DNA Methylation Signatures Predict Cytogenetic Subtype and Outcome in Pediatric Acute Myeloid Leukemia (AML)

Krali, Olga

DNA Methylation Signatures Predict Cytogenetic Subtype and Outcome in Pediatric Acute Myeloid Leukemia (AML)

Olga Krali [1,†], Josefine Palle [1,2,*,†], Christofer L. Bäcklin [3], Jonas Abrahamsson [4], Ulrika Norén-Nyström [5], Henrik Hasle [6], Kirsi Jahnukainen [7], Ólafur Gísli Jónsson [8], Randi Hovland [9], Birgitte Lausen [10], Rolf Larsson [3], Lars Palmqvist [11], Anna Staffas [11], Bernward Zeller [12] and Jessica Nordlund [1,*]

1   Department of Medical Sciences, Molecular Precision Medicine and Science for Life Laboratory, Uppsala University, 752 37 Uppsala, Sweden; Olga.Krali@medsci.uu.se
2   Department of Women's and Children's Health, Uppsala University, 752 37 Uppsala, Sweden
3   Department of Medical Sciences, Cancer Pharmacology and Computational Medicine, Uppsala University, 751 85 Uppsala, Sweden; christofer.backlin@gmail.com (C.L.B.); rolf.larsson@medsci.uu.se (R.L.)
4   Department of Pediatrics, Queen Silvia Children's Hospital, 416 85 Gothenburg, Sweden; vobjab@gmail.com
5   Department of Clinical Sciences, Pediatrics, Umeå University Hospital, 901 85 Umeå, Sweden; ulrika.noren-nystrom@umu.se
6   Department of Pediatrics, Aarhus University Hospital, DK-8200 Aarhus, Denmark; hasle@dadlnet.dk
7   Children's Hospital, Helsinki University Central Hospital, Helsinki, and University of Helsinki, 00290 Helsinki, Finland; kirsi.jahnukainen@ki.se
8   Department of Pediatrics, Landspitali University Hospital, 101 Reykjavík, Iceland; olafurgi@landspitali.is
9   Center of Medical Genetics and Molecular Medicine, Haukeland University Hospital, 5009 Bergen, Norway; randi.hovland@helse-bergen.no
10 Department of Pediatrics and Adolescent Medicine, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark; birgitte.lausen@rh.regionh.dk
11 Department of Clinical Chemistry and Transfusion Medicine, University of Gothenburg, 41346 Gothenburg, Sweden; lars.palmqvist@clinchem.gu.se (L.P.); anna.staffas@clinchem.gu.se (A.S.)
12 Division of Paediatric and Adolescent Medicine, Oslo University Hospital, 0450 Oslo, Norway; bzeller@ous-hf.no
*   Correspondence: Josefine.Palle@kbh.uu.se (J.P.); Jessica.Nordlund@medsci.uu.se (J.N.)
†   Equal contribution.

Abstract: Pediatric acute myeloid leukemia (AML) is a heterogeneous disease composed of clinically relevant subtypes defined by recurrent cytogenetic aberrations. The majority of the aberrations used in risk grouping for treatment decisions are extensively studied, but still a large proportion of pediatric AML patients remain cytogenetically undefined and would therefore benefit from additional molecular investigation. As aberrant epigenetic regulation has been widely observed during leukemogenesis, we hypothesized that DNA methylation signatures could be used to predict molecular subtypes and identify signatures with prognostic impact in AML. To study genome-wide DNA methylation, we analyzed 123 diagnostic and 19 relapse AML samples on Illumina 450k DNA methylation arrays. We designed and validated DNA methylation-based classifiers for AML cytogenetic subtype, resulting in an overall test accuracy of 91%. Furthermore, we identified methylation signatures associated with outcome in t(8;21)/*RUNX1-RUNX1T1*, normal karyotype, and *MLL/KMT2A*-rearranged subgroups ($p < 0.01$). Overall, these results further underscore the clinical value of DNA methylation analysis in AML.

Keywords: DNA methylation; pediatric AML; acute myeloid leukemia; classification; 450k array; epigenetics; subtyping

1. Introduction

Acute myeloid leukemia (AML) in children is a heterogeneous disease with variable clinical outcome. In pediatric as well as adult AML, different prognostic subtypes are

characterized either by recurrent cytogenetic aberrations or a cytogenetically normal karyotype (NK) [1]. However, as many as 44% of pediatric AML cases remain cytogenetically unclassified, resulting in a large highly heterogeneous group with not otherwise specified (NOS) cytogenetic classification [2]. Not only does molecular subtype provide prognostic information, but it is important prior knowledge for targeted assays that measure the response to induction treatment measured by minimal residual disease [3,4]. In addition, cytogenetic analysis not only yields important information on diagnosis, prognosis, and follow-up, it provides information about the initiating events in the process of leukemogenesis [5]. Furthermore, NK as determined by conventional cytogenetic analysis may conceal diagnostically relevant rearrangements [6]. Accurate subtyping is therefore exceedingly important for diagnostics as well as future exploratory research [7].

In addition to cytogenetic aberrations, DNA methylation may reflect underlying biological processes that are relevant for studying clinical outcomes. Aberrant DNA methylation has been widely observed in correlation with cytogenetic subtypes of AML, however most studies have been performed predominantly in adults [8–13], where the molecular landscapes and clinical outcomes are different from those observed in children [14,15]. However, the strong evidence for a correlation between epigenetic factors, molecular cytogenetic subtype, and outcome in adult AML renders DNA methylation an interesting molecular marker for prognostication also in pediatric AML. A handful of studies have examined DNA methylation in depth in pediatric AML in general [16,17] as well as within MLL/*KMT2A*-rearranged [18], inv(16)/*CBFB-MYH11* [19], and t(8;21)/*RUNX1-RUNX1T1* [20,21] subtypes. Several other studies have identified putative signatures predictive of outcome across subtypes [22], as well as within specific AML subtypes [19,21]. Taken together, these studies provide compelling evidence that DNA methylation captures information about underlying the various biological processes taking place within this heterogeneous disease.

In the present study, we examined 125 Nordic pediatric patients with AML genome-wide DNA methylation analysis and integrated the DNA methylation profiles with clinical data in order to investigate whether variable DNA methylation patterns hold diagnostic or prognostic information in pediatric AML.

2. Materials and Methods

2.1. DNA Samples

Bone marrow (BM) or peripheral blood (PB) from 142 samples taken at diagnosis (*n* = 123) or relapse (*n* = 19) collected from 125 unique AML patients diagnosed at Nordic Pediatric Centers between 1997–2008 were analyzed in the study (Table S1). The patients were treated according to NOPHO93 and NOPHO2004 protocols [23,24]. Diagnosis was determined by utilizing BM aspirates for morphology and immunophenotype, while G-band karyotyping, fluorescence in situ hybridization (FISH), and reverse transcription polymerase chain reaction (RT-PCR) was used for cytogenetic characterization. Leukemic cells isolated from AML patient samples were centrifuged by Ficoll-isopaque (Pharmacia, Uppsala, Sweden), pelleted and frozen in liquid nitrogen. All samples selected for the study contained ≥70% blasts as evaluated by microscopy. DNA was extracted from pelleted cells with the AllPrep DNA/RNA Mini Kit according to manufacturer's instructions (Qiagen, Hilden, Germany).

2.2. DNA Methylation Assay

First, 500 ng of genomic DNA was treated with sodium bisulfite according to the manufacturer's specifications (EZ DNA methylation Gold, Zymo Research, Irvine, CA, USA). Then, 200 ng of the converted DNA was loaded on an Infinium HumanMethylation 450k BeadChip Assay and processed according to the manufacturer's instructions (Illumina, San Diego, CA, USA).

2.3. Data Preprocessing

Data analysis was performed in Python [25], using SciPy [26] in the Jupyter Notebook environment [27]. All scripts are available on GitHub (https://github.com/Molmed/Krali-Palle_2021, accessed on 4 June 2021). The raw IDAT files were processed using methylprep (https://pypi.org/project/methylprep/, accessed on 13 January 2021) to produce a β-value matrix, where $\beta = 0$ corresponds to no methylation, $\beta = 1$ complete methylation of a given CpG site. β-values that failed to pass the p-value cutoff (<0.05) were replaced with NaNs. Normal-exponential convolution using out-of-band probes used for background correction was used to normalize the β-value distribution of Infinium type I and II probes. Probes were filtered using an empirical method previously described [28], and further filtered to remove CpG sites with >10% missing values, resulting in a total of 406,830 autosomal CpG sites for downstream analyses. Multidimensional scaling (MDS) was performed for outlier detection resulting in high quality data from 142 DNA samples for analysis. No batch effects were identified among the different batches, while running batch effect correction with Combat [29], thus the uncorrected data were used.

2.4. Subtype Prediction with DNA Methylation

Subtype classification was performed using diagnostic samples from 99 patients with the following cytogenetic subtypes: t(8;21)/*RUNX1-RUNX1T1*, inv(16)/*CBFB-MYH11*, t(15;17)*PML-RARA*, *MLL/KMT2A*-rearranged, NK, mono 7, sole +8, and 3q21q26. Samples with 11q23/*MLL/KMT2A*, t(9;11), t(10;11), t(11;19) were merged into one *MLL/KMT2A*-rearranged group. The samples with undefined cytogenetic subtype (no result and other clonal abnormalities) were excluded from the classifier design. The dataset was split into a training set used to train the classifiers (66%, N = 66) and a test set used to evaluate the classifier performance (33%, N = 33). For DNA methylation classification, we tested eight different commonly used machine learning algorithms with Scikit-learn [30] and one neural network with Keras (https://github.com/fchollet/keras, accessed on 20 February 2021). Nested k-fold cross-validation was used to identify the best hyperparameters per classification model (Figure S1) and to select the best model in terms of accuracy score on the validation dataset. During k-fold cross-validation, k-1 parts were used as a training set and the remaining part as a validation set. Each classifier was trained using an outer 5-fold cross-validation to obtain the mean accuracy per optimized classifier and 3-fold cross validation was run in the inner loop to identify the best set of hyperparameters per classifier. In each iteration, any missing DNA methylation values were imputed by using the median. A one-vs.-rest multiclass strategy was implemented where each class (subtype) was trained against all the other classes. A confusion matrix and a classification report with precision, recall, and f1-score were generated from the predictions.

Feature selection (FS) was performed prior to modeling to capture the CpG sites that represent the most variability in the dataset using the combination of two unsupervised approaches. For FS, a 5-fold cross-validation approach was followed and all of the selected CpGs per fold from both methods were retained. First, a PCA-approach was implemented where a maximum of 20 most informative CpG sites (per fold) per component from the first 15 principal components were selected, resulting in 955 CpG sites. In a second approach, CpG sites with low variance and high correlation (LVHC) were removed. LVHC thresholds of <0.1 data variance and >70% correlation were set, resulting in 456 CpG sites. The union of the two FS approaches resulted in 1300 CpG sites for downstream analysis.

The Wilcoxon signed-rank test was used to compare the performance of the models. Finally, a permutation test ($n = 1000$ permutations) was performed to validate that the model performed well due to the feature and target dependency and not due to stochastic noise. To identify differentially methylated CpG sites in the different subtypes, Mann–Whitney U tests were applied. Benjamini–Hochberg multiple testing correction was applied to p-values.

2.5. Survival Analysis

Patients were grouped within subtype based on hierarchical clustering of selected CpG sites. Relapse free survival (RFS) and overall survival (OS) where event was defined as death from any cause was used to evaluate differences in outcome. Survival plots were generated with the Kaplan–Meier method (https://github.com/erdogant/kaplanmeier, accessed on 19 March 2021). The log-rank test and Cox regression analysis were used to analyze the differences between DNA methylation groups with respect to RFS and OS.

2.6. Data Visualization

Matplotlib [31] and Seaborn [32] were used for all visualizations. Heatmaps with annotation bars ordered by hierarchical clustering were created by a custom function AnnotHeatMap (https://github.com/Molmed/Krali-Palle_2021, accessed on 4 June 2021).

3. Results

3.1. Overview of the DNA Methylome in Pediatric AML

A total of 142 DNA samples from 125 unique pediatric AML patients were analyzed on Illumina 450k DNA methylation arrays (Table S1). The samples included 123 DNA samples taken at AML diagnosis and 19 samples taken at relapse. Table 1 summarizes the cytogenetic subtypes and French-American-British (FAB) classification of the AML patients included in the study. To obtain an initial view of the variation in CpG site methylation in our dataset, we subjected the complete set of methylation data (142 samples across 406,830 autosomal CpG sites) in 2D space with the help of Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) [33], which resulted in clustering of samples by cytogenetic subtype (Figure S2). A heatmap and hierarchical clustering of all samples using the 10,000 most variable CpG sites are shown in Figure S3. Because the samples with any t(11q23) clustered together, we merged them together in one *MLL/KMT2A*-rearranged subgroup for subsequent analyses.

Table 1. Cytogenetic subtype, age range, central nervous system (CNS) involvement, stem cell transplantation (SCT), and FAB classification for the 142 pediatric AML patient samples. The number of samples taken at AML diagnosis and relapse are presented. When known, the subtype of the relapse samples is indicated as the same subtype as their diagnostic pair.

Cytogenetic Subtype	Normal Karyotype (NK)	MLL/KMT2A Rearranged	Undefined	t(8;21) RUNX1/RUNX1T1	inv(16) CBFB/MYH11	mono 7	t(15;17) PML-RARA	sole +8	3q21q26
Number of diagnostic Samples	30	25	24	19	12	5	4	3	1
Number of relapse samples	8	1	5	4	1	-	-	-	-
Age range	1–18	0–16	0–17	2–16	1–17	1–5	3–16	3–14	14
CNS involvement	-	-	-	-	-	-	-	-	-
Yes	4	1	2	6	-	-	-	-	-
No	34	25	27	17	12	4	3	3	1
Missing	-	-	-	-	1	1	1	-	-
SCT	-	-	-	-	-	-	-	-	-
Yes	8	3	1	-	-	5	-	2	-
No	30	23	27	23	13	-	3	1	1
Missing	-	-	1	-	-	-	1	-	-
FAB	-	-	-	-	-	-	-	-	-
M0	2	-	-	1	-	2	-	1	-
M1	7	-	3	1	-	1	-	1	1
M2	14	1	3	19	3	1	-	-	-
M3	1	-	1	-	-	-	4	-	-
M4	10	1	10	2	9	-	-	-	-
M5	-	24	4	-	-	1	-	1	-
M6	3	-	2	-	-	-	-	-	-
M7	-	-	4	-	-	-	-	-	-
Missing	1	-	2	-	1	-	-	-	-

After performing FS, the remaining 1300 CpG sites resulted in clearly defined clusters separating the known AML subtypes, which suggests that our FS strategy retained significant information while reducing noise (Figure 1). As expected, patients belonging to sub-

types defined by *MLL/KMT2A*-rearrangments, t(8;21)*RUNX1/RUNX1T1*, inv(16)*CBFB/MYH11*, mono 7, and t(15;17)*PML-RARA* clustered by their known cytogenetic subtype, indicating a clear relationship between cytogenetic subgroup and DNA methylation. The patients with cytogenetic "normal karyotype" (NK) formed two loosely defined clusters. The large proportion of the cytogenetically undefined samples (no result from diagnostic cytogenetic workup) in this cohort demonstrate DNA methylation similarities with the known subtypes, and not of new independent subtypes.
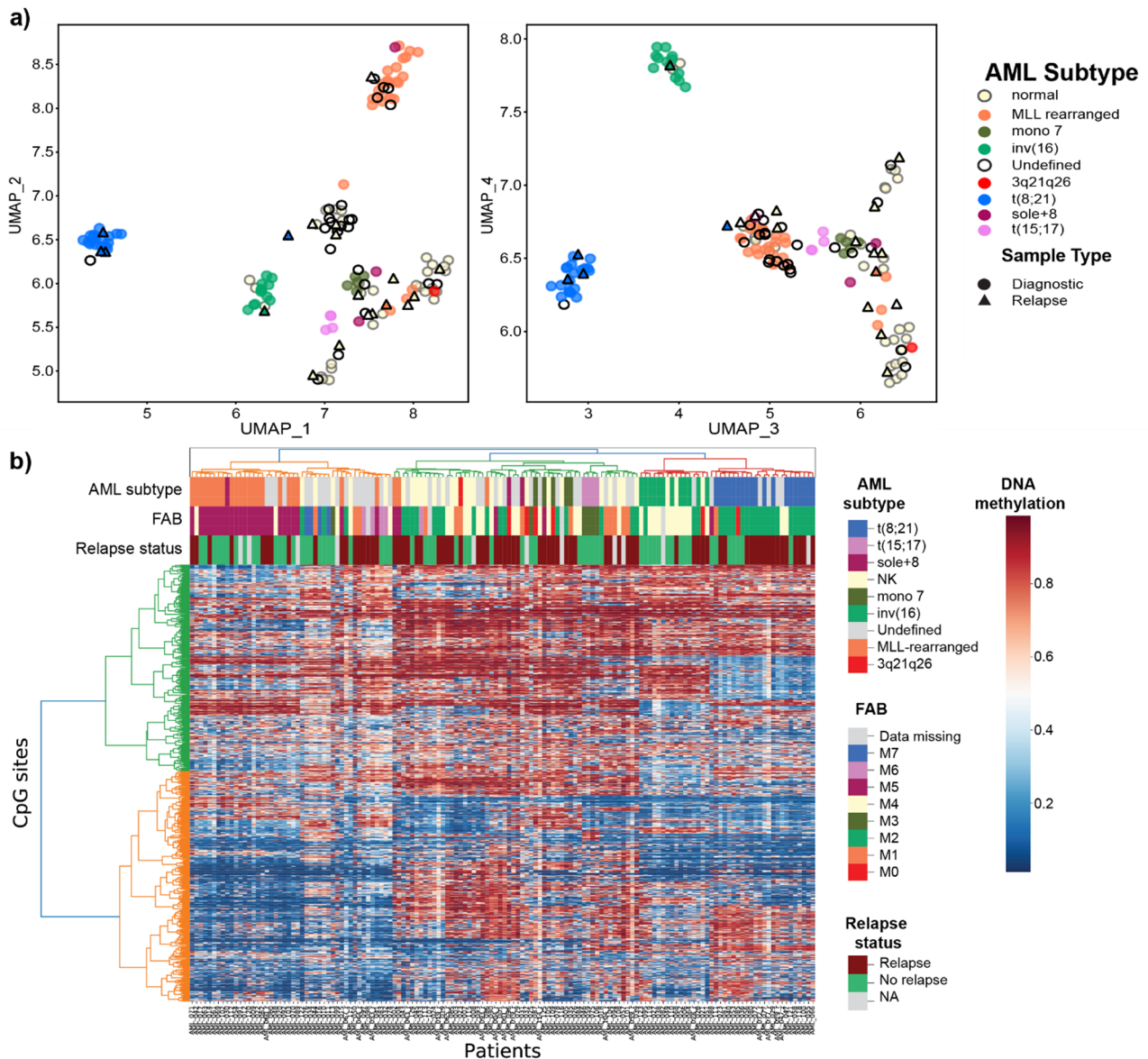
Figure 1. Data representation in space for the 142 pediatric AML patient samples for the selected 1300 CpG sites. (**a**) DNA methylation data projection in 2D space in a UMAP plot for UMAPs 1–2 (left) and 3–4 (right). Each point represents a patient sample. All samples are labeled by cytogenetic subtype, samples with unknown subtype (undefined) are shown as black circles with no fill (N = 24, diagnostic samples), and samples taken at relapse are shown by filled triangles colored by cytogenetic subtype (N = 19, relapse samples). (**b**) DNA methylation heatmap ordered by hierarchical clustering. Each row in the heatmap denotes a CpG site and each column is a patient. Cytogenetic subtype, FAB classification, and relapse status are shown as annotated bars over the plot.

3.2. Prediction of Cytogenetic Subtype with DNA Methylation

To address subtype membership of the cytogenetically undefined samples in our cohort, we built cytogenetic subtype classifiers for eight of the known cytogenetic subtypes present in our cohort using diagnostic AML samples. In total, nine different classification methods were tested and the best performing model was the neural network with mean validation accuracy of 85%, followed by nearest centroid (82%) and random forest (80%) (Figure S4, Table S2). As the neural network model performed best in terms of accuracy, we proceeded to fit the classifier on all 66 training samples and tested the performance on the remaining 33 unseen patient samples (test set). Finally, the classifier was fitted on all 99 samples of known subtype and used to predict the cytogenetic subtype of the 43 unknown samples. The classifier predicted the correct subtype in 91% (30/33) of test samples, indicating that cytogenetic subtype is highly correlated with DNA methylation levels (permuted *p*-value = 0.001; Figure S5). Our resulting DNA methylation classifier correctly predicted samples of cytogenetic subtypes *MLL/KMT2A*-rearranged, inv(16)/*CBFB-MYH11*, and t(8;21)/*RUNX1-RUNX1T1* with the highest precision and recall scores, followed by NK (Figure 2a; Table 2). The small number of samples ($n \leq 5$) precludes further evaluation of the t(15;17)/*PML-RARA*, sole +8, mono 7, and 3q21q26 classifiers, although our data provide evidence that these subtypes can be captured by DNA methylation given enough data for training (Table 2). To provide additional validation for our subtype classifier, we predicted the subtype of 14 of the matched samples taken at AML relapse where the diagnostic sample was *MLL/KMT2A*-rearranged, t(8;21)/*RUNX1-RUNX1T1*, NK or inv(16)/*CBFB-MYH11*. The relapse samples are expected to maintain the same subtype as at diagnosis, and were not included in the classifier design (Figure S6). In total, 11 of the 14 predicted subtypes matched with the diagnostic subtype, adding an extra layer of validation to evaluate the predictive performance of the classifier (Table S3).
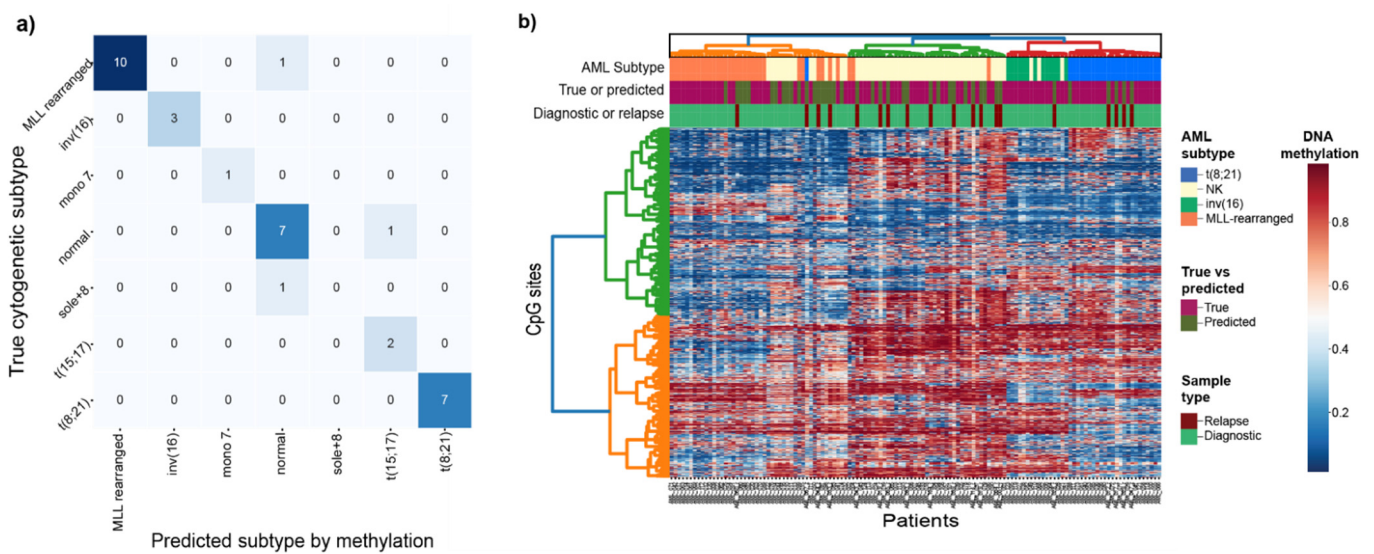
Figure 2. (**a**) Confusion matrix for the cytogenetically defined (diagnostic-known) vs. DNA methylation predicted cytogenetic subtypes (N = 33). (**b**) Heatmap and hierarchical clustering of 127 samples taken at AML diagnosis or relapse from patients belonging to the four subtypes most common subtypes; t(8;21)/*RUNX1-RUNX1T1*, inv(16)/*CBFB-MYH11*, *MLL/KMT2A*-rearranged and NK. The methylation status of the selected 1300 CpG sites are plotted in the heatmap. The samples are in columns labeled by AML subtype, whether the sample was cytogenetically defined (true/diagnostic-known) or cytogenetically undefined (DNA methylation predicted), and if the sample was taken at AML diagnosis or relapse.

Table 2. Precision, recall, F1 scores per subtype, as well as overall accuracy of the classifier for the four subtypes with >5 patients in the group. High precision scores indicate the low number of false positives (FP) while high recall scores the low number of false negatives (FN).

|	Precision	Recall	F1 Score	Total samples Test Set	Total Samples Train Set
MLL/KMT2A-rearranged	1	0.91	0.95	11	14
inv(16)/*CBFB-MYH11*	1	1	1	3	9
Normal Karyotype (NK)	0.78	0.88	0.82	8	22
t(8;21)/*RUNX1-RUNX1T1*	1	1	1	7	12

The cytogenetic subtypes of the remaining 29 cytogenetically undefined diagnostic samples were predicted by the classifier. In total, 18 were predicted as NK, seven as *MLL*-rearranged, two as t(8;21)/*RUNX1-RUNX1T1*, and two as mono7. The methylation profiles of the newly classified samples closely matched those of the group of original samples used to design the classifier and are referred to as "subtype-suspected". The heatmap and hierarchical clustering of 127 diagnostic and samples taken at relapse for both characterized and predicted subtypes belonging to the four groups show that most of the data are clustered by subtype driven by differential DNA methylation levels for the 1300 selected CpGs (Figure 2b).

For the most part, the samples and classes performed neat homogenous clusters based on the DNA methylation levels. One notable deviation is the split of the NK class into two clusters (Figure 2b). Additionally, the three cytogenetically defined *MLL*-rearranged samples (AML002, AML004, AML086) that deviated in the UMAP clustering, also cluster together with the large green cluster of NK samples and six more including the fourth AML sample that deviated (AML015, AML33_r, AML041, AML044_r, AML097, and AML099) on the other NK smaller cluster (orange). The NK samples (AML021, AML049, and AML068) cluster together with the inv(16)/*CBFB-MYH11* samples. The high subtype probability scores from the classifier (Table S1) strongly indicate that at least based on DNA methylation evidence, that these samples either were incorrectly classified at AML diagnosis or that these patients have other underlying molecular events driving their deviating DNA methylation profile.

3.3. Intra-Subtype Heterogenetity in NK-AML

As observed in Figures 1 and 2b, the NK group split into two distinctive groups based on DNA methylation levels. To further investigate intra-group heterogeneity of NK-AML, we performed unsupervised hierarchical clustering of all of the diagnostically defined NK samples (N = 30) as well as DNA methylation predicted NK samples (N = 15) using the previously defined 1300 CpG sites. Two well-defined clusters denoted as Cluster A and Cluster B emerged (Figure 3). We compared key molecular and clinical features available for this dataset, such as mutational status of genes of clinical interest (*CEBPA*, *WT1*, *FLT3*, and *NPM1*) from previous work [34] as well as features such as age, sex, and clinical outcome between the two cluster groups (Table S1). Collectively, *CEBPA*, *FLT3*, *NMP1*, and *WT1* mutations were enriched in Cluster B, with significantly higher proportion of patients in Cluster B (68%) displaying one or more mutations in comparison to only 18% of the A group (Fisher's exact *p*-value 0.002). Although no significant difference was observed when each gene was analyzed independently (Table S4), most *CEBPA* mutated patients clustered together in a sub-cluster of Cluster B, suggesting a common underlying epigenetic signature associated with this mutation. Of the clinical features analyzed, the patients in Cluster B were significantly older at AML diagnosis (mean 10.8 years) than those in Cluster A (mean 4.5 years, *t*-test *p*-value 0.00004).
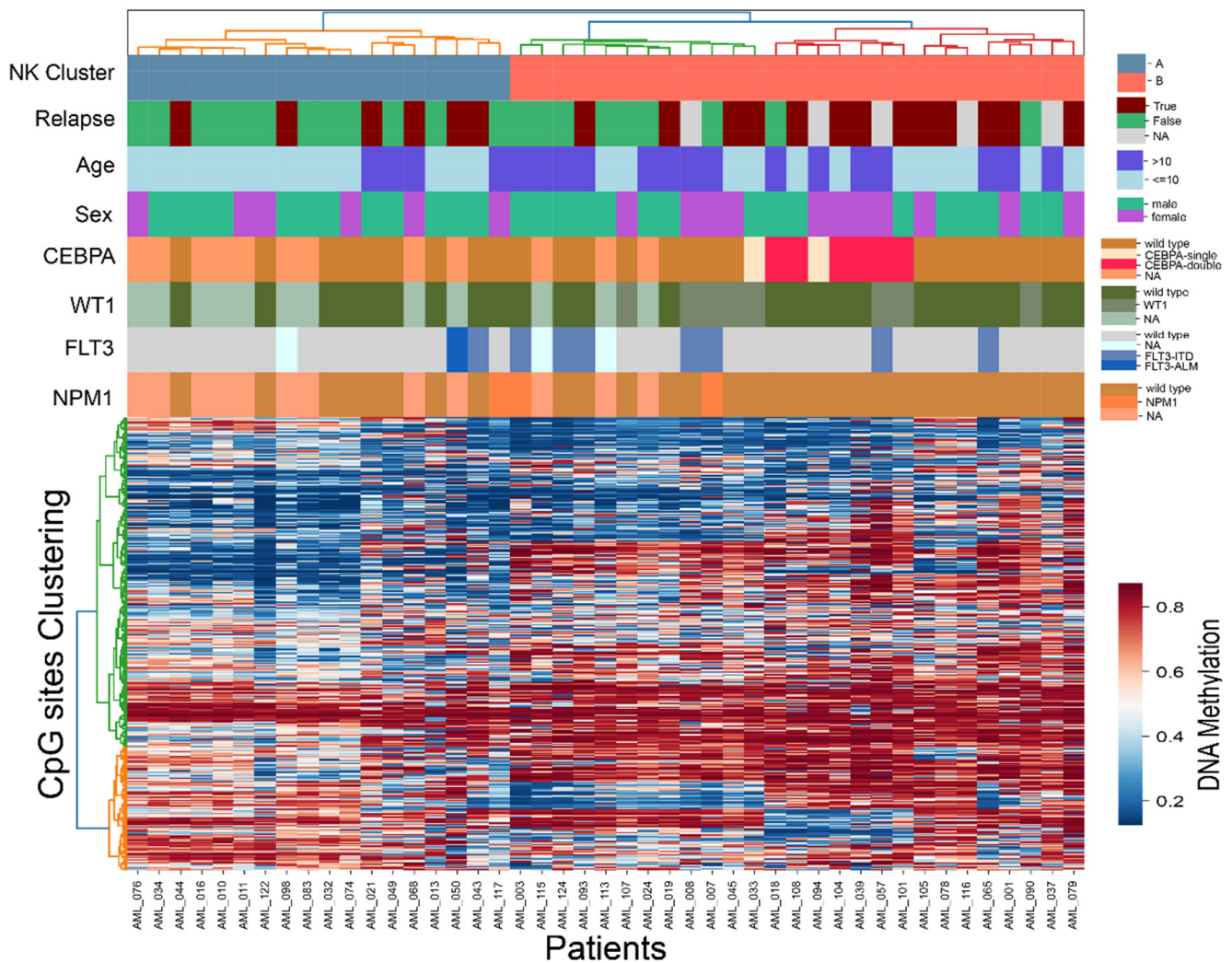
Figure 3. Heatmap and hierarchical clustering of the 45 NK samples across 1300 CpG sites, including 30 cytogenetically defined (diagnostic-known) patients and 15 DNA methylation predicted NK patients. Each row in the heatmap denotes a CpG site and each column is a patient. The labels at the top of the heatmap represent clinical and molecular features of interest, including the Cluster denotation A and B.

3.4. Cytogenetic-Specific DNA Methylation Signatures

The CpG sites important for each cytogenetic subtype were determined by analyzing the 1300 CpG sites in a one-vs.-rest approach. Information about the CpG sites (adjusted *p*-value < 0.05) that separated each subtype from all the rest and the CpG sites that were unique to each subtype are presented in Table 3. More detailed information about the 105 unique CpG sites obtained after CpG annotation to the human genome can be found in Table S5. We were unable to detect subtype-specific DNA methylation for the Sole + 8 and 3q21q26 subtypes due to low numbers of samples. Amongst the genes identified, include those genes previously known to be differentially expressed or methylated in a subtype-specific pattern such *ASB2* [35] and *PLAUR* [36] in *MLL/KMT2A*-rearranged and *DPF3* in inv(16)/*CBFB-MYH11* [17].

Table 3. Top selected CpGs and genes per subtype based on a one-vs.-rest CpG Site Selection with adjusted *p*-value threshold 0.05.

| Subtype	N CpG Sites (Adjusted *p* Value < 0.05)	N CpG Sites Unique to Subtype (N Genes)	Gene Names (CpG IDs) Unique to Subtype
Normal Karyotype	569	6 (5)	*ZNF793* (cg15139588), *APBA2* (cg15605858), *PRDM16* (cg02390319), *PPP1R14A* (cg02571816), *ACCN1* (cg03745383)
MLL/KMT2A-rearranged	873	59 (33)	*KIAA1755* (cg14003035), *PLAUR* (cg27340480), *PER3* (cg05803631), *ASB2* (cg09341793), *KLK4* (cg26827876), *BARHL2* (cg18322569), *L1TD1* (cg23049458), *ARPC1B* (cg10428938), *ST8SIA6* (cg17256364), *NKX6-2* (cg11174855), *WNT5A* (cg19554389), *HOXA5* (cg12128839, cg25307665), *NFIX* (cg06744585), *SNED1* (cg25241559, cg09991306), *TNXB* (cg12694372, cg10923662, cg16834823, cg01992382), *MSX2* (cg06013117), *MAPK8IP1* (cg08214808), *BNIP3* (cg18477674), *CASR* (cg19108881), *HECW1* (cg24384918), *PCDHA1; PCDHA2; PCDHA3; PCDHA4; PCDHA5; PCDHA6; PCDHA7; PCDHA8* (cg19596110), *PITX1* (cg00396667), *KCNN1* (cg07857792), *TMEM132D* (cg20168964), *NPSR1* (cg20276677), *LOC732275* (cg16709904), *NOM1* (cg02413092), *SPEG* (cg16440561), *EDARADD* (cg09164898), *THBS4* (cg26286839), *HOOK2* (cg06417478), *LOC254559* (cg09969277), *DCC* (cg25204852)
mono7	330	24 (12)	*BCL2* (cg25059899), *ZNF577* (cg03562414, cg24794228, cg11269599, cg10635122), *ZNF154* (cg21790626, cg27049766, cg26465391), *ARRB2* (cg07971820, cg02286380), *SKI* (cg25139649), *ERCC3* (cg06373940), *RPTOR* (cg09929238), *FBXO47* (cg04120272), *DLL1* (cg00084338), *C1orf86; LOC100128003* (cg26227225), *CYP1A1* (cg22549041), *PLD6* (cg24578857, cg19093370)
inv(16)/*CBFB-MYH11*	571	22 (15)	*DPF3* (cg13588403), *IFLTD1* (cg13134916), *BAHCC1* (cg06636541), *LEPR* (cg16987305), *AFAP1* (cg22079161), *PRHOXNB* (cg20101529), *ANK1* (cg19537719), *SHISA6* (cg13330559), *PRDM16* (cg03337482), *MUC4* (cg05834845), *C22orf34* (cg20744362), *LY96* (cg23732024), *ZNF423* (cg26929700, cg04086531), *GNG7* (cg26988138), *CNTD2* (cg08871608)

Table 3. *Cont.*

Subtype	N CpG Sites (Adjusted *p* Value < 0.05)	N CpG Sites Unique to Subtype (N Genes)	Gene Names (CpG IDs) Unique to Subtype
t(8;21)/*RUNX1-RUNX1T1*	723	27 (15)	*SMTNL2* (cg13375589), *TUSC1* (cg13811417), *PDLIM3* (cg14632696), *CYP27C1* (cg08022717), *PCDHA1; PCDHA10; PCDHA11; PCDHA2; PCDHA3; PCDHA4; PCDHA5; PCDHA6; PCDHA7; PCDHA8; PCDHA9* (cg26514430), *RYR2* (cg07790615), *TACSTD2* (cg13443627), *FBXL7* (cg26134895), *VSTM2A* (cg19868631), *MARCH11* (cg25092681, cg01791874,cg00339556, cg16150752, cg17712694), *IGSF21* (cg15564444), *TTBK1* (cg16620382), *SHROOM1* (cg21811204), *ELOVL4* (cg04107099), *SETDB1* (cg15448220)
t(15;17)/*PML-RARA*	328	8 (8)	*NFYC* (cg16167741), *C7orf50* (cg23657099), *IGDCC4* (cg00776960), *SETD7* (cg02409722), *SCHIP1* (cg23553912), *SYNE1* (cg02796568), *WBSCR17* (cg02300154), *C21orf7* (cg08854834)

3.5. Putatively Prognostic DNA Methylation Signatures

Next, we investigated if the 1300 CpGs hold prognostic information. To avoid confounding results due to the subtype-specific differences in this dataset, we performed survival analysis within each of the t(8;21)/*RUNX1-RUNX1T1*, *MLL/KMT2A*-rearranged, and NK subtypes, where there was an adequate number of patients for analysis. First, we analyzed the 1300 CpG sites for differences between patients with t(8;21)/*RUNX1-RUNX1T1* (N = 18), *MLL/KMT2A*-rearranged (N = 19) or NK (N = 26) subtypes who experienced a relapse in comparison to those who did not (top 50 CpG sites relapse vs. no relapse; *p*-value < 0.05). The three *MLL/KMT2A*-rearranged samples (AML002, AML004, AML086) that grouped together with the large NK cluster (Figures 1 and 2) were excluded from the survival analysis. We used a hierarchical clustering approach to split patients into groups based on DNA methylation patterns of the 50 top CpG sites (Figure 4, Table S6) and investigated clinical outcomes in each group of patients. In a second stage, we included all of the previously undefined diagnostic samples predicted by the DNA methylation classifier to each respective group, leading to the following sample sizes: t(8;21)/*RUNX1-RUNX1T1* (N = 19), *MLL/KMT2A*-rearranged (N = 24), and NK (N = 40), and repeated the survival analyses (Figure 4).

We identified variability in DNA methylation in the within-subtype groups of patients determined by diagnostic karyotyping and obtained two clusters of high and low methylation levels (Figure 4, heatmaps). Survival analysis on the groups of patients with subtypes known at AML diagnosis (Figure 4a,c,e right) revealed differences for all three subtypes (logrank *p*-value < 0.5) in RFS and OS while comparing the low and high methylation groups. Similar results were observed when the predicted undefined diagnostic samples were added in their corresponding subtype groups. Overall, the survival analysis revealed a relationship between patient outcome (relapse or death) and high/low methylation levels, in each of the three subtype groups (Figure 4). We included the DNA methylation group, treatment protocol, and risk group in a Cox proportional hazards regression analysis for RFS and OS. Treatment protocol had no systematic significant impact on outcome for NK or *MLL/KMT2A*-rearranged patients (Figures S7 and S8). The impact on risk grouping could only be evaluated in the patients with *MLL/KMT2A*-rearranged subtype treated on

the NOPHO-2004 protocol, also showed no difference between standard and high risk (Figure S9). The t(8;21)*RUNX1-RUNX1T1* group was not analyzed in the Cox regression because the majority (16/19) of patients were treated on the NOPHO-2004 protocol and all were treated as standard risk.
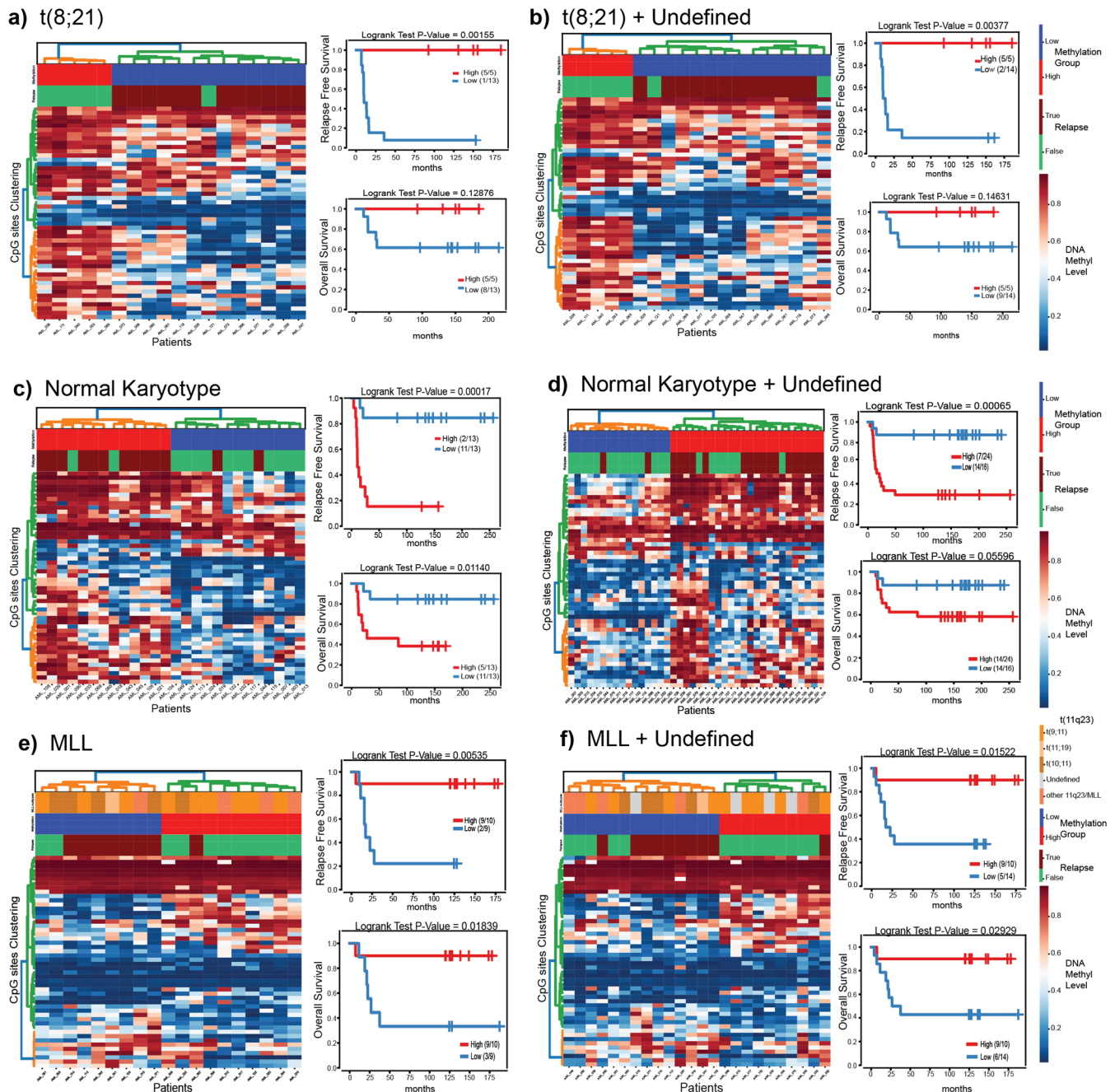
Figure 4. Survival analysis for t(8;21)/*RUNX1-RUNX1T1* (**a**,**b**), NK (**c**,**d**), and *MLL/KMT2A*-rearranged (**e**,**f**) patient samples based on the 50 most significant CpGs that separate the diagnostic samples of patients who later went on to relapse from those who did not. Panels (**a**,**c**,**e**) contain patients with a confirmed diagnostic subtype (diagnostic-known). Panels (**b**,**d**,**f**) contain diagnostic-known in addition DNA methylation predicted samples from the undefined group. In each panel, the heatmap (left), is ordered by hierarchical clustering. Samples along the *x*-axis are split into two groups color coded by high (red) or low (blue) overall methylation level. The numbers in the legend represent the fraction of patients who did not experience an event. Kaplan–Meier curve for relapse free survival analysis (upper right) and overall survival (lower right) of the two groups identified by clustering. The *x*-axis represents the time until event (months) and events are plotted on the *y*-axis.

4. Discussion

Herein, we investigated CpG site methylation in a relatively large cohort of Nordic pediatric AML patients and integrated the epigenomic data with clinical outcome. Overall, we found a large degree of DNA methylation variability between and within cytogenetic subtypes, which was informative for subtype determination as well as for identifying subsets of patients with inferior RFS and OS.

Pediatric AML represents many molecularly diverse entities. A large proportion (30%) of the samples in our cohort had undefined cytogenetic subtype, which is similar to the proportion of not otherwise specified (NOS) karyotype expected using the current classification system [2]. Our unsupervised clustering and subtype prediction modeling approaches suggest that several (N = 23) of these patients likely do belong to one of the recurrent subtypes (class probability > 80%) and that based on their DNA methylation patterns, these patients follow similar clinical paths as the subtype-confirmed cases. Although the DNA methylation classification provides a compelling case for re-annotating the subtypes of these patients, subtype-like cases in pediatric ALL [37,38] as well as adult t(8;21)-AML [8] without evidence of underlying genomic aberrations raises the possibility that the DNA methylation signatures are capturing another genetic event affecting similar pathways. Further molecular analysis of these cases will be needed in future studies to confirm their true subtype membership.

Increasing and compelling published evidence across the different hematological malignancies [39], adult [40,41] and pediatric AML [22], and the data presented herein support the notion that DNA methylation signatures at diagnosis are predictive of clinical outcome. Genes highlighted here in, such as *PRDM16* [42,43], *PER3* [44], *NOM1* [45] *BAHCC1* [46], *ZNF423* [47], *SETD7* [48], *RPTOR* [49], and others have been implicated in hematological disease development, progression or clinical outcome. Our results further strengthen two recent studies in pediatric AML focusing on genome-wide DNA methylation. First, in a similarly sized multicenter cohort (AML02) also using 450k arrays, DNA methylation patterns were associated with clinical outcome [22] and cytogenetic subtypes [17]. Reassuringly, several genes were identified in both studies, such as *CRIP2*, *LAMB2*, *USP2*. Second, a DNA hypomethylation signature associated with relapse was described in a small cohort of Italian pediatric t(8;21)/*RUNX1-RUNX1T1* patients [21]. Our data (Figure 4a,b) support and replicate this finding. Amongst the 18 t(8;21)/*RUNX1-RUNX1T1* cases presented herein, all but one patient who was in the group of patients with lower DNA methylation experienced a relapse. Further strengthening this finding is that the analytical and computational methods for DNA methylation analysis applied herein differ from those used by Zampini et al. [21]. The t(8;21)/*RUNX1-RUNX1T1* signature in particular, may reflect the presence of a CpG island methylator phenotype (CIMP), which has been associated with outcome in adult AML [41].

Although AMLs are generally known to incur fewer somatic mutations than other cancer types, several genes are well established to be recurrently mutated in AML and may hold prognostic information [34,50]. *WT1* and *FLT3-ITD* mutations are markers of poor outcome in pediatric AML, while *NPM1* and *CEBPA* mutations may have a favorable outcome. We found that our NK-AML subgroup was split into two distinctive clusters based on DNA methylation, and that mutations in *WT1*, *FLT3*, and *CEBPA* were overrepresented in one of the clusters (Cluster B), while the other cluster contained few mutations (Cluster A). Few (N = 3) *NPM1* mutations were observed in our dataset, two in Cluster B and one in Cluster A. We observed a clearly defined sub-cluster within Cluster B, containing all of the patients with *CEBPA* mutations, a pattern which has been reproduced in adult AML cohorts [8,11]. However, the limited resolution of the four gene panel applied herein (*CEBPA*, *FLT3*, *NPM1*, and *WP1*) in addition to incomplete data for 27% of the patients, may give a partial view of the mutational status and furthermore may have missed key molecular changes associated with the two DNA-methylation defined NK clusters. Future studies are warranted to more thoroughly investigate if there are any other genetic differences that be driven by other recurrent mutation(s) or cryptic translocations that disrupt pathways resulting in altered

DNA methylation landscapes. Other biological differences, perhaps by variation in DNA methylation by age of the patient, may also contribute to these findings. Another limitation of the present study is the unavailability of MRD data for the patients analyzed herein. Inclusion of patients with MRD data available would be important information to analyze in future studies to investigate if DNA methylation grouping can provide additional prognostic information.

Our study not only adds to the growing wealth of knowledge about clinical use of epigenetic signatures, but also highlights the possibility that exclusive use of cytogenetic markers at diagnosis may underestimate the variability within and between classical subtypes. The randomized clinical trials in recent years have struggled to improve outcomes for pediatric AML patients, and this failure may be in part due to applying blanket treatment strategies across AML subtypes, which may not prove effective for all patients [14]. The importance of cataloguing epigenetic changes within genetic subgroups of pediatric AML is underscored given new possible alternatives for future treatment of acute leukemias with DNA methylating agents such as Temozolomide, demethylating agents, and histone deacetylase inhibitors (HDACi). There are several ongoing studies to explore their efficacy in the treatment of hematological cancers [51]. The demethylating agents azacytidine and decitabine have been approved for treatment of myelodysplastic syndrome and a combination of demethylating agents and HDACi has shown promising results in AML patients [52,53].

5. Conclusions

In summary, we show that DNA methylation is associated with cytogenetic subtype and clinical outcome in pediatric AML. Further analyses of DNA methylation levels in t(8;21)/*RUNX1-RUNX1T1* and NK subtypes, in particular, are warranted to investigate the underlying molecular and cellular pathways leading to the intra-subtype DNA methylation variability observed in this cohort and to evaluate if DNA methylation signatures hold prognostic information in larger cohorts.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/genes12060895/s1, Figure S1: Hyperparameters used to optimize for each of the nine classification models. Figure S2: UMAP representation of the complete 450 k dataset and patients labelled by cytogenetic subtype. Figure S3: Heatmap of the 10,000 most variable CpG sites ordered by hierarchical clustering. Figure S4: Model performance accuracy scores for the nine optimized classification models. Figure S5: Permutation analysis scores for Neural Network Classification model for 1000 random permutations. Figure S6: UMAP representation of the 14 pairs of samples taken at AML diagnosis and relapse. Figure S7: Survival analysis of the NK subtype treated on the NOPHO-93 or NOPHO-2004 protocols. Figure S8: Survival analysis of the *MLL/KMT2A*-rearranged patients treated on the NOPHO-93 or NOPHO-2004 protocols. Figure S9: Survival analysis of the *MLL/KMT2A*-rearranged patients treated in the standard vs. high risk groups on the NOPHO-2004 protocol. Table S1: Patient sample information including age, sex, 2004 risk group, treatment protocol, subtype, FAB, subtype prediction, predicted class probability, RFS, OS, relapse, and death status. Table S2: Model Performance Comparison with Wilcoxon signed-rank test and Benjamini/Hochberg multiple test correction for the all nine Classification Models. Table S3: Cytogenetic subtype prediction on 14 patient samples taken from relapse with not recorded subtype and the real subtype of their diagnostic pair. Table S4: Clinical variables (age, sex, outcome) and mutational status of *CEBPA*, *FLT3*, *NPM1*, and *WT1* in the two NK DNA methylation clusters. Table S5: The unique CpG sites per subtype (adjusted *p*-value < 0.05), annotated genes and statistics per CpG. Table S6: The top CpG sites per subtype that separate the patients who relapsed from those who did not (*p*-value < 0.05), annotated genes and statistics per CpG.

Author Contributions: J.P. and J.N. conceived the study. O.K., J.P., C.L.B. and J.N. performed data analysis. O.K. performed modeling. J.P., J.A., U.N.-N., H.H., K.J., Ó.G.J., R.H., B.L., L.P., A.S., B.Z. and R.L. provided AML samples and clinical expertise. O.K., J.P. and J.N. wrote the manuscript. All authors read and approved the final version. All authors have read and agreed to the published version of the manuscript.

References

1. Zwaan, C.M.; Kolb, E.A.; Reinhardt, D.; Abrahamsson, J.; Adachi, S.; Aplenc, R.; De Bont, E.S.; De Moerloose, B.; Dworzak, M.; Gibson, B.E.; et al. Collaborative Efforts Driving Progress in Pediatric Acute Myeloid Leukemia. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2015**, *33*, 2949–2962. [CrossRef] [PubMed]

2. Sandahl, J.D.; Kjeldsen, E.; Abrahamsson, J.; Ha, S.-Y.; Heldrup, J.; Jahnukainen, K.; Jónsson, Ó.G.; Lausen, B.; Palle, J.; Zeller, B.; et al. The applicability of the WHO classification in paediatric AML. A NOPHO-AML study. *Br. J. Haematol.* **2015**, *169*, 859–867. [CrossRef]

3. Rubnitz, J.E.; Inaba, H.; Dahl, G.; Ribeiro, R.C.; Bowman, W.P.; Taub, J.; Pounds, S.; Razzouk, B.I.; Lacayo, N.J.; Cao, X.; et al. Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: Results of the AML02 multicentre trial. *Lancet Oncol.* **2010**, *11*, 543–552. [CrossRef]

4. Coustan-Smith, E.; Campana, D. Should evaluation for minimal residual disease be routine in acute myeloid leukemia? *Curr. Opin. Hematol.* **2013**, *20*, 86–92. [CrossRef]

5. Manola, K.N. Cytogenetics of pediatric acute myeloid leukemia. *Eur. J. Haematol.* **2009**, *83*, 391–405. [CrossRef]

6. Hollink, I.H.I.M.; Zwaan, C.M.; Zimmermann, M.; Arentsen-Peters, T.C.J.M.; Pieters, R.; Cloos, J.; Kaspers, G.J.L.; de Graaf, S.S.N.; Harbott, J.; Creutzig, U.; et al. Favorable prognostic impact of NPM1 gene mutations in childhood acute myeloid leukemia, with emphasis on cytogenetically normal AML. *Leukemia* **2009**, *23*, 262–270. [CrossRef]

7. Mrózek, K.; Marcucci, G.; Paschka, P.; Whitman, S.P.; Bloomfield, C.D. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: Are we ready for a prognostically prioritized molecular classification? *Blood* **2007**, *109*, 431–448. [CrossRef]

8. Figueroa, M.E.; Lugthart, S.; Li, Y.; Erpelinck-Verschueren, C.; Deng, X.; Christos, P.J.; Schifano, E.; Booth, J.; van Putten, W.; Skrabanek, L.; et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **2010**, *17*, 13–27. [CrossRef] [PubMed]

9. Akalin, A.; Garrett-Bakelman, F.E.; Kormaksson, M.; Busuttil, J.; Zhang, L.; Khrebtukova, I.; Milne, T.A.; Huang, Y.; Biswas, D.; Hess, J.L.; et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet.* **2012**, *8*, e1002781. [CrossRef] [PubMed]

10. Cancer Genome Atlas Research, N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **2013**, *368*, 2059–2074. [CrossRef]

11. Gebhard, C.; Glatz, D.; Schwarzfischer, L.; Wimmer, J.; Stasik, S.; Nuetzel, M.; Heudobler, D.; Andreesen, R.; Ehninger, G.; Thiede, C.; et al. Profiling of aberrant DNA methylation in acute myeloid leukemia reveals subclasses of CG-rich regions with epigenetic or genetic association. *Leukemia* **2019**, *33*, 26–36. [CrossRef]

12. Saied, M.H.; Marzec, J.; Khalid, S.; Smith, P.; Down, T.A.; Rakyan, V.K.; Molloy, G.; Raghavan, M.; Debernardi, S.; Young, B.D. Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. *PLoS ONE* **2012**, *7*, e33213.

13. Li, S.; Garrett-Bakelman, F.E.; Chung, S.S.; Sanders, M.A.; Hricik, T.; Rapaport, F.; Patel, J.; Dillon, R.; Vijay, P.; Brown, A.L. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat. Med.* **2016**, *22*, 792–799. [CrossRef] [PubMed]

14. Bolouri, H.; Farrar, J.E.; Triche, T., Jr.; Ries, R.E.; Lim, E.L.; Alonzo, T.A.; Ma, Y.; Moore, R.; Mungall, A.J.; Marra, M.A.; et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **2018**, *24*, 103–112. [CrossRef] [PubMed]

15. Stratmann, S.; Yones, S.A.; Mayrhofer, M.; Norgren, N.; Skaftason, A.; Sun, J.; Smolinska, K.; Komorowski, J.; Herlin, M.K.; Sundstrom, C.; et al. Genomic characterization of relapsed acute myeloid leukemia reveals novel putative therapeutic targets. *Blood Adv.* **2021**, *5*, 900–912. [CrossRef]

16. Juhl-Christensen, C.; Ommen, H.B.; Aggerholm, A.; Lausen, B.; Kjeldsen, E.; Hasle, H.; Hokland, P. Genetic and epigenetic similarities and differences between childhood and adult AML. *Pediatric Blood Cancer* **2012**, *58*, 525–531. [CrossRef] [PubMed]

17. Lamba, J.K.; Cao, X.; Raimondi, S.; Downing, J.; Ribeiro, R.; Gruber, T.A.; Rubnitz, J.; Pounds, S. DNA Methylation Clusters and Their Relation to Cytogenetic Features in Pediatric AML. *Cancers* **2020**, *12*, 3024. [CrossRef] [PubMed]

18. Koldobskiy, M.A.; Abante, J.; Jenkinson, G.; Pujadas, E.; Tetens, A.; Zhao, F.; Tryggvadottir, R.; Idrizi, A.; Reinisch, A.; Majeti, R.; et al. A Dysregulated DNA Methylation Landscape Linked to Gene Expression in MLL-Rearranged AML. *Epigenetics* **2020**, *15*, 841–858. [CrossRef]

19. Larmonie, N.S.D.; Arentsen-Peters, T.C.J.M.; Obulkasim, A.; Valerio, D.; Sonneveld, E.; Danen-van Oorschot, A.A.; de Haas, V.; Reinhardt, D.; Zimmermann, M.; Trka, J.; et al. MN1 overexpression is driven by loss of DNMT3B methylation activity in inv(16) pediatric AML. *Oncogene* **2018**, *37*, 107–115. [CrossRef]

20. Zhou, L.; Fu, L.; Lv, N.; Liu, J.; Li, Y.; Chen, X.; Xu, Q.; Chen, G.; Pang, B.; Wang, L.; et al. Methylation-associated silencing of BASP1 contributes to leukemogenesis in t(8;21) acute myeloid leukemia. *Exp. Mol. Med.* **2018**, *50*, 1–8. [CrossRef]

21. Zampini, M.; Tregnago, C.; Bisio, V.; Simula, L.; Borella, G.; Manara, E.; Zanon, C.; Zonta, F.; Serafin, V.; Accordi, B.; et al. Epigenetic heterogeneity affects the risk of relapse in children with t(8;21)RUNX1-RUNX1T1-rearranged AML. *Leukemia* **2018**, *32*, 1124–1134. [CrossRef] [PubMed]

22. Lamba, J.K.; Cao, X.; Raimondi, S.C.; Rafiee, R.; Downing, J.R.; Lei, S.; Gruber, T.; Ribeiro, R.C.; Rubnitz, J.E.; Pounds, S.B. Integrated epigenetic and genetic analysis identifies markers of prognostic significance in pediatric acute myeloid leukemia. *Oncotarget* **2018**, *9*, 26711–26723. [CrossRef] [PubMed]

23. Lie, S.O.; Abrahamsson, J.; Clausen, N.; Forestier, E.; Hasle, H.; Hovi, L.; Jonmundsson, G.; Mellander, L.; Gustafsson, G. Treatment stratification based on initial in vivo response in acute myeloid leukaemia in children without Down's syndrome: Results of NOPHO-AML trials. *Br. J. Haematol.* **2003**, *122*, 217–225. [CrossRef] [PubMed]

24. Abrahamsson, J.; Forestier, E.; Heldrup, J.; Jahnukainen, K.; Jonsson, O.G.; Lausen, B.; Palle, J.; Zeller, B.; Hasle, H. Response-guided induction therapy in pediatric acute myeloid leukemia with excellent remission rate. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **2011**, *29*, 310–315. [CrossRef]

25. Van Rossum, G.; Drake, F.L. *The Python Language Reference Manual*; Network Theory Ltd.: Bristol, UK, 2011.

26. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]

27. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S. *Jupyter Notebooks–A Publishing Format For Reproducible Computational Workflows*; IOS Press: Amsterdam The Netherlands, 2016.

28. Nordlund, J.; Backlin, C.L.; Wahlberg, P.; Busche, S.; Berglund, E.C.; Eloranta, M.L.; Flaegstad, T.; Forestier, E.; Frost, B.M.; Harila-Saari, A.; et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* **2013**, *14*, r105. [CrossRef]

29. Behdenna, A.; Haziza, J.; Azencott, C.-A.; Nordor, A. pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BioRxiv* **2020**. [CrossRef]

30. Abraham, A.; Pedregosa, F.; Eickenberg, M.; Gervais, P.; Mueller, A.; Kossaifi, J.; Gramfort, A.; Thirion, B.; Varoquaux, G. Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **2014**, *8*, 14. [CrossRef]

31. Hunter, J.D. Matplotlib: A 2D graphics environment. *IEEE Ann. Hist. Comput.* **2007**, *9*, 90–95. [CrossRef]

32. Waskom, M.L. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]

33. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:preprint/1802.03426. Available online: https://umap-learn.readthedocs.io/en/latest/ (accessed on 19 February 2021).

34. Staffas, A.; Kanduri, M.; Hovland, R.; Rosenquist, R.; Ommen, H.B.; Abrahamsson, J.; Forestier, E.; Jahnukainen, K.; Jónsson, Ó.G.; Zeller, B. Presence of FLT3-ITD and high BAALC expression are independent prognostic markers in childhood acute myeloid leukemia. *Blood J. Am. Soc. Hematol.* **2011**, *118*, 5905–5913. [CrossRef]

35. Wang, J.; Muntean, A.G.; Hess, J.L. ECSASB2 mediates MLL degradation during hematopoietic differentiation. *Blood J. Am. Soc. Hematol.* **2012**, *119*, 1151–1161. [CrossRef]

36. Stavropoulou, V.; Kaspar, S.; Brault, L.; Sanders, M.A.; Juge, S.; Morettini, S.; Tzankov, A.; Iacovino, M.; Lau, I.-J.; Milne, T.A. MLL-AF9 expression in hematopoietic stem cells drives a highly invasive AML expressing EMT-related genes linked to poor outcome. *Cancer Cell* **2016**, *30*, 43–58. [CrossRef]

37. Nordlund, J.; Bäcklin, C.L.; Zachariadis, V.; Cavelier, L.; Dahlberg, J.; Öfverholm, I.; Barbany, G.; Nordgren, A.; Övernäs, E.; Abrahamsson, J. DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia. *Clin. Epigenet.* **2015**, *7*, 1–12. [CrossRef]

38. Lilljebjörn, H.; Henningsson, R.; Hyrenius-Wittsten, A.; Olsson, L.; Orsmark-Pietras, C.; Von Palffy, S.; Askmyr, M.; Rissler, M.; Schrappe, M.; Cario, G. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nat. Commun.* **2016**, *7*, 1–13. [CrossRef]

39. Duran-Ferrer, M.; Clot, G.; Nadeu, F.; Beekman, R.; Baumann, T.; Nordlund, J.; Marincevic-Zuniga, Y.; Lönnerholm, G.; Rivas-Delgado, A.; Martin, S. The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nat. Cancer* **2020**, *1*, 1066–1081. [CrossRef] [PubMed]

40. Bullinger, L.; Ehrich, M.; Dohner, K.; Schlenk, R.F.; Dohner, H.; Nelson, M.R.; van den Boom, D. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood* **2010**, *115*, 636–642. [CrossRef]

41. Kelly, A.D.; Kroeger, H.; Yamazaki, J.; Taby, R.; Neumann, F.; Yu, S.; Lee, J.T.; Patel, B.; Li, Y.; He, R.; et al. A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome. *Leukemia* **2017**, *31*, 2011–2019. [CrossRef] [PubMed]

42. Yamato, G.; Yamaguchi, H.; Handa, H.; Shiba, N.; Kawamura, M.; Wakita, S.; Inokuchi, K.; Hara, Y.; Ohki, K.; Okubo, J. Clinical features and prognostic impact of PRDM16 expression in adult acute myeloid leukemia. *Genes Chromosomes Cancer* **2017**, *56*, 800–809. [CrossRef] [PubMed]

43. Zhou, B.; Wang, J.; Lee, S.Y.; Xiong, J.; Bhanu, N.; Guo, Q.; Ma, P.; Sun, Y.; Rao, R.C.; Garcia, B.A. PRDM16 suppresses MLL1r leukemia via intrinsic histone methyltransferase activity. *Mol. Cell* **2016**, *62*, 222–236. [CrossRef]

44. Yang, M.-Y.; Lin, P.-M.; Hsiao, H.-H.; Hsu, J.-F.; Lin, H.Y.-H.; Hsu, C.-M.; Chen, I.-Y.; Su, S.-W.; Liu, Y.-C.; Lin, S.-F. Up-regulation of PER3 expression is correlated with better clinical outcome in acute leukemia. *Anticancer. Res.* **2015**, *35*, 6615–6622.

45. Simmons, H.M.; Ruis, B.L.; Kapoor, M.; Hudacek, A.W.; Conklin, K.F. Identification of NOM1, a nucleolar, eIF4A binding protein encoded within the chromosome 7q36 breakpoint region targeted in cases of pediatric acute myeloid leukemia. *Gene* **2005**, *347*, 137–145. [CrossRef]

46. Fan, H.; Lu, J.; Guo, Y.; Li, D.; Zhang, Z.-M.; Tsai, Y.-H.; Pi, W.-C.; Ahn, J.H.; Gong, W.; Xiang, Y. BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat. Genet.* **2020**, *52*, 1384–1396. [CrossRef] [PubMed]

47. Harder, L.; Eschenburg, G.; Zech, A.; Kriebitzsch, N.; Otto, B.; Streichert, T.; Behlich, A.-S.; Dierck, K.; Klingler, B.; Hansen, A. Aberrant ZNF423 impedes B cell differentiation and is linked to adverse outcome of ETV6-RUNX1 negative B precursor acute lymphoblastic leukemia. *J. Exp. Med.* **2013**, *210*, 2289–2304. [CrossRef]

48. Batista, I.d.A.A.; Helguero, L.A. Biological processes and signal transduction pathways regulated by the protein methyltransferase SETD7 and their significance in cancer. *Signal Transduct. Target. Ther.* **2018**, *3*, 1–14. [CrossRef] [PubMed]

49. Zhang, H.; Song, G.; Song, G.; Li, R.; Gao, M.; Ye, L.; Zhang, C. Identification of DNA methylation prognostic signature of acute myelocytic leukemia. *PLoS ONE* **2018**, *13*, e0199689. [CrossRef] [PubMed]

50. Radtke, I.; Mulligan, C.G.; Ishii, M.; Su, X.; Cheng, J.; Ma, J.; Ganti, R.; Cai, Z.; Goorha, S.; Pounds, S.B. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12944–12949. [CrossRef]

51. Touzart, A.; Mayakonda, A.; Smith, C.; Hey, J.; Toth, R.; Cieslak, A.; Andrieu, G.P.; Tran Quang, C.; Latiri, M.; Ghysdael, J.; et al. Epigenetic analysis of patients with T-ALL identifies poor outcomes and a hypomethylating agent-responsive subgroup. *Sci. Transl. Med.* **2021**, *13*. [CrossRef]

52. Bots, M.; Verbrugge, I.; Martin, B.P.; Salmon, J.M.; Ghisi, M.; Baker, A.; Stanley, K.; Shortt, J.; Ossenkoppele, G.J.; Zuber, J.; et al. Differentiation therapy for the treatment of t(8;21) acute myeloid leukemia using histone deacetylase inhibitors. *Blood* **2014**, *123*, 1341–1352. [CrossRef]

53. Newcombe, A.A.; Gibson, B.E.; Keeshan, K. Harnessing the potential of epigenetic therapies for childhood acute myeloid leukemia. *Exp. Hematol.* **2018**, *63*, 1–11. [CrossRef] [PubMed]