Aging of preleukemic thymocytes drives CpG island hypermethylation in T-cell acute lymphoblastic leukemia

Juliette Roels[1,2*], Morgan Thénoz[1,2*], Bronisława Szarzyńska[3], Mattias Landfors[4], Stien De Coninck[1,2], Lisa Demoen[1,2], Lien Provez[1,2], Anna Kuchmiy[2,5], Steven Strubbe[5], Lindy Reunes[1,2], Tim Pieters[1,2], Filip Matthijssens[1,2], Wouter Van Loocke[1,2], Büşra Erarslan-Uysal[6,7], Paulina Richter-Pechańska[6,7], Ken Declerck[8], Tim Lammens[2,9], Barbara De Moerloose[2,9], Dieter Deforce[10], Filip Van Nieuwerburgh[10], Laurence C Cheung[11,12], Rishi S Kotecha[11,12], Marc R. Mansour[13], Bart Ghesquière[14], Guy Van Camp[15], Wim Vanden Berghe[8], Jerzy R. Kowalczyk[16], Tomasz Szczepański[17], Utpal P Davé[18], Andreas E. Kulozik[6,7], Steven Goossens[1,2,5], David J. Curtis[19], Tom Taghon[5], Małgorzata Dawidowska[3], Sofie Degerman[4,20] and Pieter Van Vlierberghe[1,2]

[1]Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

[2]Cancer Research Institute Ghent (CRIG), Ghent, Belgium

[3]Institute of Human Genetics, Polish Academy of Sciences, Poznań, Poland

[4]Department of Medical Biosciences, Umeå University, Umeå, Sweden

[5]Department of Diagnostic Sciences, Ghent University, Ghent, Belgium

[6]Department of Pediatric Oncology, Hematology, and Immunology, University of Heidelberg, and Hopp Children's Cancer Center at NCT Heidelberg, Heidelberg, Germany

[7]Molecular Medicine Partnership Unit (MMPU), European Molecular Biology Laboratory (EMBL), University of Heidelberg, Heidelberg, Germany

[8]Laboratory of Protein Chemistry, Proteomics and Epigenetic Signaling (PPES) and Integrated Personalized and Precision Oncology Network (IPPON), Department of Biomedical Sciences, University of Antwerp (UA), Belgium.

[9]Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, Ghent, Belgium

[10]Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium

[11]Telethon Kids Cancer Centre, Telethon Kids Institute, University of Western Australia, Perth, WA, Australia

[12]School of Pharmacy and Biomedical Sciences, Curtin University, Perth, WA, Australia

[13]Department of Haematology, University College London Cancer Institute, London, England

[14]Metabolomics Expertise Center, VIB Center for Cancer Biology, Leuven, Belgium

[15]Center of Medical Genetics, University of Antwerp, Antwerp, Belgium

[16]Department of Pediatric Hematology, Oncology and Transplantology, Medical University of Lublin, Poland

[17]Department of Pediatric Hematology and Oncology, Zabrze, Medical University of Silesia, Katowice, Poland

[18]Roudebush Veterans Affairs Medical Center and Indiana University School of Medicine, Indianapolis, USA

[19]Australian Centre for Blood Diseases (ACBD), Monash University, Australia

[20]Department of Clinical Microbiology, Umeå University, Umeå, Sweden

*authors equally contributed to this work

Correspondence

Pieter Van Vlierberghe, PhD
Ghent University, Department of Biomolecular Medicine
Corneel Heymanslaan 10
9000 Ghent, Belgium
Tel: 003293321043

Email: pieter.vanvlierberghe@ugent.be

Conflicts of Interest The authors declare no potential conflicts of interest.

Word count text: 5997
Word count abstract: 241
Figure count: 5
Table count: 1
Reference count: 67
Running title: DNA methylation uncovers the pre-malignant history of T-ALL
Key words: T-ALL, DNA methylation, aging, self-renewing thymocytes, Decitabine

Abstract

Cancer cells display DNA hypermethylation at specific CpG islands in comparison to their normal healthy counterparts, but the mechanism that drives this so-called CpG island methylator phenotype (CIMP) remains poorly understood. Here, we show that CpG island methylation in human T-cell acute lymphoblastic leukemia (T-ALL) mainly occurs at promoters of Polycomb Repressor Complex 2 (PRC2) target genes that are not expressed in normal or malignant T-cells and which display a reciprocal association with H3K27me3 binding. In addition, we revealed that this aberrant methylation profile reflects the epigenetic history of T-ALL and is established already in pre-leukemic, self-renewing thymocytes that precede T-ALL development. Finally, we unexpectedly uncover that this age-related CpG island hypermethylation signature in T-ALL is completely resistant to the FDA-approved hypomethylating agent Decitabine. Altogether, we here provide conceptual evidence for the involvement of a pre-leukemic phase characterized by self-renewing thymocytes in the pathogenesis of human T-ALL.

Statement of significance

We developed a DNA methylation signature that reveals the epigenetic history of thymocytes during T-cell transformation. This human signature was recapitulated by murine self-renewing pre-leukemic thymocytes that build an age-related CpG island hypermethylation phenotype, providing conceptual evidence for the involvement of a pre-leukemic thymic phase in human T cell leukemia.

Introduction

During the last decade, aberrant DNA methylation has been identified as a hallmark of human cancer and several studies have highlighted the promising potential of DNA methylation as a clinically or diagnostically relevant biomarker(1). In comparison to their putative normal healthy counterparts, cancer cells generally display DNA hypermethylation at specific CpG islands, but the actual mechanism that drive this so-called CpG island methylator phenotype (CIMP) remains poorly understood(2). Despite the fact that DNA hypomethylating agents are actively used in the clinic for the treatment of some hematological malignancies, their putative effects on leukemia-specific DNA methylation signatures and gene expression remain unclear(1,2).

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological cancer for which prognostically relevant CIMP subtypes have also been described(3). Indeed, based on the methylation status of about 1000 promoter-associated CpG sites, $CIMP^+$ T-ALLs were associated with a better event-free and overall survival as compared to $CIMP^-$ leukemias(3). Notably, these findings were recently confirmed in independent pediatric(4) and adult(5) T-ALL cohorts, further reinforcing the idea that aberrant DNA methylation might act as a clinically relevant biomarker in human T-ALL. However, the mechanism driving the CpG island methylator phenotype in T-cell leukemia has remained elusive so far.

In this study, we performed a comprehensive DNA methylome analysis of normal and malignant T-cells in both human and murine settings in an attempt to increase our understanding on the origin of aberrant DNA methylation in T-ALL.

Results

A CpG island and Open Sea DNA methylation signature in human T-ALL

Previous studies have shown that molecular genetic subtypes of human T-ALL are associated with an arrest at specific stages of normal human T cell differentiation(6,7). Here, we performed DNA methylation profiling using the 850k EPIC array platform on 109 primary T-ALLs(8) (**Supplementary Table 1**) and 10 stages of sorted human thymocytes (9,10), reflecting the normal counterparts of this

disease(6,7) (**Supplementary Table 2**). Using unsupervised clustering of the 5000 most variably methylated CpGs, we built a methylation-based signature that we termed COSMe (**C**pG island and **O**pen **S**ea **Me**thylation) referring to CpGs located both inside CpG islands (n=2713), outside of CpG islands (n=1245) and CpGs that flank CpG islands (CpG shores and shelves, n=1042) (**Supplementary Table 3**). This COSMe signature divides T-ALLs in two methylation-based categories, i.e. COSMe Type I and Type II (**Figure 1A, Supplementary Table 3**), based on three main clusters of CpGs (Cluster A, B and C). The subdivision of T-ALLs in COSMe-I and COSMe-II is dominated by sites located in Cluster A, which are mainly located at CpG islands and largely correspond to the CIMP classification that has previously been established in T-ALL(3) (**Figure 1A, Supplementary Table 3**). In line with previous reports on the clinical relevance of CIMP status both in pediatric(3,4) and adult(5) T-ALL, we also observed a significantly higher cumulative incidence of relapse in the $CIMP^-$ T-cell leukemias from our cohort treated according to the ALL IC-BFM 2002/2009 protocol (p=0.04, **Supplementary Figure 1, Supplementary Table 4-5**).

At the genetic level, the COSMe subtypes were differentially enriched for genetic defects previously associated with T-ALL biology and CIMP status(7,8,11) (**Figure 1A, Supplementary Figure 2**). COSMe-I T-ALLs were significantly enriched for *TAL1* rearrangements, whereas COSMe-II T-ALLs mainly consisted of leukemias with aberrant activation of *TLX1*, *TLX3*, *NKX2.1* or *HOXA* (**Figure 1A, Supplementary Figure 2, Supplementary Table 6**)(8). In addition, other genetic defects, which have previously been associated with the *TAL1* gene expression cluster(7), were also more prevalent in COSMe-I T-ALLs, including 6q deletions ($p < 0.01$) and *PTEN* deletions (p=0.071) (**Figure 1A, Supplementary Figure 2, Supplementary Table 6**). In contrast, COSMe-II T-ALLs showed enrichment for genetic aberrations previously associated with double negative or early-cortical T-ALLs(12,13), including 5q deletions and loss-of-function alterations targeting *WT1*, *CTCF* and the PRC2 complex members *EZH2*, *SUZ12* or *EED* (**Figure 1A, Supplementary Figure 2, Supplementary Table 6**).

COSMe-I T-ALLs showed characteristic low-level methylation compared to COSMe-II T-ALLs in the CpG-island dominated Cluster A (**Figure 1A**). Nevertheless, COSMe-I T-ALLs already displayed increased DNA methylation of these Cluster A

CpGs in comparison to normal developing T cells, which uniformly lack any methylation at these sites (**Figure 1B**).

Besides Cluster A, the COSMe signature also included 2 clusters that mainly consisted of Open Sea CpG sites (Clusters B and C; **Supplementary Table 3; Figure 1A**). Remarkably, Cluster B methylation gradually increased during T cell development (**Figure 1C**). Cluster B CpG sites displayed hypermethylation in almost all COSMe-I T-ALLs, but only in a subset of COSMe-II leukemias (**Figure 1A and 1C**).

To further study this, we performed clustering of COSMe-II leukemias solely based on cluster B methylation levels and identified COSMe-II Cluster B^+ and COSMe-II Cluster B^- T-ALLs (**Supplementary Figure 3A**). Of note, COSMe-II Cluster B^- leukemias showed a trend for enrichment of genetic defects previously associated with early immature T-ALL and ETP-ALL (12,13), including 5q deletions, loss-of-function alterations targeting *WT1*, *CTCF*, PRC2 (*EZH2*, *SUZ12* or *EED)*, as well as leukemias displaying aberrant activation of *HOXA* genes (**Supplementary Figure 3A-B**). In contrast, there was a trend for higher prevalence of 6q deletions, loss-of-function alterations targeting PTEN, as well as T-ALLs showing aberrant expression of *NKX2.1*, in COSMe-II Cluster B^+ leukemias (**Supplementary Figure 3A-B**).

Finally, and in contrast to Cluster A and B, Cluster C CpG methylation was more heterogeneous in normal T cells and both COSMe T-ALL subtypes (**Figure 1A**).

To further validate these findings, we subsequently profiled an independent cohort of 14 T-ALL patients and 2 thymocyte subsets ($CD34^+$ and $CD4^+CD8^+$) by EPIC sequencing (EPICseq), an alternative, sequencing-based DNA methylation profiling method (**Supplementary Figure 4A-C**). Clustering using the COSMe signature (**Supplementary Table 3**) confirmed the presence of COSMe-I and COSMe-II T-ALLs, with *TAL1*-rearranged T-ALLs being exclusively present in the COSMe-I subtype (**Supplementary Figure 4A**). Furthermore, in this series, COSMe-II T-ALLs consisted of immature ETP-ALL as well as *TLX1*$^+$, *TLX3*$^+$, *NKX2.1*$^+$ or *HOXA*$^+$ leukemias. In addition, within COSMe-II T-ALLs, the lowest levels of Cluster B methylation were also present in immature ETP-ALLs and a *HOXA*$^+$ T-ALL (**Supplementary Figure 4A**).

Altogether, we here show robust COSMe classification as a more elaborate version of the previously established CIMP classification in T-ALL, using different DNA profiling methods and two independent leukemia patient cohorts.

Characterization of CpG clusters that define COSMe in T-ALL

To understand which genes or regulatory pathways are potentially affected by COSMe methylation in T-ALL, we performed enrichment(14) and gene expression analysis(15,16) of transcripts associated with the CpG clusters mentioned above. Of note, Cluster A sites were mostly located in gene promoters, while Cluster B and C CpG sites were mostly located in gene bodies or intergenic regions (**Figure 1D**). Cluster A CpG sites were enriched for promoters of PRC2 targets (**Supplementary Figure 5**) which uniformly showed very low expression in T-ALLs as well as normal thymocytes (**Figure 1E**). Using publicly available ChIP sequencing data, we confirmed the presence of the repressive H3K27me3 histone mark at these Cluster A CpG sites in both human $CD34^+$ hematopoietic stem/precursor cells (HSPCs) and the T-ALL cell line JURKAT(17) (**Figure 1F**).

Notably, Cluster B CpG sites did not show enrichment for repressive histone modifications (**Supplementary Figure 5)** and were associated with genes higher expressed in T-ALL compared to normal developing T cells (**Figure 1E**). Cluster B sites showed significant enrichment for PU.1 (*SPI1*) binding motifs (**Supplementary Figure 5**) which corresponded to specific binding of PU.1(18,19) at these exact loci in $CD34^+$ HSPCs and the Flag-tag PU-1 transfected T-ALL cell line TALL-1(20) (**Figure 1F**). Of note, during normal T cell differentiation, we observed a significant inverse correlation between Cluster B PU.1 binding site methylation and *SPI1* (which encodes PU.1) expression (**Supplementary Figure 6**).

Cluster C CpG sites, characterized by a heterogeneous pattern of DNA methylation, also showed higher expression in T-ALL as compared to normal T cells (**Figure 1E**). Cluster C sites showed significant enrichment for genes specifically expressed in T-ALL cell lines, which was not the case for Cluster A or Cluster B associated transcripts (**Supplementary Figure 7A-C**). More specifically, Cluster C sites include CpGs associated with genes known to be involved in T-ALL disease and/or normal T

cell differentiation, such as *TLX3*, *LCK*, *CDKN2B*, *BCL2L11*, *MEF2C*, *BCL11B*, *RAG1*, *RAG2*, *CD1*, *CD28* and the *TCR* loci (**Supplementary Table 3**). Thus, Cluster C CpG sites are enriched near genes with known roles in normal and malignant T cell development.

Finally, to evaluate the progression of COSMe methylation from diagnosis to relapse, we investigated paired primary and relapsed T-ALL cases that were previously profiled by EPIC arrays(21). Out of 9 patients analyzed, we found that 6 patients were classified as COSMe-I and 3 patients as COSMe-II at diagnosis (**Figure 1G**). Patients at relapse still clustered together with the corresponding primary samples, suggesting that COSMe status is largely conserved from diagnosis to relapse (**Figure 1G**). Nevertheless, COSMe-I patients did show a significant increase in Cluster A methylation from diagnosis to relapse (**Supplementary Figure 8**).

Reciprocal DNA methylation and H3K27me3 association in COSMe-I and COSMe-II T-ALLs

As described above, COSMe Cluster A CpG sites are located in the promoters of PRC2 target genes which uniformly show low expression across all genetic subtypes of human T-ALL. PRC2 is a methyltransferase that primarily produces H3K27me3, a mark of transcriptionally silent chromatin. To validate the enrichment for PRC2 targets at Cluster A CpGs, we profiled H3K27me3 in 3 COSMe-I and 3 COSMe-II human T-ALLs using ChIPmentation(22).

Remarkably, at Cluster A sites, H3K27me3 was found more abundant in COSMe-I T-ALLs, which have lower levels of DNA methylation at these sites, as compared to COSMe-II T-ALLs (**Figure 2A**). To further study this apparent anti-correlation between DNA methylation and H3K27me3, a differential analysis was conducted comparing H3K27me3 of COSMe-II with COSMe-I T-ALLs (**Figure 2B**). This revealed that the majority of differential regions (88%) had significantly lower levels of H3K27me3 in COSMe-II compared to COSMe-I T-ALLs (**Figure 2B**)**,** as also shown in a genome browser view for the genes ESRRG and DKK2 as representative examples (**Supplementary Figure 9**). Genes with differential H3K27me3 were generally low but not differentially expressed between different T-ALL subtypes

(**Supplementary Figure 10**). 44% of all genes with differential H3K27Me3 in this comparison (**Figure 2B**) were also present in Cluster A, underlining the reciprocal association between DNA methylation and H3K27me3 at these sites. High levels of Cluster A H3K27me3 also corresponded to low levels of Cluster A DNA methylation in four patients for which paired EPICseq and H3K27me3 profiles were available (**Figure 2C**). In line with this, core components of the PRC2 complex showed higher expression in TAL-R T-ALLs (EZH2, EED, p adj <0.05, SUZ12 ns, **Supplementary Figure 11**), which are most often COSMe-I T-ALLs displaying low levels of Cluster A methylation. In contrast, deletions of PRC2 members were significantly more prevalent in COSMe-II T-ALLs, whereas amplifications were exclusively found in COSMe-I T-ALLs (**Supplementary Figure 2**).

Finally, from all differential H3K27me3 regions, those that overlapped with CpG islands did show a differential DNA methylation pattern (111 regions, 50%), whereas the methylation of Open Sea sites was not affected by the presence of differential H3K27me3 (**Figure 2D**).

Altogether, these data collectively show a reciprocal association between DNA methylation and H3K27me3 levels at Cluster A CpG islands in human T-ALL.

Cluster A CpG island methylation defines the proliferative history of human T-ALL

Cluster A CpG sites are mainly located in promoters of PRC2 target genes that are differentially covered by H3K27me3 between COSMe-I and II T-ALL subtypes. Interestingly, DNA hypermethylation at PRC2-enriched CpG sites has previously been associated with increased age and proliferative history in both hematopoietic stem cells and T-ALL(11,23). Given this, we predicted the mitotic age of our T-ALL patient cohort using the Epigenetic Timer of Cancer (Epitoc)(24) and identified a strong and significant correlation with the level of Cluster A methylation, which was not observed for Cluster B or Cluster C (**Figure 3A**).

Furthermore, using the Horvath(25) age predictor, which employs methylation of a set of CpGs to predict a person's actual age in years, an increase in epigenetic age of primary T-ALL samples was observed in comparison to the actual age at

diagnosis (**Figure 3B, Supplementary Table 7**). A significant interaction between the patient's age and the Horvath age was indeed observed (t-test, p < 2e-16). However, although the age at diagnosis was already significantly different between COSMe-I and II T-ALL patients (**Figure 3C**, p=0.02506), the predicted Horvath and Epitoc age showed a much better segregation between the two T-ALL subtypes (**Figure 3B-C**, p<2.2e-16 and 3.85e-16). Thus, Cluster A methylation defines the proliferative history of T-ALL, with COSMe Type I having a shorter history of proliferation in comparison to COSMe Type II T-ALL.

Proliferation during the timeframe from primary diagnosis to relapsed T-ALL disease should also lead to an increase in mitotic age. Indeed, a significant increase in Epitoc age was also observed by comparing 9 paired diagnosis and relapse samples(21) **(Figure 3D)**. However, the increase in age and methylation was considerably more pronounced in COSMe-I T-ALLs (Δ^{Epitoc} 0.11, p= 0.0020) as compared to COSMe-II T-ALLs (Δ^{Epitoc} 0.015, p=0.0522), yet there was no significant difference between the time of relapse between both entities (**Supplementary Table 8**, p= 0.2619). From this limited analysis(21), COSMe-I T-ALLs seem to proliferate at a faster pace than COSMe-II T-ALLs between diagnosis and relapse.

Proliferation of preleukemic thymocytes drives the aging CpG island methylation signature in T-ALL

As shown above, Cluster A CpG island methylation positively correlates with the proliferative history and mitotic age of human T-ALL cells. To investigate if two distinct trajectories towards T-ALL development might underlie the observed discrepancy in epigenetic age between COSMe-I and COSMe-II T-ALLs (**Figure 3B-C**), we used two known T-ALL mouse models that might recapitulate these features.

First, we investigated if *CD2-Lmo2* transgenic (*CD2-Lmo2*^{tg}) mice(26) could function as a model for COSMe-II T-ALLs. This *in vivo* T-ALL mouse model has a long disease latency and an immature T-ALL phenotype reminiscent of T-ALLs inside the COSMe-II subgroup. In *CD2-Lmo2*^{tg} mice, a long-term self-renewing thymocyte population has been observed many months before tumour development. In contrast to wildtype control mice, in which the thymus is continuously replenished by

progenitor cells from the bone marrow, these pre-leukemic *CD2-Lmo2*^{tg} thymocytes are self-sustaining from young age. Therefore, *CD2-Lmo2*^{tg} thymocytes should undergo a gradual aging process in the months prior to malignant transformation.

Secondly, we investigated if *Lck-Cre*$^{tg/+}$*Pten*$^{fl/fl}$ mice could function as a model for COSMe-I T-ALL. This mouse T-ALL model has a short disease latency and results in the development of more mature murine T-cell leukemias. In addition, *PTEN* mutations and deletions are also more often observed in human COSMe-I T-ALLs.

To study this, we isolated full thymus from *CD2-Lmo2*^{tg} and *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice and littermate controls at different timepoints before and after leukemia development and performed DNA methylation profiling by Reduced Representation Bisulfite Sequencing (RRBS, n=4 biological replicates per condition). The most variably methylated CpGs between *CD2-Lmo2*^{tg} and *LCK-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ thymocytes and blasts (**Supplementary Table 9**) were more often situated in CpG islands (**Figure 4A**, Cluster 1) than in Open Sea enriched regions (**Figure 4A**, Cluster 2 and 3). The CpG methylation in Cluster 1 was highly increased in pre-leukemic and leukemic *CD2-Lmo2*^{tg} mice, but not in *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice isolated before leukemia development. In leukemic *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice, Cluster 1 methylation was only moderately increased compared to *CD2-Lmo2*^{tg} samples at the same stage of disease manifestation.

Notably, Cluster 1 CpGs were mostly situated in promoter regions (**Figure 4B**) of lowly expressed genes (**Figure 4C**) and displayed enrichment for PRC2 target genes and H3K27me3(27,28) at these sites (**Supplementary Figure 12**, **Figure 4D**), thus showing similarity with the hypermethylation phenotype of T-ALLs in the human Cluster A. In contrast, we did not find any evidence in these mouse models for potential overlap between murine Cluster 2 or Cluster 3 and the human COSMe Cluster B.

To further confirm similarities between murine Cluster 1 and human Cluster A, we subsequently looked at cross-species overlap at the gene level. Notably, 726 out of 2248 CpG sites in mouse Cluster 1 overlapped with human gene orthologs in Cluster A (overlap significant at $p < 0.0001$, exact hypergeometric probability). These sites,

which we termed Cluster A^{mm} sites (**Supplementary Table 10**), were equally significantly enriched for PRC2 target genes (p adj <0.0001).

Cluster A^{mm} regions showed a gradual increase in methylation with aging in *CD2-Lmo2*^{tg} thymocytes but remained constant in corresponding wild type mice from 8, 16 and 24 weeks old (**Figure 4E**). Already in 8-week-old *CD2-Lmo2*^{tg} thymocytes, a strong increase in methylation could be observed compared to WT control cells of the same age. In contrast, in *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice at 8 weeks, no difference in Cluster A^{mm} methylation was detected. We found that Cluster A^{mm} methylation did not further increase in fully transformed *CD2-Lmo2*^{tg} leukemia, sacrificed on average at 35.75 ± 8.28 weeks, compared to pre-leukemic *CD2-Lmo2*^{tg} at 24 weeks (**Figure 4E**), whereas only a moderate increase in Cluster A^{mm} methylation was detected in leukemic *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice (sacrificed at on average 18 ± 3.83 weeks). Finally, the epigenetic age, calculated using the mouse specific calculation method of Petkovich et al.(29), increased over time in preleukemic thymocytes of the *CD2-Lmo2*^{tg} mice, which was not the case in the *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ model (**Figure 4F**).

Thus, pre-leukemic, self-renewing thymocytes display a murine CpG island DNA hypermethylation signature that recapitulates features of the CpG island hypermethylation phenotype that we observed in human COSMe-II T-ALL. Therefore, *CD2-Lmo2*^{tg} might serve as a bona fide model for COSMe-II T-ALL development. In contrast, this CpG island hypermethylation phenotype was not observed in preleukemic thymocytes of *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice, suggesting that this murine model is more similar to COSMe-I human T-ALL. Furthermore, based on these results, COSMe-I and COSMe-II T-ALLs might have followed a different trajectory towards leukemia, marked by the absence or presence of a thymic self-renewing population that preceded leukemia development.

Age-related CpG island hypermethylation is resistant to the FDA-approved hypomethylating agent Decitabine

DNA hypomethylating agents such as Azacitidine and Decitabine are approved for myelodysplastic syndrome and acute myeloid leukemia(30). Interestingly, few publications have also investigated the possible use of DNA hypomethylating agents

for the treatment of human T-ALL. One T-ALL patient responded to Decitabine and achieved complete response in a phase 1 clinical trial(31). In addition, durable remissions have been reported for few other cases of (early) T-cell precursor ALL treated with Decitabine as mono-therapy(32-34) or in combination with the BCL-2 inhibitor Venetoclax(35,36). The largely perturbed DNA methylation profile observed in T-ALL further supports the rationale of using DNA hypomethylating agents for the treatment of this disease. However, the actual mechanism of action that could explain the anti-leukemic properties of these hypomethylating agents and their putative effect on aberrant DNA methylation in T-ALL, as exemplified by the COSMe phenotype, has remained largely unclear.

To address these questions, we first evaluated the anti-leukemic properties of Decitabine in a pre-clinical setting using four different primary human T-ALL Patient Derived Xenograft (PDX) models with variable genetic backgrounds (**Supplementary Table 11A-H**). Notably, 10 days of Decitabine (5 days on, 2 days off, 0.5 mg/kg) significantly improved leukemia-free survival for all PDX samples analyzed, including a sample that originated from a STIL-TAL1^{+} mature T-ALL at second relapse (**Figure 5A**). As expected, this improved survival coincided with a significant decrease in blast percentage in the peripheral blood, as shown for one PDX T-ALL (**Supplementary Figure 13**).

We selected two PDX T-ALLs for DNA methylation profiling representative for a COSMe-I and a COSMe-II T-ALL based on the genetics (PDX#1 and PDX#2). For this, seven days after start of the treatment, leukemic blasts from control and Decitabine treated animals (n=3 each group) were collected from the spleen. We performed differential methylation analysis comparing Decitabine treatment with vehicle control for each PDX *in vivo*. As expected, a very large number of regions showed a significant decrease in DNA methylation, but also a substantial portion of CpGs displayed an increase in methylation, especially those located in CpG islands (**Figure 5B**, DESeq2 p adj <0.05,). Despite these global changes in DNA methylation, Decitabine did not drastically affect DNA methylation at the 5000 human COSMe CpGs, neither in PDX#1, which we determined to be a COSMe-Type I Cluster B^{+} T-ALL, nor in the COSMe Type II Cluster B^{+} PDX#2 (**Figure 5C**).

To obtain additional insights in the mechanism of action of hypomethylating agents in T-ALL, we used the same samples (PDX T-ALL#1 and #2, 5 days, 0.5 mg/kg, n=3 each group) for RNAseq. Decitabine treatment resulted in significant differential expression of 456 (PDX T-ALL#1) and 886 (PDX T-ALL#2) protein-coding genes (Deseq2, adjP<0.05, **Supplementary Table 12A-B**). Of these differentially expressed genes, 361 (PDX#1) and 348 (PDX#2) genes also showed a significant decrease in methylation at one or more CpGs (**Supplementary Table 13A-B**). However, the transcriptome is less disturbed than one would expect as Decitabine induced profound genome-wide DNA hypomethylation which globally followed CpG density (**Supplementary Figure 14A-B**). The genes down-regulated by Decitabine more frequently had CpG islands in their promoters (57%) compared to up-regulated or randomly selected genes (both 23%, Chi Square test, $p < 0.00001$).

As Decitabine did not revert the COSMe methylation signature, we further looked at the gene expression changes induced by this hypomethylating agent to explain its anti-leukemic properties. Both PDX samples showed significant upregulation of tumor suppressor genes (TSGs) upon Decitabine treatment, including *BCL2L11* (BIM), *BBC3* (PUMA) and *BMF* (**Supplementary Table 12A-B)**. In addition, multiple known T-ALL oncogenes were significantly downregulated by *in vivo* Decitabine treatment, including *MYC*, *HES1* and *BCL2* in PDX#1 and *NOTCH3*, *GFI1*, *IL7R*, *CISH*, *MYB* and *TOX* in PDX#2 (**Figure 5D**). Enrichment analysis revealed that Decitabine induced a global downregulation of MYC target genes in both PDX samples, as exemplified by pre-ranked Gene Set Enrichment Analysis (GSEA)(37) (**Figure 5E**, **Supplementary Table 14**). To further validate these findings, we performed additional RNAseq on 5 human T-ALL cell lines (LOUCY, PER117, PEER, MOLT16 and TALL-1) following *in vitro* Decitabine treatment (1μM, 48h). The differentially expressed genes from both PDX samples significantly overlapped with the differential data from the cell lines (33% in PDX#1 and 35% in PDX#2). Indeed, also in these cell lines, significant upregulation of TSGs like BMF and BBC3 (PUMA) and coordinate downregulation of the MYC pathway upon treatment with Decitabine was observed (**Supplementary Figure 15A-B, Supplementary Table 15A-F**).

Several studies showed that DNA methylation can alter the binding of CTCF to influence the 3D architecture of the genome(38-40). Therefore, we compared the genes downregulated by Decitabine in both PDX and T-ALL cell lines with gene

expression in CTCF depleted acute leukemia(41). Interestingly, we observed a very significant enrichment for downregulated genes in both datasets by GSEA (**Supplementary Figure 16**), providing a potential mechanism for Decitabine-induced downregulation of the MYC pathway in human T-ALL.

Discussion

In this study, we directly compared the genome-wide DNA methylation landscape in human T-ALL with human thymocytes covering the complete trajectory of normal T-cell development. This simultaneous DNA methylome analysis of T-cell leukemias with their putative cell-of-origin empowered us to distinguish cell-of-origin methylation profiles from leukemia-specific alterations at CpG islands and Open Sea sites. Integration of DNA methylation profiles with gene expression signatures, and ChIPseq allowed us to obtain novel insights in the functional relevance of differentially methylated CpG sites between normal and malignant T-cells, and to distinguish two trajectories towards T-ALL development.

We built a DNA methylation-based signature for T-ALL, termed COSMe. COSMe comprises three clusters that can divide T-ALLs in two subgroups, COSMe Type I and COSMe Type II which both have unique characteristics as summarized in **Table 1**. First, COSMe Cluster A sites were closely related to the CpG sites that have previously been used to determine CIMP status in T-ALL(42) and showed low expression in both normal developing thymocytes and human T-ALLs. An inverse correlation between Cluster A DNA methylation and H3K27me3 was observed. This interplay could explain the low expression of Cluster A genes across all genetic subtypes of human T-ALLs as repression is maintained by different mechanisms in COSMe-I (PRC2 mediated repression) and COSMe-II T-ALLs (DNA methylation mediated repression). This phenomenon has previously been observed in murine *Ezh2* knockout ETP-ALL(43) that displayed increased DNA methylation. In ETP-ALL, EZH2 inactivating events were previously linked to oncogenic active stem- and progenitor cell genes(44). Also, in our cohort, inactivating alterations targeting PRC2 members were more prevalent in COSMe-II T-ALL. However, the relation between PRC2 and DNA methylation is complex, as PRC2 has been shown to recruit both DNMT(45) and TET enzymes(46). Therefore, PRC2 might potentially affect both methylation as well as demethylation depending on the cellular context. However,

the mechanism and functional relevance of these different levels of H3K27me3 between human COSMe-I and COSMe-II T-ALLs remains to be established. In addition, additional studies will be required to further unravel how these epigenetic alterations exactly create a permissive landscape for T-ALL transformation or if they are just the result of specific cell-intrinsic anti-tumor mechanisms.

Next, our study linked DNA hypermethylation in COSMe Cluster A to the proliferative history of the cancer cells, thereby confirming (11) that COSMe-II T-ALLs are epigenetically older and displayed a longer mitotic history as compared to COSMe-I T-ALL. In line with this, we showed that relapsed COSMe-I T-ALLs displayed a larger increase in epigenetic age during their progression from diagnosis to relapsed disease. The younger mitotic age in combination with a faster rate of cellular proliferation might potentially contribute to the more aggressive nature of COSMe-I T-ALLs as compared to COSMe-II leukemias.

Furthermore, we could recapitulate these methylation features distinguishing fast and aggressive COSMe-I from the slower transforming COSMe-II T-ALLs, in two distinct T-ALL mouse models: the spontaneous CD2-Lmo2^{tg} T-ALL mouse model, which mimics immature T-ALL development and has a long disease latency, and the fast transforming *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mouse model, as a model of more mature human T-ALL, where *PTEN* deletions are more abundant. Notably, these findings establish *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ and *CD2-Lmo2*^{tg} as models to further study COSMe-I and COSMe-II T-ALL Cluster A methylation *in vivo*, but also provide evidence that a pre-leukemic self-renewing thymocyte population might also exist in human COSMe-II T-ALL. Indeed, in *CD2-Lmo2*^{tg} mice, but not in *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice, a pre-leukemic state has been described, in which thymic precursors gain self-renewing potential prior to full malignant T cell transformation. The pre-leukemic cells gain DNA methylation at these sites before transformation from 8 to 24 weeks, which was not observed in the corresponding aging thymocytes of *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice or littermate controls. The methylation at these sites did not further increase after full transformation of the pre-leukemic cells, suggesting that aberrant methylation at these sites is mainly derived from the pre-leukemic history of these tumor cells.

Besides Cluster A CpG island methylation as a surrogate marker for the epigenetic age and replicative history of tumor cells, Cluster B sites enabled additional

classification of COSMe Type-II T-ALL into two categories, based on the methylation level of specific Open Sea enriched CpG sites. Interestingly, these Cluster B loci displayed significant enrichment for the binding motif of PU.1, a transcription factor critically involved in early T-cell development(47).

Of note, we could show that these Cluster B sites are able to distinguish immature from more mature T-ALLs within the COSMe-II leukemias. Furthermore, immunophenotypically validated ETP-ALLs, which are derived from the most immature T cell precursors, also showed the lowest levels of Cluster B methylation in the validation cohort. Whereas conclusive immunophenotypic data were missing to accurately define ETP-ALs in the initial T-ALL cohort, these data do collectively suggest that both immature T-ALL and ETP-ALL could be distinguished from other T-ALLs based on the absence of Cluster B methylation within COSMe-II leukemias.

Notably, a similar correlation between PU.1 binding site methylation and *SPI1* expression was recently also identified in TCF7-SPI1 fusion positive T-ALLs(4). However, the lack of Cluster B methylation in this very aggressive subtype of PU.1 rearranged human T-ALL(4) is most probably caused by the aberrant *SPI1* expression downstream of the fusion proto-oncogene rather than being associated with its cell of origin.

Finally, Cluster C CpG sites were also enriched for Open Sea sites but showed more heterogeneity in DNA methylation between T-ALLs. Cluster C associated transcripts showed generally lower expression in normal T-cell subsets but were clearly active in at least some T-ALL tumors, and included genes involved in T-ALL disease biology (48) as well as normal T-cell differentiation. Thus, the methylation pattern at some of these Cluster C Open Sea CpG sites might be T-ALL specific and potentially hold some information on the genetic abnormalities present in each individual T-ALL.

In the last part of our work, we looked for further evidence for DNA hypomethylating agents as a promising therapeutic strategy for human T-ALL. Indeed, some case reports have previously shown that the FDA approved drug Decitabine might be effective as salvage therapy for T-ALL(31) and ETP-ALL(32-34). In line with this, Decitabine here displayed profound anti-leukemic properties in PDXs from a variety of different human T-ALLs obtained from both primary as well as relapse tumor

material. However, and most unexpectedly, these anti-tumoral effects, which were shown to be mediated by downregulating of oncogenic MYC signaling, were both observed in COSMe-I and COSMe-II samples. Indeed, Decitabine did not revert the age-related human CpG island hypermethylation phenotype *in vivo*. Instead, this hypomethylating agent triggered a profound and genome-wide hypomethylation effect on CpGs located in open sea areas, and surprisingly also hypermethylation in CpG islands. The widely altered DNA methylation profile has only limited effects on the transcriptome of these cells. Decitabine down-regulated genes are reconciled by CTCF depletion, thus DNA hypomethylating agents might alter the binding capacity of CTCF as a potential mechanism to induce changes at the transcriptomic level. Several studies also show a direct link between MYC and DNMTs(49). MYC inactivation has been shown to reduce the expression of DNMT3B in T-ALL(50), and reciprocally, reduced expression of DNMT3B resulted in reduced proliferation and tumor maintenance, reflecting the effects observed in our Decitabine treated PDX T-ALLs. Thus, down-regulation of MYC or inhibition of DNMTs by Decitabine might converge on counteracting T-ALL disease burden.

Altogether, our work identifies aging of preleukemic thymocytes as a driver of the CpG island methylator phenotype in human T-ALL, revealing different trajectories towards T-cell transformation. Our work provides evidence for the involvement of pre-leukemic thymocytes in the pathogenesis of human T-AL, which has previously only been reported in mouse T-ALL models so far. Additionally, we provide a biological explanation for the profound differences in epigenetic age between COSMe-I and COSMe-II T-ALLs and show that the FDA approved hypomethylation drug Decitabine shows a promising increase in survival of both epigenetically young and old T-ALLs but fails to revert age-related CpG island hypermethylation in human T-ALL xenografts. This very extensive DNA methylome dataset of murine and human T-ALL will be of importance to further increase our understanding of T-ALL disease biology, which could ultimately result in better treatment stratification and the development of novel and less toxic therapeutic strategies for the treatment of this aggressive hematological malignancy.

Acknowledgements

This work was supported by the following funding agencies: the European Research Council (StG-639784 to PVV), Fund for Scientific Research Flanders, Kom op tegen Kanker (Stand up to Cancer; the Flemish cancer society), Stichting Tegen Kanker (STK), Kinderkankerfonds (a non-profit childhood cancer foundation under Belgian law), Cancer Research Institute Ghent (CRIG), the National Science Center Poland (2017/24/T/NZ5/00359), the National Centre for Research and Development Poland (STRATEGMED3/304586/5/NCBR/2017), the Swedish Childhood Foundation (PR2018-0064), the Medical Faculty of Umeå University, the Kempe Foundation and the Lion's Cancer Research Foundation.

Rishi S Kotecha (NHMRC APP1142627) is supported by a Fellowship from the National Health and Medical Research Council of Australia. Marc R Mansour is a Bloodwise Bennett Fellow. The Davé Lab was supported by R01CA207530 from the National Cancer Institute (USA), I01BX001799 from the Department of Veterans Affairs (USA) and the IUSM Strategic Research Initiative. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI.

Authorship Contributions

MT, JR, SDC, BS, ML, LD, LP, AK, SS, TP, FM, KD, UPD and BG performed experiments. JR, ML and WVL performed analyses. LR, DD, GVC, WVB and FVN provided technical assistance. BE, PR, TL, LC, RSK, MM, JRK, TS, AEK and BDM collected and provided primary T-ALL patient material. PVV, SG, TT, MD, MT, DJC, JR and SD designed research and wrote the paper, with help from the other authors. All authors have seen, reviewed and approved the final version.

Materials and methods

Patient samples, normal T cells and cell lines

DNA from 109 T-ALLs was collected from the previously characterized(8,51) ALL IC-BFM 2002/2009 protocol. Exome sequencing was performed by Novogene. Clinical characteristics of patients treated according to both treatment regimens were not significantly different (Chi2test).

Thymocytes were isolated from postnatal thymus suspension of two donors each, as previously described(9,16).

Cell lines were purchased from DSMZ (Braunschweig Germany) and cultured in RPMI-1640 medium (Life Technologies, Carlsbad, CA, USA) supplemented with 10 or 20% fetal calf serum (FCS), 100 U/ml penicillin, 100 µg/ml streptomycin (Life Technologies), and 2 mM L-glutamine (Life Technologies) at 37°C with 5% CO2. CUTTL-1 and PER-117 were a kind gift from Adolfo Ferrando (Columbia University, New York, NY, USA) and Rishi Kotecha (Telethon Kids Cancer Center, Perth, Australia) respectively. Cell lines were monthly screened for mycoplasma contamination and were consistently negative.

All human samples were acquired with written informed consent according to the declaration of Helsinki and the studies were approved by the ethical committee review board of the Department of Pediatric Hemato-Oncology at Ghent Univerisity Hospital.

DNA methylation profiling

Human DNA methylation analysis of 109 T-ALLs was done with the Infinium HumanMethylationEPIC BeadChip array (Illumina). 250ng of DNA was used for bisulfite conversion by the EZ DNA methylation Kit (Zymo Research).

DNA methylation profiling of the 14 human T-ALL validation cohort was done by TruSeq Methyl Capture EPICseq using 500ng of DNA. Sequencing was done on the Hiseq3000 (PE150). Reads were trimmed using TrimGalore (v0.4.5) and aligned with Bismark(52) (v0.20.0) on hg38.

RRBS was done on thymus isolated from *CD2-Lmo2*^{tg} and *Lck-Cre*$^{tg/+}$ *Pten*$^{fl/fl}$ mice and littermate controls (n=4 mice per condition). DNA was isolated from full thymus using the QiaAMP DNA Mini kit (Qiagen). RRBS was done using MSPI digest. Bisulfite converted DNA libraries were sequenced on Illumina NextSeq500 using the NextSeq 500/550 High Output v2 kit (SE75). Reads were trimmed using TrimGalore (v0.4.5) with –rrbs and –non-directional and aligned with Bismark(52) (v0.20.0) on GRCm38.

Illumina EPIC Array and RRBS data was analyzed using RnBeads.hg19 or RnBeads.mm10 respectively(53) in R (versions >3.4). β values were obtained after filtering and normalization using the default preprocessing pipeline. EPICseq was analyzed using the package “methylkit” in R. Annotation of CpGs to the closest gene was done by ChipPeakAnno(54). Copy number variations were called from EPICarray data using the R packages “minfi” and “conumee”.

H3K27me3 Profiling

50-100 000 CD45+ sorted cells were fixed using 1% formaldehyde (FA) (Thermo Scientific 28906) and quenched by Glycin (125 mM final). Next, cell pellets were lysed in 100 µL of short-term complete lysis buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 0.25% SDS, 20mM NaBu HDAC inhibitor, 1X complete protease inhibitors cocktail EDTA-free (Roche, 5056489001)). Chromatin was sheared on the Bioruptor Pico (Diagenode) using a 15s ON / 30s OFF, 7 cycle regimen. Sheared chromatin was magnetically immunoprecipitated and tagmentated using the Auto ChIPmentation Kit for Histones (Diagenode, C01011010) on the IP-Star® Compact Automated System (Diagenode, B03000002), according to the manufacturer’s instructions. Input DNA was de-crosslinked, purified (MinElute, Qiagen) and tagmentated (Nextera DNA library prep, Illumina). Stripping, end repair and library amplification were performed according to the ChIPmentation kit guidelines. Libraries were sequenced with the Nextseq500 (SR75, High Output). Reads were trimmed by Trimmomatic and aligned to hg38 with Bowtie2 using the parameters -N 1 -k 1. Peaks were called with MACS2 with the respective input control for each patient sample.

Gene expression profiling

Total RNA was isolated using the miRNeasy mini kit (Qiagen) and evaluated on the Agilent 2100 bioanalyzer (Agilent Technologies). Library preparation was performed using QuantSeq 3’mRNA-Seq FWD for Illumina (LEXOGEN). cDNA libraries were sequenced as described above. Reads were aligned to GRCh38 using STAR2.4.2a(55) and quantified on Gencode v24.

For visualization of RNAseq data, EdgeR $\log_2$-transformed normalized counts per million were plotted unless mentioned otherwise.

Micro-array data normalized with Limma and was log_2 transformed.

In vivo treatment of xenografts

Patient derived xenografts were established in female nonobese diabetic/severe combined immunodeficient γ (NSG) mice. For the initial PDX experiments, upon disease establishment, human leukemic cells were isolated from the spleen. Secondary injections were performed in randomized NSG mice (2 groups of 5) and treated 2 cycles (5 days on, 2 days off) with vehicle-only or Decitabine (0.5 mg/kg bodyweight).

For RNA and DNA collection, mice were randomized in 2 groups of 3 and treated for 5 days with vehicle-only or Decitabine (0.5 mg/kg bodyweight). At day 7, animals were sacrificed, and tumor cells collected from the spleen.

The animal welfare ethical committee (Ghent University Hospital) approved all animal experiments.

CIMP classification

1099 CIMP CpGs were defined by filtering(3) CIMP probes to exclude CpGs within 5 base pairs from European SNPs, cross hybridizing probes, repeated regions(56) and methylation quantitative trait locus(57). Missing values were imputed using K nearest neighbor.

COSMe classification

5000 human COSMe CpGs were defined by most variably methylated CpGs in all normal thymocyte subsets and 109 T-ALLs. The classification of diagnosis and relapsed T-ALLs was obtained by hierarchical clustering of these 5000 CpGs with Euclidean distance measures and clustering method “ward.D”, using row-scaling.

Epigenetic age calculations

The Horvath(25) and EpiTOC(24) age were calculated with ‘cgageR’. Epigenetic age in mouse samples was estimated by taking the weighted average of CpGs within the 90 CpG age classifier defined by Petkovich et al.(29). Only CpGs that had enough coverage (>5) after data imputation with BoostMe(58) were retained. The weighted average was scaled and used as a relative measure of epigenetic age.

Data analysis and statistics

R or GraphPad Prism 6.0 was used for statistical analyses.

When applicable, normality was tested using a Shapiro-Wilk test.

Publicly available ChIP-seq data was retrieved from the ChIP-Atlas(59). Heatmaps of ChIP-seq data were generated by deepTools(60) computeMatrix and plotHeatmap with the options reference-point and missingDataAsZero.

Enrichment analyses were performed with GSEA(61,62), DAVID(63,64) and Enrichr(14,65). Hierarchical clustering and heatmaps were generated with R 'pheatmap' using row-scaling. Differential expression was identified by DeSEQ2(66).

Data availability

All generated data was deposited in NCBI Gene Expression Omnibus (GEO) under accession number GSE155339.

References

1. Michalak EM, Burr ML, Bannister AJ, Dawson MA. The roles of DNA, RNA and histone methylation in ageing and cancer. Nature reviews Molecular cell biology **2019**;20(10):573-89 doi 10.1038/s41580-019-0143-1.
2. Issa JP. CpG island methylator phenotype in cancer. Nature reviews Cancer **2004**;4(12):988-93 doi 10.1038/nrc1507.
3. Borssen M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J*, et al.* Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. PLoS One **2013**;8(6):e65373 doi 10.1371/journal.pone.0065373.
4. Kimura S, Seki M, Kawai T, Goto H, Yoshida K, Isobe T*, et al.* DNA methylation-based classification reveals difference between pediatric T-cell acute lymphoblastic leukemia and normal thymocytes. Leukemia **2019**;34(4):1163-8 doi 10.1038/s41375-019-0626-2.
5. Touzart A, Boissel N, Belhocine M, Smith C, Graux C, Latiri M*, et al.* Low level CpG island promoter methylation predicts a poor outcome in adult T-cell acute lymphoblastic leukemia. Haematologica **2019**;105(6):1575-81 doi 10.3324/haematol.2019.223677.
6. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC*, et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer cell **2002**;1(1):75-87.
7. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M*, et al.* Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. Cancer Cell **2011**;19(4):484-97 doi 10.1016/j.ccr.2011.02.008.
8. Szarzynska-Zawadzka B, Kunz JB, Sedek L, Kosmalska M, Zdon K, Biecek P*, et al.* PTEN abnormalities predict poor outcome in children with T-cell acute lymphoblastic leukemia treated according to ALL IC-BFM protocols. American journal of hematology **2019**;94(4):E93-E6 doi 10.1002/ajh.25396.
9. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. Curr Top Microbiol Immunol **2012**;360:75-97 doi 10.1007/82_2012_230.
10. Dolens AC, Durinck K, Lavaert M, Van der Meulen J, Velghe I, De Medts J*, et al.* Distinct Notch1 and BCL11B requirements mediate human gammadelta/alphabeta T cell development. EMBO Rep **2020**;21(5):e49006 doi 10.15252/embr.201949006.
11. Haider Z, Larsson P, Landfors M, Kohn L, Schmiegelow K, Flaegstad T*, et al.* An integrated transcriptome analysis in T-cell acute lymphoblastic leukemia links DNA methylation subgroups to dysregulated TAL1 and ANTP homeobox gene expression. Cancer Med **2019**;8(1):311-24 doi 10.1002/cam4.1917.
12. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D*, et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature **2012**;481(7380):157-63 doi 10.1038/nature10725.
13. Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR*, et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. Nat Genet **2017**;49(8):1211-8 doi 10.1038/ng.3909.
14. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z*, et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res **2016**;44(W1):W90-7 doi 10.1093/nar/gkw377.
15. Verboom K, Van Loocke W, Volders PJ, Decaesteker B, Cobos FA, Bornschein S*, et al.* A comprehensive inventory of TLX1 controlled long non-coding RNAs in T-cell acute lymphoblastic leukemia through polyA+ and total RNA sequencing. Haematologica **2018**;103(12):e585-e9 doi 10.3324/haematol.2018.190587.
16. Roels J, Kuchmiy A, De Decker M, Strubbe S, Lavaert M, Liang KL*, et al.* Distinct and temporary-restricted epigenetic mechanisms regulate human alphabeta and gammadelta T cell development. Nat Immunol **2020** doi 10.1038/s41590-020-0747-9.

17. Manser M, Sater MR, Schmid CD, Noreen F, Murbach M, Kuster N*, et al.* ELF-MF exposure affects the robustness of epigenetic programming during granulopoiesis. Sci Rep **2017**;7:43345 doi 10.1038/srep43345.
18. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O*, et al.* Topologically associating domains are stable units of replication-timing regulation. Nature **2014**;515(7527):402-5 doi 10.1038/nature13986.
19. Huang J, Liu X, Li D, Shao Z, Cao H, Zhang Y*, et al.* Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. Dev Cell **2016**;36(1):9-23 doi 10.1016/j.devcel.2015.12.014.
20. Minderjahn J, Schmidt A, Fuchs A, Schill R, Raithel J, Babina M*, et al.* Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. Nat Commun **2020**;11(1):402 doi 10.1038/s41467-019-13960-2.
21. Richter-Pechanska P, Kunz JB, Bornhauser B, von Knebel Doeberitz C, Rausch T, Erarslan-Uysal B*, et al.* PDX models recapitulate the genetic and epigenetic landscape of pediatric T-cell leukemia. EMBO molecular medicine **2018**;10(12) doi 10.15252/emmm.201809443.
22. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nature methods **2015**;12(10):963-5 doi 10.1038/nmeth.3542.
23. Beerman I, Bock C, Garrison BS, Smith ZD, Gu H, Meissner A*, et al.* Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. Cell Stem Cell **2013**;12(4):413-25 doi 10.1016/j.stem.2013.01.017.
24. Yang Z, Wong A, Kuh D, Paul DS, Rakyan VK, Leslie RD*, et al.* Correlation of an epigenetic mitotic clock with cancer risk. Genome Biol **2016**;17(1):205 doi 10.1186/s13059-016-1064-3.
25. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol **2013**;14(10):R115 doi 10.1186/gb-2013-14-10-r115.
26. Smith S, Tripathi R, Goodings C, Cleveland S, Mathias E, Hardaway JA*, et al.* LIM domain only-2 (LMO2) induces T-cell leukemia by two distinct pathways. PLoS One **2014**;9(1):e85883 doi 10.1371/journal.pone.0085883.
27. Chen S, Wang Q, Yu H, Capitano ML, Vemula S, Nabinger SC*, et al.* Mutant p53 drives clonal hematopoiesis through modulating epigenetic pathway. Nat Commun **2019**;10(1):5649 doi 10.1038/s41467-019-13542-2.
28. Yashiro-Ohtani Y, Wang H, Zang C, Arnett KL, Bailis W, Ho Y*, et al.* Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. Proc Natl Acad Sci U S A **2014**;111(46):E4946-53 doi 10.1073/pnas.1407079111.
29. Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN. Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. Cell metabolism **2017**;25(4):954-60.e6 doi 10.1016/j.cmet.2017.03.016.
30. Duchmann M, Itzykson R. Clinical update on hypomethylating agents. Int J Hematol **2019**;110(2):161-9 doi 10.1007/s12185-019-02651-9.
31. Benton CB, Thomas DA, Yang H, Ravandi F, Rytting M, O'Brien S*, et al.* Safety and clinical activity of 5-aza-2'-deoxycytidine (decitabine) with or without Hyper-CVAD in relapsed/refractory acute lymphocytic leukaemia. Br J Haematol **2014**;167(3):356-65 doi 10.1111/bjh.13050.
32. El Chaer F, Holtzman N, Binder E, Porter NC, Singh ZN, Koka M*, et al.* Durable remission with salvage decitabine and donor lymphocyte infusion (DLI) for relapsed early T-cell precursor ALL. Bone Marrow Transplant **2017**;52(11):1583-4 doi 10.1038/bmt.2017.191.
33. Yang Y, Yao S, Zhang J, Yan Z, Chu J, Wang H*, et al.* Decitabine-containing G-CSF priming regimen overcomes resistance of primary mediastinal neoplasm from early T-cell precursors to conventional chemotherapy: a case report. Onco Targets Ther **2019**;12:7039-44 doi 10.2147/OTT.S214905.

34. Cui JK, Xiao Y, You Y, Shi W, Li Q, Luo Y*, et al.* Decitabine for relapsed acute lymphoblastic leukemia after allogeneic hematopoietic stem cell transplantation. J Huazhong Univ Sci Technolog Med Sci **2017**;37(5):693-8 doi 10.1007/s11596-017-1790-0.
35. Rahmat LT, Nguyen A, Abdulhaq H, Prakash S, Logan AC, Mannis GN. Venetoclax in Combination with Decitabine for Relapsed T-Cell Acute Lymphoblastic Leukemia after Allogeneic Hematopoietic Cell Transplant. Case Rep Hematol **2018**;2018:6092646 doi 10.1155/2018/6092646.
36. Farhadfar N, Li Y, May WS, Adams CB. Venetoclax and decitabine for treatment of relapsed T-cell acute lymphoblastic leukemia: A case report and review of literature. Hematol Oncol Stem Cell Ther **2020** doi 10.1016/j.hemonc.2019.10.002.
37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A **2005**;102(43):15545-50 doi 10.1073/pnas.0506580102.
38. De Decker M, Lavaert M, Roels J, Tilleman L, Vandekerckhove B, Leclercq G*, et al.* HES1 and HES4 have non-redundant roles downstream of Notch during early human T cell development. Haematologica **2020** doi 10.3324/haematol.2019.226126.
39. Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol Cell **2017**;66(5):711-20 e3 doi 10.1016/j.molcel.2017.05.004.
40. Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm JP, Hofer T*, et al.* Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. Genome Res **2014**;24(8):1285-95 doi 10.1101/gr.164418.113.
41. Hyle J, Zhang Y, Wright S, Xu B, Shao Y, Easton J*, et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. Nucleic Acids Res **2019**;47(13):6699-713 doi 10.1093/nar/gkz462.
42. Borssen M, Haider Z, Landfors M, Noren-Nystrom U, Schmiegelow K, Asberg AE*, et al.* DNA Methylation Adds Prognostic Value to Minimal Residual Disease Status in Pediatric T-Cell Acute Lymphoblastic Leukemia. Pediatr Blood Cancer **2016**;63(7):1185-92 doi 10.1002/pbc.25958.
43. Wang C, Oshima M, Sato D, Matsui H, Kubota S, Aoyama K*, et al.* Ezh2 loss propagates hypermethylation at T cell differentiation-regulating genes to promote leukemic transformation. J Clin Invest **2018**;128(9):3872-86 doi 10.1172/jci94645.
44. Danis E, Yamauchi T, Echanique K, Zhang X, Haladyna JN, Riedel SS*, et al.* Ezh2 Controls an Early Hematopoietic Program and Growth and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia. Cell Rep **2016**;14(8):1953-65 doi 10.1016/j.celrep.2016.01.064.
45. Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C*, et al.* The Polycomb group protein EZH2 directly controls DNA methylation. Nature **2006**;439(7078):871-4 doi 10.1038/nature04431.
46. Wu H, Zhang Y. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle **2011**;10(15):2428-36 doi 10.4161/cc.10.15.16930.
47. Zhou W, Yui MA, Williams BA, Yun J, Wold BJ, Cai L*, et al.* Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. Cell Syst **2019**;9(4):321-37 e9 doi 10.1016/j.cels.2019.09.008.
48. Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute lymphoblastic leukemia. J Clin Invest **2012**;122(10):3398-406 doi 10.1172/JCI61269.
49. Cao L, Wang N, Pan J, Hu S, Zhao W, He H*, et al.* Clinical significance of microRNA-34b expression in pediatric acute leukemia. Mol Med Rep **2016**;13(3):2777-84 doi 10.3892/mmr.2016.4876.
50. Poole CJ, Zheng W, Lodh A, Yevtodiyenko A, Liefwalker D, Li H*, et al.* DNMT3B overexpression contributes to aberrant DNA methylation and MYC-driven tumor

maintenance in T-ALL and Burkitt's lymphoma. Oncotarget **2017**;8(44):76898-920 doi 10.18632/oncotarget.20176.
51. Kowalczyk JR, Zawitkowska J, Lejman M, Drabko K, Samardakiewicz M, Matysiak M*, et al.* Long-term treatment results of Polish pediatric and adolescent patients enrolled in the ALL IC-BFM 2002 trial. Am J Hematol **2019**;94(11):E307-E10 doi 10.1002/ajh.25619.
52. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **2011**;27(11):1571-2 doi 10.1093/bioinformatics/btr167.
53. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods **2014**;11(11):1138-40 doi 10.1038/nmeth.3115.
54. Zhu LJ. Integrative analysis of ChIP-chip and ChIP-seq dataset. Methods Mol Biol **2013**;1067:105-24 doi 10.1007/978-1-62703-607-8_8.
55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics **2013**;29(1):15-21 doi 10.1093/bioinformatics/bts635.
56. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Research **2016**;45(4):e22-e doi 10.1093/nar/gkw967.
57. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O*, et al.* Systematic identification of genetic influences on methylation across the human life course. Genome Biol **2016**;17:61 doi 10.1186/s13059-016-0926-z.
58. Zou LS, Erdos MR, Taylor DL, Chines PS, Varshney A, Parker SCJ*, et al.* BoostMe accurately predicts DNA methylation values in whole-genome bisulfite sequencing of multiple human tissues. BMC Genomics **2018**;19(1):390 doi 10.1186/s12864-018-4766-y.
59. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y*, et al.* ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO reports **2018**;19(12):e46255 doi 10.15252/embr.201846255.
60. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research **2014**;42(Web Server issue):W187-W91 doi 10.1093/nar/gku365.
61. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences **2005**;102(43):15545 doi 10.1073/pnas.0506580102.
62. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J*, et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet **2003**;34(3):267-73 doi 10.1038/ng1180.
63. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res **2009**;37(1):1-13 doi 10.1093/nar/gkn923.
64. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols **2009**;4(1):44-57 doi 10.1038/nprot.2008.211.
65. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV*, et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics **2013**;14:128 doi 10.1186/1471-2105-14-128.
66. 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 doi 10.1186/s13059-014-0550-8.

	Normal T cells	COSMe-I T-ALL	COSMe-II T-ALL
Cluster A PRC2 Target genes	No DNA methylation	Low DNA methylation	High DNA methylation
		High H3K27me3	Low H3K27me3
		Epigenetically young	Epigenetically old
		Lck-Cre$^{Tg/+}$ Pten$^{fl/fl}$ mouse model	CD2-Lmo2^{Tg} mouse model
Cluster B PU.1 binding sites	Low DNA methylation increases with maturation of thymocytes	High DNA methylation	Two subgroups with low or high DNA methylation
	Inversely correlated with PU.1 expression		Immature T-ALLs have low Cluster B methylation
Cluster C	Heterogeneous DNA methylation	Heterogeneous DNA methylation	Heterogeneous DNA methylation
Genetics		Enriched for TAL1 rearranged T-ALLs	Enriched for HOXA, TLX3 and NKX2-1 T-ALLs

Table 1. Summary overview of DNA methylation clusters in COSMe Type-I and COSMe Type-II T-ALLs and normal T cells with different characteristics.

Figure Legends

Figure 1. DNA methylation profiling in normal and malignant T-cells. (A) Unsupervised clustering of mean centered methylation score (β values) of 5000 most variably methylated CpGs in 109 T-ALL cases and 10 subsets of developing thymocytes (2 biological replicates each), with indication of the three main clusters of CpG probes and their location with respect to CpG context. T-ALL subtypes and genetic defects of all T-ALL cases are indicated below the heatmap. **(B-C)** Mean β values of Cluster A (B) and Cluster B (C) CpG sites in COSMe type I and II T-ALLs and normal thymic precursors. ISP: Immature Single Positive, DP: $CD4^+$ $CD8^+$ Double Positive, SP: Single Positive. **(D)** Percentage of probes located inside different genomic categories for each of the clusters defined above. **(E)** Mean expression (log2-transformed normalized counts per million, EdgeR) per sample of genes linked to CpG sites in Cluster A, B and C (A) in a T-ALL cohort (GSE110637) plotted separately for TAL-rearranged T-ALLs (enriched in COSMe-I) and T-ALLs without such rearrangements (enriched in COSMe-II). Normal thymocytes (GSE151079) were included as healthy controls. Genes in the NOTCH1 pathway were used as positive reference value for T-ALL. **(F)** ChIP-seq enrichment in cluster A, B and C for H3K27me3 in CD34+ HSPCs (GSM2277181) and in the JURKAT T-ALL cell line (GSM2279072) and for PU.1 in CD34+ HSPCs (GSM1816090) and the TALL-1 cell line transfected with FLAG-PU.1 (GSE128837). Data publicly available from the Gene Expression Omnibus. **(G)** COSMe signature in paired primary-relapsed T-ALL samples(21). Row scaled β values of CpG methylation are shown.

Figure 2. Correlation between Cluster A DNA methylation and H3K27me3. (A) H3K27me3 binding at Cluster A, B and C sites in COSMe-I and COSMe-II T-ALLs (summarized for visualization, n=3 per group). **(B)** Volcano plot of differential H3K27me3 between COSMe-II and COSMe-I T-ALLs. Gene names from the top 25 most differential genes based on adjusted p value are indicated. **(C)** Patient specific comparison between Cluster A DNA methylation levels (β values) and corresponding differential H3K27me3 (normalized counts, DESeq2) in COSMe-II vs COSMe-I. Pearson coefficient of the linear correlation with p value is shown. **(D)** The levels of DNA methylation (β values) at all differential H3K27me3 regions between COSMe-II

and COSMe-I T-ALLs that were covered by the EPICarray dataset, separately plotted for CpG islands (111 regions) and Open Sea CpGs (111 regions).

Figure 3. Epigenetic age in T-ALLs at diagnosis and relapse. (A) Correlation between Epigenetic "Epitoc" age and mean methylation per patient in cluster A, B and C. **(B)** Age at diagnosis (Dx) and age predicted by Horvath for each patient and normal T cell subset. **(C)** Mean age at diagnosis (Dx) and in different age predictors in COSMe-I and COSMe-II subgroups with P values of differences (Wilcoxon signed-rank test). **(D)** Epigenetic "Epitoc" age of paired diagnosis-relapse patients.

Figure 4. Murine *In vivo* T-ALL models recapitulate the COSMe phenotype. (A) Mean centered methylation scores (β values) of most variably methylated CpGs in RRBS profiling of aging (8, 16 and 24 weeks) preleukemic and fully transformed *CD2-Lmo2*^{tg} mice (Tg), 8 week old and fully transformed *Lck-Cre*$^{Tg/+}$*Pten*$^{fl/fl}$ mice (KO) and littermate controls (WT) with indication of three main subclusters and its relation to CpG context. Data from 4 mice per condition. **(B)** Percentage of CpGs located in different genomic categories for each of the clusters defined in (A). **(C)** Per sample mean expression of genes in the clusters defined in (A) shown in *CD2-Lmo2*^{tg} mice (Limma Log2 normalized counts, microarray data, GSE49164) or in *Lck-Cre*$^{Tg/+}$*Pten*$^{fl/fl}$ (FPKM RNAseq, GSE115346). Notch1 pathway genes are shown as positive reference values. **(D)** ChIP-seq of H3K27me3 in murine clusters 1-3 (A) in CD34+ HSPCs (GEO: GSM4067369) and mouse T-ALL (GEO: GSM1506768). **(E)** Methylation at 726 Cluster 1 CpGs (defined above in (A) that overlap with genes in COSMe Cluster A defined in Figure 1A, now termed ClusterA^{mm} sites, in aging pre-leukemic and transformed *CD2-Lmo2*^{tg}, *Lck-Cre*$^{Tg/+}$*Pten*$^{fl/fl}$ and control mice. **(F)** Relative epigenetic age estimation for (pre-)leukemic thymocytes from *CD2-Lmo2*^{tg}, *Lck-Cre*$^{Tg/+}$*Pten*$^{fl/fl}$ and control mice.

Figure 5. DNA methylation profiling of T-ALL Patient Derived Xenografts treated with Decitabine *in vivo*. (A) Mice engrafted with T-ALL patient samples were treated two cycles with vehicle or Decitabine (DAC, 0.5mg/kg body weight) for 5 consecutive days follow by 2 days off and followed for survival analysis. Kaplan-Meier analysis of leukemia-free survival is shown (Log-rank Mantel-Cox test) for each of the 4 PDX samples analyzed. **(B)** Mice engrafted with T-ALL patient sample were treated with vehicle or Decitabine (0.5mg/kg body weight) for 5 consecutive

days. Seven days after treatment initiation, leukemic blasts from control and Decitabine treated animals (n=3 each group) were collected from the spleen and used for DNA methylation profiling with EPIC arrays. Number of significantly differentially methylated CpGs are shown for each PDX comparing Decitabine treated with Vehicle treated samples, with respect to their location inside different genomic categories. Red bars indicate more methylation after Decitabine treatment, blue bars mark a decrease in methylation upon treatment. **(C)** β values of COSMe CpGs in PDX#1 and PDX#2 T-ALLs with Decitabine or vehicle treatment. Methylation at the 5000 COSMe CpG sites was used to determine COSMe-I and COSMe-II T-ALL subtypes. **(D)** RNA sequencing of samples described in (A). Volcano plots show Log2 fold change of protein coding gene expression upon Decitabine treatment for PDX T-ALL#1 and PDX T-ALL#2. Genes with adjusted p-value lower than 0.05 are represented in blue (Deseq2). **(E)** Pre-ranked Gene Set Enrichment Analysis (GSEA) of for PDX#1 and PDX#2 in Decitabine versus control conditions.

Figure 1

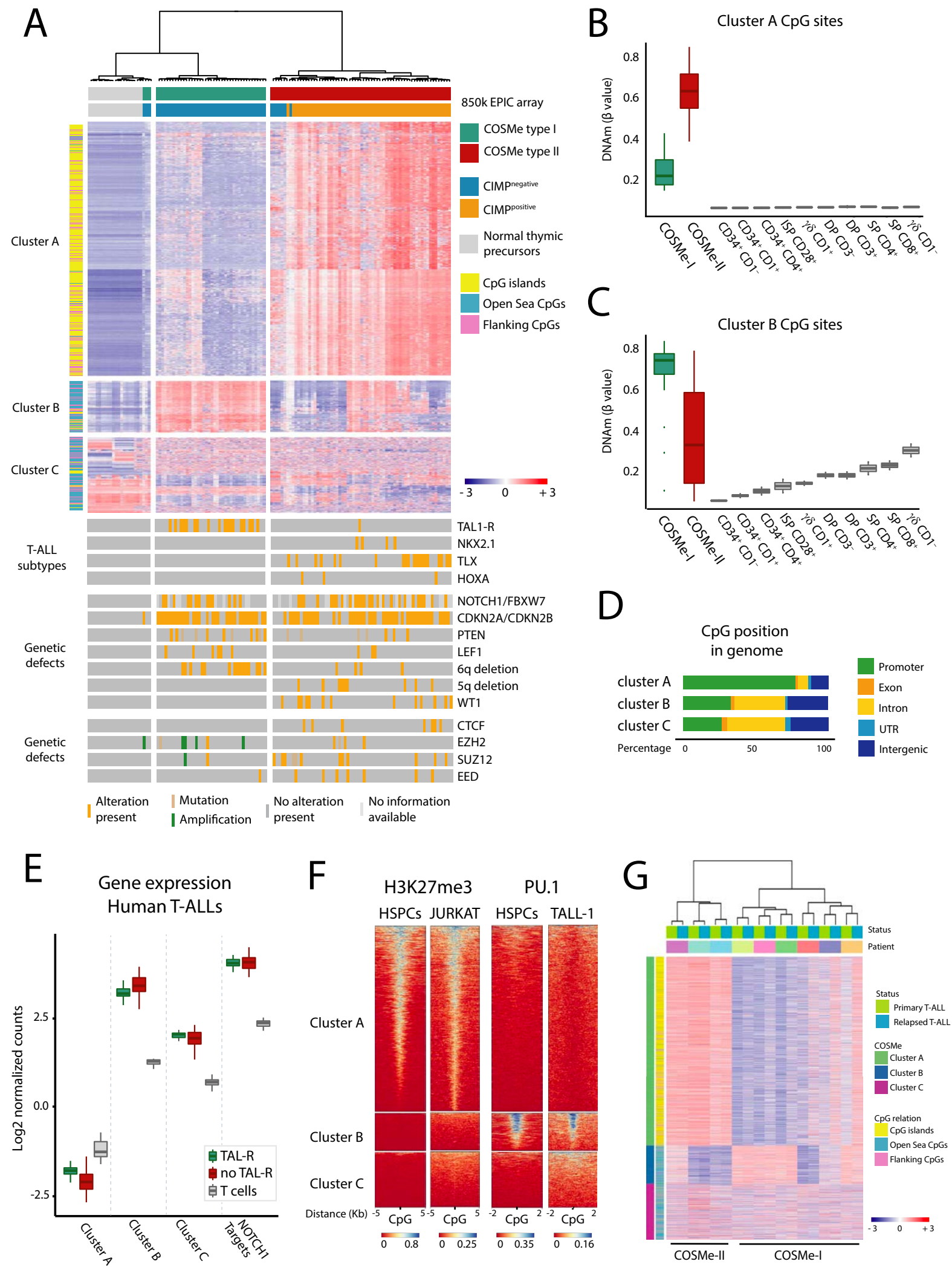

A
850k EPIC array
COSMe type I
COSMe type II
CIMP negative
CIMP positive
Normal thymic precursors
CpG islands
Open Sea CpGs
Flanking CpGs
Cluster A
Cluster B
Cluster C
- 3
0
+ 3
T-ALL subtypes
TAL1-R
NKX2.1
TLX
HOXA
Genetic defects
NOTCH1/FBXW7
CDKN2A/CDKN2B
PTEN
LEF1
6q deletion
5q deletion
WT1
Genetic defects
CTCF
EZH2
SUZ12
EED
Alteration present
Mutation
Amplification
No alteration present
No information available
B
Cluster A CpG sites
DNAm (β value)
0.8
0.6
0.4
0.2
COSMe-I
COSMe-II
CD34+ CD1-
CD34+ CD1+
CD34+ CD4+
ISP CD28+
γδ CD1+
DP CD3-
DP CD3+
SP CD4+
SP CD8+
γδ CD1-
C
Cluster B CpG sites
DNAm (β value)
0.8
0.6
0.4
0.2
COSMe-I
COSMe-II
CD34+ CD1-
CD34+ CD1+
CD34+ CD4+
ISP CD28+
γδ CD1+
DP CD3-
DP CD3+
SP CD4+
SP CD8+
γδ CD1-
D
CpG position in genome
cluster A
cluster B
cluster C
Percentage
0
50
100
Promoter
Exon
Intron
UTR
Intergenic
E
Gene expression Human T-ALLs
Log2 normalized counts
2.5
0.0
-2.5
TAL-R
no TAL-R
T cells
Cluster A
Cluster B
Cluster C
NOTCH1 Targets
F
H3K27me3
PU.1
HSPCs
JURKAT
HSPCs
TALL-1
Cluster A
Cluster B
Cluster C
Distance (Kb)
-5
CpG
5
-5
CpG
5
-2
CpG
2
-2
CpG
2
0
0.8
0
0.25
0
0.35
0
0.16
G
Status
Patient
Status
Primary T-ALL
Relapsed T-ALL
COSMe
Cluster A
Cluster B
Cluster C
CpG relation
CpG islands
Open Sea CpGs
Flanking CpGs
- 3
0
+ 3
COSMe-II
COSMe-I

Figure 2

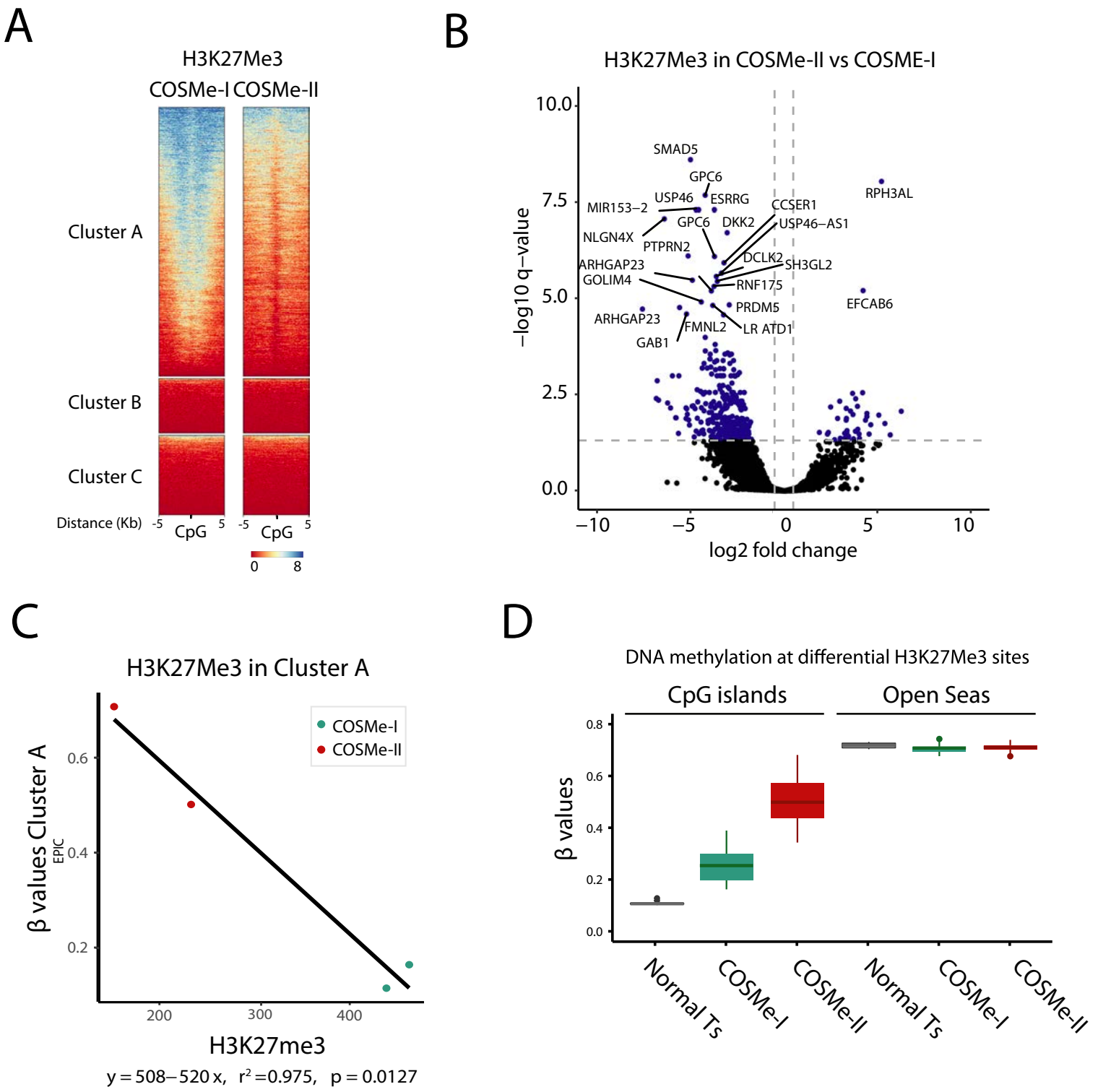
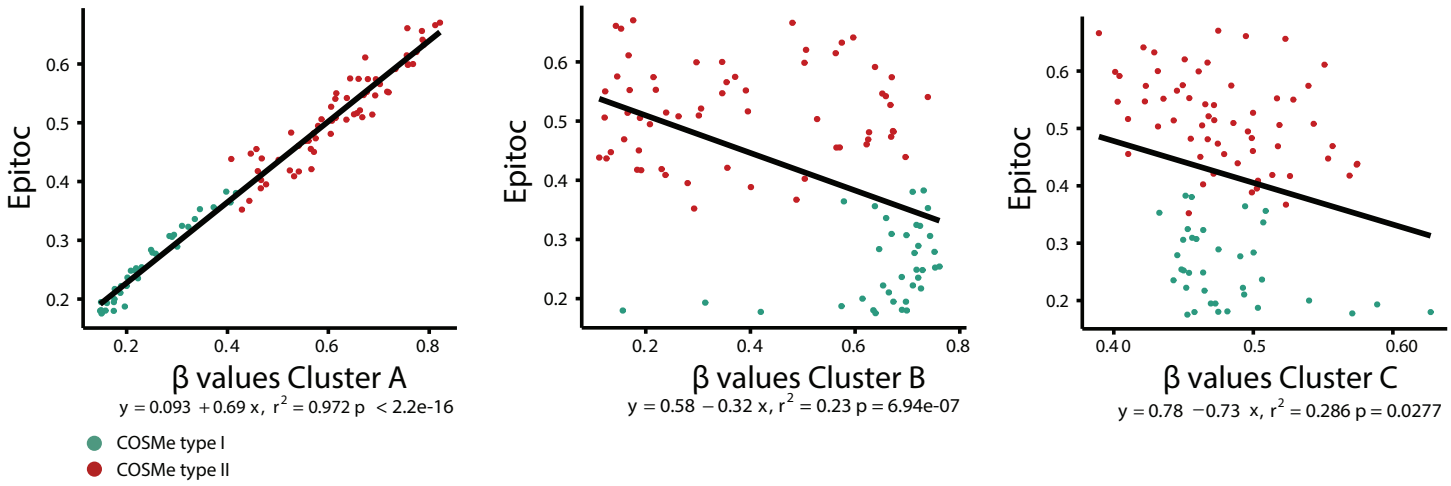
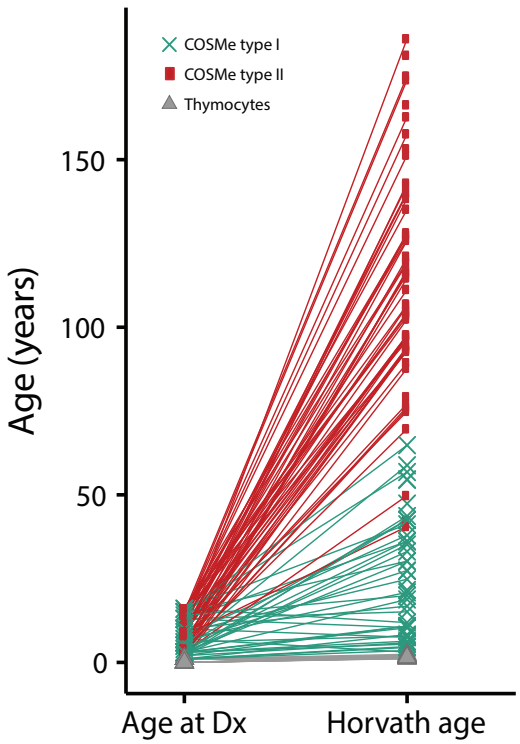

A
H3K27Me3
COSMe-I COSMe-II
Cluster A
Cluster B
Cluster C
Distance (Kb)
-5
CpG
5
0
8
B
H3K27Me3 in COSMe-II vs COSME-I
−log10 q−value
log2 fold change
SMAD5
GPC6
USP46
ESRRG
MIR153−2
CCSER1
GPC6
DKK2
USP46−AS1
NLGN4X
PTPRN2
DCLK2
SH3GL2
ARHGAP23
GOLIM4
RNF175
PRDM5
ARHGAP23
FMNL2
LR ATD1
GAB1
RPH3AL
EFCAB6
C
H3K27Me3 in Cluster A
β values Cluster A EPIC
H3K27me3
COSMe-I
COSMe-II
y = 508−520 x, r² =0.975, p = 0.0127
D
DNA methylation at differential H3K27Me3 sites
CpG islands
Open Seas
β values
Normal Ts
COSMe-I
COSMe-II
Normal Ts
COSMe-I
COSMe-II

Figure 3

A



B



C

	Age at Dx	Horvath age	Epitoc age
COSMe-I	6.75	24.7	0.26
COSMe-II	8.85	112.23	0.52
P Value	0.02506	<2.2e-16	3.85e-16

D

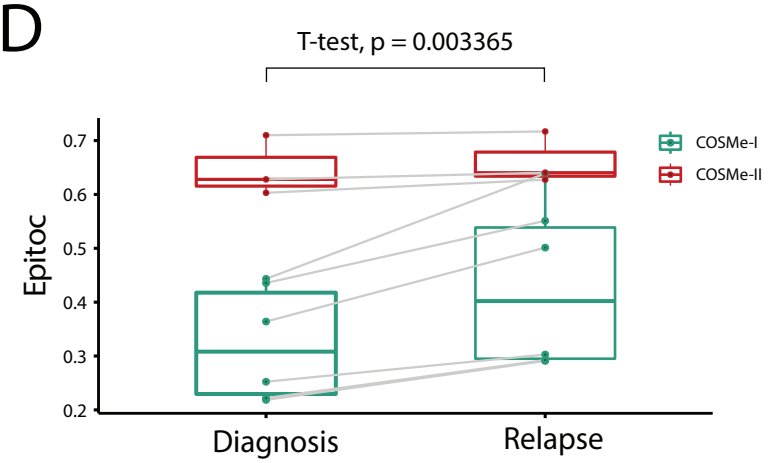


Figure 4

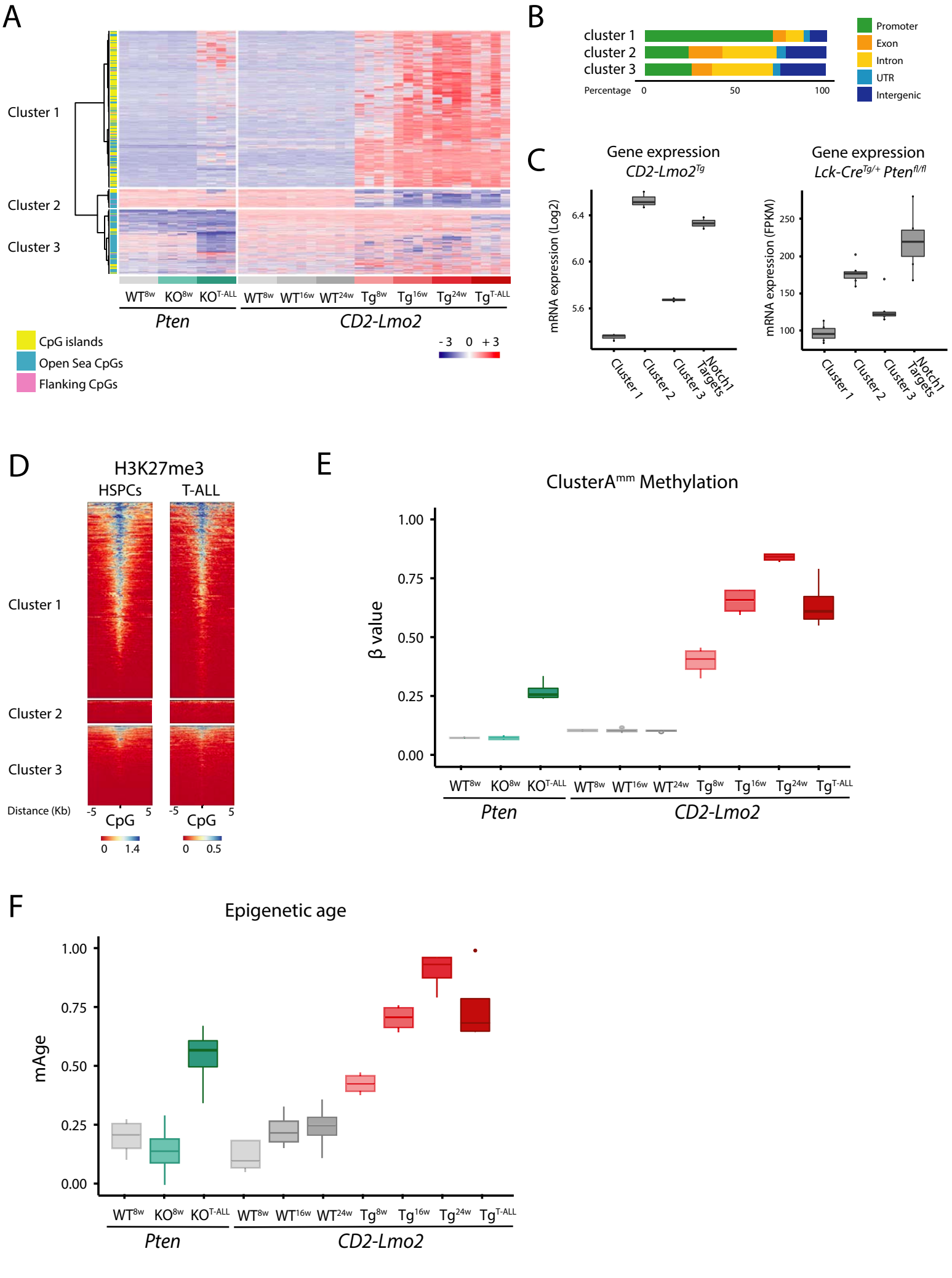
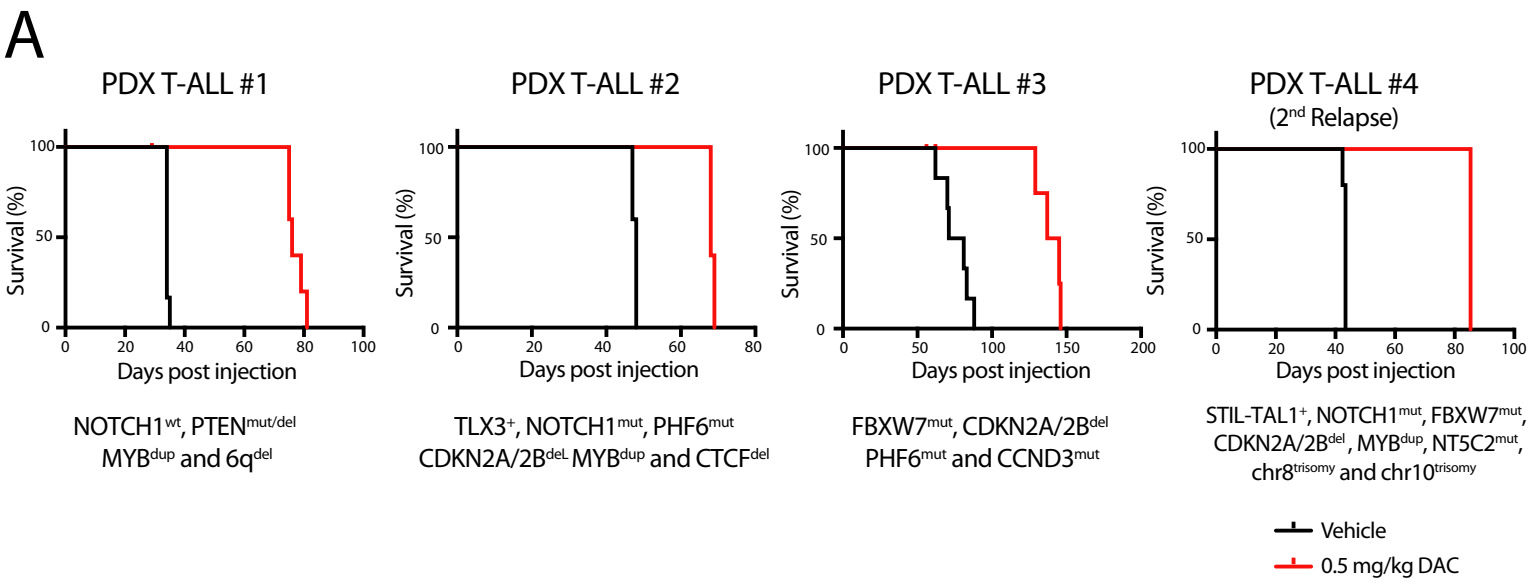
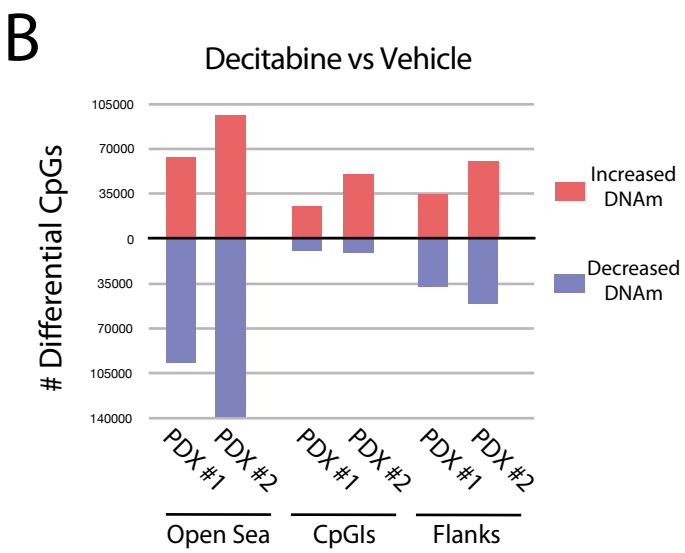
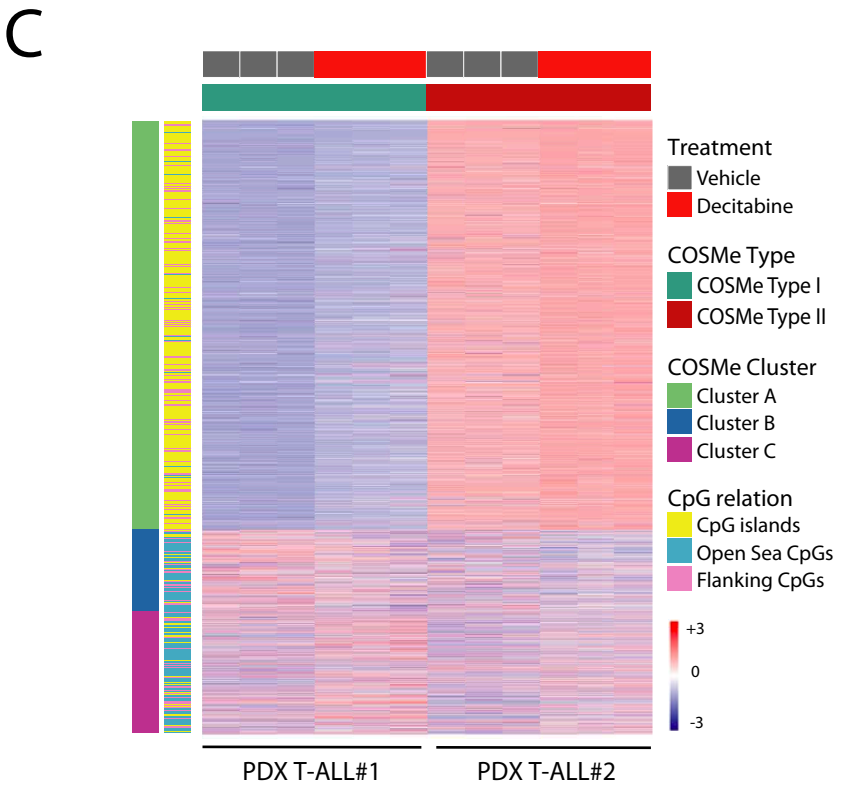
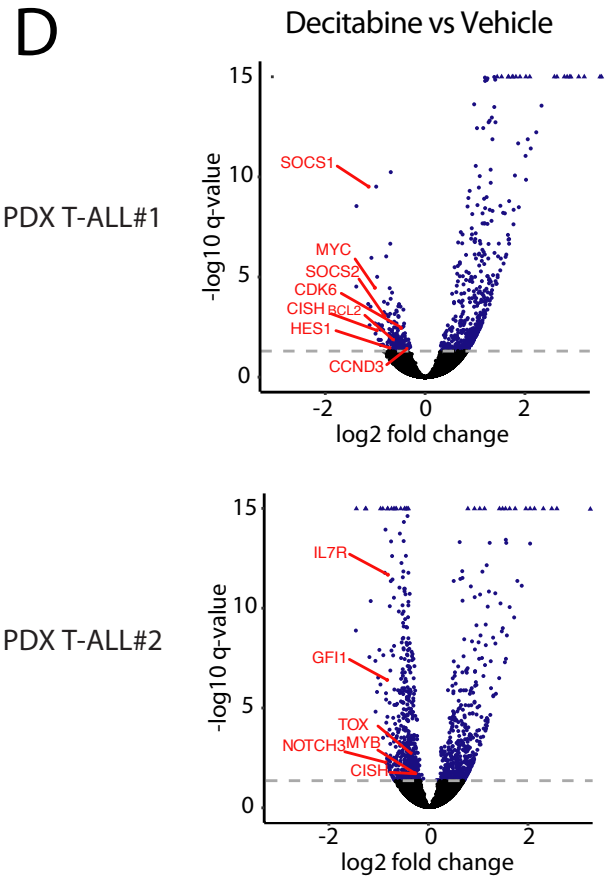

A
Cluster 1
Cluster 2
Cluster 3
WT8w KO8w KOT-ALL WT8w WT16w WT24w Tg8w Tg16w Tg24w TgT-ALL
Pten
CD2-Lmo2
CpG islands
Open Sea CpGs
Flanking CpGs
-3 0 +3
B
cluster 1
cluster 2
cluster 3
Percentage 0 50 100
Promoter
Exon
Intron
UTR
Intergenic
C
Gene expression
CD2-Lmo2Tg
mRNA expression (Log2)
6.4
6.0
5.6
Cluster 1
Cluster 2
Cluster 3
Notch1 Targets
Gene expression
Lck-CreTg/+ Ptenfl/fl
mRNA expression (FPKM)
250
200
150
100
Cluster 1
Cluster 2
Cluster 3
Notch1 Targets
D
H3K27me3
HSPCs
T-ALL
Cluster 1
Cluster 2
Cluster 3
Distance (Kb)
-5 CpG 5
0 1.4
-5 CpG 5
0 0.5
E
ClusterAmm Methylation
β value
1.00
0.75
0.50
0.25
0.00
WT8w KO8w KOT-ALL WT8w WT16w WT24w Tg8w Tg16w Tg24w TgT-ALL
Pten
CD2-Lmo2
F
Epigenetic age
mAge
1.00
0.75
0.50
0.25
0.00
WT8w KO8w KOT-ALL WT8w WT16w WT24w Tg8w Tg16w Tg24w TgT-ALL
Pten
CD2-Lmo2

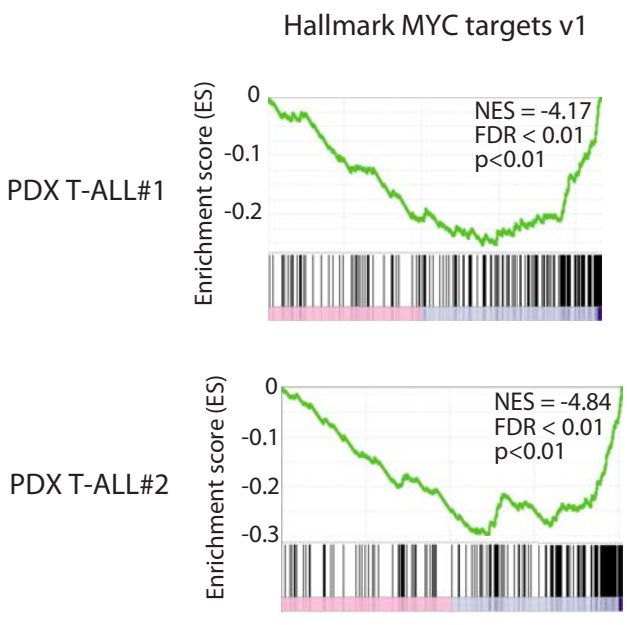
Figure 5


A
PDX T-ALL #1
PDX T-ALL #2
PDX T-ALL #3
PDX T-ALL #4
(2nd Relapse)
Survival (%)
Days post injection
NOTCH1wt, PTENmut/del
MYBdup and 6qdel
TLX3+, NOTCH1mut, PHF6mut
CDKN2A/2Bdel MYBdup and CTCFdel
FBXW7mut, CDKN2A/2Bdel
PHF6mut and CCND3mut
STIL-TAL1+, NOTCH1mut, FBXW7mut,
CDKN2A/2Bdel, MYBdup, NT5C2mut,
chr8trisomy and chr10trisomy
Vehicle
0.5 mg/kg DAC


B
Decitabine vs Vehicle
Differential CpGs
Increased DNAm
Decreased DNAm
PDX #1
PDX #2
Open Sea
CpGIs
Flanks


C
Treatment
Vehicle
Decitabine
COSMe Type
COSMe Type I
COSMe Type II
COSMe Cluster
Cluster A
Cluster B
Cluster C
CpG relation
CpG islands
Open Sea CpGs
Flanking CpGs
+3
0
-3
PDX T-ALL#1
PDX T-ALL#2


D
Decitabine vs Vehicle
PDX T-ALL#1
PDX T-ALL#2
-log10 q-value
log2 fold change
SOCS1
MYC
SOCS2
CDK6
CISH
BCL2
HES1
CCND3
IL7R
GFI1
TOX
NOTCH3
MYB
CISH


E
Hallmark MYC targets v1
PDX T-ALL#1
PDX T-ALL#2
Enrichment score (ES)
NES = -4.17
FDR < 0.01
p<0.01
NES = -4.84
FDR < 0.01
p<0.01