

1 **Prognostic significance of protein-coding and long non-coding RNA expression profile in T-cell**  
2 **acute lymphoblastic leukemia**

3 Deepak Verma<sup>1</sup>, Shruti Kapoor<sup>7</sup>, Disha Sharma<sup>7</sup>, Jay Singh<sup>1</sup>, Gunjan Sharma<sup>6</sup>, Sarita Kumari<sup>1</sup>,  
4 Mohit Arora<sup>6</sup>, Sameer Bakhshi<sup>2</sup>, Rachna Seth<sup>3</sup>, Baibaswata Nayak<sup>5</sup>, Atul Sharma<sup>2</sup>, Raja  
5 Pramanik<sup>2</sup>, Jayanth Kumar Palanichamy<sup>6</sup>, Sridhar Sivasubbu<sup>7</sup>, Vinod Scaria<sup>7</sup>, Rajive Kumar<sup>8</sup>,  
6 Anita Chopra<sup>1\*</sup>

7  
8  
9 <sup>1</sup>Laboratory Oncology, Dr BRAIRCH, All India Institute of Medical Science, New Delhi, India

10 <sup>2</sup>Medical Oncology, Dr BRAIRCH, All India Institute of Medical Science, New Delhi, India

11 <sup>3</sup>Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India

12 <sup>4</sup>Department of Biostatistic, All India Institute of Medical Sciences, New Delhi, India

13 <sup>5</sup>Department of Gastroenterology, All India Institute of Medical Science, New Delhi, India

14 <sup>6</sup>Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India

15 <sup>7</sup>CSIR- Institute of Genomics and Integrative Biology, New Delhi, India

16 <sup>8</sup>Mahavir Cancer Institute & Research Centre, Patna, India

17

18

19 \*Corresponding Author: Anita Chopra, Laboratory Oncology, Dr BRAIRCH, All India Institute  
20 of Medical Science, New Delhi, India

21

22 Email: dranitachopra@aiims.edu , chopraanita2005@gmail.com

23

24 **Abstract**

25 T-acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy associated with  
26 poor outcome. To unravel gene-expression profile of immunophenotypic subtypes of T-ALL, we did  
27 transcriptome analysis in 35 cases. We also analyzed the prognostic relevance of 23 targets: protein-  
28 coding genes, histone modifiers and long non-coding RNA (lncRNA) expression profile, identified on  
29 RNA sequencing, on an independent cohort of 99 T-ALL cases. We found high expression of *MEF2C* to  
30 be associated with prednisolone resistance ( $p=0.048$ ) and CD34 expression ( $p=0.012$ ). *BAALC* expression  
31 was associated with expression of CD34 ( $p=0.032$ ) and myeloid markers ( $p=0.021$ ). *XIST* and *KDM6a*  
32 expression levels were higher in females ( $p=0.047$  and  $0.011$ , respectively). Post-induction minimal  
33 residual disease (MRD) positivity was associated with high lncRNA PCAT18 ( $p=0.04$ ), *HHEX* ( $p=0.027$ )  
34 and *MEF2C* ( $p=0.007$ ). Early thymic precursor (ETP-ALL) immunophenotype was associated with high  
35 expression of *MEF2C* ( $p=0.003$ ), *BAALC* ( $p=0.003$ ), *LYL1* ( $p=0.01$ ), *LYN* ( $p=0.01$ ), *XIST* ( $p=0.02$ ) and

1 low levels of *ST20* ( $p=0.007$ ) and *EML4* ( $p=0.03$ ). On survival analysis, *MEF2C* high expression  
2 emerged as significant predictor of 3-year event free survival (EFS) (low  $71.78\pm 6.58\%$  vs high  
3  $36.57\pm 10.74\%$ , HR 3.5,  $p=0.0003$ ) and overall survival (OS) (low  $94.77\pm 2.96\%$  vs high  $78.75\pm 8.45\%$ ,  
4 HR 4.88,  $p=0.016$ ) in our patients. LncRNA MALAT1 low expression also emerged as predictor inferior  
5 OS (low  $76.02\pm 10.48$  vs high  $94.11\pm 3.31$ , HR 0.22,  $p=0.027$ ).

## 6 **Keywords**

7 RNA-Sequencing, T-cell acute lymphoblastic Leukemia, Early thymic precursor, LncRNA, Gene  
8 expression profile.

## 9 **Introduction**

10 T-lineage acute lymphoblastic leukemia (T-ALL) accounts for ~25% of adults and 10%–15% of pediatric  
11 ALL cases<sup>1</sup>. In the past several years immunological, cytogenetic and genomic approaches have been  
12 extensively utilized for a better understanding of the genomic organization of functional elements and  
13 their alterations in leukemia. This, combined with techniques of cell biology, applied to T-ALL has led to  
14 major advances in our understanding of this disease and has allowed the development of novel therapeutic  
15 approaches.

16 The malignant transformation that culminates in T-ALL is a multi-step process, in which both genetic and  
17 epigenetic alterations of key cellular pathways coordinate to produce the T-ALL phenotype (Peirs et al.,  
18 2015; Girardi et al., 2017). T-ALL pathogenesis is characterized by an abnormality with highly variable  
19 gene expression profiles. The genetic alterations and transcriptomic signatures are often used for  
20 classification of T-ALL patients into molecular subtypes exhibiting distinct clinical outcomes which are  
21 partly determined by their underlying oncogenic signalling pathways, such as *IL7R/JAK/STAT*,  
22 *PI3K/AKT*<sup>5-7</sup> and *RAS/MEK/ERK*<sup>8</sup> signaling. Aberrant expression of a diverse group of transcription  
23 factors such as *LYL1*, *LMO1*, *LMO2*, *TAL1*, *TLX1*, *TLX3*, *HOXA*, *NKX2.1*, *NKX2.2*, *NKX2.5*, *MYC*, *MYB*  
24 and *SPI1* has been reported in distinct T-ALL subtypes<sup>3,4</sup>. Some of the less understood facets in T-ALL

1 include epigenetic deregulation, ribosomal dysfunction, and altered expression of oncogenic miRNAs and  
2 long non-coding RNA. Apart from the genetic alterations, epigenetic alterations such as DNA  
3 methylation, histone modifications, chromosomal remodelling and chromosomal topology have also been  
4 shown to be involved in T-ALL pathogenesis and modulation of clinical response (Van der Meulen et al.,  
5 2014; Kloetgen et al., 2020).

6 More recently, a greater understanding of molecular pathophysiology and immunophenotyping methods  
7 led to the refinement of classification of T-ALL but the clinical relevance of these subtypes remains  
8 controversial<sup>1,9-11</sup>. The only subtype of T-ALL which has got a place in the 2016 revision of WHO, is  
9 early thymic precursor ALL (ETP-ALL)<sup>11,12</sup>. In this study, we analyzed the relevance of T-ALL  
10 immunophenotyping and the expression pattern of protein-coding genes, epigenetic modifiers and long  
11 non-coding RNAs in T-ALL classification and also determined their association with patient prognosis.

## 12 **Materials and Methods**

13 A total of 134 T-ALL cases, newly diagnosed based on morphology, cytochemistry and  
14 immunophenotyping, were enrolled in this study. These cases were immunophenotypically classified into  
15 immature (pro T- and pre T-), cortical and mature T-ALL based on the EGIL criteria<sup>13</sup>. ETP-ALL was  
16 recognized based on the previously defined criteria<sup>12</sup>. The Patients were divided into 2 cohorts: discovery  
17 (n=35) and validation cohort (n=99). Total RNA sequencing was done in the discovery cohort. The  
18 clinical and prognostic relevance of transcriptomic features identified in the discovery cohort were tested  
19 in the validation cohort. All patients or guardians gave informed consent for blood/bone marrow  
20 collection and biological analyses, in agreement with the Declaration of Helsinki. The study was  
21 approved by the Institutional Ethical Committee, All India Institute of Medical Sciences, New Delhi,  
22 India. Transcriptome data of pooled RNA extracted from 5 normal human thymus samples were used as a  
23 non-cancer control (kind courtesy Dr Jan Cools, Belgium).

### 24 ***Discovery cohort***

1 In the discovery cohort (n=35), there were 13 immature (including 5 ETP-ALL), 17 cortical and 5 mature  
2 T-ALL cases with a median age of 14 years (range, 1 to 55 year). There were 26 males (19 pediatrics,  
3 7 adults) and 9 females (7 pediatrics, 2 adults).

4

## 5 **RNA sequencing and analysis of the discovery cohort**

6 RNA was extracted from freshly isolated patient blood samples by the TRIzol method (Thermo Fisher  
7 Scientific, Massachusetts, USA). Transcriptome data of pooled total RNAs from 5 normal human thymus  
8 was used as a control (kindly gifted by Dr Jan Cools, Belgium) for analysis of epigenetic modifiers and  
9 miRNA. Paired-end whole transcriptome sequencing was performed on the Illumina HiSeq2000 platform  
10 using Truseq RNA sample preparation kit (Illumina, San Diego, California, U.S.). Sequence reads were  
11 processed to identify the expression profile of protein-coding and non-coding RNAs by supervised and  
12 unsupervised approaches (**supplementary methods**). To investigate the role of histone modifiers in T-  
13 ALL development, transcript abundance of epigenetic modifiers was measured across mature, cortical,  
14 and immature subtypes of T-ALL (supplementary methods). To know the novel lncRNA transcript, a  
15 computational pipeline combining Open Reading Frame (ORF) prediction coupled with coding potential  
16 calculator (CPC algorithm) to annotate the *protein-coding* potential of transcripts, was used  
17 (supplementary methods). Biomart was utilized to identify the annotated miRNA and transcripts.  
18 Transcripts with  $\geq 2$ -FC were selected as differentially expressed putative coding and non-coding  
19 transcripts in the given subgroup and heatmap were plotted using MeV. Gene Ontology enrichment and  
20 KEGG pathway analyses were used to explore the potential biological processes, cellular components,  
21 and molecular functions of differentially expressed genes (DEG). Gene functional analysis was done on  
22 DAVID 6.8 database<sup>14,15</sup>. To predict the functions of highly overexpressed validated genes for the protein-  
23 coding genes, histone modifiers, lncRNA in all three subgroups of T-ALL, a correlation network analysis  
24 was performed using the Rcorr package in R based on Spearman correlation and visualized using  
25 Cytoscape<sup>16</sup>. Detailed methods are described in the supplementary methods.



1 **Validation cohort:** Candidate genes found differentially expressed in the discovery cohort were further  
2 validated through real-time PCR in an independent cohort of 99 {46 immature (including 15 ETP-ALL),  
3 43 cortical, 10 mature} T-ALL cases including 70 pediatrics (61 males, 9 females), 29 adults (25 males, 4  
4 females). The age of the patients ranged from 3 to 65 years with a median age of 12 years. The expression  
5 levels of 23 targets: *BAALC*, *HHEX*, *MEF2C*, *FAT1*, *LYL1*, *LMO2*, *LYN*, *TAL1*, *DOT1L*, *XIST*, *PCAT18*,  
6 *PCAT14*, *LNC202*, *LNC461*, *LNC648*, *MEF2C-AS1*, *ST20*, *RAG*, *EP300*, *EML4*, *EZH2*, *MALAT1* and  
7 *KDM6A* were compared with patient characteristics and survival. (Supplementary Methods). The  
8 relative quantitation method by real-time PCR was used to calculate the fold change in gene  
9 expression relative to housekeeping gene *ABL1*. Consistently selected genes from discovery and  
10 validation cohort followed the similar distribution of expression. Detailed experimental plan used  
11 to study the discovery and validation cohorts is mentioned in Supplementary Methods.

## 12 **Treatment**

13 In the discovery cohort, 14 patients were treated with ICICLE protocol, 3 with Berlin-Frankfurt-Munster-  
14 90 (BFM-90) protocol, 1 with INCTR, 1 with Holzer's protocol and 3 with hyper CVAD. 13 patients did  
15 not receive treatment/succumbed to the disease before taking treatment.

16 In the validation cohort, seventy-one patients were treated with ICICLE protocol, 15 with Berlin-  
17 Frankfurt-Munster-90 (BFM-90) protocol and 3 with Hyper CVAD. Ten patients did not take any  
18 treatment. Two patients died during induction chemotherapy. Complete remission was defined as bone  
19 marrow blasts <5% with a recovery of blood counts at the end of 4 weeks of induction chemotherapy.  
20 Any failure to do so (including the persistence of leukemic blasts in an extramedullary site), or death  
21 during induction therapy due to any cause, was considered as induction failure. Patients who failed with  
22 one protocol were re-induced with another.

## 23 **Statistical analyses**

1 Fisher's exact test for categorical data and the nonparametric Mann-Whitney U for continuous variables  
2 were used to compare baseline clinical variables across groups in the validation cohort. A P-value  $\leq 0.05$   
3 (two-sided) was considered significant. Event-free survival (EFS) was defined as the time from diagnosis  
4 to the date of the last follow-up in complete remission or the first event (i.e., induction failure, relapse,  
5 secondary neoplasm, or death from any cause). Failure to achieve remission due to non-response was  
6 considered an event at time zero. Survival was defined as the time from diagnosis to death or the last  
7 follow-up. Patients lost to follow-up were censored at the last contact. The last follow up was carried out  
8 on April 2020. The Kaplan-Meier method was used to estimate survival rates, with the differences  
9 compared using a two-sided log-rank test. Univariate and multivariate Cox proportional hazard models  
10 were constructed for EFS and OS. Covariates included sex, WBC ( $<50 \times 10^9/L$ ,  $\geq 50 \times 10^9/L$ ), age ( $<12$   
11 years vs.  $\geq 12$  years), gene expression, immunophenotype and response to prednisolone treatment and  
12 presence of minimal residual disease after the end of induction chemotherapy. Patients with high and low  
13 expression were delineated using maximally selected rank statistics as implemented in the maxstat R  
14 package (<http://cran.r-project.org/web/packages/maxstat/index.html>) for each target (*BAALC*, *HHEX*,  
15 *MEF2C*, *FAT1*, *LYL1*, *LMO2*, *LYN*, *TAL1*, *DOT1L*, *XIST*, *PCAT18*, *PCAT14*, LINC00202, LINC00461,  
16 LINC00648, *MEF2C-AS1*, *ST20*, *RAG1*, *EP300*, *EML4*, *EZH2*, *MALAT1* and *KDM6A*). Statistical  
17 analyzed were performed using the SPSS statistical software package (version 20.0), STATA software  
18 (version 11) and R statistical software (version ).

## 19 **Data Availability**

20 The RNA-Seq raw data that support the findings of this study are available from the corresponding author  
21 upon reasonable request.

22

## 23 **RESULTS**

### 24 **A. Discovery cohort**

#### 25 **Distinct profiling of differentially expressed genes expression among T-ALL immunophenotype**

1 Comparison of gene expression profile among the three different subtypes defined by  
2 immunophenotyping revealed a total of 2,318 genes to be differentially expressed (Figure 2a)  
3 (supplementary1-4). In **immature T-ALL** subtype, transcription factors which control early  
4 hematopoiesis, such as *MEF2C*, *TP63*, *HHEX*, *RUNX2*, *HOXA10*, *HOXA9*, *RUNX1T1* and *ZBTB16*;  
5 homeobox gene like *HOPX*; *LYL1*, *LMO2* and *LYN* (a tyrosine kinase coding gene) were highly  
6 expressed. Interestingly, *LMO2* was previously reported to be expressed under the presence of  
7 *LYL1* gene only but in our data *LMO2* was found upregulated in the absence of significant  
8 overexpression of *LYL1*, a gene reported to be critical for oncogenic functions of *LMO2*. *BAALC*  
9 and *MNI* genes, previously reported in AML<sup>17</sup>, were overexpressed. *BAALC* associated genes like  
10 *IGFBP7* and *PROM1* (CD133), which are known to confer chemoresistance, were also overexpressed<sup>18</sup>.  
11 Also, *MAML3*, *NT5E* (CD73) and *ARID5B* had significant overexpression in immature T-ALL.  
12 Moreover, genes not previously described in T-ALL like *PLD4* and *TP63*, were overexpressed. In  
13 **cortical T-ALL**, the cortical thymocytes - defining *CD1A* gene was overexpressed. Homeobox domain  
14 genes, *NKX2-1*, *TLX1*, *TLX3*, involved in T-cell development were overexpressed. Also, *FAT1*, *FAT3*,  
15 *RAG1*, *EREG*, *CD1C*, *AKAP-2*, *IL-4*, *PRTG*, *TCL-6*, *ZP1* and *TRAV* genes were overexpressed.  
16 Comparison of CD1a+/sCD3+ and CD1a+/sCD3- groups, revealed that *TLX1*, in contrast to reported  
17 literature (Ferrando 2002), was overexpressed in the former group whereas *TLX3* was overexpressed in  
18 the latter. In **mature T-ALL**, *APC2*, *BCL3*, *CCR4*, *CDKN2A*, *EML4*, *HIST1H4G*, *HIST2H2B*, *NCOR2*,  
19 *ST20* and *TRAV22* genes were overexpressed (figure 2a). *TALI* expression was seen in T-ALL cases with  
20 mature immunophenotype. A complete list of DEG is shown in the supplementary file.

## 21 **Unsupervised immunophenotype analysis segregates mixed T-ALL gene expression profile** 22 **into three distinct clusters**

23 Principal component analysis (PCA) was performed to determine variabilities in gene expression profiles  
24 (GEPs) in T-ALL concerning normal thymus tissue (kind courtesy, Dr Jan Cools, Belgium). We observed  
25 that these GEPs could be broadly separated in 3 major clusters (figure 2b), in which **Cluster 1**, **Cluster 2**

1 and **Cluster 3** comprised 5, 9 and 19 cases, respectively. Three samples including one normal thymus  
2 were not present in any cluster. In major cluster 1, all 9 samples had immature T-ALL immunophenotype  
3 and further encompassed 2 sub-clusters consisting of 4 samples each while one remaining sample did not  
4 fall in any of the two sub-clusters. In 1<sup>st</sup> sub-cluster, 3 samples had ETP-ALL *CD5*<sup>-</sup> immunophenotype  
5 and one sample were *CD5*<sup>+</sup> near-ETP-ALL (pre-T-ALL). In 2<sup>nd</sup> sub-cluster, three samples were near-  
6 ETP-ALL (pre/pro-T-ALL) and one sample was ETP-ALL. Major cluster 2 consisted of 5 samples, all  
7 were cortical T-ALL. In major cluster 3, out of 19 samples, 3 were immature, 4 mature and the remaining  
8 were cortical T-ALL. Three samples that did not fall into any of the major clusters consisted of normal  
9 thymus, immature and mature T-ALL, respectively. Interestingly, principal component analysis  
10 placed the mixed T-ALL cases into three distinct categories of T-ALL subclasses.

#### 11 **mRNA expression profiling of epigenetic modifiers in T-ALL**

12 Analysis of genes which are primarily involved in the epigenetic regulation revealed overexpression of  
13 *SETD2*, *ATM*, *ASH1L*, *KDM6A*, *PHF6*, *SUZ12* and *HDAC4* in all subtypes of T-ALL concerning normal  
14 thymic tissue. Furthermore, In immature T-ALL, *HDAC9* and *SMYD3* were overexpressed while *EZH2*  
15 was under-expressed. *HDAC10* was under-expressed in cortical T-ALL. In mature T-ALL, *EP300*,  
16 *PKN1*, *EML4*, *DOT1L* were overexpressed while *HDAC7* was under-expressed in immature and cortical  
17 T-ALL. (figure 3b)

#### 18 **Functional gene annotation and pathway analysis**

19 To determine the biological role of DEG, which were observed during comparison of GEP in different  
20 subtypes, we performed gene ontology analysis. These DEG were strongly involved in the various  
21 biological processes including pathways involved in the inflammatory response, immune response, T cell  
22 co-stimulation, positive regulation of interferon-gamma production, signal transduction, cell-cell  
23 signaling, cell adhesion and migration. Their involvement may lead to abnormal function of various  
24 cellular components comprising plasma membrane, cell surface, extracellular space, an integral  
25 component of the plasma membrane, extracellular region, transcription factor complex, an

1 integral component of the membrane and secretory granules. A complete list of biological processes  
2 and cell components are mentioned in figure 4a,b. We further tried to disintegrate the molecular pathways  
3 by KEGG in which the immunophenotype associated DEG could be involved. This analysis demonstrated  
4 enrichment of cellular pathways such as hematopoietic cell lineage, allograft rejection, transcriptional  
5 dysregulation in cancer, toll-like receptor signaling pathway, graft-versus-host disease, cytokine-cytokine  
6 receptor interaction, cell adhesion molecules (CAMs). Detailed pathways information showed in figure  
7 4c.

### 8 **Co-expression network analysis**

9 Co-expression networks were constructed for differentially expressed genes between the immature,  
10 cortical and mature subgroups (with correlation score $\geq$ 0.9). The genes included in the analysis  
11 were highly expressed in our RNA seq data of T-ALL patients. **Table 1** shows the relation among  
12 the *protein-coding* and *non-coding genes*. Coexpression networks of *BAALC*, *MEF2C*, lncRNA and  
13 epigenetic modifiers are described in supplementary figures 1a-b, 2a-b respectively.

### 14 **Long non-coding RNA (lncRNA)**

#### 15 **RNA-seq analysis post annotation and differential expression analysis resulted in 2,243 lncRNAs**

16 **which were expressed in the T-ALL samples. A total of 223 lncRNAs** were filtered based on the  
17 criteria of  $>2$ FPKM scores. We observed differentially enriched lncRNA in the subgroup of T-ALL such  
18 as *HOTTIP* in immature T-ALL; LINC01221, LINC00202, LINC00461, LINC00648 in cortical T-ALL  
19 and *MALAT1*, *ST20* and *TRBV11* in mature T-ALL, to be overexpressed. Interestingly, these lncRNAs  
20 have not been earlier reported in T-ALL (figure 3a). X-inactive specific transcript (*XIST*), which is known  
21 to have a role in multiple cancers, was expressed in both immature and cortical T-ALL<sup>19</sup>. LUNAR1,  
22 which is known to be a specific *NOTCH1*-regulated lncRNA, was expressed in cortical and mature T-  
23 ALL<sup>20</sup>. The co-expression network for *HOTTIP*, which was highly expressed in immature T-ALL, was  
24 constructed and observed to exhibit a strong positive correlation with important transcription factors,  
25 having a role in/ previously reported to be involved in T-ALL pathogenesis. (supplementary figure 2a). In

1 comparison with normal thymus transcriptome, we found lncRNAs, PCAT14 and PCAT18, to be  
2 significantly overexpressed in T-ALL cases.  
3 In addition to the annotated lncRNA though less explored in T-ALL, we also identified 1,290  
4 novel putative *non-coding* transcripts or lncRNA in our data. All transcripts which overlapped  
5 coding potential were removed from the analysis as these could potentially contribute to false-  
6 positive annotations.

7

### 8 ***Validation cohort***

#### 9 **Correlation between gene expression, response to chemotherapy and outcome**

10 The mRNA expression levels of differentially expressed 23 targets selected from discovery cohort based  
11 on their variable expression pattern in distinct immunophenotype, including protein-coding genes,  
12 epigenetic modifiers and lncRNA transcripts: *BAALC*, *HHEX*, *MEF2C*, *FAT1*, *LYL1*, *LMO2*, *LYN*, *TAL1*,  
13 *DOTIL*, *XIST*, *PCAT18*, *PCAT14*, *LNC202*, *LNC461*, *LNC648*, *MEF2C-AS1*, *ST20*, *RAG*, *EP300*, *EML4*,  
14 *EZH2*, *MALAT1* and *KDM6A*. These genes were further validated in the validation cohort to assess their  
15 clinical significance. Interestingly inconsistency to the results obtained in RNA seq analysis we observed  
16 the similar pattern of overexpression in distinct immunophenotypes in the validation cohort. High  
17 expression of *BAALC* ( $p=0.001$ ), *MEF2C* ( $p=0.002$ ), *LYL1* ( $p=0.018$ ), *HHEX* ( $p=0.007$ ) and low expression  
18 of *EZH2* ( $p=0.005$ ) were found significantly associated with the immature T-ALL immunophenotype.  
19 *RAG1* and *FAT1* expression were higher in cortical T-ALL ( $p=0.004$  and  $0.033$ , respectively). *DOTIL*  
20 expression was higher in mature T-ALL ( $p=0.025$ ). ETP-ALL immunophenotype was associated with  
21 high levels of *BAALC* ( $p=0.003$ ), *MEF2C* ( $p=0.003$ ), *LYL1* ( $p=0.01$ ), *LYN* ( $p=0.01$ ), *XIST* ( $p=0.02$ ) and  
22 lower levels of *ST20* ( $p=0.007$ ) and *EML4* ( $p=0.03$ ). We also observed an association between CD34  
23 positivity on immunophenotyping with expression levels of *BAALC* ( $p=0.032$ ) and *MEF2C* ( $p=0.012$ ).  
24 Myeloid markers (CD13/CD33) expression on immunophenotyping was associated with high *BAALC*  
25 ( $p=0.021$ ) and low *ST20* ( $p=0.007$ ) and low *KDM6A* ( $p=0.026$ ). We did not find any significant

1 association between T-ALL subtype and expression levels of PCAT14, PCAT18, *TAL1*, *LMO2*, *XIST*,  
2 *ST20*, *EP300*, *EML4*, *KDM6A*, LINC00202, LINC00461 and LINC00648. Of the 99 T-ALL patients in  
3 the validation cohort, RNA sample was inadequate quality and quantity was available for the  
4 determination of *MEF2C* in 99; *BAALC*, *HHEX*, *LYL1*, *TAL1*, *FAT1*, *XIST* and *TAL1* in 87 cases; *LMO2*,  
5 *DOT1L* and *LYN* in 78; LINC00648, *PCAT18* and LINC00461 in 72; *MEF2C-AS1*, *PCAT14* and  
6 LINC00202 in 76; *ST20*, *RAG1*, *EP300* and *EML4* in 84; *EZH2* and *KDM6A* in 82 and *MALAT1* in 81  
7 cases due to inadequacy of the samples.

### 8 **Association of protein and non-coding RNAs levels with patient variables**

9 On analysis of the potential association of patient's characteristics with expression levels of protein and  
10 non-coding RNAs, we found an association between *RAG1* expression and age of the patients. *RAG1*  
11 expression was higher in patients <12 years as compared to age  $\geq 12$  years ( $p=0.034$ ). *XIST* and *KDM6A*  
12 expression were higher in females ( $p=0.047$ ; 0.011, respectively). We did not find any association  
13 between WBC count at diagnosis and all parameters tested. Patients with low *XIST* expression and high  
14 *TAL1* more frequently had NCI high risk ( $p=0.01$ ). Prednisolone resistance was associated with high  
15 *MEF2C* expression ( $p=0.048$ ). Post-induction MRD positivity ( $\geq 0.01\%$ ) was associated with high  
16 expression of PCAT18 ( $p=0.04$ ), *HHEX* ( $p=0.027$ ) and *MEF2C* ( $p=0.007$ ). (Table 2)

### 17 **Survival analysis**

18 Complete remission was achieved in 78 (87.64%) patients with induction chemotherapy. Median follow  
19 up was 22 months. The 3 year EFS ( $\pm$ SE) and OS ( $\pm$ SE) was  $62.23 \pm 5.86\%$  and  $90.40 \pm 3.24\%$ ,  
20 respectively. On univariate analysis, we observed high *MEF2C* expression (low  $71.78 \pm 6.58\%$  vs high  
21  $36.57 \pm 10.74$ , HR 3.5, 95% confidence interval 1.67-7.3,  $p=0.0003$ ), high *LYL1* expression (low  
22  $63.14 \pm 6.65\%$  vs  $26.67 \pm 15.9\%$ , HR 2.69, 95% confidence interval 1.08-6.69,  $p=0.029$ ), low *ST20* (low  
23  $43.21 \pm 13.56$  vs high  $61.02 \pm 7.26$ , HR 0.45, 95% confidence interval 0.19-1.03,  $p=0.049$ ), low *RAG1*  
24 expression (low  $41.67 \pm 12.9$  vs high  $61.24 \pm 7.49$ , HR 0.45, 95% confidence interval 0.20-0.99,  $p=0.037$ ),



1 low *EML4* expression (low  $45.29 \pm 8.92$  vs  $83.33 \pm 8.78$ , HR 0.26, 95% confidence interval 0.078-0.88,  
2  $p=0.018$ ) and low *KDM6A* expression (low  $50.48 \pm 7.58$  vs high  $83.33 \pm 8.78$ , HR 0.27, 95% confidence  
3 interval 0.062-1.12,  $p=0.049$ ) were significantly associated with poor 3-year EFS (Table 3). In addition,  
4  $\text{age} \geq 12$  years and ETP-ALL immunophenotype were also associated with poor 3-year EFS (Table 3). We  
5 also found high *MEF2C* expression (low  $94.77 \pm 2.96$  vs high  $78.75 \pm 8.45$ , HR 4.88, 95% confidence  
6 interval 1.16-20.40,  $p=0.016$ ), low *DOTIL* expression (low  $68.38 \pm 3.15$  vs high  $92.64 \pm 3.55$ , HR 0.22,  
7 95% confidence interval 0.06-0.88,  $p=0.019$ ), low *RAG1* expression (low  $75 \pm 10.83$  vs high  $92.54 \pm 3.6$ ,  
8 HR 0.25, 95% confidence interval 0.06-0.98,  $p=0.03$ ) and low MALAT1 expression (low  $76.02 \pm 10.48$  vs  
9 high  $94.11 \pm 3.31$ , HR 0.22, 95% confidence interval 0.048-0.97,  $p=0.027$ ) to be significantly associated  
10 with poor 3-year OS (Table 3). On multivariate analysis for EFS, we found high *MEF2C* expression (HR  
11 3.25,  $p=0.017$ ) to be significantly associated with inferior EFS (Table 4a). We also found *MEF2C*  
12 expression (HR 6.73,  $p=0.04$ ) and low MALAT1 expression (HR 0.16,  $p=0.031$ ) to be significantly  
13 associated with inferior OS (Table 4b).

## 14 **Discussion**

15 With the advancement of molecular techniques, T-ALL has been extensively molecularly characterized.  
16 Although genetically heterogeneous, T-ALL can be categorized into various subtypes based on gene  
17 expression profiles. Although, unlike B-ALL, molecular features in T-ALL have not been utilized for risk  
18 stratification in clinical practice. Furthermore, only a limited number of studies have recently reported  
19 prognostic relevance of lncRNAs and epigenetic modifiers in T-ALL. In this study on 134 T-ALL cases,  
20 we performed high throughput RNA sequencing in the discovery cohort ( $n=35$ ) and identified several  
21 protein-coding and non-coding transcripts which exhibit differential expression among  
22 immunophenotypic subtypes of T-ALL viz. immature, cortical and mature T-ALL. Furthermore, we also  
23 validated the expression of 23 identified targets in T-ALL patients and assessed their clinical significance.  
24 The key genes which served as transcription factors in early hematopoiesis like *MEF2C*, *LYL1*, *LMO2*,  
25 *HHEX*, *RUNX2*, *HOXA10*, *HOXA9*, *RUNX1T1* and *ZBTB16* were upregulated in immature T-ALL.



1 *MEF2C* dysregulation has been previously shown in immature T-ALL<sup>21-27</sup>. Our previous study also  
2 showed the clinical importance of *MEF2C* in predicting prognosis of the ETP-ALL T-ALL patient (Singh  
3 et al 2020). Colomer-Lahiguera S, et al<sup>22</sup>, reported that *MEF2C* dysregulation in T-ALL is associated with  
4 *CDKN1B* deletion and poor response to prednisolone therapy. We also found an association between  
5 prednisolone resistance and high *MEF2C* expression. Starza et al<sup>28</sup> showed its upregulation in T-ALL  
6 cases with interstitial deletion of 5q. Nagel et al<sup>27</sup> proposed distinct mechanisms for aberrant *MEF2C*  
7 gene expression, either by *NKX2-5* signaling or by chromosomal deletion of 5q. They also showed that  
8 *MEF2C* inhibits *BCL2*-regulated apoptosis by inhibition of *NR4A1/NUR77*<sup>27</sup>. In addition to this,  
9 Kawashima-Goto et al, 2015<sup>23</sup>, reported that *BCL2* inhibitors may be useful for treating T-ALL with high  
10 expression levels of *MEF2C*. On network analysis, *MEF2C* gene expression was found to interact with  
11 protein and non-protein-coding partners: *HOPX*, *KIT*, *BAALC*, *HHEX*, *EMP1*, *LYL1*, *BCL*, *SMYD3*,  
12 *HDAC9*, *HOTTIP*, *HOTAIR*, *LINC01021*, *XIST*, *MIR3142HG*, *MIR3132*, *MIR4741*, *SNORD100*,  
13 *SNORD101*.

14 In our study, *MNI*, *BAALC* and *IGFBP7* were overexpressed in immature T-ALL. The upregulation of  
15 these genes is believed to arise from T-cell progenitors retaining myeloid differentiation potential<sup>29-31</sup>.  
16 Previous studies suggest that overexpression of these genes is associated with poor outcome and  
17 resistance to chemotherapy<sup>29,32-35</sup>. Baldus et al, 2007<sup>33</sup>, reported that high *BAALC* expression was  
18 associated with poor long-term survival in T-ALL. On contrary to their observation, we did not find any  
19 significant association between *BAALC* expression and prednisolone sensitivity. Like previous studies, we  
20 found *BAALC* overexpression to be associated with the expression of *CD34* and myeloid markers<sup>29,31</sup>.  
21 However, we did not find any association between *BAALC* expression and patient outcome. We also  
22 found overexpression of *ZBTB16* (*PLZF*) in our patients of immature T-ALL, although not stressed in  
23 previous western studies, was a notable finding in a recently reported study<sup>21,36</sup>. *ZBTB16* (or  
24 promyelocytic leukemia zinc finger, *PLZF*) contains one BTB domain and nine zinc finger motifs. Its  
25 overexpression was shown in this study to be a result of *ZBTB16-ABL1* translocation and occurred in

1 different patients along with other mutations, including *NOTCH1*, *ZEB2*, *PTEN*, *MYCN*, and *PIK3CD*.  
2 Laboratory studies with both *in vitro* and mouse model suggest *ZBTB16-ABL1* be a driver leukemogenic  
3 lesion that causes increased proliferation and a heightened protein tyrosine kinase (PTK) activity that is  
4 amenable to tyrosine kinase inhibitor (TKI) activity<sup>21</sup>. These findings indicate that our *ZBTB16*-  
5 expressing patients have the same translocation. Our finding is of significance also because along with  
6 *LYN* overexpression, T-ALL patients with *ZBTB16* overexpression may also benefit from TKIs.  
7 Apart from these known genes, we identified aberrantly expression of some genes which have not been  
8 reported in T-ALL patients such as *RUNX1T1*, *RUNX2*, *PLD4*, *NT5E (CD73)*, *HOPX*, *TP63*, *HOXA11*-  
9 *AS*. A role for *RUNX2* in T-ALL has been suggested in a study by Nagel et al, 2011, who, to uncover  
10 additional target genes, investigated in detail the aberrant expression of *MEF2C* mediated by complex  
11 deletion at 5q, del(5)(q14) in T-ALL cell line<sup>26</sup>. This could be an evidential proof where *RUNX2* instead  
12 of *RUNX1* could be involved in the manifestation of ETP-ALL that allows *in vivo* functional evaluation  
13 of putative oncogenes and allows preclinical drug testing.  
14 Further, some of the observed differentially expressed gene in cortical T-ALL such as *CD1A*, *CD1C*,  
15 *CD4*, *CFTR*, *FAT3*, *NKX2-1*, *TLX1*, *TLX3* and *RAG1* have been previously reported while we observed  
16 three additional genes, *EREG*, *PAX* and, *ZIC2* to be upregulated in the present study.  
17 Neumann et al 2013, in a study of adult ETP-ALL, showed that cadherins *FAT1* (25%) and *FAT3*  
18 (20%) were mutated, implicating alterations in cell adhesion, and activation of the Wnt  
19 pathway<sup>37</sup>. Neumann et al, 2014<sup>38</sup>, showed that *FAT1* expression was correlated with a more  
20 mature leukemic immunophenotype in T-ALL, with 74% patients with thymic T-ALL being  
21 *FAT1* positive compared with 45% of patients with mature T-ALL and only 4% of early T-ALL  
22 patients. This is in line with our results, as we observed that *FAT1* was associated with cortical  
23 immunophenotype in our study. Like previous study<sup>38</sup>, we did not find any correlation between  
24 *FAT1* expression and patient outcome.

1 Mature T-ALL is a rare subgroup and immunophenotypically diagnosed by CD1a- and sCD3+.

2 Molecularly *TALI* has been identified as a driver gene for late cortical T-ALL<sup>1</sup>. We observed *TALI* be

3 overexpressed in both mature and cortical T-ALL in our study. Among the protein-coding genes, *APC2*,

4 *BCL3*, *CCR4*, *ST20*, *EML4* and *NCOR2* were some of the key upregulated genes.

5 Aberrant histone modifications are the hallmark for cancer and are associated with dysregulated

6 expression of histone modifiers. Therefore, we also studied their expression pattern of epigenetic

7 modifiers to identify a set of histones modifying enzymes to be upregulated or downregulated specifically

8 to the subtype of T-ALL. *EZH2*, a member of the polycomb repressor complex, which is known to be

9 under-expressed in our immature T-ALL cases. This may be related to their previously reported mutations

10 in immature T-ALL<sup>39</sup>. Danis et al<sup>40</sup>, mechanistically linked *EZH2* inactivation to stem-cell-associated

11 transcriptional programs and increased growth/survival signaling, features that convey an adverse

12 prognosis in patients. However, we did not observe the association between *EZH2* expression and clinical

13 outcome. Loss-of-function mutations and deletions in *SETD2* have been shown to lead to chemotherapy

14 tolerance and clonal survival by cell cycle arrest followed by apoptosis. Further, overexpression of

15 *SETD2* has been demonstrated to confer chemotherapy resistance in a variety of cancers including

16 leukemias<sup>39,41-43</sup>. We also observed overexpression of *SETD2* in T-ALL as compared to the normal

17 thymus. Therefore, the potential role of *SETD2* overexpression in therapeutic resistance in T-ALL

18 requires further investigation. In pediatric cases, higher expression of *HDAC7* and *HDAC9* in ALL is

19 associated with poor prognosis. In our study, we observed overexpression of *HDAC9* in our immature and

20 cortical cases. *CREBBP*, *EP300*, *ASH1L*, *ATM*, *PKN1*, *KDM2B*, *KDM4B* and *DOT1L* showed

21 significant differential expression in mature T-ALL. In the context of transcription coactivation,

22 *EP300* and *CREBBP* have lysine acetyltransferase activity<sup>44-47</sup>. Targeted histone lysine

23 acetylation of *EP300* and *CREBBP* can influence chromatin conformation<sup>46</sup>, and concomitant

24 binding of *EP300* and acetylation of H3K27 is a hallmark of promoter or enhancer activation<sup>48</sup>.

1 We also found low expression of *ST20* (figure 5c) and *EML4* (figure 5e) to be associated with  
2 poor EFS. This has not been reported before.

3 *DOT1*-like (*DOTIL*) histone lysine methyltransferase methylates H3K79 and plays an important  
4 role in embryogenesis and hematopoiesis (Ref). Its aberrant activation is associated with acute  
5 leukemias<sup>49,50</sup>, but its function is unknown in T-ALL. *DOTIL* catalytic activity depends on the  
6 mono-ubiquitination of lysine120 in histone H2B (H2BK120Ub), which provides crosstalk  
7 between various histone post-translational modifications<sup>51</sup>. Recent studies suggested the role of  
8 *DOTIL* in H3K79 methylation and mono-ubiquitination of lysine (H2BK120Ub) that may pave  
9 the way for the development of novel *DOTIL*-driven anti-leukemia therapies.<sup>52-56</sup> *DOTIL* was  
10 overexpressed in our mature T-ALL patients and it may be worth investigating if they could be  
11 subjects for *DOTIL*-driven anti-leukemia therapy. We found *DOTIL* low expression to be  
12 associated with poor OS. This has never been reported before.

13 Apart from proteins, non-coding transcript's/ RNA repertoire forms another layer of regulatory paradigm  
14 in normal cell hemostasis. Using RNA-seq, we identified the differentially expressed non-coding RNAs  
15 especially lncRNAs which have been very well documented earlier for their role in cancers. Our analysis  
16 revealed 223 lncRNAs, showing differential expression among various T-ALL subtypes. *NOTCH1*-  
17 regulated lncRNA, LUNAR1, was overexpressed in cortical and mature T-ALL<sup>20</sup>. This may be related to  
18 a higher incidence of activating *NOTCH1* mutations in this T-ALL subtypes<sup>1</sup>. *HOTTIP* and *MEF2C*-  
19 *AS1* were overexpressed in immature; LINC00202, LINC0648, LINC00461 in cortical T-ALL and  
20 *MALAT1* in mature T-ALL. These have not been reported in T-ALL before in the English literature. We  
21 did not find any significant association between their expression with immunophenotypic subtypes.  
22 *HOTTIP* has been reported to be aberrantly activated in AML. It promotes hematopoietic stem cell  
23 renewal leading to AML-like disease in mice.<sup>57</sup> This may explain its overexpression in immature T-ALL  
24 which has myeloid potential in our study. *MALAT1* is known to be involved in a plethora of

1 biological processes ranging from alternative splicing, nuclear organization, epigenetic  
2 regulation of gene expression. It is also associated with various pathological complications like  
3 breast cancer, lung adenocarcinomas, hepatocellular carcinomas, bladder cancers and diabetes  
4 etc<sup>58-60</sup>. Several studies suggest MALAT1 expression as a prognostic marker for various cancer  
5 types<sup>61</sup>. At a molecular level, MALAT1 plays an important role in modulating several signaling  
6 pathways like *MAPK/ERK*, *PI3K/AKT*, *WNT* and *NF-kB* leading to a modification of  
7 proliferation, cell death, cell cycle, migration, invasion, immunity, angiogenesis, and  
8 tumorigenicity. The exact mechanism of how MALAT1 helps in cancer development and  
9 progression is not fully known. MALAT1 can be a therapeutic target, potential diagnostic and  
10 prognosis biomarker for cancers<sup>59,62,63</sup>.

11 Out of the 23 targets tested, we found that *MEF2C* gene expression emerged as a significant predictor of  
12 EFS and OS (figure 5a and 6a). Although *MEF2C* overexpression is associated with chemoresistance and  
13 poor outcome in AML, its prognostic relevance in T-ALL has not been reported to the best of our  
14 knowledge. MALAT1 low expression also emerged as a marker for the poor OS. This has also not been  
15 reported before. Apart from these, the prognostic relevance of other markers like, *ST20* (figure 5c)  
16 *DOT1L*, *RAG1* (figure 5d), *EML4* (figure 5e) and *KDM6a* (figure 5f) should be studied in a larger number  
17 of patients. Previous studies on the utilization of high throughput sequencing and microarray gene  
18 expression have shown that the immature gene signature is associated with inferior survival in T-ALL.  
19 Both of these methods are time, labour and cost-intensive. Based on our results, we recommend *MEF2C*  
20 gene expression analysis by real-time PCR is a reliable and cheap alternative, therefore, can be easily  
21 integrated into routine clinical practice.

22 Taken together, our study provides a comprehensive transcriptional map of coding as well as long  
23 noncoding RNAs. We have identified unique gene signatures that were not discovered in the western  
24 population. This may be related to the ethnic variation of Indian patients. Along with protein-coding  
25 genes, we identified novel as well as known lncRNAs which were differentially expressed in T-ALL

1 patients. Experimental validation and survival analysis for some of the candidates confirmed the RNA-seq  
2 results while co-expression analysis gave an insight into the putative functional roles and pathways  
3 involved in T-ALL. *MEF2C* high expression emerged as a significant predictor of poor EFS and OS.

#### 4 **Acknowledgements**

5 None

#### 6 **Conflict of Interest**

7 The authors declare that there are no competing interests.

#### 8 **Ethics approval and consent to participate**

9 I confirm all relevant ethical guidelines have been followed, and necessary ethical approval approvals  
10 (IESC/T-395/28-11-2014) have been obtained from Institute Ethics Committee For Post Graduate  
11 Research, All India Institute of Medical Sciences New Delhi.

#### 12 **Author Contribution**

13 DV performed the wetlab experiments related to RNA-Seq and also analyzed the data with SKp,  
14 DS, JS, GS, SKr, MA performed the validation experiments and data collection. SB, RS, BN,  
15 AS, RP, JP, RK contributed to project design, clinical data, patient recruitment, and experiment  
16 management. SS, VS supported through highthroughput computing facility. AC is the principal  
17 project investigator.

#### 18 **References**

- 19 1. Belver, L. and A. Ferrando (2016). "The genetics and mechanisms of T cell acute lymphoblastic  
20 leukaemia." *Nature Reviews Cancer*16: 494.
- 21 2. Van Vlierberghe, P., A. Ambesi-Impiombato, K. De Keersmaecker, M. Hadler, E. Paietta, M. S.  
22 Tallman, J. M. Rowe, C. Forne, M. Rue and A. A. Ferrando (2013). "Prognostic relevance of  
23 integrated genetic profiling in adult T-cell acute lymphoblastic leukemia." *Blood*122(1): 74.
- 24 3. Gianni F, Belver L, Ferrando A. The Genetics and Mechanisms of T-Cell Acute Lymphoblastic  
25 Leukemia. *Cold Spring Harb Perspect Med.* 2020 Mar 2;10(3). pii:a035246.

- 1 4. Van Vlierberghe, P. and A. Ferrando (2012). "The molecular basis of T cell acute lymphoblastic  
2 leukemia." *J Clin Invest*122(10): 3398-3406.
- 3 5. Correia, N. C., A. Girio, I. Antunes, L. R. Martins and J. T. Barata (2014). "The multiple layers of  
4 non-genetic regulation of PTEN tumour suppressor activity." *Eur J Cancer*50(1): 216-225.
- 5 6. Fragoso, R. and J. T. Barata (2014). "PTEN and leukemia stem cells." *Advances in Biological  
6 Regulation*56: 22-29.
- 7 7. Fragoso, R. and J. T. Barata (2015). "Kinases, tails and more: Regulation of PTEN function by  
8 phosphorylation." *Methods*77-78: 75-81.
- 9 8. Martelli, A. M., G. Tabellini, F. Ricci, C. Evangelisti, F. Chiarini, R. Bortul, J. A. McCubrey and  
10 F. A. Manzoli (2012). "PI3K/AKT/mTORC1 and MEK/ERK signaling in T-cell acute  
11 lymphoblastic leukemia: New options for targeted therapy." *Advances in Biological  
12 Regulation*52(1): 214-227.
- 13 9. Girardi T, Vicente C, Cools J, De Keersmaecker K. The genetics and molecular biology of T-  
14 ALL. *Blood*. 2017 Mar 2;129(9):1113-1123.
- 15 10. Iacobucci I, Mullighan CG. Genetic Basis of Acute Lymphoblastic Leukemia. *J Clin Oncol*. 2017  
16 Mar 20;35(9):975-983.
- 17 11. Taylor J, Xiao W, Abdel-Wahab O. Diagnosis and classification of hematologic malignancies on  
18 the basis of genetics. *Blood*. 2017 Jul 27;130(4):410-423.
- 19 12. Coustan-Smith, E., C. G. Mullighan, M. Onciu, F. G. Behm, S. C. Raimondi, D. Pei, C. Cheng,  
20 X. Su, J. E. Rubnitz, G. Basso, A. Biondi, C.-H. Pui, J. R. Downing and D. Campana (2009).  
21 "Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia."  
22 *The Lancet Oncology*10(2): 147-156.
- 23 13. Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A, van't Veer MB. Proposals for  
24 the immunological classification of acute leukemias. European Group for the Immunological  
25 Characterization of Leukemias (EGIL). *Leukemia*. 1995 Oct;9(10):1783-6.

- 1 14. Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Bioinformatics enrichment tools:  
2 paths toward the comprehensive functional analysis of large gene lists." *Nucleic Acids Res*37(1):  
3 1-13.
- 4 15. Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of  
5 large gene lists using DAVID bioinformatics resources." *Nat Protoc*4(1): 44-57.
- 6 16. Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B.  
7 Schwikowski and T. Ideker (2003). "Cytoscape: a software environment for integrated models of  
8 biomolecular interaction networks." *Genome Res*13(11): 2498-2504.
- 9 17. Haferlach C, Kern W, Schindela S, Kohlmann A, Alpermann T, Schnittger S,Haferlach T. Gene  
10 expression of BAALC, CDKN1B, ERG, and MN1 adds independent prognostic information to  
11 cytogenetics and molecular mutations in adult acute myeloid leukemia. *Genes Chromosomes*  
12 *Cancer*. 2012 Mar;51(3):257-65.
- 13 18. Heesch, S., C. Schlee, M. Neumann, A. Stroux, A. Kuhn, S. Schwartz, T. Haferlach, N.  
14 Goekbuget, D. Hoelzer, E. Thiel, W. K. Hofmann and C. D. Baldus (2010). "BAALC-associated  
15 gene expression profiles define IGFBP7 as a novel molecular marker in acute leukemia."  
16 *Leukemia*24(8): 1429-1436.
- 17 19. Yin S, Dou J, Yang G, Chen F. Long non-coding RNA XIST expression as a prognostic factor in  
18 human cancers: A meta-analysis. *Int J Biol Markers*. 2019 Dec;34(4):327-333.
- 19 20. Trimarchi T, Bilal E, Ntziachristos P, Fabbri G, Dalla-Favera R, Tsigirgos A,Aifantis I. Genome-  
20 wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia.  
21 *Cell*. 2014 Jul 31;158(3):593-606.
- 22 21. Chen B, Jiang L, Zhong ML, Li JF, Li BS, Peng LJ, Dai YT, Cui BW, Yan TQ,Zhang WN, Weng  
23 XQ, Xie YY, Lu J, Ren RB, Chen SN, Hu JD, Wu DP, Chen Z, Tang JY,Huang JY, Mi JQ, Chen  
24 SJ. Identification of fusion genes and characterization of transcriptome features in T-cell acute  
25 lymphoblastic leukemia. *Proc Natl Acad Sci U S A*. 2018 Jan 9;115(2):373-378.



- 1 22. Colomer-Lahiguera S, Pisecker M, König M, Nebral K, Pickl WF, Kauer MO, Haas OA,  
2 Ullmann R, Attarbaschi A, Dworzak MN, Strehl S. *MEF2C*-dysregulated pediatric T-cell acute  
3 lymphoblastic leukemia is associated with *CDKN1B* deletions and a poor response to  
4 glucocorticoid therapy. *Leuk Lymphoma*. 2017 Dec;58(12):2895-2904.
- 5 23. Kawashima-Goto S, Imamura T, Tomoyasu C, Yano M, Yoshida H, Fujiki A, Tamura S, Osone  
6 S, Ishida H, Morimoto A, Kuroda H, Hosoi H. BCL2 Inhibitor (ABT-737): A Restorer of  
7 Prednisolone Sensitivity in Early T-Cell Precursor-Acute Lymphoblastic Leukemia with High  
8 *MEF2C* Expression? *PLoS One*. 2015 Jul 14;10(7):e0132926.
- 9 24. Zuurbier L, Gutierrez A, Mullighan CG, Canté-Barrett K, Gevaert AO, de Rooi J, Li Y, Smits  
10 WK, Buijs-Gladdines JG, Sonneveld E, Look AT, Horstmann M, Pieters R, Meijerink JP.  
11 Immature *MEF2C*-dysregulated T-cell leukemia patients have an early T-cell precursor acute  
12 lymphoblastic leukemia gene signature and typically have non-rearranged T-cell receptors.  
13 *Haematologica*. 2014 Jan;99(1):94-102.
- 14 25. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, Vuerhard M, Buijs-  
15 Gladdines J, Kooi C, Klous P, van Vlierberghe P, Ferrando AA, Cayuela JM, Verhaaf B, Beverloo  
16 HB, Horstmann M, de Haas V, Wiekmeijer AS, Pike-Overzet K, Staal FJ, de Laat W, Soulier J,  
17 Sigaux F, Meijerink JP. Integrated transcript and genome analyses reveal *NKX2-1* and *MEF2C* as  
18 potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell*. 2011 Apr 12;19(4):484-  
19 97.
- 20 26. Nagel S, Venturini L, Meyer C, Kaufmann M, Scherr M, Drexler HG, Macleod RA.  
21 Transcriptional deregulation of oncogenic myocyte enhancer factor 2C in T-cell acute  
22 lymphoblastic leukemia. *Leuk Lymphoma*. 2011 Feb;52(2):290-7.
- 23 27. Nagel S, Meyer C, Quentmeier H, Kaufmann M, Drexler HG, MacLeod RA. *MEF2C* is activated  
24 by multiple mechanisms in a subset of T-acute lymphoblastic leukemia cell lines. *Leukemia*.  
25 2008 Mar;22(3):600-7.

- 1 28. La Starza R, Barba G, Demeyer S, Pierini V, Di Giacomo D, Gianfelici V, Schwab C, Matteucci  
2 C, Vicente C, Cools J, Messina M, Crescenzi B, Chiaretti S, Foà R, Basso G, Harrison CJ,  
3 Mecucci C. Deletions of the long arm of chromosome 5 define subgroups of T-cell acute  
4 lymphoblastic leukemia. *Haematologica*. 2016 Aug;101(8):951-8.
- 5 29. Heesch S, Bartram I, Neumann M, Reins J, Mossner M, Schlee C, Stroux A, Haferlach T,  
6 Goekbuget N, Hoelzer D, Hofmann WK, Thiel E, Baldus CD. Expression of IGFBP7 in acute  
7 leukemia is regulated by DNA methylation. *Cancer Sci*. 2011 Jan;102(1):253-9.
- 8 30. Neumann M, Heesch S, Gökbüget N, Schwartz S, Schlee C, Benlasfer O, Farhadi-Sartangi N,  
9 Thibaut J, Burmeister T, Hoelzer D, Hofmann WK, Thiel E, Baldus CD. Clinical and molecular  
10 characterization of early T-cell precursor leukemia: a high-risk subgroup in adult T-ALL with a  
11 high frequency of FLT3 mutations. *Blood Cancer J*. 2012 Jan;2(1):e55.
- 12 31. Heesch S, Schlee C, Neumann M, Stroux A, Kühnl A, Schwartz S, Haferlach T, Goekbuget N,  
13 Hoelzer D, Thiel E, Hofmann WK, Baldus CD. *BAALC*-associated gene expression profiles  
14 define IGFBP7 as a novel molecular marker in acute leukemia. *Leukemia*. 2010 Aug;24(8):1429-  
15 36.
- 16 32. Carow, C. E., M. Levenstein, S. H. Kaufmann, J. Chen, S. Amin, P. Rockwell, L. Witte, M. J.  
17 Borowitz, C. I. Civin and D. Small (1996). "Expression of the hematopoietic growth factor  
18 receptor FLT3 (STK- 1/Flk2) in human leukemias." *Blood*87(3): 1089.
- 19 33. Baldus, C. D., P. Martus, T. Burmeister, S. Schwartz, N. Gokbuget, C. D. Bloomfield, D.  
20 Hoelzer, E. Thiel and W. K. Hofmann (2007). "Low ERG and BAALC expression identifies a  
21 new subgroup of adult acute T-lymphoblastic leukemia with a highly favorable outcome." *J Clin*  
22 *Oncol*25(24): 3739-3745.
- 23 34. Baldus, C. D., S. M. Tanner, D. F. Kusewitt, S. Liyanarachchi, C. Choi, M. A. Caligiuri, C. D.  
24 Bloomfield and A. d. l. Chapelle (2003). "BAALC, a novel marker of human hematopoietic  
25 progenitor cells." *Experimental Hematology*31(11): 1051-1056.

- 1 35. Heuser, M., B. Argiropoulos, F. Kuchenbauer, E. Yung, J. Piper, S. Fung, R. F. Schlenk, K.  
2 Dohner, T. Hinrichsen, C. Rudolph, A. Schambach, C. Baum, B. Schlegelberger, H. Dohner, A.  
3 Ganser and R. K. Humphries (2007). "MN1 overexpression induces acute myeloid leukemia in  
4 mice and predicts ATRA resistance in patients with AML." *Blood* 110(5): 1639.
- 5 36. Ferrando, A. A., D. S. Neuberg, J. Staunton, M. L. Loh, C. Huard, S. C. Raimondi, F. G. Behm,  
6 C.-H. Pui, J. R. Downing, D. G. Gilliland, E. S. Lander, T. R. Golub and A. T. Look (2002).  
7 "Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic  
8 leukemia." *Cancer Cell* 1(1): 75-87.
- 9 37. Neumann, M., E. Coskun, L. Fransecky, L. H. Mochmann, I. Bartram, N. F. Sartangi, S. Heesch,  
10 N. Gokbuget, S. Schwartz, C. Brandts, C. Schlee, R. Haas, U. Dührsen, M. Griesshammer, H.  
11 Dohner, G. Ehninger, T. Burmeister, O. Blau, E. Thiel, D. Hoelzer, W. K. Hofmann and C. D.  
12 Baldus (2013). "FLT3 mutations in early T-cell precursor ALL characterize a stem cell like  
13 leukemia and imply the clinical use of tyrosine kinase inhibitors." *PLoS One* 8(1): e53190.
- 14 38. Neumann, M., M. Seehawer, C. Schlee, S. Vosberg, S. Heesch, E. K. von der Heide, A. Graf, S.  
15 Krebs, H. Blum, N. Gökbuget, S. Schwartz, D. Hoelzer, P. A. Greif and C. D. Baldus (2014).  
16 "FAT1 expression and mutations in adult acute lymphoblastic leukemia." *Blood Cancer Journal* 4:  
17 e224.
- 18 39. Zhang, J., L. Ding, L. Holmfeldt, G. Wu, S. L. Heatley, D. Payne-Turner, J. Easton, X. Chen, J.  
19 Wang, M. Rusch, C. Lu, S. C. Chen, L. Wei, J. R. Collins-Underwood, J. Ma, K. G. Roberts, S.  
20 B. Pounds, A. Ulyanov, J. Becksfort, P. Gupta, R. Huether, R. W. Kriwacki, M. Parker, D. J.  
21 McGoldrick, D. Zhao, D. Alford, S. Espy, K. C. Bobba, G. Song, D. Pei, C. Cheng, S. Roberts,  
22 M. I. Barbato, D. Campana, E. Coustan-Smith, S. A. Shurtleff, S. C. Raimondi, M. Kleppe, J.  
23 Cools, K. A. Shimano, M. L. Hermiston, S. Doulatov, K. Eppert, E. Laurenti, F. Notta, J. E. Dick,  
24 G. Basso, S. P. Hunger, M. L. Loh, M. Devidas, B. Wood, S. Winter, K. P. Dunsmore, R. S.  
25 Fulton, L. L. Fulton, X. Hong, C. C. Harris, D. J. Dooling, K. Ochoa, K. J. Johnson, J. C.  
26 Obenauer, W. E. Evans, C. H. Pui, C. W. Naeve, T. J. Ley, E. R. Mardis, R. K. Wilson, J. R.

- 1           Downing and C. G. Mullighan (2012). "The genetic basis of early T-cell precursor acute  
2           lymphoblastic leukaemia." *Nature*481(7380): 157-163.
- 3           40. Danis E, Yamauchi T, Echanique K, Zhang X, Haladyna JN, Riedel SS, Zhu N, Xie H, Orkin SH,  
4           Armstrong SA, Bernt KM, Neff T. *EZH2* Controls an Early Hematopoietic Program and Growth  
5           and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia. *Cell Rep*. 2016  
6           Mar 1;14(8):1953-65.
- 7           41. Mar, B. G., L. B. Bullinger, K. M. McLean, P. V. Grauman, M. H. Harris, K. Stevenson, D. S.  
8           Neuberg, A. U. Sinha, S. E. Sallan, L. B. Silverman, A. L. Kung, L. Lo Nigro, B. L. Ebert and S.  
9           A. Armstrong (2014). "Mutations in epigenetic regulators including SETD2 are gained during  
10           relapse in paediatric acute lymphoblastic leukaemia." *Nat Commun*5: 3469.
- 11           42. Zhu, X., F. He, H. Zeng, S. Ling, A. Chen, Y. Wang, X. Yan, W. Wei, Y. Pang, H. Cheng, C.  
12           Hua, Y. Zhang, X. Yang, X. Lu, L. Cao, L. Hao, L. Dong, W. Zou, J. Wu, X. Li, S. Zheng, J.  
13           Yan, J. Zhou, L. Zhang, S. Mi, X. Wang, L. Zhang, Y. Zou, Y. Chen, Z. Geng, J. Wang, J. Zhou,  
14           X. Liu, J. Wang, W. Yuan, G. Huang, T. Cheng and Q.-f. Wang (2014). "Identification of  
15           functional cooperative mutations of SETD2 in human acute leukemia." *Nature Genetics*46: 287.
- 16           43. Wang, S., X. Yuan, Y. Liu, K. Zhu, P. Chen, H. Yan, D. Zhang, X. Li, H. Zeng, X. Zhao, X.  
17           Chen, G. Zhou and S. Cao (2019). "Genetic polymorphisms of histone methyltransferase SETD2  
18           predicts prognosis and chemotherapy response in Chinese acute myeloid leukemia patients." *J*  
19           *Transl Med*17(1): 101.
- 20           44. Ait-Si-Ali, S., S. Ramirez, F. X. Barre, F. Dkhissi, L. Magnaghi-Jaulin, J. A. Girault, P. Robin,  
21           M. Knibiehler, L. L. Pritchard, B. Ducommun, D. Trouche and A. Harel-Bellan (1998). "Histone  
22           acetyltransferase activity of CBP is controlled by cycle-dependent kinases and oncoprotein E1A."  
23           *Nature*396(6707): 184-186.
- 24           45. Iyer, N. G., H. Ozdag and C. Caldas (2004). "p300/CBP and cancer." *Oncogene*23(24): 4225-  
25           4231.

- 1 46. Liu, X., L. Wang, K. Zhao, P. R. Thompson, Y. Hwang, R. Marmorstein and P. A. Cole (2008).  
2 "The structural basis of protein acetylation by the p300/CBP transcriptional coactivator."  
3 Nature451(7180): 846-850.
- 4 47. Qian, M., H. Zhang, S. K. Kham, S. Liu, C. Jiang, X. Zhao, Y. Lu, C. Goodings, T. N. Lin, R.  
5 Zhang, T. Moriyama, Z. Yin, Z. Li, T. C. Quah, H. Ariffin, A. M. Tan, S. Shen, D. Bhojwani, S.  
6 Hu, S. Chen, H. Zheng, C. H. Pui, A. E. Yeoh and J. J. Yang (2017). "Whole-transcriptome  
7 sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant  
8 genomic abnormalities of *EP300* and *CREBBP*." Genome Res27(2): 185-195.
- 9 48. Vo, N. and R. H. Goodman (2001). "CREB-binding protein and p300 in transcriptional  
10 regulation." J Biol Chem276(17): 13505-13508.
- 11 49. Bernt, Kathrin M., N. Zhu, Amit U. Sinha, S. Vempati, J. Faber, Andrei V. Krivtsov, Z. Feng, N.  
12 Punt, A. Daigle, L. Bullinger, Roy M. Pollock, Victoria M. Richon, Andrew L. Kung and  
13 Scott A. Armstrong (2011). "MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79  
14 Methylation by DOT1L." Cancer Cell20(1): 66-78.
- 15 50. Zhang, Y. and T. G. Kutateladze (2019). "Methylation of Histone H3K79 by Dot1L Requires  
16 Multiple Contacts with the Ubiquitinated Nucleosome." Mol Cell74(5): 862-863.
- 17 51. McGinty, R. K., J. Kim, C. Chatterjee, R. G. Roeder and T. W. Muir (2008). "Chemically  
18 ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation."  
19 Nature453: 812.
- 20 52. Anderson, C. J., M. R. Baird, A. Hsu, E. H. Barbour, Y. Koyama, M. J. Borgnia and R. K.  
21 McGinty (2019). "Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L  
22 Methyltransferase." Cell Rep26(7): 1681-1690 e1685.
- 23 53. Jang, S., C. Kang, H. S. Yang, T. Jung, H. Hebert, K. Y. Chung, S. J. Kim, S. Hohng and J. J.  
24 Song (2019). "Structural basis of recognition and destabilization of the histone H2B ubiquitinated  
25 nucleosome by the DOT1L histone H3 Lys79 methyltransferase." Genes Dev33(11-12): 620-625

- 1 54. Valencia-Sanchez, M. I., P. De Ioannes, M. Wang, N. Vasilyev, R. Chen, E. Nudler, J. P.  
2 Armache and K. J. Armache (2019). "Structural Basis of Dot1L Stimulation by Histone H2B  
3 Lysine 120 Ubiquitination." *Mol Cell*74(5): 1010-1019 e1016.
- 4 55. Worden, E. J., N. A. Hoffmann, C. W. Hicks and C. Wolberger (2019). "Mechanism of Cross-  
5 talk between H2B Ubiquitination and H3 Methylation by Dot1L." *Cell*176(6): 1490-1501 e1412.
- 6 56. Yao, T., W. Jing, Z. Hu, M. Tan, M. Cao, Q. Wang, Y. Li, G. Yuan, M. Lei and J. Huang (2019).  
7 "Structural basis of the crosstalk between histone H2B monoubiquitination and H3 lysine 79  
8 methylation on nucleosome." *Cell Res*29(4): 330-333.
- 9 57. Luo H, Zhu G, Xu J, Lai Q, Yan B, Guo Y, Fung TK, Zeisig BB, Cui Y, Zha J, Cogle C, Wang F,  
10 Xu B, Yang FC, Li W, So CWE, Qiu Y, Xu M, Huang S. HOTTIP lncRNA Promotes  
11 Hematopoietic Stem Cell Self-Renewal Leading to AML-like Disease in Mice. *Cancer Cell*. 2019  
12 Dec 9;36(6):645-659.e8.
- 13 58. Yoshimoto, R., A. Mayeda, M. Yoshida and S. Nakagawa (2016). "MALAT1 long non-coding  
14 RNA in cancer." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*1859(1):  
15 192-199.
- 16 59. Jianghua Liu, W.-X., Peng, Yin-Yuan Mo, Dianzhong Luo (2017). "MALAT1-mediated  
17 tumorigenesis." *Frontiers in Bioscience*.
- 18 60. Sun, W., Y. Yang, C. Xu and J. Guo (2017). "Regulatory mechanisms of long noncoding RNAs  
19 on gene expression in cancers." *Cancer Genetics*216: 105-110.
- 20 61. Wei, Y. and B. Niu (2015). "Role of MALAT1 as a Prognostic Factor for Survival in Various  
21 Cancers: A Systematic Review of the Literature with Meta-Analysis." *Disease markers*2015:  
22 164635-164635.
- 23 62. Zhao, M., S. Wang, Q. Li, Q. Ji, P. Guo and X. Liu (2018). "MALAT1: A long non-coding RNA  
24 highly associated with human cancers." *Oncology letters*16(1): 19-26.
- 25 63. Li, Z. X., Q. N. Zhu, H. B. Zhang, Y. Hu, G. Wang and Y. S. Zhu (2018). "MALAT1: a potential  
26 biomarker in cancer." *Cancer Manag Res*10(1179-1322 (Print)): 6757-6768.

1 **Legends to Tables and figures**

2 **Table 1:** Co-expression network analysis of few selected protein-coding and non-coding genes

3 **Table 2:** Association of expression of various protein and non-protein coding RNAs with  
4 patients' characteristics

5 **Table 3:** Association of patients' characteristics with EFS and OS

6 **Table 4: (a)** Multivariate Cox regression analysis for EFS **(b)** Multivariate Cox regression  
7 analysis of OS

8 **Figure 1:** Workflow for identification and validation of molecular markers in T-ALL patients.

9 **Figure 2: (a)** Radar graph of Differentially expressed genes among T-ALL subgroups: immature  
10 (Blue), cortical (Green), mature (Gray). each circle represents the fold change for the  
11 differentially expressed genes. **(b)** PCA analysis of Gene expression: Figure depicts the  
12 clustering of samples in 3 major clusters according to their normalized FPKM count. Cluster 1  
13 comprises of 9, Cluster 2 comprises 5 and Cluster 3 comprises 19 T-ALL samples.

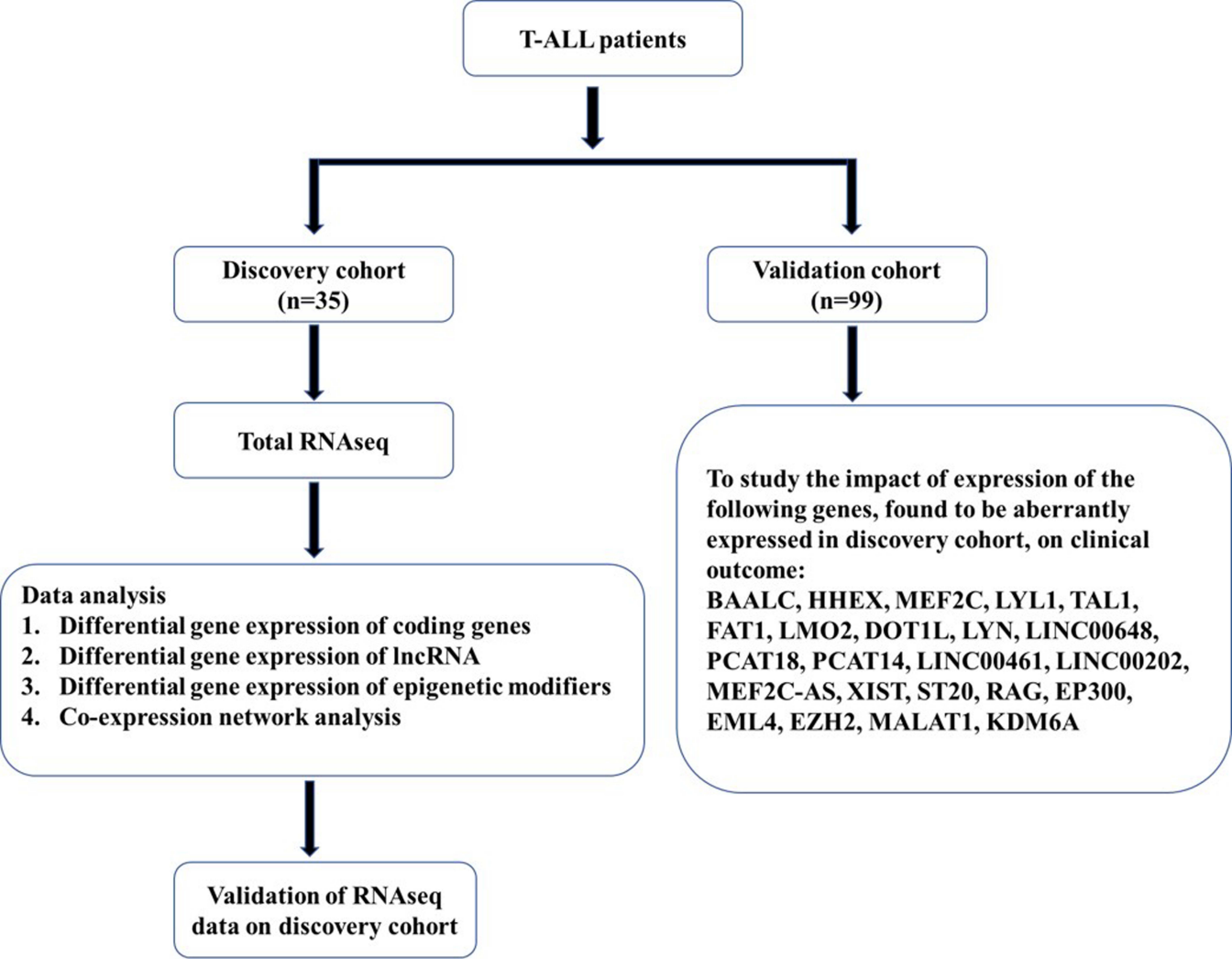
14 **Figure 3: (a)** Hierarchical clustering of differential expressed lncRNA among three T-ALL  
15 subgroups. **(b)** Hierarchical clustering differentially expressed epigenetic modifiers in T-ALL.;  
16 Each row represents gene symbol. The "heat map" indicates high (red) or low (green) level of  
17 expression according to the scale shown.

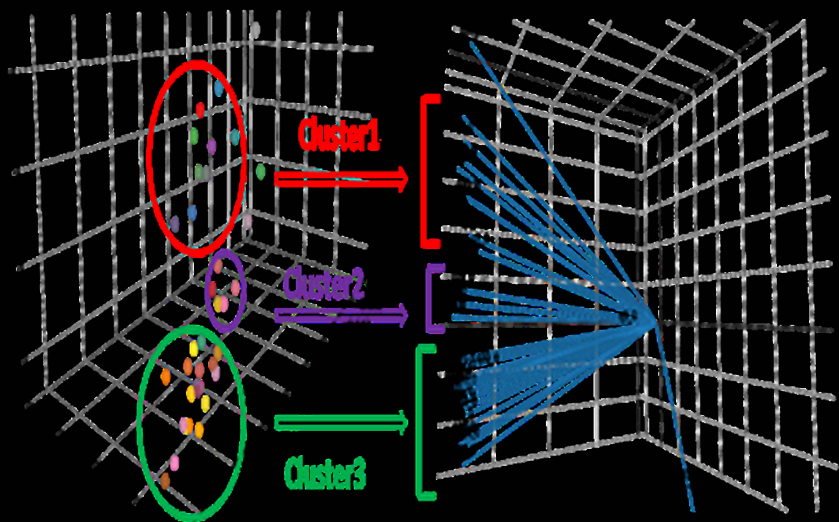
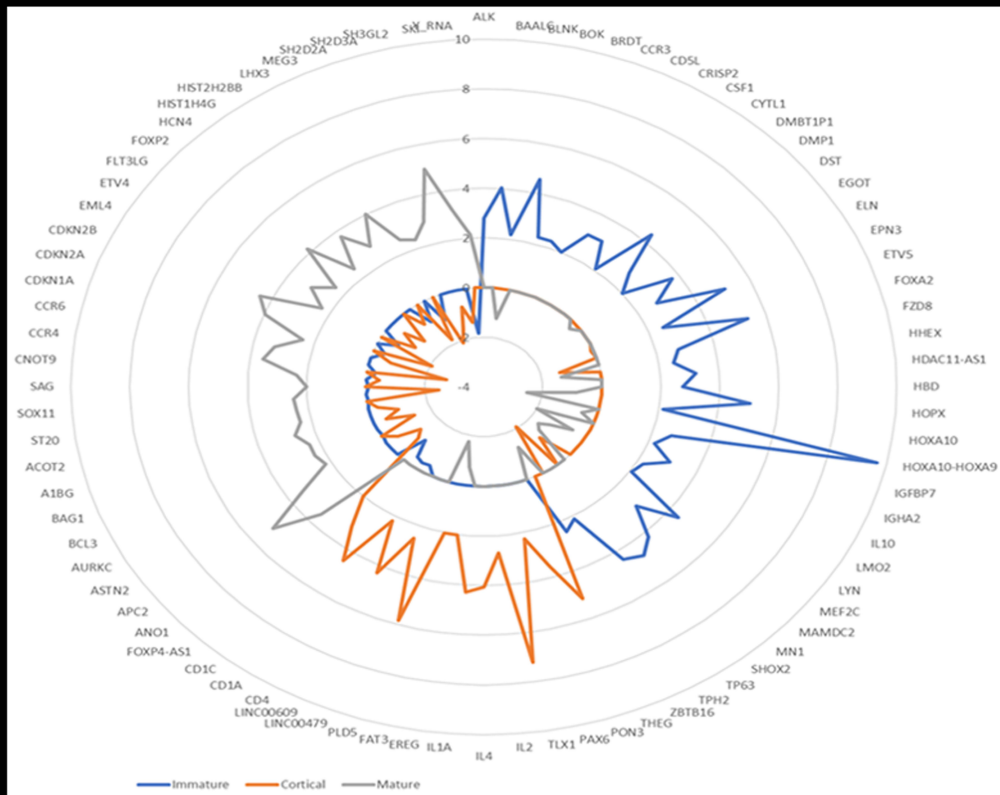
18 **Figure 4:** Enriched gene ontology functions of differentially expressed genes, DAVID results  
19 the number of involved genes from differentially expressed genes in T-ALL for biological  
20 process **(a)**, cellular components **(b)** and KEGG pathways**(c)**. ( $p < 0.005$ ). Complete list of genes  
21 involved in the biological process and cellular component of are shown in Supplementary

22 **Figure 5.** Kaplan Meier analysis for event free survival for expression of **a.** *MEF2C*, **b.** *LYL1*, **c.**  
23 *ST20*, **d.** *RAG1*, **e.** *EML4* and **f.** *KDM6A* genes.

- 1 **Figure 6.** Kaplan Meier analysis for overall survival for expression of **a. *MEF2C*, b. *DOT1L*, c.**
- 2 ***RAG1*, d. *MALATI*.**

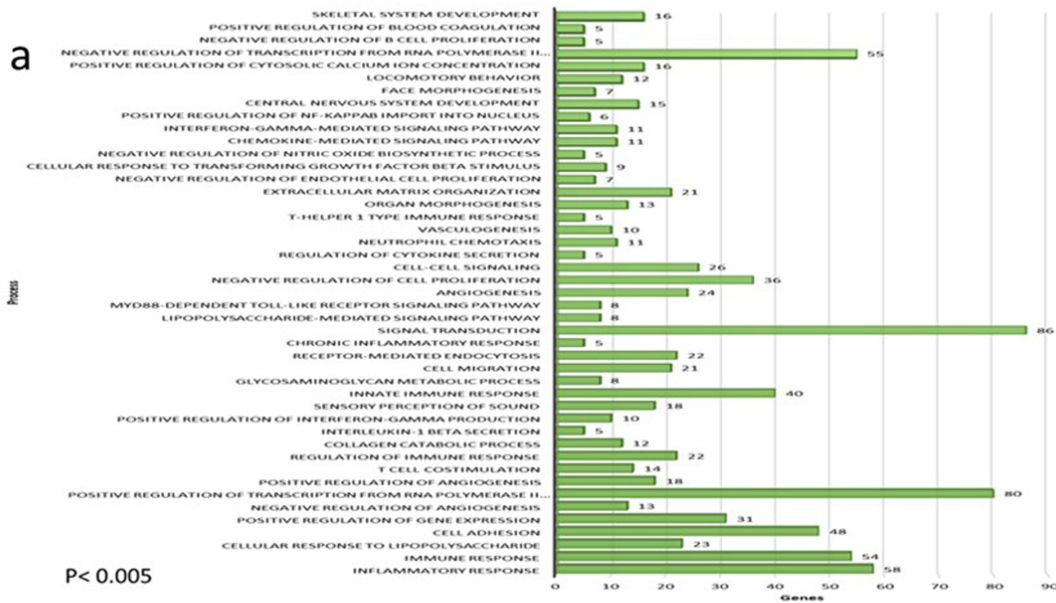




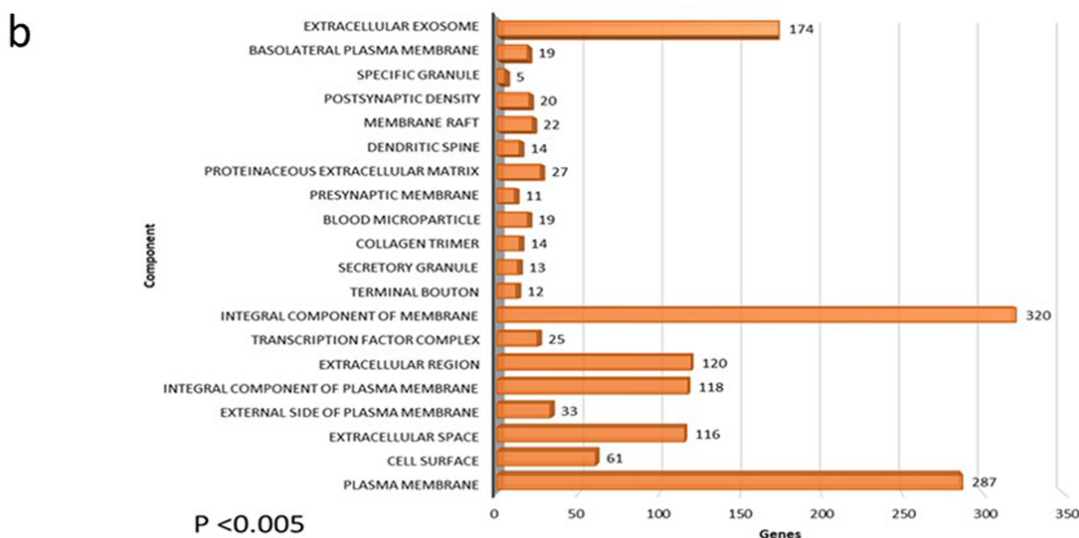




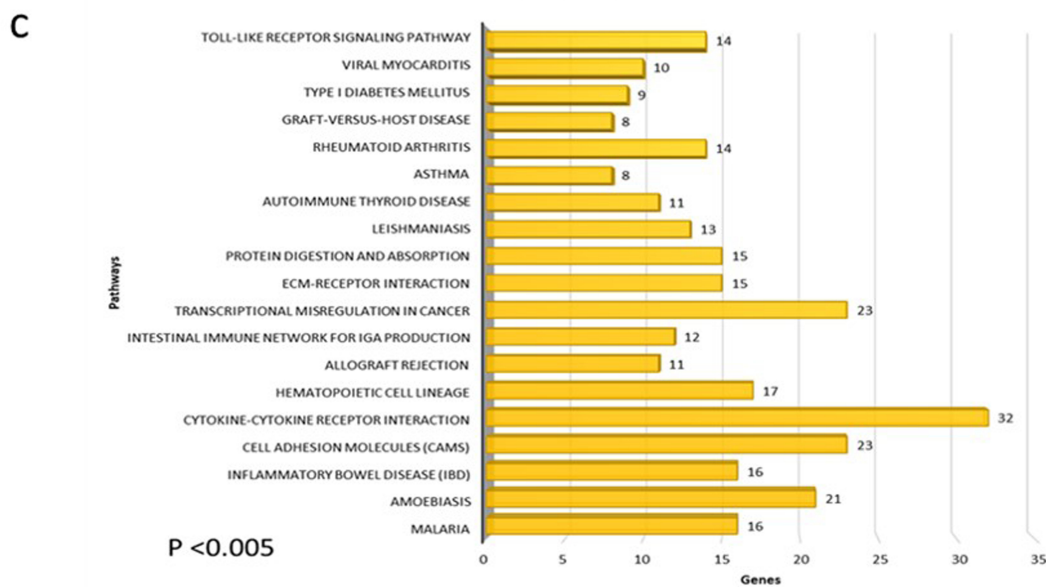
## Biological Process: T-ALL

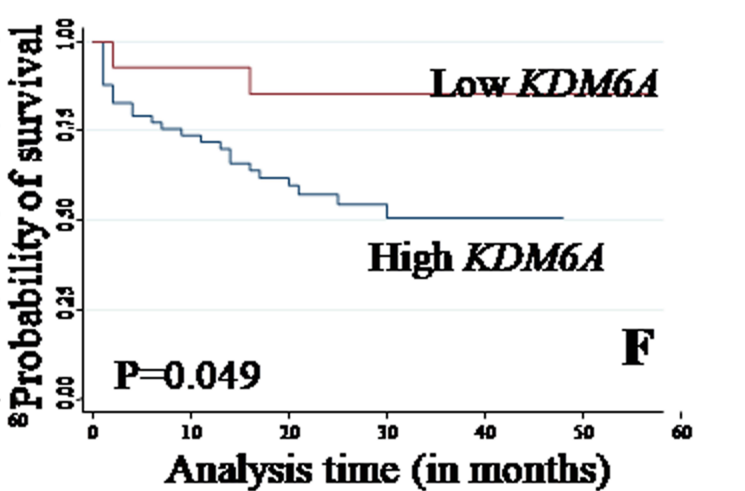
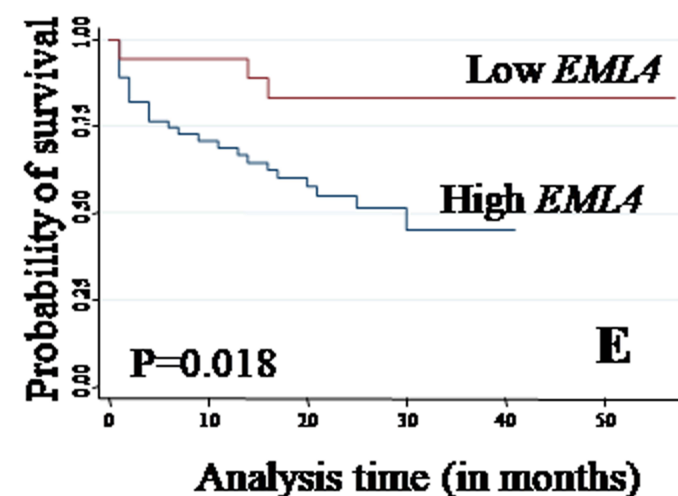
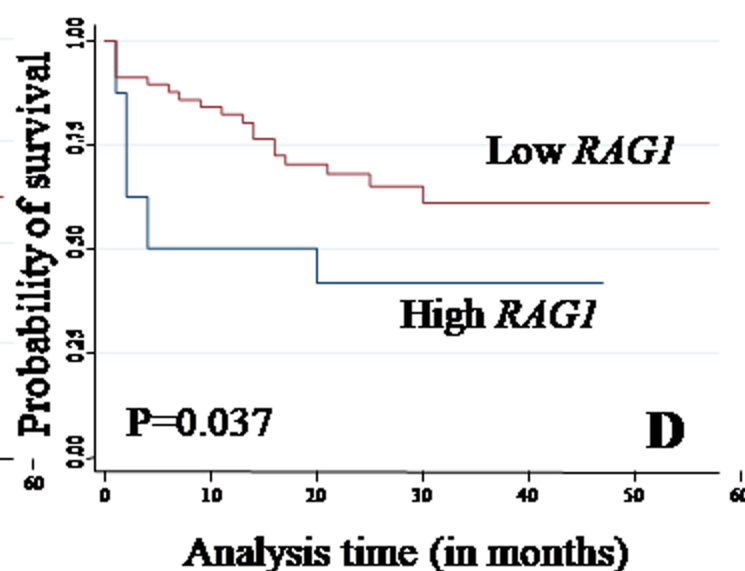
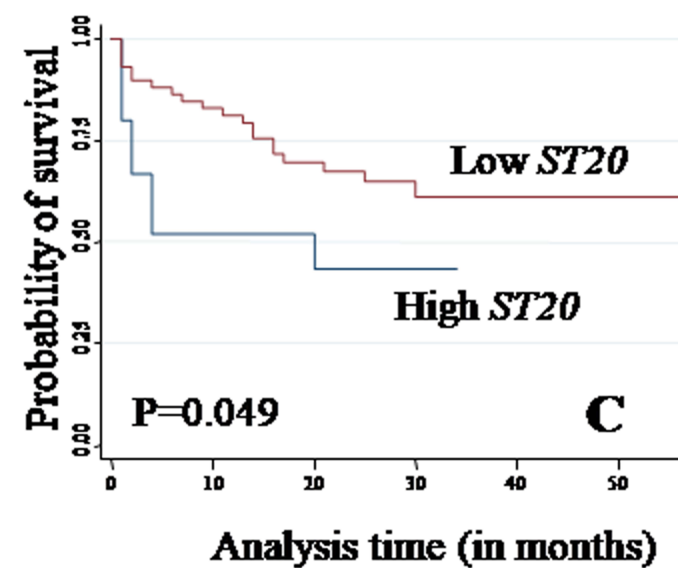
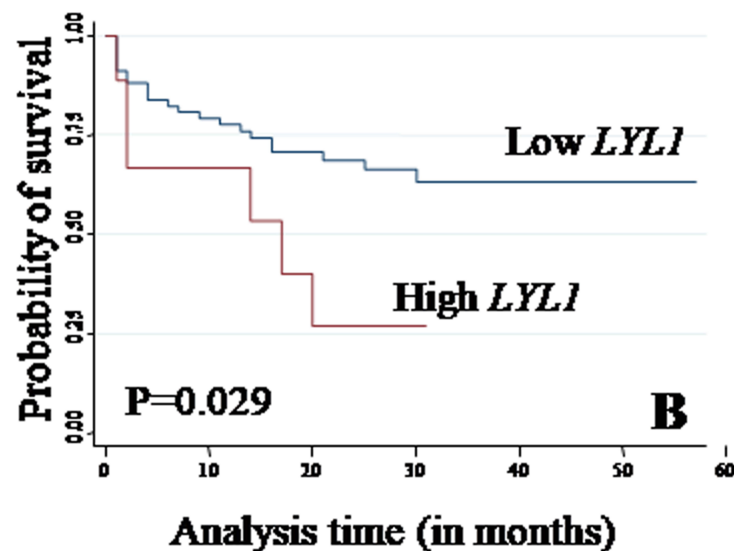
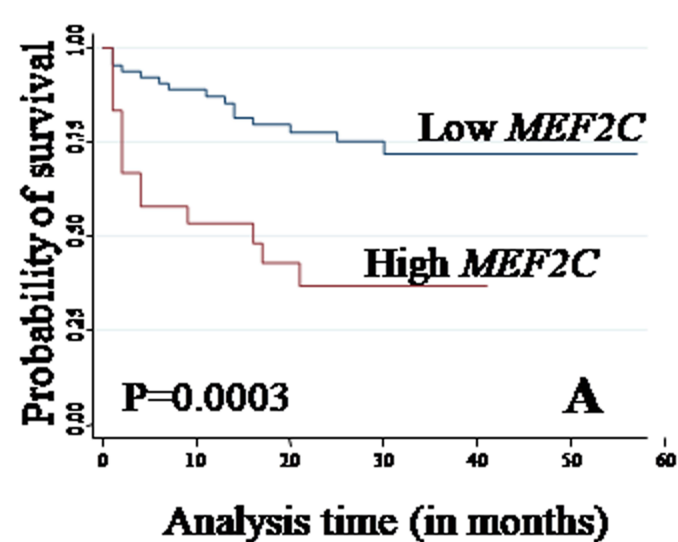


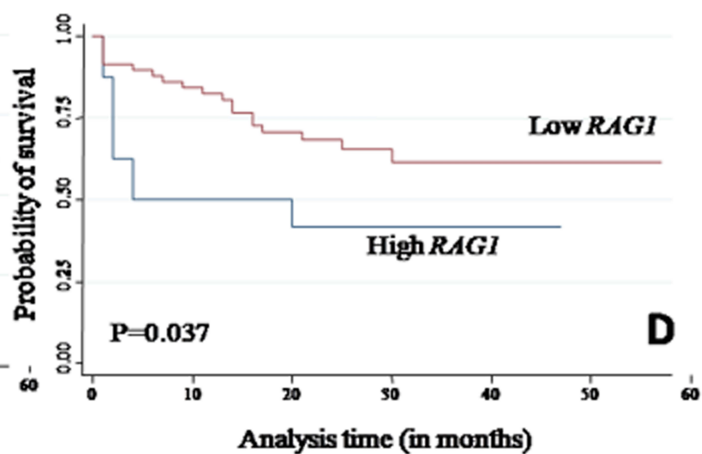
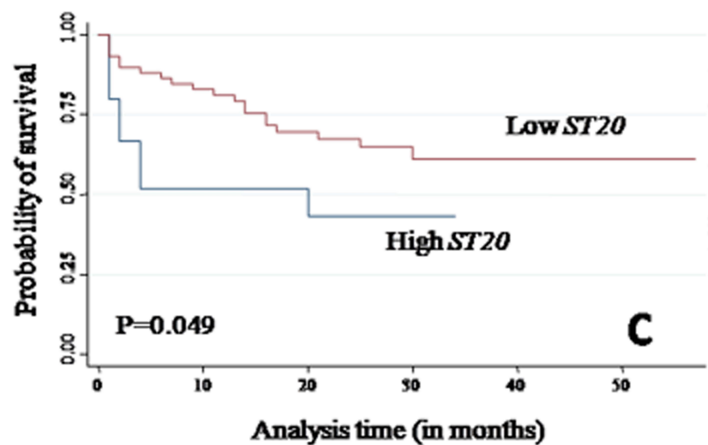
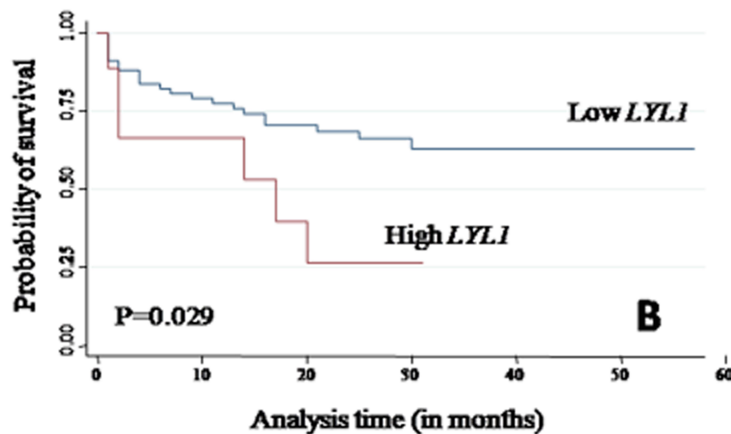
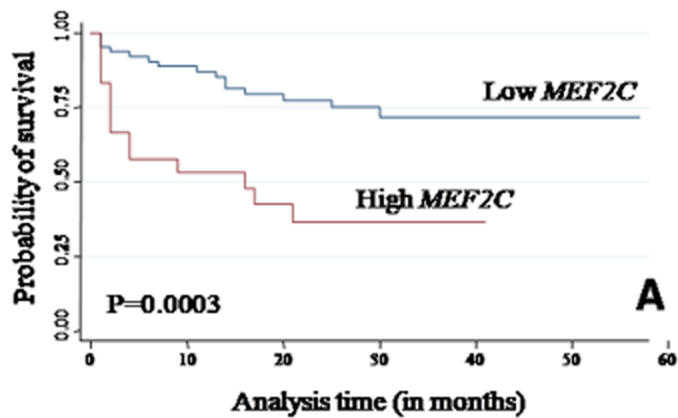
## Cellular component: T-ALL



## KEGG Pathways: T-ALL







**Table 1: Co-expression network analysis of few selected protein-coding and non-coding genes**

<b>Co-expression network</b>	<b>Protein-coding genes</b>	<b>Non-coding genes</b>
<b>BAALC</b>	<i>MEF2C, HOPX, IGFBP7, LYL1, HDAC9, LMO2, BCL2, FLT3, LYN, KIT</i>	<i>HOTTIP, HOTAIR, SNORD101, MIR6804, MIR939, XIST</i>
<b>HHEX</b>	<i>MEF2C, FLT3, KIT, LYL1, LMO2, TRGC1, HOPX</i>	<i>HOTTIP, HOTAIR, LINC01021, XIST, SNORD86C, SNORD88A, SNORD104, MIR629, MIR3132, MIR3692, MIR600</i>
<b>MEF2C</b>	<i>HOPX, KIT, BAALC, HHEX, EMP1, LYL1, BCL, SMYD3, HDAC9</i>	<i>HOTTIP, HOTAIR, LINC01021, XIST, MIR3142HG, MIR3132, MIR4741, SNORD100, SNORD101</i>
<b>FGR</b>	<i>LEF1, RAG1, CD4, CD8, TRBV2,</i>	<i>LINC01221, LINC01225, LINC01226, SNORA50A, SNORA21, SNORA31</i>
<b>FAT1</b>	<i>PRC1, PLK44, RAD51, LMO1, HIST2H2BF, HIST1H2BB, HIST1H4L</i>	<i>LINC00977, SNOR79, LINC01572, LNIC1934</i>
<b>EML4</b>	<i>CDKN2A, NCOR2, TRAV30, CRIP, TBP, BACH2RNF-4, TGFB1</i>	<i>LINC00672, LINC00664, SNORA78, SNORD30, MIR3960, MIR562, MIR573</i>
<b>HOTTIP</b>	<i>LYL1, BAALC, TRGC1, LMO2, HHEX, MEF2C, FLT3, EMP1, IGFBP7, KIT, GBP1, TLR1, HDAC9, SMYD3</i>	<i>LINC01021, XIST, HOTAIR), miRNA (MIR5692C2, MIR3917, MIR3692, MIR6658</i>

**Table 2: Association of expression of various protein and non-protein coding RNAs with patients' characteristics**

Variables	BAALC			HHEX			MEF2C			LYLI			TALI			FATI			LMO2			DOTIL		
	Low (n=43) (%)	High (n=44) (%)	P	Low (n=43) (%)	High (n=44) (%)	P	Low (n=73) (%)	High (n=26) (%)	P	Low (n=77) (%)	High (n=10) (%)	P	Low (n=45) (%)	High (n=42) (%)	P	Low (n=43) (%)	High (n=44) (%)	P	Low (n=67) (%)	High (n=11) (%)	P	Low (n=17) (%)	High (n=61) (%)	P
<b>Age (in years)</b>																								
<12	26 (60.5)	18 (40.9)	0.087	25 (58.1)	19 (43.2)	0.2	40 (54.8)	10 (38.5)	0.18	41 (53.2)	3 (30)	0.15	24 (53.3)	20 (47.6)	0.67	19 (44.2)	25 (56.8)	0.29	34 (50.7)	4 (36.4)	0.52	7 (41.2)	31 (50.8)	0.59
≥12	17 (39.5)	26 (59.1)		18 (41.9)	25 (56.8)		33 (45.2)	16 (61.5)		36 (46.8)	7 (70)		21 (46.7)	22 (52.4)		24 (55.8)	19 (43.2)		33 (49.3)	7 (63.6)		10 (58.8)	30 (49.2)	
<b>Sex</b>																								
Male	37 (86)	38 (86.4)	1	37 (86)	38 (86.4)	1	63 (86.3)	23 (88.5)	1	66 (85.7)	9 (90)	1	37 (82.2)	38 (90.5)	0.36	37 (86)	38 (86.4)	1	57 (85.1)	10 (90.9)	1	14 (82.4)	53 (86.9)	0.69
Female	6 (14)	6 (13.6)		6 (14)	6 (13.6)		10 (13.7)	3 (11.5)		11 (14.3)	1 (10)		8 (17.8)	4 (9.5)		6 (14)	6 (13.6)		10 (14.9)	1 (9.1)		3 (17.6)	8 (13.1)	
<b>TLC (X10<sup>9</sup>/L)</b>																								
<50	22 (51.2)	27 (61.4)	0.39	23 (53.5)	26 (59.1)	0.67	40 (54.8)	19 (73.1)	0.16	43 (55.8)	6 (60)	0.19	28 (62.2)	21 (50)	0.29	25 (58)	24 (54.5)	0.83	39 (58.2)	5 (45.5)	0.52	10 (58.8)	34 (55.7)	1
≥50	21 (48.8)	17 (38.6)		20 (46.5)	18 (40.9)		33 (45.2)	7 (26.9)		34 (44.2)	4 (40)		17 (37.8)	21 (50)		18 (41.9)	20 (45.5)		28 (41.8)	6 (45.5)		7 (41.2)	27 (44.3)	
<b>NCI risk</b>																								
Standard	8 (18.6)	6 (13.6)	0.572	8 (18.6)	6 (13.6)	0.57	14 (19.2)	5 (19.2)	1	11 (14.3)	3 (30)	1	11 (24.4)	3 (7.1)	0.04	4 (9.3)	10 (22.7)	0.14	11 (16.4)	1 (9.1)	1	1 (5.9)	11 (18)	0.45
High	35 (81.4)	38 (86.4)		35 (81.4)	38 (86.4)		59 (80.8)	21 (80.8)		66 (85.7)	7 (70)		34 (75.6)	39 (92.9)		39 (90.7)	34 (77.3)		56 (83.6)	10 (90.9)		16 (94.1)	50 (82)	
<b>ETP-ALL IPT</b>																								
No	41 (95.3)	31 (70.5)	0.003	39 (90.7)	33 (75)	0.08	67 (91.8)	17 (65.4)	0.003	67 (87)	5 (50)	0.012	34 (75.6)	38 (90.5)	0.09	34 (79.1)	38 (86.4)	0.41	54 (80.6)	10 (90.9)	0.68	12 (70.6)	52 (85.2)	0.17
Yes	2 (4.7)	13 (29.5)		4 (9.3)	11 (25)		6 (8.2)	9 (34.6)		10 (13)	5 (50)		11 (24.4)	4 (9.5)		9 (20.9)	6 (13.6)		13 (19.4)	1 (9.1)		5 (29.4)	9 (14.2)	
<b>CD34</b>																								
Negative	28 (65.1)	18 (40.9)	0.032	27 (88.4)	19 (43.2)	0.08	44 (60.3)	8 (30.8)	0.012	41 (53.3)	5 (50)	1	21 (46.7)	25 (59.5)	0.28	23 (53.5)	23 (52.3)	1	32 (47.8)	8 (72.7)	0.194	10 (58.8)	30 (49.2)	0.59
Positive	15 (34.9)	26 (59.1)		16 (37.2)	25 (56.8)		29 (39.7)	18 (69.2)		36 (46.8)	5 (50)		24 (53.3)	17 (40.5)		20 (46.5)	21 (47.7)		35 (52.2)	3 (27.3)		7 (41.2)	31 (50.8)	
<b>Myeloid markers</b>																								
Negative	40 (93)	32 (72.7)	0.021	38 (88.4)	34 (77.3)	0.26	63 (86.3)	19 (73.1)	0.14	66 (85.7)	6 (60)	0.065	38 (84.4)	34 (81)	0.78	33 (76.7)	39 (88.6)	0.17	54 (80.6)	10 (90.9)	0.68	16 (94.1)	48 (78.7)	0.28
Positive	3 (7)	12 (27.3)		5 (11.6)	10 (22.7)		10 (13.7)	7 (26.9)		11 (14.3)	4 (40)		7 (15.6)	8 (19)		10 (23.3)	5 (11.4)		13 (19.4)	1 (9.1)		1 (5.9)	13 (21.3)	
<b>Prednisolone sensitivity</b>																								
Sensitive	23 (71.9)	21 (61.8)	0.441	24 (82.8)	20 (54)	0.018	41 (77.4)	12 (54.5)	0.048	39 (67.2)	5 (62.5)	1	22 (62.9)	22 (71)	0.60	21 (63.6)	23 (69.7)	0.79	34 (68)	7 (70)	1	7 (53.8)	34 (63.3)	0.31
Resistant	9 (28.1)	13 (38.2)		5 (17.2)	17 (45.9)		12 (22.6)	10 (45.5)		19 (32.8)	3 (37.5)		13 (37.1)	9 (29)		12 (36.4)	10 (30.3)		16 (32)	3 (30)		6 (46.2)	13 (21.3)	
<b>Post-induction MRD</b>																								
Negative	21 (75)	16 (15.31)	0.107	21 (80.8)	16 (50)	0.027	38 (76)	6 (37.5)	0.007	32 (62.7)	5 (71.4)	1	17 (58.6)	20 (69)	0.59	16 (53.3)	21 (75)	0.11	29 (61.7)	6 (66.7)	1	6 (60)	29 (63)	1.00
Positive	7 (25)	14 (46.7)		5 (17.2)	16 (50)		12 (24)	10 (62.5)		19 (37.3)	2 (28.6)		12 (41.4)	9 (31)		14 (46.7)	7 (25)		18 (38.3)	3 (33.3)		4 (40)	17 (37)	



Variables	LYN			LINC00648			PCAT18			LINC00461			LINC00202			MEF2C-ASI			XIST			PCAT14		
	Low (n=39) (%)	High (n=39) (%)	P	Low (n=55) (%)	High (n=17) (%)	P	Low (n=20) (%)	High (n=52) (%)	P	Low (n=36) (%)	High (n=36) (%)	P	Low (n=38) (%)	High (n=38) (%)	P	Low (n=64) (%)	High (n=12) (%)	P	Low (n=70) (%)	High (n=17) (%)	P	Low (n=38) (%)	High (n=38) (%)	P
<b>Age (in years)</b>																								
<12	15 (38.5)	23 (59)	0.11	24 (43.6)	14 (82.4)	1	11 (55)	24 (46.2)	0.6	18 (50)	17 (47.2)	1	19 (50)	17 (44.7)	0.82	27 (42.2)	9 (75)	0.06	34 (48.6)	10 (58.8)	0.75	17 (44.7)	19 (50)	0.82
≥12	24 (61.5)	16 (41)		31 (56.4)	3 (17.6)		9 (45)	28 (53.8)		18 (50)	19 (52.8)		19 (50)	21 (55.3)		37 (57.8)	3 (25)		36 (51.4)	7 (41.2)		21 (55.3)	19 (50)	
<b>Sex</b>																								
Male	34 (87.2)	33 (84.6)	1	47 (85.5)	15 (88.2)	1	16 (80)	46 (88.5)	0.45	31 (86.1)	31 (86.1)	1	32 (84.2)	34 (89.5)	0.74	56 (87.5)	10 (83.3)	0.65	66 (94.3)	9 (52.9)	<0.01	32 (84.2)	34 (89.5)	0.74
Female	5 (12.8)	6 (15.4)		8 (14.5)	2 (11.8)		4 (20)	6 (11.5)		5 (13.9)	5 (13.9)		6 (15.8)	4 (10.5)		8 (12.5)	2 (16.7)		4 (5.7)	8 (47.1)		6 (15.8)	4 (10.5)	
<b>TLC (X10<sup>9</sup>/L)</b>																								
<50	22 (56.4)	22 (56.4)	1	29 (52.7)	11 (64.7)	0.42	8 (40)	32 (61.5)	0.11	21 (58.3)	19 (52.8)	0.81	22 (57.9)	21 (55.3)	1	36 (56.2)	7 (58.3)	1	40 (57.1)	9 (52.9)	0.19	21 (55.3)	22 (57.9)	1
≥50	17 (43.6)	17 (43.6)		26 (47.3)	6 (35.3)		12 (60)	20 (38.5)		15 (41.7)	17 (47.2)		16 (42.1)	17 (44.7)		28 (43.8)	5 (41.7)		30 (42.9)	8 (47.1)		17 (44.7)	16 (42.1)	
<b>NCI risk</b>																								
Standard	4 (10.3)	8 (20.5)	0.35	9 (16.4)	1 (5.9)	0.43	3 (15)	7 (13.5)	1	6 (16.7)	4 (11.1)	0.74	6 (15.8)	5 (13.2)	1	10 (15.6)	1 (8.3)	1	8 (11.4)	6 (35.3)	0.01	5 (13.2)	6 (15.8)	1
High	35 (89.7)	31 (79.5)		46 (83.6)	16 (94.1)		17 (85)	45 (86.5)		30 (83.3)	32 (88.9)		32 (84.2)	33 (86.8)		54 (84.4)	11 (91.7)		62 (88.6)	11 (64.7)		33 (86.8)	32 (84.2)	
<b>ETP-ALL</b>																								
No	37 (94.9)	27 (69.2)	0.01	45 (81.8)	14 (82.4)	1	16 (80)	43 (82.7)	0.75	31 (86.1)	28 (77.8)	0.54	32 (84.2)	31 (81.6)	1	52 (81.2)	11 (91.7)	0.68	61 (87.1)	11 (64.7)	0.02	31 (81.6)	32 (84.2)	1
Yes	2 (5.1)	12 (30.8)		10 (18.2)	3 (17.6)		4 (20)	9 (17.3)		5 (13.9)	8 (22.2)		6 (15.8)	7 (18.4)		12 (18.8)	1 (8.3)		9 (12.9)	6 (35.3)		7 (18.4)	6 (15.8)	
<b>CD34</b>																								
Negative	21 (53.8)	19 (48.7)	0.82	30 (54.5)	7 (41.2)	0.41	11 (55)	26 (50)	0.79	23 (63.9)	14 (38.9)	0.06	23 (60.5)	16 (42.1)	0.17	34 (53.1)	5 (41.7)	0.54	38 (54.3)	8 (47.1)	0.85	21 (55.3)	18 (47.4)	0.65
Positive	18 (46.2)	20 (51.3)		25 (45.5)	10 (58.8)		9 (45)	25 (50)		13 (36.1)	22 (61.1)		15 (39.5)	22 (57.9)		30 (46.9)	7 (58.3)		32 (45.7)	9 (52.9)		17 (44.7)	20 (52.6)	
<b>Myeloid markers</b>																								
Negative	34 (87.2)	30 (76.9)	0.38	46 (83.6)	14 (82.4)	1	17 (85)	43 (82.7)	1	32 (88.9)	28 (77.8)	0.34	34 (89.5)	29 (76.3)	0.22	51 (79.7)	12 (100)	0.11	59 (84.3)	13 (76.5)	0.75	32 (84.2)	31 (81.6)	1
Positive	5 (12.8)	30 (76.9)		9 (16.4)	3 (17.6)		3 (15)	9 (17.3)		4 (11.1)	8 (22.2)		4 (10.5)	9 (23.7)		13 (20.3)	0		11 (15.7)	4 (23.5)		6 (15.8)	7 (18.4)	
<b>Prednisolone sensitivity</b>																								
Sensitive	22 (75.9)	30 (76.9)	0.38	28 (68.3)	11 (78.6)	0.73	12 (70.6)	27 (71.1)	1	22 (75.9)	17 (65.4)	0.55	21 (70)	19 (67.9)	1	31 (66)	9 (81.8)	0.47	37 (69.8)	7 (53.8)	0.06	20 (74.1)	20 (64.5)	
Resistant	5 (12.8)	9 (23.1)		13 (31.7)	3 (21.4)		5 (29.4)	11 (28.9)		7 (24.1)	9 (34.6)		9 (30)	9 (32.1)		16 (34)	2 (18.2)		16 (30.2)	6 (46.2)		7 (25.9)	11 (35.5)	0.57
<b>Post-induction MRD</b>																								
Negative	16 (59.3)	19 (65.5)	0.78	24 (64.9)	7 (50)	0.36	14 (82.4)	17 (50)	0.04	15 (65.2)	16 (57.1)	0.58	20 (74.1)	14 (51.9)	0.16	28 (62.2)	6 (66.7)	1	30 (65.2)	7 (58.3)	0.37	18 (66.7)	16 (59.3)	0.78
Positive	11 (40.7)	10 (34.5)		13 (35.1)	7 (50)		3 (17.6)	17 (50)		8 (34.8)	12 (42.9)		7 (25.9)	13 (48.1)		17 (37.8)	3 (33.3)		16 (34.8)	5 (41.7)		9 (33.3)	11 (40.3)	

Variables	ST20			RAG			EP300			EML4			EZH2			MALAT1			KDM6A		
	Low (n=16) (%)	High (n=68) (%)	P	Low (n=18) (%)	High (n=66) (%)	P	Low (n=60) (%)	High (n=24) (%)	P	Low (n=66) (%)	High (n=18) (%)	P	Low (n=63) (%)	High (n=19) (%)	P	Low (n=20) (%)	High (n=61) (%)	P	Low (n=67) (%)	High (n=15) (%)	P
<b>Age (in years)</b>																					
<12	7 (43.8)	36 (52.9)	0.59	5 (27.8)	38 (57.6)	0.034	31 (51.7)	12 (50)	1.00	32 (48.5)	11 (61.1)	0.43	30 (47.6)	12 (63.2)	0.299	8 (40)	33 (54)	0.31	31 (46.3)	11 (73.3)	0.086
≥12	9 (56.2)	32 (47.1)		13 (72.2)	28 (42.4)		29 (48.3)	12 (50)		34 (51.5)	7 (38.9)		33 (52.4)	7 (36.8)		12 (60)	28 (45.9)		36 (53.7)	4 (26.7)	
<b>Sex</b>																					
Male	12 (75)	60 (88.2)	0.23	18 (100)	54 (81.8)	0.06	51 (85)	21 (87.5)	1.00	56 (84.8)	16 (88.9)	1.00	55 (87.8)	15 (78.9)	0.46	18 (90)	51 (83.6)	0.72	59 (88.1)	11 (73.3)	0.22
Female	4 (25)	8 (11.8)		0 (18.2)	12 (18.2)		9 (15)	3 (12.5)		10 (15.2)	2 (11.1)		8 (12.7)	4 (21.1)		2 (10)	10 (16.4)		8 (11.9)	4 (26.7)	
<b>TLC (X10<sup>9</sup>/L)</b>																					
<50	12 (75)	36 (52.9)	0.16	12 (66.7)	36 (54.5)	0.43	33 (55)	15 (62.5)	0.63	36 (54.5)	12 (66.7)	0.43	34 (54)	12 (63.2)	0.60	16 (80)	30 (49.2)	0.02	37 (55.2)	9 (60)	0.78
≥50	4 (25)	32 (47)		6 (33.3)	30 (45.5)		27 (45)	9 (37.5)		30 (45.5)	6 (33.3)		29 (46)	7 (36.8)		4 (20)	31 (50.8)		30 (44.8)	6 (40)	
<b>NCI risk</b>																					
Standard	4 (25)	10 (14.7)	0.45	3 (16.7)	11 (16.7)	1.00	10 (16.7)	4 (16.7)	1.00	12 (18.2)	2 (11.1)	0.72	8 (12.7)	6 (31.6)	0.08	3 (15)	11 (18)	1.00	8 (11.9)	6 (40)	0.018
High	12 (75)	58 (85.3)		15 (83.3)	55 (83.3)		50 (83.3)	20 (83.3)		54 (81.8)	16 (88.9)		55 (87.5)	13 (68.4)		17 (85)	50 (82)		59 (88.1)	9 (60)	
<b>ETP-ALL IPT</b>																					
No	9 (56.2)	60 (88.2)	0.007	13 (72.2)	56 (84.8)	0.29	46 (76.7)	23 (95.8)	0.056	51 (77.3)	18 (100)	0.033	51 (81)	16 (84.2)	1.00	16 (80)	50 (82)	1.00	55 (82.1)	12 (80)	1.00
Yes	7 (43.8)	8 (11.8)		5 (27.8)	10 (15.2)		14 (23.3)	1 (4.2)		15 (22.7)	0		12 (9)	3 (15.8)		4 (20)	11 (18)		12 (17.9)	3 (20)	
<b>CD34</b>																					
Negative	8 (50)	36 (52.9)	1.00	9 (50)	35 (53)	1.00	32 (53.3)	12 (50)	0.81	34 (51.5)	10 (55.6)	0.79	32 (50.8)	10 (52.6)	1.00	8 (40)	33 (54)	0.31	33 (49.3)	9 (60)	0.57
Positive	8 (50)	32 (47.1)		9 (50)	31 (47)		28 (46.7)	12 (50)		32 (48.5)	8 (44.4)		31 (49.2)	9 (47.4)		12 (60)	28 (45.9)		34 (50.7)	6 (40)	
<b>Myeloid markers</b>																					
Negative	9 (56.2)	60 (88.2)	0.007	13 (72.2)	56 (84.8)	0.29	50 (83.3)	19 (79.2)	0.75	54 (81.8)	15 (83.3)	1.00	52 (82.5)	15 (78.9)	0.74	17 (85)	49 (80.3)	0.75	58 (86.6)	9 (60)	0.026
Positive	7 (43.8)	8 (11.8)		5 (27.8)	10 (15.2)		10 (16.7)	5 (20.8)		12 (18.2)	3 (16.7)		11 (17.5)	4 (21.1)		3 (15)	12 (19.7)		9 (13.4)	6 (40)	
<b>Prednisolone sensitivity</b>																					
Sensitive	7 (53.8)	35 (68.6)	0.343	8 (50)	34 (70.8)	0.14	27 (61.4)	15 (75)		29 (61.7)	13 (76.5)	0.38	32 (68.1)	8 (53.3)	0.36	8 (53.3)	31 (67.4)	0.37	31 (62)	9 (75)	
Resistant	6 (46.2)	16 (31.4)		8 (50)	14 (29.2)		17 (38.6)	4 (21.1)	0.39	18 (38.3)	4 (23.5)		15 (31.9)	7 (46.7)		7 (46.7)	15 (32.6)		19 (38)	3 (25)	0.51
<b>Post-induction MRD</b>																					
Negative	4 (40)	32 (66.7)	0.16	5 (38.5)	31 (68.9)	0.059	21 (53.8)	15 (78.9)	0.09	23 (54.8)	13 (81.2)	0.077	27 (64.3)	9 (60)	0.77	7 (58.3)	29 (64.4)	0.75	27 (61.4)	9 (69.2)	0.75
Positive	6 (60)	16 (33.3)		8 (61.5)	14 (31.1)		18 (46.2)	4 (21.1)		19 (45.2)	3 (18.8)		15 (35.7)	6 (40)		5 (41.7)	16 (35.6)		17 (38.6)	4 (30.8)	

**Table 3: Association of patients' characteristics with EFS and OS**

Variable	Event free survival				Overall survival			
	3yr-EFS (%)±SE	HR	95% CI	P value	3yr-OS (%)±SE	HR	95% CI	P value
Age (in years)								
<12 (n=47)	76.49±6.59	1	1.16-5.45	0.0132	90.81±4.4	1	0.28-4.46	0.88
≥12 (n=42)	42.19±9.36	2.52			90.09±4.71	1.11		
TLC (X 10 <sup>9</sup> /L)								
<50 (n=54)	63.36±7.27	1	0.51-2.24	0.86	91.52±4.07	1	0.39-6.28	0.52
≥50 (n=35)	60.28±9.75	1.068			88.38±5.47	1.57		
NCI risk								
Standard (n=18)	79.9±10.59	1	0.74-8.08	0.12	100	1	0	0.15
High (n=71)	57.98±6.71	2.44			88.01±4.0	7.43e-14		
BAALC								
Low (n=35)	62.45±8.68	1	0.56-2.51	0.65	90.39±5.31	1	0.33-5.75	0.66
High (n=42)	56.17±8.87	1.88			88.03±5.02	1.37		
HHEX2								
Low (n=36)	50.73±10.53	1	0.43-1.88	0.76	89.99±5.52	1	0.35-6.16	0.59
High (n=41)	63.83±7.87	0.89			87.80±5.11	1.47		
MEF2C								
Low (n=65)	71.78±6.52	1	1.67-7.3	0.0003	94.77±2.96	1	1.16-20.40	0.016
High (n=24)	36.57±10.74	3.5			78.75±8.45	4.88		
LYL1								
Low (n=68)	63.14±6.65	1	1.08-6.69	0.024	89.14±3.89	1	0.14-9.24	0.91
High (n=9)	26.67±15.9	2.69			88.89±10.48	1.14		
TAL1								
Low (n=39)	52.25±9.77	1	0.39-1.74	0.61	85.78±5.95	1	0.14-2.51	0.48
High (n=38)	64.35±8.01	0.613			92.11±4.37	0.59		
FAT1								
Low (n=39)	54.78±8.62	1	0.36-1.59	0.45	86.74±5.53	1	0.14-2.54	0.49
High (n=38)	62.23±9.61	0.75			91.58±4.	0.61		

					68			
LYN								
Low (n=33)	58.86±9.39	1	0.52-2.41	0.78	86.86±6.19	1	0.22-3.47	0.84
High (n=36)	58.98±8.54	1.12			88.71±5.32	0.87		
LMO2								
Low (n=28)	64.46±6.65	1	0.94-5.33	0.058	89.11±4.2	1	0.38-9.4	0.42
High (n=41)	29.99±15.73	2.24			81.82±1.63	1.89		
DOT1L								
Low (n=13)	44.87±14.14	1	0.25-1.44	0.25	68.38±1.315	1	0.06-0.88	0.019
High (n=56)	62.52±6.94	0.61			92.64±3.55	0.22		
LNC648								
Low (n=31)	54.81±6.65	1	0.18-1.55	0.23	87.3±4.85	1	0.19-4.81	0.97
High (n=33)	70±15.73	0.53			85.94±9.31	0.97		
PCAT18								
Low (n=19)	59.71±8.67	1	0.45-2.14	0.95	89.47±7.04	1	0.26-6.43	0.75
High (n=45)	56.40±9.67	0.98			85.89±5.38	1.29		
LNC461								
Low (n=30)	57.96±9.33	1	0.45-2.16	0.97	86.02±6.50	1	0.23-3.63	0.89
High (n=34)	58.34±9.67	0.99			88.03±5.62	0.91		
LNC202								
Low (n=34)	55.29±8.62	1	0.35-1.67	0.49	82.04±6.65	1	0.07-1.74	0.18
High (n=33)	59.80±9.86	0.77			93.94±4.15	0.35		
MEF2C-AS1								
Low (n=58)	53.92±6.94	1	0.028-1.52	0.082	85.72±4.69	1	0	0.25
High (n=9)	85.71±13.23	0.21			100±0	1.82e-16		
PCAT14								
Low (n=34)	54.27±8.83	1	0.37-1.74	0.56	82.04±6.65	1	0.07-1.74	0.18
High (n=33)	62.14±9.32	0.79			93.94±4.15	0.35		
XIST								
Low (n=62)	58.36±7.17	1	0.51-3.11	0.61	91.21±3.78	1	0.59-10.30	0.21

High (n=16)	61.88±12.30	1.26			80.77±10	2.45		
ST20								
Low (n=15)	43.21±13.56	1	0.19-1.03	0.049	86.15±9.11	1	0.15-3.59	0.69
High (n=59)	61.02±7.26	0.45			89.32±4.14	0.72		
RAG								
Low (n=16)	41.67±12.9	1	0.20-0.99	0.037	75±10.83	1	0.06-0.98	0.03
High (n=58)	61.24±7.49	0.45			92.54±3.6	0.25		
EP300								
Low (n=51)	46.71±9.3	1	0.19-1.15	0.085	87.75±4.7	1	0.14-3.42	0.65
High (n=23)	73.66±9.25	0.46			91.1±6.01	0.69		
EML4								
Low (n=56)	45.29±8.92	1	0.078-0.88	0.018	88.89±4.29	1	0.19-4.86	0.98
High (n=18)	83.33±8.78	0.26			88.89±7.41	0.98		
EZH2								
Low (n=55)	52.02±7.66	1	0.18-1.51	0.22	90.57±4.02	1	0.24-6.44	0.79
High (n=17)	76.02±10.48	0.52			87.84±8.07	1.25		
MALAT1								
Low (n=17)	39.93±13.66	1	0.23-1.12	0.081	76.02±10.48	1	0.048-0.97	0.027
High (n=54)	61.73±7.62	0.502			94.11±3.31	0.22		
KDM6A								
Low (n=58)	50.48±7.58	1	0.062-1.12	0.049	89.35±4.12	1	0.075-5.2	0.66
High (n=14)	85.12±9.73	0.27			92.31±7.39	0.63		
ETP-ALL immunophenotype								
No (n=76)	65.80±6.25	1	0.99-5.46	0.043	90.20±3.54	1	0.09-6.24	0.80
Yes (n=13)	43.96±14.26	2.32			92.31±7.39	0.76		
MRD								
Negative (n=44)	73.78±7.28	1	0.72-4.66	0.19	95.01±3.45	1	0.29-14.89	0.45
Positive (n=22)	60.78±11.15	1.83			90.48±6.41	2.09		

**Table 4a: Multivariate Cox regression analysis for EFS**

<b>Variable</b>	<b>Hazard ratio</b>	<b>95% CI</b>	<b>P value</b>
Age (<12 years vs $\geq$ 12 years)	1.21	0.46-3.14	0.703
ETP-ALL immunophenotype (yes vs no)	1.23	0.39-3.85	0.72
MEF2C expression(low versus high)	3.25	1.24-8.56	0.017
LYL1 expression(low versus high)	1.06	0.33-3.42	0.92
LMO2 expression(low versus high)	1.81	0.57-5.71	0.31
MEF2C- AS1 expression(low versus high)	0.23	0.027-1.93	0.18
ST20 (low versus high)	1.16	0.31-4.35	0.82
RAG (low versus high)	0.53	0.18-1.51	0.23
EP300 (low versus high)	0.53	0.18-1.5	0.23
MALAT1 (low versus high)	0.53	0.19-1.45	0.21
KDM6A (low versus high)	0.48	0.10-2.24	0.35
<b>Table 4b: Multivariate Cox regression analysis of OS</b>			
MEF2C expression(low versus high)	6.73	1.09-41.52	0.04
DOT1L expression(low versus high)	0.16	0.022-1.06	0.058
LNC202 expression(low versus high)	0.19	0.02-1.75	0.14
PCAT14 expression(low versus high)	1.12	0.15-8.56	0.91
MALAT1 expression(low versus high)	0.16	0.031-0.85	0.031
RAG expression (low versus high)	1.36	0.21-8.83	0.75