RNA-seq-based miRNA signature as an independent predictor of relapse in pediatric B-cell acute lymphoblastic leukemia

Hirohito Kubota,[1] Hiroo Ueno,[1] Keiji Tasaka,[1] Tomoya Isobe,[2,3] Satoshi Saida,[1] Itaru Kato,[1] Katsutsugu Umeda,[1] Mitsuteru Hiwatari,[2,4] Daiichiro Hasegawa,[5] Toshihiko Imamura,[6] Nobuyuki Kakiuchi,[7,8] Yasuhito Nannya,[7,9] Seishi Ogawa,[7,10,11] Hidefumi Hiramatsu,[1,*] Junko Takita[1,*]

[1]Department of Pediatrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [2]Department of Pediatrics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Department of Hematology, Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [4]Department of Pediatrics, School of Medicine, Teikyo University, Tokyo, Japan. [5]Department of Hematology and Oncology, Hyogo Prefectural Kobe Children Hospital, Hyogo, Japan. [6]Department of Pediatrics, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan. [7]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [8]The Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan. [9]Division of Hematopoietic Disease Control, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [10]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto, Japan. [11]Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden. *Corresponding authors.

Running title: A novel prognostic miRNA signature in BCP-ALL

Key words: miRNA signature, integrative analysis, pediatric BCP-ALL

Word counts: main text 3948 words (abstract 230 words)

Figure count: 6 (Supplementary Figures: 10)

Supplementary Table count: 16

Reference count: 58

Key Points

- We identified a novel miRNA signature in BCP-ALL patients at high-risk for relapse, which is independent of known genetic subtypes.
- This miRNA signature is often newly acquired at relapse and is associated with the upregulation of MYC target genes.

Author notes

The miRNA-seq, mRNA-seq, and WES data in the Japanese cohort that support the findings of this study were deposited in the Japanese Genotype-phenotype Archive (JGA) under accession code JSUB000865.

Publicly available RNA-seq data for normal B cell progenitors were downloaded from the NCBI GEO database under accession number GSE115656.

For other data, please contact the corresponding authors.

The full-text version of this article contains a data supplement.

Abstract

Aberrant miRNA expression profiles have been associated with disease progression and clinical outcome in pediatric cancers. However, few studies have analyzed genome-wide dysregulation of miRNAs and mRNAs in pediatric B cell precursor acute lymphoblastic leukemia (BCP-ALL). To identify novel prognostic factors, we comprehensively investigated miRNA and mRNA sequencing (miRNA-seq and mRNA-seq) data in poor-outcome pediatric BCP-ALL samples. We analyzed 180 patients, including 43 matched pairs at diagnosis and relapse. Consensus clustering of miRNA expression data revealed a distinct profile characterized by mainly downregulation of miRNAs (referred to as an miR-low cluster; MLC). The MLC profile was not associated with any known genetic subgroups. Intriguingly, patients classified as MLC had significantly shorter event-free survival (median 21 vs 33 months, log-rank $P = 3 \times 10^{-5}$). Furthermore, this poor prognosis was retained even in hyperdiploid ALL. This poor prognostic MLC profiling was confirmed in the validation cohort. Notably, non-MLC profiling at diagnosis often (n= 9/23, Fisher's exact test, $P = 0.039$) changed into MLC profiling at relapse in the same patient. Integrated analysis of miRNA-seq and mRNA-seq data revealed that the transcriptional profile of MLC was characterized by enrichment of MYC target and oxidative phosphorylation genes, reduced intron retention, and low expression of *DICER1*. Thus, our miRNA-mRNA integration approach yielded a truly unbiased molecular stratification of pediatric BCP-ALL cases based on a novel prognostic miRNA signature, which may lead to better clinical outcomes.

Introduction

Recent genomic and transcriptomic analyses of pediatric B-cell precursor acute lymphoblastic leukemia (BCP-ALL) have revealed novel genetic subgroups.[1] These novel genetic subgroups and their driver mutations affect prognosis; therefore, treatment stratification by risk classification has been proposed to improve treatment outcomes.[2,3] However, nearly 20% of pediatric BCP-ALL patients relapse and the prognosis remains poor; indeed, less than 50% of relapsed patients survive long-term,[4] whereas, about half of relapsed patients have a favorable prognosis; these include patients with hyperdiploidy and *ETV6::RUNX1*.[5] Although minimal residual disease (MRD), a marker of response to therapy, is the strongest prognostic factor,[6] many patients who achieve undetectable MRD experience relapse; therefore, current prognostic factors are not thought to be sufficiently stratified.

miRNAs regulate post-transcriptional gene expression by functioning as mRNA translation repressors or degraders.[7] Aberrant expression of miRNAs by tumor cells is a novel biomarker or therapeutic target for various malignancies.[8] Recent studies have revealed characteristic expression patterns according to genetic subgroups of pediatric BCP-ALL, as well as their prognostic significance.[9,10] In addition, novel prognostic factors identified by integrated analysis, including miRNAs and mRNAs, have been reported in several types of malignant tumors[11]; however, few studies have conducted an integrated analysis focusing on miRNA profiles in a large cohort of pediatric BCP-ALL.

Here, we aimed to clarify the expression profile of miRNAs in pediatric BCP-ALL by performing an integrative analysis of miRNA sequencing (miRNA-seq) data. We then used this information to investigate their clinical significance and identify

prognostic factors, on which we can base a more refined risk stratification approach.

Methods

Patients

The TARGET ALL-P2 cohorts have been published previously, and data were used with permission from the database of Genotypes and Phenotypes (dbGaP Study Accession: phs000218).[1] The study included 111 primary samples of BCP-ALL (of which 35 patients had paired sample data at relapse) from the TARGET cohort, a high-risk cohort with clinical information for which both miRNA-seq and mRNA-sequencing (mRNA-seq) data were available (Supplementary Tables 1 and 2). Of these 111 cases, 105 had SNP array data (Affymetrix SNP Array 6.0), and 53 had methylation data (HELP Assay with Roche NimbleGen). The samples sequenced in this study were derived from 69 primary samples (including eight paired samples at relapse) provided by multiple institutions in Japan (Supplementary Tables 1 and 3). All samples were obtained with written consent from the patient or parent/guardian under the protocols approved by the institutional review board of each institution. All studies were conducted in accordance with the international ethical guidelines for biomedical research involving human subjects. The chemotherapy regimens used in the TARGET cohort were AALL0232[12] and AALL0331[13], and in the Japanese cohort were based on JACLS[14] or TCCSG[15] protocols.

miRNA-seq

miRNA-seq was performed in the Japanese cohort (validation cohort). Total RNA was extracted using the miRNeasy Kit (Qiagen). Libraries for miRNA-seq were generated

using the NEXTflex Small RNA Sequencing kit v3 (PerkinElmer) and sequenced using the Illumina NextSeq 550 using a 75 bp single-end read protocol. miRNA-seq analysis was performed using the nf-core/smrnaseq pipeline v1.0.0.[16] Briefly, all steps consisted of the removal of 3' adapter sequences by TrimGalore v0.6.5, mapping to mature and hairpin miRNAs using miRbase v22.1, and mapping to the GRCh37 human reference genome using Bowtie v1.3.0. After alignment and trimming, sorted BAM files were used for further analysis using edgeR v3.32 for reads per million (RPM) normalization, and mirtop v0.4.23 for raw counts. Normalized count data obtained from the regularized logarithm function of the R package DESeq2 v1.26.0 were used for clustering analysis. Cluster stability was determined by consensus clustering with 1000 iterations using the R package ConsensusClusterPlus v1.50.0. The function removeBatchEffect from the R package limma package v3.42.2 was used to remove batch effects for merged miRNA expression during unsupervised clustering analysis. Differential expression analysis was performed using the DESeq2. After the rlog gene expression level was calculated by DESeq2, Rtsne in the R package was used to map the samples to a two-dimensional tSNE plot of the top 500 most variable miRNAs (on the basis of median absolute deviation), and the tSNE perplexity parameter was set to 30.

mRNA-seq

RNA was isolated from fresh frozen samples using the NucleoSpin Triprep (Macherey-Nagel). Sequencing libraries were prepared using a the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England BioLabs). The library was sequenced using an Illumina HiSeq X platform using a 150 bp paired-end read protocol. mRNA-seq was performed in the Japanese cohort. Sequence alignment and read counting were

conducted using the Genomon v2.6.2 pipeline (http://genomon.readthedocs.io/ja/latest/index.html). Read counting was also normalized to obtain fragments per kilobase million (FPKM) values. Fusion transcripts were detected by the CICERO software (Supplementary Table 4).[17] To identify *BCR::ABL1*-like and *ETV6::RUNX1* ALL, 198 genes were identified using the ROSE method,[18] including the top 25 genes in Clusters 1–8 in the TARGET cohort (Supplementary Figure 1). Ward's hierarchical clustering method was then used for clustering analysis of the ROSE gene set. Samples in Cluster 5 had a signature similar to that of *BCR::ABL1* ALL and were therefore labeled *BCR::ABL1*-like, whereas samples in Cluster 8 had a signature similar to *ETV6::RUNX1* ALL and were labeled *ETV6::RUNX1*-like.

We used mRNA-seq data to examine driver mutations in variant calls following the GATK Best Practice for calling variants (https://gatk.broadinstitute.org/hc/). Variants were called by the GATK HaplotypeCaller v.4.1 according to the GATK best practices for variant calling in mRNA-seq, as implemented in the CRG-CNAG/CalliNGS-NF pipeline (https://github.com/CRG-CNAG/CalliNGS-NF). Overall, 110 known or putative driver genes, and target genes of activation-induced cytidine deaminase in BCP-ALL were filtered (Supplementary Table 5).[3] The filtered variants were then annotated against the COSMIC database v92 (a cancer mutation catalogue) from the GRCh37 assembly to retain those variants previously identified as cancer mutations.[19]

Additional details for methods sections are provided in the Supplementary data.

Results

Consensus clustering of miRNAs reveals a novel high-risk subgroup with a distinct miRNA profile

To characterize the miRNA profile of poor-outcome BCP-ALL patients, we focused on

the TARGET ALL-P2 cohort, which mostly consisted of cases with early bone marrow relapse. We included 111 cases for which miRNA-seq, mRNA-seq data, and clinical information were all available and used them as the discovery cohort (Supplementary Tables 1 and 2). First, we performed unsupervised consensus clustering of miRNA expression data obtained from primary samples. The analysis revealed a strong cluster (n = 31) that was not associated with well-known genetic subtypes (Cluster 2 in Figure 1A and B; Supplementary Figure 2A and B). This cluster included diverse genetic subtypes such as hyperdiploidy (n = 7) and *ETV6::RUNX1* (n = 4) fusions. Differential expression analysis of miRNAs with high average expression across the whole population comprising this subgroup demonstrated that they were mainly down-regulated (Figure 1C, D and Supplementary Table 6). Hereafter, we refer to this subgroup with a distinct miRNA expression profile as the miR-low cluster (MLC). Notable differentially expressed miRNAs in the MLC included tumor suppressive (e.g., miR-26b,[20] miR-101,[21] miR-29, miR-142,[22] miR-146b,[23] and miR-326[24]) and oncogenic (e.g., miR-130b[25] and miR-92a[26]) miRNAs linked to leukemia. Consecutive two-step unsupervised consensus clustering of the MLC and non-MLC group identified four clusters respectively, which generally clustered together with well-known subtypes (Supplementary Figure 2C), further highlighting the distinct features of the aberrant miRNA expression profile identified in the MLC.

Next, we assessed the impact of the miRNA signature of the MLC on the clinical outcome of pediatric BCP-ALL patients. In the TARGET ALL cohort, 83 patients experienced relapse. We found that the MLC group had significantly shorter event-free survival (EFS) than the non-MLC group (median, 21 months *vs.* 33 months, respectively; 95% confidence interval (CI) = 11.1–26.8; log-rank $P = 3 \times 10^{-5}$) (Figure

1E). Although not significant, the MLC group also showed a trend toward shorter overall survival (OS) (median, 29 months *vs.* 64 months, respectively; 95% CI = 21.2–84.9; log-rank P =0.09). In addition, when the analysis was restricted to hyperdiploid patients (n=23), the MLC group (7 of 23 patients) had significantly shorter EFS (median, 28 months; 95% CI = 3.3–39.1; log-rank $P = 5 \times 10^{-4}$). Although the difference in OS was not statistically significant, there was a trend toward shorter OS in the MLC group of hyperdiploid patients (log-rank $P = 0.15$). Multivariable Cox regression analysis of MLC status and established prognostic factors, including end-induction MRD assessed by multi-color flow cytometry,[27,28] showed that MLC status was an independent predictor of relapse (hazard ratio, 2.59; 95% CI = 1.56–4.3; $P < 0.001$) (Figure 1F). This was also true in the analysis in which each genetic subtype was included as a covariate (Supplementary Figure 3). Taken together, these results suggest that the miRNA signature of MLC could be a novel prognostic factor that predicts relapse better than current prognostic factors such as MRD.

Validation of the poor prognostic MLC profile in the validation cohort

To validate the poor prognostic miRNA profile of MLC, we evaluated the miRNA profile in the Japanese BCP-ALL cohort. We collected primary samples of 69 patients, 34 of whom relapsed (validation cohort; Supplementary Tables 1 and 3). Two-step consecutive unsupervised consensus clustering (Supplementary Figure 4A and B) identified a cluster (n = 19) that was independent of known genetic subtypes (Cluster 2 in Supplementary Figure 4C), as observed for the MLC in the TARGET cohort. To identify patients similar to the MLC in this small Japanese cohort, we applied a statistical regression algorithm based on the least absolute shrinkage selection operator

(LASSO).[19,29] This analysis identified a 19-miRNA signature characterizing the MLC (Figure 2A). Next, the weighted sum of the expression levels of these 19 miRNAs was calculated for each patient in the validation cohort and was expressed as the MLC score. The group identified by the two-step consecutive unsupervised clustering method described above (Supplementary Figure 4C) had MLC scores almost exclusively in the highest tertile within the validation cohort; therefore, we refered to this group as MLC-like patients (Figure 2B). Compared with the non-MLC-like group, MLC-like patients were more likely to relapse (Figure 2C) and had a significantly shorter EFS (median, 25 months; 95% CI = 16.1–33.0; log-rank $P = 1 \times 10^{-6}$) and a significantly shorter OS (median, 42.6 months; 95% CI = 22.3–78.4; log-rank $P = 8 \times 10^{-4}$) (Figure 2D). Furthermore, evaluation of the area under the curve (AUC) of the time-dependent receiver operating characteristic (ROC) curve, which measures the prognostic ability of the MLC score and miRNAs to predict EFS and OS at 2 years, showed that the MLC score was one of the best predictors of poor prognosis compared with each single miRNA in both the TARGET and Japanese cohorts (Figure 2E and Supplementary Tables 7 and 8). In particular, among the elements of the MLC score, high expression of miR-92b-3p and miR-1304-3p were good predictors of poor prognosis (Supplementary Figure 5A and B). Taken together, these data confirm the utility of the MLC profile for predicting poor prognosis in the validation cohort.

MLC status is not associated with previously reported genetic alterations

To investigate the relationship between MLC status and previously reported genetic alterations in BCP-ALL, we examined driver mutations in short variants and copy number alterations from the discovery cohort (Supplementary Tables 9 and 10). In

general, the distribution of these genetic alterations did not differ significantly according to MLC status, although alterations in *IKZF1* or *TP53,* and *ZNF384*-rearranged were significantly mutually exclusive with MLC status (Figure 3A, B and Supplementary Figure 6A). Variant call and copy number analysis were also performed using whole exome sequencing (WES) from the validation cohort (Supplementary Figure 6B, C and Supplementary Tables 11 and 12), but we found no abnormalities characterizing MLC-like cases. Analysis of focal copy number (Supplementary Figure 7) and DNA methylation data (Supplementary Figure 8) from the TARGET ALL cohort did not identify characteristic abnormalities for MLC. In addition, the poor risk chromosomal profile for hyperdiploidy[30] was observed in only one case in the MLC (Supplementary Table 2). Thus, the MLC miRNA profile does not seem to be associated with any of the well-known high-risk genetic subgroups, driver sequence mutations, copy number alterations such as *IKZF1* or *IKZF1*^{plus}, or aberrant DNA methylation, making it a strong candidate for novel risk stratification.

Non-MLC samples frequently acquired MLC-like miRNA signature at relapse

Next, we analyzed paired samples (obtained at diagnosis and relapse) to determine whether the miRNA profile changes at the point of relapse. Unsupervised clustering of 43 paired primary and relapse samples identified two clusters (Figure 4A). Cluster 1 contained all 20 primary samples classified as MLC at diagnosis; this was considered to be the group with the MLC profile. Eighteen of these were also categorized into the same cluster at relapse (Figure 4B and Supplementary Table 13). By contrast, nine of 23 patients classified as having a non-MLC profile at diagnosis showed a change in their miRNA profile: from Cluster 2 at diagnosis to Cluster 1 at relapse (Fisher's exact test, *P*

= 0.039). This suggests that non-MLC patients can acquire an MLC profile at relapse, which is consistent with the fact that MLC scores were significantly higher at relapse than at diagnosis in these nine patients (Wilcoxon signed-rank test, $P = 0.021$) (Figure 4C). Furthermore, evaluation of the MLC scores for the paired primary and relapse samples from the 23 non-MLC cases showed that the MLC score increased significantly at relapse (Wilcoxon signed-rank test, $P = 0.014$) (Figure 4D). These results imply that an MLC profile may confer resistance to treatment and contribute to disease progression and relapse.

The MLC is characterized by enrichment of MYC target and oxidative phosphorylation genes, and by reduced intron retention (IR)

To identify genes targeted by deregulated miRNAs in the MLC, we performed an integrated analysis of miRNAs and mRNAs in the TARGET cohort (Figure 5A and Supplementary Table 14).[11] Gene set enrichment analysis (GSEA) of the upregulated genes in the 2331 pairs showing negatively correlated between miRNAs and mRNAs revealed significant enrichment of MYC target and oxidative phosphorylation genes (Figure 5B), as well as genes related to RNA splicing (Figure 5C). Furthermore, GSEA of genes upregulated in the paired primary and relapse samples from the nine patients who acquired an MLC profile at relapse also showed significant enrichment of MYC target genes and genes related to oxidative phosphorylation (Figure 5D and E). Collectively, these results suggest that these pathways contribute to the aggressive phenotype of the MLC.

Given the enrichment of RNA splicing terms within the genes upregulated in the MLC, we next focused on splicing alterations in the MLC. We identified almost all

of the alternative-splicing events in MLC cases were reductions in intron retention (IR) (Figure 5F and Supplementary Table 15), which was also observed when compared with normal pro-B cells, and in nine relapsed cases that acquired the MLC-like profile at relapse (Supplementary Figure 9A and B). Genes showing decreased IR and increased expression due to a reduction in miRNA expression may be crucial for the MLC; therefore, we looked for genes that met all of the following criteria: (1) genes for which increased expression in MLC compared with non-MLC samples (Supplementary Table 16); (2) genes for which an miRNA reduction in the MLC may be associated with an increase in mRNA expression (compared with non-MLCs); and (3) genes with decreased IR (compared with non-MLCs and normal pro-B cells). This resulted in identification of the *MTA1* gene, an RNA-binding protein gene that post-transcriptionally activates MYC[31] (Figure 5G). Notably, miR-30e-5p, a previously identified MTA1 suppressor,[32] showed a significant inverse correlation with *MTA1* expression in the TARGET cohort (Figure 5H) and may play a role in increasing MYC activity in the MLC. Next, we performed gene set variation analysis.[33] Based on the B lineage cluster-specific gene set,[34] the gene expression pattern of MLC leukemic blasts resembled that of pre-B cells; this was not the case for non-MLC patterns (Figure 5I). A previous study shows that MYC-driven transcriptional programs contribute to pre-B cell transformation, and that the pre-B cell receptor signaling expression pattern could serve as a prognostic marker for high-risk BCP-ALL.[34,35] Similarly for MLC, it is suggested that increased pre-B cell signaling upon activation of the MYC pathway (caused by miRNA aberrations) may be associated with a poor prognosis.

Downregulation of *DICER1* may dysregulate miRNA expression in the MLC

Finally, to identify the mechanism underlying the downregulation in key miRNA expression in the MLC, we explored abnormalities in the miRNA biogenesis pathway (Supplementary Figure 10A). We found that expression of *DICER1*, also a crucial factor for B cell development,[36] differed significantly between MLC, non-MLC, and normal B progenitor cells (Kruskal-Wallis test, $P = 5.0 \times 10^{-6}$); *DICER1* was downregulated significantly in the MLC compared with the non-MLC samples (Wilcoxon rank-sum test, Bonferroni-corrected, $P = 2.0 \times 10^{-5}$) (Figure 6A), which was also the case in the MLC-like patients of the Japanese cohort (Wilcoxon rank-sum test, $P = 0.0153$) (Figure 6B). In addition, *DICER1* was downregulated significantly at relapse compared with the paired primary sample in all nine patients who acquired an MLC profile at relapse (Wilcoxon signed-rank test, $P = 7.81 \times 10^{-3}$) (Figure 6C). Furthermore, *DICER1* expression was significantly lower in patients at relapse than at diagnosis in the other B-ALL cohort[37] (Supplementary Figure 10B) and patients with low expression of *DICER1* were associated with poor prognosis, although there was no significant difference between the intermediate and high expression groups in the TARGET cohort, probably due to the high proportion of relapsed cases (Supplementary Figure 10C). Indeed, sequencing of mature miRNAs revealed a marked reduction in the 5p- to 3p- mature miRNA ratio in MLC compared with non-MLC samples (Wilcoxon signed-rank test, $P = 6.59 \times 10^{-15}$) (Figure 6D), which is consistent with the finding that *DICER1* loss-of-function mutations lead to depletion of 5p miRNAs in other cancers.[38] *In vitro* experiments showed that while *DICER1* knockdown of REH cell lines had no effect on cell growth, it did confer resistance to cytarabine, for which oxidative phosphorylation has been reported to be involved in resistance,[39] suggesting that *DICER1* is crucial for chemoresistance of BCP-ALL (Figure 6E and F). Taken together, these observations suggest that the

downregulation of *DICER1* is a potential mechanism underlying the miRNA dysregulation that characterizes the MLC. Although reduced expression of *DICER1* due to DNA methylation or copy number alterations promotes tumorigenesis in other tumors,[40] these abnormalities were not observed in MLC (Supplementary Figure 10D).

Discussion

This study describes a novel RNA-sequencing based miRNA signature that identifies pediatric BCP-ALL patients at high risk of relapse and could lead to optimized treatment stratification as a novel risk factor. Unlike mRNA expression profiles that are dependent on existing subtypes,[41] the miRNA profile of the MLC was independent of existing subtypes, driver mutations, and methylation changes, meaning that it cannot be identified in the absence of miRNA-seq data. Several studies report prognostic prediction based on miRNA signatures in pediatric ALL.[42,43] However, these studies were limited to analysis of specific miRNAs by quantitative PCR or microarrays, and did not analyze their association with comprehensive genetic alterations in patients with BCP-ALL. Previous studies of miRNA expression in cases of pediatric ALL are inconsistent with respect to methodology and study design, which reduces reproducibility and prevents adequate meta-analysis; therefore, no definitive prognostic miRNAs have been identified.[10] Among miRNA profiling methods, miRNA-seq is the most accurate, sensitive, and specific, and it can quantify expression over a wide dynamic range.[44] Thus, comprehensive integrative analysis using miRNA-seq allowed us to identify an aberrant miRNA signature in the MLC.

Predicting relapse is particularly important for standard-risk BCP-ALL. Recently, several novel prognostic factors for ALL have been reported. However, we

found that *IKZF1*^{plus}[45] and a focal 22q11.22 deletion combined with *IKZF1* alterations,[46] both of which are poor prognostic factors, were not associated with the MLC. Other recently identified poor prognostic factors, including the poor risk chromosomal profile for hyperdiploidy[30] and the prognostic epigenetic regulator mutations[1] (i.e., *TBL1XR1* and *SETD2*), were not associated with the MLC. The miRNA profile observed in the MLC could be a predictor of relapse even in hyperdiploid ALL; indeed, it may be a better predictor of relapse than MRD. More precise identification of patients at high risk for relapse at the time of diagnosis based on miRNA expression would improve the prognosis of pediatric BCP-ALL through risk-stratified treatment, including induction of immunotherapy.[47] As evaluation of the miRNA transcriptome is still difficult in clinical practice, it was suggested that high expression of miRNA-92b-3p and miRNA-1304-3p could alternatively predict MLC profile and prognosis, suggesting the potential utility of their individual assessment, but further validation is needed.

Our integrative miRNA-mRNA analysis showed that a distinctive reduction in miRNA expression in the MLC may lead to increased expression of MYC target genes and genes related to oxidative phosphorylation, which are associated with a poor prognosis and cytarabine resistance in BCP-ALL.[48] Given the challenges associated with direct targeting of MYC, exploring the therapeutic potential of drugs targeting the oxidative phosphorylation pathway, such as venetoclax[39], warrants further investigation. Elucidating the functional basis for the association between the MLC miRNA profile and adverse outcomes and the founding genetic lesion underlying *DICER1* downregulation is beyond the scope of this study, but further research is warranted.

The MLC was characterized by reduced IR, accompanied by increased expression of RNA processing genes. Aberrant AS is relevant to BCP-ALL.[49] In addition,

a recent study of BCP-ALL found that recurrent alterations in the RNA machinery genes,[1] as well as RNA splicing alterations,[50] have prognostic implications. IR is a type of AS event that occurs when an intron is retained in the final mature mRNA transcript. Many malignancies are characterized by increased IR; however, we found (unexpectedly) that the MLC had lower levels of IR than the non-MLC and normal B cell precursors. Decreased IR has been reported in *SF3B1*-mutated myelodysplastic syndrome[51] and breast cancer[52] and an inverse relationship between IR and cell proliferation has been observed during B cell development.[53] Reduced IR in the MLC is presumably due to altered expression of spliceosome components through dysregulated miRNAs or to increased MYC activity.[54] Since mRNA transcripts containing introns are degraded by nonsense-mediated decay of host genes, reduced IR is associated with the upregulation of gene products. In the MLC, reduced IR of the *MTA1* gene, an activator of MYC, has been implicated in disease pathogenesis. Overexpression of *MTA1* results in spontaneous development of B cell lymphoma, and PAX5 has been identified as a potential target of MTA1[55] which could be associated with increased pre-BCR signaling. Pre-BCR signaling may be a useful therapeutic target because a subset of BCP-ALL depends on it for survival,[56] and BTK inhibitors such as ibrutinib are effective.[57]

This study has several limitations. First, we did not perform functional analyses of the aberrant miRNAs in MLC. The phenotypic changes conferred by these miRNA alterations in BCP-ALL cells *in vitro* and *in vivo* need to be investigated. Second, since this study was a retrospective cohort study with a relatively small sample size and various treatment contexts, caution should be exercised when interpreting the results. Future validation is needed to determine whether the results can be replicated in the modern era when immunotherapy is available for relapsed cases. As only limited miRNA-seq data

for pediatric BCP-ALL are still available, further data accumulation in large prospective cohorts will be required to optimize prognostic models for clinical application. Recently, whole genome sequencing identified new genetic abnormalities in BCP-ALL[58] and future analyses combining transcriptome and whole-genome sequencing are needed to obtain a complete overview of the genomics of miRNAs in BCP-ALL.

In conclusion, our miRNA-mRNA integration analysis comprehensively identified the miRNA profiles associated with pediatric BCP-ALL and clearly identified the novel prognostic miRNA signature in a group of patients at high risk of relapse. Risk stratification based on miRNA expression profiles is expected to improve the treatment outcomes for pediatric patients with BCP-ALL.

Acknowledgments: The authors would thank the patients and their families for their cooperation. We also thank K. Kodama for technical assistance. This work was supported by JSPS KAKENHI Grant Numbers JP17H04224, JP18K19467, JP20H00528, and JP21K19405 (to JT), the Project for Cancer Research and Therapeutic Evolution (P-CREATE; grant no. JP19cm0106509h9904), the Project for Promotion of Cancer Research and Therapeutic Evolution (P-PROMOTE; grant no. JP22cm0106xxxh000x and 23ama221505h0002), the Practical Research for Innovative Cancer Control (grant no. JP19ck0106468h0001) from the Japan Agency for Medical Research and Development (AMED), the Princess Takamatsu Cancer Research Fund (all to JT), the Takeda Hosho Grants for Research in Medicine (to JT), and Ishizue from Kyoto University Research Administration (to JT).

Authorship

Contributions: H.K., H.U., H.H., and J.T. designed the study. M.H., D.H., and T.I. provided specimens and collected clinical data. H.K. and K.T. prepared the samples. H.K., N.K., Y.N., and S.O. performed sequencing. H.K., H.U., and T.I. performed bioinformatics analysis. S.S., I.K., and K.U. participated in helpful discussions and commented on the research direction. H.K., H.H., and J.T. wrote the manuscript. All authors reviewed the manuscript during preparation.

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

Correspondence: Junko Takita, Department of Pediatrics, Graduate School of Medicine, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto, 606-8507, Japan; Phone: +81-75-751-3290; Fax: +81-75-752-2361; E-mail address: jtakita@kuhp.kyoto-u.ac.jp; and Hidefumi Hiramatsu, Department of Pediatrics, Graduate School of Medicine, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto, 606-8507, Japan; Phone: +81-75-751-3290; Fax: +81-75-752-2361; E-mail address: hiramatu@kuhp.kyoto-u.ac.jp

References

1. Brady SW, Roberts KG, Gu Z, Shi L, Pounds S, Pei D, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat Genet*. 2022;54(9):1376–1389.
2. Jeha S, Choi J, Roberts KG, Pei D, Coustan-Smith E, Inaba H, et al. Clinical Significance of Novel Subtypes of Acute Lymphoblastic Leukemia in the Context of Minimal Residual Disease–Directed Therapy. *Blood Cancer Discov*. 2021;2(4):326–337.
3. Ueno H, Yoshida K, Shiozawa Y, Nannya Y, Iijima-Yamashita Y, Kiyokawa N, et al. Landscape of driver mutations and their clinical impacts in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood Adv*. 2020;4(20):5165–5173.
4. Teachey DT, Pui CH. Comparative features and outcomes between paediatric T-cell and B-cell acute lymphoblastic leukaemia. *Lancet Oncol*. 2019;20(3):e142–e154.
5. Irving JAE, Enshaei A, Parker CA, Sutton R, Kuiper RP, Erhorn A, et al. Integration of genetic and clinical risk factors improves prognostication in relapsed childhood B-cell precursor acute lymphoblastic leukemia. *Blood*. 2016;128(7):911–922.
6. Teachey DT, Hunger SP. Predicting relapse risk in childhood acute lymphoblastic leukaemia. *Br J Haematol*. 2013;162(5):606–620.
7. Pasquinelli AE. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet*. 2012;13(4):271–282.
8. Peng Y, Croce CM. The role of microRNAs in human cancer. *Signal Transduct Target Ther*. 2016;1:15004.

9. Ziętara KJ, Lejman J, Wojciechowska K, Lejman M. The Importance of Selected Dysregulated microRNAs in Diagnosis and Prognosis of Childhood B-Cell Precursor Acute Lymphoblastic Leukemia. *Cancers (Basel)*. 2023;15(2):428.

10. Kyriakidis I, Kyriakidis K, Tsezou A. MicroRNAs and the Diagnosis of Childhood Acute Lymphoblastic Leukemia: Systematic Review, Meta-Analysis and Re-Analysis with Novel Small RNA-Seq Tools. *Cancers (Basel)*. 2022;14(16):3976.

11. Lim EL, Trinh DL, Ries RE, Wang J, Gerbing RB, Ma Y, et al. MicroRNA expression-based model indicates event-free survival in pediatric acute myeloid leukemia. *Journal of Clinical Oncology*. 2017;35(35):3964–3977.

12. Larsen EC, Devidas M, Chen S, Salzer WL, Raetz EA, Loh ML, et al. Dexamethasone and high-dose methotrexate improve outcome for children and young adults with high-risk B-acute lymphoblastic leukemia: A report from children's oncology group study AALL0232. *Journal of Clinical Oncology*. 2016;34(20):2380–2388.

13. Maloney KW, Devidas M, Wang C, Mattano LA, Friedmann AM, Buckley P, et al. Outcome in Children With Standard-Risk B-Cell Acute Lymphoblastic Leukemia: Results of Children's Oncology Group Trial AALL0331. *J Clin Oncol*. 2019;38:602–612.

14. Hasegawa D, Imamura T, Yumura-Yagi K, Takahashi Y, Usami I, Suenobu S ichi, et al. Risk-adjusted therapy for pediatric non-T cell ALL improves outcomes for standard risk patients: results of JACLS ALL-02. *Blood Cancer J*. 2020;10(2):23.

15. Takahashi H, Kajiwara R, Kato M, Hasegawa D, Tomizawa D, Noguchi Y, et al. Treatment outcome of children with acute lymphoblastic leukemia: the Tokyo Children's Cancer Study Group (TCCSG) Study L04-16. *Int J Hematol*.

2018;108(1):98–108.

16. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol*. 2020;38(3):276–278.

17. Tian L, Li Y, Edmonson MN, Zhou X, Newman S, McLeod C, et al. CICERO: A versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol*. 2020;21(1):1–18.

18. Harvey RC, Mullighan CG, Wang X, Dobbin KK, Davidson GS, Bedrick EJ, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010;116(23):4874–4884.

19. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature 2016 540:7633*. 2016;540(7633):433–437.

20. Yuan T, Yang Y, Chen J, Li W, Li W, Zhang Q, et al. Regulation of PI3K signaling in T-cell acute lymphoblastic leukemia: a novel PTEN/Ikaros/miR-26b mechanism reveals a critical targetable role for PIK3CD. *Leukemia*. 2017;31(11):2355–2364.

21. Qian L, Zhang W, Lei B, He A, Ye L, Li X, et al. MicroRNA-101 regulates T-cell acute lymphoblastic leukemia progression and chemotherapeutic sensitivity by targeting Notch1. *Oncol Rep*. 2016;36(5):2511–2516.

22. Wang X-S, Gong J-N, Yu J, Wang F, Zhang X-H, Yin X-L, et al. MicroRNA-29a and microRNA-142-3p are regulators of myeloid differentiation and acute myeloid leukemia. *Blood*. 2012;119(21):4992–5004.

23. Mitsumura T, Ito Y, Chiba T, Matsushima T, Kurimoto R, Tanaka Y, et al. Ablation of miR-146b in mice causes hematopoietic malignancy. *Blood Adv*. 2018;2(23):3483–3491.

24. Ghodousi ES, Aberuyi N, Rahgozar S. Simultaneous changes in expression levels of BAALC and miR-326: A novel prognostic biomarker for childhood ALL. *Jpn J Clin Oncol*. 2020;50(6):671–678.

25. Malouf C, Antunes ETB, O'Dwyer M, Jakobczyk H, Sahm F, Landua SL, et al. miR-130b and miR-128a are essential lineage-specific codrivers of t(4;11) MLL-AF4 acute leukemia. *Blood*. 2021;138(21):2066–2092.

26. Li Y, Vecchiarelli-Federico LM, Li YJ, Egan SE, Spaner D, Hough MR, et al. The miR-17-92 cluster expands multipotent hematopoietic progenitors whereas imbalanced expression of its individual oncogenic miRNAs promotes leukemia in mice. *Blood*. 2012;119(19):4486–4498.

27. Borowitz MJ, Group13 for the CO, Devidas M, Group13 for the CO, Hunger SP, Group13 for the CO, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood*. 2008;111(12):5477–5485.

28. Borowitz MJ, Wood BL, Devidas M, Loh ML, Raetz EA, Salzer WL, et al. Prognostic significance of minimal residual disease in high risk B-ALL: a report from Children's Oncology Group study AALL0232. *Blood*. 2015;126(8):964–971.

29. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1–22.

30. Enshaei A, Vora A, Harrison CJ, Moppett J, Moorman A V. Defining low-risk high hyperdiploidy in patients with paediatric acute lymphoblastic leukaemia: a

retrospective analysis of data from the UKALL97/99 and UKALL2003 clinical trials. *Lancet Haematol*. 2021;8(11):e828–e839.

31. Li Y-T, Liu C-J, Kao J-H, Lin L-F, Tu H-C, Wang C-C, et al. Metastatic tumor antigen 1 contributes to hepatocarcinogenesis posttranscriptionally through RNA-binding function. *Hepatology*. 2023;77(2):379–394.

32. Deng L, Tang J, Yang H, Cheng C, Lu S, Jiang R, et al. MTA1 modulated by miR-30e contributes to epithelial-to-mesenchymal transition in hepatocellular carcinoma through an ErbB2-dependent pathway. *Oncogene*. 2017;36(28):3976–3985.

33. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. 2013;14(1):7.

34. Chen D, Zheng J, Gerasimcik N, Lagerstedt K, Sjögren H, Abrahamsson J, et al. The expression pattern of the Pre-B cell receptor components correlates with cellular stage and clinical outcome in acute lymphoblastic leukemia. *PLoS One*. 2016;11(9):e0162638.

35. Lee RD, Munro SA, Knutson TP, LaRue RS, Heltemes-Harris LM, Farrar MA. Single-cell analysis identifies dynamic gene expression networks that govern B cell development and transformation. *Nat Commun*. 2021;12(1):6843.

36. Koralov SB, Muljo SA, Galler GR, Krek A, Chakraborty T, Kanellopoulou C, et al. Dicer Ablation Affects Antibody Diversity and Cell Survival in the B Lymphocyte Lineage. *Cell*. 2008;132(5):860–874.

37. McLeod C, Gout AM, Zhou X, Thrasher A, Rahbarinia D, Brady SW, et al. St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov*. 2021;11(5):1082–1099.

38. Vedanayagam J, Chatila WK, Aksoy BA, Majumdar S, Skanderup AJ, Demir E, et al. Cancer-associated mutations in DICER1 RNase IIIa and IIIb domains exert similar effects on miRNA biogenesis. *Nat Commun*. 2019;10(1):3682.

39. Chen C, Hao X, Lai X, Liu L, Zhu J, Shao H, et al. Oxidative phosphorylation enhances the leukemogenic capacity and resistance to chemotherapy of B cell acute lymphoblastic leukemia. *Sci Adv*. 2021;7(11):eabd6280.

40. Foulkes WD, Priest JR, Duchaine TF. DICER1: mutations, microRNAs and mechanisms. *Nat Rev Cancer*. 2014;14(10):662–672.

41. Li JF, Dai YT, Lilljebjörn H, Shen SH, Cui BW, Bai L, et al. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc Natl Acad Sci U S A*. 2018;115(50):E11711–E11720.

42. Chang Y-H, Jou S-T, Yen C-T, Lin C-Y, Yu C-H, Chang S-K, et al. A microRNA signature for clinical outcomes of pediatric ALL patients treated with TPOG protocols. *Am J Cancer Res*. 2022;12(10):4764–4774.

43. Rashed WM, Hamza MM, Matboli M, Salem SI. MicroRNA as a prognostic biomarker for survival in childhood acute lymphoblastic leukemia: a systematic review. *Cancer Metastasis Rev*. 2019;38(4):771–782.

44. Godoy PM, Barczak AJ, DeHoff P, Srinivasan S, Etheridge A, Galas D, et al. Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA Quantification Platforms. *Cell Rep*. 2019;29(12):4212-4222.e5.

45. Stanulla M, Dagdan E, Zaliova M, Möricke A, Palmi C, Cazzaniga G, et al. IKZF1 plus defines a new minimal residual disease-dependent very-poor prognostic profile in pediatric b-cell precursor acute lymphoblastic leukemia. *Journal of*

Clinical Oncology. 2018;36(12):1240–1249.

46. Mangum DS, Meyer JA, Mason CC, Shams S, Maese LD, Gardiner JD, et al. Association of Combined Focal 22q11.22 Deletion and IKZF1 Alterations With Outcomes in Childhood Acute Lymphoblastic Leukemia. *JAMA Oncol*. 2021;7(10):1521–1528.

47. Locatelli F, Zugmaier G, Rizzari C, Morris JD, Gruhn B, Klingebiel T, et al. Effect of Blinatumomab vs Chemotherapy on Event-Free Survival Among Children With High-risk First-Relapse B-Cell Acute Lymphoblastic Leukemia: A Randomized Clinical Trial. *JAMA*. 2021;325(9):843–854.

48. Ahmadi SE, Rahimi S, Zarandi B, Chegeni R, Safa M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. J Hematol Oncol. 2021;9:121.

49. Black KL, Naqvi AS, Asnani M, Hayer KE, Yang SY, Gillespie E, et al. Aberrant splicing in B-cell acute lymphoblastic leukemia. *Nucleic Acids Res*. 2018;46(21):11357–11369.

50. Closa A, Reixachs-Solé M, Fuentes-Fayos AC, Hayer KE, Melero JL, Adriaanse FRS, et al. A convergent malignant phenotype in B-cell acute lymphoblastic leukemia involving the splicing factor SRRM1. *NAR Cancer*. 2022;4(4):zcac041.

51. Shiozawa Y, Malcovati L, Gallì A, Sato-Otsubo A, Kataoka K, Sato Y, et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat Commun*. 2018;9(1):3649.

52. Shah JS, Milevskiy MJG, Petrova V, Au AYM, Wong JJL, Visvader JE, et al. Towards resolution of the intron retention paradox in breast cancer. *Breast Cancer Research*. 2022;24(1):1–11.

53. Ullrich S, Guigó R. Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during B-cell development. *Nucleic Acids Res*. 2021;48(3):1327–1340.

54. Hsu TYT, Simon LM, Neill NJ, Marcotte R, Sayad A, Bland CS, et al. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. *Nature*. 2015;525(7569):384–388.

55. Balasenthil S, Gururaj AE, Talukder AH, Bagheri-Yarmand R, Arrington T, Haas BJ, et al. Identification of Pax5 as a target of MTA1 in B-cell lymphomas. *Cancer Res*. 2007;67(15):7132–7138.

56. Geng H, Hurtz C, Lenz KB, Chen Z, Baumjohann D, Thompson S, et al. Self-Enforcing Feedback Activation between BCL6 and Pre-B Cell Receptor Signaling Defines a Distinct Subtype of Acute Lymphoblastic Leukemia. *Cancer Cell*. 2015;27(3):409–425.

57. Kim E, Hurtz C, Koehrer S, Wang Z, Balasubramanian S, Chang BY, et al. Ibrutinib inhibits pre-BCR+ B-cell acute lymphoblastic leukemia progression by targeting BTK and BLK. *Blood*. 2017;129(9):1155–1165.

58. Ryan SL, Peden JF, Kingsbury Z, Schwab CJ, James T, Polonen P, et al. Whole genome sequencing provides comprehensive genetic testing in childhood B-cell acute lymphoblastic leukaemia. *Leukemia*. 2023;37(3):518–528.

Figure Legends

Figure. 1. miRNA expression clusters in 111 BCP-ALL cases based on unsupervised consensus clustering and clinical impact of the MLC miRNA signature.

(A) Heatmap of miRNA expression by 111 primary samples from the TARGET cohort, along with the clinical information for each case. Two stable clusters were identified by consensus clustering of the 111 samples using 500 miRNAs (the distance method: Pearson) (Supplementary Figure 2A). (B) Gene expression profiling of the 111 primary samples is shown as a two-dimensional tSNE plot. Each dot represents one sample. The top 1,000 most variable mRNA genes (based on the median absolute deviation) (top row) and the top 500 most variable miRNAs (bottom row) were selected and processed by the tSNE algorithm, with a perplexity score of 30. Genetic BCP-ALL subtypes are highlighted in different colors. Circles and triangles indicate Cluster 1 and Cluster 2, respectively. (C) MA plot for differential expression analysis between Cluster 1 and Cluster 2 generated by DESeq2: for each miRNA, the log2 (fold change) (log2(Cluster 2/Cluster 1)) is plotted (A, y axis) against the log2 (average normalized expression) of the gene in the two clusters (M, x axis). Significantly differentially expressed genes with an adjusted *P*-value < 0.05 are shown in red. (D) Volcano plot showing differences in miRNAs expressed in Cluster 2 samples (compared with Cluster 1 samples) among the highly expressed miRNAs with a median of 25 or more reads per million. Significantly differentially expressed miRNAs showing a log2-fold change (Cluster 2/Cluster 1) > 1 or < -1, and an adjusted *P*-value < 0.05, are shown in red. (E) Kaplan–Meier survival curves of event-free survival (EFS) and overall survival (OS) for 111 BCP-ALL cases and 23 hyperdiploid BCP-ALL cases with and without the MLC signature. *P* values were

calculated by the log-rank test. (F) Forest plot showing multivariate Cox regression analysis of the effect of different parameters on EFS. Squares represent the hazard ratio, and horizontal lines represent the confidence interval (CI). WBC at Dx = white blood cell count at diagnosis; MRD = minimal residual disease.

Figure. 2. Validation of the MLC signature in the Japanese cohort.

(A) Strategy used to identify and test the MLC score for the 19-miRNA signature. RPM = reads per million. (B) Heatmap of miRNA expression by 69 primary samples from the Japanese cohort, along with clinical information for each case. Six clusters were identified by consensus clustering of the 69 samples using 500 miRNAs (the distance method: Pearson). Cases with an MLC score in the third quartile or higher were defined as having a high MLC score. Samples in Cluster 2 had a high MLC score and were labeled as MLC-like (n = 19). (C) The distribution of MLC scores and event-free survival (EFS) in the Japanese cohort. (D) Kaplan–Meier survival curves of event-free survival (EFS) and overall survival (OS) for BCP-ALL cases with and without an MLC-like signature. *P* values are based on the log-rank test. (E) Time-dependent receiver operating characteristic curve analysis and AUC (area under the curve) of the predictive power of MLC score, miR-92b-3p, and miR-1304-3p expression on 2-year EFS and OS in the TARGET and Japanese cohort. TP = True positive rate; FP = False positive rate.

Figure. 3. Gene mutations in the MLC BCP-ALL cases.

(A) Recurrent driver mutations identified by variant calls from RNA sequencing, and copy number variations from the SNP array, along with clinical information, for 111 BCP-ALL primary samples from the TARGET cohort. Cases are grouped as MLC or non-MLC according to miRNA expression clusters. WBC = white blood cells. (B) Gene mutations and MLC signature pairs showing co-occurrence or exclusivity in the 111 BCP-ALL primary samples from the TARGET cohort are illustrated as a triangular matrix. Green indicates a tendency toward co-occurrence, whereas brown indicates exclusivity. The point indicates $P < 0.1$, and the asterisk indicates $P < 0.05$.

Figure. 4. The MLC signature of relapsed BCP-ALL.

(A) Heatmap showing miRNA expression by 86 paired primary and relapse samples from the TARGET cohort (35 cases) and the Japanese cohort (eight cases), along with clinical information for each case. Two clusters were identified by consensus clustering of the 86 samples using 400 miRNAs (the distance method: Pearson). Primary samples with an MLC or MLC-like signature are denoted as MLC at diagnosis. (B) An alluvial plot illustrates the transition of the miRNA cluster from the primary to relapse samples from 43 cases; red indicates primary samples classified as MLC (20 cases), and blue indicates those classified as non-MLC (23 cases). (C and D) Box plots comparing the MLC scores between paired primary and relapse samples (C) in the nine cases that acquired MLC-like miRNA profiles at relapse, and (D) in the 23 non-MLC cases at diagnosis. *P* values are calculated using the Wilcoxon signed-rank test.

Figure. 5. Integrative miRNA-mRNA analysis of BCP-ALL.

(A) Workflow for the integrative miRNA-mRNA analysis of 111 BCP-ALL cases from the TARGET ALL cohort. (B and C) Dot plot showing the top enriched (B) Hallmark gene sets, and (C) Gene Ontology terms ranked according to gene ratio based on upregulated genes in sample pairs showing decreased miRNA expression and increased mRNA expression in the 111 BCP-ALL primary samples from the TARGET cohort. The adjusted *P*-value is represented by the color scale, and the size of the dot reflects the number of genes assigned to the gene sets. Over-representation analysis was performed based on a one-sided Fisher's exact test conducted using the R package clusterProfiler. (D) Workflow for gene set enrichment analysis of genes significantly upregulated in relapse samples compared with primary samples from the nine cases that acquired an MLC profile at relapse. FDR: false discovery rate. (E) Dot plot showing the top enriched Hallmark gene sets, ranked according to the gene ratio of the upregulated genes shown in (D). (F) Type and number of splicing alterations significantly associated with MLC compared with non-MLC cases in the TARGET cohort. The bars on the right denote alternative splicing events that were increased in MLC cases. The bars on the left denote those that were decreased in MLC cases. (G) Venn diagram of decreased intron retention in the MLC compared with that in normal pro-B cell and non-MLC samples, genes showing increased expression in the MLC compared with non-MLC samples, and genes with negatively correlated miRNA-mRNA pairs in MLC samples. (H) miR-30e-5p and *MTA1* show inversely correlated expression patterns in 111 cases from the TARGET ALL cohort. Red dots represent MLC cases and black dots represent non-MLC cases. RPM = reads per million; FPKM = fragments per kilobase million. (I) Heat map and clustering based on gene set variation analysis enrichment scores for genes upregulated in normal

B-precursor cells from 111 cases in the TARGET ALL cohort (left). Gene set enrichment analysis and enrichment plot for MLC and non-MLC cases, showing enrichment of genes upregulated in pre-B cells from MLC cases (right). CLP = common lymphoid progenitor; NES = normalized enrichment score; FDR = false discovery rate.

Figure. 6. *DICER1* expression in BCP-ALL.

(A) Box plots comparing *DICER1* expression (expressed as fragments per kilobase million (FPKM)) by normal pro-B cells (n = 4), pre-B cells (n = 4), immature B cells (n = 4), non-MLC cases (n = 80), and MLC cases (n = 31) from the TARGET ALL cohort. The *P*-value was calculated using the Wilcoxon rank-sum test, adjusted using Bonferroni-correction. (B) Box plots comparing *DICER1* expression (FPKM) in non-MLC-like cases (n = 50) and MLC-like cases (n = 19) from the Japanese cohort. The *P*-value was calculated using the Wilcoxon rank-sum test. (C) Box plots comparing *DICER1* expression (FPKM) in paired primary and relapse samples from nine cases that acquired an MLC signature at relapse. The *P*-value was calculated using the Wilcoxon signed-rank test. (D) Box plots showing quantification of miRNA processing, based on the median 5p to (5p + 3p) miRNA ratio in non-MLC cases (n = 80) and MLC cases (n = 31) from the TARGET ALL cohort. The *P*-value was calculated using the Wilcoxon rank-sum test. (E) Immunoblotting to detect DICER1 and GAPDH expression by REH cells transduced with control (shLuc) or *DICER1* shRNAs (shDICER1). Cells were treated for 72 h with 3 μM doxycycline (DOX) or an equivalent amount of dimethyl sulfoxide (DMSO). (F) Box plots comparing the viability of REH cells transduced with control or shDICER1 after exposure to cytarabine. REH cells were transduced with shLuc or shDICER1 and then cultured in the presence of 3 μM doxycycline, or an equivalent amount of DMSO, for 72 h. After the indicated concentrations of cytarabine were added to the culture medium, cell viability (%) relative to that of control wells containing culture medium and the equivalent amount of vehicle was measured after 48 h in a WST-8 assay using Cell Counting Kit-8 (n = 6). *P* values were calculated using the Wilcoxon signed-rank test. **P* < 0.05, ***P* < 0.01.

Figure. 1

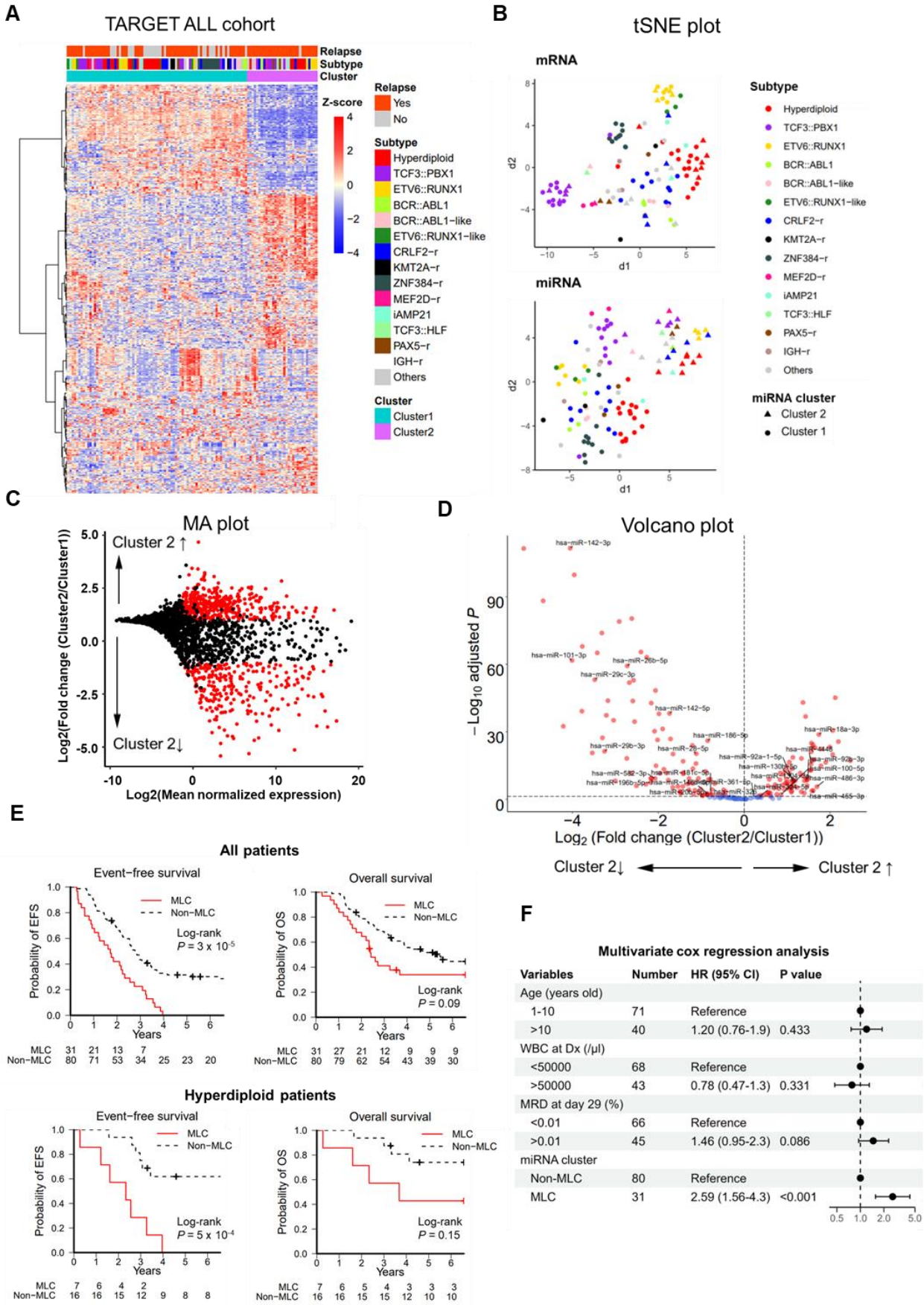

A
TARGET ALL cohort
Relapse
Subtype
Cluster
Z-score
Relapse
Yes
No
Subtype
Hyperdiploid
TCF3::PBX1
ETV6::RUNX1
BCR::ABL1
BCR::ABL1−like
ETV6::RUNX1−like
CRLF2−r
KMT2A−r
ZNF384−r
MEF2D−r
iAMP21
TCF3::HLF
PAX5−r
IGH−r
Others
Cluster
Cluster1
Cluster2
B
tSNE plot
mRNA
miRNA
d1
d2
Subtype
miRNA cluster
Cluster 2
Cluster 1
C
MA plot
Cluster 2 ↑
Cluster 2↓
Log2(Fold change (Cluster2/Cluster1))
Log2(Mean normalized expression)
D
Volcano plot
− Log10 adjusted P
Log2 (Fold change (Cluster2/Cluster1))
Cluster 2↓
Cluster 2 ↑
E
All patients
Event−free survival
Overall survival
Probability of EFS
Probability of OS
MLC
Non−MLC
Log-rank
P = 3 x 10-5
P = 0.09
Years
Hyperdiploid patients
P = 5 x 10-4
P = 0.15
F
Multivariate cox regression analysis
Variables
Number
HR (95% CI)
P value
Age (years old)
1-10
71
Reference
>10
40
1.20 (0.76-1.9)
0.433
WBC at Dx (/μl)
<50000
68
Reference
>50000
43
0.78 (0.47-1.3)
0.331
MRD at day 29 (%)
<0.01
66
Reference
>0.01
45
1.46 (0.95-2.3)
0.086
miRNA cluster
Non-MLC
80
Reference
MLC
31
2.59 (1.56-4.3)
<0.001

Figure. 2

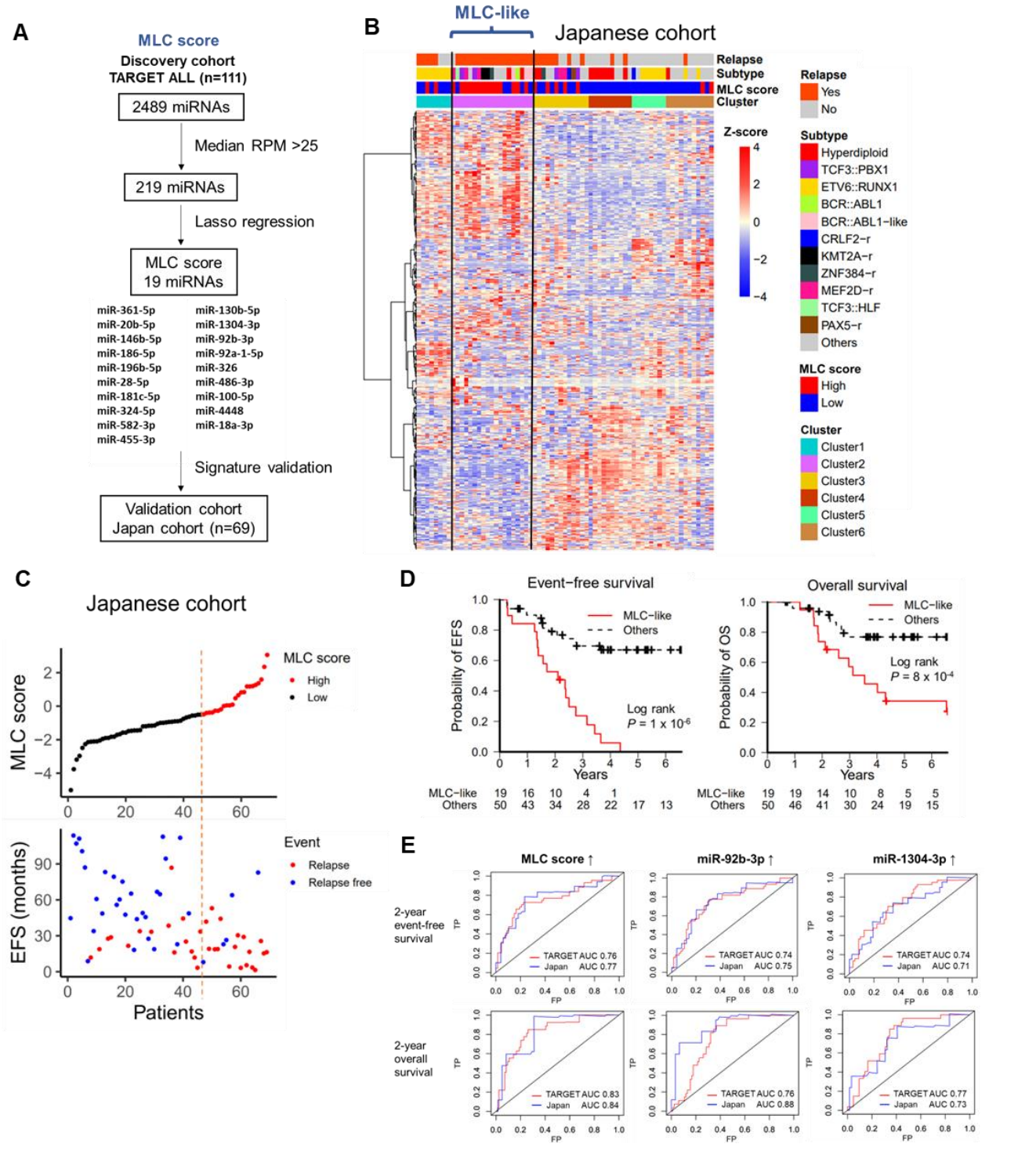

A
MLC score
Discovery cohort
TARGET ALL (n=111)
2489 miRNAs
Median RPM >25
219 miRNAs
Lasso regression
MLC score
19 miRNAs
miR-361-5p
miR-20b-5p
miR-146b-5p
miR-186-5p
miR-196b-5p
miR-28-5p
miR-181c-5p
miR-324-5p
miR-582-3p
miR-455-3p
miR-130b-5p
miR-1304-3p
miR-92b-3p
miR-92a-1-5p
miR-326
miR-486-3p
miR-100-5p
miR-4448
miR-18a-3p
Signature validation
Validation cohort
Japan cohort (n=69)
B
MLC-like
Japanese cohort
Relapse
Subtype
MLC score
Cluster
Z-score
Relapse
Yes
No
Subtype
Hyperdiploid
TCF3::PBX1
ETV6::RUNX1
BCR::ABL1
BCR::ABL1−like
CRLF2−r
KMT2A−r
ZNF384−r
MEF2D−r
TCF3::HLF
PAX5−r
Others
MLC score
High
Low
Cluster
Cluster1
Cluster2
Cluster3
Cluster4
Cluster5
Cluster6
C
Japanese cohort
MLC score
High
Low
Event
Relapse
Relapse free
Patients
D
Event−free survival
Probability of EFS
MLC−like
Others
Log rank
P = 1 x 10−6
Years
MLC−like 19 16 10 4 1
Others 50 43 34 28 22 17 13
Overall survival
Probability of OS
MLC−like
Others
Log rank
P = 8 x 10−4
Years
MLC−like 19 19 14 10 8 5 5
Others 50 46 41 30 24 19 15
E
MLC score ↑
miR-92b-3p ↑
miR-1304-3p ↑
2-year event-free survival
TARGET AUC 0.76
Japan AUC 0.77
TARGET AUC 0.74
Japan AUC 0.75
TARGET AUC 0.74
Japan AUC 0.71
2-year overall survival
TARGET AUC 0.83
Japan AUC 0.84
TARGET AUC 0.76
Japan AUC 0.88
TARGET AUC 0.77
Japan AUC 0.73
TP
FP

Figure. 3

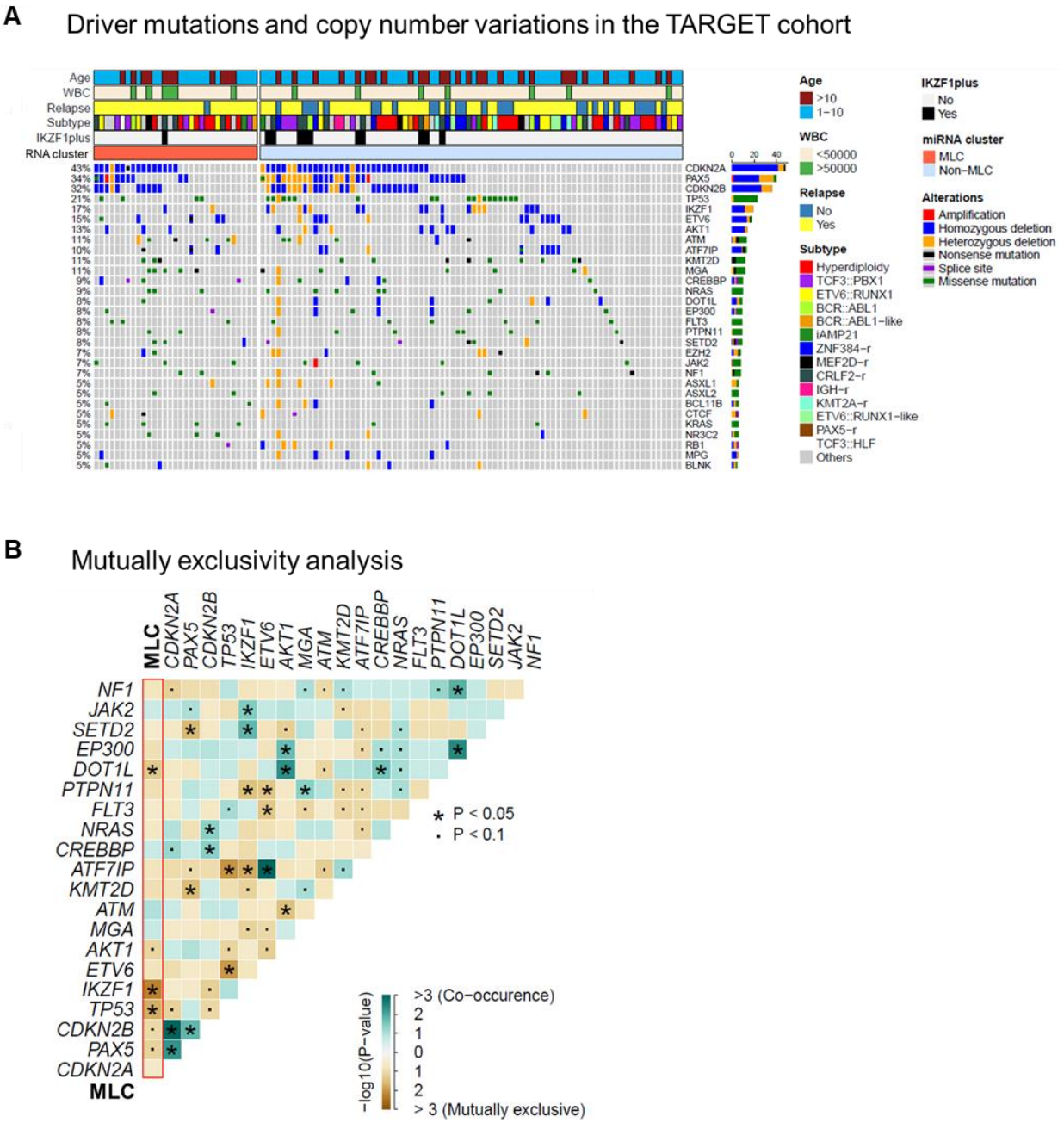

A
Driver mutations and copy number variations in the TARGET cohort
Age
WBC
Relapse
Subtype
IKZF1plus
RNA cluster
CDKN2A
PAX5
CDKN2B
TP53
IKZF1
ETV6
AKT1
ATM
ATF7IP
KMT2D
MGA
CREBBP
NRAS
DOT1L
EP300
FLT3
PTPN11
SETD2
EZH2
JAK2
NF1
ASXL1
ASXL2
BCL11B
CTCF
KRAS
NR3C2
RB1
MPG
BLNK
Age
>10
1–10
WBC
<50000
>50000
Relapse
No
Yes
Subtype
Hyperdiploidy
TCF3::PBX1
ETV6::RUNX1
BCR::ABL1
BCR::ABL1–like
iAMP21
ZNF384–r
MEF2D–r
CRLF2–r
IGH–r
KMT2A–r
ETV6::RUNX1–like
PAX5–r
TCF3::HLF
Others
IKZF1plus
No
Yes
miRNA cluster
MLC
Non–MLC
Alterations
Amplification
Homozygous deletion
Heterozygous deletion
Nonsense mutation
Splice site
Missense mutation
B
Mutually exclusivity analysis
MLC
CDKN2A
PAX5
CDKN2B
TP53
IKZF1
ETV6
AKT1
MGA
ATM
KMT2D
ATF7IP
CREBBP
NRAS
FLT3
PTPN11
DOT1L
EP300
SETD2
JAK2
NF1
* P < 0.05
· P < 0.1
−log10(P−value)
>3 (Co−occurence)
2
1
0
1
2
> 3 (Mutually exclusive)

Figure. 4

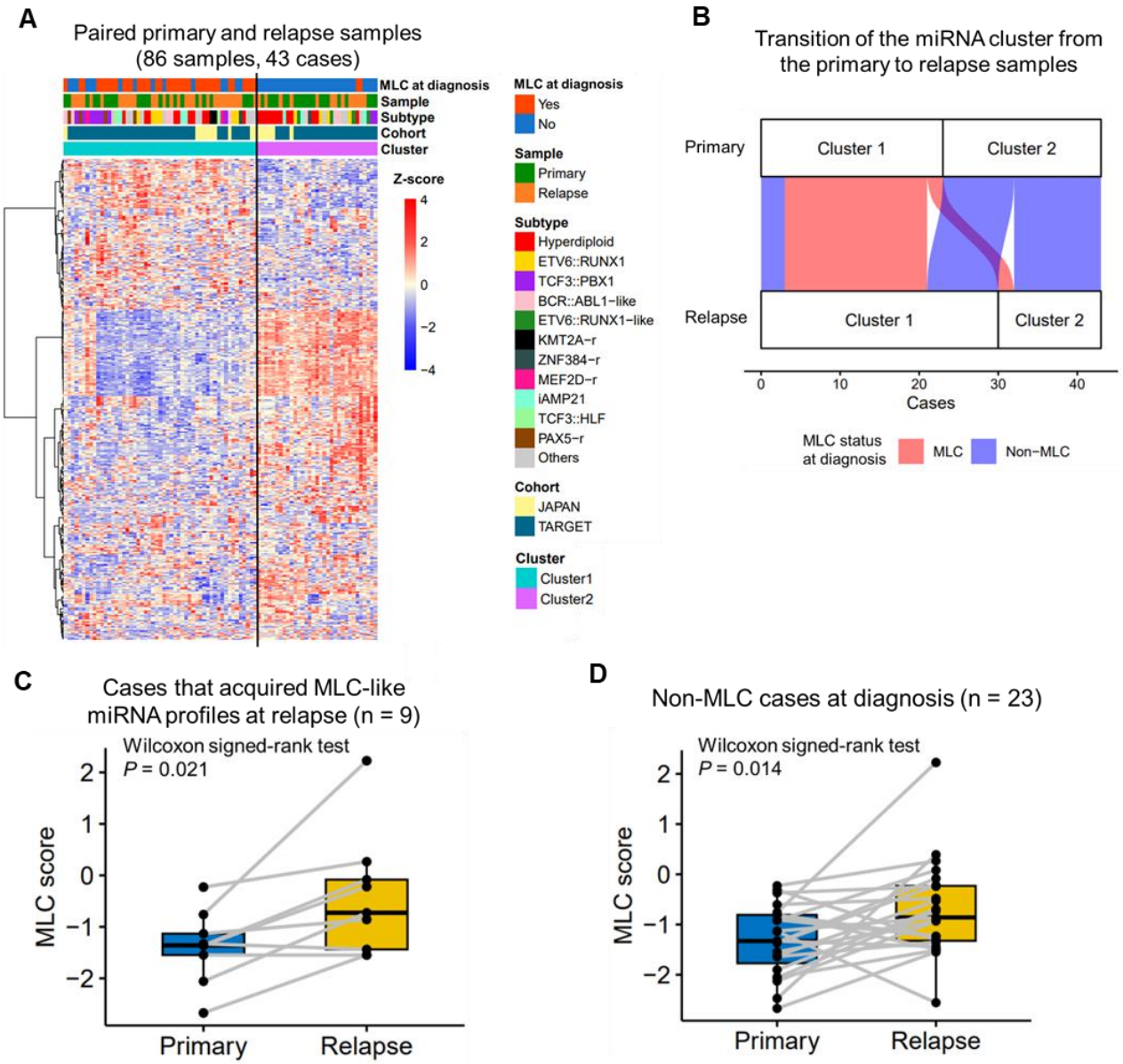

A
Paired primary and relapse samples
(86 samples, 43 cases)
MLC at diagnosis
Sample
Subtype
Cohort
Cluster
Z-score
4
2
0
−2
−4
MLC at diagnosis
Yes
No
Sample
Primary
Relapse
Subtype
Hyperdiploid
ETV6::RUNX1
TCF3::PBX1
BCR::ABL1−like
ETV6::RUNX1−like
KMT2A−r
ZNF384−r
MEF2D−r
iAMP21
TCF3::HLF
PAX5−r
Others
Cohort
JAPAN
TARGET
Cluster
Cluster1
Cluster2
B
Transition of the miRNA cluster from
the primary to relapse samples
Primary
Relapse
Cluster 1
Cluster 2
Cluster 1
Cluster 2
0
10
20
30
40
Cases
MLC status
at diagnosis
MLC
Non−MLC
C
Cases that acquired MLC-like
miRNA profiles at relapse (n = 9)
Wilcoxon signed-rank test
P = 0.021
MLC score
2
1
0
−1
−2
Primary
Relapse
D
Non-MLC cases at diagnosis (n = 23)
Wilcoxon signed-rank test
P = 0.014
MLC score
2
1
0
−1
−2
Primary
Relapse

Figure. 5

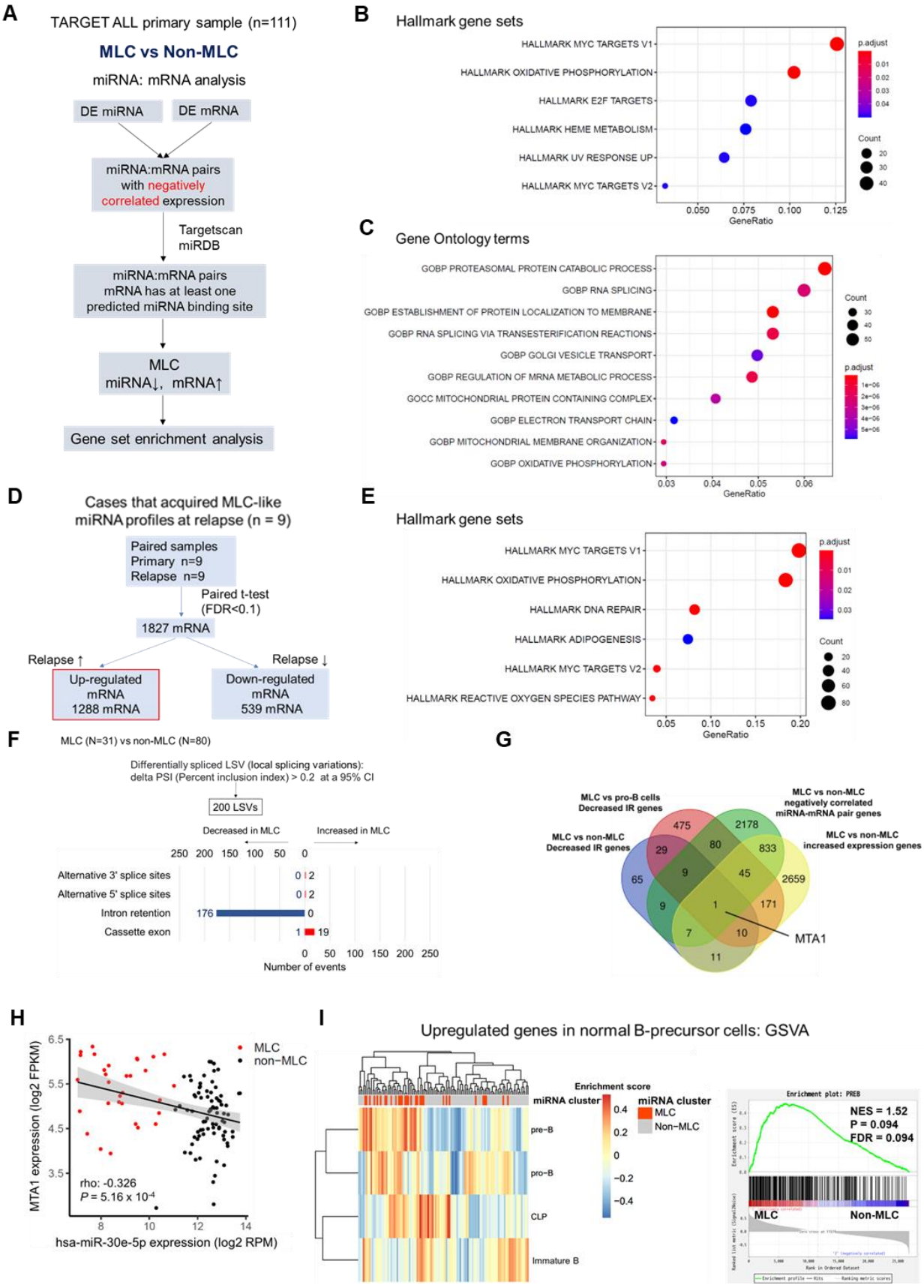
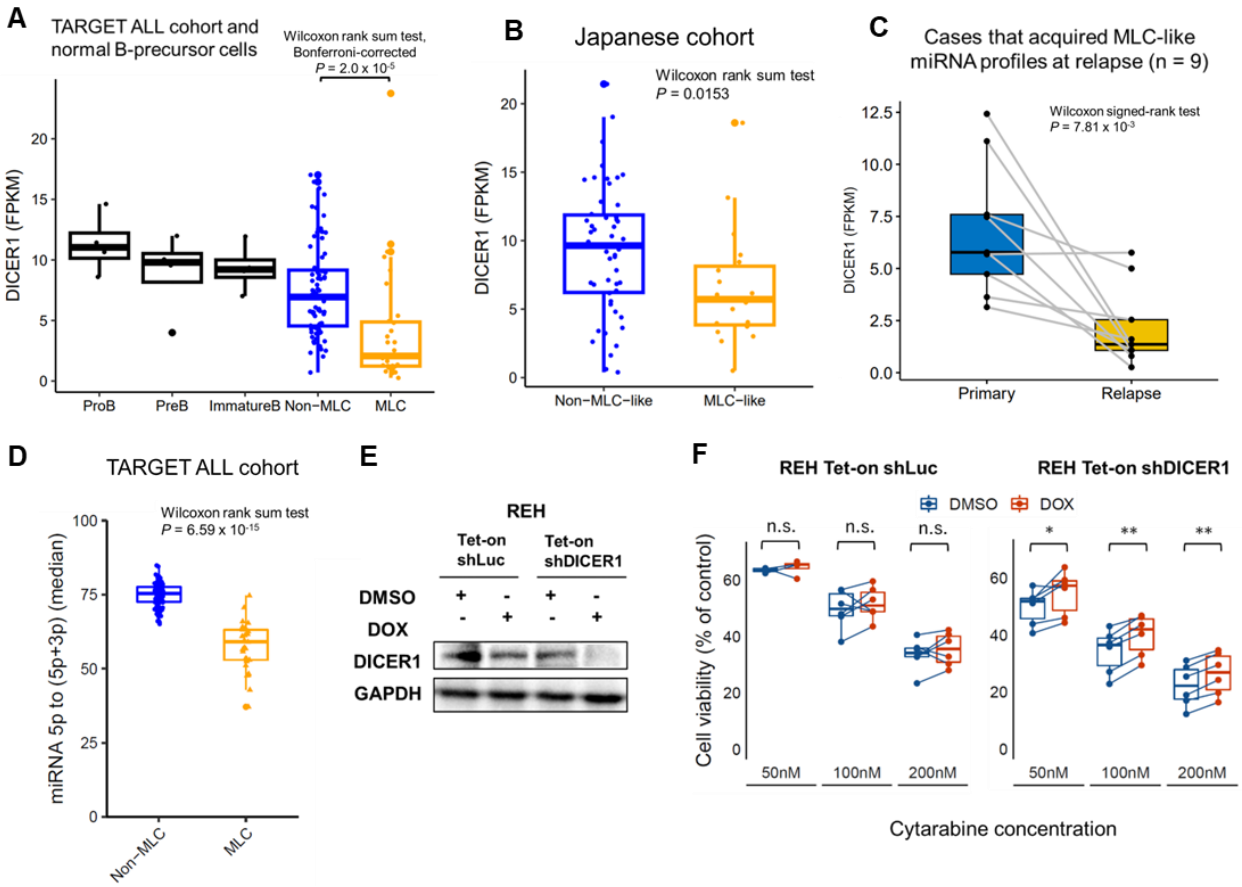

A
TARGET ALL primary sample (n=111)
MLC vs Non-MLC
miRNA: mRNA analysis
DE miRNA
DE mRNA
miRNA:mRNA pairs with negatively correlated expression
Targetscan miRDB
miRNA:mRNA pairs mRNA has at least one predicted miRNA binding site
MLC miRNA↓, mRNA↑
Gene set enrichment analysis
B
Hallmark gene sets
HALLMARK MYC TARGETS V1
HALLMARK OXIDATIVE PHOSPHORYLATION
HALLMARK E2F TARGETS
HALLMARK HEME METABOLISM
HALLMARK UV RESPONSE UP
HALLMARK MYC TARGETS V2
p.adjust
Count
GeneRatio
C
Gene Ontology terms
GOBP PROTEASOMAL PROTEIN CATABOLIC PROCESS
GOBP RNA SPLICING
GOBP ESTABLISHMENT OF PROTEIN LOCALIZATION TO MEMBRANE
GOBP RNA SPLICING VIA TRANSESTERIFICATION REACTIONS
GOBP GOLGI VESICLE TRANSPORT
GOBP REGULATION OF MRNA METABOLIC PROCESS
GOCC MITOCHONDRIAL PROTEIN CONTAINING COMPLEX
GOBP ELECTRON TRANSPORT CHAIN
GOBP MITOCHONDRIAL MEMBRANE ORGANIZATION
GOBP OXIDATIVE PHOSPHORYLATION
Count
p.adjust
GeneRatio
D
Cases that acquired MLC-like miRNA profiles at relapse (n = 9)
Paired samples Primary n=9 Relapse n=9
Paired t-test (FDR<0.1)
1827 mRNA
Relapse ↑
Relapse ↓
Up-regulated mRNA 1288 mRNA
Down-regulated mRNA 539 mRNA
E
Hallmark gene sets
HALLMARK MYC TARGETS V1
HALLMARK OXIDATIVE PHOSPHORYLATION
HALLMARK DNA REPAIR
HALLMARK ADIPOGENESIS
HALLMARK MYC TARGETS V2
HALLMARK REACTIVE OXYGEN SPECIES PATHWAY
p.adjust
Count
GeneRatio
F
MLC (N=31) vs non-MLC (N=80)
Differentially spliced LSV (local splicing variations): delta PSI (Percent inclusion index) > 0.2 at a 95% CI
200 LSVs
Decreased in MLC
Increased in MLC
Alternative 3' splice sites
Alternative 5' splice sites
Intron retention
Cassette exon
Number of events
G
MLC vs pro-B cells Decreased IR genes
MLC vs non-MLC negatively correlated miRNA-mRNA pair genes
MLC vs non-MLC Decreased IR genes
MLC vs non-MLC increased expression genes
MTA1
H
MLC
non-MLC
MTA1 expression (log2 FPKM)
rho: -0.326
P = 5.16 x 10-4
hsa-miR-30e-5p expression (log2 RPM)
I
Upregulated genes in normal B-precursor cells: GSVA
Enrichment score
miRNA cluster
MLC
Non-MLC
pre-B
pro-B
CLP
Immature B
Enrichment plot: PREB
NES = 1.52
P = 0.094
FDR = 0.094
MLC
Non-MLC

Figure. 6


A
TARGET ALL cohort and normal B-precursor cells
Wilcoxon rank sum test, Bonferroni-corrected
P = 2.0 x 10-5
DICER1 (FPKM)
ProB
PreB
ImmatureB
Non−MLC
MLC
B
Japanese cohort
Wilcoxon rank sum test
P = 0.0153
DICER1 (FPKM)
Non−MLC−like
MLC−like
C
Cases that acquired MLC-like miRNA profiles at relapse (n = 9)
Wilcoxon signed-rank test
P = 7.81 x 10-3
DICER1 (FPKM)
Primary
Relapse
D
TARGET ALL cohort
Wilcoxon rank sum test
P = 6.59 x 10-15
miRNA 5p to (5p+3p) (median)
Non−MLC
MLC
E
REH
Tet-on shLuc
Tet-on shDICER1
DMSO
DOX
DICER1
GAPDH
F
REH Tet-on shLuc
REH Tet-on shDICER1
DMSO
DOX
n.s.
*
**
Cell viability (% of control)
50nM
100nM
200nM
Cytarabine concentration

Supplementary data

Title: RNA-seq-based miRNA signature as an independent predictor of relapse in pediatric B-cell acute lymphoblastic leukemia

Authors: Hirohito Kubota, Hiroo Ueno, Keiji Tasaka, Tomoya Isobe, Satoshi Saida, Itaru Kato, Katsutsugu Umeda, Mitsuteru Hiwatari, Daiichiro Hasegawa, Toshihiko Imamura, Nobuyuki Kakiuchi, Yasuhito Nannya, Seishi Ogawa, Hidefumi Hiramatsu, Junko Takita

Supplementary Methods

Survival analysis

Overall survival (OS) was calculated from the date of diagnosis to the date of death from any cause, and shorter event-free survival (EFS) was calculated from the date of diagnosis until the date of the first event (failure to achieve remission, relapse, secondary malignancy, or death by any cause) or the date of the last follow-up. The Kaplan–Meier method was used to estimate OS, EFS and their associated 95% confidence intervals (CIs), and the log-rank test was used to assess differences between the groups. Analyses were performed using the R package survival v3.2.13. The EFS was also evaluated using the Cox proportional-hazards model. The model included age, white blood cell count, MRD status at day 29, and miRNA cluster as covariates. P values of less than 0.05 were considered statistically significant. The R package survivalROC v1.0.3.1 was used to plot and calculate the area under the curve (AUC) of the time-dependent receiver operating characteristic (ROC) curve to evaluate the predictive power of prognostic model and miRNA expression.

miRNA signature training and validation

To extract a core subset of miRNAs from among the 219 most highly expressed (RPM > 25) miRNAs in the TARGET ALL cohort that best explained the miR-low cluster (MLC) signature in the discovery cohort, a linear regression technique based on the Lasso algorithm was used (as implemented in the glmnet v4.1.2 R package) while still enabling leave-one-out cross-validation to fit a Cox regression model.[1] The MLC score was calculated for each patient as a linear combination of the z-score-normalized expression (reads per million; RPM), using the mean and standard deviations from the training data (the TARGET ALL cohort) for the 19 miRNAs, weighted by regression coefficients estimated from the training data, as follows: MLC score = (miR-100-5p × 0.123) + (miR-1304-3p × 0.0116) + (miR-130b-5p × 0.0592) + (miR-146b-5p × −0.313)

+ (miR-181c-5p × -0.0303) + (miR-186-5p × -0.00347) + (miR-18a-3p × 0.119) + (miR-196b-5p × 0.104) + (miR-20b-5p × −0.0169) + (miR-28-5p × −0.0222) + (miR-324-5p × -0.108) + (miR-326 × −0.492) + (miR-361-5p × -0.0511) + (miR-4448 × 0.0496) + (miR-455-3p × 0.0556) + (miR-486-3p × 0.165) + (miR-582-3p × 0.054) + (miR-92a-1-5p x 0.00299) + (miR-92b-3p x 0.188) - 1.117. In the test cohort (the Japanese cohort), a group with a score in the highest tertile of the population was defined as having a high MLC score.

Copy number analysis based on SNP array data

Of the 111 cases in the TARGET ALL cohort, 105 had available SNP array data (Affymetrix SNP Array 6.0). Raw copy number segmentation results from Affymetrix Genomewide SNP6 arrays were downloaded from (https://target-data.nci.nih.gov/Public/ALL/copy_number_array/Phase2/L3/). The GISTIC 2.0 method was used to identify significant copy number variations. G-scores were plotted along each chromosome using a 99% CI at a false discovery threshold of 5%. Copy number variants arising from immunoglobulin/T cell receptor gene rearrangements were excluded. *IKZF1*^{plus} was defined as *IKZF1* deletions co-occurring with deletions in *CDKN2A*, *CDKN2B*, *PAX5*, or *PAR1* in the absence of *ERG* deletion.[2]

DNA methylation analysis

Of the 111 cases in the TARGET ALL cohort, 53 had available DNA methylation data (HELP Assay with Roche NimbleGen). DNA methylation was quantified using the median normalized log2 ratio of the HpaII to MspI signal intensity from the HELP Assay.

Whole exome sequencing (WES) and mutation calling

WES libraries were prepared using the Lotus DNA Library Prep Kit (IDT) and the xGen

Exome Research Panel (IDT), followed by sequencing of enriched exon fragments on an MGI DNBSEQ platform using a 150 bp paired-end read protocol. Candidate somatic mutations were identified using the Genomon pipeline v2.6.2. Variant calling of the WES data was conducted as follows: candidate mutations showing (1) a Fisher's exact test *P*-value of ≤ 0.1, (2) an Empirical Bayesian Mutation Calling *P*-value ≤ 0.001, and (3) a variant allele frequency in tumor samples > 5% were selected. These variants were further filtered by excluding (1) synonymous single nucleotide variants (SNVs), (2) SNVs in genes whose structures were not correctly annotated, and (3) SNVs listed as polymorphisms in the ExAc, ToMMO, tommo35K, 1000 Genomes Project, ESP6500, and HGVD databases, with a minor allele frequency of ≥ 0.001. In addition, mapping errors were removed by visual inspection using the Integrative Genomics Viewer browser (https://software.broadinstitute.org/software/igv/).

Analysis of copy number alterations identified by WES

The genomic copy number in the samples examined by WES was analyzed using the signal data calculated from the sequence reads. The copy number alterations were evaluated using an in-house CNACS pipeline (https://github.com/papaemmelab/toil_cnacs). Depth was calculated from the weighted sum of the fragments accounting for length and GC biases during sequencing library amplification. The depths were compared with those of the pooled controls. The genomic copy number was estimated from the depth ratios using the circular binary segmentation method. The allele frequencies of the heterozygous SNPs assessed allelic imbalance covered by >50 reads.

Co-occurrence and mutual exclusivity analyses of genetic alterations

A Poisson–Binomial distribution-based analysis implemented in the R package Rediscover v0.2.0 was performed to identify co-occurring or mutually exclusive pairs of

genetic alterations.

Pathway enrichment analysis

Pathway enrichment analysis of both ranked gene lists was performed using GSEA software (version 4.0.3; https://www.gsea-msigdb.org/gsea/index.jsp) and the R package clusterProfiler v.4.6.2. Gene set enrichment analysis (GSEA) was performed on normalized counts (FPKM). The analysis used the h.all.v.7.4.symbols.gmt (Hallmark) and c5.all.v.7.4.symbols.gmt (Gene Ontology) gene set databases. To identify B cell developmental pathways associated with BCP-ALL, gene set variation analysis[3] was performed on genes differentially upregulated in normal B progenitor cells[4] and relative pathway activity at the level of individual samples was determined using the R package GSVA v1.46.0.

Integrative miRNA-mRNA expression analysis

Integrated analyses of the miRNA and mRNA expression profiles were carried out using the differentially expressed miRNAs and mRNAs. Correlations between miRNA and mRNA expression were identified using the miRComb package v0.9.6. Spearman's correlation coefficients between a given miRNA and its predicted target mRNA were calculated and matched by target prediction using two databases (TargetScan and miRDB). Since miRNAs function as negative regulators, miRNA-mRNA pairs that correlated negatively (adjusted $P < 0.05$), and appeared in at least one of the two databases, were selected.

Splicing analysis

The MAJIQ tool (version 2.3) was used to detect local splice variations (LSVs).[5] MAJIQ was run on the Ensembl-based GFF annotations, disallowing *de novo* calls. Changes in the relative inclusion of LSVs (delta percent inclusion index, dPSI) of at

least 20% at the 95% CI between two given conditions were selected for further analysis. For comparative analysis of splicing with that of the normal B progenitor, the pro-B fraction was used as a control, as described previously.[6]

Cell lines and culture

Human REH cell lines were cultured at 37°C/5% CO_2 in RPMI 1640 medium supplemented with 10% fetal calf serum and 1% penicillin/streptomycin.

Western blotting

Protein samples were separated by SDS-PAGE, transferred to a polyvinylidene difluoride membrane, and probed with specific primary antibodies, followed by horseradish peroxidase (HRP)-conjugated secondary antibodies. Specific bands were visualized by enhanced chemiluminescence using the Clarity Western ECL Substrate (Bio-Rad) and quantified by the ChemiDoc XRS1 System and Image Laboratory software (Bio-Rad). Mouse monoclonal anti-DICER1 (F-10, 1:1,000; Santa Cruz Biotechnology) and mouse polyclonal anti-GAPDH (1:4,000; FUJIFILM Wako) were used as primary antibodies for western blotting. HRP-conjugated anti-mouse IgG Abs (1:4,000; Jackson ImmunoResearch Laboratories) were used as secondary antibodies.

Transduction of REH cell lines with doxycycline-inducible short hairpin RNA

Specific shRNAs targeting human *DICER1* and non-targeting control shRNA (for Luc), were designed and subcloned into the pENTR4-H1tetOx1 and CS-RfA-ETV vectors (RIKEN BRC). The following target sequences were used: *DICER1*-shRNA, 5′-ATTGGCTTCCTCCTGGTTA-3′; Luc-shRNA, 5′-CGTACGCGGAATACTTCGA-3′. For the production of lentivirus, HEK293T cells were transfected with lentiviral vectors by polyethylenimine (Sigma-Aldrich). Forty-eight hours after transfection, viral supernatants were harvested and immediately used for infection, followed by sorting of

successfully transduced cells with FACS Aria III (BD Biosciences). For tetracycline-inducible shRNA expression, doxycycline was added to the cultures at a final concentration of 3 μM for 72h. Knockdown efficiency was determined by western blotting.

Measurement of cell viability

Cells were plated in a normal growth medium in a 96-well cell culture plate (50000 cells per well). Different concentrations of cytarabine (FUJIFILM Wako) were added to the culture medium, and cell viability (%) relative to that of control wells containing cell culture medium and an equivalent amount of vehicle was measured at 48 h in a WST-8 assay using the Cell Counting Kit-8 (Dojin). Absorbance was measured using an iMark Microplate Absorbance Reader (Bio-Rad).

Statistical analysis

Statistical analyses were performed using R (v3.5.0, v3.6.3, and v4.1.0). All *P* values were calculated by two-sided analysis, unless otherwise specified. Student t-test, a paired t-test, Fisher's exact test, the Wilcoxon rank-sum test, and the Wilcoxon signed-rank test were used for group comparisons, and comparisons among multiple groups were analyzed by the Kruskal-Wallis test. Multiple testing was corrected using the Benjamini–Hochberg method. A *P*-value of <0.05 was considered statistically significant.

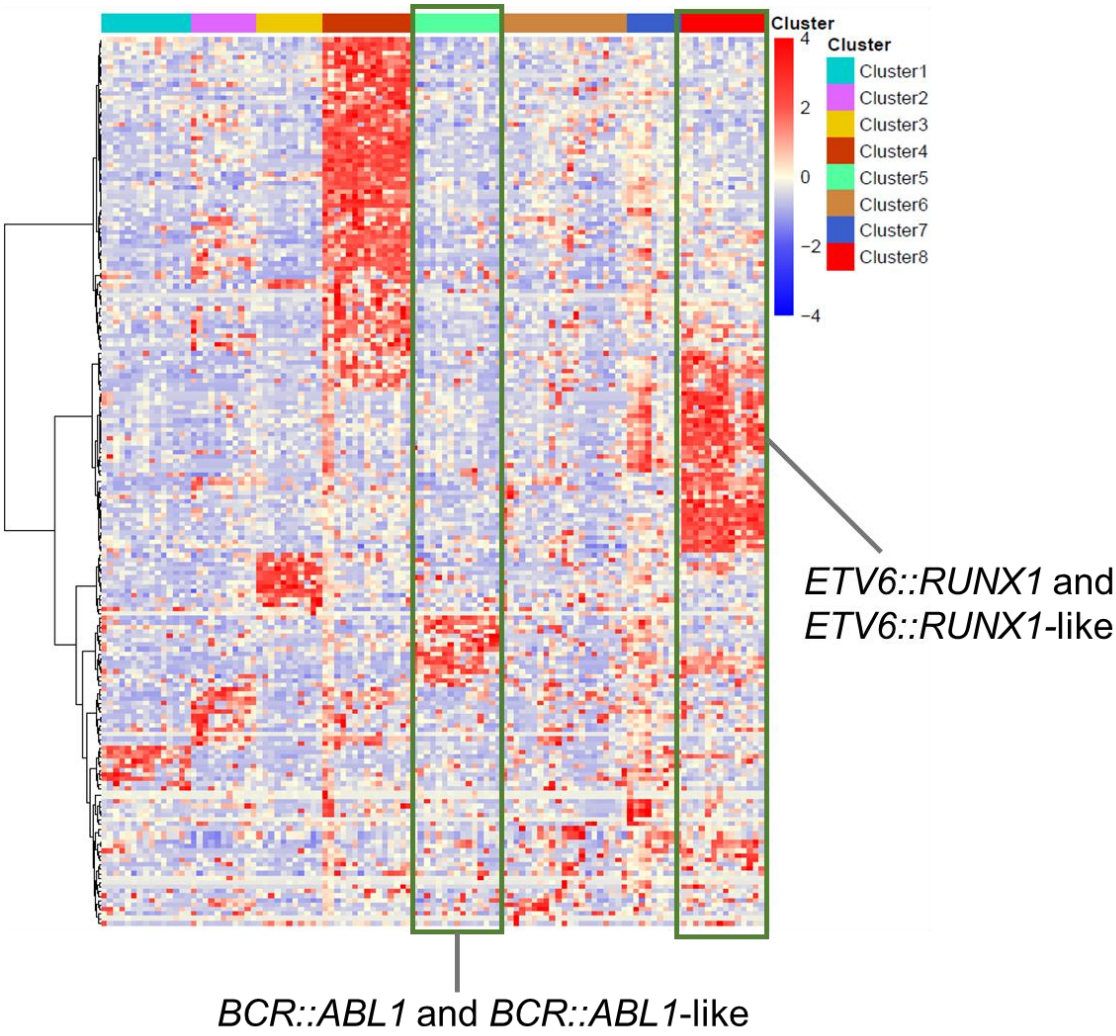
References

1. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature 2016 540:7633*. 2016;540(7633):433–437.
2. Stanulla M, Dagdan E, Zaliova M, Möricke A, Palmi C, Cazzaniga G, et al. IKZF1 plus defines a new minimal residual disease-dependent very-poor prognostic

profile in pediatric b-cell precursor acute lymphoblastic leukemia. *Journal of Clinical Oncology*. 2018;36(12):1240–1249.

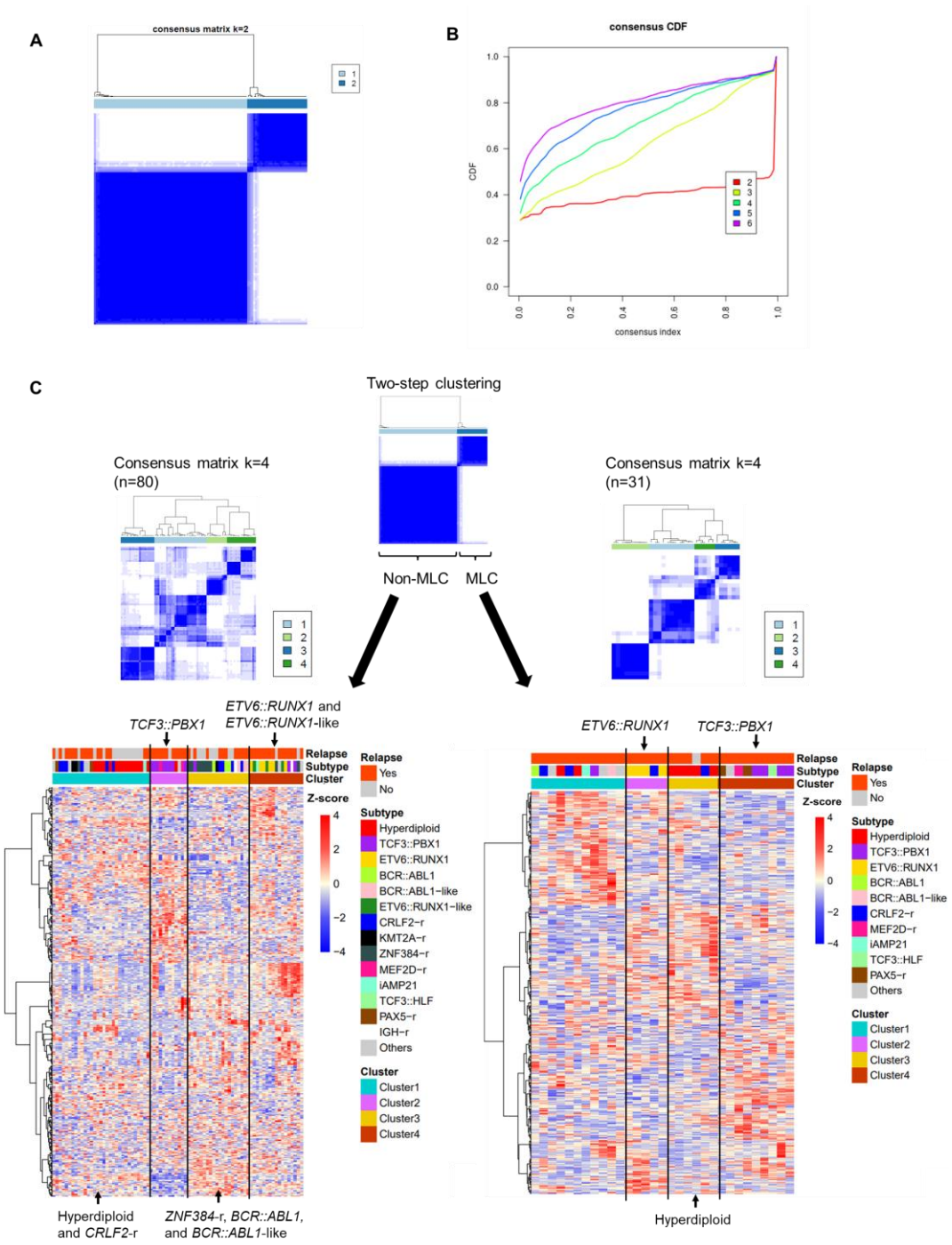
3. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. 2013;14(1):7.
4. Chen D, Zheng J, Gerasimcik N, Lagerstedt K, Sjögren H, Abrahamsson J, et al. The expression pattern of the Pre-B cell receptor components correlates with cellular stage and clinical outcome in acute lymphoblastic leukemia. *PLoS One*. 2016;11(9):e0162638.
5. Norton SS, Vaquero-Garcia J, Lahens NF, Grant GR, Barash Y. Outlier detection for improved differential splicing quantification from RNA-Seq experiments with replicates. *Bioinformatics*. 2018;34(9):1488–1497.
6. Black KL, Naqvi AS, Asnani M, Hayer KE, Yang SY, Gillespie E, et al. Aberrant splicing in B-cell acute lymphoblastic leukemia. *Nucleic Acids Res*. 2018;46(21):11357–11369.

Supplementary Figure 1.



Gene expression clusters in the 111 BCP-ALL samples from the TARGET cohort. The ROSE method identified 198 genes, including the top 25 genes in clusters 1-8 of the TARGET cohort. Ward's hierarchical clustering method was used for clustering analysis of the ROSE gene set. Four samples had a *BCR::ABL1*-like signature with no known fusion genes, and three samples had an *ETV6::RUNX1*-like signature.

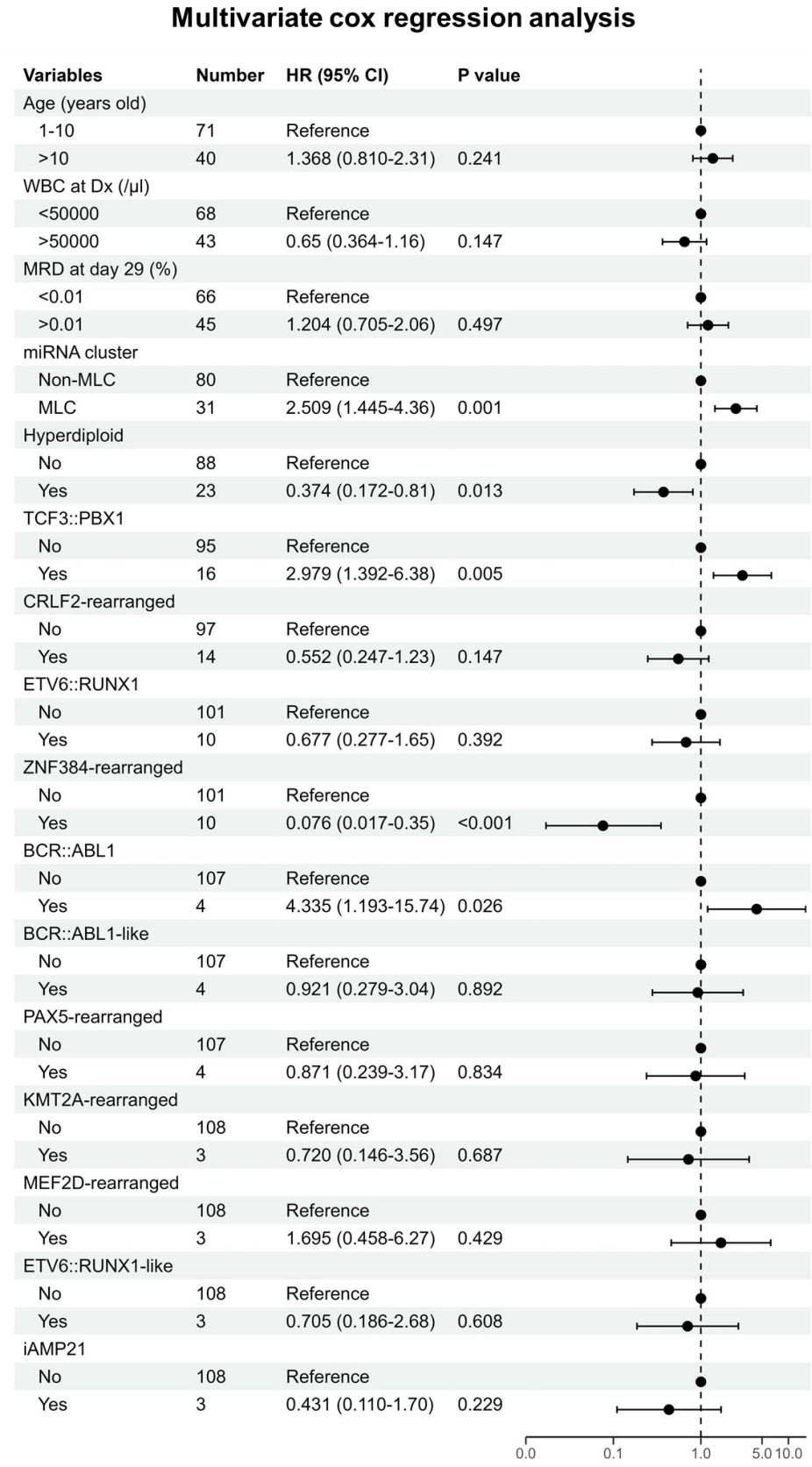
Supplementary Figure 2.



Consensus clustering analysis of miRNA expression in 111 BCP-ALL cases. (A) Consensus matrix and (B) CDF plots showing consensus clustering of all 111 samples. (C) Two-step unsupervised consensus clustering restricted to the non-MLC cases (n = 80) and MLC cases (n = 31) respectively. Consensus matrix and miRNA

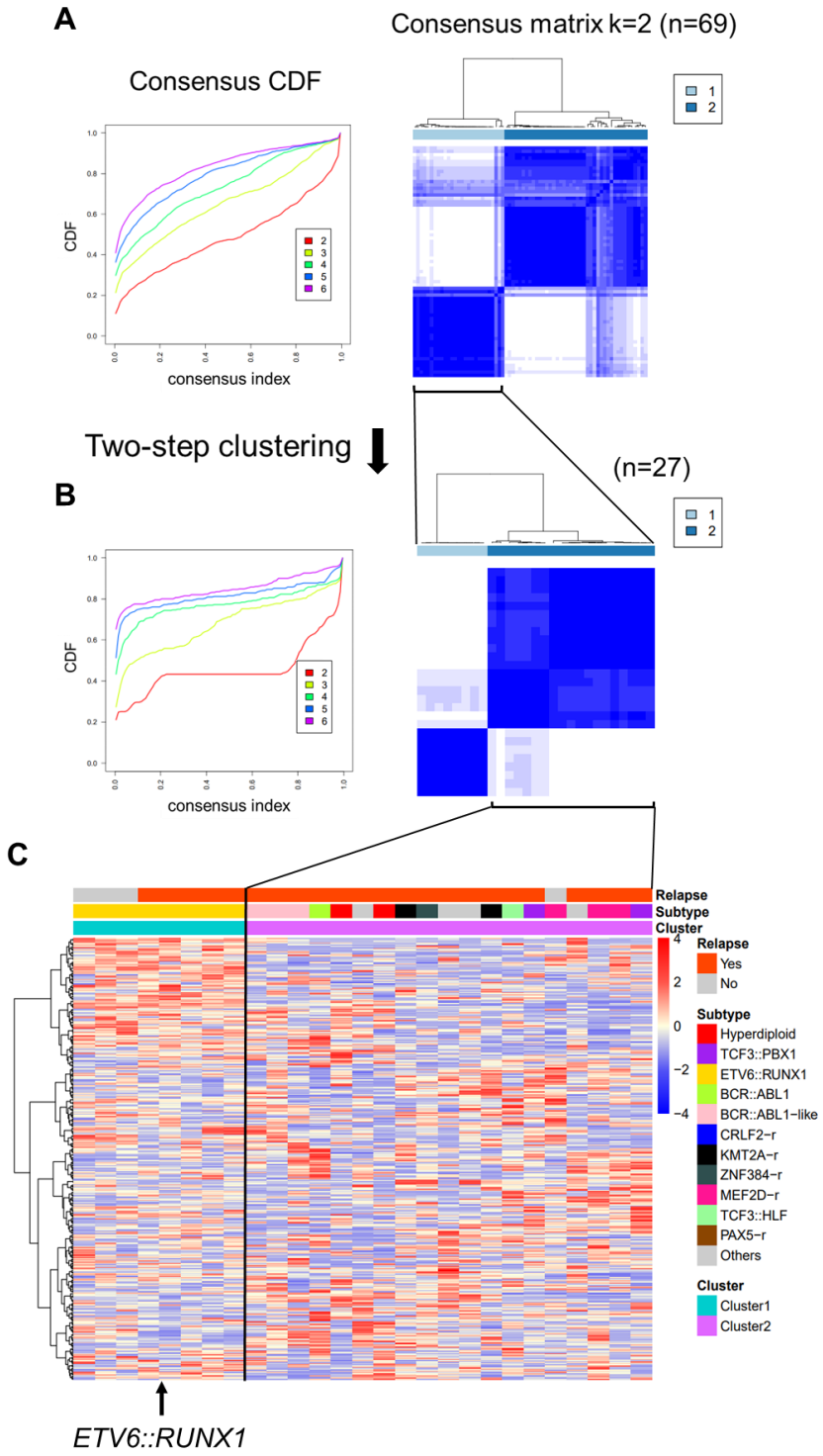
expression heatmap generated using 500 probes across 80 non-MLC primary samples, along with clinical information for each case.

Supplementary Figure 3.



Forest plot showing the results of multivariate Cox regression analysis of the effect of different parameters, including genetic subtype, on event-free survival.
Squares represent the hazard ratio and horizontal lines represent the confidence interval (CI). WBC at Dx, white blood cell count at diagnosis; MRD, minimal residual disease.

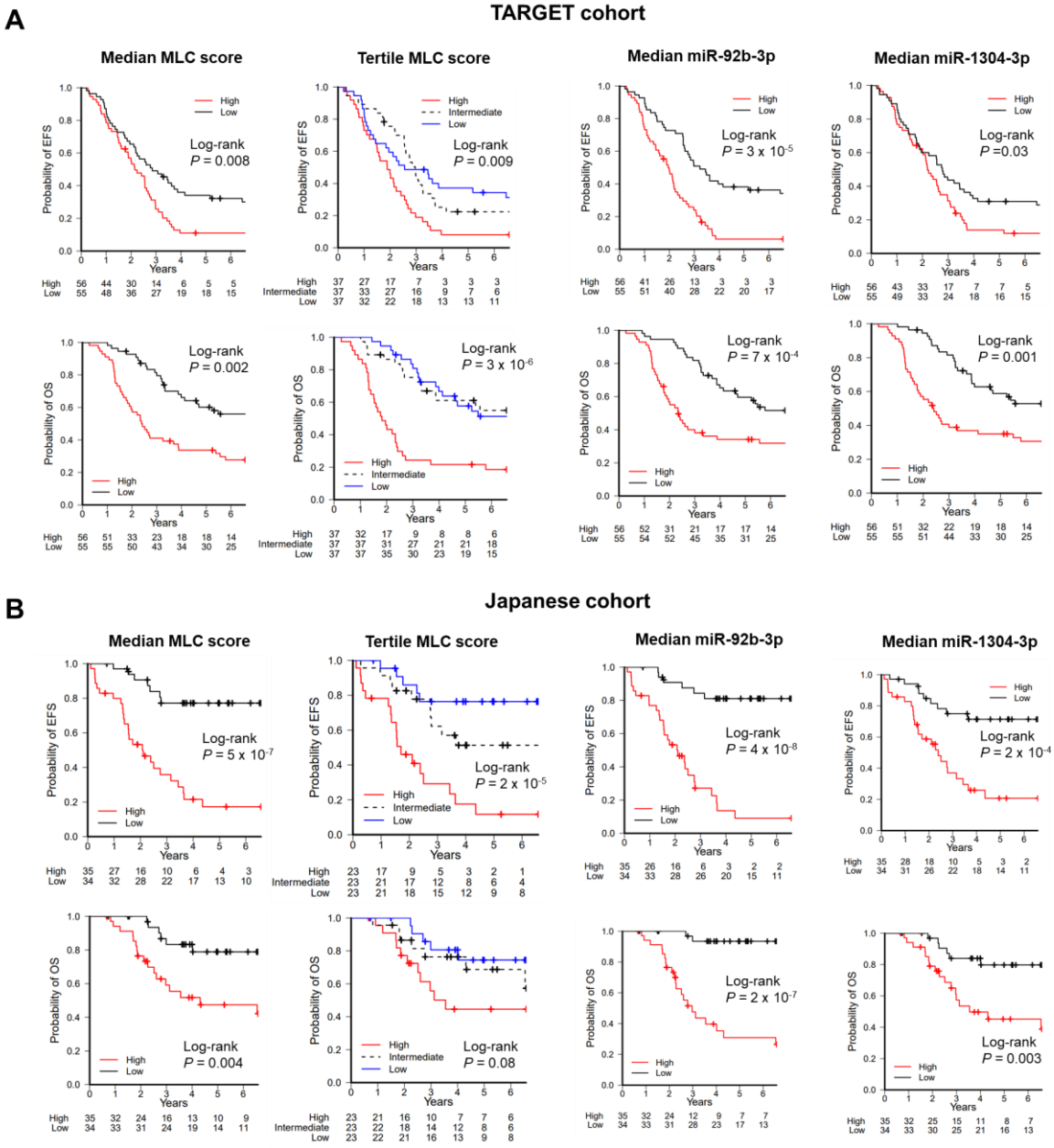
Supplementary Figure 4.



Two-step consecutive unsupervised consensus clustering of miRNA expression by 69 BCP-ALL cases in the Japanese cohort.

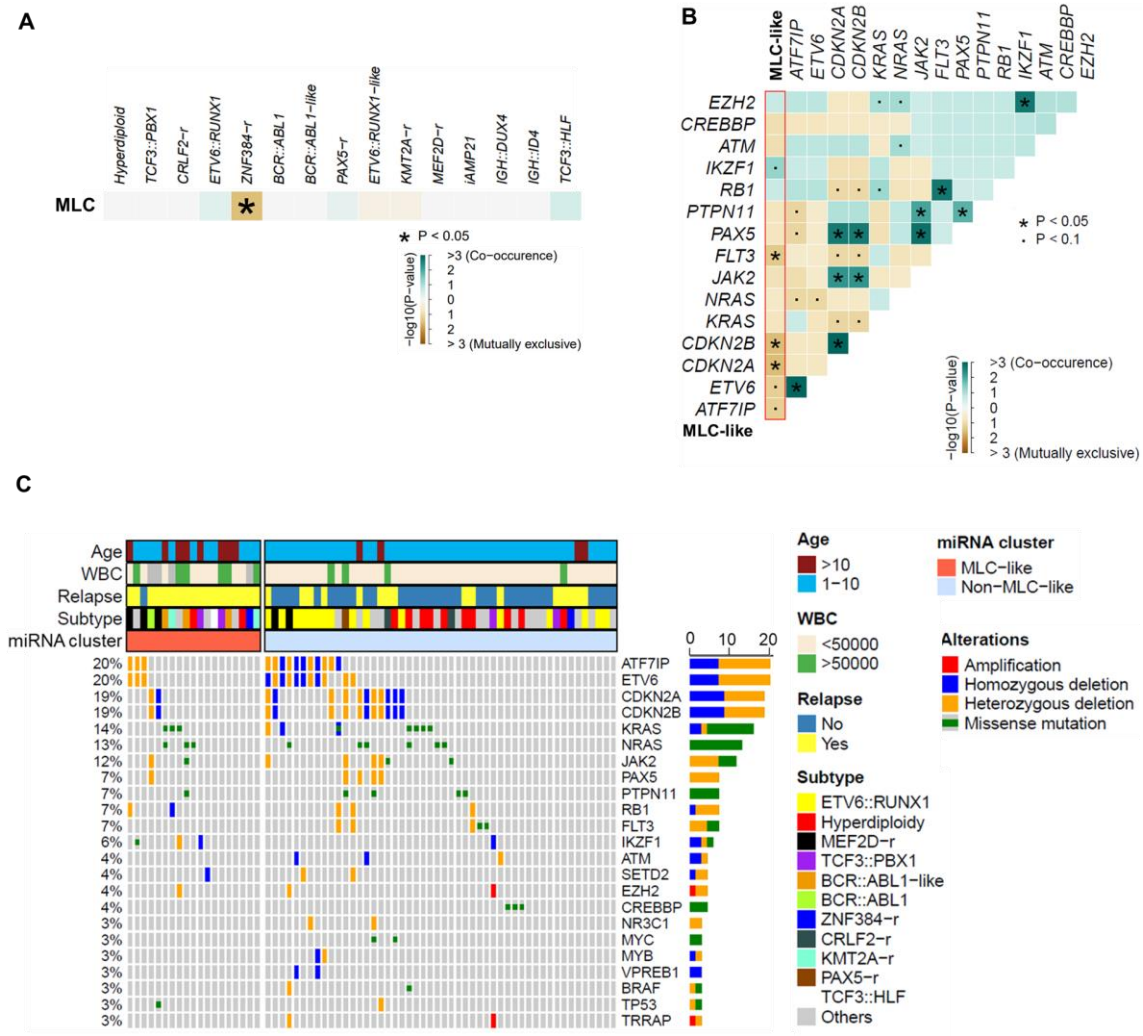
(A) CDF plots and consensus matrix showing consensus clustering of the 69 samples. (B) CDF plots and consensus matrix showing consensus clustering of 27 samples confined to the first cluster shown in (A). (C) Heatmap of miRNA expression generated using 400 probes across 27 samples, along with clinical information for each case.

Supplementary Figure 5.



Prognostic value of MLC scores, miR-92b-3p and miR-1304-3p in BCP-ALL cases. Kaplan–Meier survival curves of event-free survival (EFS) and overall survival (OS) for BCP-ALL cases with high or low level of the MLC score, miR-92b-3p and miR-1304-3p in the TARGET cohort (A) and the Japanese cohort (B). The cut-off for stratification was set to median or tertile for the MLC score and median for miRNAs. P values are based on the log-rank test.

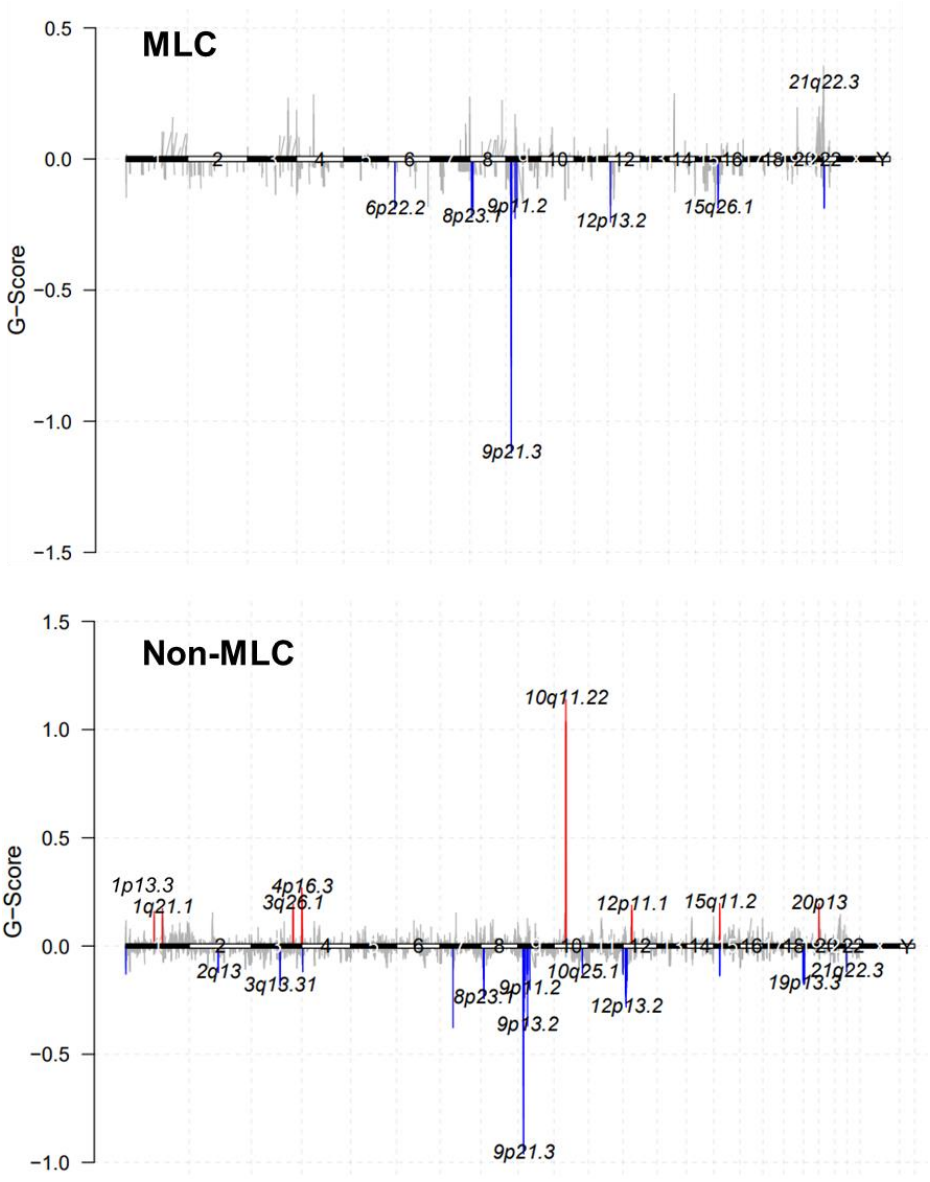
Supplementary Figure 6.



Gene mutations in MLC and MLC-like BCP-ALL cases.

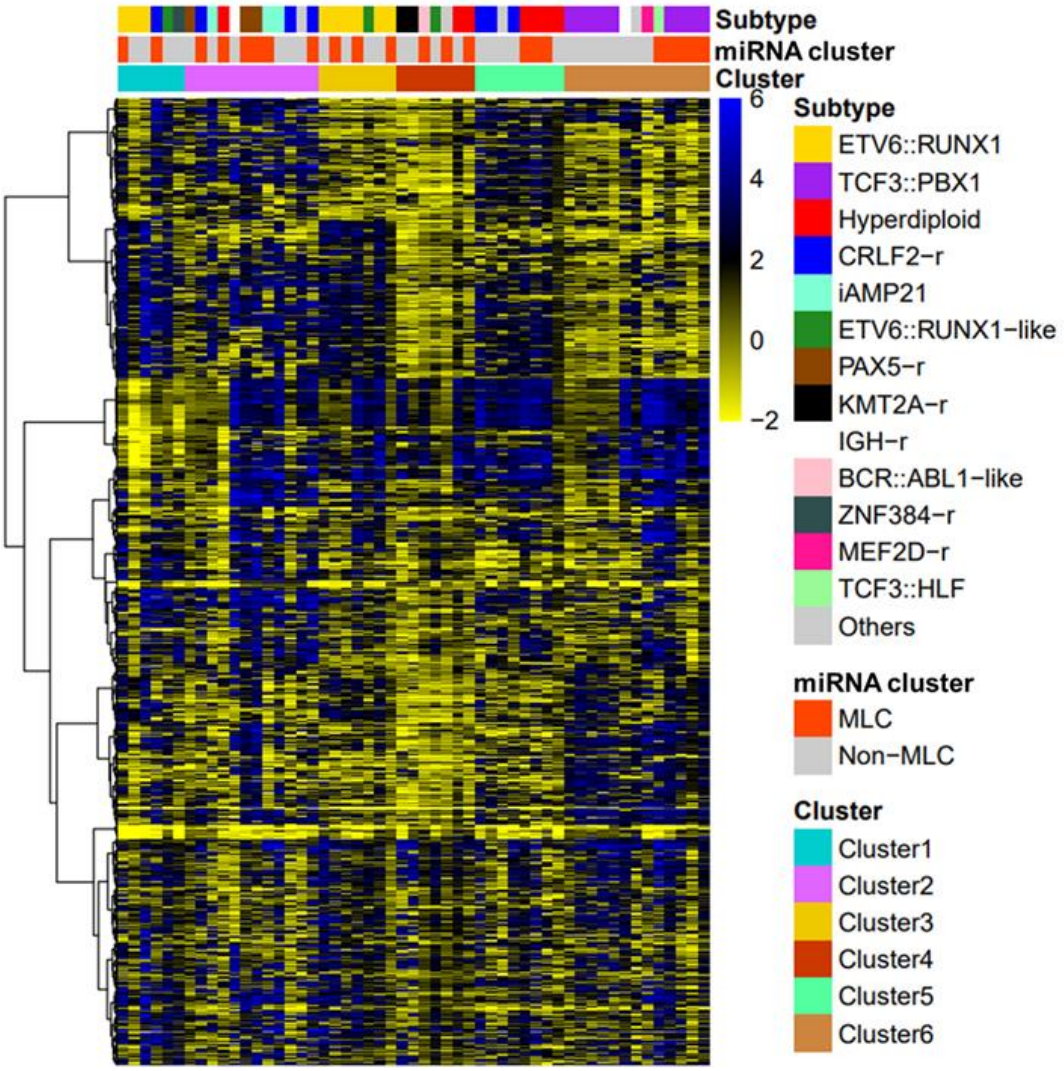
Co-occurrence or mutual exclusivity between genetic subtype and MLC signature pairs in 111 BCP-ALL primary samples from the TARGET cohort (A), and co-occurrence or mutual exclusivity between gene mutations and MLC-like signature pairs in 69 BCP-ALL primary samples from the Japanese cohort (B). Green indicates a tendency toward co-occurrence, and brown indicates exclusivity. The point indicates $P < 0.1$, and the asterisk indicates $P < 0.05$. (C) Recurrent driver mutations and copy number alterations identified by WES, based on clinical information for the 69 BCP-ALL primary samples from the Japanese cohort. Cases are grouped as MLC-like and non-MLC-like according to miRNA expression cluster. WBC, white blood cell count.

Supplementary Figure 7.



Copy number alterations in MLC and non-MLC of BCP-ALL cases.
Statistically significant copy number gains and losses detected by the GISTIC algorithm are shown for 105 BCP-ALL primary samples (31 MLC and 74 non-MLC) from TARGET cohort SNP array data. MLC in the top panel; non-MLC in the bottom panel.

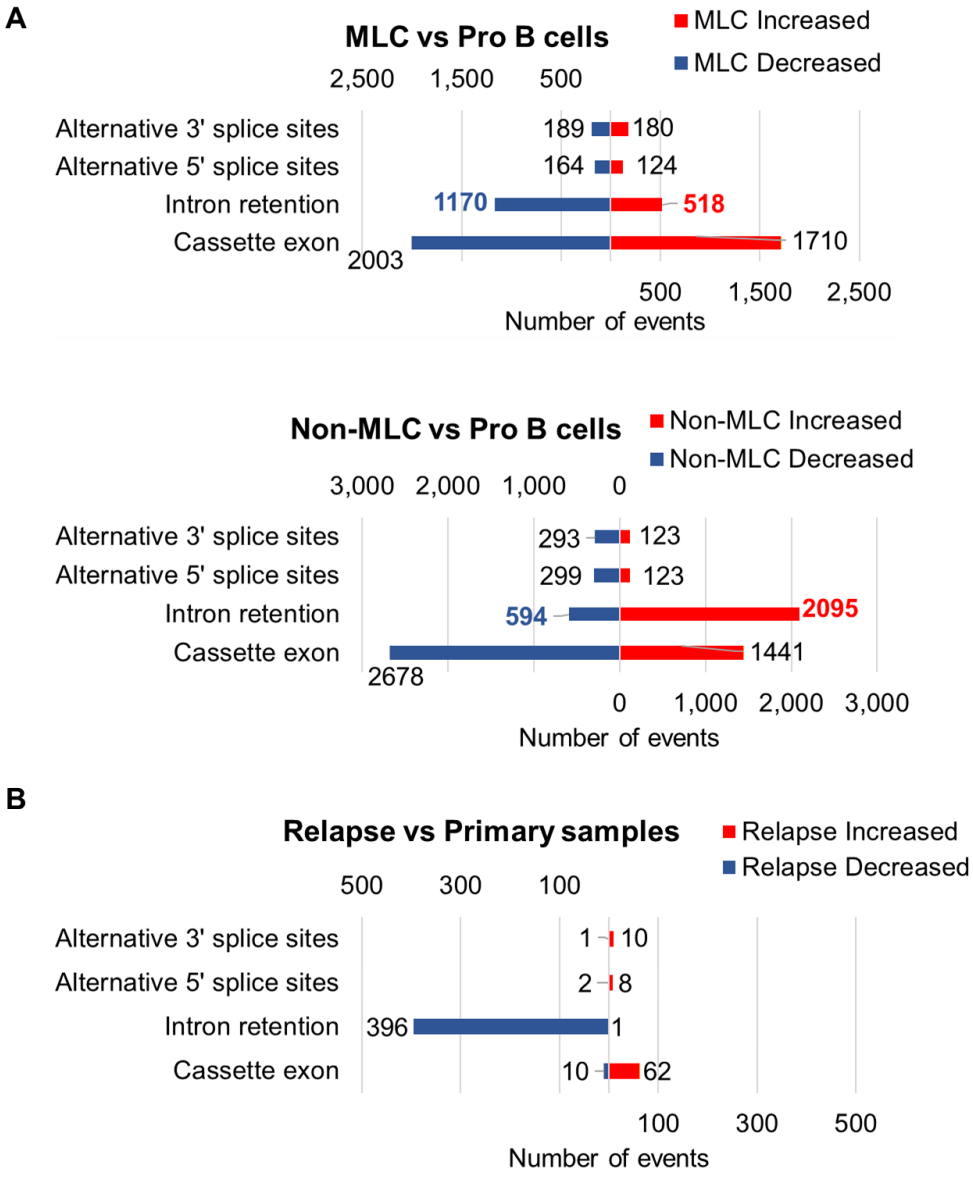
Supplementary Figure 8.



Hierarchical clustering of DNA methylation profiles for the MLC and non-MLC BCP-ALL cases.

The heatmap shows the DNA methylation profiles of 53 BCP-ALL cases (22 MLC and 31 non-MLC) based on unsupervised hierarchical clustering using the HpaII to MspI ratio from the HELP Assay along with Roche NimbleGen from the TARGET cohort. Six clusters were identified by consensus clustering with 1000 probes (distance method: Pearson).

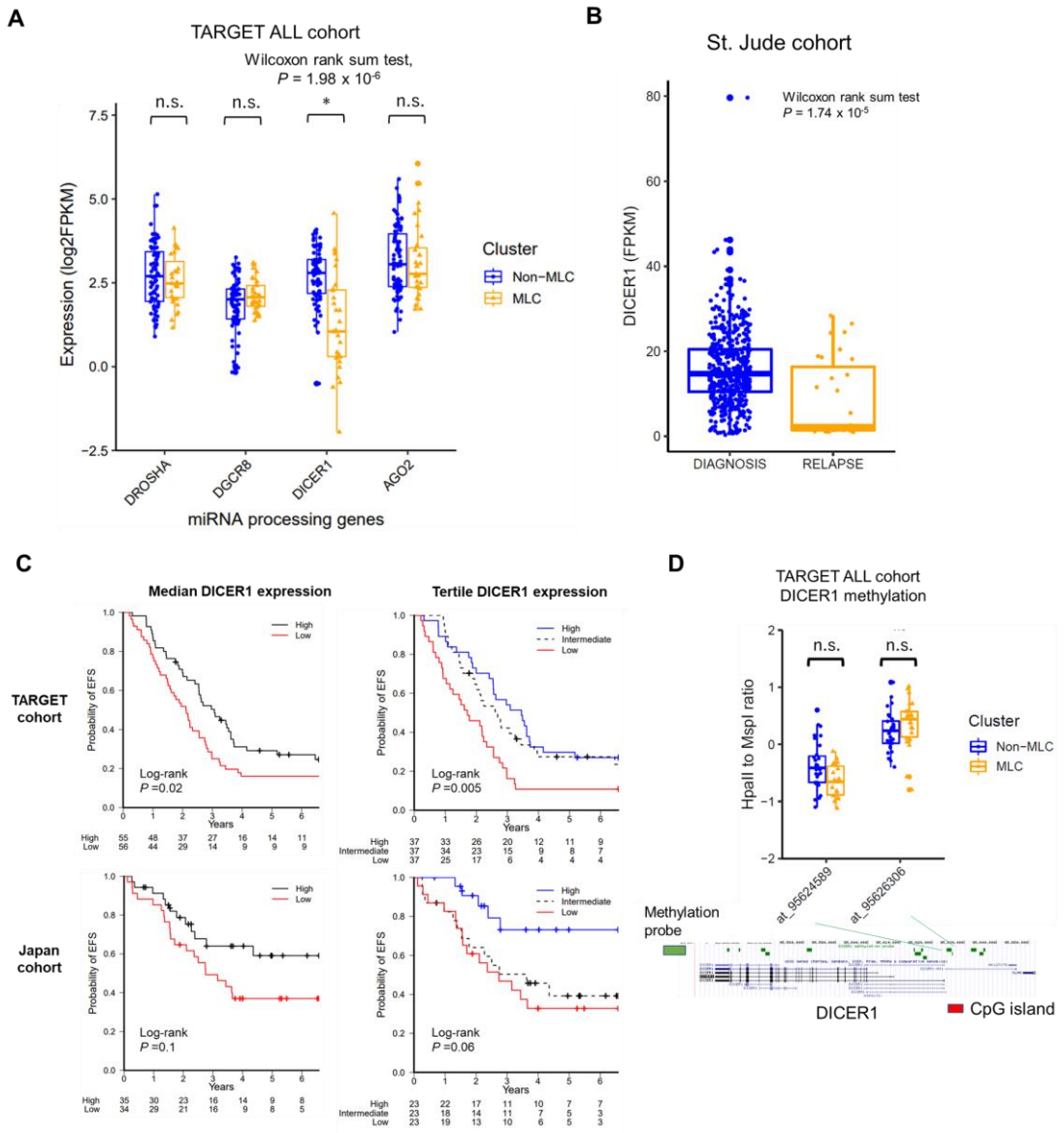
Supplementary Figure 9.



Alternative RNA splicing of MLC and relapsed cases.

(A) Type and number of splicing alterations significantly associated with MLC cases compared with normal pro-B cells (top), and with non-MLC cases compared with normal pro-B cells (bottom). (B) Type and number of splicing alterations significantly associated with relapsed samples compared with paired-primary samples from nine relapsed cases that acquired the MLC-like profile at relapse. Changes in the relative inclusion of local splice variations of at least 20% at the 95% confidence interval between two given conditions are shown. Bars on the right indicate alternative splicing events that were increased, and bars on the left indicate those that were decreased.

Supplementary Figure 10.



***DICER1* expression and methylation in BCP-ALL.**

(A) Box plots comparing miRNA processing genes (*DROSHA, DGCR8, DICER1* and *AGO2*) expression (expressed as log2 of fragments per kilobase million (FPKM)) by non-MLC cases (n = 80), and MLC cases (n = 31) from the TARGET ALL cohort. The P-value was calculated using the Wilcoxon rank-sum test. (B) Box plots comparing *DICER1* expression (expressed as FPKM) in primary samples (n = 501) and relapse samples (n = 31) from the B-ALL cohort (n = 532) from the St. Jude Cloud database. The *P* value was calculated using the Wilcoxon rank-sum test. (C) Kaplan–Meier survival curves of event-free survival (EFS) and overall survival (OS) for BCP-ALL

cases with high or low level of the *DICER1* expression in the TARGET ALL and Japanese cohorts. The cut-off for stratification was set to median or tertile for *DICER1* expression. (D) Box plot showing quantification of DNA methylation using the HpaII to MspI ratio from the HELP Assay, along with Roche NimbleGen of the CpG island probe targeting the *DICER1* promoter (at_95624589 and at_95626306) in non-MLC cases (n = 31) and MLC cases (n = 22) from the TARGET ALL cohort. The Wilcoxon rank-sum test was used to evaluate the differences. n.s., not significant.

Supplementary Table

Title: RNA-seq-based miRNA signature as an independent predictor of relapse in pediatric B-cell acute lymphoblastic leukemia

Authors: Hirohito Kubota, Hiroo Ueno, Keiji Tasaka, Tomoya Isobe, Satoshi Saida, Itaru Kato, Katsutsugu Umeda, Mitsuteru Hiwatari, Daiichiro Hasegawa, Toshihiko Imamura, Nobuyuki Kakiuchi, Yasuhito Nannya, Seishi Ogawa, Hidefumi Hiramatsu, Junko Takita

Here is the website where you can download the Supplementary table data.

https://doi.org/10.1182/bloodadvances.2023011583