrograms and heatmap of β-values are presented in Fig. 6. In both analyses, samples were well separated according to IGHV mutation status, with the bigger cluster consisting only of U-CLL cases and the smaller cluster comprising all M-CLL cases
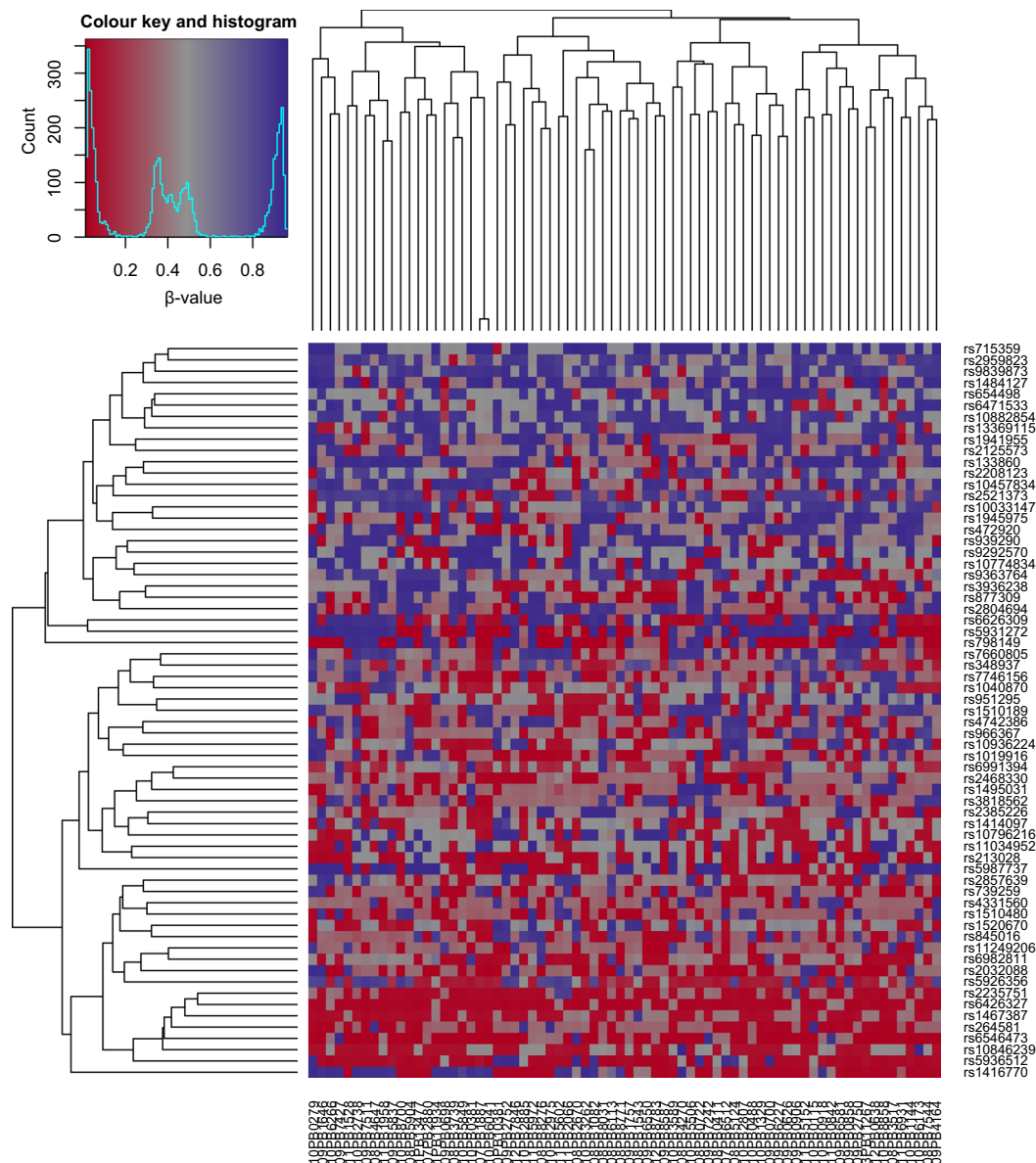
Fig. 3 Heatmap of signal from the 65 probes on the methylation arrays that distinguish single nucleotide polymorphisms (SNPs). Two samples show highly similar patterns, as they originate from the same patient.

plus five U-CLL cases, three of which had a considerable level of IGHV hypermutation (98.2–98.3% homology to germline; see also Online-only Table 1), thus possibly belonging to the intermediate CLL group as defined by Kulis *et al.*[7].

The highly similar DNA methylation profiles of the two serial samples from one patient (Fig. 6) demonstrated the reproducibility and robustness of the whole analytical procedure (the two samples were hybridized to different BeadChips) and in addition supported previous observations that evolution of DNA methylation in CLL is limited to cases that acquire high-risk genetic alterations[31].

Usage notes

In our exploratory analysis, we could not observe obvious clustering of samples based on patients' sensitivity or resistance to chemotherapy (Figs. 5 and 6). Accordingly, a differential methylation test using the limma method with the threshold for difference of β-values set to 0.1 and the false discovery rate set to 0.05 could not find any differentially methylated CpGs between chemoresistant and untreated/chemosensitive patients, neither in the subgroup of patients with del(17p)/*TP53*mut (groups B and C vs. group A), nor in the subgroup without these aberrations (group E vs. group D). Nevertheless, as multiple testing correction inflates the type II error rate considerably, we cannot exclude that some of the tested CpGs have truly different methylation between the groups and a causative role in chemoresistance development. In this regard, our dataset can be a valuable resource for conducting hypothesis-driven research addressing questions of chemoresistance in CLL.
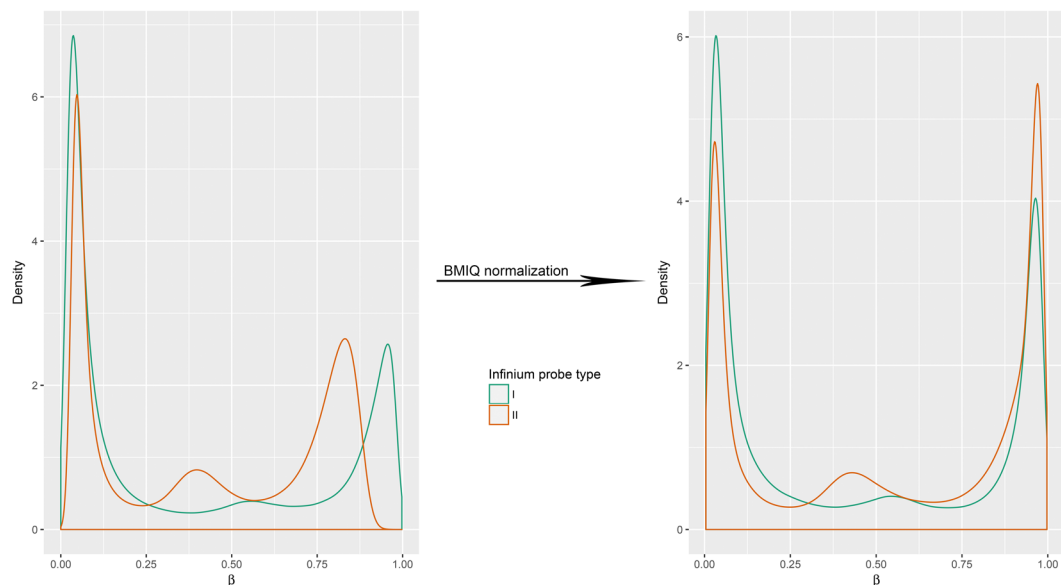
Fig. 4 Density plots of the β-values distribution for the Infinium I and II probes before and after background subtraction and beta-mixture quantile normalization (BMIQ).
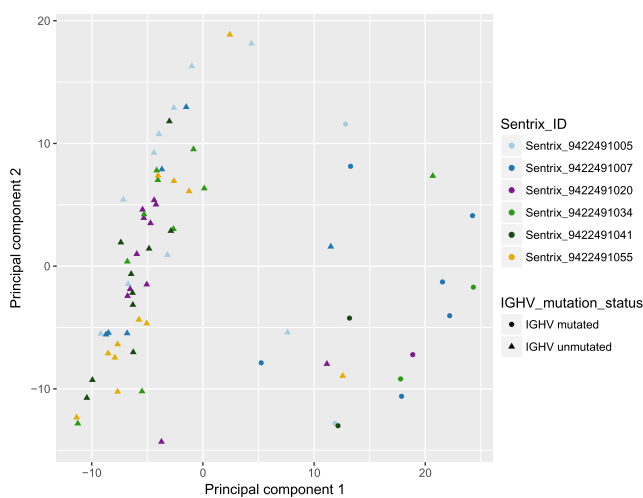

Fig. 5 Principal component analysis (PCA) showing grouping of the 72 samples based on the 10000 most variable CpGs within the dataset. Samples are coloured according to the BeadChip (Sentrix_ID) that they were applied on. Samples of any given colour (batch) do not form separate clusters, whereas principal component 1 distinguishes M-CLL from U-CLL patients ($P = 7.15 \times 10^{-8}$, Wilcoxon rank sum test).

	Principal component							
Trait	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
BeadChip number (Sentrix_ID)*	0.15	0.26	0.21	0.22	0.24	0.68	0.65	0.27
IGHV mutation status^	7.15×10^{-8}	0.08	0.15	0.73	0.78	0.65	0.33	0.62

Table 2. Table of p-values for associations of the first 8 principal components with *IGHV* mutation status as an intrinsic characteristic of the samples or with the BeadChip that the samples were applied on as an extrinsic variable (batch). *P-values from a Kruskal-Wallis one-way analysis of variance. ^P-values from a two-sided Wilcoxon rank sum test.

Additionally, thanks to the rich annotation of the samples, our dataset can be used to explore associations of DNA methylation with other markers of prognostic or predictive value in CLL, e.g. presence of specific chromosome aberrations (see Table 1 and Online-only Table 1). Such analyses should be performed with the necessary
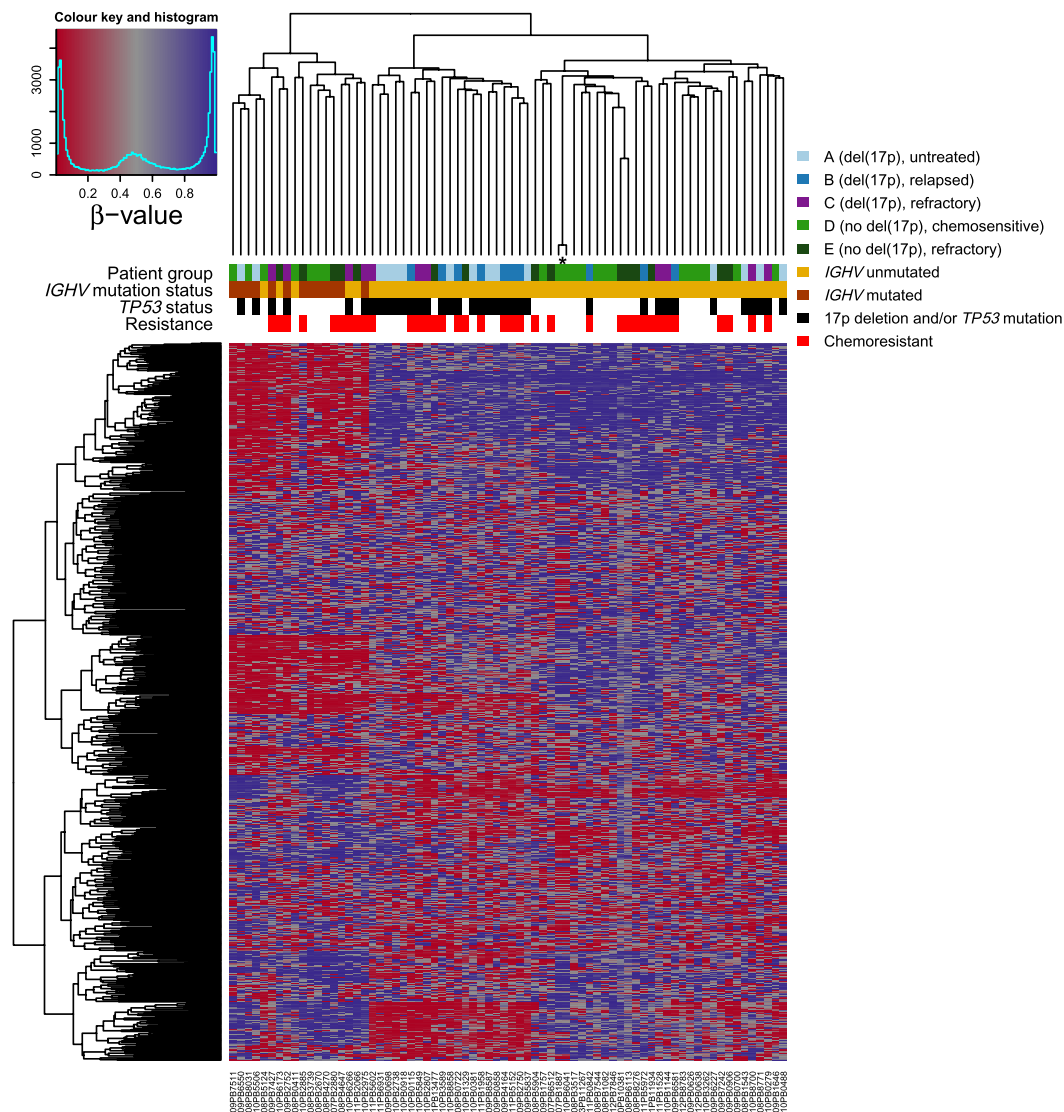
Fig. 6 Unsupervised hierarchical clustering of the 72 CLL samples based on the β-values of the 5000 most variable CpG sites. Clustering was based on Manhattan distance with complete linkage. Columns represent samples and rows CpGs. Two of the samples originate from the same patient (see the main text) and are marked by an asterisk. The high similarity of DNA methylation patterns between them affirms the reproducibility of the methodology.

caution and subsequent validation of the findings in an independent clinical cohort[32]. Conversely, our dataset can serve as the validation dataset for findings originating from other clinical cohorts.

DNA methylation is highly informative when analyzed together with additional layers of the epigenomic regulatory landscape. That is why we recommend that any CpGs of interest found to be differentially methylated between groups be analyzed in the context of published whole-genome maps of histone modifications, chromatin accessibility and chromatin states of CLL and normal B cells[33]. Data from the ENCODE project can also be useful if parallels with a broader variety of cell types are sought[34,35]. Additional insights can be gained if tools like HOMER are used to identify transcription factor recognition motifs around CpGs of interest.

References

1. Rai, K. R. & Jain, P. Chronic lymphocytic leukemia (CLL)-Then and now. *Am J Hematol* **91**, 330–340 (2016).
2. German Population-based Cancer Registry. *German Centre for Cancer Registry Data (ZfKD) at the Robert Koch Institute*, https://www.krebsdaten.de/Krebs/EN/Database/databasequery_step1_node.html (2017).
3. Zenz, T., Mertens, D., Kuppers, R., Dohner, H. & Stilgenbauer, S. From pathogenesis to treatment of chronic lymphocytic leukaemia. *Nat Rev Cancer* **10**, 37–50 (2010).
4. Dohner, H. *et al*. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* **343**, 1910–1916 (2000).

5. Nadeu, F. *et al*. Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia* **32**, 645–653 (2018).
6. Zenz, T. *et al*. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol* **28**, 4473–4479 (2010).
7. Kulis, M. *et al*. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* **44**, 1236–1242 (2012).
8. Eichhorst, B. *et al*. First-line chemoimmunotherapy with bendamustine and rituximab versus fludarabine, cyclophosphamide, and rituximab in patients with advanced chronic lymphocytic leukaemia (CLL10): an international, open-label, randomised, phase 3, non-inferiority trial. *Lancet Oncol* **17**, 928–942 (2016).
9. Fischer, K. *et al*. Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. *Blood* **127**, 208–215 (2016).
10. Fischer, K. *et al*. Bendamustine in combination with rituximab for previously untreated patients with chronic lymphocytic leukemia: a multicenter phase II trial of the German Chronic Lymphocytic Leukemia Study Group. *J Clin Oncol* **30**, 3209–3216 (2012).
11. Keating, M. J. *et al*. Early results of a chemoimmunotherapy regimen of fludarabine, cyclophosphamide, and rituximab as initial therapy for chronic lymphocytic leukemia. *J Clin Oncol* **23**, 4079–4088 (2005).
12. Shanafelt, T. D. *et al*. Ibrutinib–Rituximab or Chemoimmunotherapy for Chronic Lymphocytic Leukemia. *N Engl J Med* **381**, 432–443 (2019).
13. Yosifov, D. Y., Wolf, C., Stilgenbauer, S. & Mertens, D. From Biology to Therapy: The CLL Success Story. *HemaSphere* **3**, e175 (2019).
14. Hilal, T., Betcher, J. A. & Leis, J. F. Economic Impact of Oral Therapies for Chronic Lymphocytic Leukemia-the Burden of Novelty. *Curr Hematol Malig Rep* **13**, 237–243 (2018).
15. Shanafelt, T. D. *et al*. Impact of ibrutinib and idelalisib on the pharmaceutical cost of treating chronic lymphocytic leukemia at the individual and societal levels. *J Oncol Pract* **11**, 252–258 (2015).
16. te Raa, G. D. *et al*. Assessment of p53 and ATM functionality in chronic lymphocytic leukemia by multiplex ligation-dependent probe amplification. *Cell Death Dis* **6**, e1852 (2015).
17. Zenz, T. *et al*. Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the contribution of 17p deletion, TP53 mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial. *Blood* **114**, 2589–2597 (2009).
18. Yu, L. *et al*. Survival of Del17p CLL Depends on Genomic Complexity and Somatic Mutation. *Clin Cancer Res* **23**, 735–745 (2017).
19. Moussay, E. *et al*. Determination of genes and microRNAs involved in the resistance to fludarabine *in vivo* in chronic lymphocytic leukemia. *Mol Cancer* **9**, 115 (2010).
20. Pandzic, T. *et al*. Transposon Mutagenesis Reveals Fludarabine Resistance Mechanisms in Chronic Lymphocytic Leukemia. *Clin Cancer Res* **22**, 6217–6227 (2016).
21. Hallek, M. *et al*. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376**, 1164–1174 (2010).
22. Malcikova, J. *et al*. Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival, and response to DNA damage. *Blood* **114**, 5307–5314 (2009).
23. Yosifov, D. Y. *et al*. DNA methylation of chronic lymphocytic leukemia with differential response to chemotherapy. *Gene Expression Omnibus*, https://identifiers.org/geo:GSE136724 (2020).
24. Assenov, Y. *et al*. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* **11**, 1138–1140 (2014).
25. Price, M. E. *et al*. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**, 4 (2013).
26. Triche, T. J. Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90 (2013).
27. Teschendorff, A. E. *et al*. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
28. Sartor, M. A. *et al*. Genome-wide methylation and expression differences in HPV(+) and HPV(−) squamous cell carcinoma cell lines are consistent with divergent mechanisms of carcinogenesis. *Epigenetics* **6**, 777–787 (2011).
29. Wang, X., Laird, P. W., Hinoue, T., Groshen, S. & Siegmund, K. D. Non-specific filtering of beta-distributed data. *BMC Bioinformatics* **15**, 199 (2014).
30. Dedeurwaerder, S. *et al*. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* **3**, 771–784 (2011).
31. Oakes, C. C. *et al*. Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov* **4**, 348–361 (2014).
32. Kraft, P., Zeggini, E. & Ioannidis, J. P. Replication in genome-wide association studies. *Stat Sci* **24**, 561–573 (2009).
33. Beekman, R. *et al*. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat Med* **24**, 868–880 (2018).
34. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
35. Ernst, J. *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

Acknowledgements

Author contributions

Study design: D.M., J.B., S.S. and D.Y.Y. Laboratory work: D.Y.Y. Data processing and analysis: D.Y.Y., J.B. and D.M. Data interpretation: all authors. Manuscript drafting: D.Y.Y. and D.M. Review of manuscript content: all authors. Approval of the final version: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41597-020-0456-0.

Correspondence and requests for materials should be addressed to D.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.