# Long non-coding RNAs defining major subtypes of B cell precursor acute lymphoblastic leukemia

Alva Rani James[1,2,3], Michael P Schroeder[1], Martin Neumann[1,2,3], Lorenz Bastian[1,2,3], Cornelia Eckert[2,3,4], Nicola Gökbuget[2,3,5], Jutta Ortiz Tanchez[1], Cornelia Schlee[1], Konstandina Isaakidis[1], Stefan Schwartz[1], Thomas Burmeister[6], Arend von Stackelberg[2,3,4], Michael A Rieger[2,3,5], Stefanie Göllner[7], Martin Horstman[8], Martin Schrappe[9], Renate Kirschner-Schwabe[2,3,4], Monika Brüggemann[10], Carsten Müller-Tidow[7], Hubert Serve[2,3,5], Altuna Akalin[11], Claudia D Baldus[1,2,3*]

1. Charité, University Hospital Berlin, Department of Hematology and Oncology, Campus Benjamin Franklin, Berlin, 12203, Germany
2. German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany
3. German Cancer Consortium (DKTK), Heidelberg, 69120, Germany
4. Charité, University Hospital Berlin, Department of Pediatric Hematology/Oncology, Campus Rudolf Virchow, 13353, Berlin, Germany
5. Goethe University Hospital, Department of Medicine II, Department of Hematology/Oncology, Frankfurt/M, 60590, Germany
6. Charité Universitätsmedizin Berlin, Campus Virchow-Klinikum, Department. of Hematology, Oncology and Tumor Immunology, Berlin, 13353, Germany
7. University Clinic Heidelberg, Department of Hematology, Oncology & Rheumatology, Heidelberg, 69120, Germany
8. Research Institute Children's Cancer Center, Department of Pediatric Hematology and Oncology, University Medical Center Hamburg, Hamburg, 20251,Germany
9. University Hospital Schleswig-Holstein, Campus Kiel, Department of Pediatrics, Kiel, 24105, Germany.
10. University Hospital Schleswig-Holstein, Campus Kiel, Department of Hematology and Oncology, Kiel, 24105, Germany
11. Bioinformatics platform, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center (MDC), Berlin, 13125, Germany

*Corresponding author
Email: claudia.baldus@charite.de (CDB)

## Abstract

34 Recent studies implicated that long non-coding RNAs (lncRNAs) may play a role in the

35 progression and development of acute lymphoblastic leukemia, however, this role is not yet

36 clear. In order to unravel the role of lncRNAs associated with B-cell precursor Acute

37 Lymphoblastic Leukemia (BCP-ALL) subtypes, we performed transcriptome sequencing and

38 DNA methylation array across 82 BCP-ALL samples from three molecular subtypes (DUX4,

39 Ph-like, and Near Haploid or High Hyperdiploidy). Unsupervised clustering of BCP-ALL

40 samples on the basis of their lncRNAs on transcriptome and DNA methylation profiles

41 revealed robust clusters separating three molecular subtypes. Using extensive

42 computational analysis, we developed a comprehensive catalog of 1235 aberrantly

43 dysregulated BCP-ALL subtype-specific lncRNAs with altered expression and methylation

44 patterns from three subtypes of BCP-ALL. By analyzing the co-expression of subtype-specific

45 lncRNAs and protein-coding genes, we inferred key molecular processes in BCP-ALL

46 subtypes. A strong correlation was identified between the DUX4 specific lncRNAs and

47 activation of TGF-β and Hippo signaling pathways. Similarly, Ph-like specific lncRNAs were

48 correlated with genes involved in activation of PI3K-AKT, mTOR, and JAK-STAT signaling

49 pathways. Interestingly, the relapse-specific differentially expressed lncRNAs correlated

50 with the activation of metabolic and signaling pathways. Finally, we showed a set of

51 epigenetically altered lncRNAs facilitating the expression of tumor genes located at their *cis*

52 location. Overall, our study provides a comprehensive set of novel subtype and relapse-

53 specific lncRNAs in BCP-ALL. Our findings suggest a wide range of molecular pathways are

54 associated with lncRNAs in BCP-ALL subtypes and provide a foundation for functional

55 investigations that could lead to new therapeutic approaches.

2

## 56  **Author Summary**

57  Acute lymphoblastic leukemia is a heterogeneous blood cancer, with multiple molecular

58  subtypes, and with high relapse rate. We are far from the complete understanding of the

59  rationale behind these subtypes and high relapse rate. Long non-coding (lncRNAs) has

60  emerged as a novel class of RNA due to its diverse mechanism in cancer development and

61  progression. LncRNAs does not code for proteins and represent around 70% of human

62  transcripts. Recently, there are a number of studies used lncRNAs expression profile in the

63  classification of various cancers subtypes and displayed their correlation with genomic,

64  epigenetic, pathological and clinical features in diverse cancers. Therefore, lncRNAs can

65  account for heterogeneity and has independent prognostic value in various cancer subtypes.

66  However, lncRNAs defining the molecular subtypes of BCP-ALL are not portrayed yet. Here,

67  we describe a set of relapse and subtype-specific lncRNAs from three major BCP-ALL

68  subtypes and define their potential functions and epigenetic regulation. Our data uncover

69  the diverse mechanism of action of lncRNAs in BCP-ALL subtypes defining how lncRNAs are

70  involved in the pathogenesis of disease and the relevance in the stratification of BCP-ALL

71  subtypes.

## INTRODUCTION

B-cell Precursor Acute Lymphoblastic Leukemia (BCP-ALL) is the most prevalent disease in children and affects also adults. Despite improvements in treatment regimens such as chemotherapy and allogeneic hematopoietic stem cell transplantation, the prognosis remains poor for patients in high-risk groups and at relapse (1). Various risk subtypes have been established based on the cytogenetic analysis and molecular genetics studies. These subtypes are classified based on the presence of high hyperdiploidy (51-65 chromosomes) (2), hypodiploidy (less than 44 chromosomes)(3) and fusion genes (for example BCR-ABL, ETV6-RUNX, MLL, etc)(4). About 70-80% of both adults and pediatric cases of BCP-ALL constitute these subtypes, although the frequency may differ (5).

Recent efforts taking advantage of whole transcriptome sequencing (RNA-Seq) have refined this classification by identifying novel BCP-ALL subtypes. RNA-Seq analysis identified cytogenetically non-detectable recurrent rearrangements and gene fusions, which allowed characterization of additional subtypes based on distinct gene expression profiles (6). For example, the DUX4 (7) subtype is defined mainly by the IGH-DUX4 or ERG-DUX4 gene fusions; the Ph-like (8) subtype is a high-risk subtype with a gene expression profile similar to Ph-positive ALL; however, lacking BCR-ABL1 fusion gene; and the Near Haploid/High Hyperdiploid (NH-HeH) (51–67 chromosomes) subtype (9,10) is a common subtype, comprising 30% of all pediatric BCP-ALL. These subtypes are clinically relevant with distinct gene expression profile and have been widely studied in the recent past. Nevertheless, we are far from complete understanding of BCP-ALL subtypes and their heterogeneity and its associated molecular mechanisms, which pose a major challenge for improving diagnosis and therapy. Recent studies have suggested that long non-coding RNAs (lncRNAs) and small non-coding RNAs (e.g. microRNAs) might play a key role in

4

96   development and progression of leukemia (11) and thus constitute as new biomarkers and

97   potential targets for novel therapies (12).

98   LncRNAs are arbitrarily defined as transcripts longer than 200 base pairs and lacking an

99   extended protein-coding open reading frame (ORF). It has become apparent that lncRNAs

100  are frequently spliced and polyadenylated and are mainly transcribed by RNA polymerase II

101  (13). LncRNAs expression has been reported as highly tissue-specific even though the

102  expression abundance is generally lower compared to protein-coding genes (14). The

103  expression specificity has been extended to a wide variety of physiological and pathological

104  mechanisms like cancer development and Pluripotency (15). LncRNAs can act either

105  proximally (in the cis region) or distally (in the trans region) for the transcriptional

106  regulation of protein-coding genes (16). Like proteins, various lncRNAs are attributed to

107  oncogenic or tumor-suppressive (17) activities exerting various cellular functions (18). In

108  addition, lncRNAs regulate gene expression at the epigenetic (19) and post-transcription

109  (20) levels. Genome-wide association studies in cancer have disclosed that 80% of cancer-

110  associated single-nucleotide polymorphisms (SNPs) are in non-coding regions (21),

111  including lncRNAs, suggesting that a significant portion of the genetic etiology of cancer

112  can be related to lncRNAs (22). Moreover, lncRNAs are reported to be useful for disease

113  prognosis, exemplified by the lncRNA HOTAIR (HOX transcript antisense RNA), which is

114  up-regulated in acute myeloid leukemia (AML) patients (23). So far, the majority of studies

115  explored the role of single lncRNAs in leukemia including AML (24), chronic lymphocytic

116  leukemia (CLL) (25) and pediatric ALL (26). Yet a comprehensive genomic and epigenetic

117  delineation of lncRNAs deregulations in BCP-ALL subtypes, and their molecular and

118  functional insights are lacking.

119  In the present study, we explored lncRNA landscapes in DUX4, Ph-like, and NH-HeH BCP-

5

120 ALL subtypes and extracted novel biological and functional insights of BCP-ALL subtype-

121 specific lncRNAs and their epigenetic activity. On the basis of RNA-seq transcriptional and

122 DNA methylation survey of lncRNAs, we have determined 1235 subtype-specific, relapse-

123 specific markers and epigenetically altered lncRNAs and demonstrated their relevance in

124 BCP-ALL subtype classification. From our in-depth analyses, we have inferred the potential

125 functions of subtype-specific lncRNAs. Overall, this work provides a most comprehensive

126 and integrative resource which highlights the impact of lncRNAs on relevant pathways that

127 are dysregulated in the molecular subgroups of BCP-ALL and may provide new approaches

128 for prognosis and treatment.

129 **RESULTS**

130 **Unique lncRNAs expression profiles characterize BCP-ALL subtypes**

131 To identify BCP-ALL subtype-specific lncRNAs, we analysed transcriptome profiles from

132 paired initial diagnosis (ID) and relapse (REL) samples of 26 pediatric and 22 adult BCP-

133 ALL patients lacking known chromosomal translocations like BCR-ABL. Based on DNA

134 mutations and chromosomal translocations combined with RNA expression and DNA

135 methylation profiles the samples were classified into different molecular subtypes (Table

136 S1), namely DUX4 (n = 23), Ph-like (n = 21), Near Haploid or High Hyperdiploid (NH-

137 HeH) (n = 16), and low-hypodiploid (LH) (n = 6) and others (n = 18).

138 When the distribution of lncRNAs gene expression levels across all BCP-ALL samples was

139 compared with that of protein-coding genes, the former generally showed lower expression

140 levels than the latter (27) (Fig S1A, Table S1). The principal component analysis (PCA) on

141 the expression (FPKM value) of 13,860 GENCODE lncRNAs revealed three major BCP-ALL

142 subtypes, DUX4, Ph-like and NH-HeH with a distinct separation (Fig 1A). This observation

143 is in concordance to the predefined molecular classification. In particular, samples of the

6

144   DUX4 subtype showed robust separation compared to the remaining samples implying a

145   subtype-specific lncRNAs signatures.

146   To unveil differentially expressed (DE) lncRNAs across these three major molecular

147   subtypes, we performed DE analysis between subtypes. We obtained 1235 significant DE

148   subtype-specific lncRNAs ($P$-value $<= 0.01$ and absolute Fold change $>= +- 1.5$) defining

149   signatures of three subtypes (Fig 1B, Fig S2A-C, Table S1). Of these, 24 lncRNAs were

150   commonly detected in all 3 BCP-ALL subtypes (Fig 1C), about 523 (Hypergeometric $P$-value

151   $= 9.2E-29$) subtype-specific lncRNAs overlapped with deregulated lncRNAs from 12 other

152   cancer types (Fig S2D, Table S1) (28). The remaining 46% (n = 713) of BCP-ALL subtype-

153   specific DE lncRNAs were novel and specific to our subtypes. Out of the overlapped DE

154   lncRNAs (n = 523), 23 (Table S1) were cross-validated in independent cohorts from

155   lnc2cancer (29) database and found to be enriched for oncogenic class of lncRNAs. For

156   example, oncogenic lncRNAs *PVT1* (30) and *GAS5* (31)  are differentially up-regulated in

157   the DUX4 subgroup, and *CRNDE* (32) is DE in Ph-like subgroup. Together, subtype-specific

158   lncRNAs signatures assigned molecular subgroups of BCP-ALL.

159   **Identification and inferred functions of lncRNAs associated with molecular subtypes**

160   **of BCP-ALL**

161   As lncRNAs can function by regulating protein-coding genes in *cis* and/or *trans* (33–36)

162   regions, we performed functional enrichment analyses using guilt-by-association approach

163   based on the correlation between neighbouring (*cis*) and distally (*trans*) located protein-

164   coding (PC) genes (within $\pm$ 100 kb cis and $>\pm$ 100 kb window for trans) of the subtype-

165   specific lncRNAs (see materials and methods). Expression of both *cis* and *trans* PC genes

166   showed a higher tendency towards positive correlation with the expression of the

167   corresponding lncRNAs (Table 1).

7

168 **Table 1: Number of BCP-ALL subtype-specific lncRNAs co-expressed with its *cis* and**

169 ***trans* PC genes.**

| Subtypes | *Cis* PC genes (n = 929) | *Cis* co-expressed DE lncRNAs (n = 62) | *Trans* PC genes (n = 753) | *Trans* co-expressed DE lncRNAs (n = 552) |
|---|---|---|---|---|
| Ph-like | 260 | 170 (383) | 261 | 173 (383) |
| DUX4 | 669 | 451 (736) | 492 | 379 (736) |

170     The table represents the number of DE lncRNAs showed *cis* (<= 100 Kb proximity)

171     and *trans* (>= 100 Kb) protein coding genes and the number of DE lncRNAs co-

172     expressed with them. The numbers shown within the bracket is the total number of

173     DE lncRNAs corresponding to the respective subtypes.

174 Significantly co-expressed (Pearson correlation coefficient => 0.55, 2-tailed *P*-value <=

175 0.05) *cis* and *trans* protein-coding genes associated with DUX4 (n = 58 in *cis* and n = 127

176 in *trans*) and Ph-like (n = 24 in *cis* and n = 20 in *trans*) specific DE lncRNAs demonstrated

177 activation of key signalling pathways involved in proliferation, apoptosis, and

178 differentiation in leukemia (Table S2). For example, in the *cis* based analysis, we identified

179 a strong correlation between DUX4 specific lncRNAs and genes involved in the TGF-beta,

180 Hippo, and P53 signalling pathways (Fig 2A, Table S2). Whereas, the Ph-like specific

181 lncRNAs were correlated with genes involved in JAK-STAT, mTOR, and PIK3-AKT signalling

182 pathways (Fig 2B, Table S2). The *trans* based analysis revealed same vital signalling

183 pathways in DUX4 subtype (Fig S3A-B, Table S2), whereas in Ph-like subtype we identified

184 additional signalling pathways, including, P53 and mitogen-activated protein kinase

185 (MAPK) pathways (Fig S3C, Table S2). The strongly co-expressed *cis* PC genes with DE

186 lncRNAs (n = 32) includes oncogenes including, *IL2RA* (37), *TGFB2* (38), and *CDK6* (39)

187 activated in signalling pathways from DUX4 and Ph-like subgroups (Fig S4A-D, Table 2-3).

188

8

189 **Table 2. Subtype-specific lncRNAs and oncogenes.**

| Subtype-specific lncRNAs | Pearson correlation coefficient | P-value | Oncogene |
|---|---|---|---|
| RP11-347C18.3 | 0.56 | 3.25E-008 | CDK6 |
| RP11-461F16.3 | 0.62 | 5.21E-010 | |
| RP11-96H19.1 | 0.62 | 3.89E-010 | |
| RP11-228B15.4 | 0.64 | 7.68E-011 | |
| MME-AS1 | 0.56 | 3.68E-008 | |
| CTB-39G8.3 | 0.57 | 1.78E-008 | |
| AC002454.1 | 0.72 | 2.21E-014 | |
| RP11-582J16.4 | 0.55 | 8.08E-008 | |
| AC009970.1 | 0.64 | 6.23E-011 | |
| RP11-229P13.20 | 0.66 | 1.44E-011 | |
| LINC00114 | 0.57 | 3.06E-008 | |
| CTB-118N6.3 | 0.61 | 9.70E-010 | |
| SOCS2-AS1 | 0.62 | 4.94E-010 | |
| CTD-2561B21.10 | 0.61 | 9.91E-010 | |
| RP11-413E1.4 | 0.56 | 4.36E-008 | |
| KB-1460A1.1 | 0.55 | 7.77E-008 | |
| AC012309.5 | 0.59 | 4.10E-009 | |
| RP11-37B2.1 | 0.59 | 4.76E-009 | |
| ASB16-AS1 | 0.65 | 3.86E-011 | |
| LINC00426 | 0.62 | 6.32E-010 | |
| LINC01071 | 0.57 | 2.46E-008 | |
| RP11-536K7.5 | 0.74 | 5.11E-15 | IL2RA |
| RP11-224O19.2 | 0.98 | 1.08E-061 | TGFB2 |
| AC004837.5 | 0.83 | 6.11E-023 | |
| RP11-251M1.1 | 0.79 | 7.39E-019 | |
| CTD-2571L23.8 | 0.75 | 2.94E-016 | |
| RP11-35O15.1 | 0.65 | 3.36E-011 | |
| AC139100.3 | 0.58 | 1.00E-008 | |
| RP11-158M2.3 | 0.58 | 1.50E-008 | |
| RP11-672A2.5 | 0.56 | 4.68E-008 | |
| CTD-2357A8.3 | 0.55 | 7.46E-008 | |
| RP11-677M14.3 | 0.55 | 6.68E-008 | |

190 Positively correlating novel *cis* subtype-specific lncRNAs with oncogenes, *CDK6*, *TGFB2*,

191 and *IL2RA* from Ph-like and DUX4 subtypes.

192 **Table 3: Subtype-specific novel DE lncRNAs co-expressed with oncogenes, which are**

193 **associated with vital molecular pathways.**

| Subtype-specific lncRNAs | *Cis* PC | Pearson correlation coefficient | Associated pathways |
|---|---|---|---|
| RP11-224O19.2<br>AC004837.5<br>RP11-251M1.1<br>CTD-2571L23.8 | TGFB2 | 0.98<br>0.83<br>0.79<br>0.75 | Hippo<br>TGF-β<br>Endocytosis |

9

| AC093818.1<br>AC078883.3 | ITGA6 | 0.95<br>0.68 | PI3K-Akt |
|---|---|---|---|
| U62631.5 | CD22 | 0.78 | Cell adhesion molecules (CAMs)<br>B cell receptor signaling pathway |
| CTD-2267D19.2<br>RP11-486L19.2 | RARA | 0.89<br>0.70 | Pathways in cancer<br>Transcriptional mis-regulation in cancer pathways |

194   The table represents the novel subtype specific DE lncRNAs co-expressed with its *cis*

195   genes such as *TGFB2, ITGA6, CD22,* and *RARA* genes, which were enriched in vital

196   molecular pathways in BCP-ALL.

197   However, there were no significant pathways identified within NH-HeH subtype. The

198   subtype-specific *cis* and *trans* acting lncRNAs which are up-regulated and correlated with

199   genes involved in signalling pathways from DUX4 and Ph-like subtypes were hinting their

200   gene expression regulatory activity.

201   We next related the functions of DUX4 and Ph-like specific DE lncRNAs obtained from *cis*

202   based analysis to those functions identified with DE PC genes. We observed an overlap of

203   100% (n = 18, Table S2) of pathways from the DUX4 subtype between lncRNAs based and

204   PC based functional enrichment analysis (Fig 2C). Whereas, in Ph-like subtype, we

205   identified 60% (9 out of 15) same pathways between DE PC based and DE lncRNAs based

206   functional enrichment analysis (Table S2 and Fig 2D). However, we identified Ph-like

207   specific lncRNAs to be more strongly correlated with genes involved in key signalling

208   pathways than Ph-like specific protein-coding genes. For example, we identified mTOR and

209   PI3K-AKT exclusively in the Ph-like specific lncRNAs based analysis. Together, our analyses

210   highlight important functions of BCP-ALL subtype-specific lncRNAs whose expression

211   correlates tightly with that of cancer-related oncogenes.

212   **Relapse-specific lncRNAs driving BCP-ALL progression**

213   To gain insights into the possible role of lncRNAs driving BCP-ALL progression, we

214   investigated dysregulation of lncRNAs at relapse. For each molecular BCP-ALL subtype, we

10

216   performed a differential expression analysis of lncRNAs between ID and REL samples (Fig

217   3). 947 lncRNAs (Table S3) emerged as significantly DE  (absolute Fold change >= +-1.5;

218   *P*-value <= 0.01) between ID and REL from the three subtypes. Around 20% (n = 186) of

219   those DE lncRNAs were up-regulated and 80% were down-regulated at relapse. The

220   hierarchical clustering on relapse-specific lncRNAs within each subtype (DUX4, Ph-like, NH-

221   HeH) identified clear separation between ID and REL (Fig 3A-C). We observed 19% (183)

222   relapse-specific lncRNAs identified here overlapped with subtype-specific lncRNAs (Fig 3E).

223   The putative molecular functions of relapse-specific lncRNAs were identified using the

224   previously mentioned guilt-by-association approach. Relapse-specific lncRNAs within Ph-

225   like and NH-HeH subtypes did not show any significant correlation with activation of

226   pathways. In contrast, in the DUX4 subtype, we identified 56% (n = 321) relapse-specific

227   lncRNAs correlated with *cis* PC genes (Table S3). These strongly correlated relapse-specific

228   lncRNAs showed activation of PC genes involved in vital signalling pathways and metabolic

229   pathways, including NF-kappa B-signalling pathway, cell adhesions molecule (CAMS) and

230   metabolic pathways (number of genes involved >= 3 and *P*-value <= 0.05) (Fig 3D, Table

231   S3). These results indicate that relapse-specific markers from DUX4 subtype may be

232   functionally engaged in metabolic and signalling pathways.

233   **Subtype specific BCP-ALL lncRNAs show epigenetic alterations**

234   For the analysis of the methylation status of loci located at the lncRNAs genomic position in

235   the BCP-ALL subtypes, we used DNA methylation array data (collected from Illumina 450k

236   methylation array) from the same patients (n = 46) including matched diagnosis (ID) and

237   relapse (REL) samples (n = 82). The distribution of DNA methylation levels of CpG sites (n

238   = 60,021, Table S4) associated with 7,160 lncRNAs was compared with CpG sites

239   associated with PC genes across all BCP-ALL samples. Unlike the expression levels, the

11

240    distribution of DNA methylation (hypo-methylation or hyper-methylation) between

241    lncRNAs and PC genes were similar (Fig S1B). Given the robust separation of BCP-ALL

242    subtypes on DNA methylation profile of CpGs associated with lncRNAs on the PCA analysis

243    (Fig 4A), we next studied the differential hypo-methylated (methylation difference value <

244    0; *P*-value <= 0.05) and hyper-methylated (methylation difference value > 0.2; *P*-value

245    <= 0.05) CpGs associated with lncRNAs in each subtype (see materials and method). The

246    hierarchical clustering of differentially methylated (DM) lncRNAs showed distinct

247    methylation patterns of each subtype, concordant with the DE lncRNAs signature (Fig 4B-D,

248    Table S4). In the DUX4 and NH-HeH subtypes the number of hypo-methylated lncRNAs

249    (differential methylation value < 0, *P*-value <= 0.05) were higher compared to the

250    number of hyper-methylated lncRNAs. We classified the DM lncRNAs based on their

251    genomic regions as gene body methylated and promoter-TSS methylated. In the promoter

252    methylated lncRNAs we identified the same trend with high degree of hypo-methylated and

253    lower number hyper-methylated lncRNAs in DUX4 and NH-HeH subtypes. However, the Ph-

254    like subtype has shown a higher degree of hyper-methylated DM lncRNAs than hypo-

255    methylated DM lncRNAs. The list of subtype-specific DM lncRNAs from three subtypes

256    contained previously defined epigenetically altered lncRNAs from other cancer types, for

257    example, we observed the oncogenic lncRNAs *LINC00312* (40), *PVT1*, and *TCL6* (41),

258    which are differentially methylated in at least one of the three subtypes. Together, this data

259    illustrates that epigenetically altered lncRNAs within three BCP-ALL subtypes.

260    **Correlation between differentially expressed and differentially methylated lncRNAs**

261    In order to define whether the subtype-specific promoter methylation impacts on the

262    expression level, we compared the promoter-TSS differential CpG methylated lncRNAs (n =

263    227) with its differential expression signature. We observed 44 lncRNAs with differential

12

264   methylation pattern in their promoter region with differential expression pattern at RNA

265   level. Out these, lncRNAs harboring significant hypo-methylation and hyper-methylation

266   pattern (Pearson correlation, 2-tailed *P*-value <= 0.05) at the promoter region accounted

267   for 23 (Table 4) lncRNAs across the three BCP-ALL subtypes.

268   **Table 4: The list of significantly correlated DNA methylation and expression for**

269   **promoter methylated lncRNAs (n = 23) from BCP-ALL subtypes.**

| DM lncRNAs | Pearson correlation coefficient | *P*-value | Methylation | Absolute Fold change | Subtypes |
|---|---|---|---|---|---|
| AC003075.4 | -0.31 | 0.004 | 1.43 | -1.26 | DUX4 |
| AC099754.1 | -0.32 | 0.002 | -1.74 | 3.2 | |
| AC104655.3 | -0.26 | 0.017 | -2.27 | 2.07 | |
| CACNA1C-AS1 | -0.45 | 2.03E-05 | 1.97 | -1.62 | |
| CTB-25B13.9 | -0.26 | 0.016 | -1.73 | 1.46 | |
| IGF2-AS | -0.24 | 0.028 | -1.33 | 4.95 | |
| LINC01006 | -0.39 | 0.001 | -2.06 | 2.53 | |
| PVT1 | -0.40 | 0.001 | -2.13 | 1.15 | |
| RGMB-AS1 | -0.26 | 0.0193 | -1.48 | 5.96 | |
| RP11-125B21.2 | -0.35 | 0.001 | -1.75 | 4.11 | |
| RP11-138M12.1 | -0.70 | 5.21E-13 | -5.98 | 3.77 | |
| RP11-367G6.3 | -0.30 | 0.004 | 1.98 | -1.63 | |
| RP11-624M8.1 | -0.50 | 1.34E-06 | -3.34 | 4.13 | |
| RP11-789C17.3 | -0.36 | 0.001 | -2.27 | 3.2 | |
| SERTAD4-AS1 | -0.25 | 0.0232 | -1.98 | 1.79 | |
| LINC01006 | -0.38 | 0.0003 | 1.44 | -1.56 | Ph-like |
| RP11-138M12.1 | -0.70 | 5.21E-13 | 2.06 | -1.44 | |
| RP11-305F18.1 | -0.64 | 5.36E-11 | 1.76 | -2.08 | |
| AC099754.1 | -0.33 | 0.002 | 1.21 | -1.36 | |
| ACVR2B-AS1 | -0.36 | 0.0009 | 2.18 | -1.75 | |
| LINC00996 | -0.39 | 0.0003 | -1.56 | 2.11 | |
| ERICH1-AS1 | -0.40 | 0.0006 | -1.82 | 2.21 | |
| DIO3OS | -0.31 | 0.0037 | -1.76 | 4.05 | NH-HeH |
| U3 | -0.83 | 1.346E-22 | -2.01 | 2.43 | |

270   The lncRNAs are promoter differentially methylated and differentially expressed in their

13

271    corresponding subtypes. DM: Differentially methylated. The significance is calculated

272    based on Pearson correlation rate and two -tailed *P*-value <= 0.05.

273    Of these 23 putative epigenetically facilitated lncRNAs, 15 were related to the DUX4

274    subgroup (Fig 5A) including the novel lncRNAs, *R11-138M12.1* and *RP11-624M8.1*. These

275    were significantly hypo-methylated at their promoter region and transcriptionally up-

276    regulated in the DUX4 subgroup (Pearson correlation coefficient = -0.69; *P*-value = 5.1E-

277    13 for *R11-138M12.1*; Pearson correlation coefficient = -0.50; *P*-value = 1.36E-06 for

278    *RP11-624M8.1*; Fig 5B and 5C). In the Ph-like subtype, we observed 7 lncRNAs with

279    promoter methylation (Fig 5D); interestingly, the same lncRNA *R11-138M12.1* showed

280    significant hypermethylation at the promoter region and a concordant down-regulation in

281    the Ph-like subgroup (Fig 5E). Besides these novel lncRNAs, we identified lncRNAs

282    previously reported in the context of different cancers from our epigenetically altered

283    results. The lncRNAs *PVT1* (Pearson correlation coefficient = -0.40, 2-tailed *P*-value <=

284    0.001), and *DIO3OS* (42) (Pearson correlation coefficient = -0.31, 2-tailed *P*-value =

285    0.0037) are examples, which we observed in the DUX4 and NH-HeH subtype with

286    significant anti-correlation (2-tailed *P*-value <= 0.01) to its expression level. Around 46%

287    (n = 512) of DM subtype-specific lncRNAs are localized in the intronic and intergenic

288    genomic regions. We next aimed to investigate whether these lncRNAs regions has

289    chromatin markers encoded within their genomic location. Recent human genome-wide

290    chromatin marker study (43) has provided us with a rich resource to identify chromatin

291    markers. Genome-wide mapping of B-lymphocyte cell line by searching for epigenetic

292    markers within our DM subtype-specific intronic and intergenic regions revealed a

293    significant number of lncRNAs (n=53) (Table S4, Fisher extract test P-value = 2.2E-16)

294    with enchacer and insulator markers (Table S4). Out of these, lncRNAs, *RP11-134O21.1,*

14

295 *RP11-398B16.2, RP11-689B22.2, CTC-458I2.2* and *LINC00880* were DE expressed, with a

296 significant negative correlation between DNA methylation and expression levels in the

297 DUX4 subtype (Table 5) .

298 **Table 5: The list of significantly correlated DNA methylation and expression for**

299 **intronic and Intergenic methylated lncRNAs (n = 5) from DUX4 BCP-ALL subtypes.**

| DM lncRNAs | Absolute Fold change | Methylation value | Pearson correlation rate | *P*-value | Epi-markers | Biotype |
|---|---|---|---|---|---|---|
| RP11-134O21.1 | 2.54 | -1.56 | -0.63 | 1.9E-010 | Enhancer | Intron |
| RP11-398B16.2 | 2.08 | -1.85 | -0.47 | 0.0007 | Insulator | |
| RP11-689B22.2 | 1.52 | -3.37 | -0.47 | 0.008 | Enhancer | |
| CTC-458I2.2 | -1.16 | 3.38 | -0.42 | 0.0001 | Enhancer | |
| LINC00880 | -1.45 | 2.23 | -0.25 | 0.02 | Enhancer | Intergenic |

300 The significance is calculated based on Pearson correlation rate and two -tailed *P*-value

301 <= 0.05. The lncRNAs are promoter differentially methylated and differentially

302 expressed in their corresponding subtypes. These lncRNAs are with enchancer and

303 insulator epigenetic markers. DM: Differentially methylated.

304 These findings suggest that epigenetic silencing and activation of promoter lncRNAs may be

305 a mechanism that contributes to the dysregulation of expression of lncRNAs. In addition to

306 that, both intronic and intergenic DM lncRNAs associated with strong enhancer and

307 insulator regions can accelerate its expression at the epigenetic level.

308 **Epigenetic alterations of subtype-specific lncRNAs are associated with evelvated**

309 **expression of tumor genes located at their *cis* region**

310 We next investigated the relationship between the epigenetic alterations of DM subtype-

311 specific lncRNAs (n = 1118) and the aberrant expressions of their *cis* PC genes. We found

15

312  78 protein-coding genes located in their *cis* region, out of these 33 protein-coding genes

313  have shown a significant up-regulated and down-regulated expression pattern in their

314  corresponding subtypes (Table 6).

315  **Table 6: The list of DNA methylated lncRNAs and differentially up and down-**

316  **regulated *cis* PC genes from three subtypes.**

| Subtype-specific Methylated lncRNAs | DNA methylation | Cis PC | Absolute Fold change | Subtype |
|---|---|---|---|---|
| CTD-2231H16.1 | -1.6803385981 | **PLEKHG4B** | 2.4840131576 | DUX4 |
| RP11-80H8.4 | -4.8712973818 | **CHST2** | 5.021248353 | |
| IGF2-AS | -1.5149692084 | **IGF2** | 6.7074586227 | |
| RP11-332H18.4 | 1.7840373497 | **TBX2** | 5.6459057649 | |
| RP11-332H18.4 | 1.7840373497 | TBX4 | 1.5056957579 | |
| RP11-624M8.1 | -3.3400278742 | HEY2 | 3.5653881965 | |
| CTB-25B13.9 | -1.7341024339 | REEP6 | 1.0135168425 | |
| RGMB-AS1 | -1.4711332705 | **RGMB** | 4.4287098777 | |
| AC073316.1 | -2.1346425435 | **SDK1** | 4.2539854375 | |
| CTB-35F21.2 | -3.3996007483 | CXXC5 | 1.5322350503 | |
| CTB-35F21.2 | -3.5262894443 | PSD2 | 2.0348265047 | |
| AC099754.1 | -1.7432607805 | LRRC3B | 3.2956570803 | |
| AC078883.4 | -3.4089445499 | CORO1C | 0.797282079 | |
| AC078883.4 | -3.4089445499 | **ITGA6** | 3.2489394471 | |
| RP11-314O13.1 | -1.5432811428 | CDYL2 | 1.7070802978 | |
| RP3-455J7.4 | -3.0695619042 | CREG1 | 1.5618184051 | |
| RP11-125B21.2 | -1.749750681 | VLDLR | 4.2858349208 | |
| RP11-676J12.8 | -1.8927665898 | GLOD4 | 0.6471353534 | |
| LINC00114 | 3.4320004844 | ETS2 | -0.7176247124 | |
| RP11-367G6.3 | 1.9798401851 | THBS4 | -1.5796823436 | |
| CTC-458I2.2 | 3.3822300323 | CTGF | -1.892559675 | |
| RP11-69I8.3 | 2.355315514 | AHR | -2.4815111685 | |
| RP11-293A21.1 | -3.6963992484 | STIM2 | 2.487452138 | NH-HeH |
| LINC00114 | -2.9731205546 | ETS2 | 0.9908037963 | |
| ACVR2B-AS1 | 2.3094716973 | ACVR2B | -2.4697071313 | Ph-like |
| AGAP1-IT1 | 2.0926102582 | AGAP1 | -2.9857451706 | |
| RP11-69I8.3 | -3.1689478827 | CTGF | 1.8492946882 | |
| LINC00996 | -1.553999019 | GIMAP8 | 1.4145369937 | |
| AC099754.1 | 1.2121139088 | LRRC3B | -2.0406367867 | |

16

| CTB-79E8.2 | -2.2285747372 | NEURL1B | 1.1469246971 | |
| AL133493.2 | 2.0838061516 | PCBP3 | -3.0795988736 | |
| RP11-420G6.4 | -2.4300037457 | SERPINB1 | 0.5848645847 | |

317   The table represents hypo-methylated and hyper methylated lncRNAs from three

318   subtypes elevating and diminishing the expression of the protein-coding genes localized

319   at their cis regions. The PC genes differentially up and down regulated in the respective

320   subtypes include tumor genes as well. The highlighted rows are some examples of

321   oncogenes with up-regulated expression profile within DUX4 subtype.

322   Intriguingly, the up-regulated PC genes include known tumor genes from various cancer

323   types. For example, *IGF2* (44) (absolute fold change = 6.70, *adj*.P-value = 0.0061), *CTGF*

324   (45) (absolute fold change = 1.85, *adj*.P-value = 0.02) and *ETS2* (46) (absolute fold

325   change = 0.99, *adj*.P-value = 0.01) from DUX4, Ph-like and NH-HeH subtypes respectively

326   (Fig 6 A-C). Together, this illustrates a set of lncRNAs which are capable of epigenetically

327   elevating and silencing the expression profile of tumor genes localized in its *cis* region in

328   BCP-ALL subtypes.

## DISCUSSION

330   Although previous studies have demonstrated the involvement of lncRNAs in acute

331   leukemias (25,26) comprehensive characterization of the transcriptome, epigenetic

332   regulation and functional contribution of lncRNAs in distinct BCP-ALL subtypes are lacking.

333   lncRNAs, as the novel class of functional molecules involved in cancer biology, is defined in

334   the stratification of different molecular subtypes in various cancers (47–49). However, their

335   role in BCP-ALL subtypes has not been investigated. Here, we unravel the lncRNAs

336   landscape using transcriptome and methylome data from 46 (adult and pediatric) relapsed

337   BCP-ALL patients focusing on the three molecular subtypes namely, DUX4, Ph-like, and NH-

338   HeH. Our integrated transcriptomic analyses using RNA-seq and DNA methylation brings

17

339  significant insights and advances over other studies: it provides the most comprehensive

340  novel datasets so far for BCP-ALL subtypes, a resource of subtype-specific and relapse-

341  specific lncRNAs, potential lncRNAs functions and uncovers their epigenetic alterations of

342  the BCP-ALL subtypes. We identified 1235 DE subtype-specific lncRNAs dysregulated in at

343  least one of the three subtypes. Compared to the pan-cancer comprehensive set of

344  aberrantly expressed lncRNAs we found 66% (712 out of 1564) of our DE lncRNAs were

345  more specific for our subtypes introducing novel insight in the non-protein-coding part of

346  the genome in BCP-ALL subtypes.

347  Another important aspect of our study is the identification of relapse-specific dysregulated

348  lncRNAs across three BCP-ALL subtypes. A closer look into the relapse-specific lncRNAs

349  signature identified lncRNAs previously described as oncogenic: lncRNAs including, *RP11-*

350  *701P16.5 (50)*, *SLC38A3 (51)*, and *LINC00312 (40)*,  which are up-regulated in relapsed

351  samples within DUX4 subtype (Table 7).

352  **Table 7: Previously reported lncRNAs identified as relapse-specific lncRNAs in BCP-**

353  **ALL subtypes.**

| Relapse-specific lncRNAs | Disease association |
|---|---|
| TCL6 (DUX4) | Chromosomal translocations T-cell leukaemia/lymphoma (2) |
| LINC00312 (DUX4, Ph-like, NH-HeH) | Proliferation, invasion, and migration of thyroid cancer, Nasopharyngeal carcinoma (3) |
| miR-17-92a-1 (DUX4, Ph-like, NH-HeH) | Development, progression, and aggressiveness of colorectal cancer (4) |

354  The differentially expressed lncRNAs between relapse (REL) and initial diagnosis

355  (ID), from three subtypes, which were previously reported for its disease association,

356  selected representative examples from relapse-specific lncRNAs, which are previously

357  identified in other diseases.

18

358  Importantly, apoptosis suppressor lncRNA in myc-driven lymphomas (52) *miR-17/92*

359  cluster host gene (*MIR17HG*) is up-regulated in relapse samples within the Ph-like subtype

360  and down regulated in relapsed samples within DUX4 and NH-HeH subtypes. Overall, the

361  relapse-specific lncRNAs highlights the oncogenic relevance in BCP-ALL subtypes.

362  Besides the oncogenic properties, lncRNAs can act as prognostic markers (53) and aid for

363  disease diagnosis and treatment. A subset of our relapse-specific lncRNAs (n = 61, Table

364  S3) is recently identified as prognostic markers in 14 Pan-Cancer data (36) types, including

365  Lung Cancer Associated Transcript 1 (*LUCAT1*), which is previously reported for its drug

366  resistance in solid cancer (54). Within the DUX4 subtype, we identified up-regulated

367  expression of *LUCAT1* at relapse, potentially providing a novel insight into treatment

368  resistance for BCP-ALL subtypes. Together, this illustrates the catalog of relevant lncRNAs

369  in different subtypes of BCP-ALL as subtype-specific and relapse-specific markers with the

370  potential of RNA based treatments in the treatment of BCP-ALL subtypes.

371  The dissection of the regulatory pathways mediated by the molecular subtype-specific and

372  relapse-specific lncRNAs revealed the activation of pivotal signalling pathways across three

373  BCP-ALL subtypes. The functional analysis using guilt-by-association approach highlights

374  the subtype-specific and relapse-specific lncRNAs associated with activation of signaling

375  pathways and metabolic pathways that are associated with leukemogenesis including TGF-

376  Beta, hippo, P53, and JAK-STAT, cytokine-cytokine receptor, endocytosis, mTOR and

377  metabolic pathways. Characterization of the lncRNAs involved in this pathway may

378  potentially reveal novel targets in molecular therapies.

379  The functional insights of relapse-specific and subtype-specific lncRNAs revealed biological

380  relevance to BCP-ALL subtypes including cell cycle functions, signal transduction, cell

381  migration and metabolic processes. Some of the functions predicted here corroborate

19

382   previous studies emphasizing the strengths of the employed guilt-by-association. For

383   example, lncRNA *AC002454.1*, which we associated to the PIK3-AKT pathway in Ph-like

384   subtype, was recently reported to regulate cyclin-dependent kinase (*CDK6*) to participate in

385   cell cycle disorder (55). The *CDK6* gene appears to be frequently dysregulated in

386   hematopoietic malignancies (39) and is hence attributed a critical role in tumorigenesis,

387   also shown in ALL driven by mixed lineage leukemia fusion proteins (56). In Ph-like

388   subtype, both *CDK6* and *AC002454.1* are correlated and up-regulated specifically in Ph-like

389   samples, suggesting they displayed enhancer-like functions. We identified 8 relapse-specific

390   lncRNAs (Table S3) associated with metabolic pathways in the DUX4 subtype overlapping

391   with lncRNAs reported (57) to synergistically dysregulate metabolic pathways in multiple

392   tumour context.

393   Besides known lncRNAs, we also identified novel lncRNAs associated with activation of key

394   signalling pathways. For instance, in DUX4 subtype, we identified a set of novel lncRNAs

395   associated with TGF-beta pathway, including the antisense *RP11-224019.2*, with a

396   significant positive correlation to the *TGFB* gene. Recently, there are a number of lncRNAs

397   documented to be associated with TGFß signalling pathway, including MEG3 regulating the

398   TGFB2 pathway in breast cancer (58). However, lncRNAs associated with the TGFß

399   pathway in BCP-ALL subtypes have not been reported. The co-expression of *RP11-224019.2*

400   and *TGFB* in DUX4 subtype may indicate their functional relatedness or regulatory

401   relationships. In addition to that, a notable observation was a strong correlation between

402   relapse-specific lncRNAs with genes involved in the activation of metabolic pathways in the

403   DUX4 subtype. We identified 112 relapse-specific lncRNAs co-expressed with 29 (Table S3)

404   PC genes activated in metabolic pathways, including previously reported 8 biomarker

405   lncRNAs. For example, we identified oncogenic lncRNA *LUCAT1* reported to be associated

20

406     with poor prognosis in lung cancer (54). However, the *LUCAT1* has not yet been reported in

407     the BCP-ALL context. The global co-expression analysis and gene-expression profiling

408     suggest important and previously unappreciated roles of lncRNAs in the BCP-ALL subtypes.

409     Our analyses provide important functions of subtype-specific and relapse-specific lncRNA

410     genes whose expression correlates tightly with oncogenic coding genes.

411     Although we observed that subtype-specific lncRNAs and subtype-specific protein-coding

412     genes were predicted to activate or inhibit the same pathways, some important exclusivity

413     was observed. For instance, the signalling pathways such as the PI3K and mTOR in Ph-like

414     subtype was enriched only in the lncRNAs based enrichment analysis, whereas these

415     pathways did not appear to be enriched/dysregulated in the mRNA based analysis. The

416     PI3K and mTOR signalling pathways control proliferation, differentiation, and survival of

417     hematopoietic cells (59). Consistent with our studies, other studies indicated the potency of

418     lncRNAs facilitating the cancer cell growth through mTOR and PI3K signalling pathways

419     (36,47,60) yet reports on BCP-ALL subtypes are lacking. Considering the functional nexus

420     between Ph-like specific lncRNAs and the activation of pathways such as mTOR and PI3K

421     signalling pathways, targeting those lncRNAs may be a promising novel therapeutic target

422     for BCP-ALL subtypes.

423     Our work additionally underscores the importance of epigenetic alterations in modulating

424     lncRNAs transcriptional activities. Although previous studies (16, 57) have demonstrated

425     cross-talk between DNA methylation and transcriptional activities of lncRNAs, their role in

426     the etiology of BCP-ALL subtypes has not been investigated. DNA methylation analyses of

427     lncRNAs revealed that DNA methylation might underlie the differential expression of BCP-

428     ALL subtype-specific lncRNAs. Subtype-specific lncRNAs identified here have been reported

429     by previous studies. For example, *SOX2-OT* (62), *LINC00312* (63)*, TCL6* and *PVT1*, are

21

430 onco-lncRNAs, which are promoter methylated in one of the three subtypes. The lncRNA,

431 *PVT1* was reported for its MYC activity (64,65) and as oncogenic lncRNA with multiple

432 roles in cell growth, dysfunction, and differentiation in AML (66). Both lncRNAs,

433 *LINC00312* and *TCL6* have been extensively investigated on expression levels but not on the

434 epigenetic level. The promoters of both *TCL6* and *LINC00312* were observed to be hyper-

435 methylated with corresponding diminished expression in the DUX4 and NH-HeH samples.

436 Notably, the DNA methylation analysis of lncRNAs revealed that DNA methylation might

437 underlie the differential expression of subtype-specific lncRNAs. Our analysis identified 23

438 subtype-specific lncRNAs showing hypo-methylation and hyper-methylation pattern at their

439 promoter region that are significantly correlated with their diminished and increased

440 expression in respective subtypes. In addition to that, we have identified 33 epigenetically

441 co-regulated oncogenes localized in the *cis* regions of hypo-methylated and hyper-

442 methylated lncRNAs from three subtypes. Interestingly, the oncogene associated with

443 leukemia, IGF2 (44) has shown an elevated expression level in Ph-like subtypes

444 corresponding to the hypo-methylation of its antisense, IGF2-AS1. These findings suggest

445 that epigenetic silencing of lncRNAs genes may be a mechanism that contributes to the

446 dysregulation of expression of lncRNAs and their *cis* genes in BCP-ALL subtypes.

447 Overall, our study provides an in-depth analysis of the lncRNA transcriptome and

448 epigenome in BCP-ALL subtypes and provides numerous new lncRNAs markers associated

449 with subtype and relapse-specificity and with epigenetic alterations in BCP-ALL subtypes.

450 Additionally, we also demonstrated these lncRNAs might contribute to the regulation of key

451 signalling pathways involved in BCP-ALL. In summary, our study provides a comprehensive

452 set of dysregulated lncRNAs from BCP-ALL subtypes derived using different integrative

453 approaches. This can serve as a major resource of BCP-ALL subtype-specific lncRNAs and

22

454 their mechanisms of action in detail that might pave the way for the future studies to

455 investigate key biomarkers and potential therapeutic targets in BCP-ALL subtypes.

## Materials and Methods

### Patient samples

458 Patients (n = 46) used in this project were negatively selected for fusion genes detectable

459 by routine diagnostic workup (BCR-ABL, MLL translocations, ETV6-RUNX1) from 26

460 pediatric and 22 adult patients. From these patients we collected 44 samples at initial

461 diagnosis (ID) and 44 samples with relapse (REL). All patients were treated in population-

462 based German study trials (GMALL for adult and BFM for pediatric patients). A written

463 informed consent to participate in these trails according to the Declaration of Helsinki was

464 obtained from all patients. The studies were approved by the ethics board of Charité,

465 Berlin.

### Overview of RNA-seq and DNA methylation array data

467 To generate transcriptome profiles of patient samples, mRNA was isolated by using Trizol

468 reagent (Life Technologies, Grand Island, NY) procedure from the bone marrow

469 mononuclear cells (MNCs) of the ID and REL samples. The paired-end RNA sequencing was

470 done on Illumina HiSeq4000 platform (multiplexing) in the high throughput sequencing

471 core facility, German Cancer Research Center, Heidelberg, Germany. The RNA-Seq was

472 performed by using six samples per lane, which resulted in an average of 64 Million

473 mapped paired reads per sample. For methylation, genomic DNA was isolated using

474 unstranded Allprep extraction (Qiagen, Hilden, Germany) from the bone marrow of same

475 patients (ID and REL samples) was then hybridized onto an Illumina Infinium

476 HumanMethylation450K. From the DNA methylation chip we identified 60,021 probes

477 annotated to 7,190 lncRNAs.

23

478 **RNA-seq read alignment and quantitative extraction**

479 The STAR aligner (version 2.4.0.1) (67) (2-pass alignment parameters) was used to align

480 paired-end reads to the human genome reference. The human genome reference files used

481 for processing RNA-seq samples were the hg19 (GRCh37) genome version for alignment

482 and transcript annotation from GENCODE version 19 (equivalent Ensembl GRCh37). The

483 transcriptome construction and gene-level counts for each sample were obtained using

484 StringTie (68). The read count information from the files generated by StringTie was

485 extracted using the "prepDE.py" python script provided by the StringTie. We detected 84%

486 of 13,860 lncRNAs (including 23,898 transcripts) annotated by GENCODE (V19) from our

487 samples (FPKM > 0 for multi-exon lncRNAs and FPKM > 0 for single exonic lncRNAs)

488 showing that our sequencing depth was good.

489 **Sample clustering and differential expression analysis for subtype- specific and**

490 **relapse-specific lncRNAs**

491 We performed PCA using the *prcomp* R function on 13,860 lncRNAs from RNA-seq and

492 60,021 CpG's on lncRNAs from DNA methylation datasets. The PCA plots were plotted

493 using python matplotlib axes3D function. The R bioconductor package Linear Models for

494 Microarray (*LIMMA) Voom* (69) was used on gene-level expression data for identifying the

495 subtype-specific and relapse-specific differentially expressed (DE) lncRNAs. The subtype-

496 specific DE lncRNAs were identified by implementing separate design matrix for the three

497 subtypes (DUX4, Ph-like and NH-HeH). Within each subtype, we used using all subtype

498 samples versus the rest of the cohort. Within our cohort (82 samples from 46 patients), not

499 all patients had matching ID and REL samples, and moreover, we wanted to compare across

500 subtypes. LIMMA voom leveraged the sample imbalances and confounder (patient and

501 samples) with its *duplicatecorrelation* function. We implemented *duplicatecorrelation*

24

502  function which addressed all patient effects by estimating correlations of multiple samples

503  from the same patient while allowing us to compare across the subtypes. Additionally, we

504  included the ID and REL time factors into the design (*makeContrasts*) to avoid the inflation

505  of the variance due to time factor for each subtype. The relapse-specific DE lncRNAs within

506  each subtype were identified by analysing DE lncRNAs ID versus REL samples within each

507  subtype separately. The significant DE genes were assigned based on the p-value < 0.01

508  and Fold change of >= +-1.5. The lncRNAs from GENCODE version 19 (equivalent

509  Ensembl GRCh37) were used as reference annotation. The heatmaps and correlation based

510  (Spearman method) hierarchical clustering of DE lncRNAs were performed on z-score

511  transformed LIMMA normalized gene expression values using the R Bioconductor package

512  ComplexHeatmap.

513  **Differential methylation data analysis**

514  The ID and REL samples from the same patients have been assayed with the Illumina 450k

515  methylation array. All the beta values have been normalized using the Subset-quantile

516  Within Array Normalization (SWAN) method. In order to detect differentially methylated

517  regions, we used the R package bumphunter (70) using the most variant quartile of the

518  CpG probes. Bumphunter searches for differentially methylated regions in an annotation-

519  unbiased manner. Separate bumphunter runs have been performed for ID and REL samples

520  for every three subtypes (DUX4, Ph-like, and NH-HeH), using all subtype samples versus

521  the rest of the cohort. We associated the differentially methylated regions from three BCP-

522  ALL subtypes using HOMER (hypergeometric optimization of motif enrichment) suite of

523  tool with the reference file GRCh37.74, using the -gene parameter. HOMER provided us

524  with annotation of each probe, we separated lncRNAs from the output. The genomic

525  regions were divided into promoter (+-2kb from transcription start site, TSS) and gene

25

526   body. The gene body was defined if the CpGs were annotated in exonic, intronic or

527   transcription termination site (TTS). The regions mapped to lncRNAs were then used for

528   analysis. The significantly differentially hyper-methylated (Methylation difference value

529   $>= 0.2$; *P*-value $<= 0.05$) and hypo-methylated (Methylation difference value $=< 0$; *P*-

530   value $<= 0.05$) regions were used for further analysis. The intronic and intergenic

531   differentially methylated (DM) lncRNAs were then mapped using 'BedTools' with the B-

532   lymphocyte cell line "wgEncodeBroadHmmGm12878HMM.bed" in order to find the

533   epigenetic markers. The significance of enrichment was calculated using Fisher's exact test.

534   The epigenetically altered lncRNAs were assigned if promoter methylated lncRNAs  were

535   differentially expressed and their DNA methylation values (log-transformed Beta values)

536   and expression values (log-transformed FPKM values) are correlated. The most significant

537   correlations (Pearson correlations coefficient , 2-tailed *P*-value $<= 0.05$) were classified

538   later called as epigenetically altered lncRNAs.

539   **Functional predictions using guilt-by-association approach**

540   In our study, we used the "guilt-by-association" (71) approach by establishing the pairwise

541   expression correlations between DE lncRNAs (from all BCP-ALL subtypes) and it's *cis* and

542   *trans* protein-coding (PC) genes to predict the functions of subtype-specific lncRNAs. We

543   located the *cis* and *trans* PC genes of DE lncRNAs using the GREAT tool (version v3.0.0)

544   (72). All PC genes from GENCODE v19 annotation (n = 20698) were used in the analysis.

545   The individual *cis* and *trans* genes for each DE lncRNAs were located within a genomic

546   window of 100 kb and greater >100kb respectively. From each dataset, we then computed

547   the pairwise expression correlation using Pearson correlation method between each

548   lncRNAs and its *cis* and *trans* coding gene. The significantly co-expressed PC genes (Pearson

549   correlation coefficient $>= 0.55$ and 2-tailed *P*-value $<= 0.05$) were further used for

26

550 functional enrichment analysis using GeneSCF v1.0 (73). The functional enrichment

551 analysis was performed using the KEGG database with a background of all protein-coding

552 genes from GENCODE v19 (20,345). The functional terms were considered significant only

553 if it is enriched with *P*-value $<= 0.05$.

## ACKNOWLEDGEMENT

## Author Contributions

558 The project was conceived and designed by: ARJ, CDB, MN, AA. ARJ developed

559 bioinformatics pipeline and analyzed RNA-seq data. MN normalized the methylation data.

560 MSP performed DNA methylation data analysis. MPS, LB, MN and CDB performed the

561 analyses of clinical and molecular data. JOT, CS and KI performed the sample preparation.

562 CMT, MAR, MB, NG, RKS, AVS, MS, MH, TB, SS, HS, SG, RKS, CE were involved in the

563 sample collection, genetic characterization and provided molecular diagnostic data. All

564 authors were involved in writing and reviewing the manuscript.

## REFERENCES

1.  Portell CA, Wenzell CM, Advani AS. Clinical and pharmacologic aspects of blinatumomab in the treatment of B-cell acute lymphoblastic leukemia. Clinical Pharmacology: Advances and Applications. 2013. p. 5–11.

2.  Ito C, Kumagai M, Manabe A, Coustan-Smith E, Raimondi SC, Behm FG, et al. Hyperdiploid acute lymphoblastic leukemia with 51 to 65 chromosomes: a distinct biological entity with a marked propensity to undergo apoptosis. Blood. 1999;93(1):315–20.

3.  Heerema NA, Nachman JB, Sather HN, Sensel MG, Lee MK, Hutchinson R, et al. Hypodiploidy with less than 45 chromosomes confers adverse risk in childhood acute lymphoblastic leukemia: a report from the children's cancer group. Blood. 1999;94(12):4036–45.

4.  Dun KA, Vanhaeften R, Batt TJ, Riley LA, Diano G, Williamson J. BCR-ABL1 gene rearrangement as a subclonal change in ETV6-RUNX1-positive B-cell acute lymphoblastic leukemia. Blood Adv [Internet]. 2016 Dec;1(2):132—138. Available from: http://europepmc.org/articles/PMC5737165

5.  Lilljebjörn H, Fioretos T. New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. Blood. 2017. p. 1395–401.

6.  Mullighan CG. Genomic profiling of B-progenitor acute lymphoblastic leukemia. Best Pract Res Clin Haematol [Internet]. 2011;24(4):489–503. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1521692611000843

7.  Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, Olsson L, Orsmark-Pietras C, Von Palffy S, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. Nat Commun. 2016;7.

8.  Jain N, Roberts KG, Jabbour E, Patel K, Eterovic AK, Chen K, et al. Ph-like acute lymphoblastic leukemia: A high-risk subtype in adults. Blood. 2017;129(5):572–81.

9.  Paulsson K, Lilljebjörn H, Biloglav A, Olsson L, Rissler M, Castor A, et al. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. Nat Genet. 2015;47(6):672–6.

10. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet. 2013;45(3):242–52.

11. Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, et al.

28

598    Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl
599    Acad Sci U S A [Internet]. 2014;111(52):18679–84. Available from:
600    http://www.pubmedcentral.nih.gov/articlerender.fcgi?
601    artid=4284555&tool=pmcentrez&rendertype=abstract

602  12.  Chen S, Liang H, Yang H, Zhou K, Xu L, Liu J, et al. Long non-coding RNAs: The
603    novel diagnostic biomarkers for leukemia. Environmental Toxicology and
604    Pharmacology. 2017. p. 81–6.

605  13.  Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: Past, present, and future.
606    Genetics. 2013;193(3):651–69.

607  14.  Ward M, McEwan C, Mills JD, Janitz M. Conservation and tissue-specific transcription
608    patterns of long noncoding RNAs. J Hum Transcr [Internet]. 2015;1(1):2–9.
609    Available from:
610    http://www.tandfonline.com/doi/full/10.3109/23324015.2015.1077591

611  15.  Wang P, Ning S, Zhang Y, Li R, Ye J, Zhao Z, et al. Identification of lncRNA-
612    associated competing triplets reveals global patterns and prognostic markers for
613    cancer. Nucleic Acids Res. 2015;43(7):3478–89.

614  16.  Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs.
615    Nature. 2012. p. 339–46.

616  17.  Liu Q, Huang J, Zhou N, Zhang Z, Zhang A, Lu Z, et al. LncRNA loc285194 is a p53-
617    regulated tumor suppressor. Nucleic Acids Res. 2013;41(9):4976–87.

618  18.  Spizzo R, Almeida MI, Colombatti A, Calin GA. Long non-coding RNAs and cancer: A
619    new frontier of translational research. Oncogene. 2012. p. 4577–87.

620  19.  Sati S, Ghosh S, Jain V, Scaria V, Sengupta S. Genome-wide analysis reveals distinct
621    patterns of epigenetic features in long non-coding RNA loci. Nucleic Acids Res.
622    2012;40(20):10018–31.

623  20.  Yoon J-H, Abdelmohsen K, Gorospe M. Posttranscriptional Gene Regulation by Long
624    Noncoding RNA. J Mol Biol. 2013;425(19):3723–30.

625  21.  Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the
626    genetics of cancer. Br J Cancer [Internet]. 2013;108(12):2419–25. Available from:
627    http://www.nature.com/doifinder/10.1038/bjc.2013.233

628  22.  Gong J, Liu W, Zhang J, Miao X, Guo A-Y. lncRNASNP: a database of SNPs in
629    lncRNAs and their potential functions in human and mouse. Nucleic Acids Res
630    [Internet]. 2015;43(D1):D181–6. Available from:
631    https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku1000

632  23.  Zhang Y-Y, Huang S-H, Zhou H-R, Chen C-J, Tian L-H, Shen J-Z. Role of HOTAIR in

29

633       the diagnosis and prognosis of acute leukemia. Oncol Rep. 2016;36(6):3113–22.

634   24.   Wei S, Wang K. Long noncoding RNAs: pivotal regulators in acute myeloid leukemia.
635       Exp Hematol Oncol [Internet]. 2015;5(1):30. Available from:
636       http://ehoonline.biomedcentral.com/articles/10.1186/s40164-016-0059-9

637   25.   Ronchetti D, Manzoni M, Agnelli L, Vinci C, Fabris S, Cutrona G, et al. LncRNA
638       profiling in early-stage chronic lymphocytic leukemia identifies transcriptional
639       fingerprints with relevance in clinical outcome. Blood Cancer J. 2016;6(9).

640   26.   Gioia R, Drouin S, Ouimet M, Caron M, St-Onge P, Richer C, et al. LncRNAs
641       downregulated in childhood acute lymphoblastic leukemia modulate apoptosis, cell
642       migration, and DNA damage response. Oncotarget [Internet]. 2017; Available from:
643       http://www.oncotarget.com/fulltext/20817

644   27.   Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE
645       v7 catalog of human long noncoding RNAs: Analysis of their gene structure,
646       evolution, and expression. Genome Res. 2012;22(9):1775–89.

647   28.   Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, et al. Comprehensive Genomic
648       Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell.
649       2015;28(4):529–40.

650   29.   Ning S, Zhang J, Wang P, Zhi H, Wang J, Liu Y, et al. Lnc2Cancer: A manually
651       curated database of experimentally supported lncRNAs associated with various
652       human cancers. Nucleic Acids Res. 2016;44(D1):D980–5.

653   30.   Colombo T, Farina L, Macino G, Paci P. PVT1: A rising star among oncogenic long
654       noncoding RNAs. BioMed Research International. 2015.

655   31.   Pickard MR, Mourtada-Maarabouni M, Williams GT. Long non-coding RNA GAS5
656       regulates apoptosis in prostate cancer cell lines. Biochim Biophys Acta.
657       2013;1832(10):1613–23.

658   32.   Petri A, Dybkær K, Bøgsted M, Thrue CA, Hagedorn PH, Schmitz A, et al. Long
659       noncoding RNA expression during human B-cell development. PLoS One.
660       2015;10(9).

661   33.   Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. Nat Struct Mol Biol
662       [Internet]. 2012;19(11):1068–75. Available from:
663       http://www.nature.com/doifinder/10.1038/nsmb.2428

664   34.   Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, et al. MEG3 long
665       noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA
666       triplex structures. Nat Commun. 2015;6.

667   35.   Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, et al.

668    Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional
669    Silencing through Chromatin-Level Regulation. Mol Cell. 2008;32(2):232–46.

670 36. Ali MM, Akhade VS, Kosalai ST, Subhash S, Statello L, Meryet-Figuiere M, et al. PAN-
671    cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and
672    biomarkers. Nat Commun. 2018;9(1).

673 37. Sadras T, Heatley SL, Kok CH, Dang P, Galbraith KM, McClure BJ, et al. Differential
674    expression of MUC4, GPR110 and IL2RA defines two groups of CRLF2-rearranged
675    acute lymphoblastic leukemia patients with distinct secondary lesions. Cancer Lett.
676    2017;408:92–101.

677 38. Dong M, Blobe GC. Role of transforming growth factor-beta in hematologic
678    malignancies. Blood [Internet]. 2006;107(12):4589–96. Available from:
679    http://www.pubmedcentral.nih.gov/articlerender.fcgi?
680    artid=1895802&tool=pmcentrez&rendertype=abstract

681 39. Scheicher R, Hoelbl-Kovacic A, Bellutti F, Tigan A-S, Prchal-Murphy M, Heller G, et
682    al. CDK6 as a key regulator of hematopoietic and leukemic stem cell activation.
683    Blood. 2015. 90-101 p.

684 40. Wang Y-Y, Wu Z-Y, Wang G-C, Liu K, Niu X-B, Gu S, et al. LINC00312 inhibits the
685    migration and invasion of bladder cancer cells by targeting miR-197-3p. Tumor Biol
686    [Internet]. 2016;37(11):14553–63. Available from:
687    http://link.springer.com/10.1007/s13277-016-5303-8

688 41. Saitou M, Sugimoto J, Hatakeyama T, Russo G, Isobe M. Identification of the TCL6
689    genes within the breakpoint cluster region on chromosome 14q32 in T-cell leukemia.
690    Oncogene. 2000;19:2796–802.

691 42. Mitra R, Chen X, Greenawalt EJ, Maulik U, Jiang W, Zhao Z, et al. Decoding critical
692    long non-coding RNA in ovarian cancer epithelial-to-mesenchymal transition. Nat
693    Commun. 2017;8(1).

694 43. ENCODE Project Consortium AIE of DE in the H. An integrated encyclopedia of DNA
695    elements in the human genome. Nature. 2012;489(7414):57–74.

696 44. Wu HK, Weksberg R, Minden MD, Squire JA. Loss of imprinting of human insulin-like
697    growth factor II gene, IGF2, in acute myeloid leukemia. Biochem Biophys Res
698    Commun. 1997;231(2):466–72.

699 45. Lu H, Kojima K, Battula VL, Korchin B, Shi Y, Chen Y, et al. Targeting connective
700    tissue growth factor (CTGF) in acute lymphoblastic leukemia preclinical models:
701    Anti-CTGF monoclonal antibody attenuates leukemia growth. Ann Hematol.
702    2014;93(3):485–92.

703 46. Fu L, Fu H, Wu Q, Pang Y, Xu K, Zhou L, et al. High expression of ETS2 predicts poor

704 prognosis in acute myeloid leukemia and may guide treatment decisions. J Transl
705 Med. 2017;15(1).

706 47. Van Grembergen O, Bizet M, de Bony EJ, Calonne E, Putmans P, Brohée S, et al.
707 Portraying breast cancers with long noncoding RNAs. Sci Adv [Internet].
708 2016;2(9):e1600220. Available from:
709 http://www.ncbi.nlm.nih.gov/pubmed/27617288%5Cnhttp://www.pubmedcentral.
710 nih.gov/articlerender.fcgi?artid=PMC5010371

711 48. Zhao W, Luo J, Jiao S. Comprehensive characterization of cancer subtype associated
712 long non-coding RNAs and their clinical implications. Sci Rep [Internet].
713 2014;4:6591. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?
714 artid=4194441&tool=pmcentrez&rendertype=abstract

715 49. Chen H, Xu J, Hong J, Tang R, Zhang X, Fang JY. Long noncoding RNA profiles
716 identify five distinct molecular subtypes of colorectal cancer with clinical relevance.
717 Mol Oncol. 2014;8(8):1393–403.

718 50. Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, et al. A potential signature of eight
719 long non-coding RNAs predicts survival in patients with non-small cell lung cancer. J
720 Transl Med [Internet]. 2015;13(1):231. Available from: http://www.translational-
721 medicine.com/content/13/1/231

722 51. Wang Y, Fu L, Cui M, Wang Y, Xu Y, Li M, et al. Amino acid transporter SLC38A3
723 promotes metastasis of non-small cell lung cancer cells by activating PDK1. Cancer
724 Lett. 2017;393:8–15.

725 52. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: A comprehensive update on its
726 genomics, genetics, functions and increasingly important and numerous roles in
727 health and disease. Cell Death Differ [Internet]. Nature Publishing Group;
728 2013;20(12):1603–14. Available from: http://dx.doi.org/10.1038/cdd.2013.125

729 53. Zhong L, Lou G, Zhou X, Qin Y, Liu L, Jiang W. A six-long non-coding RNAs signature
730 as a potential prognostic marker for survival prediction of ER-positive breast cancer
731 patients. Oncotarget. 2017;8(40):3762–72.

732 54. Han Z, Shi L. Long non-coding RNA LUCAT1 modulates methotrexate resistance in
733 osteosarcoma via miR-200c/ABCB1 axis. Biochem Biophys Res Commun [Internet].
734 Elsevier Ltd; 2018;495(1):947–53. Available from:
735 https://doi.org/10.1016/j.bbrc.2017.11.121

736 55. Wang Y, Li Y, Yang Z, Liu K, Wang D. Genome-wide microarray analysis of long non-
737 coding RNAs in Eutopic secretory endometrium with endometriosis. Cell Physiol
738 Biochem. 2015;37(6):2231–45.

739 56. Placke T, Faber K, Nonami A, Putwain SL, Salih HR, Heidel FH, et al. Requirement

32

740    for CDK6 in MLL-rearranged acute myeloid leukemia. Blood. 2014;124(1):13–23.

741  57.  Chiu HS, Somvanshi S, Patel E, Chen TW, Singh VP, Zorman B, et al. Pan-Cancer
742    Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each
743    Tumor Context. Cell Rep. 2018;23(1):297–312.e12.

744  58.  Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, et al. MEG3 long
745    noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA
746    triplex structures. Nat Commun [Internet]. 2015;6:7743. Available from:
747    http://www.nature.com/ncomms/2015/150724/ncomms8743/full/ncomms8743.ht
748    ml
749    %5Cnhttp://www.nature.com/doifinder/10.1038/ncomms8743%5Cnhttp://www.nc
750    bi.nlm.nih.gov/pubmed/26205790

751  59.  Neri LM, Cani A, Martelli AM, Simioni C, Junghanss C, Tabellini G, et al. Targeting
752    the PI3K/Akt/mTOR signaling pathway in B-precursor acute lymphoblastic leukemia
753    and its therapeutic potential. Leukemia. 2014;28(4):739–48.

754  60.  Hui Z, Xianglin M, Access O. Association of HOTAIR expression with PI3K / Akt
755    pathway activation in adenocarcinoma of esophagogastric junction. 2016;36–40.

756  61.  Feng N, Ching T, Wang Y, Liu B, Lin H, Shi O, et al. Analysis of microarray data on
757    gene expression and methylation to identify long non-coding RNAs in non-small cell
758    lung cancer. Sci Rep. 2016;6.

759  62.  Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, et al.
760    Emerging role of long non-coding RNA SOX2OT in SOX2 regulation in breast cancer.
761    PLoS One. 2014;9(7).

762  63.  Zhang W, Huang C, Gong Z, Zhao Y, Tang K, Li X, et al. Expression of LINC00312, a
763    long intergenic non-coding RNA, is negatively correlated with tumor size but
764    positively correlated with lymph node metastasis in nasopharyngeal carcinoma. J Mol
765    Histol. 2013;44(5):545–54.

766  64.  Li JR, Sun CH, Li W, Chao RF, Huang CC, Zhou XJ, et al. Cancer RNA-Seq Nexus: A
767    database of phenotype-specific transcriptome profiling in cancer cells. Nucleic Acids
768    Res. 2016;44(D1):D944–51.

769  65.  Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, et al. PVT1
770    dependence in cancer with MYC copy-number increase. Nature. 2014;512(1):82–6.

771  66.  L´Abbate A, Tolomeo D, Cifola I, Severgnini M, Turchiano A, Augello B, et al. MYC-
772    containing amplicons in acute myeloid leukemia: genomic structures, evolution, and
773    transcriptional consequences. Leukemia. 2018;1–15.

774  67.  Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast
775    universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

33

776  68.  Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie
777      enables improved reconstruction of a transcriptome from RNA-seq reads. Nat
778      Biotechnol. 2015;33(3):290–5.

779  69.  Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential
780      expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.
781      2015;43(7):e47.

782  70.  Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting
783      to identify differentially methylated regions in epigenetic epidemiology studies. Int J
784      Epidemiol. 2012;41(1):200–9.

785  71.  Guo X, Gao L, Liao Q, Xiao H, Ma X, Yang X, et al. Long non-coding RNAs function
786      annotation: A global prediction method based on bi-colored networks. Nucleic Acids
787      Res. 2013;41(2).

788  72.  McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT
789      improves functional interpretation of cis-regulatory regions. Nat Biotechnol.
790      2010;28(5):495–501.

791  73.   Subhash S, Kanduri C. GeneSCF: A real-time based functional enrichment tool with
792  support for multiple organisms. BMC Bioinformatics. 2016;17(1).

793

34

## Supporting Information caption

**Fig S1: The distribution of lncRNAs and PC gene expression and DNA methylation levels across samples. (**A) The level of distribution of expression between 13460 lncRNAs and 20,135 PC genes across 82 BCP-ALL samples. (B) The level of distribution of DNA methylation rate between 60,022 CpGs probes associated with lncRNAs region and 120,000 CpGs probes associated with PC genes across 82 BCP-ALL samples.

**Fig S2: BCP-ALL subtype-specific differentially expressed lncRNAs**. (A-C) The hierarchical clustering representing lncRNAs clustering and expression differences of the compared subtypes DUX4, Ph-like and NH-HeH; corresponding to 736, 383, and 445 subtype-specific DE lncRNAs in DUX4, Ph-like and NH-HeH subtypes, respectively. In the DUX4 subtype, 100% of samples clustered together based on the DE lncRNAs signature. The hierarchical clustering of the subtype-specific DE lncRNAs revealed that 90% (19 out of 21 samples) of Ph-like samples clustered within the predefined Ph-like subtype. For the NH-HeH subtype 69% (11 out of 16 samples) of samples correlated and clustered together using the respective DE lncRNA signature. The BCP-ALL samples box representing the number of samples within each subtypes and versus (vs) the other samples used as control group in DE analysis. (D) The overlap between DE subtype specific lncRNAs from three subtypes versus public list of dysregulated lncRNAs from 12 different cancer types comprehensive cancer genome (CGC)

**Fig S3: Comparison of molecular pathways from cis and trans based analysis on subtype-specific DE lncRNAs.** (A) Molecular pathway analysis from functional enrichment analysis on *trans* ($>=$ 100 kb) protein-coding genes correlated (Pearson correlation coefficient $>=$ 0.55 and two-tailed *P*-value $<=$ 0.05) with DE lncRNAs in DUX4 subtype.

35

817   (B) The molecular pathways overlapped between *cis* (< 100 kb proximity) and *trans* (>

818   100 kb) based functional enrichment analysis in the DUX4 subtype. (C) Molecular pathway

819   analysis from functional enrichment analysis on *trans* (> 100 kb) protein-coding genes

820   correlated (Pearson correlation coefficient >= 0.55 and two-tailed *P*-value <= 0.05) with

821   DE lncRNAs in Ph-like subtype. CAMs : Cell adhesion molecules, CML: Chronic myeloid

822   leukemia , AML: Acute myeloid leukemia.

823   **Fig S4: The subtype-specific lncRNAs co-expressed with oncogenes involved in key**

824   **signaling pathways in DUX4 and Ph-like subtypes.** (A-B) Antisense *RP11-224O19.2*

825   (absolute Fold change = 2.786, P-value = 9.74E-08) and its *cis* oncogene *TGFB2* (absolute

826   Fold change = 3.84, P-value = 2.74E-10) are significantly up-regulated in DUX4 samples.

827   (C) Antisense lncRNAs *R11-536K7*.5 located at *cis* region of oncogene *IL2RA*. Expression of

828   antisense lncRNA *RP11-536K7.5* showed significant co-expression with expression of its *cis*

829   oncogene *IL2RA*. Both *RP11-536K7.5* (absolute Fold change = 2.79, *P*-value = 3.07E-008)

830   and *IL2RA* (absolute Fold change = 3.11, *P*-value = 3.97e-1) are up-regulated in Ph-like

831   samples. (D) The expression of *cis* antisense lncRNA *AC002454.1* significant co-expressed

832   with its *cis* oncogene *CDK6* in Ph-like subtype. Both *CDK6* (absolute Fold change = 1.01, *P*-

833   value = 0.0005) and antisense lncRNA *AC002454.1* (absolute Fold change = 1.79, *P*-value

834   = 0.00015) are up-regulated in Ph-like samples.

835   **Table S1. RNA-seq data used for analysis and subtype-specific lncRNAs from three**

836   **subtypes**

837   **Table S2. The functionally involved subtype-specific lncRNAs from DUX4 and Ph-like**

838   **subtypes. The trans and *cis*-acting subtype-specific lncRNAs**

839   **Table S3. The relapse-specific lncRNAs from three subtypes. The lncRNAs involved in**

840   **functions from DUX4 subtype**

36

841   **Table S4. DNA methylation array dataset. The differentially methylated lncRNAs from**

842   **three subtypes. List of cis-acting epigenetically active lncRNAs.**

## Figure captions

844   **Fig 1: BCP-ALL subtype-specific lncRNA expression signatures.**

845   (A) PCA plot constructed from expression FPKM values of lncRNAs from 82 BCP-ALL

846   samples obtained from RNA-seq. Each point represents a BCP-ALL sample. DUX4, Ph-like,

847   NH-HeH,  LH subtype and others are represented by orange, rose, blue, green and gray

848   respectively. (B) Heatmap illustrates hireachial clustering DE subtype-specific lncRNAs

849   (absolute Fold change >= +- 1.5, *P*-value <= 0.01) signature based on z-score

850   transformed LIMMA normalized expression values on 930 subtype-specific lncRNAs from

851   DUX4 (n = 450), Ph-like (n = 193), and NH-HeH (n = 287) subtypes. (C) The venn

852   diagram illustrates the overlap between subtype-specific lncRNAs from three subtypes,

853   showing 24 lncRNAs are to be common for all three subtypes.

854   **Fig 2: The molecular pathways of lncRNAs involved in the DUX4 and Ph-like BCP-ALL**

855   **subgroups.**

856   (A) The plot depicts the molecular pathway analysis from the functional enrichment

857   analysis for nearby (<= 100 kb proximity) cis protein-coding genes correlated (Pearson

858   correlation coefficient >= 0.55 and 2-tailed *P*-value <= 0.05) with DE lncRNAs in the

859   DUX4 subtype. (B) The plot depicts the molecular pathway analysis from the functional

860   enrichment analysis for nearby (<= 100 kb proximity) cis protein-coding genes correlated

861   (Pearson correlation coefficient >= 0.55 and 2-tailed *P*-value <= 0.05) with DE lncRNAs

862   in the Ph-like subtype.  (C) The heatmap depicts the concordance between the protein-

863   coding and lncRNAs based predictions for DUX4 subtypes. (D) The heatmap depicts the

864   overlapping pathways from both lncRNAs and protein-coding in the Ph-like subtype.  The

37

865    KEGG pathways or biological functions presented in the heatmaps and barplots show with

866    *P*-value <= 0.05 and > 2 genes involved in each pathways. The hypergeometric p-values

867    are obtained from GeneSCF for the pathways. CAMs : Cell adhesion molecules, CML :

868    Chronic myeloid leukemia, AML : Acute myeloid leukemia.

869    **Fig 3: Relapse-specific DE lncRNAs from BCP-ALL subtypes.**

870    (A-C) Heatmap depicting the hierarchical clustering on relapse-specific DE lncRNAs

871    signature on Z-score transformed LIMMA normalized expression values from DUX4, Ph-like

872    and NH-HeH subtypes. Each heatmap shows the up and down regulated lncRNAs specific to

873    ID and REL samples. (D) Molecular pathway analysis with the number of genes involved in

874    each pathway from the enrichment analysis of the nearby (< 100 kb proximity) cis protein-

875    coding genes correlated (Pearson correlation > 0.55 and P-value <= 0.05) with relapse-

876    specific DE lncRNAs in the DUX4 subtype. The legend box indicates the number of ID and

877    REL samples within each group. CAMs : Cell adhesion molecules. (E) The overlap between

878    relapse-specific and subtype-specific lncRNAs from three subtypes.

879    **Fig 4: Hierarchical clustering of CGID's associated with DM lncRNAs.**

880    (A) PCA of CpG's associated with lncRNAs on SWAN normalized β values on 82 BCP-ALL

881    samples obtained from DNA methylation array. Each point represents a BCP-ALL sample.

882    DUX4, Ph-like, NH-HeH, LH and others are represented by orange, rose, blue, green and

883    gray, respectively. (B) The heatmap representing hierarchal clustering on 544 differentially

884    methylated (DM) CGID's associated with 434 lncRNAs in DUX4 subtype. In the DUX4

885    subtype, we identified 328 (76%) differentially hypo-methylated and 106 (25%) hyper-

886    methylated lncRNAs. (C) The heatmap representing hierarchal clustering on 518 DM

887    CGID's associated with 450 lncRNAs in the Ph-like subtype. In Ph-like subtype, we observed

888    302 (67%) hyper-methylated lncRNAs and 148 (33%) hypo-methylated lncRNAs. (D) The

38

889     heatmap representing hierarchal clustering on 295 DM CGID's associated with 234 lncRNAs

890     in NH-HeH subtype. In the NH-HeH subtype, we identified 200 (86%) hypo-methylated

891     and 34 (14%) hyper-methylated lncRNAs. The heatmap is ploted using SWAN normalized

892     beta values. The barplots below each heatmap represents the distribution of DM lncRNAs in

893     the genome (Promoter-TSS and gene body) lncRNAs from each subtype. The distribution

894     DM Promoter-TSS lncRNAs are as follows: 25%, 29% and 39% in DUX4, Ph-like, and NH-

895     HeH subtype, respectively.

896     **Fig 5: The epigenetically altered promoter methylated lncRNAs and their expression.**

897     (A) The promoter methylated lncRNAs with significant negative correlation with DE

898     expression profile from the DUX4 subtypes. (B-C) Two representative examples of hypo-

899     methylated lncRNAs with increased expression profile from DUX4 subtype. LncRNAs, *RP11-*

900     *138M12.1* (Pearson correlation coefficient = -0.69, 2-tailed *P*-value = 5.21e-13), *RP11-*

901     *624MB.1* (Pearson correlation coefficient =  -0.50, *P*-value = 1.36e-06) are examples with

902     hypo-methylation and up-regulated expression pattern with significant inverse correlation

903     between DNA methylation and expression levels. (D) The promoter methylated lncRNAs

904     with significant negative correlation with DE expression profile from the Ph-like subtypes.

905     (E) A representative example of the promoter hyper-methylated lncRNA, *RP11-138M12.1*

906     (Pearson correlation coefficient = -0.69, 2-tailed *P*-value = 5.21e-13) with down-regulated

907     expression pattern, and with inverse correlation within the Ph-like subtype.

908     **Fig 6: Differentially methylated lncRNAs epigenetically altered expression levels of the**

909     ***cis* oncogenes.**

910      (A) The upper panel of boxplot represents the DNA methylated lncRNAs, the boxplot

911     below that represents their corresponding *cis* ocogenes which are up-regulated in DUX4

39

912    subytpe. The barplox shows representative examples of hypo-methylated lncRNAs, *RP11-*

913    *80H8.4* (DNA methylation value = -4.87 , *P*-value = 0.0001), *IGF-AS2* (DNA methylation

914    value = -1.52, *P*-value = 0.011)*, RP11-332H18.4* (DNA methylation value = 1.79 , *P*-value

915    = 0.0057 ), *RGMB-AS1* (DNA methylation value = -1.47, *P*-value = 0.007), and *RP11-*

916    *125B21.2* (DNA methylation value = -1.75, *P*-value = 0.007) and its corresponding

917    significantly up-regulated *cis* oncogenes, *CHT2* (absolute log fold change = 5.021, FDR =

918    2.39E-08), IGF2 (absolute log fold change = 6.71, FDR = 41.412E-15), *TBX* (absolute log

919    fold change =5.64, FDR = 3.97E-13), and RGMB (absolute log fold change = 4.42, FDR =

920    3.02E-16) within DUX4 subytpe. (B**)** The upper panel of boxplot represents the DNA

921    methylated lncRNAs, the boxplot below that  represents their corresponding *cis* ocogenes

922    which are up-regulated in Ph-like subytpe. The barplox shows representative examples of

923    hypo-methylated lncRNAs, *RP11-69l8.3* (DNA methylation value= -3.16, P-value = 4.46E-

924    08), *LINC00996* (DNA methylation value = -1.55, *P*-value = 0.02 ), *CTB-79E8.2* ( DNA

925    methylation value = -2.22, *P*-value = 0.009 ), *RP11-420G6.4* (DNA methylation value =

926    -2.43, *P*-value = 0.004 ) and its corresponding significatly up-regulated *cis* oncogenes,

927    *CTGF* ( absolute log fold change = 1.85, = 0.02 ), *GIMAP8* (absolute log fold change =

928    1.14, FDR = 0.004), *NEURLB* (absolute log fold change = 1.14, FDR = 0.09), *SERPINB1*

929    (absolute log fold change = 1.63, FDR = 0.0004) within Ph-like subtype**.** (C) The upper

930    panel of boxplot represents the DNA methylated lncRNAs, the boxplot below that

931    represents their *cis* ocogenes which are up-regulated in NH-HeH subytpe. The barplot

932    shows representative examples of hypo-methylated lncRNAs, *LINC00114* (DNA methylation

933    value = -2.97, P-value = 0.0003), *RP11-293A21.1* (DNA methylation value = -3.69, P-

934    value = 0.00071) and its corresponding significantly up-regulated *cis* oncogenes, *ETS2*

935    (absolute Fold change = 0.99, FDR = 0.019) and *STIM2* (absolute Fold change = 2.48,

936   FDR = 6.42E-10) within NH-HeH subtype. False discovery rate: FDR

## 937   AVAILABILITY OF DATA AND ACCESSION NUMBERS

938   All sequencing data used in this study is available at the European Genome

939   phenome Archive (accession number to be provided after acceptance of the manuscript)

## 940   FUNDING

41

A  DUX4 (n=23)

146

426

B  Ph-like (n=21)

101

14

C  NH-HeH (n=16)

134

26

Number of ID and REL samples in each
BCP-ALL subtype

| Ph-like | DUX4 | NH-HeH |
|---|---|---|
| ID (n=10) | ID (n=12) | ID (n=7) |
| REL (n=11) | REL (n=11) | REL (n=9) |

Z-score
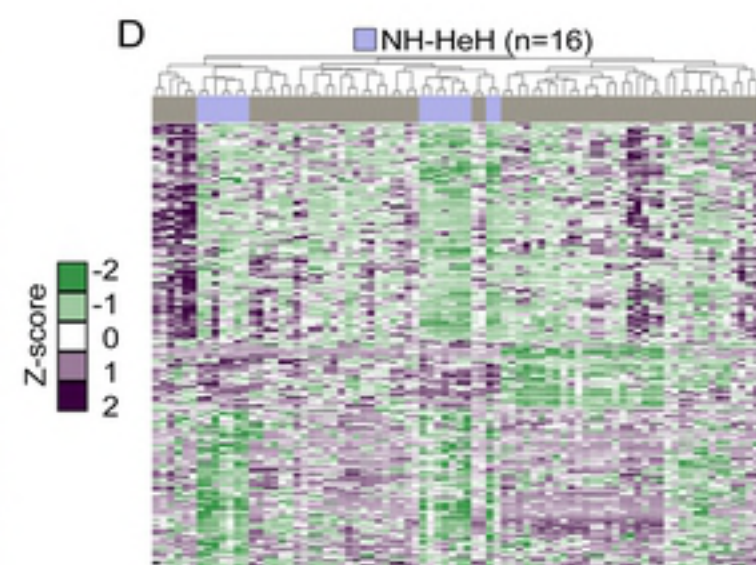
2  1  0  -1  -2

D

Pathways enriched based on relapse specific DUX4 lncRNAs

Metabolic pathways — 29
MAPK signaling pathway — 8
RNA transport — 6
Hematopoietic cell lineage — 4
Drug metabolism-other enzymes — 3
CAMs — 5
mTOR signaling pathway — 3
Hippo signaling pathway — 5
Pyrimidine metabolism — 4
T-cell receptor — 4

0    1.0   2.0   3.0   4.0
-log10 (p-value)

E
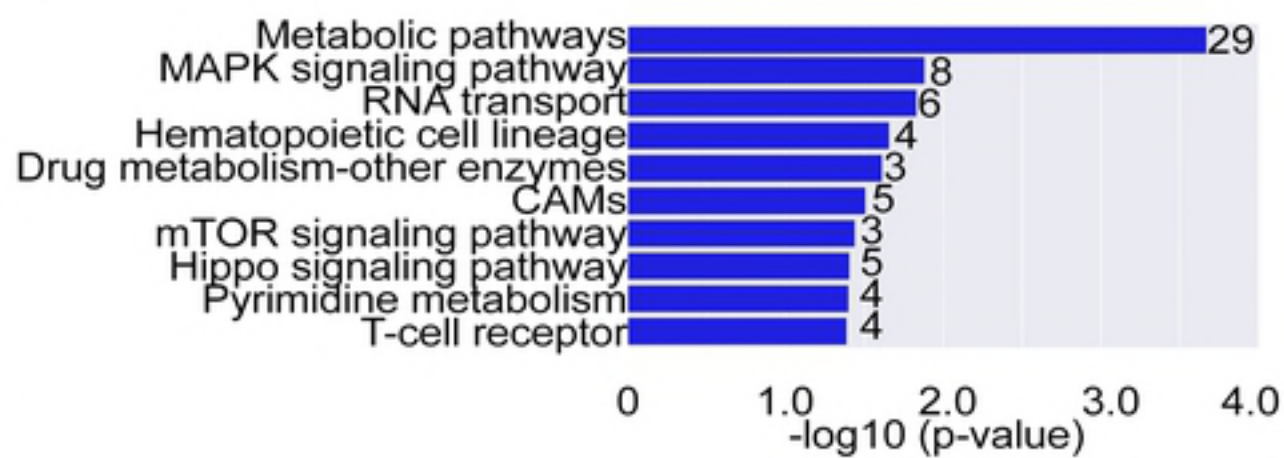
Relapse-specific lncRNAs (867)    Subtype-specific lncRNAs (1235)

684      183      1052

**A**

DUX4 (n=23)
Ph-like (n=21)
NH-HeH (n=16)
Others (n=16)
LH (n=6)

**B** DUX4 (n=23)

**C** Ph-like (n=21)

**D** NH-HeH (n=16)

Z-score
-2
-1
0
1
2

Hyper-methylation

Hypo-methylation

Distribution of DM lncRNAs in different genomic regions
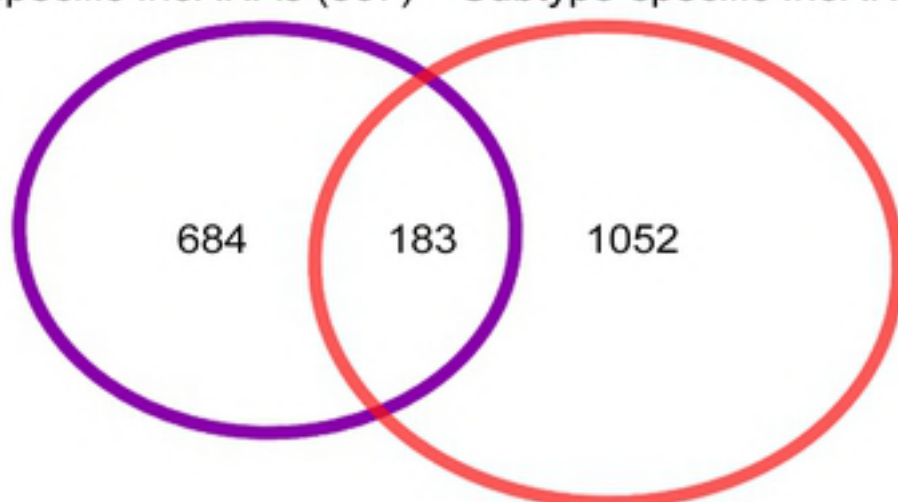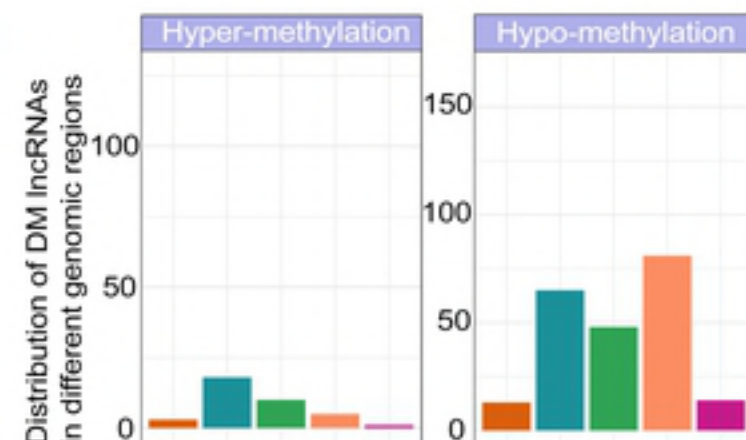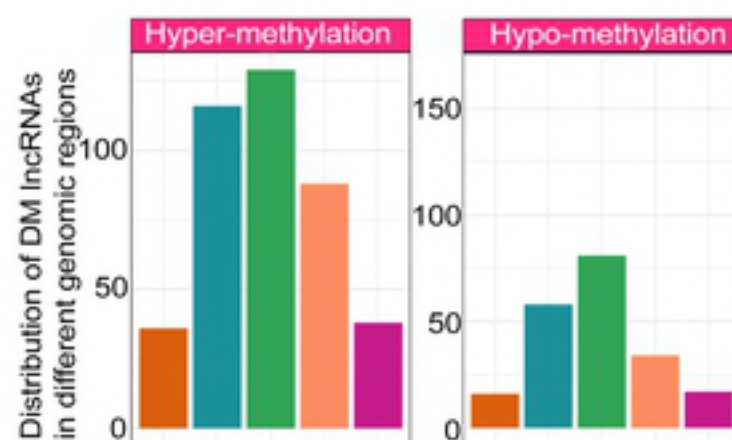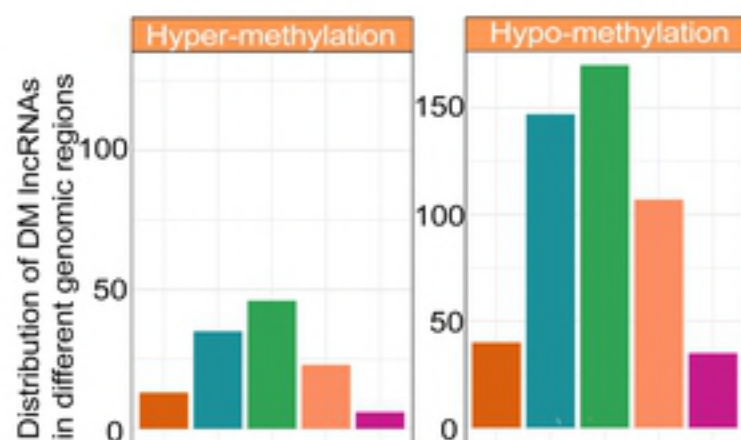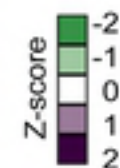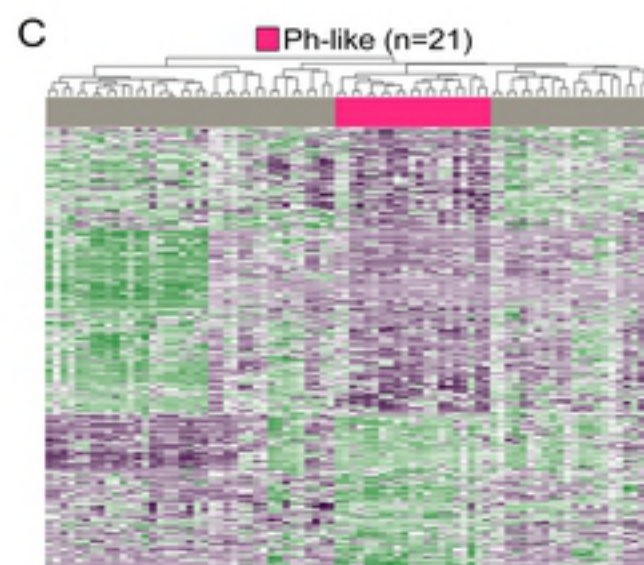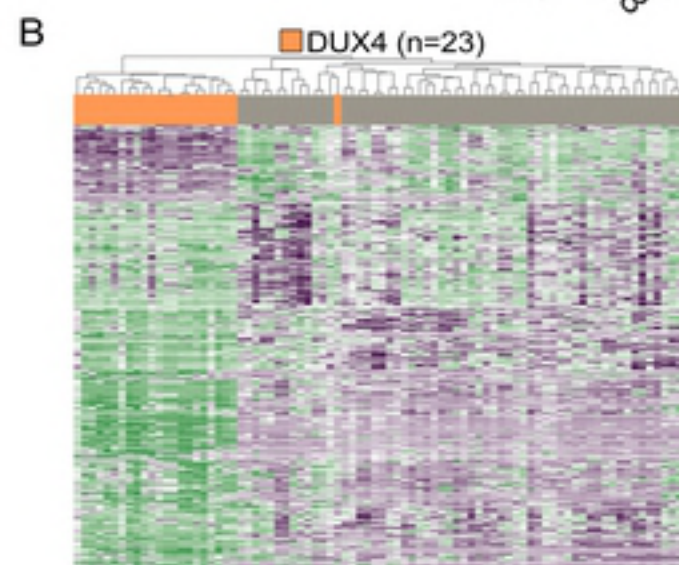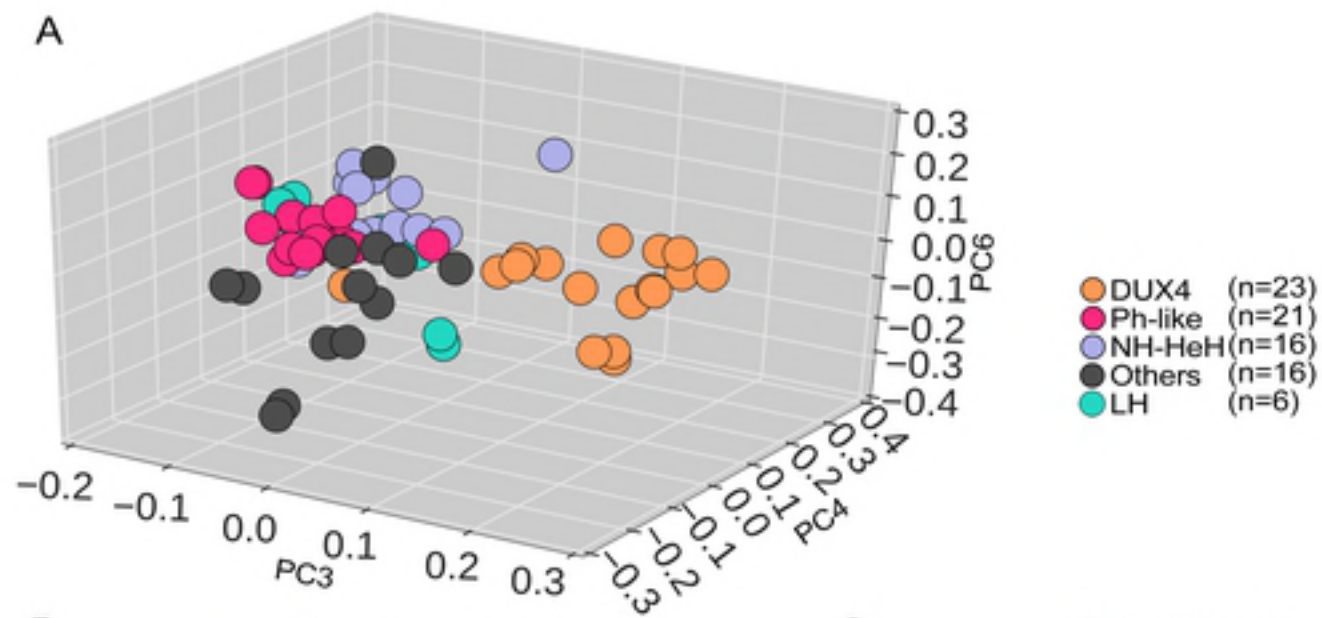
Exon   Intergeneic   Intron   Promoter-TSS   TTS

A — DUX4 (n=15)
B — RP11-624M8.1
C — RP11-138M12.1
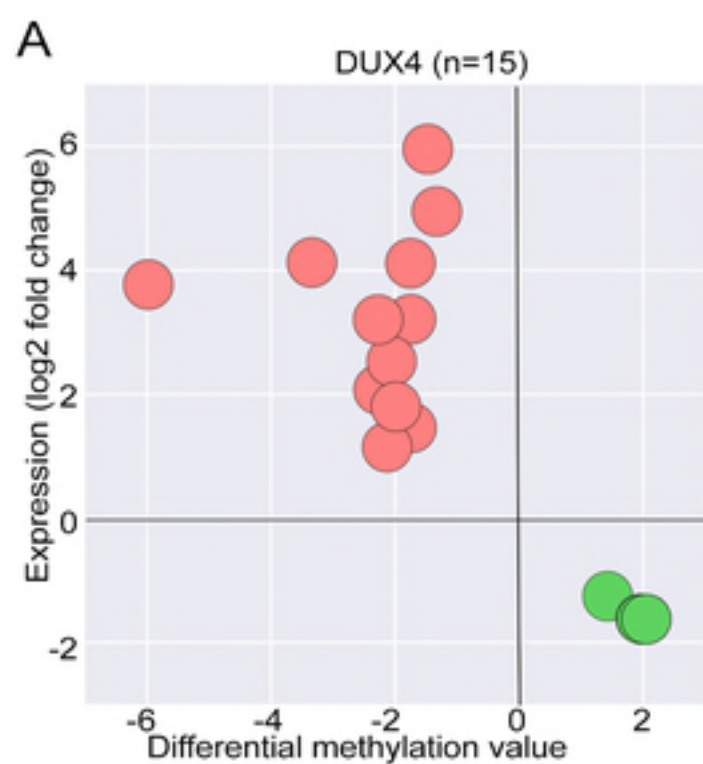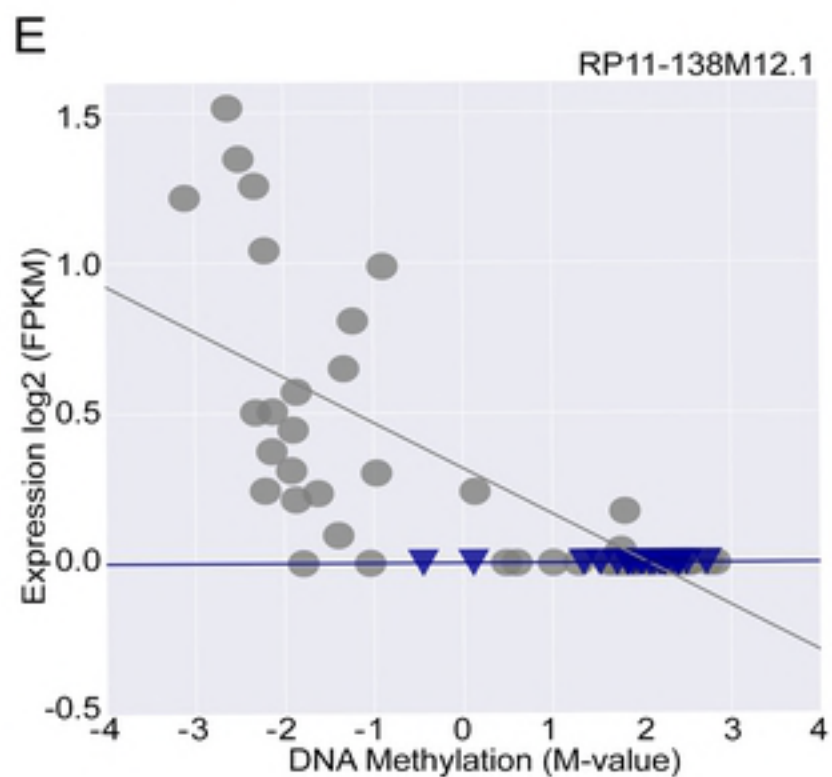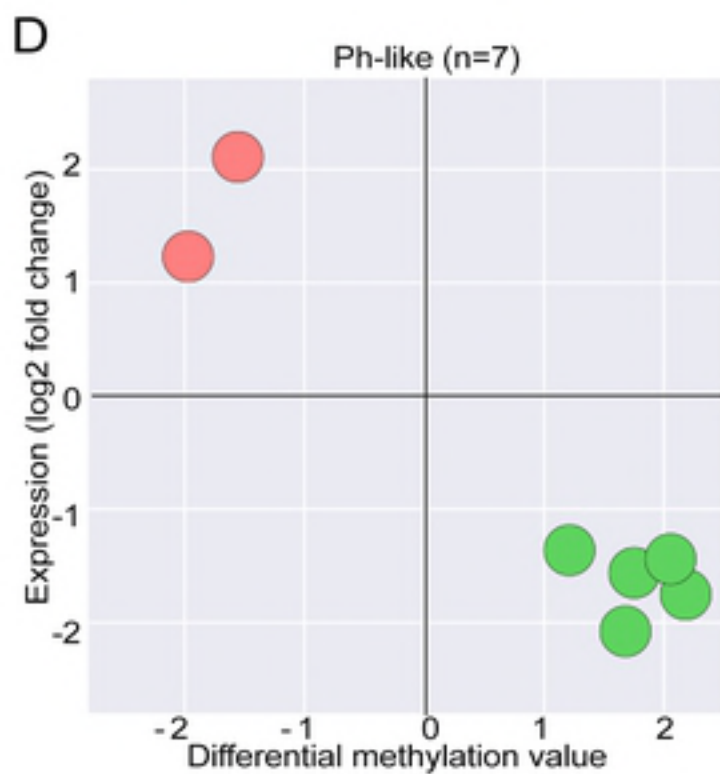
Hypermethylation & downregulation in subtypes
Hypomethylation & upregulation in subtypes

▼ DUX4 (n=23)    ● Others (n=59)

D — Ph-like (n=7)
E — RP11-138M12.1

▼ Ph-like (n=21)    ● Others (n=61)

**A** — DUX4 specific hypo-methylated lncRNAs

RP11-80H8.4, IGF2-AS, RP11-332H18.4, RGMB-AS1, RP11-125B21.2

Up-regulated protein coding genes located within the cis region of DUX4 specific methylated lncRNAs

CHT2, IGF2, TBX2, RGMB, VLDLR

**B** — Ph-like specific hypo-methylated lncRNAs

RP11-69I8.3, LINC00996, RP11-420G6.4

Up-regulated protein coding genes located within the cis region of Ph-like specific methylated lncRNAs

CTGF, GIMAP8, SERPINB9

**C** — NH-HeH specific hypo-methylated lncRNAs

LINC00114, RP11-293A21.1

Up-regulated protein coding genes located within the cis region of NH-HeH specific methylated lncRNAs

ETS2, STIM2