Epigenetic analysis of patients with T-ALL identifies poor outcomes and a hypomethylating agent-responsive subgroup

Aurore Touzart, Anand Mayakonda, Charlotte Smith, Joschka Hey, Reka Toth, Agata Cieslak, Guillaume Andrieu, Christine Tran Quang, Mehdi Latiri, Jacques Ghysdael, et al.

Title: Epigenetic blueprint identifies poor outcome and hypomethylating agent-responsive T-ALL subgroup

Authors: Aurore Touzart[1,2]†, Anand Mayakonda[1,3]†, Charlotte Smith[2], Joschka Hey[1,3,4], Reka Toth[1], Agata Cieslak[2], Guillaume P. Andrieu[2], Christine Tran Quang[5], Mehdi Latiri[2], Jacques Ghysdael[5], Salvatore Spicuglia[6], Hervé Dombret[7], Norbert Ifrah[8], Elizabeth Macintyre[2], Pavlo Lutsik[1,9], Nicolas Boissel[7], Christoph Plass[1,9*] and Vahid Asnafi[2*]

Affiliations:

[1]Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany

[2]Université de Paris (Descartes), Institut Necker -Enfants Malades (INEM), Institut national de la santé et de la recherche médicale (Inserm) U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris, Hôpital Necker Enfants-Malades, Paris, France

[3]Faculty of Biosciences, Heidelberg University, 69120 Heidelberg, Germany

[4]Germany-Israeli Helmholtz Research School in Cancer Biology

[5]Institut Curie, CNRS UMR3348, University Paris Sud - Paris-Saclay, Orsay, France

[6]Aix-Marseille University, Inserm, Theories and Approaches of Genomic Complexity (TAGC), Equipe labellisée Ligue, UMR1090, 13288 Marseille, France

[7]Université Paris Diderot, Institut Universitaire d'Hématologie, EA-3518, Assistance Publique-Hôpitaux de Paris, University Hospital Saint-Louis, Paris, France.

[8]PRES LUNAM, CHU Angers service des Maladies du Sang et INSERM U 892, Angers, France,

[9]German Cancer Research Consortium (DKTK)

* To whom correspondence should be addressed: e-mail: vahid.asnafi@aphp.fr (V.A); c.plass@dkfz-heidelberg.de (C.P)

†Contributed equally

One Sentence Summary: In-depth characterization of normal and adult T-ALL methylomes identify clinically aggressive hypermethylated subgroup sensitive to DNA hypomethylating agents.

Abstract: Adult T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy still associated with poor outcome calling for new therapeutic options. DNA methylation landscapes of adult T-ALL remain largely uncharacterized. Here, we systematically analyzed the DNA methylation profiles of normal thymic sorted T-cell subpopulations and 143 primary adult T-ALLs as part of French GRAALL 2003-2005 trial. Our results indicated that T-ALL is epigenetically distinct consisting of five major subtypes (C_1-C_5) which were either associating with co-occurring *DNMT3A/IDH2* mutations (C_1), *TAL1* deregulation (C_2), *TLX3* (C_3), *TLX1*/in *cis-HOXA9* (C_4) or in *trans-HOXA9* (C_5) overexpression. Integrative analysis of DNA methylation and Gene expression identified potential cluster-specific oncogenes and tumor suppressor genes. Importantly, in addition to an aggressive hypomethylated subgroup (C_1), our data identified an unexpected subset of hypermethylated T-ALL (C_5) associated with poor

outcome and primary therapeutic response. Using mouse xenografts, we showed that hypermethylated T-ALL samples displayed therapeutic response to the DNA hypomethylating agent, 5-Azacitidine, which significantly delayed tumor progression suggesting epigenetic-based therapies as a novel treatment option in hypermethylated T-ALL

[Main Text:]

Introduction

T-ALL (T-cell Acute Lymphoblastic Leukemia) is an aggressive hematological malignancy, most common among children (10-15% of ALL) and young adults (20-25% of ALL) (*1, 2*). Although intensive chemotherapy results in high cure rates, survival remains particularly poor in relapsing patients (*3*) calling for novel targeted therapeutic approaches. Both pediatric and adult T-ALLs harbor common oncogenic driver mutations leading to deregulation of specific transcription factors (TFs), including basic helix-loop-helix family members (*TAL1, TAL2, LYL1* and *BHLHB1*), LIM-only domain family members (*LMO1* and *LMO2*), homeobox genes (*TLX1, TLX3, HOXA, NKX2-1, NKX2-2*) as well as oncogenic TFs such as *MYB* and *MYC* (*4*). Investigation of transcriptional signatures identified molecular subgroups with specific driver oncogenes and maturation arrest stages including *TAL1/LMO* (mature or late cortical stage), *TLX1, TLX3* and *NKX2-1/NKX2-2* (early cortical stage), *LYL1, MEF2C* and *HOXA* (immature stage) (*5*). In addition to these oncogenic drivers, numerous additional recurrent genetic alterations have been described, including mutations in genes involved in the NOTCH1 pathway (*NOTCH1, FBXW7*) and deletion or epigenetic silencing of *CDKN2A*, both observed in more than 70% of T-ALLs (*6*). Past work has demonstrated that cancer-specific changes in DNA methylation (DNAm) include both global hypomethylation and focal hypermethylation changes at discrete loci, associated with distinct histone modification patterns and chromatin conformations, creating functional chromatin states (*7*). T-ALL is among a subset of cancers with the highest frequency of mutations in enzymes regulating epigenetic processes, suggesting that deregulation of the epigenetic landscape may play an important role in T-ALL leukemogenesis (*8, 9*). Notably, genes involved in establishing DNAm patterns are frequently mutated in adult T-ALL. *DNMT3A* mutations are found in about 18% of cases and are associated with an adverse outcome (*10, 11*). Mutations in TET family genes and *IDH1/2* are also frequent (*12-14*).

Several studies have utilized DNAm as a marker to characterize cancer samples in search of markers for relapse prediction, molecular classification or mechanisms involved in disease transformation (*15-17*). In pediatric T-ALL, DNAm analysis identified two distinct CpG island methylator phenotype (CIMP) groups, in which CIMP-negative patients displayed a significantly higher cumulative incidence of relapse as compared to CIMP-positive patients, suggesting prognostic relevance of aberrant DNAm profiles in T-ALL (*18-20*). Similar observations have also been made recently in adult T-ALL (*21*). Given the potential therapeutic impact of epigenetic anomalies, in particular their reversible nature by hypomethylating agents, we analyzed genome-wide DNAm patterns of primary adult T-ALL samples. Our results indicate that T-ALL is heterogeneous and consists of epigenetically distinct subgroups with characteristic clinical features. Of note, we identify an unexpected hypermethylated T-ALL subgroup of poor clinical outcome, which is likely to benefit from targeted therapy.

Results

DNA methylation blueprints identify distinct epigenetic subgroups of T-ALL

To better understand the role of DNAm in T-ALL leukemogenesis, we performed genome-wide DNAm analysis using Illumina Infinium Methylation EPIC BeadChip (EPIC arrays) in a series of 143 primary adult T-ALL samples, as well as on sorted normal thymic T-cell subpopulations (see **Methods** section). Patient clinical, phenotypic, and oncogenetic characteristics are described in **Table 1**. Using unsupervised clustering, we classified T-ALL samples within the cohort into five distinct, robust and stable sub-clusters **(Figure 1A, B** and **see methods** and **Supplementary Figure S1)**. These data-driven *de novo* clusters showed variable degrees of genome-wide DNAm levels compared to normal thymic populations **(Figure 1C)**. Based on their genome-wide DNAm level, we named the clusters C_1, C_2, C_3, C_4 and C_5, with C_1 displaying the lowest level of DNAm and C_5 the highest.

Additional characterization of the clusters revealed a significant association with thymic maturation arrest stages and key oncogenetic alterations **(Figure 1D-G)**. C_1 represented the smallest cluster ($n = 14$ samples, 9.8%) and was characterized by an Early Thymic Precursor (ETP) phenotype (7/11; $P < 0.01$), an immature maturation arrest stage (9/12; $P < 0.01$) and under-representation of classical T-ALL oncogenic drivers (*HOXA, TAL1, TLX1* and *TLX3*). C_2 was associated with a more mature αβ-lineage maturation arrest stage (28/31; $P < 0.01$) and was mainly characterized by *TAL1* deregulation (SIL-TAL1 or upstream-neoenhancer positive) (16/33; $P < 0.01$). Of note, all but one of the *TAL1* deregulated cases (16/17) clustered together, suggesting that they display similar DNAm signature regardless of the molecular mechanism leading to *TAL1* deregulation. Importantly, overall genome-wide methylation levels of C_1 and C_2 showed no significant changes compared to normal thymic subpopulations despite the differences in maturation arrest stages or genetic alterations **(Figure 1C)**. In contrast, clusters C_3, C_4, and C_5 displayed significant genome-wide hypermethylation compared to normal thymic subpopulations (t-test, $P < 0.001$; **Figure 1C**) and were enriched for overexpression of HOX gene family members. Although *TLX1* and *TLX3* proteins are very similar, *TLX1* and *TLX3* deregulated leukemia clustered separately. C_3 contained samples enriched for overexpression of *TLX3* (14/22) and a TCRγδ phenotype whereas C_4 contained samples displaying *TLX1* overexpression (25/38), and *cis*-activated (resulting from *cis*-deregulation by translocation into the TCRβ locus) *HOXA9* overexpression (7/37) along with IMβ/pre-αβ (αβ lineage) maturation arrest stages. C_5 samples were most significantly hypermethylated compared to normal thymic subpopulations (**Figure 1C**; t-test, $P < 0.001$) and were characterized by *trans*-activated (translocations involving MLL, MLLT10 rearrangements, SET-NUP214) *HOXA9* overexpression (16/30) and ETP and immature maturation arrest stages, as described (*22*). Importantly, ETP-ALLs clustered in C_1 and C_5, which displayed distinct hypo-, and hyper- methylation profiles respectively, suggesting a high degree of heterogeneity within ETP-ALLs, and that an ETP phenotype alone cannot drive or result from methylation changes.

Besides the classical T-ALL driver events - CNVs and mutational alterations for a targeted panel of genes that are frequently altered in T-ALL were also studied **(see methods; Supplementary Figure S2A)** (*22*). Among genes involved in establishing DNA methylation patterns, *DNMT3A* was the most frequently mutated (14%) followed by *TET2* (5%), *IDH2* (4%), *TET3* (3%) and *IDH1* (1%). Comparison of mutational events within each cluster against all other

clusters identified thirteen genes, which were preferentially mutated in particular epigenetic clusters (**Figure 1H**; Fisher's exact test, *FDR* < 0.05; **Supplementary Table S1**). Whereas *DNMT3A, TET2/3, IDH1* mutations alone did not seem to be significantly associated with any specific methylation cluster, all *IDH2* mutations co-occurred with *DNMT3A* mutations. A large fraction of these co-occurring *IDH2/DNMT3A* mutations (5/6) were found within the C_1 subgroup, suggesting their potential synergistic role for the underlying leukemogenesis (*23*) **(Figure 1H)**. In the entire cohort, *DNMT3A* mutations were largely of loss-of-function/dominant-negative type, with 30% (6/20) involving the R882 hotspot, thereby affecting the methyltransferase domain **(Supplementary Figure S2B)**. Mutations in *IDH2* occurred at the known oncogenic hotspot R140 **(Supplementary Figure S2C)** (*24*).

PTEN alterations were enriched in C_2 (enriched in *TAL1* deregulated cases). *PTEN* mutations were mostly loss-of-function INDELs thereby dysregulating the PI3K signaling pathway, in line with the higher frequency of *PTEN* alterations in *TAL1* deregulated and mature T-ALL cases (*25, 26*). The C_2 subgroup was also characterized by a lower frequency of mutations affecting both epigenomic regulators in general and the JAK-STAT pathway. C_3 was particularly enriched for *WT1* and *STAT5B* mutations affecting the JAK-STAT pathway (*25*). C_4 demonstrated a significantly higher frequency of *BCL11B* mutations. Importantly, *SUZ12* and *EZH2* alterations, both critical components of the PRC2 complex, clustered in the C_5 subgroup. A large fraction of *NF1* alterations and a low frequency of *CDKN2A* deletions were also observed in this subgroup.

Altogether, based on methylation profiles we have identified five robust subgroups of T-ALL, and demonstrated that epigenetic clusters correlate with the driver oncogenes and are associated with distinct maturation arrest stages. In particular, we have identified a novel uncharacterized subgroup (C_1) with co-occurring *DNMT3A/IDH2* mutations.

T-ALL methylomes differ from thymic T-cell subpopulations but DNA methylation predicts maturation arrest stages of epigenetic clusters

We compared T-ALL samples with a series of sorted normal thymic subpopulations including thymic CD34+, 4ISP (Immature Simple positive CD4+), DP and TCR-/CD3- (Double Positive CD4+ CD8+ TCR- CD3-), DP and TCR+/CD3+ (Double Positive CD4+ CD8+ TCR+ CD3+), SP4 (Single positive CD4+) and SP8 (Single Positive CD8+) cells (see **Methods** section). We first analyzed the dynamics of DNAm changes during normal T-cell maturation; principal component analysis (PCA) robustly separated different subpopulations, suggesting significant heterogeneity in the T-cell precursor methylomes and thereby implicating epigenetic programming as an underlying part of T-cell development (**Figure 2A**). We next performed differential methylation analysis, which identified a set of differentially methylated probes (DMPs) for each thymic subpopulation (tDMPs) **(Supplementary Figure S3A)(Supplementary Table S2)**. CD34+ cells contained the highest number of tDMPs (n = 7385), followed by the more mature SP4/SP8 cells (n = 5148 and n = 6145 respectively) with varying degrees of overlap between the subpopulations **(Supplementary Figure S3B)**. Genes associated with precursor CD34+ cell-specific tDMPs were affiliated with the gene ontology terms including chemotaxis or cell migration, in line with the process by which migratory hematopoietic precursors from the bone marrow respond to homing signals from thymic epithelial cells before restricting towards the thymic lineage **(Figure 2B)**. In contrast, 4ISP cells were significantly enriched for terms associated with lymphocyte activation and differentiation, and immune system development, suggesting lineage commitment towards T-

cell development **(Figure 2C)**. Similar terms were also enriched for the mature subpopulations **(Supplementary Figure S3C-F)**.

To further characterize tDMPs, we identified eight sub-clusters (tDMP1-tDMP8) based on the kinetics of DNAm changes **(Figure 2D** and **2E) (Supplementary Table S3)**. tDMP1 and tDMP2 consisted of hypermethylated probes which became slightly hypomethylated in terminal stages, and were associated with processes such as cellular morphogenesis and cell development **(Supplementary Figure S4A, B)**. Clusters tDMP3, tDMP4, and tDMP5 showed a continuous gain of DNAm, suggesting silencing of cellular pathways associated with cell adhesion, cell communication, regulation of GTP-ase activity and the PI3k pathway **(Figure 2E and Supplementary Figure S4C, D, E)**. In contrast, clusters tDMP6, tDMP7 and tDMP8 contained genes associated with DMPs which showed progressive hypomethylation during differentiation. These clusters were exclusively associated with T-cell differentiation and activation and the adaptive immune system, implying the presence of a core set of CpGs whose hypomethylation is critical for T-cell maturation **(Supplementary Figure S4F-H)**.

As a complementary to the DNAm analysis of thymopoiesis, we also analyzed the gene expression data from 6 distinct thymic cell types generated by Roels et al (CD34+CD1-, ISPCD28+, DPCD3-, DPCD3+, SPCD4+ and SPCD8+) (*27*). Gene expression profiles of well-known candidate genes involved in thymopoiesis showed an inverse correlation with their promoter DNAm levels **(Supplementary Figure S5A)**. We further characterized the dynamics of gene expression and promoter DNAm profiles by comparing each mature thymic cell type to the CD34+ immature early thymic precursors. Both differential expression and differential promoter DNAm analysis revealed a linearly increasing number of differentially regulated genes along the maturation stages **(Supplementary Figure S5B)**. Next we identified a common set of genes which were both differentially expressed and showed a differential promoter methylation - which further revealed a core subset of genes which are both transcriptionally and epigenetically regulated in thymic developmental processes (N = 942) **(Supplementary Figure S5B-C, Supplementary Table S4)**. In addition, expression of these genes showed significant inverse correlation with their promoter DNAm levels with high expressed genes showing least methylation and vice-versa **(Supplementary Figure S5D, E)**. Large fraction of these genes also overlapped with the genes identified in recent single cell based assays thereby further validating the results (data not shown) (*28*).

Next, we sought to investigate whether tDMPs can recapitulate the known hierarchy of T-cell maturation stages. Using an unsupervised, distance-based neighbor joining approach, tDMPs robustly reconstructed T-cell developmental stages, with the phylogenetic tree leading to either mature SP4 or SP8 cells originating from precursor thymic CD34+ cells **(Figure 2F)**. Phylogenetic trees constructed using random CpGs failed to infer the same, thereby hinting at a critical role of tDMP-associated genes in T-cell development **(Supplementary Figure S6A and S6B)**. Of note, similar to DNAm, DEGs involved in thymopoiesis were also able to reconstruct the thymic developmental trajectory **(Figure 2F, inset plot)** suggesting a possible coordinated regulation of DNAm and gene expression profiles. We then hypothesized that T-ALL samples display distinct DNA methylomes compared to normal thymic subpopulations and that these DMPs are distinct from tDMPs associated with T-cell differentiation. To address this, we performed PCA on the entire sample cohort of primary T-ALL and normal T-cell subpopulations. This analysis showed clear separation of T-ALL samples from normal T-cell subpopulations, irrespective of their maturation stage, inferring that T-ALL are highly heterogeneous in their methylome **(Figure 2G;**

left panel**)**. We repeated the PCA analysis by excluding tDMPs and did not observe any significant changes, suggesting that T-ALL differed in their DNA methylomes compared to thymic subpopulations **(Figure 2G;** right panel**)**. Unsupervised hierarchical clustering showed similar results, suggesting that in addition to the DNAm signature obtained from the cell-of-origin, a significant number of DNAm changes in T-ALL are primarily related to leukemogenesis and are T-ALL specific. Interestingly, a subset of leukemic samples clustered close to normal subpopulations, implying a lower deviation of their methylome **(Supplementary Figure S6C, D)**.

Finally, we asked how T-ALL clusters diverge from normal T-cell development. In order to investigate this in an unsupervised way, we first performed PCA on the entire cohort (thymic subpopulations and T-ALL) using only tDMPs as reference. This analysis recovered the known developmental trajectory of T-cell subpopulations, and the possible ordering of T-ALL clusters relative to the trajectory **(Figure 2H)**. We next projected DNAm of T-ALL samples onto the normal thymic developmental tree, which highlighted the possible deviation of T-ALL samples from the normal developmental trajectory, thereby indicating potential maturation arrest stages **(Figure 2I)**. We then calculated the average methylation of T-ALL samples belonging to the same one of earlier identified clusters C_1 to C_5, and rebuilt the developmental trajectory. This display showed the hierarchical ordering of clusters along the normal T-cell developmental path, with C_5, and C_1 being derived from earlier T-cell maturation stages, followed by C_3, C_4, and C_2 - in line with the respective maturation arrest stages **(Figure 2I;** inlet plot**)**. Of note, all five epigenetic clusters were positioned between ISP and DP TCR- stages of T-cell development suggesting an early epigenetic divergence occurring between the cortex to subcapsular thymic zones (*29*). Taken together, our results suggest the potential role of DNAm in predicting the normal T-cell developmental trajectory and the order of maturation arrest stages for the five epigenetic T-ALL subgroups.

Methylation changes in normal and T-ALL-associated regulatory elements

To further understand the heterogeneity among the epigenetic clusters, we compared each cluster with the normal thymic subpopulations to identify DMPs (*FDR* < 0.05, |meth change| > 0.2) **(Supplementary Table S5)**. This analysis revealed varying degrees of DMPs with C_1 (34424 hyper/21478 hypo) and C_2 (48538 hyper/ 27757 hypo) displaying approximately equal proportion of hyper- and hypo- methylated probes. In contrast C_3 (76873 hyper/ 22272 hypo), C_4 (97297 hyper/ 26728 hypo), and C_5 (101198 hyper/ 14503 hypo) contained large number of hypermethylated probes **(Figure 3A)**. Further genomic and CpG annotations of the DMPs showed no significant differences in their overall distribution within clusters however, hypermethylated DMPs were strongly enriched in promoters and CpG islands across all the clusters **(supplementary Figure S7A)**. Recent evidence showed that DNAm changes in regulatory regions play a critical role in maintaining cellular identity, and malignant transformation in both solid and leukemic cancers (*30, 31*). Consequently, we generated the promoter (H3K4me3) and enhancer (H3K27ac/H3K4me1) landscape of normal thymic cell types at various developmental stages (Immature CD34+, EC (Early Cortical CD4+ CD8+ TCR/CD3-), LC (Late Cortical; CD4+ CD8+ TCR/CD3+), SP4 (Single Positive; CD4+, TCRαβ+), SP8 (Single Positive CD8+, TCRαβ+)) using the data collected from the BLUEPRINT consortium (*32*). We defined three types of enhancers (active, poised and putative) based on the presence or absence of H3K27ac, and H3K4me1 histone marks **(Supplementary Figure S7B and S7C)**. In addition, we also collected H3K27ac and H3K4me3 ChIP-seq profiles for 12 primary T-ALLs belonging to four distinct methylation clusters (7 to C_2, 2 to C_4 and 3 to C_5). PCA analysis based on the histone marks still retained the

DNAm-based clustering (**Figure 3B and 3C**). We next identified super-enhancers (SEs) and typical enhancers (TEs) for each normal T-cell and T-ALL methylation cluster. SE-associated genes from both normal and T-ALL contained genes known to be involved in T-cell maturation, and T-ALL leukemogenesis, respectively **(Supplementary Figure S7D) (Supplementary Table S6)**. Interestingly, a number of SEs identified in T-cell subtypes increased with the developmental stages but, we observed significantly higher number of SEs in T-ALL compared to normal T-cells ($P < 0.001$; t-test) (**Figure 3D**).

Next, we performed an enrichment analysis, which further revealed the dynamics of DNAm changes in T-ALL associated regulatory elements **(Figure 3E)**. Hypermethylated DMPs across all five clusters were mainly enriched within known TSS (+/-1000 base pairs) and active promoters (H3K4me3 peaks) derived from normal T-cell developmental datasets. Of note, we also observed enrichment of hypermethylated DMPs within normal thymus derived poised enhancers (H3K4me1+/H3K27ac-), suggesting a potential shut-down of enhancers that could be primed for activation in later T-cell developmental stages (*33*). In contrast, hypomethylated DMPs were enriched within intergenic regions. T-ALL associated SEs and TEs were largely enriched for hypomethylated DMPs suggesting active downstream expression of SE associated genes. Given the role of SEs in disease progression, hypomethylation of T-ALL specific regulatory elements suggests a DNAm mediated activation of SEs and further cancer progression (*34*).

Several studies have reported the binding of core oncogenic or cell specific master TFs within enhancer elements (*35, 36*). We therefore checked for the enrichment of TF motifs within 100 bps of hypomethylated DMPs located within T-ALL derived SEs and TEs. This revealed specific TF dynamics within each cluster **(Figure 3F) (Supplementary Table S7)**. For example, C_2 was enriched for TAL1, RUNX1/2, GATA4/6, MYB TF motifs, which have been reported to form auto-regulatory loops in *TAL1* deregulated T-ALLs thereby driving oncogenesis (*37*). C_1 was primarily associated with bZIP and "myeloid like" motifs such as CEBP, CHOP, GATA1/2, PU1 and AP1, in accordance with the samples being enriched for Immature/ETP geno/phenotypes **(Figure 1D)**. Similarly, C_5 was enriched for HOXA cluster-specific homeobox TFs.

Altogether, our results show diverging roles of DNAm in chromatin regulation by which disease associated enhancers are frequently hypomethylated, leading to probable chromatin activation, whereas normal developmental associated promoters are possibly repressed by hypermethylation. In addition, enrichment of several key oncogenic TF motifs within disease associated enhancers highlights the active ongoing oncogenic transcriptional programs.

Integrative analysis of DNAm and Gene expression among epigenetic subgroups

To understand the influence of DNAm on gene expression in T-ALL clusters, we generated the expression profiles by means of RNA-sequencing for a total of 48 samples ($C_1 = 4$, $C_2 = 13$, $C_3 = 6$, $C_4 = 7$, $C_5 = 14$) including 4 normal total thymus control. In the PCA analysis of the RNA-seq data, although the samples tended to cluster according to their methylation cluster, the distinction between the different leukemia groups appeared much less clear than based on their methylation signature. Interestingly, total thymus samples were clustered with C_2 samples similar to DNAm results. **(Figure 4A)**. In addition, clusters showed clear expression of their candidate driver transcription factors **(Figure 4B)**. We performed differential gene expression (DGE) analysis by comparing each cluster against total thymus samples. This analysis revealed a varying

degrees of gene expression changes with C_1 displaying the lowest and C_5 the highest number of DE genes (DEGs) (**Supplementary Table S8**). Significant portion of these DE genes were up-regulated suggesting an activated gene expression profiles **(Figure 4C)**. In order to measure the genome-wide influence of DNAm on gene expression, we performed correlation analysis between gene expression and the corresponding promoter DNAm levels of all protein coding genes (N = 15,912 genes) across all 44 T-ALL samples **(Figure 4D)**. Only a fraction of genes (N = 235) significantly correlated with their promoter DNAm levels (FDR < 0.1). The fact that DNA methylation levels at promoter regions do not necessarily correlate with gene expression levels has also been observed in multiple studies(*38, 39*). However, except for C_1 cluster, we globally found an expected inverse correlation between DE genes and their promoter DNAm levels - with up-regulated genes displaying lower methylation levels compared to down regulated genes **(Figure 4E)**. Therefore, in order to better depict DNAm clusters, we identified cluster-specific genes with significant inverse correlation between DNAm and gene expression in a robust manner. We used all the DMPs (from **Figure 3A**) within promoter regions and studied their corresponding gene expression **(Figure 4F)**. Careful cataloging of these genes revealed several interesting cluster-specific genes, already involved in oncogenic processes, with an inverse correlation between DNAm and gene expression. In C_1 cluster - enriched in immature, ETP-ALL, co-occurring DNMT3A/IDH2 mutations and myeloid features - among genes hypomethylated and overexpressed, we identified the myeloid factors, *AZU1*, *CSF3,* the oncogenic genes *EGFL7(40)*, *FES(41), SLC2A5(42)* involved in AML*, S100A6* involved in several solid tumor types(*43-45*) and *JDP2(46)*. Similarly, we also found several other cluster-specific candidate oncogenes such as *CD160(47), CD47(48), FOSL1(49)* in C_2, *MAPK8(50-53)* and *MYEOV(54, 55)* in C_3, and *CAPG, RGS17(56, 57), FAM83A(58), LGALS(59), NCR1(60) and EMP1(61)* in C_5. On the other hand, DNA hypermethylation, a hallmark of cancer, has been extensively involved in tumor suppressor genes (TSG) inactivation. In line with this concept, we found among genes hypermethylated and down regulated several candidate TSGs, especially in cluster C_5, which displayed the highest number of repressed DEGs: *NPTX1(62)* in C_4, *ALS2CL(63), AMPH(64), CMTM8(65), DEPDC7(66), HOOK1(67), MITF(68), MPPED2(69), PCDH9(70), RARRES1(71), RASEF(72), RNF180(73), S100A16(74)* and *SLFN5(75)* in C_5. Despite the limited number of RNAseq data available in this series, our gene expression analysis reinforces the biological relevance of the methylation-based clustering and leads to the identification of potential cluster-specific oncogenes and tumor suppressor genes.

Subgroup C_5 associates with poor outcome in GRAALL-treated patients

CIMP-negative T-ALL patients have been shown to be significantly associated with higher cumulative incidences of relapse compared to CIMP-positive patients, suggesting a prognostic relevance of DNAm profiles in both pediatric and adult T-ALL (*19, 21*). Evaluation of the prognostic relevance of the five DNAm clusters revealed diverse levels of overall survival (OS) and event-free survival (EFS) (P = 0.08 and P = 0.045, respectively) with C_4 and C_1 displaying the most favorable and most unfavorable outcomes respectively (P = 0.002 OS and P = 0.01 EFS) **(Supplementary Figure S8A and S8B)**. Based on genome-wide DNAm levels (**see Figure 1C**), we classified five clusters into three groups with significantly different methylation levels, $C_{(1+2)}$, $C_{(3+4)}$ and $C_{(5)}$ which displayed hypomethylation, intermediate-hypermethylation and high-hypermethylation respectively. In line with previous reports, patients in the hypomethylated $C_{(1+2)}$ subgroup demonstrated shorter OS (5-year OS probability; 50% [95% CI = 35% to 63%] *vs* 71.5%

[95% CI = 58% to 81%]; *P* = 0.031) and EFS (5-year EFS probability; 44% [95% CI = 30% to 57%] vs 69% [95% CI = 55% to 79%]; *P* = 0.015) as compared to the intermediate-hypermethylated $C_{(3+4)}$ subgroup. Interestingly, the high-hypermethylated $C_{(5)}$ subgroup displayed distinct clinical outcome and had significantly poorer survival probability as compared to the $C_{(3+4)}$ subgroup (5-year OS probability; 53% [95% CI = 34% to 68%] *vs* 72% [95% CI = 58% to 81%]; P = 0.037) and EFS (5-year EFS probability; 46% [95% CI = 28% to 62%] vs 69% [95% CI = 55% to 79%]; P = 0.045). Patients in both C_5 and $C_{(1+2)}$ showed similar shorter OS and EFS **(Figure 5A and 5B)**. Importantly, the adverse prognosis associated with $C_{(5)}$ and $C_{(1+2)}$ did not seem to be related to the same clinical parameters. C_5 patients demonstrated poor early response with significantly slower D8 prednisone and D15 bone marrow responses, and higher post-induction Minimal Residual Disease (MRD) levels (prednisone response; 28.1% vs 54.2%, Bone marrow response; 21.9% vs 66.7% and negative post-induction MRD; 38.1% vs 78.3% respectively) unlike to the $C_{(1+2)}$ patients who relapsed despite a good initial response (**Table 2**). In both EFS and OS multivariate analyses, the difference between $C_{(1+2)}$ and $C_{(3+4)}$ remained statistically significant as the difference between $C_{(5)}$ and $C_{(3+4)}$ lost this significance **(Supplementary Table S9)**. Overall, these data identified a subset of hypermethylated T-ALL with poor primary therapeutic response, paving the way for specific targeted therapeutic schedules.

In vivo responses to hypomethylating agents

As we observed significant differences in the DNAm patterns and overall levels in the leukemic subgroups, we reasoned that epigenetic treatment with DNMT inhibitors (hypomethylating agents) could be of potential interest in the treatment of, particularly hypermethylated T-ALL. The *TLX1*+ ALL-SIL cell line (belonging to C_4) was transduced to stably express luciferase and transplanted in NSG mice. At day 2 post-transplantation, mice (*n* = 3-5 per group) received a daily dose of 5mg/kg intra-peritoneal 5-azacytidine for two courses of five days with an interval of two days, with bioluminescence measurement of bulk leukemic engraftment. This preventive 5-azacytidine *in vivo* treatment significantly delayed leukemic engraftment and increased the overall survival of the treated mice (*P* = 0.0067 respectively) **(Figure 6A, B)**. In order to test all cluster subtypes using more relevant models of human T-ALL, we established Patients Derived Xenografts (PDX) from five primary T-ALL samples **(Supplementary Figure S9)**. Mice were treated *in-vivo* in a "curative-like" setting (two courses, starting when peripheral blood blast counts exceeded 1%). Leukemic blast expansion was regularly measured by flow cytometry analysis. For the PDX lines derived from three patients UPNT-M149, UPNT-670 and UPNT-529, belonging respectively to the hypermethylated C_3, C_4, and C_5 clusters, 5-azacytidine treatment significantly increased survival, delayed tumor expansion (*P* = 0.001, *P* = 0.01 and *P* = 0.02 respectively) and led to drastic, albeit transient, decreases in leukemic burdens **(Figure 6C, 6D and 6E)**. In contrast, for the two hypomethylated C_2 patients UPNT-525 and UPNT-894 **(Figure 6F, 6G)**, curative treatment had no significant effect and in contrast was associated with a rapid increase of peripheral blast counts and death of the recipient mice. Of note, we observed increase in the tumor burden upon termination of the treatment which might be due to the short duration of the treatment (2 weeks) being insufficient in the complete clearance of blast cells thereby resulting in the recurrence of residual clones surviving the treatment.

Altogether, these results, although obtained in a limited number of mice, indicate the efficacy of DNMT inhibitor treatment in T-ALLs with a hypermethylated profile, including those belonging to the poor prognosis C_5 cluster. Aberrant hypermethylation could be a predictive factor

for therapeutic response and the use of hypomethylating agents for a subset of T-ALLs with unfavorable prognosis.

Discussion

Historically, epigenetic studies in T-ALL have focused on promoter regions (*18, 19, 76, 77*). We now provide extensive DNAm analysis in a large cohort of 143 primary adult T-ALL samples using Illumina Infinium Methylation EPIC arrays covering over 850,000 CpGs across promoters and regulatory regions. Using an unbiased approach, we identified five epigenetic clusters, four of which are associated with activation of distinct oncogenic TFs identified by expression-based studies (*5, 25*). In addition, we identified a novel and likely adult-specific subgroup of T-ALL with co-occurring *DNMT3A* and/or *IDH2* mutations (*78*), which showed hypomethylation, and an immature/ETP phenotype. Patients in this group were associated with poor prognosis. These results highlight the role of developmental stage specific oncogenic events, which remodel the underlying methylome to drive leukemogenesis. Importantly, features associated with these DNAm clusters complement the previously reported gene expression-based analyses (*25*). For example, expression-based studies have shown extensive overlap in expression signatures between *TLX1* and *TLX3* T-ALLs, whereas these two oncogenic subgroups are robustly separated in their DNAm patterns and methylation-based arrest stages, suggesting more profound differences in deregulated pathways. *HOXA9* deregulated samples showed large differences in their DNAm based on the mode of overexpression. Samples with *HOXA9* deregulation clustered in two distinct subgroups, those displaying deregulation in *cis* (under the influence of the TCRβ enhancer) clustered with *TLX1* deregulated samples (C_4) displaying early cortical maturation arrest stages whereas, *HOXA9* samples in *trans* (under the influence of *SET-NUP214*, *MLLT10 or MLL* fusions) formed a distinct hypermethylation cluster (C_5), associated with a more immature maturation arrest stage (*22*). C_5 was also enriched in PRC2 alterations (*SUZ12* and *EZH2* mutation/deletion). Murine models of ETP-ALL driven by Ezh2 and Tp53 inactivation have also been characterized by hypermethylation of T-cell developmental genes, thereby promoting leukemic transformation (*79*). The loss of Ezh2 in murine ETPs induced the pathogenic epigenetic switch from H3K27me3 to DNA hypermethylation of pivotal T-cell development regulators, such as *Runx1*, but sustained the expression of stem cell signature genes and promoted the transformation of ETPs, supporting a pathogenic role for inactivating PRC2 mutations in ETP-ALL. It would be interesting to compare these observations with human primary T-ALLs. DNAm clearly separated immature/ETP ALLs into C_1 hypomethylated *DNMT3A/IDH2*-rich and C_2 hypermethylated *trans*-HOXA/PRC2 groups, paving the way for distinct therapy. *TAL1* deregulated samples, irrespective of the mechanism of deregulation (*SIL-TAL1*, or upstream neo-enhancers (*80*)) clustered together, suggesting a disruption of common downstream pathways (*81, 82*). *TAL1* is one of the most frequently deregulated driver oncogene in T-ALL (approximately 30% of cases) either through deletions (*SIL-TAL1*; 10-20%), oncogenic neoenhancer (20%) (*80*) or rare V(D)J-mediated translocations [t(1;14); 1-2%]. In about 40% of *TAL1*+ cases, however, such lesions are absent, and *TAL1* is overexpressed due to unknown mechanisms. In cluster C_2, only 16/34 samples displayed either the *SIL-TAL1* deletion or oncogenic neoenhancer suggesting that this cluster may also contain samples with *TAL1* overexpression due to as yet unexplained/unidentified mechanisms of deregulation.

Cluster C_1, enriched for co-occurring *DNMT3A/IDH2* mutants, was completely devoid of any previously recognized TF overexpression, but all *DNMT3A/IDH2* mutated cases also harbored *NOTCH1* mutations, in line with previous observations, highlighting the synergistic role of deregulated epigenome and *NOTCH1* signaling pathways in disease progression (*83, 84*). High variant allele frequency of *DNMT3A* mutations indicated a large fraction of homozygous or

compound heterozygous mutations (data not shown) suggesting a distinct *DNMT3A* mutation spectrum in T-ALL as compared to acute myeloid leukemia cases (*83-86*). In the context of disease progression, higher variant allele frequency combined with significantly older age group **(**data not shown**)** suggests that *DNMT3A* clones could have arisen early in hematopoiesis, prior to T-cell lineage restriction. This hypothesis is further corroborated by recent evidence for age related *DNMT3A* mutations in T-ALL (*11*). This early acquisition of *DNMT3A* in T-cell progenitors further could acquire secondary genetic hits (such as *NOTCH1/IDH2*) upon thymic entry, thereby deregulating the methylome during T-cell development. Interestingly, unlike co-occurring *IDH2/DNMT3A* mutations, single mutations in epigenetic factors (*DNMT3A, IDH1, TET2/3*) involved in DNAm were not associated with a specific methylome, raising the question of the role played by such genomic alterations in leukemogenesis. Cluster-specific methylome analysis revealed non-random and systematic distribution of DMPs, with significant hypermethylation of T-cell developmental associated active promoters and poised enhancers, whereas T-ALL associated enhancers showed significant hypomethylation. Our results echo similar observations of enhancer hypomethylation in solid tumors as well as in leukemia (*30, 31*). Moreover, identification of disease-associated TFs within the enhancer elements suggests an enhancer dependent active oncogenic transcriptional program mediated by hypomethylated chromatin (*34, 36*). Similar to cluster specific somatic and epigenetic alterations, our integrated gene expression and DNAm analysis further highlights the combined effects of DNAm and gene expression in T-ALL pathogenesis. For example, *SLC2A5* which is hypomethylated as well as overexpressed in C_1, is also overexpressed in a subset of AML and in childhood Philadelphia chromosome positive ALL and associated with poor outcome. The encoded protein which is a fructose transporter responsible for fructose uptake by the small intestine and enhances fructose use of leukemic cells (*42, 87*). *JDP2* – a bZIP family transcription factor specific to C_1 - has been identified as a novel oncogene in ETP T-ALL and shown to be associated with poor prognosis (*46*). Importantly, enrichment of myeloid like genes and bZIP family transcription factors in C_1 is also corroborated by independent motif analysis results **(Figure 3F).** Another gene - *EMP1* specifically hypomethylated/overexpressed in C_5, has been shown to be associated with poor outcome in pediatric ALL and to confer prednisolone resistance (*61*). Interestingly, our clinical study showed also an association of C_5 with a poor early prednisone response.

Another significant finding of this study is the recapitulation of the maturation arrest stages based on DNAm. Lymphopoiesis is a complex process mediated by expression of several TFs and cell surface markers in a stage specific manner (*29*). Our DNAm based phylogenetic trees robustly captured these developmental stages, thereby illustrating the prominent role of DNAm in regulating thymopoiesis. These results further highlight previous observations in mouse hematopoiesis, wherein DNAm dynamics orchestrate myelopoiesis and lymphopoiesis to form different terminal DNAm patterns (*88*). Similar observations have also been made in human lymphopoiesis, in which DNAm dynamics during B cell maturation display a continuum of changes (*89*). By combining T-ALL samples, these phylogenetic trees robustly recaptured the known clinically relevant subgroups, thereby reflecting developmental arrest at different stages of thymic maturation. Similar results have also been observed in an independent analysis using open chromatin signals highlighting the conserved epigenetic marks at several layers (*90*). Overall, our results show an early occurrence of C_5, and C_1 clusters, characterized by in *trans* overexpression of HOXA and mutations in myeloid-like genes such as *DNMT3A/IDH2* respectively, whereas C_3 (enriched in *TLX3* cases) and C_4 (enriched in *TLX1* and cis HOXA deregulated cases) showed

intermediate arrest stages. Finally, C_2 enriched in *TAL1* deregulated samples occurred latest, in keeping with the known late cortical maturation stage arrest of the *TAL1* deregulated T-ALL subgroup (*5*). The combined observation of hypomethylated disease enhancers and maturation arrest stages further highlight the plasticity induced by DNAm in tumor progression.

Importantly, unlike previous transcriptional studies, we identified patient clusters with clinical relevance and an unexpected subgroup of highly hypermethylated patients (C_5) associated with unfavorable outcome similar to hypomethylated patients and who could benefit from targeted therapy (*91, 92*). Epigenetic drugs such as (*93*)(*94*)(*95*). Our results from use of hypomethylating treatment of T-ALLs with known methylation status suggest that 5-azacytine could be a promising therapeutic option in hypermethylated T-ALL. In contrast, hypomethylated samples showed poor response to treatment, suggesting that a certain degree of aberrant hypermethylation is required to observe treatment efficacy. Follow-up studies are required to confirm the therapeutic potential of hypomethylating agents in T-ALL and it would be interesting to study their efficacy in combination with other compounds such as conventional chemotherapy or epigenetic therapy, such as an *EZH2* inhibitor.

Materials and Methods

Study Design: Blood or bone marrow from a total of 143 adult patients diagnosed with T-ALL (15-60 years old) from two successive French ALL cooperative groups (n = 22 from GRAALL-2003 and n = 121 from GRAALL-2005) were analyzed. The GRAALL-2003 protocol was a multicenter Phase II trial, which enrolled 76 adults with T-ALL between November 2003 and November 2005, of whom 50 had sufficient diagnostic tumor material available for NGS study but only 22 had enough high-quality DNA for EPIC array (*96*). The multicenter randomized GRAALL-2005 Phase III trial was very similar to the GRAALL-2003 trial, with the addition of a randomized evaluation of an intensified sequence of hyper-fractionated cyclophosphamide during induction and late intensification (*97*). Between May 2006 and May 2010, 261 adults with T-ALL were randomized in the GRAALL-2005, of which 185 had available diagnostic material for NGS study but only 121 had enough high-quality DNA for EPIC array. In this study, the main criterion for inclusion was the availability of high-quality DNA. Survival outcomes of the 143 patients did not differ from those of the remaining 194 T-ALL patients. As expected in retrospective studies, initial white blood cell count (WBC) was higher in the study cohort. A full comparison of the clinical features of each group is shown in **Supplementary Table S10**. All samples contained >80% blasts. Phenotypic and oncogenetic characteristics were obtained as described (*22, 98, 99*). Informed consent was obtained from all patients at enrolment. All trials were conducted in accordance with the declaration of Helsinki, approved by local and multicenter research ethical committees. The GRAALL-2003 and -2005 studies were registered at http://www.clinicaltrials.gov as #NCT00222027 and #NCT00327678, respectively.

Univariate and multivariate survival analysis for OS and EFS: Since methylation subgroups were strongly associated with maturation arrest and tumor biology, we only considered age, log(WBC), CNS involvement, prednisone response, and D8 bone marrow response as covariates to avoid multicollinearity. Methylation clusters were reduced to a 3-class variable with intermediate methylation clusters ($C_{(3+4)}$) considered as baseline. Covariates finally used in the multivariate cox model were those associated with outcome in univariate analyses ($P < 0.1$). As shown in **Supplementary Table S9**, univariate analysis for EFS identified age, prednisone response, BM response, CNS involvement, and methylation clustering as covariates in the multivariate model.

Univariate analysis for OS identified age, prednisone response, CNS involvement, and methylation clustering as covariates for multivariate analysis.

Thymic subpopulations sorting: Human thymic subpopulations were obtained from children undergoing cardiac surgery. Informed consent was obtained from parents and the study was approved by the Ethics committee (CPP Ile de France II, project number 2012-03-04). Thymii were dissociated into single cell suspensions. Six specific subpopulations were purified by sorting using FACS ARIA III (BD Biosciences): thymic CD34+, 4ISP (Immature Simple positive CD4+), DP-TCR-/CD3- (Double Positive CD4+ CD8+ TCR- CD3-), DP-TCR+/CD3+ (Double Positive CD4+ CD8+ TCR+ CD3+), SP4 (Single positive CD4+) and SP8 (Single Positive CD8+). For the immature subpopulation, thymocytes were pre-purified by depletion of CD3 and CD8 positive fraction by magnetic activated cell sorting using CD3 MicroBeads (Miltenyi, Bergisch Gladbach, Germany, Cat #130-050-101) and CD8 MicroBeads (Miltenyi, Cat#130-045-201) respectively. Antibodies used for cell labelling of CD3-/CD8- thymic fraction: CD1a FITC, clone NA1/34 (Dako, Glostrup, Denmark Cat#F7141), CD34 APC (BD Biosciences, Erembodegem, Belgium Cat#345804), CD8 PC7 (Beckman Coulter, Brea, CA, USA, Cat#737661), CD3 Alexa-700 (BD Biosciences, Cat#557943), CD4 V450, clone RPA-T4 (BD Biosciences, Cat#60345), CD45 V500, clone HI30 (BD Biosciences, Cat#560777). Final purity after sorting was over 95%. Total thymus DNA was used as control.

Targeted Next-generation sequencing: Nextera XT (Illumina, San Diego, CA, USA) DNA Libraries were prepared according to the manufacturer's instructions and sequenced using the Illumina MiSeq sequencing system. The custom NGS panel comprised genes coding for factors involved in molecular pathways known to be mutated in T-ALL, namely cytokine receptor and RAS signaling (*NRAS, KRAS, JAK1, JAK3, STAT3, STAT5B, IL7R, BRAF, NF1, SH2B3* and *PTPN11*), hematopoietic development (*RUNX1, ETV6, GATA3, IKZF1, EP300)*, chemical modification of histones (*SUZ12, EED, EZH2, KMT2A, KMT2D and SETD2*) and DNAm (*DNMT3A, IDH1, IDH2, TET2, TET3*). This panel was originally inspired by the genes found to be preferentially altered in pediatric ETP-ALL (Zhang et al. 2012) and we have reported a subset of the results described in the current paper in previous clinic-biological and genetic analyses (*11, 13*). Mutational data was later integrated with copy number and analyzed with maftools Bioconductor package.

MLPA T-ALL panel: CNV for a targeted panel of genes were studied using SALSA MLPA P383 T-ALL probemix (MRC-Holland, Amsterdam, Netherland).

PDX and 5-azacytidine treatment of xenografted mice: 5-azacytidine was purchased from Sigma Aldrich (St. Louis, MO, USA) and was prepared in 0.9% NaCl and stored as recommended by the manufacturer. Treatment protocol**:** at Day 0 (D0), $1x10^6$ human leukemic cells were resuspended in 100 µL of PBS buffer and injected retro-orbitary in NSG mice. For ALL-SIL cell line, at D2, a group of mice received as "preventive" treatment 5 mg/kg/day 5-Azacytidine intra-peritoneal (IP) for 5 days twice with 2 days break. Untreated mice received the same volume of NaCl IP. Primary PDX models were treated with the same 5-azacytidine courses but in a "curative-like" setting, starting when peripheral blood blast counts exceeded 1%. Leukemic engraftment in peripheral blood was studied by flow cytometry using mouse-CD45 and human-CD45 antibodies. Mice were sacrificed when the blast fraction was more than 80% or if mice presented clinical signs of disease (loss of weight >10%, neurological signs, tumor development). During the treatment, mice were constantly supervised for weight loss, behavioral disorders such as hyperactivity, physiological disorders such as increased respiratory rate, hair loss, neurological signs which revealed no

obvious treatment induced side effects. Mouse study was approved by national ethic committee: PROJET APAFIS # 8853 N° 2017020814103710.

Data analysis:

EPIC arrays: Intensity Data (IDAT) files from EPIC arrays were processed and normalized with RnBeads Bioconductor package (*100*). Probes with high SNP probability and bad detection P-values were filtered out. EPIC arrays were also used for copy number analysis using conumee Bioconductor package, and further processed with GISTIC2 to identify recurrent copy number aberrations. Probes located in regions with altered copy numbers were excluded from downstream analysis **(Supplementary Figure S1C)**.

Clustering and dimension reduction: To identify epigenetic clusters, we used Non Negative Factorization (NMF) on the top 5% of most variable probes (n = 38,583) (*101*). NMF decomposes a matrix into two smaller matrices whose product sufficiently recomposes the original matrix. Critical step in NMF is identifying the number of factors. We used the method described by Brunet et al where, NMF is run on a range of values and a cophenetic correlation coefficient (measure of goodness of fit) was determined (*102*). An optimal number of clusters was identified for which cophenetic correlation reaches its maximum value **(Supplementary Figure S1A)**. Furthermore, to measure the fitness of identified clusters, we repeated the above clustering procedure for varying number of probes (5000 up to 38583). Clusters were stable and reached maximum cophenetic correlation coefficient at n = 5. Finally, we used 5 clusters generated by using N = 38,583 probes for all downstream analysis. These 5 epigenetic clusters were also robust as measured by rand-index, and samples showed little to no changes in cluster assignments as the number of probes used for clustering increased **(Supplementary Figure S1B)**. Moreover we randomly sampled 80% of the initial cohort (N = 114) and repeated the clustering. Rand Index - a similarity score between two clustering results - was measured between original clusters and new clusters generated on subsamples. We repeated the entire process 10 times and results showed high similarity (Rand index 0.85 – 0.99) to the original clustering (data not shown). Clusters were later visualized using Uniform Manifold Approximation and Projection (UMAP).

Statistical analysis: DMPs (FDR < 0.1 and |methylation change| > 20%) were identified for every cluster by comparing them to 12 normal thymic samples using Limma Bioconductor package (*103*). Differential gene expression analysis was performed using DESeq2 Bioconductor package adjusting for batch covariates wherever required. Survival analysis was performed with Surv R-package. For predicting T-cell developmental trajectory and T-ALL maturation arrest stages, we calculated Manhattan distance between samples with *dist* function in R. Phylogenetic trees were later constructed from distance matrix using neighbor joining method, and visualized in R with ape package. All data analysis, visualization, and test statistics were performed in R statistical environment (version 4.0), and complete source code used for data analysis is provided as **Supplementary Data S1.**

ChIP-seq analysis: BAM files for ChIP-seq datasets were obtained from the BLUEPRINT (*32*). Peak calling was performed using MACS2 with parameters *--bdg --SPMR --nomodel --extsize 200 –q 0.05*. Peaks overlapping with UCSC blacklisted regions were excluded. ChIP-seq results were visualized using deepTools. SEs were identified and annotated for every T-ALL and normal thymic samples with Rose software. Consensus SEs for each T-ALL cluster **(Supplementary Figure S7D)** were generated by merging H3K27ac BAM files from samples belonging to same cluster. Enrichment analysis of DMPs within regulatory regions was performed with LOLA

Bioconductor package. Enrichment of all vertebrate motifs around 100 base-pairs of hypo-methylated CpG probes was done using findMotifsGenome.pl from homer tools. As a background we randomly selected a subset of ca. 850,000 EPIC probes (n = 183,896; corresponding to total number of unique DMPs identified across all 5 clusters). Significant motifs were chosen based on P < 1e-5. Since homer motif database lacked the motifs for *TAL1* TF, we re-analyzed public datasets containing ChIP-seq for TAL1 (GSM935496) thereby defining de-novo motif. This was later included in the homer database for downstream analysis.

Supplementary Materials

Supplementary Figures

- Supplementary Figure-S1. Identification of epigenetic clusters.
- Supplementary Figure-S2: Genomic landscape of adult T-ALL.
- Supplementary Figure-S3: Gene expression analysis of T-ALL clusters
- Supplementary Figure-S4: Differentially Methylated Probes associated with T-cell maturation (tDMPs).
- Supplementary Figure-S5: Characterization of Differentially Methylated Probes associated with T-cell maturation (tDMPs) based on methylation dynamics.
- Supplementary Figure-S6: Gene expression dynamics during Thymopoiesis
- Supplementary Figure-S7: Influence of thymic maturation associated Differentially Methylated Probes (tDMPs)
- Supplementary Figure-S8: Defining normal and disease associated enhancers.
- Supplementary Figure-S9: Overall Survival (OS) and Event Free Survival (EFS) of methylation clusters. A and B.

Supplementary Tables

- **Supplementary Table S1:** Significantly mutated genes in every cluster compared to rest of the clusters (Related to Figure 1H). (XLSX)
- **Supplementary Table S2:** Differentially methylated probes identified in each of the normal thymic cell-types (Related to Figure-2). (XLSX)
- **Supplementary Table S3:** Differentially methylated probes identified in each of the normal thymic cell-types classified into 8 subgroups based on methylation dynamics (Related to Figure-2D, E, and Supplementary Figure S4). (XLSX)
- **Supplementary Table S4:** Genes which are both differentially expressed and showing a differential promoter methylation in normal thymic cell types (Related to Supplementary Figure S5B). (XLSX)
- **Supplementary Table S5:** Differentially methylated probes found in each cluster along with the genomic annotations (Related to Figure 3A). (XLSX)

- **Supplementary Table S6:** Super and Typical Enhancers identified in normal thymic cell-types, and in T-ALL clusters (Related to Figure 3D, and Supplementary Figure S7D). (XLSX)
- **Supplementary Table S7:** Significantly enriched TF motifs detected within 100 base pair vicinity of hypo-methylated probes located within the super or typical enhancer regions from T-ALL clusters (Related to Figure 3F). (XLSX)
- **Supplementary Table S8:** Differentially expressed genes in cluster compared to total thymus (Related to Figure 4C). (XLSX)
- **Supplementary Table S9:** Univariate and Multivariate analyses of EFS and Overall Survival.
- **Supplementary Table S10:** Clinical characteristics and outcome of the study cohort versus non-investigated patients (GRAALL-2003/05 trials).

Supplementary Data S1: Complete source used for data analysis.

References:

1. S. P. Hunger, C. G. Mullighan, Acute Lymphoblastic Leukemia in Children. *N Engl J Med* **373**, 1541-1552 (2015).
2. M. R. Litzow, A. A. Ferrando, How I treat T-cell acute lymphoblastic leukemia in adults. *Blood* **126**, 833-841 (2015).
3. R. Sutton, P. J. Shaw, N. C. Venn, T. Law, A. Dissanayake, T. Kilo, M. Haber, M. D. Norris, C. Fraser, F. Alvaro, T. Revesz, T. N. Trahair, L. Dalla-Pozza, G. M. Marshall, T. A. O'Brien, Persistent MRD before and after allogeneic BMT predicts relapse in children with acute lymphoblastic leukaemia. *Br J Haematol* **168**, 395-404 (2015).
4. L. Belver, A. Ferrando, The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat Rev Cancer* **16**, 494-507 (2016).
5. A. A. Ferrando, D. S. Neuberg, J. Staunton, M. L. Loh, C. Huard, S. C. Raimondi, F. G. Behm, C. H. Pui, J. R. Downing, D. G. Gilliland, E. S. Lander, T. R. Golub, A. T. Look, Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75-87 (2002).
6. A. P. Weng, A. A. Ferrando, W. Lee, J. P. t. Morris, L. B. Silverman, C. Sanchez-Irizarry, S. C. Blacklow, A. T. Look, J. C. Aster, Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269-271 (2004).
7. H. Shen, P. W. Laird, Interplay between the cancer genome and epigenome. *Cell* **153**, 38-55 (2013).
8. R. Huether, L. Dong, X. Chen, G. Wu, M. Parker, L. Wei, J. Ma, M. N. Edmonson, E. K. Hedlund, M. C. Rusch, S. A. Shurtleff, H. L. Mulder, K. Boggs, B. Vadordaria, J. Cheng, D. Yergeau, G. Song, J. Becksfort, G. Lemmon, C. Weber, Z. Cai, J. Dang, M. Walsh, A. L. Gedman, Z. Faber, J. Easton, T. Gruber, R. W. Kriwacki, J. F. Partridge, L. Ding, R. K. Wilson, E. R. Mardis, C. G. Mullighan, R. J. Gilbertson, S. J. Baker, G. Zambetti, D. W. Ellison, J. Zhang, J. R. Downing, The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun* **5**, 3630 (2014).
9. S. Peirs, J. Van der Meulen, I. Van de Walle, T. Taghon, F. Speleman, B. Poppe, P. Van Vlierberghe, Epigenetics in T-cell acute lymphoblastic leukemia. *Immunol Rev* **263**, 50-67 (2015).
10. P. Van Vlierberghe, A. Ambesi-Impiombato, K. De Keersmaecker, M. Hadler, E. Paietta, M. S. Tallman, J. M. Rowe, C. Forne, M. Rue, A. A. Ferrando, Prognostic relevance of integrated genetic profiling in adult T-cell acute lymphoblastic leukemia. *Blood* **122**, 74-82 (2013).

11. J. Bond, A. Touzart, S. Lepretre, C. Graux, M. Bargetzi, L. Lhermitte, G. Hypolite, T. Leguay, Y. Hicheri, G. Guillerm, K. Bilger, V. Lheritier, M. Hunault, F. Huguet, Y. Chalandon, N. Ifrah, E. Macintyre, H. Dombret, V. Asnafi, N. Boissel, DNMT3A mutation is associated with increased age and adverse outcome in adult T-cell acute lymphoblastic leukemia. *Haematologica* **104**, 1617-1625 (2019).
12. Z. Kalender Atak, K. De Keersmaecker, V. Gianfelici, E. Geerdens, R. Vandepoel, D. Pauwels, M. Porcu, I. Lahortiga, V. Brys, W. G. Dirks, H. Quentmeier, J. Cloos, H. Cuppens, A. Uyttebroeck, P. Vandenberghe, J. Cools, S. Aerts, High accuracy mutation detection in leukemia on a selected panel of cancer genes. *PLoS One* **7**, e38463 (2012).
13. J. Zhang, L. Ding, L. Holmfeldt, G. Wu, S. L. Heatley, D. Payne-Turner, J. Easton, X. Chen, J. Wang, M. Rusch, C. Lu, S. C. Chen, L. Wei, J. R. Collins-Underwood, J. Ma, K. G. Roberts, S. B. Pounds, A. Ulyanov, J. Becksfort, P. Gupta, R. Huether, R. W. Kriwacki, M. Parker, D. J. McGoldrick, D. Zhao, D. Alford, S. Espy, K. C. Bobba, G. Song, D. Pei, C. Cheng, S. Roberts, M. I. Barbato, D. Campana, E. Coustan-Smith, S. A. Shurtleff, S. C. Raimondi, M. Kleppe, J. Cools, K. A. Shimano, M. L. Hermiston, S. Doulatov, K. Eppert, E. Laurenti, F. Notta, J. E. Dick, G. Basso, S. P. Hunger, M. L. Loh, M. Devidas, B. Wood, S. Winter, K. P. Dunsmore, R. S. Fulton, L. L. Fulton, X. Hong, C. C. Harris, D. J. Dooling, K. Ochoa, K. J. Johnson, J. C. Obenauer, W. E. Evans, C. H. Pui, C. W. Naeve, T. J. Ley, E. R. Mardis, R. K. Wilson, J. R. Downing, C. G. Mullighan, The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157-163 (2012).
14. J. Feng, Y. Li, Y. Jia, Q. Fang, X. Gong, X. Dong, K. Ru, Q. Li, X. Zhao, K. Liu, M. Wang, Z. Tian, Y. Jia, Y. Wang, D. Lin, H. Wei, K. Tang, Y. Mi, J. Wang, Spectrum of somatic mutations detected by targeted next-generation sequencing and their prognostic significance in adult patients with acute lymphoblastic leukemia. *J Hematol Oncol* **10**, 61 (2017).
15. D. Capper, D. T. W. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, C. Koelsche, F. Sahm, L. Chavez, D. E. Reuss, A. Kratz, A. K. Wefers, K. Huang, K. W. Pajtler, L. Schweizer, D. Stichel, A. Olar, N. W. Engel, K. Lindenberg, P. N. Harter, A. K. Braczynski, K. H. Plate, H. Dohmen, B. K. Garvalov, R. Coras, A. Holsken, E. Hewer, M. Bewerunge-Hudler, M. Schick, R. Fischer, R. Beschorner, J. Schittenhelm, O. Staszewski, K. Wani, P. Varlet, M. Pages, P. Temming, D. Lohmann, F. Selt, H. Witt, T. Milde, O. Witt, E. Aronica, F. Giangaspero, E. Rushing, W. Scheurlen, C. Geisenberger, F. J. Rodriguez, A. Becker, M. Preusser, C. Haberler, R. Bjerkvig, J. Cryan, M. Farrell, M. Deckert, J. Hench, S. Frank, J. Serrano, K. Kannan, A. Tsirigos, W. Bruck, S. Hofer, S. Brehmer, M. Seiz-Rosenhagen, D. Hanggi, V. Hans, S. Rozsnoki, J. R. Hansford, P. Kohlhof, B. W. Kristensen, M. Lechner, B. Lopes, C. Mawrin, R. Ketter, A. Kulozik, Z. Khatib, F. Heppner, A. Koch, A. Jouvet, C. Keohane, H. Muhleisen, W. Mueller, U. Pohl, M. Prinz, A. Benner, M. Zapatka, N. G. Gottardo, P. H. Driever, C. M. Kramm, H. L. Muller, S. Rutkowski, K. von Hoff, M. C. Fruhwald, A. Gnekow, G. Fleischhack, S. Tippelt, G. Calaminus, C. M. Monoranu, A. Perry, C. Jones, T. S. Jacques, B. Radlwimmer, M. Gessi, T. Pietsch, J. Schramm, G. Schackert, M. Westphal, G. Reifenberger, P. Wesseling, M. Weller, V. P. Collins, I. Blumcke, M. Bendszus, J. Debus, A. Huang, N. Jabado, P. A. Northcott, W. Paulus, A. Gajjar, G. W. Robinson, M. D. Taylor, Z. Jaunmuktane, M. Ryzhova, M. Platten, A. Unterberg, W. Wick, M. A. Karajannis, M. Mittelbronn, T. Acker, C. Hartmann, K. Aldape, U. Schuller, R. Buslei, P. Lichter, M. Kool, C. Herold-Mende, D. W. Ellison, M. Hasselblatt, M. Snuderl, S. Brandner, A. Korshunov, A. von Deimling, S. M. Pfister, DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469-474 (2018).
16. X. Hao, H. Luo, M. Krawczyk, W. Wei, W. Wang, J. Wang, K. Flagg, J. Hou, H. Zhang, S. Yi, M. Jafari, D. Lin, C. Chung, B. A. Caughey, G. Li, D. Dhar, W. Shi, L. Zheng, R. Hou, J. Zhu, L. Zhao, X. Fu, E. Zhang, C. Zhang, J. K. Zhu, M. Karin, R. H. Xu, K. Zhang, DNA methylation markers for diagnosis and prognosis of common cancers. *Proc Natl Acad Sci U S A* **114**, 7414-7419 (2017).
17. M. Borssen, J. Nordlund, Z. Haider, M. Landfors, P. Larsson, J. Kanerva, K. Schmiegelow, T. Flaegstad, O. G. Jonsson, B. M. Frost, J. Palle, E. Forestier, M. Heyman, M. Hultdin, G. Lonnerholm, S. Degerman, DNA methylation holds prognostic information in relapsed precursor B-cell acute lymphoblastic leukemia. *Clin Epigenetics* **10**, 31 (2018).
18. M. Borssen, L. Palmqvist, K. Karrman, J. Abrahamsson, M. Behrendtz, J. Heldrup, E. Forestier, G. Roos, S. Degerman, Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One* **8**, e65373 (2013).
19. M. Borssen, Z. Haider, M. Landfors, U. Noren-Nystrom, K. Schmiegelow, A. E. Asberg, J. Kanerva, H. O. Madsen, H. Marquart, M. Heyman, M. Hultdin, G. Roos, E. Forestier, S. Degerman, DNA Methylation

Adds Prognostic Value to Minimal Residual Disease Status in Pediatric T-Cell Acute Lymphoblastic Leukemia. *Pediatr Blood Cancer* **63**, 1185-1192 (2016).
20. Z. Haider, P. Larsson, M. Landfors, L. Kohn, K. Schmiegelow, T. Flaegstad, J. Kanerva, M. Heyman, M. Hultdin, S. Degerman, An integrated transcriptome analysis in T-cell acute lymphoblastic leukemia links DNA methylation subgroups to dysregulated TAL1 and ANTP homeobox gene expression. *Cancer Med* **8**, 311-324 (2019).
21. A. Touzart, E. Lengline, M. Latiri, M. Belhocine, C. Smith, X. Thomas, S. Spicuglia, D. Puthier, F. Pflumio, T. Leguay, C. Graux, Y. Chalandon, F. Huguet, S. Lepretre, N. Ifrah, H. Dombret, E. Macintyre, M. Hunault, N. Boissel, V. Asnafi, Epigenetic Silencing Affects l-Asparaginase Sensitivity and Predicts Outcome in T-ALL. *Clin Cancer Res* **25**, 2483-2493 (2019).
22. J. Bond, T. Marchand, A. Touzart, A. Cieslak, A. Trinquand, L. Sutton, I. Radford-Weiss, L. Lhermitte, S. Spicuglia, H. Dombret, E. Macintyre, N. Ifrah, J. F. Hamel, V. Asnafi, An early thymic precursor phenotype predicts outcome exclusively in HOXA-overexpressing adult T-cell acute lymphoblastic leukemia: a Group for Research in Adult Acute Lymphoblastic Leukemia study. *Haematologica* **101**, 732-740 (2016).
23. M. Khan, J. Cortes, T. Kadia, K. Naqvi, M. Brandt, S. Pierce, K. P. Patel, G. Borthakur, F. Ravandi, M. Konopleva, S. Kornblau, H. Kantarjian, K. Bhalla, C. D. DiNardo, Clinical Outcomes and Co-Occurring Mutations in Patients with RUNX1-Mutated Acute Myeloid Leukemia. *Int J Mol Sci* **18**, (2017).
24. C. L. Green, C. M. Evans, L. Zhao, R. K. Hills, A. K. Burnett, D. C. Linch, R. E. Gale, The prognostic significance of IDH2 mutations in AML depends on the location of the mutation. *Blood* **118**, 409-412 (2011).
25. Y. Liu, J. Easton, Y. Shao, J. Maciaszek, Z. Wang, M. R. Wilkinson, K. McCastlain, M. Edmonson, S. B. Pounds, L. Shi, X. Zhou, X. Ma, E. Sioson, Y. Li, M. Rusch, P. Gupta, D. Pei, C. Cheng, M. A. Smith, J. G. Auvil, D. S. Gerhard, M. V. Relling, N. J. Winick, A. J. Carroll, N. A. Heerema, E. Raetz, M. Devidas, C. L. Willman, R. C. Harvey, W. L. Carroll, K. P. Dunsmore, S. S. Winter, B. L. Wood, B. P. Sorrentino, J. R. Downing, M. L. Loh, S. P. Hunger, J. Zhang, C. G. Mullighan, The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet* **49**, 1211-1218 (2017).
26. A. M. Martelli, F. Paganelli, A. Fazio, C. Bazzichetto, F. Conciatori, J. A. McCubrey, The Key Roles of PTEN in T-Cell Acute Lymphoblastic Leukemia Development, Progression, and Therapeutic Response. *Cancers (Basel)* **11**, (2019).
27. J. Roels, A. Kuchmiy, M. De Decker, S. Strubbe, M. Lavaert, K. L. Liang, G. Leclercq, B. Vandekerckhove, F. Van Nieuwerburgh, P. Van Vlierberghe, T. Taghon, Distinct and temporary-restricted epigenetic mechanisms regulate human alphabeta and gammadelta T cell development. *Nat Immunol* **21**, 1280-1292 (2020).
28. J. Le, J. E. Park, V. L. Ha, A. Luong, S. Branciamore, A. S. Rodin, G. Gogoshin, F. Li, Y. E. Loh, V. Camacho, S. B. Patel, R. S. Welner, C. Parekh, Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage Specification Trajectories and a Commitment Spectrum in T Cell Development. *Immunity* **52**, 1105-1118 e1109 (2020).
29. U. Koch, F. Radtke, Mechanisms of T cell development and transformation. *Annu Rev Cell Dev Biol* **27**, 539-562 (2011).
30. R. E. Bell, T. Golan, D. Sheinboim, H. Malcov, D. Amar, A. Salamon, T. Liron, S. Gelfman, Y. Gabet, R. Shamir, C. Levy, Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. *Genome Res* **26**, 601-611 (2016).
31. L. Benetatos, G. Vartholomatos, Enhancer DNA methylation in acute myeloid leukemia and myelodysplastic syndromes. *Cell Mol Life Sci* **75**, 1999-2009 (2018).
32. D. Adams, L. Altucci, S. E. Antonarakis, J. Ballesteros, S. Beck, A. Bird, C. Bock, B. Boehm, E. Campo, A. Caricasole, F. Dahl, E. T. Dermitzakis, T. Enver, M. Esteller, X. Estivill, A. Ferguson-Smith, J. Fitzgibbon, P. Flicek, C. Giehl, T. Graf, F. Grosveld, R. Guigo, I. Gut, K. Helin, J. Jarvius, R. Kuppers, H. Lehrach, T. Lengauer, A. Lernmark, D. Leslie, M. Loeffler, E. Macintyre, A. Mai, J. H. Martens, S. Minucci, W. H. Ouwehand, P. G. Pelicci, H. Pendeville, B. Porse, V. Rakyan, W. Reik, M. Schrappe, D. Schubeler, M. Seifert, R. Siebert, D. Simmons, N. Soranzo, S. Spicuglia, M. Stratton, H. G. Stunnenberg, A. Tanay, D. Torrents, A. Valencia, E. Vellenga, M. Vingron, J. Walter, S. Willcocks, BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* **30**, 224-226 (2012).
33. A. Rada-Iglesias, R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn, J. Wysocka, A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283 (2011).

34. H. Heyn, E. Vidal, H. J. Ferreira, M. Vizoso, S. Sayols, A. Gomez, S. Moran, R. Boque-Sastre, S. Guil, A. Martinez-Cardus, C. Y. Lin, R. Royo, J. V. Sanchez-Mut, R. Martinez, M. Gut, D. Torrents, M. Orozco, I. Gut, R. A. Young, M. Esteller, Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biol* **17**, 11 (2016).
35. M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K. K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, M. Snyder, Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100 (2012).
36. V. Saint-Andre, A. J. Federation, C. Y. Lin, B. J. Abraham, J. Reddy, T. I. Lee, J. E. Bradner, R. A. Young, Models of human core transcriptional regulatory circuitries. *Genome Res* **26**, 385-396 (2016).
37. T. Sanda, L. N. Lawton, M. I. Barrasa, Z. P. Fan, H. Kohlhammer, A. Gutierrez, W. Ma, J. Tatarek, Y. Ahn, M. A. Kelliher, C. H. Jamieson, L. M. Staudt, R. A. Young, A. T. Look, Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209-221 (2012).
38. R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
39. G. A. Challen, D. Sun, M. Jeong, M. Luo, J. Jelinek, J. S. Berg, C. Bock, A. Vasanthakumar, H. Gu, Y. Xi, S. Liang, Y. Lu, G. J. Darlington, A. Meissner, J. P. Issa, L. A. Godley, W. Li, M. A. Goodell, Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet* **44**, 23-31 (2011).
40. D. Papaioannou, C. Shen, D. Nicolet, B. McNeil, M. Bill, M. Karunasiri, M. H. Burke, H. G. Ozer, S. A. Yilmaz, N. Zitzer, G. K. Behbehani, C. C. Oakes, D. J. Steiner, G. Marcucci, B. L. Powell, J. E. Kolitz, T. H. Carter, E. S. Wang, K. Mrozek, C. M. Croce, M. A. Caligiuri, C. D. Bloomfield, R. Garzon, A. M. Dorrance, Prognostic and biological significance of the proangiogenic factor EGFL7 in acute myeloid leukemia. *Proc Natl Acad Sci U S A* **114**, E4641-E4647 (2017).
41. Y. L. Zhang, J. H. Ren, X. N. Guo, J. N. Zhang, Y. Wang, S. K. Qiao, F. R. Lin, [Expression of c-fes gene in leukemia cells and its clinical significance]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi* **17**, 1429-1433 (2009).
42. B. Lai, Y. Lai, Y. Zhang, M. Zhou, L. Sheng, G. OuYang, The Solute Carrier Family 2 Genes Are Potential Prognostic Biomarkers in Acute Myeloid Leukemia. *Technol Cancer Res Treat* **19**, 1533033819894308 (2020).
43. Chen, L. Luo, C. Liang, Aberrant S100A16 expression might be an independent prognostic indicator of unfavorable survival in non-small cell lung adenocarcinoma. *PLoS One* **13**, e0197402 (2018).
44. X. He, X. Xu, A. Q. Khan, W. Ling, High Expression of S100A6 Predicts Unfavorable Prognosis of Lung Squamous Cell Cancer. *Med Sci Monit* **23**, 5011-5017 (2017).
45. S. Zheng, H. Shen, Q. Jia, C. Jing, J. Lin, M. Zhang, X. Zhang, B. Zhang, Y. Liu, S100A6 promotes proliferation of intrahepatic cholangiocarcinoma cells via the activation of the p38/MAPK pathway. *Future Oncol* **13**, 2053-2063 (2017).
46. M. R. Mansour, S. He, Z. Li, R. Lobbardi, B. J. Abraham, C. Hug, S. Rahman, T. E. Leon, Y. Y. Kuang, M. W. Zimmerman, T. Blonquist, E. Gjini, A. Gutierrez, Q. Tang, L. Garcia-Perez, K. Pike-Overzet, L. Anders, A. Berezovskaya, Y. Zhou, L. I. Zon, D. Neuberg, A. K. Fielding, F. J. T. Staal, D. M. Langenau, T. Sanda, R. A. Young, A. T. Look, JDP2: An oncogenic bZIP transcription factor in T cell acute lymphoblastic leukemia. *J Exp Med* **215**, 1929-1945 (2018).
47. J. F. Lesesve, S. Tardy, B. Frotscher, V. Latger-Cannard, P. Feugier, M. De Carvalho Bittencourt, Combination of CD160 and CD200 as a useful tool for differential diagnosis between chronic lymphocytic leukemia and other mature B-cell neoplasms. *Int J Lab Hematol* **37**, 486-494 (2015).
48. S. Pai, O. A. Bamodu, Y. K. Lin, C. S. Lin, P. Y. Chu, M. H. Chien, L. S. Wang, M. Hsiao, C. T. Yeh, J. T. Tsai, CD47-SIRPalpha Signaling Induces Epithelial-Mesenchymal Transition and Cancer Stemness and Links to a Poor Prognosis in Patients with Oral Squamous Cell Carcinoma. *Cells* **8**, (2019).
49. X. Jiang, H. Xie, Y. Dou, J. Yuan, D. Zeng, S. Xiao, Expression and function of FRA1 protein in tumors. *Mol Biol Rep* **47**, 737-752 (2020).

50. S. Newman, A. Pappo, S. Raimondi, J. Zhang, R. Barnhill, A. Bahrami, Pathologic Characteristics of Spitz Melanoma With MAP3K8 Fusion or Truncation in a Pediatric Cohort. *Am J Surg Pathol* **43**, 1631-1637 (2019).
51. B. D. Lehmann, T. M. Shaver, D. B. Johnson, Z. Li, P. I. Gonzalez-Ericsson, V. Sanchez, Y. Shyr, M. E. Sanders, J. A. Pietenpol, Identification of Targetable Recurrent MAP3K8 Rearrangements in Melanomas Lacking Known Driver Mutations. *Mol Cancer Res* **17**, 1842-1853 (2019).
52. J. S. Pyo, M. J. Park, C. N. Kim, TPL2 expression is correlated with distant metastasis and poor prognosis in colorectal cancer. *Hum Pathol* **79**, 50-56 (2018).
53. A. Chorzalska, N. Ahsan, R. S. P. Rao, K. Roder, X. Yu, J. Morgan, A. Tepper, S. Hines, P. Zhang, D. O. Treaba, T. C. Zhao, A. J. Olszewski, J. L. Reagan, O. Liang, P. A. Gruppuso, P. M. Dubielecka, Overexpression of Tpl2 is linked to imatinib resistance and activation of MEK-ERK and NF-kappaB pathways in a model of chronic myeloid leukemia. *Mol Oncol* **12**, 630-647 (2018).
54. J. Moreaux, D. Hose, A. Bonnefond, T. Reme, N. Robert, H. Goldschmidt, B. Klein, MYEOV is a prognostic factor in multiple myeloma. *Exp Hematol* **38**, 1189-1198 e1183 (2010).
55. R. A. de Almeida, T. Heuser, R. Blaschke, C. R. Bartram, J. W. Janssen, Control of MYEOV protein synthesis by upstream open reading frames. *J Biol Chem* **281**, 695-704 (2006).
56. C. R. Bodle, J. H. Schamp, J. B. O'Brien, M. P. Hayes, M. Wu, J. A. Doorn, D. L. Roman, Screen Targeting Lung and Prostate Cancer Oncogene Identifies Novel Inhibitors of RGS17 and Problematic Chemical Substructures. *SLAS Discov* **23**, 363-374 (2018).
57. L. Li, H. S. Luo, G-Protein Signaling Protein-17 (RGS17) Is Upregulated and Promotes Tumor Growth and Migration in Human Colorectal Carcinoma. *Oncol Res* **26**, 27-35 (2018).
58. J. Yu, M. Hou, T. Pei, FAM83A Is a Prognosis Signature and Potential Oncogene of Lung Adenocarcinoma. *DNA Cell Biol* **39**, 890-899 (2020).
59. T. C. Shih, Y. Fan, S. Kiss, X. Li, X. N. Deng, R. Liu, X. J. Chen, R. Carney, A. Chen, P. M. Ghosh, K. S. Lam, Galectin-1 inhibition induces cell apoptosis through dual suppression of CXCR4 and Ras pathways in human malignant peripheral nerve sheath tumors. *Neuro Oncol* **21**, 1389-1400 (2019).
60. M. Cheminant, J. Bruneau, G. Malamut, D. Sibon, N. Guegan, T. van Gils, S. Cording, A. Trinquand, V. Verkarre, L. Lhermitte, N. Brousse, A. S. Jannot, S. Khater, L. Frenzel, R. Delarue, F. Suarez, A. Marcais, C. J. Mulder, E. Macintyre, V. Asnafi, L. Pouyet, C. Bonnafous, F. Lhospice, T. J. Molina, B. Meresse, C. Cellier, N. Cerf-Bensussan, O. Hermine, C. n. degrees, NKp46 is a diagnostic biomarker and may be a therapeutic target in gastrointestinal T-cell lymphoproliferative diseases: a CELAC study. *Gut* **68**, 1396-1405 (2019).
61. I. M. Aries, I. S. Jerchel, R. E. van den Dungen, L. C. van den Berk, J. M. Boer, M. A. Horstmann, G. Escherich, R. Pieters, M. L. den Boer, EMP1, a novel poor prognostic factor in pediatric leukemia regulates prednisolone resistance, cell proliferation, migration and adhesion. *Leukemia* **28**, 1828-1837 (2014).
62. Y. Zhao, Y. Yu, W. Zhao, S. You, M. Feng, C. Xie, X. Chi, Y. Zhang, X. Wang, As a downstream target of the AKT pathway, NPTX1 inhibits proliferation and promotes apoptosis in hepatocellular carcinoma. *Biosci Rep* **39**, (2019).
63. D. J. Lee, F. Schonleben, V. E. Banuchi, W. Qiu, L. G. Close, A. M. Assaad, G. H. Su, Multiple tumor-suppressor genes on chromosome 3p contribute to head and neck squamous cell carcinoma tumorigenesis. *Cancer Biol Ther* **10**, 689-693 (2010).
64. H. Yang, Z. Wan, C. Huang, H. Yin, D. Song, AMPH-1 is a tumor suppressor of lung cancer by inhibiting Ras-Raf-MEK-ERK signal pathway. *Lasers Med Sci* **34**, 473-478 (2019).
65. S. Zhang, X. Pei, H. Hu, W. Zhang, X. Mo, Q. Song, Y. Zhang, K. Xu, Y. Wang, Y. Na, Functional characterization of the tumor suppressor CMTM8 and its association with prognosis in bladder cancer. *Tumour Biol* **37**, 6217-6225 (2016).
66. Z. Liao, X. Wang, X. Wang, L. Li, D. Lin, DEPDC7 inhibits cell proliferation, migration and invasion in hepatoma cells. *Oncol Lett* **14**, 7332-7338 (2017).
67. X. Sun, Q. Zhang, W. Chen, Q. Hu, Y. Lou, Q. H. Fu, J. Y. Zhang, Y. W. Chen, L. Y. Ye, Y. Wang, S. Z. Xie, L. Q. Hu, T. B. Liang, X. L. Bai, Hook1 inhibits malignancy and epithelial-mesenchymal transition in hepatocellular carcinoma. *Tumour Biol* **39**, 1010428317711098 (2017).
68. Y. Vivas-Garcia, P. Falletta, J. Liebing, P. Louphrasitthiphol, Y. Feng, J. Chauhan, D. A. Scott, N. Glodde, A. Chocarro-Calvo, S. Bonham, A. L. Osterman, R. Fischer, Z. Ronai, C. Garcia-Jimenez, M. Holzel, C. R. Goding, Lineage-Restricted Regulation of SCD and Fatty Acid Saturation by MITF Controls Melanoma Phenotypic Plasticity. *Mol Cell* **77**, 120-137 e129 (2020).

69. S. Gu, S. Lin, D. Ye, S. Qian, D. Jiang, X. Zhang, Q. Li, J. Yang, X. Ying, Z. Li, M. Tang, J. Wang, M. Jin, K. Chen, Genome-wide methylation profiling identified novel differentially hypermethylated biomarker MPPED2 in colorectal cancer. *Clin Epigenetics* **11**, 41 (2019).
70. J. Lv, P. Zhu, X. Zhang, L. Zhang, X. Chen, F. Lu, Z. Yu, S. Liu, PCDH9 acts as a tumor suppressor inducing tumor cell arrest at G0/G1 phase and is frequently methylated in hepatocellular carcinoma. *Mol Med Rep* **16**, 4475-4482 (2017).
71. A. Roy, M. Ramalinga, O. J. Kim, J. Chijioke, S. Lynch, S. Byers, D. Kumar, Multiple roles of RARRES1 in prostate cancer: Autophagy induction and angiogenesis inhibition. *PLoS One* **12**, e0180344 (2017).
72. W. Maat, S. H. Beiboer, M. J. Jager, G. P. Luyten, N. A. Gruis, P. A. van der Velden, Epigenetic regulation identifies RASEF as a tumor-suppressor gene in uveal melanoma. *Invest Ophthalmol Vis Sci* **49**, 1291-1298 (2008).
73. J. Deng, J. Guo, X. Guo, Y. Hou, X. Xie, C. Sun, R. Zhang, X. Yu, H. Liang, Mediation of the malignant biological characteristics of gastric cancer cells by the methylated CpG islands in RNF180 DNA promoter. *Oncotarget* **7**, 43461-43474 (2016).
74. J. Zhang, W. Y. Lu, J. M. Zhang, R. Q. Lu, L. X. Wu, Y. Z. Qin, Y. R. Liu, Y. Y. Lai, H. Jiang, Q. Jiang, B. Jiang, L. P. Xu, X. H. Zhang, X. J. Huang, G. R. Ruan, K. Y. Liu, S100A16 suppresses the growth and survival of leukaemia cells and correlates with relapse and relapse free survival in adults with Philadelphia chromosome-negative B-cell acute lymphoblastic leukaemia. *Br J Haematol* **185**, 836-851 (2019).
75. G. Wan, Y. Liu, J. Zhu, L. Guo, C. Li, Y. Yang, X. Gu, L. L. Deng, C. Lu, SLFN5 suppresses cancer cell migration and invasion by inhibiting MT1-MMP expression via AKT/GSK-3beta/beta-catenin pathway. *Cell Signal* **59**, 1-12 (2019).
76. L. Milani, A. Lundmark, A. Kiialainen, J. Nordlund, T. Flaegstad, E. Forestier, M. Heyman, G. Jonmundsson, J. Kanerva, K. Schmiegelow, S. Soderhall, M. G. Gustafsson, G. Lonnerholm, A. C. Syvanen, DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* **115**, 1214-1225 (2010).
77. J. Nordlund, C. L. Backlin, P. Wahlberg, S. Busche, E. C. Berglund, M. L. Eloranta, T. Flaegstad, E. Forestier, B. M. Frost, A. Harila-Saari, M. Heyman, O. G. Jonsson, R. Larsson, J. Palle, L. Ronnblom, K. Schmiegelow, D. Sinnett, S. Soderhall, T. Pastinen, M. G. Gustafsson, G. Lonnerholm, A. C. Syvanen, Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol* **14**, r105 (2013).
78. B. Chen, L. Jiang, M. L. Zhong, J. F. Li, B. S. Li, L. J. Peng, Y. T. Dai, B. W. Cui, T. Q. Yan, W. N. Zhang, X. Q. Weng, Y. Y. Xie, J. Lu, R. B. Ren, S. N. Chen, J. D. Hu, D. P. Wu, Z. Chen, J. Y. Tang, J. Y. Huang, J. Q. Mi, S. J. Chen, Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **115**, 373-378 (2018).
79. D. C. Lin, M. R. Wang, H. P. Koeffler, Genomic and Epigenomic Aberrations in Esophageal Squamous Cell Carcinoma and Implications for Patients. *Gastroenterology* **154**, 374-389 (2018).
80. J. M. Navarro, A. Touzart, L. C. Pradel, M. Loosveld, M. Koubi, R. Fenouil, S. Le Noir, M. A. Maqbool, E. Morgado, C. Gregoire, S. Jaeger, E. Mamessier, C. Pignon, S. Hacein-Bey-Abina, B. Malissen, M. Gut, I. G. Gut, H. Dombret, E. A. Macintyre, S. J. Howe, H. B. Gaspar, A. J. Thrasher, N. Ifrah, D. Payet-Bornet, E. Duprez, J. C. Andrau, V. Asnafi, B. Nadel, Site- and allele-specific polycomb dysregulation in T-cell leukaemia. *Nat Commun* **6**, 6094 (2015).
81. O. Bernard, P. Guglielmi, P. Jonveaux, D. Cherif, S. Gisselbrecht, M. Mauchauffe, R. Berger, C. J. Larsen, D. Mathieu-Mahul, Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. *Genes Chromosomes Cancer* **1**, 194-208 (1990).
82. M. R. Mansour, B. J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, L. B. Silverman, M. L. Loh, S. P. Hunger, T. Sanda, R. A. Young, A. T. Look, Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373-1377 (2014).
83. V. Grossmann, C. Haferlach, S. Weissmann, A. Roller, S. Schindela, F. Poetzinger, K. Stadler, F. Bellos, W. Kern, T. Haferlach, S. Schnittger, A. Kohlmann, The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes Chromosomes Cancer* **52**, 410-422 (2013).
84. A. C. Kramer, A. Kothari, W. C. Wilson, H. Celik, J. Nikitas, C. Mallaney, E. L. Ostrander, E. Eultgen, A. Martens, M. C. Valentine, A. L. Young, T. E. Druley, M. E. Figueroa, B. Zhang, G. A. Challen, Dnmt3a regulates T-cell development and suppresses T-ALL transformation. *Leukemia* **31**, 2479-2490 (2017).

85. S. Aref, N. El Menshawy, M. S. El-Ghonemy, T. A. Zeid, M. A. El-Baiomy, Clinicopathologic Effect of DNMT3A Mutation in Adult T-Cell Acute Lymphoblastic Leukemia. *Clin Lymphoma Myeloma Leuk* **16**, 43-48 (2016).
86. T. J. Ley, L. Ding, M. J. Walter, M. D. McLellan, T. Lamprecht, D. E. Larson, C. Kandoth, J. E. Payton, J. Baty, J. Welch, C. C. Harris, C. F. Lichti, R. R. Townsend, R. S. Fulton, D. J. Dooling, D. C. Koboldt, H. Schmidt, Q. Zhang, J. R. Osborne, L. Lin, M. O'Laughlin, J. F. McMichael, K. D. Delehaunty, S. D. McGrath, L. A. Fulton, V. J. Magrini, T. L. Vickery, J. Hundal, L. L. Cook, J. J. Conyers, G. W. Swift, J. P. Reed, P. A. Alldredge, T. Wylie, J. Walker, J. Kalicki, M. A. Watson, S. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, P. Westervelt, M. H. Tomasson, D. C. Link, T. A. Graubert, J. F. DiPersio, E. R. Mardis, R. K. Wilson, DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433 (2010).
87. P. Zhao, J. Huang, D. Zhang, D. Zhang, F. Wang, Y. Qu, T. Guo, Y. Qin, J. Wei, T. Niu, Y. Zheng, SLC2A5 overexpression in childhood philadelphia chromosome-positive acute lymphoblastic leukaemia. *Br J Haematol* **183**, 242-250 (2018).
88. H. Ji, L. I. Ehrlich, J. Seita, P. Murakami, A. Doi, P. Lindau, H. Lee, M. J. Aryee, R. A. Irizarry, K. Kim, D. J. Rossi, M. A. Inlay, T. Serwold, H. Karsunky, L. Ho, G. Q. Daley, I. L. Weissman, A. P. Feinberg, Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338-342 (2010).
89. C. C. Oakes, M. Seifert, Y. Assenov, L. Gu, M. Przekopowitz, A. S. Ruppert, Q. Wang, C. D. Imbusch, A. Serva, S. D. Koser, D. Brocks, D. B. Lipka, O. Bogatyrova, D. Weichenhan, B. Brors, L. Rassenti, T. J. Kipps, D. Mertens, M. Zapatka, P. Lichter, H. Dohner, R. Kuppers, T. Zenz, S. Stilgenbauer, J. C. Byrd, C. Plass, DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet* **48**, 253-264 (2016).
90. B. Erarslan-Uysal, J. B. Kunz, T. Rausch, P. Richter-Pechanska, I. A. van Belzen, V. Frismantas, B. Bornhauser, D. Ordonez-Rueada, M. Paulsen, V. Benes, M. Stanulla, M. Schrappe, G. Cario, G. Escherich, K. Bakharevich, R. Kirschner-Schwabe, C. Eckert, T. Loukanov, M. Gorenflo, S. M. Waszak, J. P. Bourquin, M. U. Muckenthaler, J. O. Korbel, A. E. Kulozik, Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors. *EMBO Mol Med* **12**, e12104 (2020).
91. J. Soulier, E. Clappier, J. M. Cayuela, A. Regnault, M. Garcia-Peydro, H. Dombret, A. Baruchel, M. L. Toribio, F. Sigaux, HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* **106**, 274-286 (2005).
92. I. Homminga, R. Pieters, A. W. Langerak, J. J. de Rooi, A. Stubbs, M. Verstegen, M. Vuerhard, J. Buijs-Gladdines, C. Kooi, P. Klous, P. van Vlierberghe, A. A. Ferrando, J. M. Cayuela, B. Verhaaf, H. B. Beverloo, M. Horstmann, V. de Haas, A. S. Wiekmeijer, K. Pike-Overzet, F. J. Staal, W. de Laat, J. Soulier, F. Sigaux, J. P. Meijerink, Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* **19**, 484-497 (2011).
93. S. Heerboth, K. Lapinska, N. Snyder, M. Leary, S. Rollinson, S. Sarkar, Use of epigenetic drugs in disease: an overview. *Genet Epigenet* **6**, 9-19 (2014).
94. S. Ivanoff, B. Gruson, S. P. Chantepie, E. Lemasle, L. Merlusca, V. Harrivel, A. Charbonnier, P. Votte, B. Royer, J. P. Marolleau, 5-Azacytidine treatment for relapsed or refractory acute myeloid leukemia after intensive chemotherapy. *Am J Hematol* **88**, 601-605 (2013).
95. A. M. Cseh, C. M. Niemeyer, A. Yoshimi, A. Catala, M. C. Fruhwald, H. Hasle, M. M. van den Heuvel-Eibrink, M. Lauten, B. De Moerloose, O. P. Smith, T. Bernig, B. Gruhn, A. E. Kulozik, M. Metzler, L. Olcay, M. Suttorp, I. Furlan, B. Strahm, C. Flotho, Therapy with low-dose azacitidine for MDS in children and young adults: a retrospective analysis of the EWOG-MDS study group. *Br J Haematol* **172**, 930-936 (2016).
96. F. Huguet, S. Chevret, T. Leguay, X. Thomas, N. Boissel, M. Escoffre-Barbe, P. Chevallier, M. Hunault, N. Vey, C. Bonmati, S. Lepretre, J. P. Marolleau, T. Pabst, P. Rousselot, A. Buzyn, J. Y. Cahn, V. Lheritier, M. C. Bene, V. Asnafi, E. Delabesse, E. Macintyre, Y. Chalandon, N. Ifrah, H. Dombret, A. L. L. Group of Research on Adult, Intensified Therapy of Acute Lymphoblastic Leukemia in Adults: Report of the Randomized GRAALL-2005 Clinical Trial. *J Clin Oncol* **36**, 2514-2523 (2018).
97. S. Maury, S. Chevret, X. Thomas, D. Heim, T. Leguay, F. Huguet, P. Chevallier, M. Hunault, N. Boissel, M. Escoffre-Barbe, U. Hess, N. Vey, J. M. Pignon, T. Braun, J. P. Marolleau, J. Y. Cahn, Y. Chalandon, V. Lheritier, K. Beldjord, M. C. Bene, N. Ifrah, H. Dombret, G. for, Rituximab in B-Lineage Adult Acute Lymphoblastic Leukemia. *N Engl J Med* **375**, 1044-1053 (2016).

98. V. Asnafi, A. Buzyn, S. Le Noir, F. Baleydier, A. Simon, K. Beldjord, O. Reman, F. Witz, T. Fagot, E. Tavernier, P. Turlure, T. Leguay, F. Huguet, J. P. Vernant, F. Daniel, M. C. Bene, N. Ifrah, X. Thomas, H. Dombret, E. Macintyre, NOTCH1/FBXW7 mutation identifies a large subgroup with favorable outcome in adult T-cell acute lymphoblastic leukemia (T-ALL): a Group for Research on Adult Acute Lymphoblastic Leukemia (GRAALL) study. *Blood* **113**, 3918-3924 (2009).
99. J. Bergeron, E. Clappier, I. Radford, A. Buzyn, C. Millien, G. Soler, P. Ballerini, X. Thomas, J. Soulier, H. Dombret, E. A. Macintyre, V. Asnafi, Prognostic and oncogenic relevance of TLX1/HOX11 expression level in T-ALLs. *Blood* **110**, 2324-2330 (2007).
100. Y. Assenov, F. Muller, P. Lutsik, J. Walter, T. Lengauer, C. Bock, Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods* **11**, 1138-1140 (2014).
101. R. Gaujoux, C. Seoighe, A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
102. J. P. Brunet, P. Tamayo, T. R. Golub, J. P. Mesirov, Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* **101**, 4164-4169 (2004).
103. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

Acknowledgments: The authors would like to thank the Biomis core facility at Institute Pasteur, Paris and the Genomics and Proteomics Core facilities at DKFZ. **Funding**: The Study was supported in part by the Baden-Württemberg Stiftung to CP and the Helmholtz-Foundation. The GRAALL was supported by grants P0200701 and P030425/AOM03081 from Le Programme Hospitalier de Recherche Clinique, Ministère de l'Emploi et de la Solidarité, France and the Swiss Federal Government in Switzerland. AT was supported by a grant from the French CARPEM (CAncer Research for PErsonalized Medicine) SIRIC program and the DKFZ Postdoctoral fellowship. The Work in VA lab was supported by ARC-Labellisation and the associations "Force hémato" and "Laurette Fugain". This study was also supported in VA and SS lab by grants from INCA PLBIO18-031, PLBIO 2018-00252 and ITMO Cancer Epig-2015. **Author contributions**: A.T, V.A, and C.P conceived and designed the study; A.T, and C.S performed the flow cytometry, library preparation for NGS, and mice experiments. A.M analysed the data and performed statistical analysis. All authors have critically reviewed the manuscript. **Competing interests**: The authors declare no potential conflicts of interest. **Data and materials availability**: All microarray raw IDAT files have been deposited to gene expression omnibus (GEO) under the accession number GSE147667. A complete reproducible source code used for data analysis is provided as **Supplementary Data S1.**

Figures:

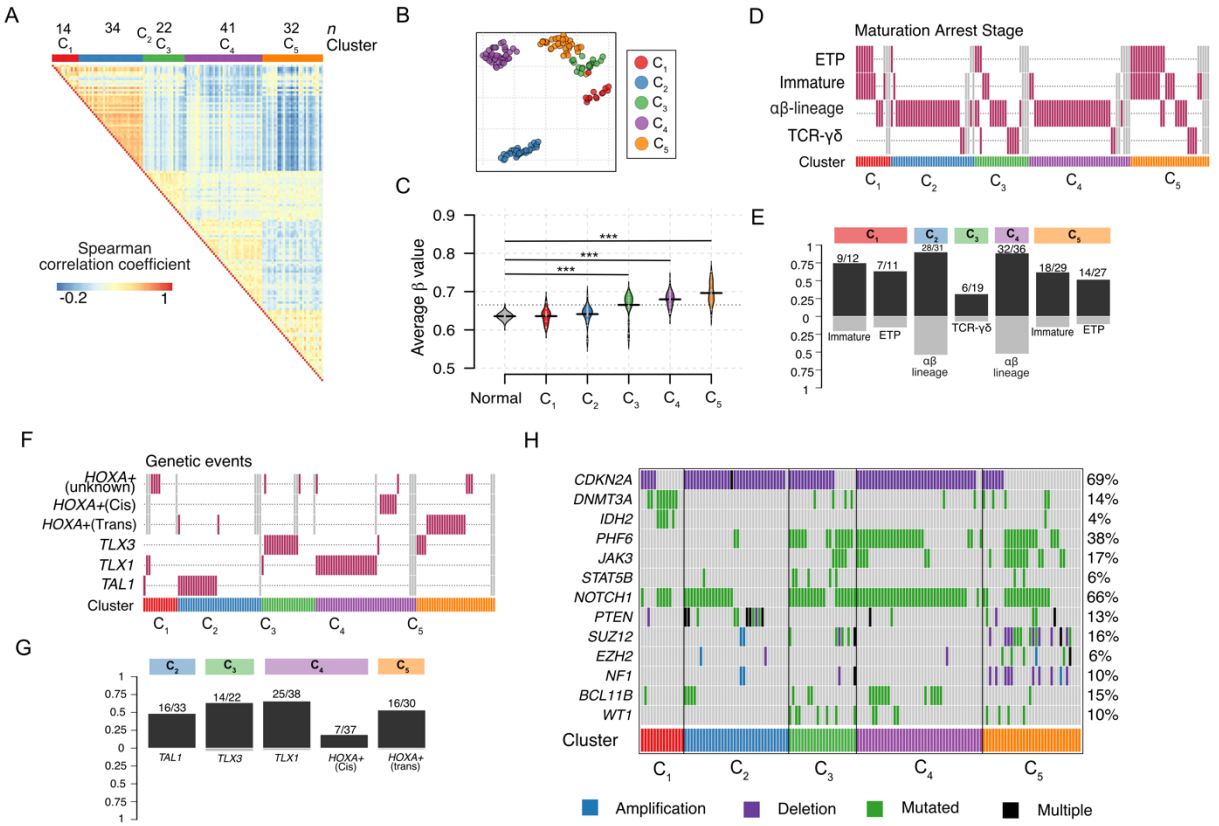


Figure-1: Epigenetic clusters in T-ALL. A. Heatmap of spearman correlation coefficient values between all leukemic samples. Top annotation bars indicate cluster sizes (*n*) and titles respectively (top to bottom). **B.** Uniform Manifold Approximation and Projection (UMAP) plot of T-ALL samples color-coded according to the cluster. **C.** Violin plots depicting genome-wide DNAm values for every cluster. (*** $P < 0.001$ two tailed t-test for differences in mean). **D.** Heatmap showing association between maturation arrest stages (row) and T-ALL samples (columns). Each column represents a sample, and their corresponding cluster assignments are displayed as a colour coded bottom annotation bar. Rows are ordered according to T-cell maturation stages (from top most immature to bottom most mature). Gray bars indicate data not available. **E.** Bar plots displaying significant association between clusters and maturation arrest stages (Fisher's exact test; $P < 0.01$). Bars are annotated with the ratio of samples belonging to a maturation arrest stage (as indicated by the bar title), and total number of samples within the clusters for which annotations are available, with the corresponding fraction on Y-axis. Bottom gray bars indicate number of samples in background with the highlighted maturation arrest stage. **F.** Heatmap showing association between genetic events (row) and T-ALL samples (columns). Each column represents a sample, and their corresponding cluster assignments are displayed as a color-coded bottom annotation bar. Rows are ordered according to genetic events (either somatic epigenetic mutations or deregulated TFs). Gray bars indicate data not available. **G.** Bar plots displaying significant association between clusters and genetic events (Fisher's exact test; $P < 0.01$). Bars are annotated with the ratio of samples with the (as indicated by the bar title), and total number of samples within the clusters for which annotations are available, with the corresponding fraction on Y-axis. Bottom gray

bars indicate number of samples in background with the highlighted genetic event. **H.** Cluster specific mutations found across the cohort (Fisher's exact test; *FDR* < 0.05).

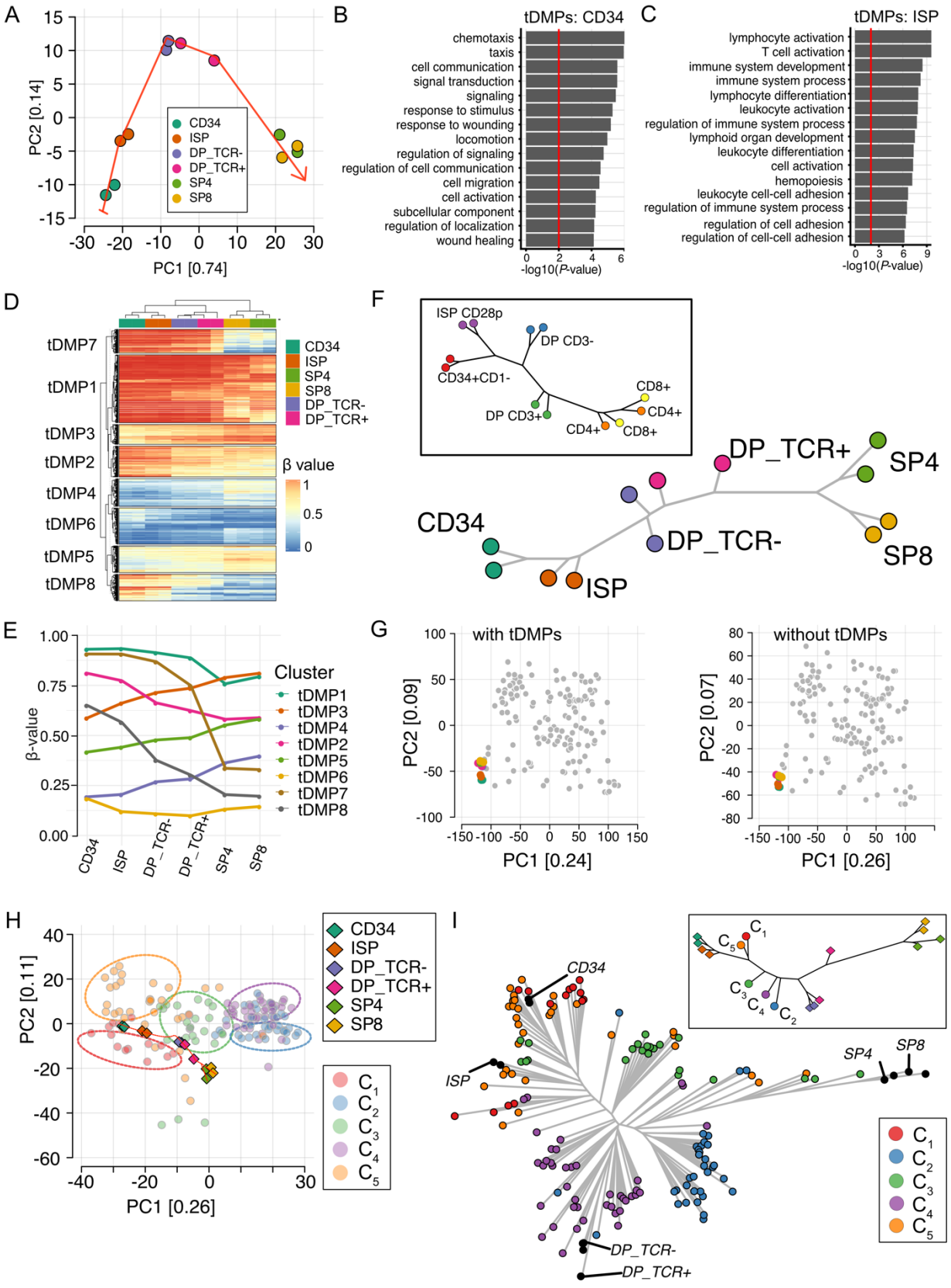

A
PC2 [0.14]
PC1 [0.74]
CD34
ISP
DP_TCR-
DP_TCR+
SP4
SP8
B
tDMPs: CD34
chemotaxis
taxis
cell communication
signal transduction
signaling
response to stimulus
response to wounding
locomotion
regulation of signaling
regulation of cell communication
cell migration
cell activation
subcellular component
regulation of localization
wound healing
-log10(P-value)
C
tDMPs: ISP
lymphocyte activation
T cell activation
immune system development
immune system process
lymphocyte differentiation
leukocyte activation
regulation of immune system process
lymphoid organ development
leukocyte differentiation
cell activation
hemopoiesis
leukocyte cell-cell adhesion
regulation of immune system process
regulation of cell adhesion
regulation of cell-cell adhesion
-log10(P-value)
D
tDMP7
tDMP1
tDMP3
tDMP2
tDMP4
tDMP6
tDMP5
tDMP8
β value
F
ISP CD28p
DP CD3-
CD34+CD1-
CD8+
CD4+
DP CD3+
CD4+
CD8+
DP_TCR+
SP4
CD34
DP_TCR-
SP8
ISP
E
β-value
Cluster
tDMP1
tDMP3
tDMP4
tDMP2
tDMP5
tDMP6
tDMP7
tDMP8
CD34
ISP
DP_TCR-
DP_TCR+
SP4
SP8
G
with tDMPs
without tDMPs
PC2 [0.09]
PC1 [0.24]
PC2 [0.07]
PC1 [0.26]
H
PC2 [0.11]
PC1 [0.26]
CD34
ISP
DP_TCR-
DP_TCR+
SP4
SP8
C1
C2
C3
C4
C5
I
CD34
ISP
SP4
SP8
DP_TCR-
DP_TCR+
C1
C2
C3
C4
C5

Figure-2: DNA methylation predicts T-cell differentiation and maturation arrest stages of T-ALL clusters. A. Principal Component Analysis (PCA) shows separation and ordering of T-cell subtypes according to different maturation stages (clockwise from bottom left to bottom right). The red curve indicates the direction of increasing T-cell maturation stages. **B and C.** Functional characterization of Differentially Methylated Probes (DMPs) from CD34 cells (B) and ISP cells (C), shows enrichment of cell-type specific biological processes. Vertical red bar indicates $P = 0.001$ (Hypergeometric test). **D.** Heatmap of DMPs from T-cell maturation stages categorized into distinct subgroups (tDMP1-tDMP8) based on dynamics of changes in DNAm across cell-types. **E.** Average DNAm of tDMP subgroups across T-cell subtypes (x-axis). **F.** T-cell developmental phylogenetic tree inferred from tDMPs shows the ordering of T-cell subtypes. Inset plot shows similar phylogenetic tree inferred from 1000 most differentially variably expressed genes. **G.** PCA of normal (colored) and T-ALL samples (in gray) with (left) and without (right) tDMPs shows little to no changes in variance explained by first two principal components. **H.** PCA of normal T-cell and T-ALL methylomes using tDMPs. Normal T-cell subpopulations ($n = 12$) are depicted in diamond shapes and T-ALLs ($n = 143$) are circles - color coded according to the cluster ($C_1 – C_5$). The known T-cell developmental trajectory starting from CD34 is shown as a red curve overlaid on top of normal T-cell population. **I.** Phylogenetic tree of the entire cohort ($n = 143$ T-ALLs; $n = 12$ normal T-cells) constructed using tDMPs shows ordering of T-ALL samples (color coded according to their corresponding clusters) along T-cell developmental pathway (left to right). Normal T-cell are in thick black circles labelled for cell-types. Inlet panel shows a simplified phylogenetic tree constructed with average DNAm levels of epigenetic clusters along with normal T-cell types (in diamond shapes) shows order of maturation arrest stages of T-ALL subtypes (in circles) along the T-cell developmental pathway (left to right).

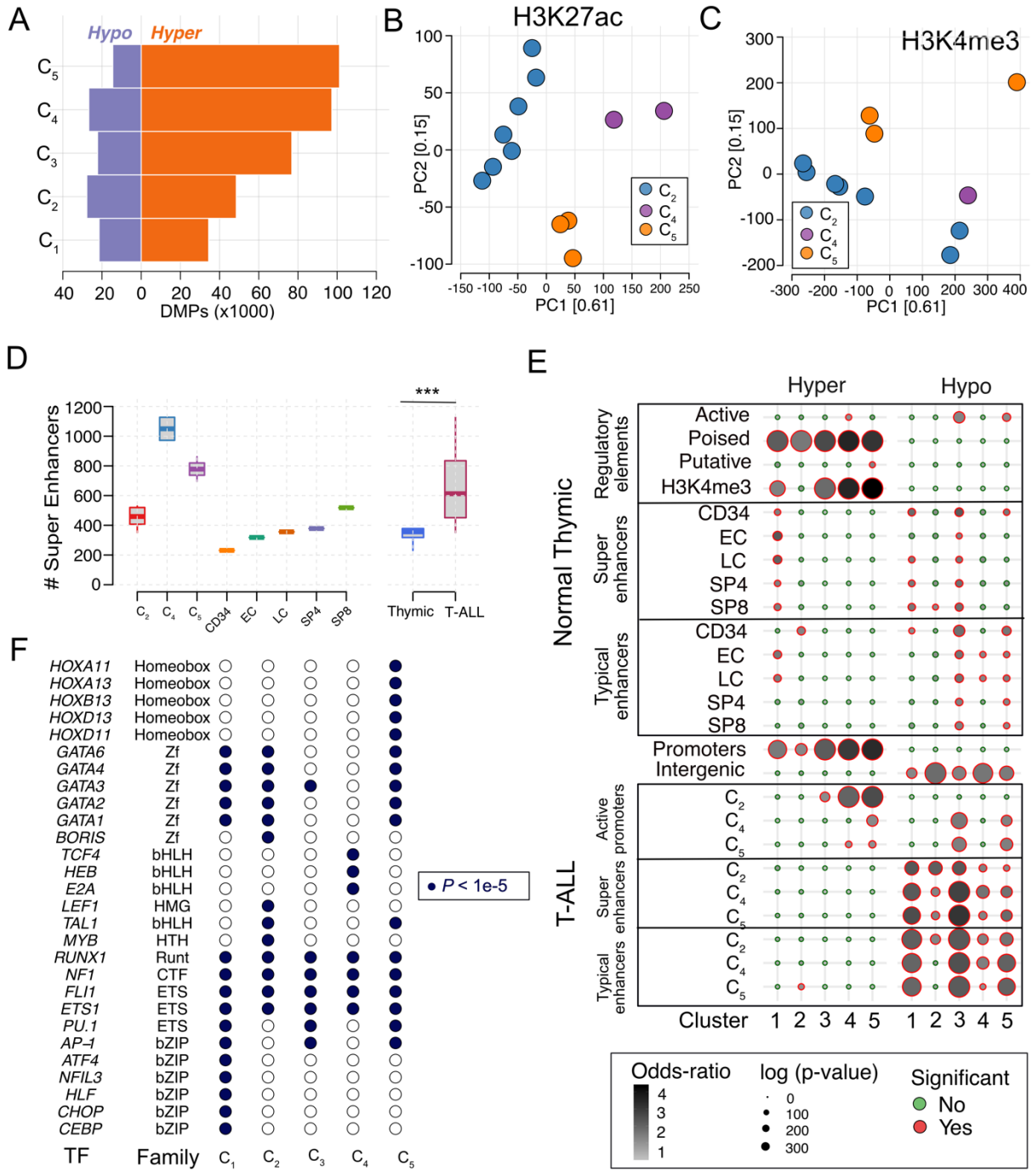


Figure-3: Characterization of Differentially Methylated Probes (DMPs) associated with T-ALL clusters. A. Numbers of hyper- and hypomethylated DMPs identified in each T-ALL cluster (C_1- C_5) as compared to normal pooled thymic subpopulations. **B and C.** Principal component analysis of primary T-ALL samples based on active H3K27ac peaks (peaks 2500 bp away from known TSS) (B) and H3K4me3 peaks (C). Samples are color coded according to their corresponding cluster. **D.** Dot plot for enrichment of hyper- and hypomethylated DMPs across all 5 clusters (X-axis) in various regulatory regions (Y-axis)

as highlighted in the left side of the plot. Dots are color coded for significance and size of the dots represents *P* values in log10 scale. **E.** Boxplot of number of super enhancers identified in each T-ALL samples, clusters and thymic cell types (X-axis) (*** t-test; $P < 0.01$). **F.** Disease associated TF motifs detected among hypomethylated DMPs (+/- 100bp) enriched within T-ALL associated super and typical enhancer regions.

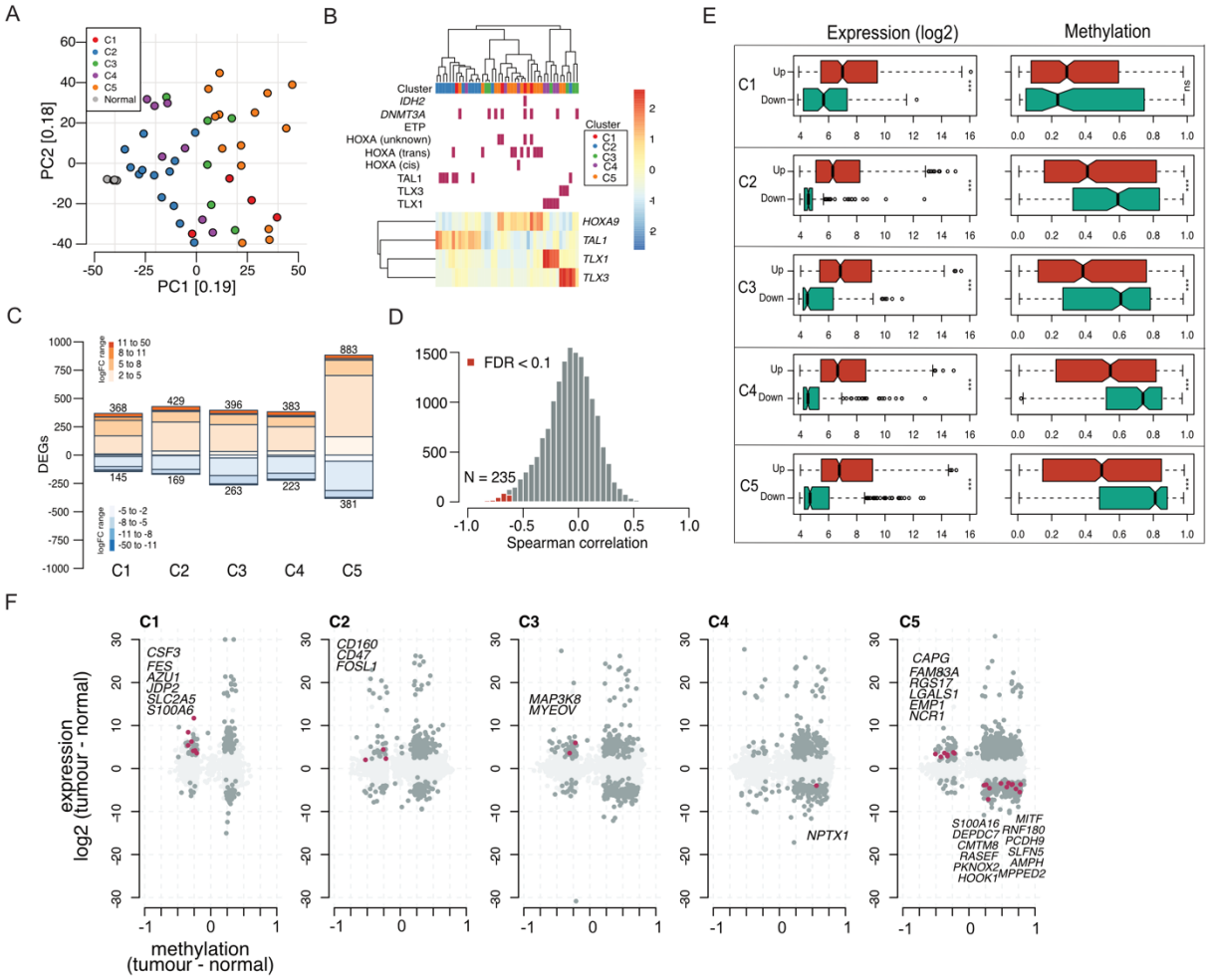


Figure-4: Integrated analysis of DNA methylation and gene expression. A. Principal component analysis of gene expression data for 48 samples (44T-ALL + 4 total thymus). Samples are color coded for their T-ALL cluster. **B.** Heatmap of known T-ALL oncogenes. Top annotation bars depict validated genetic annotations. **C.** Barplot of number of differentially expressed genes (DEGs) in each cluster compared to total thymus samples (*FDR* < 0.1). Up and down regulated genes are color coded for their fold change range. **D.** Histogram of Pearson correlation coefficient between promoter* DNAm and gene expressions of all protein coding genes (*N* = 15,912 genes) across all tumor samples (*N* = 44). **E.** Distribution of gene expression (Up and Down regulated genes from **panel C**) and their corresponding promoter* DNAm for every cluster. Notches indicate 95% CI around the median. (*** $P < 0.001$, t-test for differences in mean). **F.** Scatter plot of gene expression and DNAm differences in DMPs between tumor and normal samples for all 5 clusters

(C1-C5). Dark gray color indicates genes also found to be significantly differentially expressed (FDR < 0.1; from panel C). Selected genes known to be associated with the T-ALL pathogenesis are highlighted in red and annotated.
*Promoter DNAm is estimated by averaging the beta values from probes within 1200 bp upstream and 800 bp downstream of known TSS.

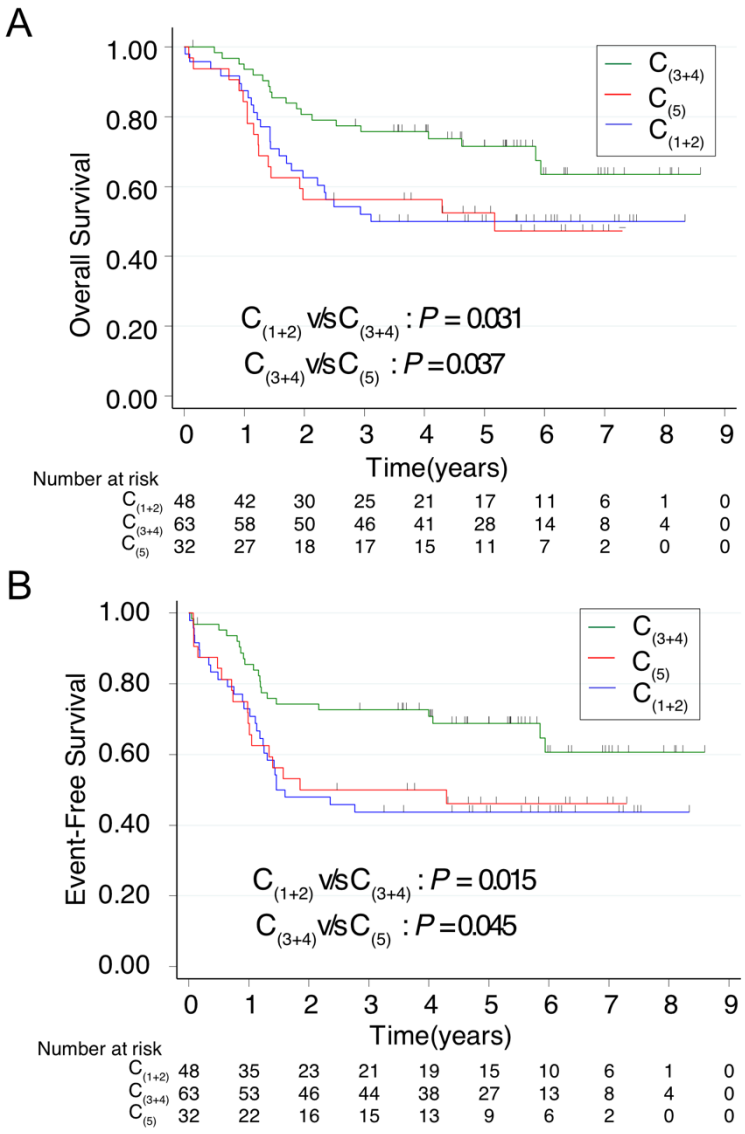


Figure-5: Prognostic impact of DNA methylation. A. Overall Survival (OS) for patients classified into hypomethylated (C_{1+2}), intermediate hypermethylated (C_{3+4}), and hypermethylated (C_5) subgroups. Risk table indicates number of individuals at risk for a given time point. *P*-values are derived from log-rank test. **B.** Event Free Survival (EFS) for patients classified into C_{1+2}, C_{3+4}, and C_5 subgroups.

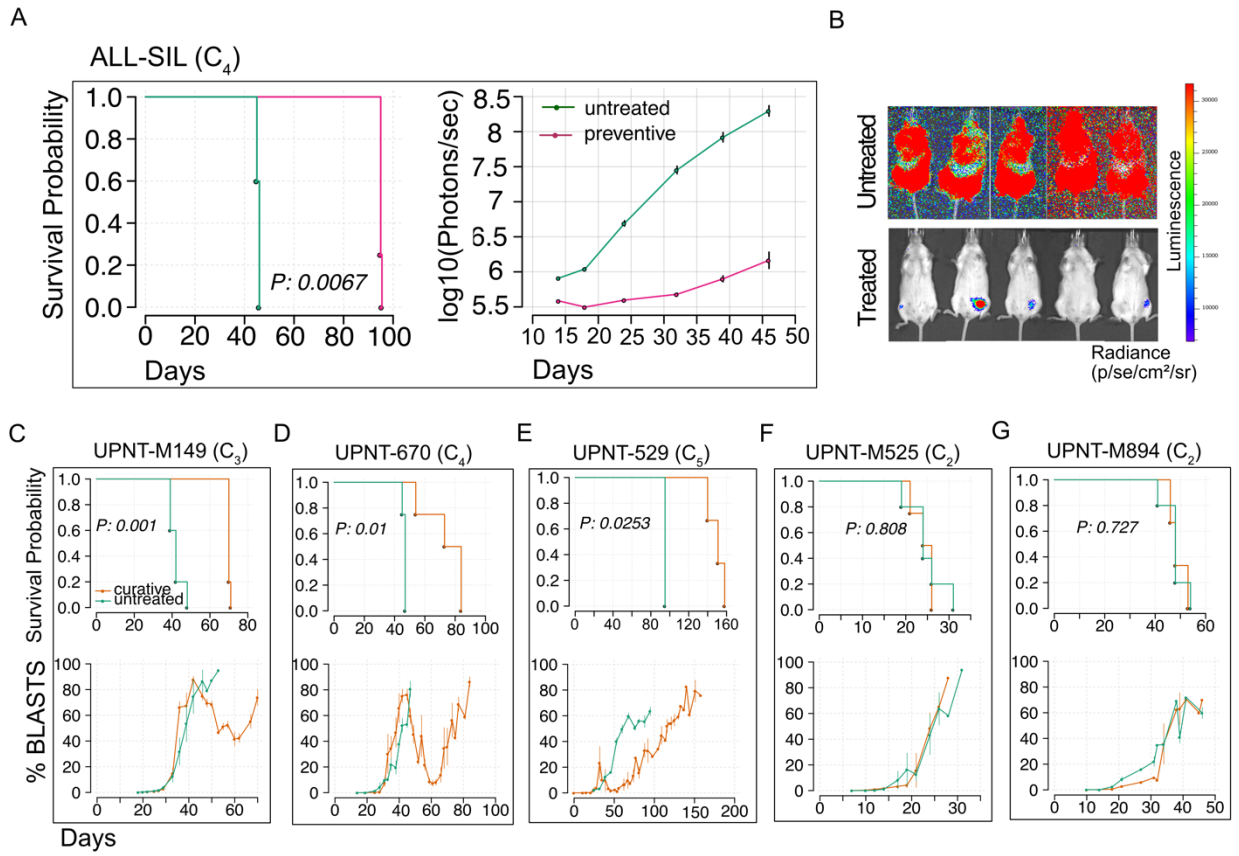


Figure-6: Efficacy of 5-Azacytidine treatment among T-ALL clusters. 5-Azacytidine treatment in PDX mice models. Leukemic cells (n = 1,000,000) were injected into NSG mice (day-0). **A.** Survival (left panel) and tumor burden measured by bioluminescence (right panel) for NSG mice injected with ALL-SILL cell line (predicted cluster C_4; n = 5 for untreated; n = 4 for treated) after 2 weeks of preventive treatment with 5-Azacytidine (treatment initiation at day 2 post-injection, 5mg/kg/day for 5 days a week for two weeks) (red line) or NaCl (green line). Tumor burden was measured as a function of luminescence (Y-axis) from tumor cells at different time points (X-axis). **B**. Bioluminescence radiance signals (photons/cm^2/sr) for mice treated with 5-Azacytidine or NaCl (untreated) at day 45 post-transplantation for ALL-SIL cell line.. **C-G.** Survival (top panel) and leukemic burden (%blast cells) measured by flow cytometry in peripheral blood (bottom panel) for five PDX mice models. **C.** UPNT-M149 (predicted C3; n = 4 to 5 per group), **D.** UPNT-670 (cluster C_4; n = 4 to 5 per group), **E.** UPNT-529 (C5; n = 3 per group), **F.** UPNT525 (cluster C_2; n = 4 to 5 per group) and **G.** UPNT-894 (cluster C_2; n = 3 to 5 per group) under untreated (NaCl; green line) or treated for 2 weeks with 5-Azacytidine in a curative-like manner (when blasts count reached 1% in blood, 5mg/kg/day for 5 days a week for two weeks) (orange line). *P*-values are calculated by comparing untreated group with rest of the groups using log-rank test.

Table 1: Characteristics of the 143 T-ALL patients[^]

Characteristic	Value
Median age at study entry (range) – year	29.9 (16.3-59.1)
Sex ratio (Male/Female) – no.	107/36
T-cell Receptor subsets analyzed – no./total no. (%)	
Immature (IM0. IMδ. IMγ)	33/127 (26)
IMβ/pre-αβ	66/127 (52)
TCRαβ+	14/127 (11)
TCRγδ+	14/127 (11)
Early T-cell precursor (ETP) Immunophenotype – no./total no. (%)	25/125 (20)
High risk patients* – no./total no. (%)	61/143 (43)
Oncogenetic category – no./total no. (%)	
TLX1	28/137 (20)
TLX3	19/137 (14)
SIL-TAL1/TAL1-neoenhancer	17/137 (12)
CALM-AF10	6/137 (4)
None of the above	67/137 (49)
HOXA deregulation[+] – no./total no. (%)	
Cis	7/127 (6)
Trans	17/127 (13)
Unknown	11/127 (9)
Mutations in epigenetic factors – no./total no. (%)	
DNMT3A	20/143 (14)
IDH1	3/143 (2)
IDH2	6/143 (4)
TET2	7/143 (5)
TET3	4/143 (3)
Early response – no./total no. (%)	
Prednisone response	80/143 (56)
Bone marrow response	78/140 (56)
Complete remission	134/143 (94)

^Percentages may not total 100 because of rounding up.

*High risk: *NOTCH1/FBXW7*^{WT} OR *NOTCH1*^{mut} + [*KRAS, NRAS, PTEN*]^{mut}

[+]Cis: *HOXA* over expression under the influence of TCRβ enhancer

Trans: *HOXA* over expression under the influence of *SET-NUP214, MLLT10* or MLL fusion

Table 2: Clinical characteristics of the three prognostic subgroups

	C(1+2)	C(3+4)	C(5)	*P*-value†
	N= 48	N=63	N=32	
Clinical Subsets Analyzed				
Median age in years (range)	26.2 (16.3-59.1)	30.5 (16.4-57.2)	31.4 (18.8-59)	0,3
age >45 years- no./total no. (%)	6/48 (13%)	6/63 (10%)	5/32 (15%)	0,67
Sex ratio, Male/Female- no.	38/10	46/17	23/9	0,7
WBC (G/L), median (range)	69.9 (0.9-604.4)	37 (4.1-645)	31.2 (2.2-241.6)	0,3
CNS involvement- no./total no. (%)	7/48 (15%)	8/63 (13%)	5/32 (16%)	0,9
Early Response				
Prednisone response- no./total no. (%)	26/48 (54%)	44/63 (70%)	9/32 (28%)	**0,001**
Bone marrow response- no./total no. (%)	30/45 (67%)	41/63 (65%)	7/32 (22%)	**<0.001**
MRD (TP1) <10-4- no./total no. (%)	18/23 (78%)	33/39 (85%)	8/21 (38%)	**0,001**
Complete remission- no./total no. (%)	44/48 (92%)	61/63 (99%)	29/32 (94%)	0,4

WBC (G/L), white blood cells; CNS, central nervous system; MRD (TP1), post-induction minimal residual disease; †Fisher's exact test or Mann-Whitney tests were used where appropriate.

Supplementary Material

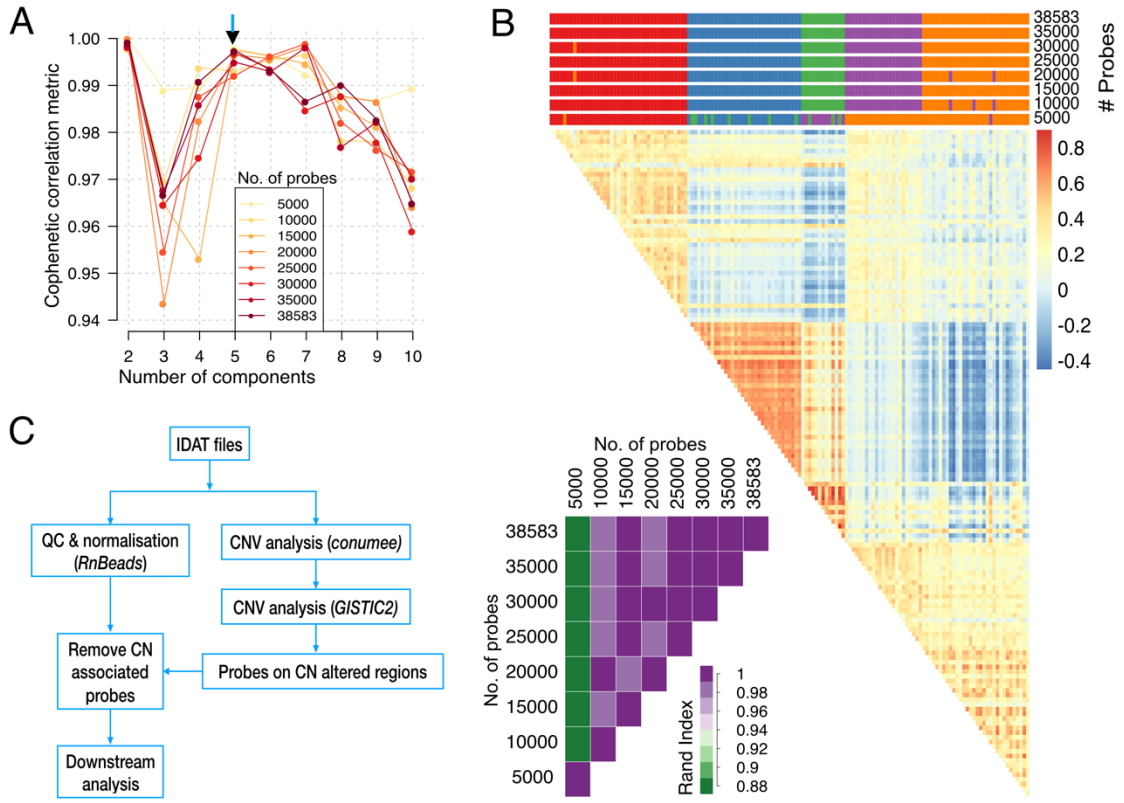
Epigenetic blueprint identifies poor outcome and hypomethylating agent-responsive T-ALL subgroup

Aurore Touzart[1,2]*, Anand Mayakonda[1,3]*, Charlotte Smith[2], Joschka Hey[1,3,4], Reka Toth[1], Agata Cieslak[2], Guillaume P. Andrieu[2], Christine Tran Quang[5], Mehdi Latiri[2], Jacques Ghysdael[5], Salvatore Spicuglia[6], Hervé Dombret[7], Norbert Ifrah[8], Elizabeth Macintyre[2], Pavlo Lutsik[2,9], Nicolas Boissel[7], Christoph Plass[1,9§] and Vahid Asnafi[2§]

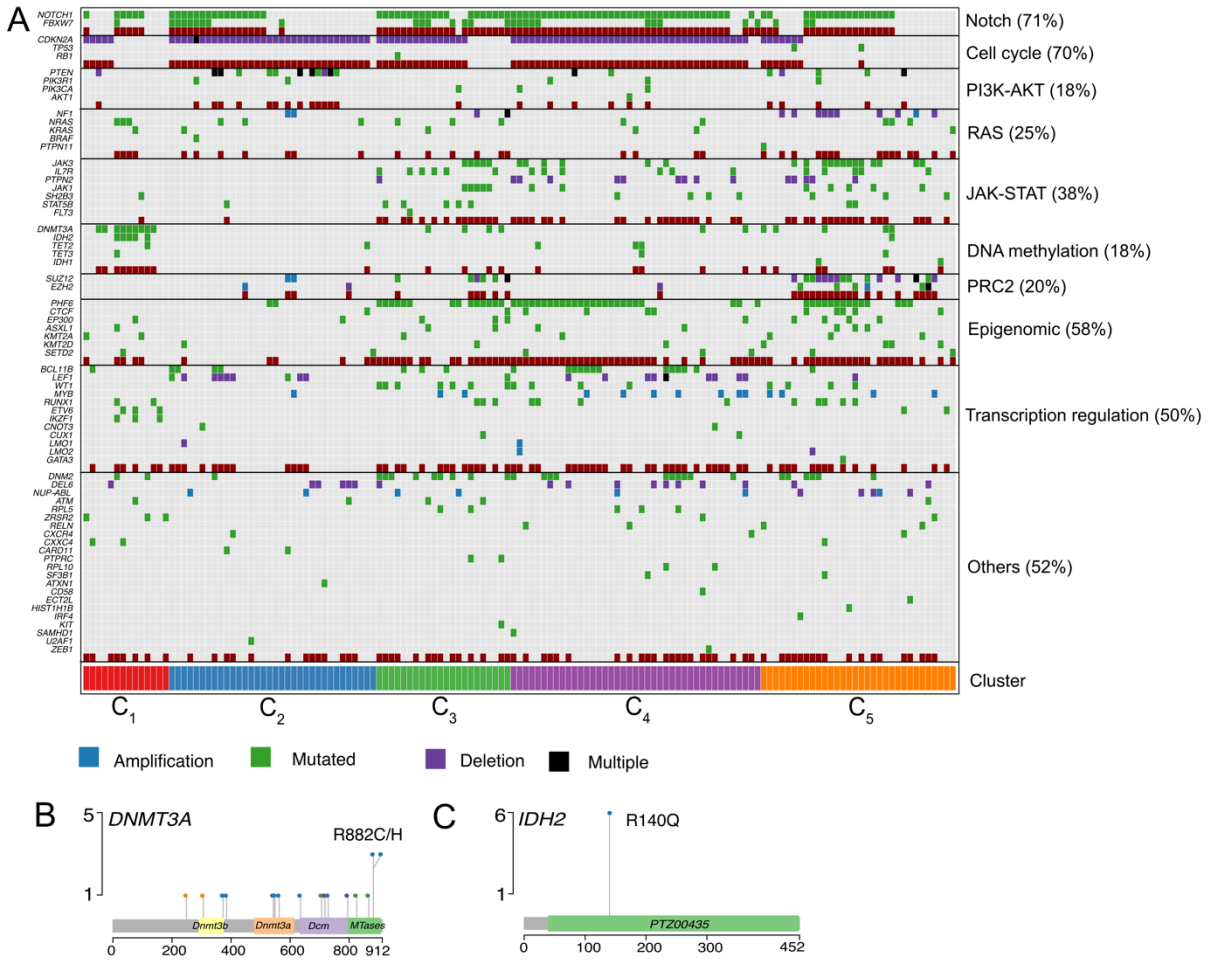
[1]Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]Université de Paris (Descartes), Institut Necker -Enfants Malades (INEM), Institut national de la santé et de la recherche médicale (Inserm) U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris, Hôpital Necker Enfants-Malades, Paris, France
[3]Faculty of Biosciences, Heidelberg University, 69120 Heidelberg, Germany
[4]Germany-Israeli Helmholtz Research School in Cancer Biology
[5]Institut Curie, CNRS UMR3348, University Paris Sud - Paris-Saclay, Orsay, France
[6]Aix-Marseille University, Inserm, Theories and Approaches of Genomic Complexity (TAGC), Equipe labellisée Ligue, UMR1090, 13288 Marseille, France
[7]Université Paris Diderot, Institut Universitaire d'Hématologie, EA-3518, Assistance Publique-Hôpitaux de Paris, University Hospital Saint-Louis, Paris, France.
[8]PRES LUNAM, CHU Angers service des Maladies du Sang et INSERM U 892, Angers, France,
[9]German Cancer Research Consortium (DKTK)
*Contributed equally
[§]Co-corresponding authors

Corresponding Authors
-Vahid Asnafi, Hôpital Necker Enfants-Malades, Laboratoire d'onco-hématologie, 149 rue de Sèvres, 75015 Paris, France. e-mail: vahid.asnafi@aphp.fr. Phone: (+33) 144 49 49 14; Fax: (+33) 144 38 17 45
- Christoph Plass, e-mail: c.plass@dkfz-heidelberg.de. German Cancer Research Center (DKFZ), Division of Cancer Epigenomics, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

Table of Contents:



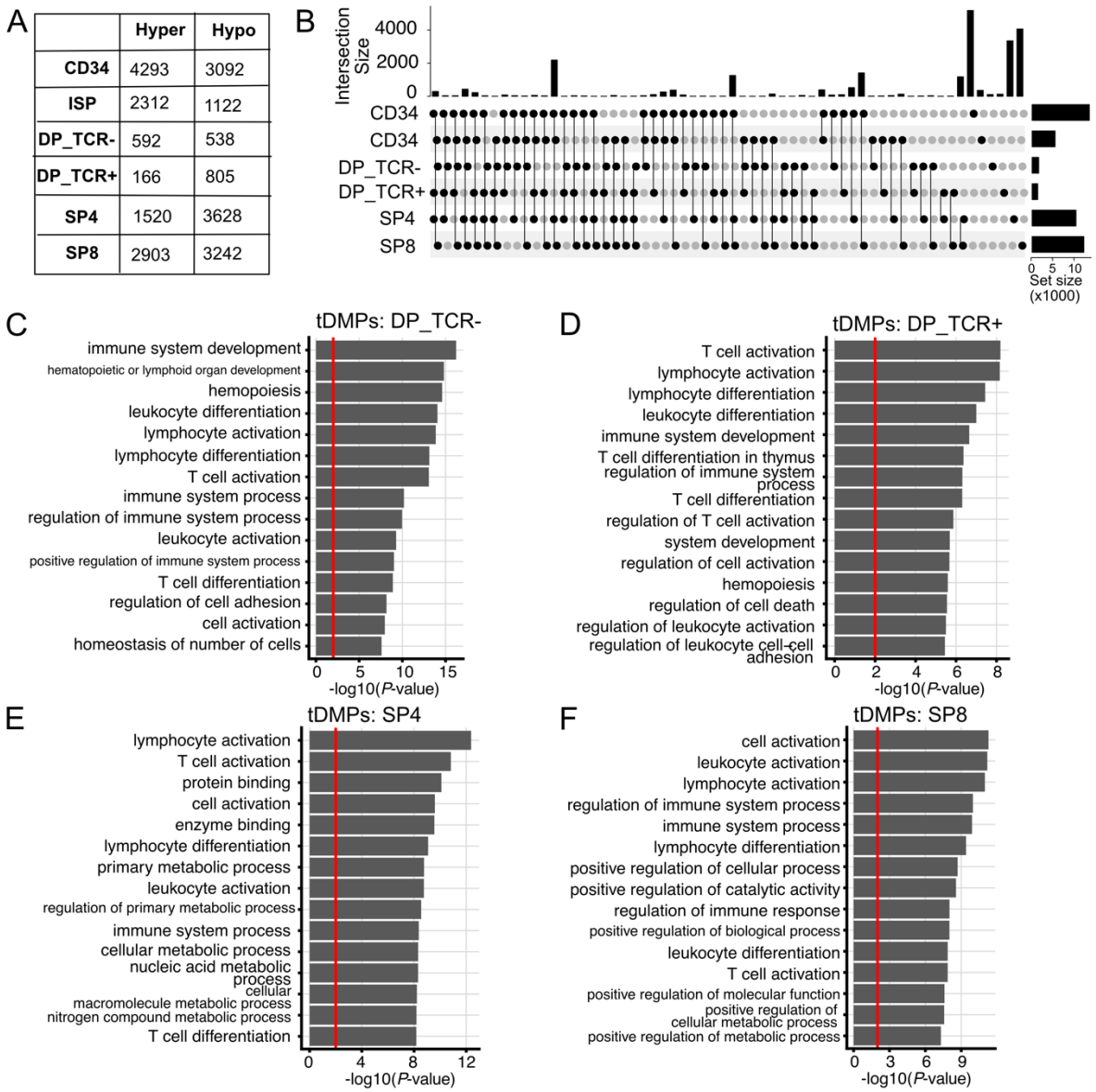
Supplementary Figure-S1: Identification of epigenetic clusters. A. Cophenetic correlation (Y-axis) measured for a range of values (2...10, X-axis). Optimum value is chosen at which the correlation value reaches maximum followed by no-change or decrease in correlation metric. Same step was performed for varying number of probes as represented by the lines with color gradient. Arrowhead represents the optimal value chosen for downstream analysis (N = 5). **B.** Cluster stability was measured by re-running clustering with n=5 for different number of probes as indicated in the top annotation bar. Heatmap represents spearman-correlation coefficient between samples. Inner panel shows pairwise Rand-index values which indicates similarity between two clustering results. **C.** Pre-processing and QC steps involved in analyzing raw Illumina EPIC arrays. R packages used in particular steps are mentioned within parenthesis in italics.



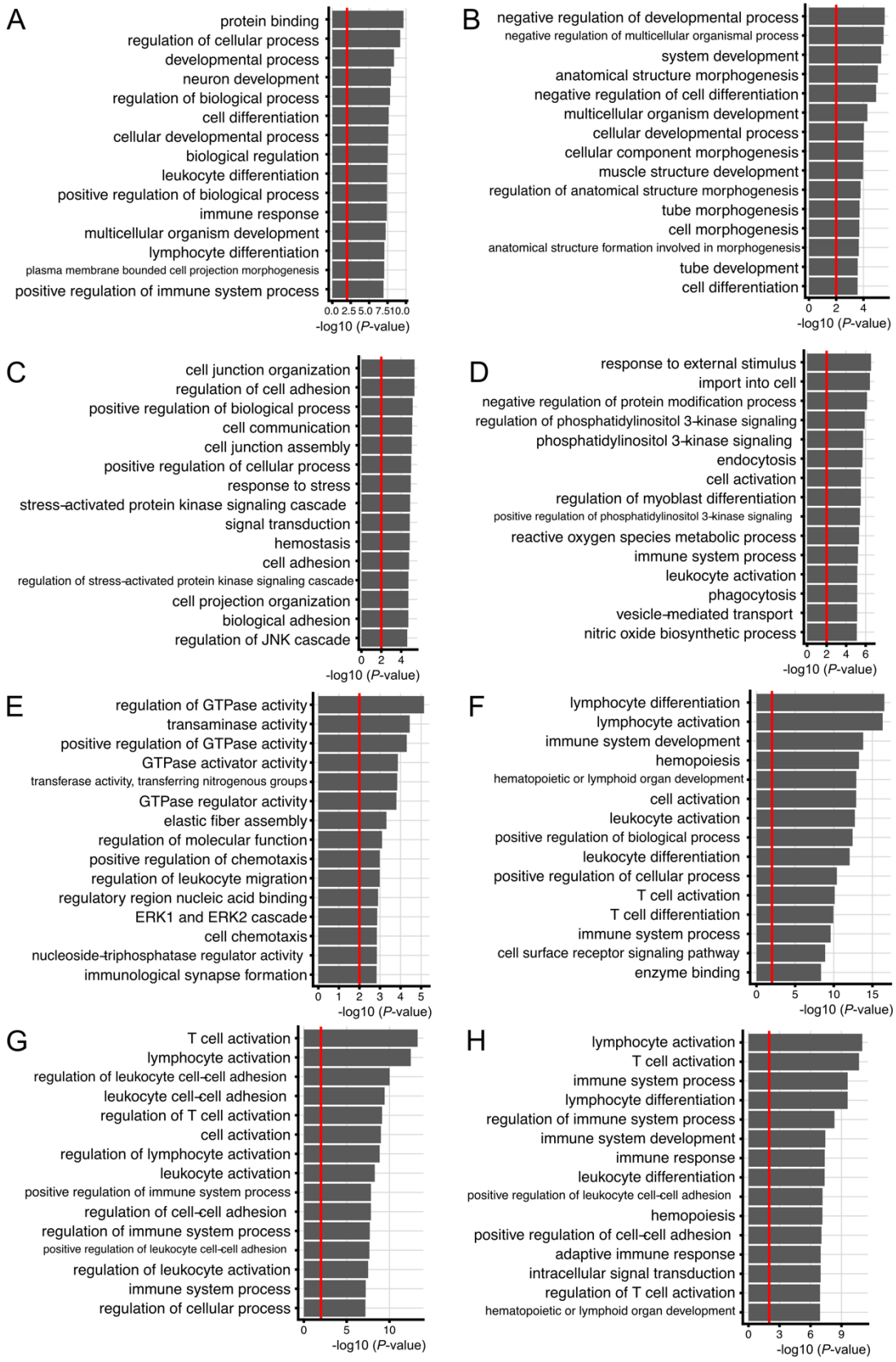
Supplementary Figure-S2: Genomic landscape of adult T-ALL. A. Oncoplot depicting genetic alterations for the targeted panel of genes known to be frequently altered in T-ALL. Each row represents a gene and each column represents a sample. Genes are grouped according to the pathway as annotated to the right. Samples are sorted according to the clusters - indicated by the bottom annotation bar. Cells are color coded for mutation and copy-number status. **B and C.** Lollipop plot for *DNMT3A* (B) and *IDH2* (C). Recurrent R882 and R140 mutational hotspots in *DNMT3A* and *IDH2* are named.

A

	Hyper	Hypo
CD34	4293	3092
ISP	2312	1122
DP_TCR-	592	538
DP_TCR+	166	805
SP4	1520	3628
SP8	2903	3242

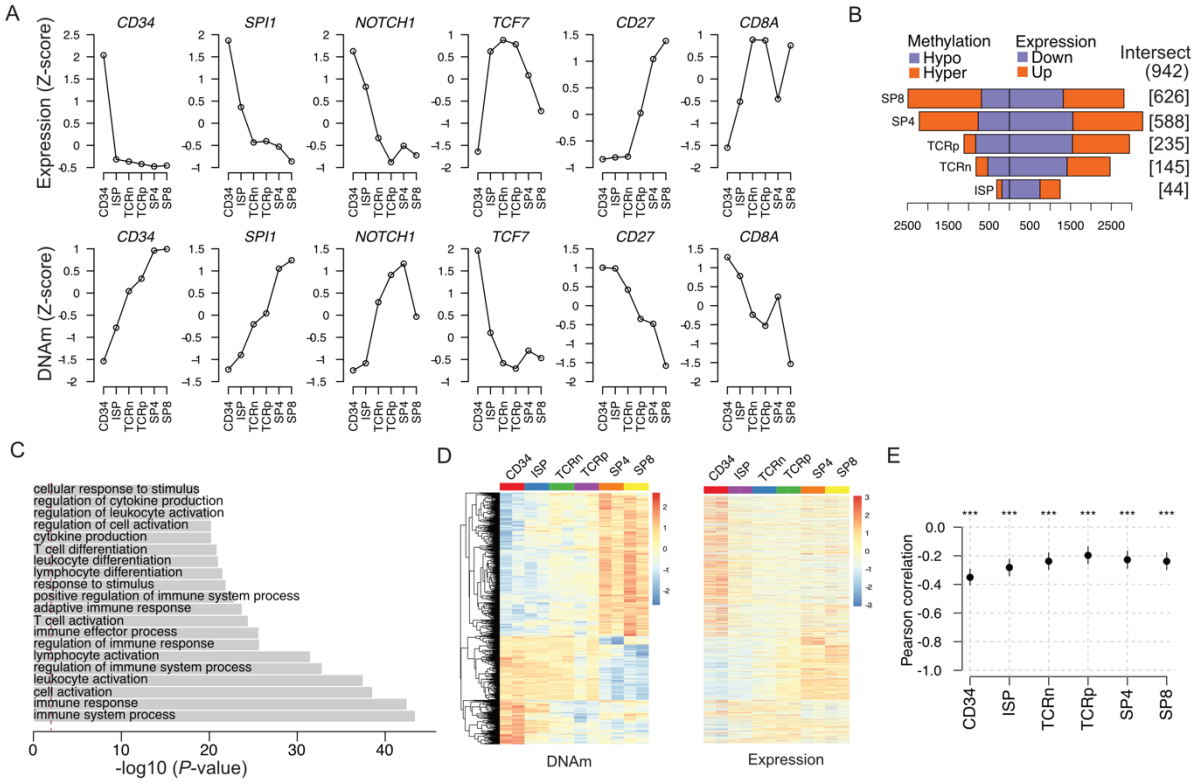


Supplementary Figure-S3: Differentially Methylated Probes associated with T-cell maturation (tDMPs). A. Number of hyper- and hypo methylated tDMPs identified for each thymic cell type at various T-cell maturation stages. **B.** Upset plot showing overlap of tDMPs between T-cell subtypes. **C-F.** Gene ontology terms enriched for tDMPs associated with DP_TCR- (C), DP_TCR+ (D), SP4 (E), and SP8 T-cell subtypes (F). Vertical red bar indicates $P = 0.001$ (Hypergeometric test).

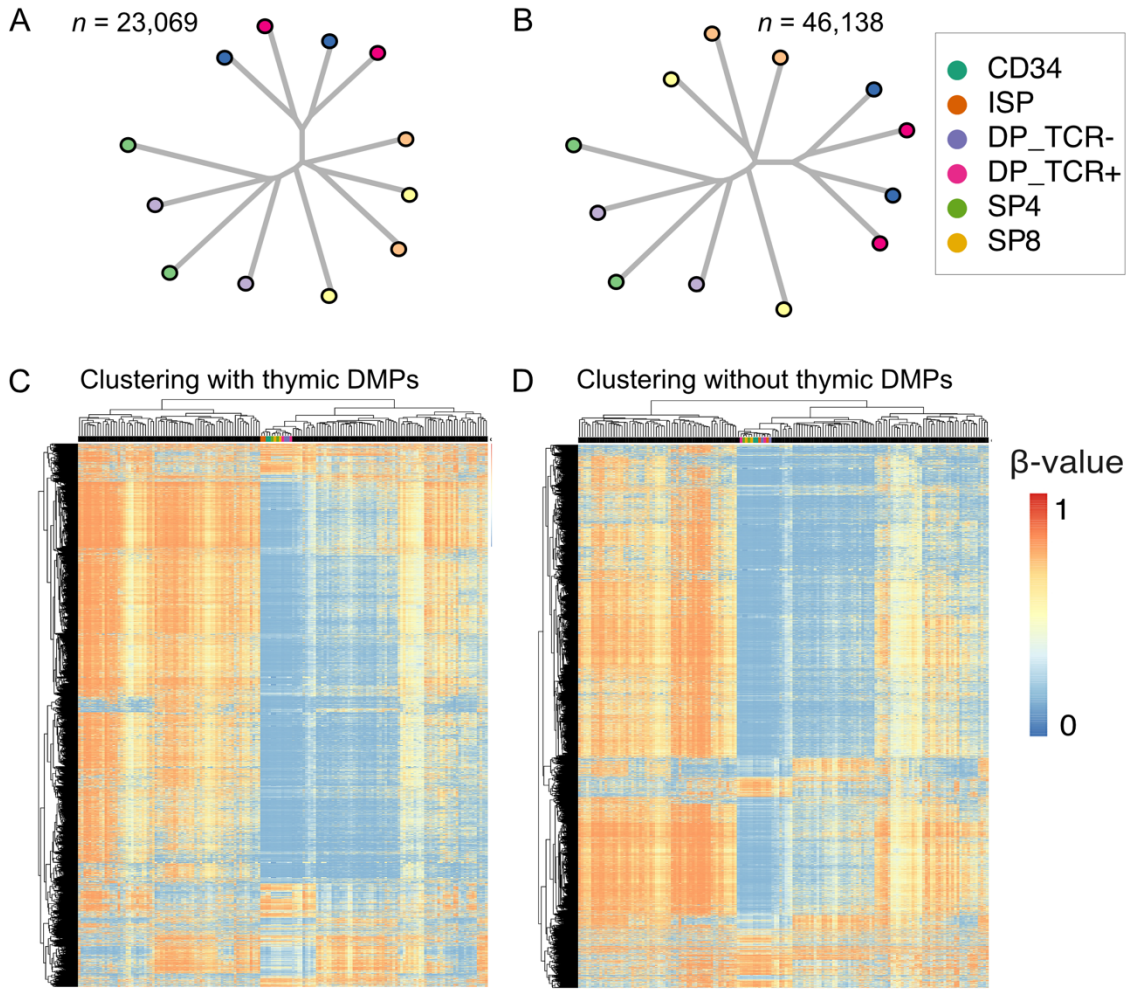

A
protein binding
regulation of cellular process
developmental process
neuron development
regulation of biological process
cell differentiation
cellular developmental process
biological regulation
leukocyte differentiation
positive regulation of biological process
immune response
multicellular organism development
lymphocyte differentiation
plasma membrane bounded cell projection morphogenesis
positive regulation of immune system process
0.0 2.5 5.0 7.5 10.0
-log10 (P-value)
B
negative regulation of developmental process
negative regulation of multicellular organismal process
system development
anatomical structure morphogenesis
negative regulation of cell differentiation
multicellular organism development
cellular developmental process
cellular component morphogenesis
muscle structure development
regulation of anatomical structure morphogenesis
tube morphogenesis
cell morphogenesis
anatomical structure formation involved in morphogenesis
tube development
cell differentiation
0 2 4
-log10 (P-value)
C
cell junction organization
regulation of cell adhesion
positive regulation of biological process
cell communication
cell junction assembly
positive regulation of cellular process
response to stress
stress-activated protein kinase signaling cascade
signal transduction
hemostasis
cell adhesion
regulation of stress-activated protein kinase signaling cascade
cell projection organization
biological adhesion
regulation of JNK cascade
0 2 4
-log10 (P-value)
D
response to external stimulus
import into cell
negative regulation of protein modification process
regulation of phosphatidylinositol 3-kinase signaling
phosphatidylinositol 3-kinase signaling
endocytosis
cell activation
regulation of myoblast differentiation
positive regulation of phosphatidylinositol 3-kinase signaling
reactive oxygen species metabolic process
immune system process
leukocyte activation
phagocytosis
vesicle-mediated transport
nitric oxide biosynthetic process
0 2 4 6
-log10 (P-value)
E
regulation of GTPase activity
transaminase activity
positive regulation of GTPase activity
GTPase activator activity
transferase activity, transferring nitrogenous groups
GTPase regulator activity
elastic fiber assembly
regulation of molecular function
positive regulation of chemotaxis
regulation of leukocyte migration
regulatory region nucleic acid binding
ERK1 and ERK2 cascade
cell chemotaxis
nucleoside-triphosphatase regulator activity
immunological synapse formation
0 1 2 3 4 5
-log10 (P-value)
F
lymphocyte differentiation
lymphocyte activation
immune system development
hemopoiesis
hematopoietic or lymphoid organ development
cell activation
leukocyte activation
positive regulation of biological process
leukocyte differentiation
positive regulation of cellular process
T cell activation
T cell differentiation
immune system process
cell surface receptor signaling pathway
enzyme binding
0 5 10 15
-log10 (P-value)
G
T cell activation
lymphocyte activation
regulation of leukocyte cell-cell adhesion
leukocyte cell-cell adhesion
regulation of T cell activation
cell activation
regulation of lymphocyte activation
leukocyte activation
positive regulation of immune system process
regulation of cell-cell adhesion
regulation of immune system process
positive regulation of leukocyte cell-cell adhesion
regulation of leukocyte activation
immune system process
regulation of cellular process
0 5 10
-log10 (P-value)
H
lymphocyte activation
T cell activation
immune system process
lymphocyte differentiation
regulation of immune system process
immune system development
immune response
leukocyte differentiation
positive regulation of leukocyte cell-cell adhesion
hemopoiesis
positive regulation of cell-cell adhesion
adaptive immune response
intracellular signal transduction
regulation of T cell activation
hematopoietic or lymphoid organ development
0 3 6 9
-log10 (P-value)

1

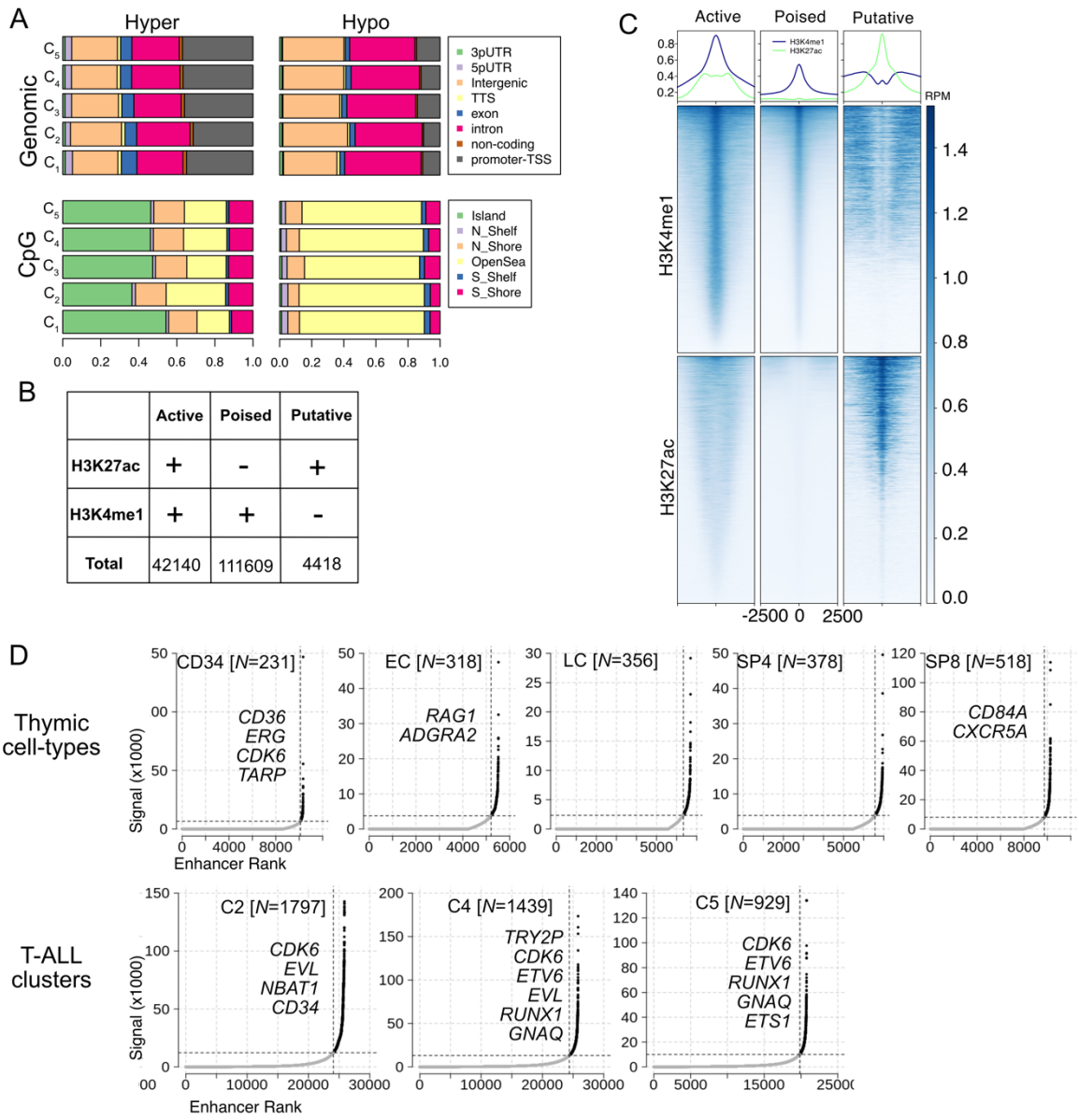
Supplementary Figure-S4: Characterization of Differentially Methylated Probes associated with T-cell maturation (tDMPs) based on methylation dynamics. A-H. Gene ontology terms enriched among tDMPs classified into 8 subgroups (tDMP1-tDMP8 respectively). Vertical red bar indicates $P = 0.001$ (Hypergeometric test).



Supplementary Figure-S5: Gene expression dynamics during Thymopoiesis A. Gene expression profiles of selected marker genes and their corresponding promoter DNAm dynamics associated with the specific thymic cell types. Centered and scaled data are plotted. (Gene expression data obtained from gene expression omnibus dataset with accession number GSE151081). **B.** Number differentially methylated promoters, and differentially expressed genes for each mature thymic cell type compared to CD34 positive ETP cells. Number in parenthesis indicate common genes showing differential expression as well as differential promoter DANm. $N = 942$ indicates unique genes from all 5 cell types. **C.** Gene ontology terms associated with the common genes which show both differential promoter methylation and expression ($N = 942$). **D.** Heatmaps of promoter DNAm and their corresponding gene expression profiles indicate inverse correlation as quantified by Pearson correlation **(E)**. *** $P < 0.001$; Bars indicate 95% confidence interval of correlation coefficient.



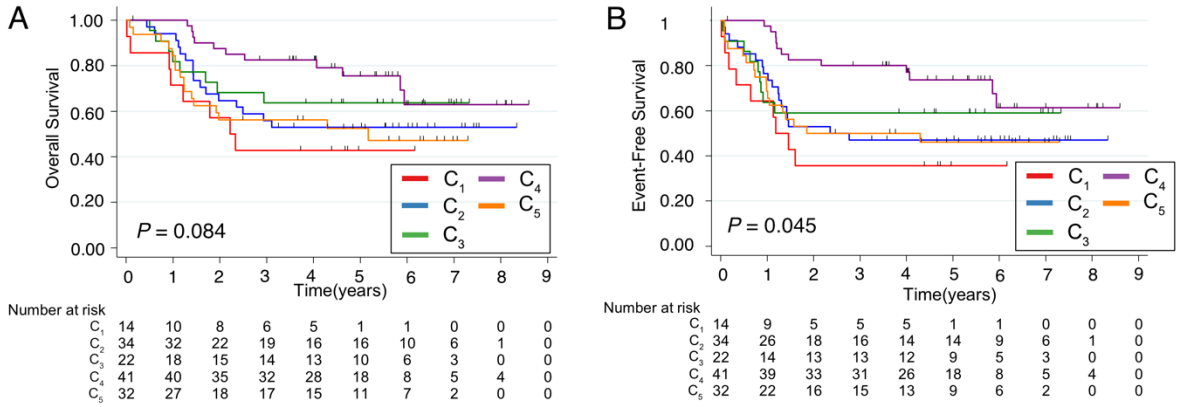
Supplementary Figure-S6: **Influence of thymic maturation associated Differentially Methylated Probes (tDMPs). A.** Unsupervised T-cell phylogenetic tree constructed from randomly selected probes (*n* = 23,069 corresponding to number of unique tDMPs identified during T-cell development) from EPIC arrays fail to infer known T-cell developmental trajectories. **B.** Unsupervised T-cell phylogenetic tree constructed from randomly selected probes (*n* = 46,138 corresponding to twice the number of unique tDMPs identified during T-cell development) from EPIC arrays fail to infer known T-cell developmental trajectories. **C and D.** Unsupervised clustering (top 5% of most variable methylation probes) of normal thymic subpopulations and T-ALL samples with (C) and without (D) tDMPs associated with T-cell maturation respectively. T-ALL samples are annotated in black in top annotation bar.



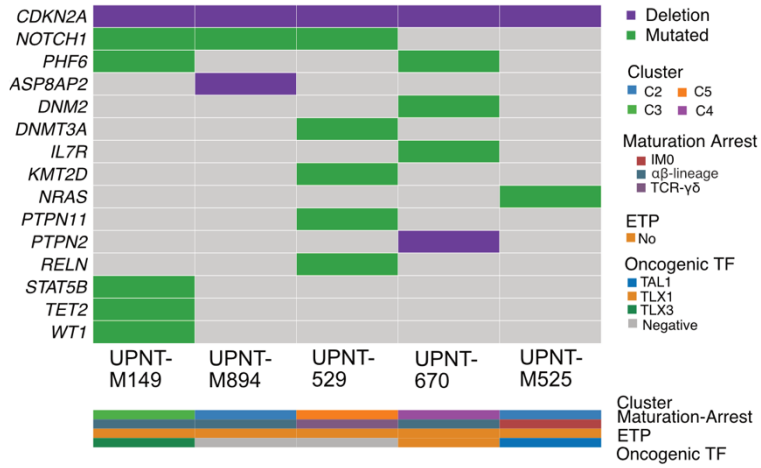
	Active	Poised	Putative
H3K27ac	+	-	+
H3K4me1	+	+	-
Total	42140	111609	4418

Supplementary Figure-S7: Defining normal and disease associated enhancers. A. CpG and genomic annotations of hyper and hypo differentially methylated probes identified in each cluster compared to pooled normal thymic cells. **B.** Number of active, poised, and putative enhancers identified based on the presence or absence of H3K4me1 and H3K27ac histone marks in Thymic cell types. **C.** ChIP-seq binding profiles of H3K4me1 and H3K27ac histone marks for three distinct regulatory genomic regions (Active, Poised and Putative enhancers) in CD34+ thymic cells. Regulatory regions are displayed as 5 kb regions centered around the peak. Top line plot shows average signal density whereas bottom heat maps display ChIP-seq signal for individual peaks. Color gradient reflects the density of ChIP-seq signal. **D.** Super Enhancers (SEs) identified for each thymic subpopulation (top panel), and T-ALL clusters (bottom panel) respectively. Key SE associated genes are highlighted in italics. SEs are highlighted in black whereas typical enhancers

are in gray; EC = Early Committer [$CD3^{-}CD4^{+}CD8^{+}$]; LC = Late Committer [$CD3^{+}CD4^{+}CD8^{+}$]; SP4 = Single CD4 Positive [$CD4^{+}$], SP8 = Single CD8 Positive [$CD8^{+}$].



Supplementary Figure-S8: Overall Survival (OS) and Event Free Survival (EFS) of methylation clusters. A and B. OS (A) and EFS (B) for five epigenetic clusters (C_1-C_5). *P*-values are estimated from log-rank test.



Supplementary Figure-S9: Key genetic alterations and cluster associations of the xenografted samples.

Supplementary tables:

- **Supplementary Table S1:** Significantly mutated genes in every cluster compared to rest of the clusters (Related to Figure 1H). (XLSX)
- **Supplementary Table S2:** Differentially methylated probes identified in each of the normal thymic cell-types (Related to Figure-2). (XLSX)
- **Supplementary Table S3:** Differentially methylated probes identified in each of the normal thymic cell-types classified into 8 subgroups based on methylation dynamics (Related to Figure-2D, E, and Supplementary Figure S4). (XLSX)
- **Supplementary Table S4:** Genes which are both differentially expressed and showing a differential promoter methylation in normal thymic cell types (Related to Supplementary Figure S5B). (XLSX)
- **Supplementary Table S5:** Differentially methylated probes found in each cluster along with the genomic annotations (Related to Figure 3A). (XLSX)
- **Supplementary Table S6:** Super and Typical Enhancers identified in normal thymic cell-types, and in T-ALL clusters (Related to Figure 3D, and Supplementary Figure S7D). (XLSX)
- **Supplementary Table S7:** Significantly enriched TF motifs detected within 100 base pair vicinity of hypo-methylated probes located within the super or typical enhancer regions from T-ALL clusters (Related to Figure 3F). (XLSX)
- **Supplementary Table S8:** Differentially expressed genes in cluster compared to total thymus (Related to Figure 4C). (XLSX)

Univariate and multivariate analysis								
	Univariate				Multivariate			
EFS	HR	95%CI		p	HR	95%CI		p
Age*	1.03	(1.00	; 1.05)	0.019	1.03	(1.01	; 1.06)	0.004
Log(WBC)*	1.28	(0.84	; 1.95)	0.243	-	-	-	-
Prednisone response	0.55	(0.34	; 0.89)	0.016	0.67	(0.37	; 1.21)	0.18
Bone marrow response	0.62	(0.38	; 1.02)	0.06	0.70	(0.39	; 1.26)	0.234
CNS involvement	1.74	(0.95	; 3.20)	0.073	2.11	(1.14	; 3.93)	0.018
C(3+4) (baseline)	1	-	-	-	1	-	-	-
C(1+2)	2.02	(1.14	; 3.58)	0.016	2.19	(1.22	; 3.94)	0.009
C(5)	1.93	(1.02	; 3.66)	0.045	1.37	(0.68	; 2.77)	0.38
OS	HR	95%CI		p	HR	95%CI		p
Age*	1.03	(1.00	; 1.05)	0.018	1.03	(1.01	; 1.06)	0.009
Log(WBC)*	1.26	(0.82	; 1.94)	0.291	-	-	-	-
Prednisone response	0.55	(0.33	; 0.92)	0.024	0.57	(0.33	; 0.99)	0.046
Bone marrow response	0.67	(0.40	; 1.12)	0.129	-	-	-	-
CNS involvement	1.86	(0.98	; 3.51)	0.056	1.92	(1.01	; 3.63)	0.046
C(3+4) (baseline)	1	-	-	-	1	-	-	-
C(1+2)	1.93	(1.06	; 3.52)	0.033	1.92	(1.05	; 3.54)	0.036
C(5)	2.04	(1.05	; 3.97)	0.035	1.56	(0.77	; 3.16)	0.22

Supplementary Table S9: Univariate and Multivariate analyses of EFS and Overall Survival: *continuous variables. Abbreviations: EFS: event-free survival, OS: overall survival, HR: hazard ratio, CI: confidence interval, CNS: central nervous system

GRAALL(2003/05)	Study Cohort (N=143)	Non-investigated (N=194)	*P*-value
Baseline characteristics			
Male	107	132	0.18
Median Age (range) - Years	29.9 (16.3-59.1)	34.1 (16.8-59.5)	0.01
Median WBC ccount (Range)	40.5 (0.9-645.0)	19.8 (0.9-573.0)	<0.001
CNS involvement – no/total (%)	20/143 (14%)	15/194 (8%)	0.07
Outcome characteristics			
Prednisone response – no/total (%)	79/143 (55%)	126/194 (65%%)	0.09
CR -– no/total (%)	134/143 (94%)	181/194 (93%)	0.99
Allo-SCT - – no/total (%)	53/143 (37%)	48/194 (30%)	0.20
5y-EFS (95%CI)	55% (47-63)	57% (50-64)	0.72
5y-OS (95%CI)	60% (51-68)	67% (60-73)	0.30

Supplementary Table S10: Clinical characteristics and outcome of the study cohort versus non-investigated patients (GRAALL-2003/05 trials). WBC, white blood cell count; CNS, central nervous system; CR, complete remission; EFS, event-free survival; OS, overall survival; CI, confidence interval; Allo-SCT, allogeneic stem cell transplantation.