Regular Article

DNA methylation-based profiling of pediatric T-ALL reveals *SPI1*/PU.1 involvement in T-ALL and T-cell differentiation

Running head: DNA methylation analysis in pediatric T-ALL

Shunsuke Kimura,[1,2] Masafumi Seki,[1] Tomoko Kawai,[3] Kenichi Yoshida,[4] Tomoya Isobe,[1] Masahiro Sekiguchi,[1] Kentaro Watanabe,[1] Yasuo Kubota,[1] Yasuhito Nannya,[4] Hiroo Ueno,[4,5] Yusuke Shiozawa,[4] Hiromichi Suzuki,[4] Yuichi Shiraishi,[6] Kentaro Ohki,[7] Motohiro Kato,[7] Katsuyoshi Koh,[8] Ryoji Kobayashi,[9] Takao Deguchi,[10] Yoshiko Hashii,[11] Toshihiko Imamura,[12] Atsushi Sato,[13] Nobutaka Kiyokawa,[7] Atsushi Manabe,[14] Masashi Sanada,[15] Marc R. Mansour,[16] Akira Ohara,[17] Keizo Horibe,[15] Masao Kobayashi,[2] Akira Oka,[1] Yasuhide Hayashi,[18] Satoru Miyano,[6] Kenichiro Hata,[3] Seishi Ogawa,[4,19] and Junko Takita[1,5]

[1]Department of Pediatrics, The University of Tokyo, Tokyo, Japan; [2]Department of Pediatrics, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan; [3]Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan; [4]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [5]Department of Pediatrics,

Kyoto University, Kyoto, Japan; [6]Human Genome Center Institute of Medical Science, The University of Tokyo, Japan; [7]Department of Pediatric Hematology and Oncology Research, National Research Institute for Child Health and Development, Tokyo, Japan; [8]Department of Hematology/Oncology, Saitama Children's Medical Center, Saitama, Japan; [9]Department of Pediatrics, Sapporo Hokuyu Hospital, Sapporo, Japan; [10]Children's Cancer Center, National Center for Child Health and Development, Tokyo, Japan; [11]Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan; [12]Department of Pediatrics, Kyoto Prefectural University of Medicine, Graduate School of Medical Science, Kyoto, Japan; [13]Department of Hematology and Oncology, Miyagi Children's Hospital, Sendai, Japan; [14]Department of Pediatrics, St. Luke's International Hospital, Tokyo, Japan; [15]Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan; [16]Department of Haematology, University College London Cancer Institute, London, United Kingdom; [17]Department of Pediatrics, Toho University, Tokyo, Japan; [18]Gunma Children's Medical Center, Maebashi, Japan; [19]Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institutet, Huddinge, Sweden

Correspondence: Junko Takita, Department of Pathology and Tumor Biology, Graduate

37 School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501,

38 Japan; E-mail: jtakita@kuhp.kyoto-u.ac.jp, Phone: +81-75-753-7531, Fax: +81-75-752-

39 2361

40

41 **Word counts**

42 Main text: 3092

43 Abstract: 242

44 **Number of figures and tables**: 5 figures, 0 tables

45 **Number of references**: 41

46

47

Key Points

* T-ALL classification by DNA methylation status showed strong association with T-cell differentiation, prognosis, and genetic features.

*Gain of methylation of SPI1/PU.1 binding sites is important for T-cell lineage commitment as well as silencing of *SPI1* expression.

Abstract

The genetic profiles of pediatric T-cell acute lymphoblastic leukemia (T-ALL) have been well studied; however, epigenetic profiles and their potential contribution to the clinicopathological features of T-ALL remain poorly elucidated. We performed genome-wide DNA methylation analysis using an EPIC methylation array in 79 pediatric T-ALL cases and combined these results with our previous expression and mutation data. Pediatric T-ALL clustered into four distinct subtypes of DNA methylation profiles exhibiting remarkable correlation with genetic signatures, expression features, differentiation status, and clinical outcomes. We identified several important differentially expressed genes regulated by DNA methylation in each cluster, such as *TAL1* and its downstream *ALDH1A2*. Notably, cases with *SPI1* fusions and *SPI1* high expression formed a specific cluster (Cluster 4), being hypomethylated and indicative of

66 a significantly worse prognosis (log-rank $P = 4.4 \times 10^{-7}$) with enrichment of the
67 RAS/MAPK and myeloid/NKT-cell-related pathways. Correlating with a previous DNA
68 methylation study using the 450K methylation array, the methylation status at EZH2
69 binding sites classified T-ALL cases into CpG island methylator phenotype-positive
70 (Cluster 2 and Cluster 3) or -negative (Cluster 1 and Cluster 4). Furthermore, using
71 probes at *SPI1*/PU.1 binding sites, we distinguished T-ALL before (Cluster 2 and Cluster
72 4) and after (Cluster 1 and Cluster 3) T-cell commitment, an important event in T-cell
73 differentiation. In addition to *SPI1* silencing, changes in methylation status at
74 *SPI1*/PU.1 binding sites during T-cell commitment suggest a prominent role of
75 *SPI1*/PU.1 in T-cell differentiation, and thus greatly impact pediatric T-ALL profiling.
76
77

Introduction

In recent years, the development of next-generation sequencing technologies has enabled genome-wide analysis of pediatric cancers,[1,2] including T-cell acute lymphoblastic leukemia (T-ALL).[3-5] Genetic alterations of *NOTCH1* and *CDKN2A/B* inactivation have been identified as T-ALL hallmarks,[6] reported to appear later during leukemic development.[7,8] Additionally, almost half of T-ALL cases harbor in-frame fusions such as *STIL-TAL1*.[9] These genetic alterations and fusion genes are well correlated with expression profiling in T-ALL.[3,4,10,11] In our previous study, we identified recurrent *SPI1*/PU.1 fusions predicting uniformly dismal outcomes in approximately 4% of Japanese pediatric T-ALL cases.[4] The *SPI1* fusion cases exhibited unique cytological and gene expression profiles different from previously reported subtypes (early T-cell precursor [ETP] phenotype, TLX, TAL1-RA, and TAL1-RB).

Over the past decades, global DNA methylation patterns in pediatric T-ALL have been investigated using 27K and 450K methylation arrays.[12,13] These studies classified T-ALL into two groups based on the CpG island methylator phenotype (CIMP), namely, a hypermethylated CIMP-positive group and a hypomethylated CIMP-negative group. Probes related to polycomb repressive complex (PRC) targets were particularly enriched in the selected CIMP probes. Moreover, CIMP-negative cases exhibited a

methylation pattern similar to normal CD3+ T-cells and CD34+ cells and exhibited a worse prognosis among minimal residual disease (MRD)-positive high-risk T-ALL patients (MRD ≥ 0.1% at day 29). The development of more precise methylation analysis using the EPIC methylation array and whole genome bisulfite sequencing has improved the genome coverage of regulatory regions and genes. The EPIC array contains almost twice the number of probes compared with the previous 450K array, covering more than 850,000 CpG sites.[14] These advances are expected to generate a more profound elucidation of the molecular basis of T-ALL.

In the present study, we investigated the epigenetic profiles of pediatric T-ALL patients and their potential contribution to the clinicopathological features of pediatric T-ALL using the EPIC array in combination with our previous whole transcriptome sequencing (WTS) and targeted capture sequencing (TCS) data.[4]

Methods

Patient samples

We enrolled 79 pediatric T-ALL patients (Supplemental Table 1), mainly from two large cohorts from the Tokyo Children's Cancer Study Group (TCCSG)[15] and the Japan Association of Childhood Leukemia Study (JACLS).[16] All cases were previously analyzed

by WTS (Supplemental Table 2) and TCS for 158 ALL-related genes and regions (Supplemental Tables 3 and 4).[4] Samples for DNA methylation analysis were collected from the study participants after receiving written, informed consent according to the protocols approved by the Human Genome, Gene Analysis Research Ethics Committee of the University of Tokyo and other participating institutes.

DNA methylation analysis

DNA methylation profiles were analyzed using the Infinium MethylationEPIC BeadChip Kit (Illumina, San Diego, CA) after bisulfite conversion with the EpiTect Fast Bisulfite Conversion Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The raw data used for this study are available from the DNA Data Bank of Japan (DDBJ) (accession number JGAS00000000138). Methylation data for CD3+ T-cells, CD34+ cells, and human embryonic stem cells (ESCs) obtained from the Gene Expression Omnibus (GSE49618[17]) and Zendo (Zenodo.1095572[18]) public databases were used as normal reference samples. The ß- and M-values were calculated after normalization and imputation using the R package ChAMP 2.8.9 (Bioconductor, https://www.bioconductor.org/).[19] Cluster stability was ascertained via consensus clustering with 1,000 iterations using the R package ConsensusClusterPlus (Bioconductor).[20] Heatmaps were generated by the R package pheatmap (CRAN,

https://cran.r-project.org/) using ß-values. Differentially methylated probes (DMPs) and regions were identified by ChAMP software using M- and ß-values, respectively, and differentially expressed genes were extracted from our previous expression data[4] using the R package DESeq2 (Bioconductor)[21] Differential gene expression and methylation were assessed for each probe, and starburst plots were generated by comparing a target cluster with the other clusters.

Data from our previous study

Our current study used our previously reported sequencing data of TCS (mutations, structural variants, and copy number alterations) and WTS (gene fusions and expression) as well as clinical and flow cytometry data.[4] As before, the normalized read counts data from WTS was used as expression data. We also adopted five expression clusters (ETP phenotype, TLX, TAL1-RA, TAL1-RB, and *SPI1* fusions) based on the WTS expression data described in our last study (Supplemental Table 1).

Functional enrichment analysis

Significantly enriched pathways were analyzed using the R package gage (Bioconductor)[22] via a one-on-one comparison of all possible combinations between a target cluster and the other clusters. The R package LOLA (Bioconductor)[23] was used to determine the DMP-related transcription factor binding and other elements of known

150 regulatory function of the target cluster with a specific level of methylation (high or low)
151 compared to the other clusters.

152 *Statistical analysis*

153 Statistical analyses were performed using R 3.4.2. We used Fisher's exact test to compare
154 categorical variables. Multiple comparisons using the Holm method were performed with
155 the R package fmsb (CRAN). To compare nonparametric continuous variables, Mann–
156 Whitney U–tests were performed. The Kaplan–Meier method was used to estimate
157 survival rates, and the log-rank test was used for comparing survival curves. Statistical
158 results were considered significant at $P < 0.05$.

159

160 **Results**

161 *Classification of pediatric T-ALL based on DNA methylation profiling*

162 We performed genome-wide DNA methylation profiling using the EPIC BeadChip
163 (Illumina) covering more than 850,000 methylation sites (probes). Following filtering,
164 imputation, and normalization using the default ChAMP parameters[19] to remove failed
165 probes and avoid gender bias, 796,832 probes remained. Of them, we selected 939 of the
166 most variable methylated probes based on a standard deviation (SD) > 0.37 for the ß-
167 values across the samples (Supplemental Table 5). Unsupervised consensus clustering

of the 79 pediatric T-ALL cases (including seven *SPI1* fusion cases) using these 939 probes identified four distinct sample clusters, showing an intermediate-methylation pattern in Cluster 1 (n = 39) and Cluster 2 (n = 20), hypermethylation in Cluster 3 (n = 11), and hypomethylation in Cluster 4 (n = 9) (Figure 1; Supplemental Figure 1). Intriguingly, the methylation status was completely opposite between Cluster 1 and Cluster 2 in the selected probes.

Genetic features of each methylation cluster

To investigate the genetic profiles in the newly-obtained DNA methylation clusters, we performed a combined analysis of the methylation data with our previous WTS and TCS data (Figure 1). Cluster 1, showing an intermediate-methylation status, was characterized by a high expression of *TAL1* with *STIL-TAL1* fusion and heterozygous somatic mutations in a specific non-coding site upstream of the *TAL1* locus.[24] *PTEN* and *PIK3R1* mutations were particularly enriched in this cluster, indicating involvement of the PI3K-AKT pathway, and there were abnormalities in cell cycle regulators such as *CDKN2A*. Almost all cases of TAL1-RA and TAL1-RB expression clusters were classified into this methylation Cluster 1. The TAL1-RA and TAL1-RB expression clusters were first defined by Soulier[11] and are both associated with strong *TAL1* expression but exhibit different T-cell developmental stages.[4]

Mutations in the JAK-STAT pathway and epigenetic regulator genes were characteristic of methylation Cluster 2. Most cases of TLX and ETP phenotype clusters as per the previous expression classification[4] were grouped together in methylation Cluster 2. Although there were cases of gene fusions involving *TLX1* and *TLX3* regions, no other fusions were detected in this cluster except *RUNX1-AFF3* and *NUP214-ABL1,* due to inefficient alignment to the T-cell receptor (TCR) and *BCL11B* regions in WTS.

In hypermethylated Cluster 3, JAK-STAT pathway and epigenetic regulator abnormalities were frequently observed as in Cluster 2, but the mutated genes were different (*IL7R*, *JAK1*, *JAK3*, *PHF6*, and *SUZ12* in Cluster 2; *DNM2*, *PHF6*, and *EZH2* in Cluster 3). The detection of HOX-related fusions, such as *PICALM-MLLT10*, were consistent with high *HOXA9* expression; however, the mixed presence of cases of TLX, ETP phenotype, and TAL1-RA expression clusters were indicative of the heterogeneous features of Cluster 3.

All *SPI1* fusion cases (n = 7) were classified into hypomethylation Cluster 4 as well as two cases that did not have *SPI1* fusions and were previously classified into the ETP phenotype expression cluster,[4] representing very high *SPI1* expression induced by hypomethylation in the *SPI1* promoter region (Supplemental Figure 2). RAS, *CDKN2A*, and *GATA3* alterations were enriched in Cluster 4.

204 ***Detection of differentially expressed genes regulated by DNA methylation in each***

205 ***methylation cluster***

206 To explore differentially expressed genes regulated by DNA methylation, we combined

207 our previous expression data from WTS[4] with the methylation data. In Cluster 1, the

208 expression of *TAL1* and its downstream *ALDH1A2* in addition to the PI3K-AKT pathway

209 and *SIX6*-related genes was associated with promoter hypomethylation when compared

210 with the other three clusters (Figure 2A; Supplemental Figure 3). Gene set enrichment

211 analysis revealed significant enrichment in cell cycle and TAL1-related pathways

212 (Supplemental Table 6). On the other hand, *ALDH1A2* expression was extremely low

213 with hypermethylation status in Cluster 2 and resulted in the suppression of TAL1-

214 related pathways (Figure 2B; Supplemental Figure 4). This is consistent with the

215 opposite methylation status in the selected probes between Cluster 1 and Cluster 2. In

216 contrast, *DEPTOR*, a direct NOTCH1 target and one of the components of the mTOR

217 complex, was highly expressed with hypomethylation status in Cluster 2. (Figure 2B;

218 Supplemental Figure 4). Because of its heterogeneity, we were unable to identify notable

219 genes and pathways characterizing Cluster 3 (Figure 2C; Supplemental Figure 5). In

220 addition to *SPI1*, the expression of RAS, NF-κB, and cell growth-related genes was

221 regulated by the methylation status of Cluster 4 (Figure 2D; Supplemental Figure 6).

222 Pathway analysis revealed significant enrichment in the myeloid/NKT cell-related, *SPI1*,
223 and RAS/MAPK-related pathways (Supplemental Table 7). Thus, our classification
224 based on DNA methylation correlated with the results of T-ALL profiling of fusions,
225 mutations, and gene expression.

226 ***Relevance of T-cell differentiation in DNA methylation classification of pediatric T-ALL***

227 Analysis of immunophenotypic data that was available in 57 of the T-ALL cases studied
228 revealed that most cases in Cluster 1 and Cluster 3 were CD4/8 double-positive (DP),
229 whereas most cases in Cluster 2 were CD4/8 double-negative (DN) (Figure 3). Among the
230 DP cases, TCRα/ß was positive in Cluster 1 but negative in Cluster 3 except for one case,
231 indicating that TAL1-related Cluster 1 was more differentiated. The DN cases in Cluster
232 2 frequently expressed B-cell marker CD79a or myeloid markers CD11b, CD13, and
233 CD33 other than T-cell markers. The negativity of CD1a and TCR-α/ß was also indicative
234 of undifferentiated T-ALL before commitment, consistent with a high expression of
235 Phase 1 genes such as *LYL1*, *HHEX*, and *MEF2C* (Supplemental Figures 7 and 8). Most
236 Cluster 4 cases, especially those with *SPI1* fusions, also represented an uncommitted DN
237 profile and were CD1a-positive. Taken together, these results suggested that Cluster 2
238 cases were in an earlier T-cell development stage.

239 ***Clinical features of each methylation cluster***

240 Among the four methylation clusters, relatively high white blood cell counts >10 × 10^9/L
241 were frequently observed in Cluster 1 (Supplemental Table 8). There was no significant
242 difference in clinical features according to sex, age at diagnosis, response to steroids, and
243 central nervous system involvement. The clinical outcomes of each methylation subgroup
244 revealed extremely dismal outcomes of Cluster 4 cases irrespective of the presence of
245 *SPI1* fusions (log-rank $P = 4.4 \times 10^{-7}$; Figure 4). Eight out of nine patients (except one
246 *STMN1-SPI1* fusion case) relapsed and died within three years of diagnosis. Although
247 the number of analyzed cases was small (n = 11), all Cluster 3 patients were alive despite
248 our treatment protocol without MRD-directed risk stratification.

249 ***Methylation status at EZH2 and SPI1 binding sites impacted the methylation-based***
250 ***classification of T-ALL***

251 To explore the differences between the selected CIMP probes in a previously reported T-
252 ALL methylation study using the 450K array[13] and our present study EPIC CIMP probes,
253 we applied available previously reported 450K CIMP probes to our cohort. Among the
254 1293 probes selected for CIMP classification in the 450K methylation study,[13] 1205 were
255 also contained in the EPIC array. After normalization and imputation, 1178 probes were
256 available for further analysis (Figure 5; Supplemental Table 9). Only 23 probes were
257 shared in common between the previous CIMP probes (n = 1178) and present EPIC CIMP

258 probes (n = 939) because the EPIC CIMP probes contained more newly designed probes
259 in the opensea regions that were particularly enriched in Region C (Figure 5). Similar to
260 the previous studies,[13] supervised hierarchical clustering of these 1178 probes classified
261 the 79 cases of T-ALL in our cohort as either CIMP-positive or -negative (Supplemental
262 Figure 9A). Cluster 1 and Cluster 4 were CIMP-negative as well as normal CD3+ T-cells
263 and CD34+ cells,[17] whereas Cluster 2 and Cluster 3 were CIMP-positive. CIMP-negative
264 cases showed significantly poor prognosis, even after excluding *SPI1* fusion-containing
265 Cluster 4 cases (Supplemental Figures 9B–9E). These results agreed with the previous
266 studies.[25]

267 We further compared the EPIC methylation data of our T-ALL cohort to that of
268 normal cells (ESCs, CD34+ hematopoietic cells, and CD3+ T-cells) in a public database[18]
269 by applying our available CIMP probes (n = 846) (Figure 5D). In immature CD34+ cells,
270 probes in Region A and Region C showed a hypomethylation status similar to the
271 methylation profiles of Cluster 4 cases, although the probes in Region A were highly
272 methylated in immunophenotypically more immature Cluster 2 cases. LOLA[23] genomic
273 locus overlap enrichment analysis revealed that the binding sites of PRC target genes or
274 components such as EZH2 were significantly enriched in probes in Region A (Figures 5C
275 and 5D; Supplemental Table 10). On the other hand, mature cells after commitment

(Cluster 1, Cluster 3, and CD3+ T-cells) showed a higher methylation status in Region C probes compared with uncommitted immature cells (Cluster 2, Cluster 4, and CD34+ cells), although ESCs before hematopoietic commitment showed hypermethylation status in Region C (Figure 5D). Most Region C probes targeted the opensea regions, where transcriptional factor binding sites related to T-cell differentiation, such as *SPI1*/PU.1, were significantly enriched (Supplemental Table 11).

Discussion

In this study, we classified pediatric T-ALL cases into four distinct clusters based on DNA methylation profiles using the EPIC array. These methylation subgroups each contained T-cells at distinctive differentiation stages, being, in order of maturity (from undifferentiated to mature) Cluster 2, Cluster 4, Cluster 3, and Cluster 1 (Figure 3B), which was also confirmed by the expression of specific transcription factors in each differentiation stage[26] (Supplemental Figures 7 and 8). In Cluster 1, *ALDH1A2* was highly expressed due to hypomethylation status of the promoter region. In T-ALL, *ALDH1A2* transcription is induced by a cryptic promoter present in the second intron through the binding of a TAL1-LMO2-GATA3 complex.[27] *ALDH1A2* is one of the TAL1 targets,[27-29] encoding an enzyme that catalyzes the synthesis of retinoic acid (RA) from

294 retinaldehyde. RA is reported to promote T-cell proliferation and survival,[30,31] while
295 inhibiting thymocyte differentiation,[32] which explains the effects of citral (3,7-dimethyl-
296 2,6-octadienal), an ALDH1A inhibitor.[31,33] This is in contrast to other tumor cells, such
297 as those of neuroblastoma, in which RA exerts an antiproliferative and prodifferentiative
298 effect.[34] In addition to PI3K-AKT inhibitor, citral might be a suitable target therapy for
299 this subgroup. We observed a high expression of *DEPTOR*, a direct NOTCH1 target, in
300 Cluster 2. Activated DEPTOR has been reported to promote cell proliferation and
301 survival via AKT activation in T-ALL,[35] suggesting that the PI3K-AKT pathway also
302 plays an important role in Cluster 2. This is consistent with the frequent detection of
303 *NOTCH1* mutations in this cluster despite few *PETN* mutations. Thus, a combination of
304 target therapies against JAK-STAT and PI3K-AKT pathways might be promising in
305 Cluster 2 cases. In Cluster 3, abnormalities of epigenetic regulators, such as *EZH2*, were
306 often detected with relatively higher expression levels of *HOXA9*, which is usually
307 expressed only during the immature stage. HOXA9 dysregulation induces ectopic
308 activation of FOS/JUN via the STAT5 pathway.[36] *DNM2* mutations were also frequently
309 observed in this cluster, leading to activation of the IL7R/JAK/STAT pathway.[37] These
310 mechanisms might be involved in leukemogenesis of this cluster. Cluster 4 was
311 characterized by a hypomethylation status, showing a similar methylation status to

312 normal CD34+ cells. Notably, all *SPI1* fusion positive T-ALL cases were classified into
313 this methylation subtype, indicating their unique methylation profiles. *SPI1* expression
314 was reported to maintain stemness of leukemia stem cells in murine T-ALL model,[38]
315 which might be associated with an extremely poor prognosis in Cluster 4 cases.
316 Enrichment of the RAS/MAPK pathway and the high expression of NF-κB-related genes
317 might be a possible target for this intractable subgroup.

318 A previous methylation study using 27K and 450K arrays[12,13] classified T-ALL
319 into CIMP-positive and -negative based on the methylation status at the binding sites of
320 PRC target genes. Hypomethylation status at these sites (CIMP-negative) was related
321 to a worse outcome; however, no significant association between ETP phenotype and
322 CIMP class was found. [13] These results were consistent with our results showing worse
323 outcomes in Cluster 1 and Cluster 4 cases with hypomethylation status in Region A,
324 where the binding sites of PRC components were significantly enriched. Our more
325 precise DNA methylation analysis using the EPIC array enabled us to further classify T-
326 ALL based on the methylation status in opensea regions (Region C). This unveiled
327 drastic changes in methylation status at *SPI1* binding sites before and after commitment.
328 Commitment is an important checkpoint during T-cell development, inducing drastic
329 changes in gene networks, including the repression of *SPI1*.[26,39,40] Our findings of a loss

of *SPI1* binding sites due to gain of methylation after commitment as well as silencing of *SPI1* expression suggest that DNA methylation strongly impacts T-cell differentiation, even in T-ALL. Thus, the four methylation clusters in our study were defined by the methylation status at the binding sites of both EZH2 and *SPI1*.

Intriguingly, Region C was methylated in ESCs but not in CD34+ cells or uncommitted T-ALL cases. Changes of methylation status from hyper- to hypomethylation at *SPI1* binding sites might be required for the development of ESCs to immature hematopoietic cells. The importance of *SPI1* in this line could be explained by the fact that *SPI1* is a pioneer transcription factor that has the ability to bind its target sites even when these sites are located within nucleosome-packed chromatin.[26,41] *SPI1* silencing during commitment might induce the release of *SPI1* from its target sites, reverting them back to hypermethylation status. However, further research is warranted to elucidate the association of *SPI1* with DNA methylation in T-cell development.

In the present study, we classified pediatric T-ALL based on DNA methylation profiles. The more precise EPIC array provided important findings that indicated a strong association between the methylation classification of T-ALL, T-cell differentiation, prognosis, and genetic features. Thus, further integrated analyses using sorted thymocytes from different stages of T-cell development are necessary. Although our

cohort was limited, our results suggested that the biological phenotype of T-ALL is mediated by both genetic and epigenetic mechanisms. To develop a new therapeutic strategy for T-ALL, explorations for aberrant DNA methylation along with genetic alterations might be helpful.

Acknowledgments

The authors are grateful to M. Matsumura, K. Yin, and F. Saito for their excellent technical assistance. The authors also wish to express their appreciation to K. Chiba, and H. Tanaka (The University of Tokyo) for the supercomputer. This work was supported by KAKENHI grant numbers 17H04224 (J.T.) and 15H05909 (S.O.) from the Japan Society of Promotion of Science, by Japan Agency for Medical Research and Development (AMED) Practical Research for Innovative Cancer Control and Project for Cancer Research and Therapeutic Evolution (P-CREATE) (16cm0106509h001 (J.T.), and by the Friends of Leukemia Research Fund (S.K.). This research also used the computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (hp140230, hp160219, and hp150232) (S.M.).

Authorship Contributions

Contribution: S.K. and J.T. wrote the manuscript; S.K., M. Seki, K.Y., T. Isobe, Y.N., H.U., and M. Sanada analyzed the data; M. Kato, K.K., R.K., Y. Hashii, T. Imamura, A.S., N.K., A.M., A. Ohara, and K. Horibe collected the data and samples; S.K., M. Seki, T.K., K.O., T.D., N.K., and K. Hata performed the experiments; Y. Shiozawa, H.S., Y. Shiraishi, and S.M. developed the bioinformatics pipelines; M. Kobayashi, A. Oka, Y. Hayashi, S.O., and J.T. gave conceptual advice; J.T. designed the study. All authors read and approved the final manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

References

1. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018;555(7696):371–376.

2. Gröbner SN, Weischenfeldt J, Worst BC, et al. The landscape of genomic alterations across childhood cancers. 2018;555(7696):321–327.

3. Liu Y, Easton J, Shao Y, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. 2017;49(8):1211–1218.

4. Seki M, Kimura S, Isobe T, et al. Recurrent SPI1 (PU.1) fusions in high-risk

384 pediatric T cell acute lymphoblastic leukemia. Nat Genet. 2017;49(8):1274–1281.

385 5. Girardi T, Vereecke S, Sulima SO, et al. The T-cell leukemia-associated

386 ribosomal RPL10 R98S mutation enhances JAK-STAT signaling. Leukemia.

387 2018;32(3):809–819.

388 6. Belver L, Ferrando A. The genetics and mechanisms of T cell acute

389 lymphoblastic leukaemia. Nat. Rev. Cancer. 2016;16(8):494–507.

390 7. De Bie J, Demeyer S, Alberti-Servera L, et al. Single-cell sequencing reveals the

391 origin and the order of mutation acquisition in T-cell acute lymphoblastic leukemia.

392 Leukemia. 2018.

393 8. Mansour MR, Duke V, Foroni L, et al. Notch-1 mutations are secondary events

394 in some patients with T-cell acute lymphoblastic leukemia. Clinical Cancer Research.

395 2007;13(23):6964–6969.

396 9. Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute

397 lymphoblastic leukemia. 2012;122(10):3398–3406.

398 10. Ferrando AA, Neuberg DS, Staunton J, et al. Gene expression signatures define

399 novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell.

400 2002;1(1):75–87.

401 11. Soulier J, Clappier E, Cayuela JM, et al. HOXA genes are included in genetic

402 and biologic networks defining human acute T-cell leukemia (T-ALL). Blood.
403 2005;106(1):274–286.
404 12. Borssén M, Palmqvist L, Karrman K, et al. Promoter DNA methylation pattern
405 identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. PLoS
406 ONE. 2013;8(6):e65373.
407 13. Borssén M, Haider Z, Landfors M, et al. DNA Methylation Adds Prognostic
408 Value to Minimal Residual Disease Status in Pediatric T-Cell Acute Lymphoblastic
409 Leukemia. Pediatric Blood & Cancer. 2016;63(7):1185–1192.
410 14. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina
411 MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling.
412 Genome Biol. 2016;17(1):208.
413 15. Takahashi H, Kajiwara R, Kato M, et al. Treatment outcome of children with
414 acute lymphoblastic leukemia: the Tokyo Children's Cancer Study Group (TCCSG) Study
415 L04-16. Int J Hematol. 2018;108(1):98–108.
416 16. Sakamoto K, Imamura T, Kihira K, et al. Low Incidence of Osteonecrosis in
417 Childhood Acute Lymphoblastic Leukemia Treated With ALL-97 and ALL-02 Study of
418 Japan Association of Childhood Leukemia Study Group. J. Clin. Oncol. 2018;36(9):900–
419 907.

17. Qu X, Davison J, Du L, et al. Identification of differentially methylated markers among cytogenetic risk groups of acute myeloid leukemia. Epigenetics. 2015;10(6):526–535.

18. Tejedor JR, Bueno C, Cobo I, et al. Epigenome-wide analysis reveals specific DNA hypermethylation of T cells during human hematopoietic differentiation. Epigenomics. 2018;10(7):903–923.

19. Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics. 2017;33(24):3982–3984.

20. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–1573.

21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

22. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009;10(1):161.

23. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 2016;32(4):587–589.

24. Mansour MR, Abraham BJ, Anders L, et al. Oncogene regulation. An oncogenic

438 super-enhancer formed through somatic mutation of a noncoding intergenic element.
439 Science. 2014;346(6215):1373–1377.
440 25. Gabriel AS, Schwalbe EC, Lafta FM, et al. Epigenetic landscape correlates with
441 genetic subtype but does not predict outcome in childhood acute lymphoblastic leukemia.
442 Epigenetics. 2015;10(8):717–726.
443 26. Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the
444 course to T cell identity. Nat Rev Immunol. 2014;14(8):529–545.
445 27. Ono Y, Fukuhara N, Yoshie O. TAL1 and LIM-only proteins synergistically
446 induce retinaldehyde dehydrogenase 2 expression in T-cell acute lymphoblastic leukemia
447 by acting as cofactors for GATA3. Molecular and Cellular Biology. 1998;18(12):6939–6950.
448 28. Sanda T, Lawton LN, Barrasa MI, et al. Core transcriptional regulatory circuit
449 controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer
450 Cell. 2012;22(2):209–221.
451 29. Tan SH, Lawton LN, Yam AWY, et al. TRIB2 reinforces the oncogenic
452 transcriptional program controlled by the TAL1 complex in T-cell acute lymphoblastic
453 leukemia. 2015;30(4):959–962.
454 30. Engedal N, Gjevik T, Blomhoff R, Blomhoff HK. All-trans retinoic acid
455 stimulates IL-2-mediated proliferation of human T lymphocytes: early induction of cyclin

456 D3. J. Immunol. 2006;177(5):2851–2861.

457 31. Longville BAC, Anderson D, Welch MD, Kees UR, Greene WK. Aberrant
458 expression of aldehyde dehydrogenase 1A (ALDH1A) subfamily genes in acute
459 lymphoblastic leukaemia is a common feature of T-lineage tumours. Br. J. Haematol.
460 2015;168(2):246–257.

461 32. Zhou X, Wang W, Yang Y. The expression of retinoic acid receptors in thymus of
462 young children and the effect of all-transretinoic acid on the development of T cells in
463 thymus. J. Clin. Immunol. 2008;28(1):85–91.

464 33. Kikonyogo A, Abriola DP, Dryjanski M, Pietruszko R. Mechanism of inhibition
465 of aldehyde dehydrogenase by citral, a retinoid antagonist. Eur. J. Biochem.
466 1999;262(3):704–712.

467 34. Sidell N, Altman A, Haussler MR, Seeger RC. Effects of retinoic acid (RA) on
468 the growth and phenotypic expression of several human neuroblastoma cell lines. Exp.
469 Cell Res. 1983;148(1):21–30.

470 35. Hu Y, Su H, Liu C, et al. DEPTOR is a direct NOTCH1 target that promotes cell
471 proliferation and survival in T-cell leukemia. Oncogene. 2017;36(8):1038–1047.

472 36. de Bock CE, Demeyer S, Degryse S, et al. HOXA9 Cooperates with Activated
473 JAK/STAT Signaling to Drive Leukemia Development. 2018;8(5):616–631.

474 37. Tremblay CS, Brown FC, Collett M, et al. Loss-of-function mutations of
475 Dynamin 2 promote T-ALL by enhancing IL-7 signalling. Leukemia. 2016;30(10):1993–
476 2001.
477 38. Zhu H, Zhang L, Wu Y, et al. T-ALL leukemia stem cell "stemness" is
478 epigenetically controlled by the master regulator SPI1. Elife. 2018;7:285.
479 39. Yui MA, Feng N, Rothenberg EV. Fine-scale staging of T cell lineage
480 commitment in adult mouse thymus. J. Immunol. 2010;185(1):284–293.
481 40. Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic
482 transformations of genome-wide epigenetic marking and transcriptional control
483 establish T cell identity. Cell. 2012;149(2):467–482.
484 41. Montecino-Rodriguez E, Casero D, Fice M, Le J, Dorshkind K. Differential
485 Expression of PU.1 and Key T Lineage Transcription Factors Distinguishes Fetal and
486 Adult T Cell Development. J. Immunol. 2018;200(6):2046–2056.
487

Figure Legends

Figure 1. Cluster classification of pediatric T-ALL based on the DNA methylation status of the EPIC array combined with previous expression and mutation data. Following normalization using the ChAMP software package, we selected 939 of the most variable methylated probes based on the ß-values (SD > 0.37) across the samples. Unsupervised consensus clustering classified 79 cases of pediatric T-ALL into four distinct clusters (top panel). Five expression clusters (TAL1-RA, TAL1-RB, ETP phenotype, TLX, and *SPI1* fusion) were defined in our previous study[2] based on WTS expression data. The expression data (Z-score) of each gene was calculated using DESeq2. All fusion and mutation data were reported previously.[2]

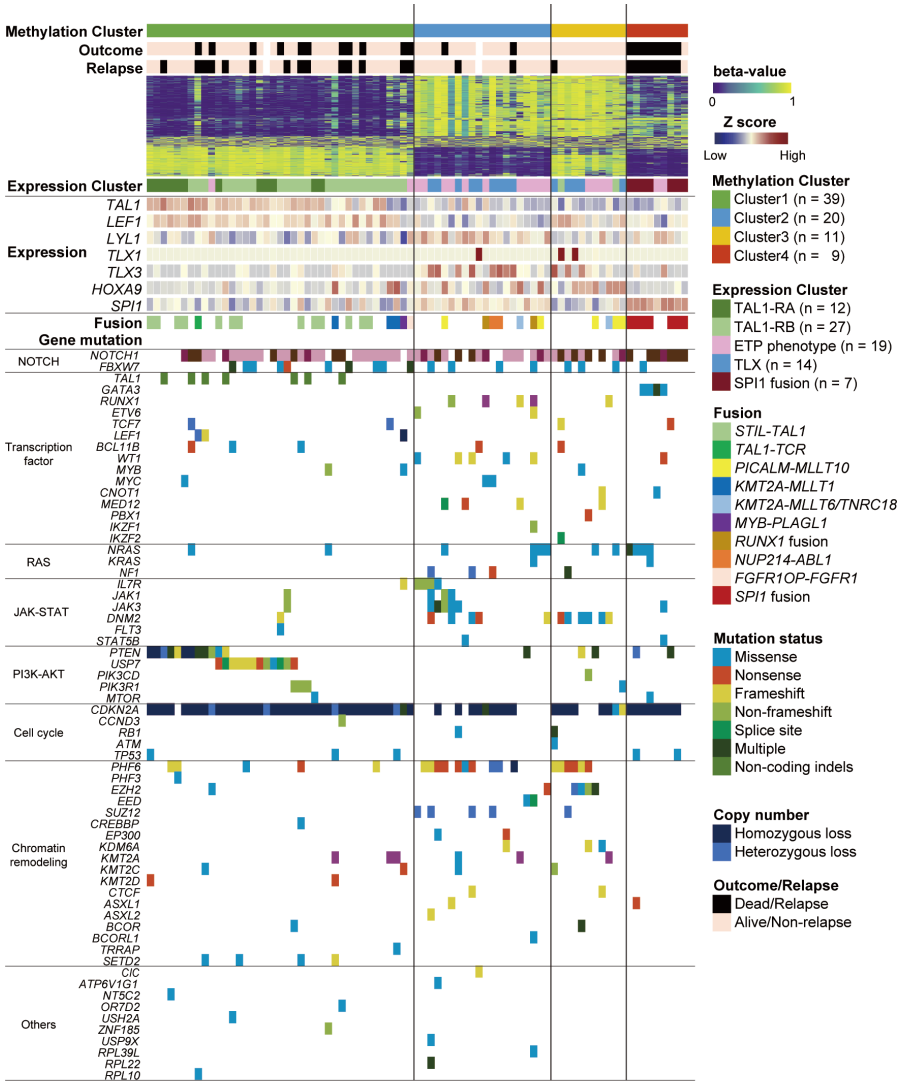
Methylation Cluster
Outcome
Relapse
Expression Cluster
Expression
TAL1
LEF1
LYL1
TLX1
TLX3
HOXA9
SPI1
Fusion
Gene mutation
NOTCH
NOTCH1
FBXW7
Transcription factor
TAL1
GATA3
RUNX1
ETV6
TCF7
LEF1
BCL11B
WT1
MYB
MYC
CNOT1
MED12
PBX1
IKZF1
IKZF2
RAS
NRAS
KRAS
NF1
JAK-STAT
IL7R
JAK1
JAK3
DNM2
FLT3
STAT5B
PI3K-AKT
PTEN
USP7
PIK3CD
PIK3R1
MTOR
Cell cycle
CDKN2A
CCND3
RB1
ATM
TP53
Chromatin remodeling
PHF6
PHF3
EZH2
EED
SUZ12
CREBBP
EP300
KDM6A
KMT2A
KMT2C
KMT2D
CTCF
ASXL1
ASXL2
BCOR
BCORL1
TRRAP
SETD2
Others
CIC
ATP6V1G1
NT5C2
OR7D2
USH2A
ZNF185
USP9X
RPL39L
RPL22
RPL10
beta-value
0
1
Z score
Low
High
Methylation Cluster
Cluster1 (n = 39)
Cluster2 (n = 20)
Cluster3 (n = 11)
Cluster4 (n = 9)
Expression Cluster
TAL1-RA (n = 12)
TAL1-RB (n = 27)
ETP phenotype (n = 19)
TLX (n = 14)
SPI1 fusion (n = 7)
Fusion
STIL-TAL1
TAL1-TCR
PICALM-MLLT10
KMT2A-MLLT1
KMT2A-MLLT6/TNRC18
MYB-PLAGL1
RUNX1 fusion
NUP214-ABL1
FGFR1OP-FGFR1
SPI1 fusion
Mutation status
Missense
Nonsense
Frameshift
Non-frameshift
Splice site
Multiple
Non-coding indels
Copy number
Homozygous loss
Heterozygous loss
Outcome/Relapse
Dead/Relapse
Alive/Non-relapse

Figure 2. Starburst plot for DNA methylation and expression in each DNA methylation cluster against the other clusters. Each circle in the starburst plots indicates a probe of the DNA methylation array. The circle size represents the *P* value of the corresponding differentially expressed gene. The horizontal axis indicates the difference (FC) in methylation status, and the vertical axis shows the difference in expression level on a log scale. Probes in the upper left section were hypomethylated with high expression compared with the other clusters. Colored genes indicate those characteristic of (A) Cluster 1, (B) Cluster 2, (C) Cluster 3, and (D) Cluster 4 . FC, fold change.

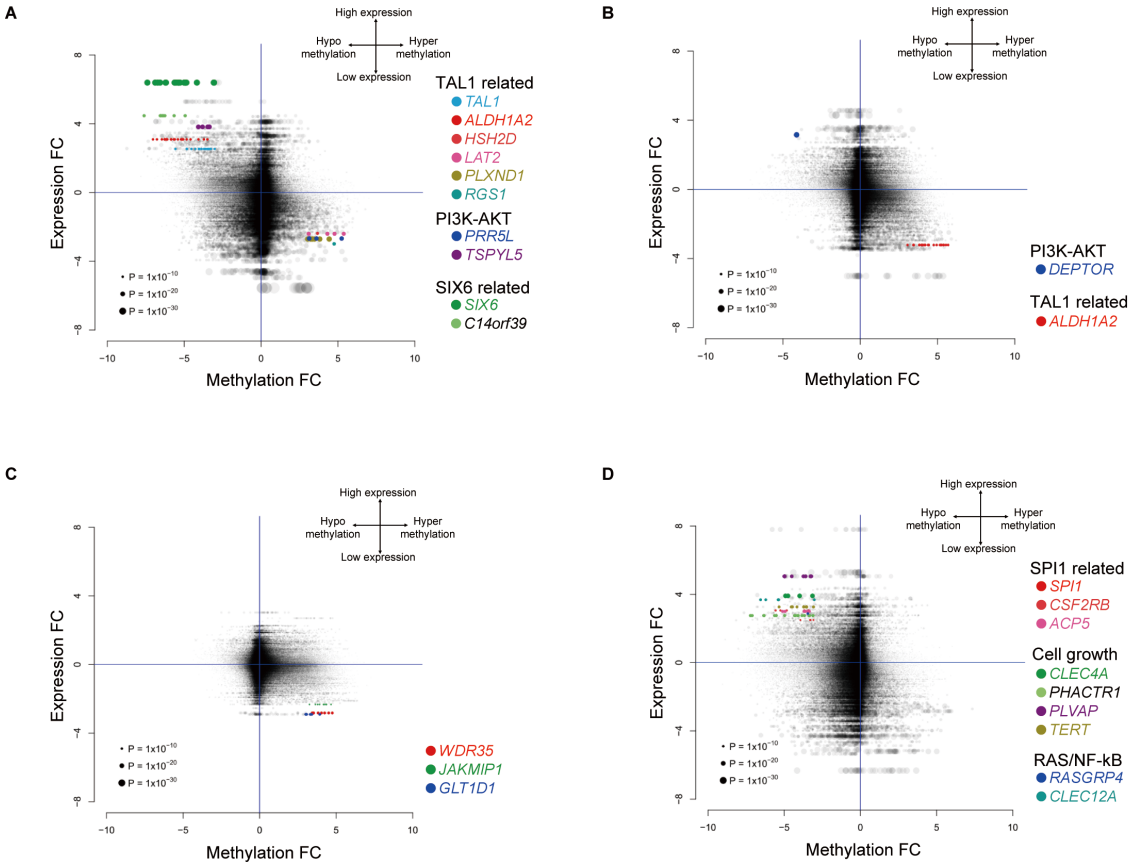

Figure 3. Relationship between DNA methylation profile clusters and T-cell differentiation. Immunophenotypic data of each methylation cluster (A) revealed that most cases in Cluster 1 and Cluster 3 were in the CD4/8 DP stage. (B) Cases in Cluster 1 exhibited a late cortical thymocyte profile after TCR rearrangement (TCRαβ+). The majority of Cluster 2 and Cluster 4 cases showed uncommitted DN T-cell profiles, but those of Cluster 2 were in an earlier CD1a-negative stage.

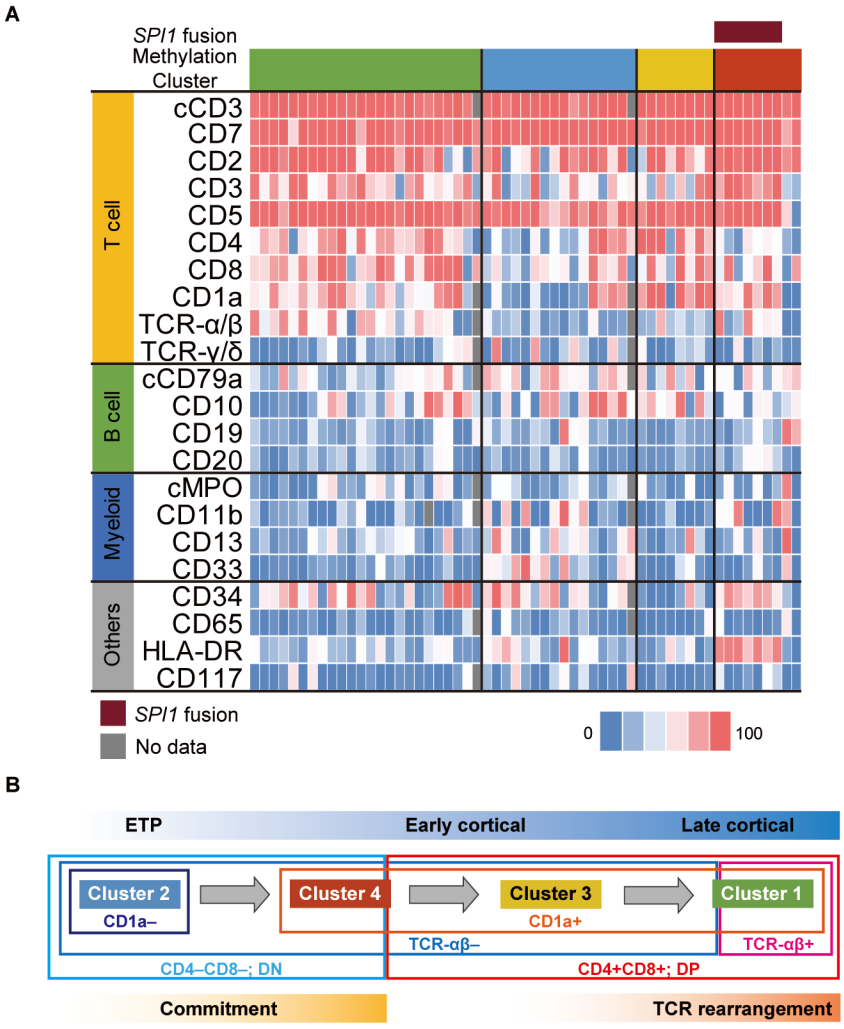

Figure 4. Clinical outcomes of each methylation cluster. (A) Overall survival and (B) disease-free survival of each methylation cluster revealed a significantly poor prognosis in Cluster 4 cases, whereas all cases in Clusters 2 and 3 represented favorable outcomes (median follow-up of 6 years). *P* values are based on the log-rank test.

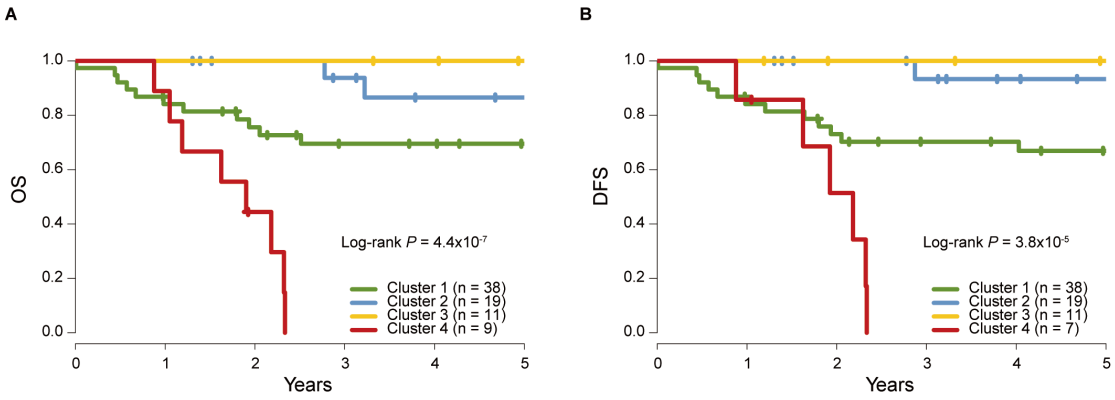

Figure 5. Features of selected probes in the EPIC array and comparison of methylation status in normal cells and T-ALL cases. (A) Selected CIMP probes in a previous 450K study and our EPIC study. Only 23 probes were in common. (B) Among the selected CIMP probes in both the 450K and EPIC array, more than half targeted CpG islands. The EPIC CIMP probes contained more probes targeted at the open sea region than the 450K probes. (C) The EPIC CIMP probes were classified into three groups (Regions A–C) according to the differences in methylation status among the T-ALL cases. Almost all probes in Region A were targeted at CpG islands. Region B probes were the most variable methylated probes across all samples but were useless for T-ALL classification. The

Region C probes consisted of a lot of opensea region probes. (D) Methylation status of normal cells (human ESC cells, CD34 positive cells, and CD3-positive SP T-cells) using public EPIC methylation data. Following normalization and imputation of our available selected CIMP probes, most probes in Region B were lost, and the remaining 846 probes were used for depicting heatmaps. SP, single positive.