

# Supplementary Figures

## DNA methylation signatures predict cytogenetic subtype and outcome in pediatric acute myeloid leukemia (AML)

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

<sup>1</sup> Department of Medical Sciences, Molecular Precision Medicine and Science for Life Laboratory, Uppsala University, Sweden

<sup>2</sup> Department of Women's and Children's Health, Uppsala University, Sweden

<sup>3</sup> Department of Medical Sciences, Cancer Pharmacology and Computational Medicine, Uppsala University, Sweden

<sup>4</sup> Department of Pediatrics, Queen Silvia Children's Hospital, Gothenburg, Sweden

<sup>5</sup> Department of Clinical Sciences, Pediatrics, Umeå University Hospital, Umeå, Sweden

<sup>6</sup> Department of Pediatrics, Aarhus University Hospital, Denmark

<sup>7</sup> Children's Hospital, Helsinki University Central Hospital, Helsinki, and University of Helsinki, Finland

<sup>8</sup> Department of Pediatrics, Landspítali University Hospital, Reykjavík, Iceland

<sup>9</sup> Center of Medical Genetics and Molecular Medicine, Haukeland University Hospital, Norway

<sup>10</sup> Department of Pediatrics and Adolescent Medicine, Rigshospitalet, University of Copenhagen, Denmark

<sup>11</sup> Department of Clinical Chemistry and Transfusion Medicine, University of Gothenburg, Sweden

<sup>12</sup> Division of Paediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway

\* Authors to whom correspondence should be addressed

<sup>†</sup> Equal contribution

```

hyperparameters = {'NeuralNetwork': {'NeuralNetwork__epochs':[100,200],
                                     'NeuralNetwork__batch_size':[8, 16],
                                     'NeuralNetwork__activation' : ['relu', 'sigmoid'],
                                     'NeuralNetwork__dim1' : [100, 200]},

'NearestCentroid':{'parameters': None},
'KNeighborsClassifier':{'KNeighborsClassifier__estimator__n_neighbors': [5,10,15,20],
                       'KNeighborsClassifier__estimator__weights':['uniform', 'distance'],
                       'KNeighborsClassifier__estimator__algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
'SVC':{'SVC__estimator__kernel':['linear', 'poly', 'rbf', 'sigmoid'],
       'SVC__estimator__degree':[3,4],
       'SVC__estimator__C':[1, 0.1, 0.01],
       'SVC__estimator__gamma':['scale', 'auto']}},
'GaussianProcessClassifier':{'parameters': None},

'DecisionTreeClassifier':{'DecisionTreeClassifier__estimator__criterion':['gini', 'entropy'],
                          'DecisionTreeClassifier__estimator__max_features':['auto', 'sqrt', 'log2'],
                          'DecisionTreeClassifier__estimator__splitter':['best','random']},

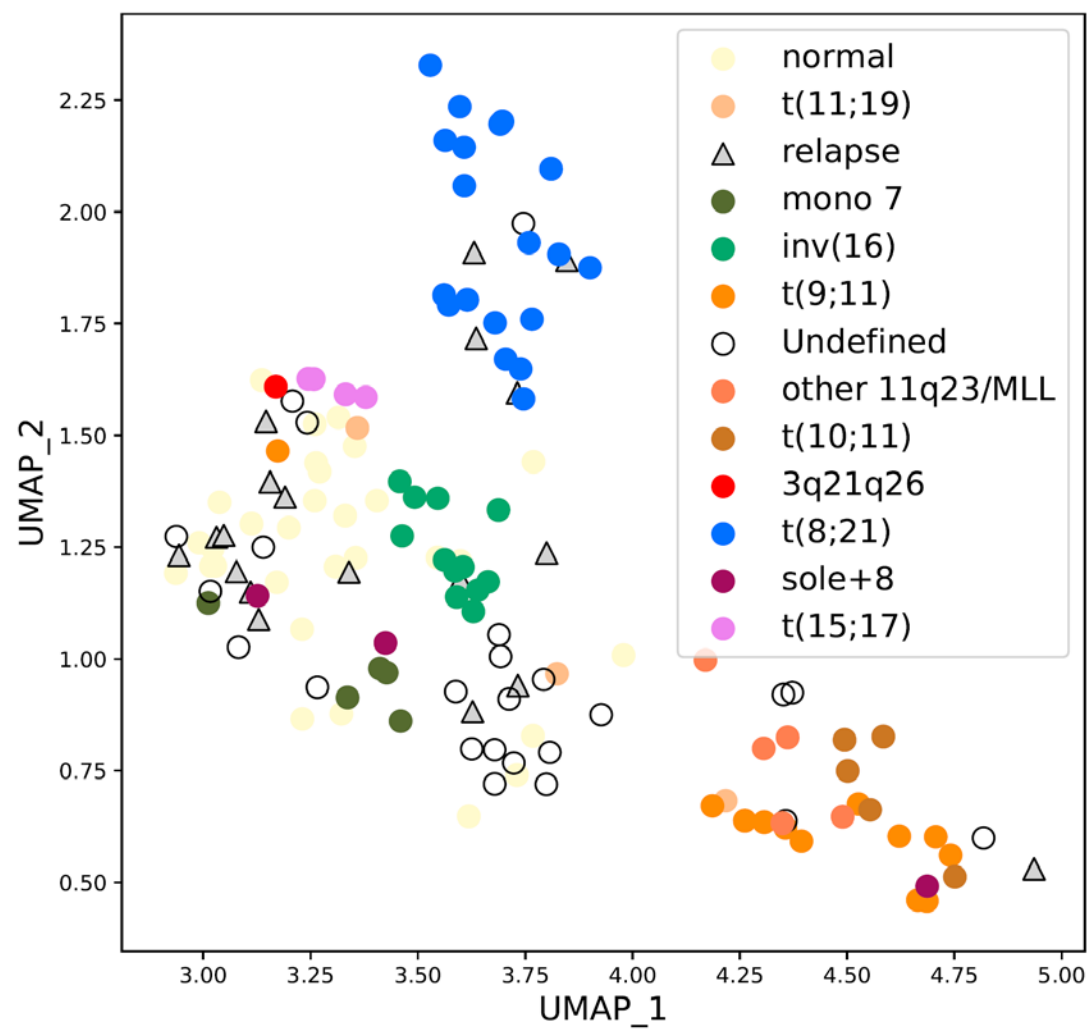
'RandomForestClassifier':{'RandomForestClassifier__estimator__n_estimators':[10, 50, 100, 200],
                          'RandomForestClassifier__estimator__criterion':['gini', 'entropy'],
                          'RandomForestClassifier__estimator__max_features':['auto', 'sqrt', 'log2']},

'AdaBoostClassifier':{'AdaBoostClassifier__estimator__n_estimators':[10, 50, 100],
                      'AdaBoostClassifier__estimator__learning_rate':[0.01, 0.1, 1]},

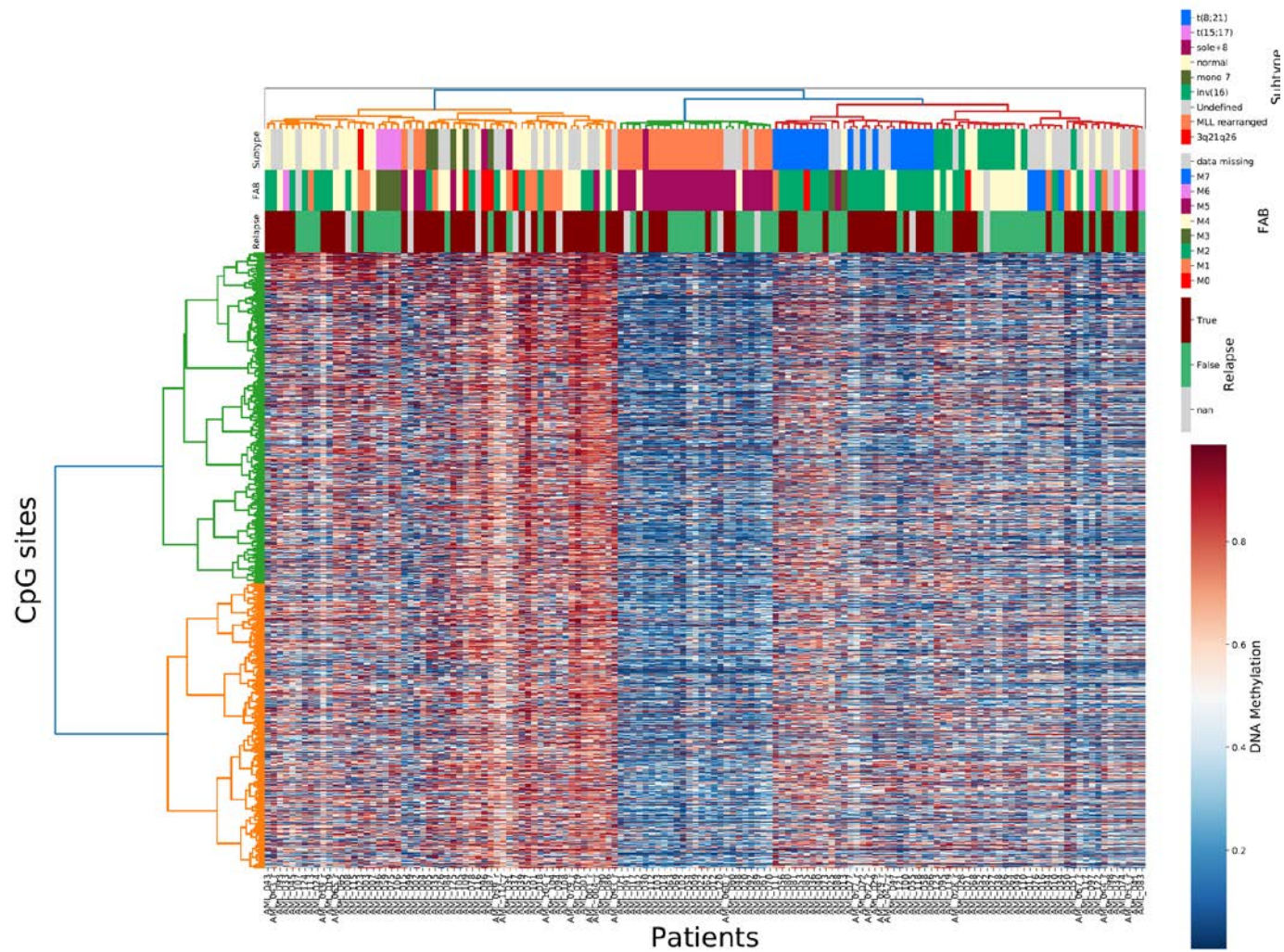
'GaussianNB':{'parameters': None}
}

```

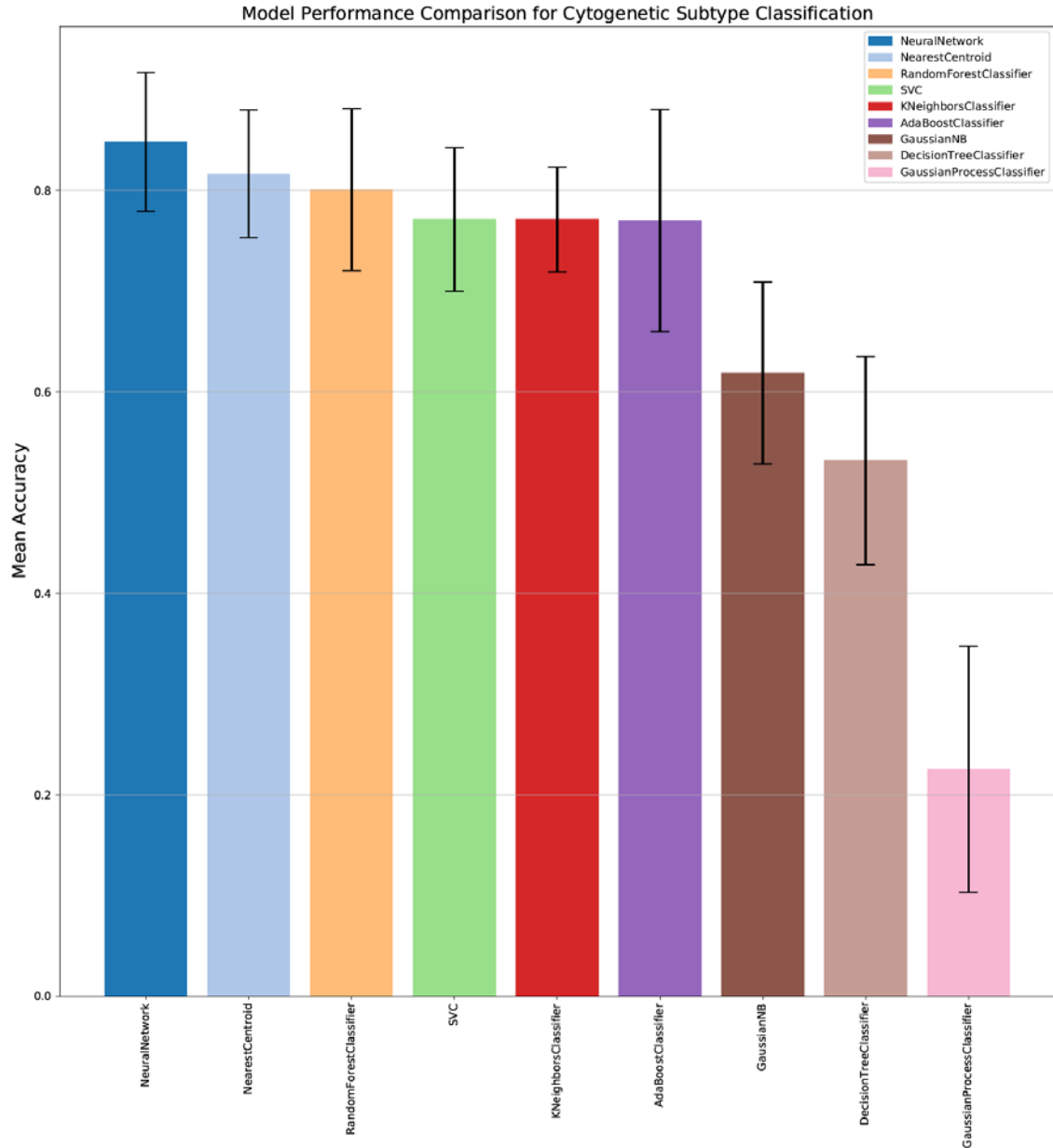
**Figure S1.** Hyperparameters used to optimize for each of the nine classification models.



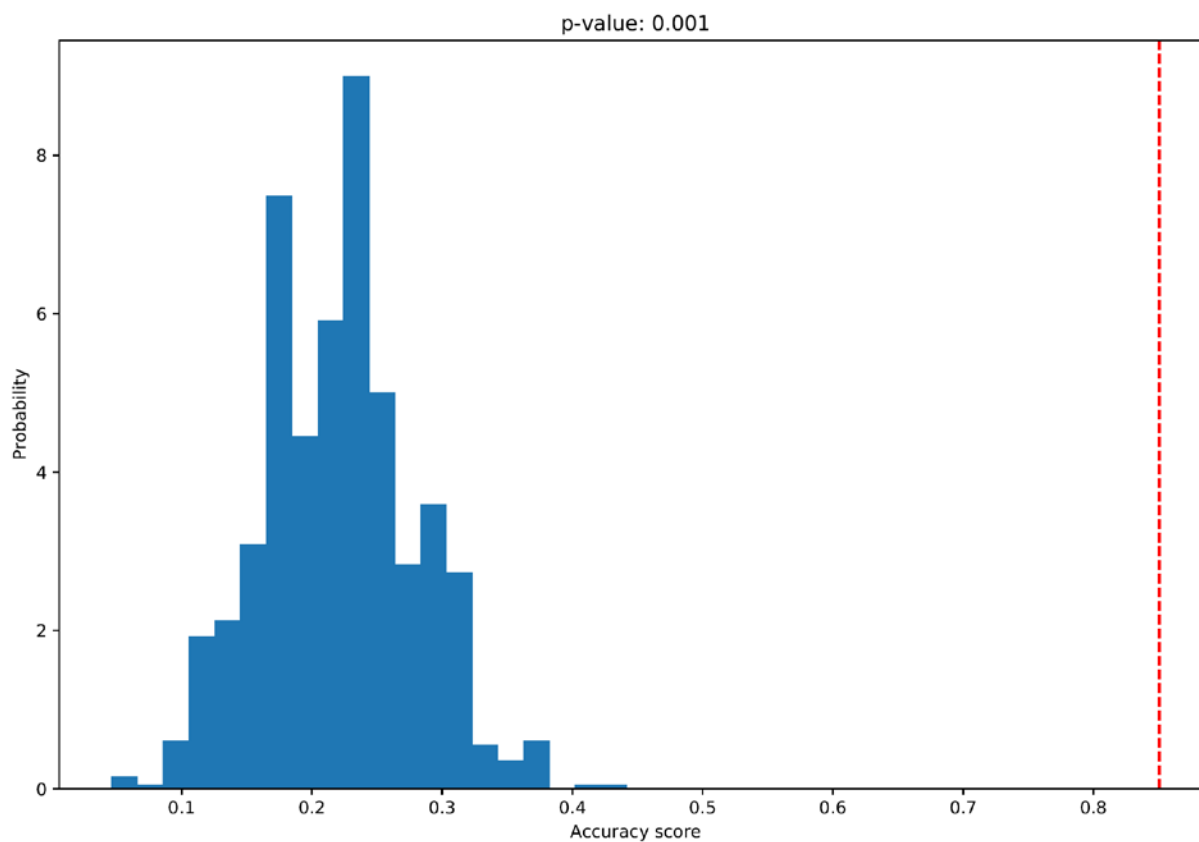
**Figure S2.** UMAP representation of the complete 450k dataset and patients labelled by cytogenetic subtype.



**Figure S3.** Heatmap of the 10,000 most variable CpG sites ordered by hierarchical clustering.

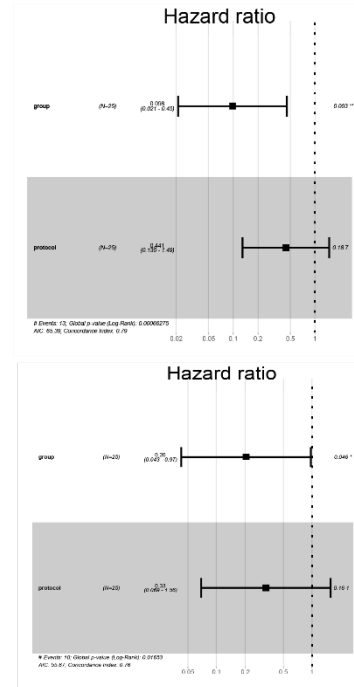
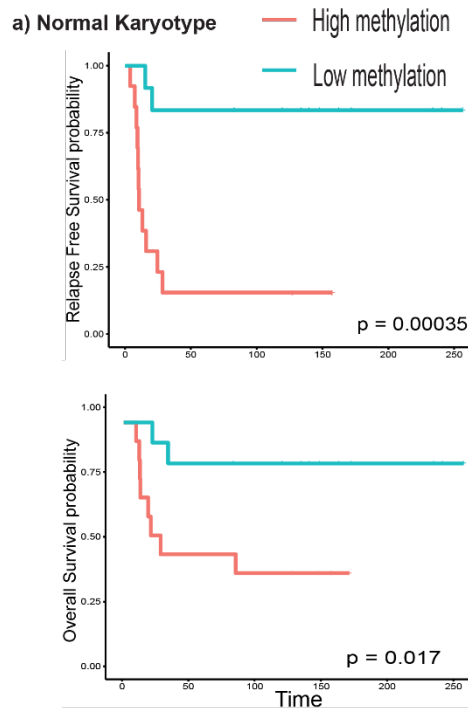


**Figure S4:** Model performance accuracy scores for the nine optimized classification models. Mean accuracy of the classifier and standard deviation (error bars) is shown on the y-axis. The nine classification models are plotted along the x-axis.

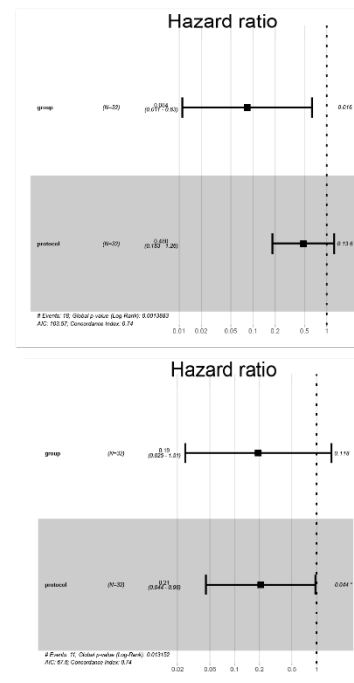
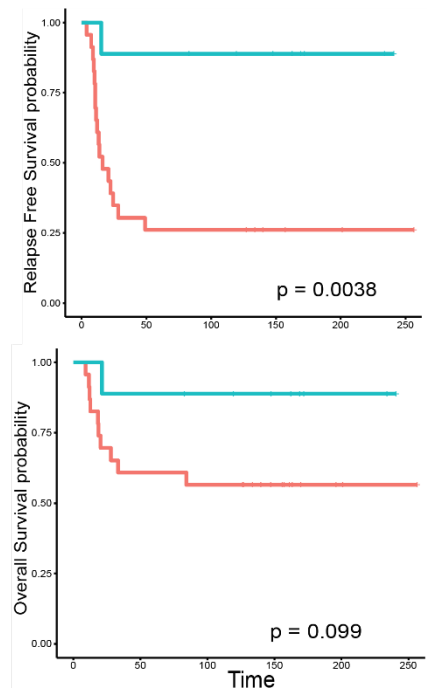


**Figure S5:** Permutation analysis scores for Neural Network Classification model for 1000 random permutations. The red dashed line demonstrates the model score on the test data. The p-value from the Wilcoxon signed-rank test is shown above the plot.



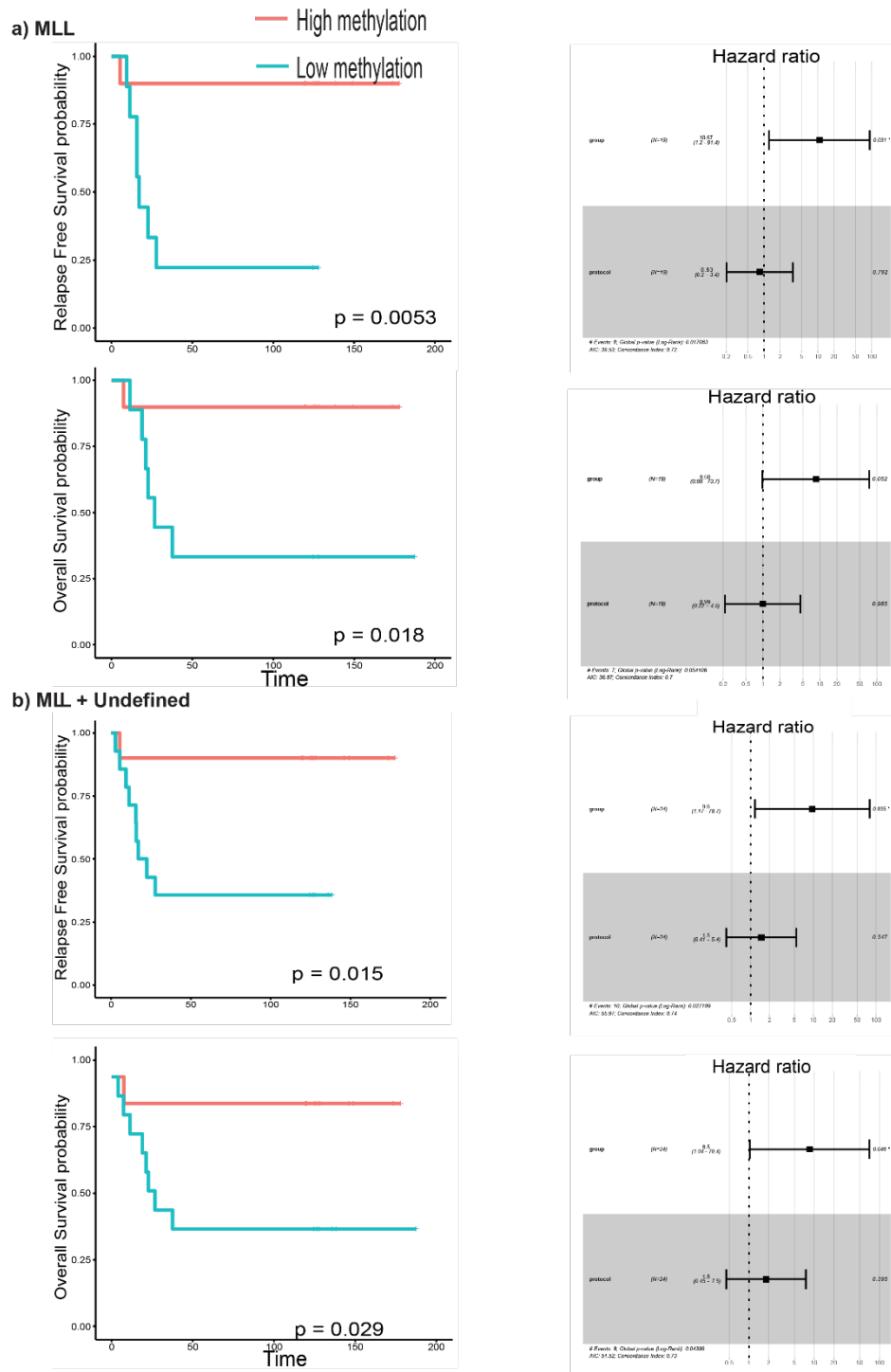


**b) Normal Karyotype + Undefined**

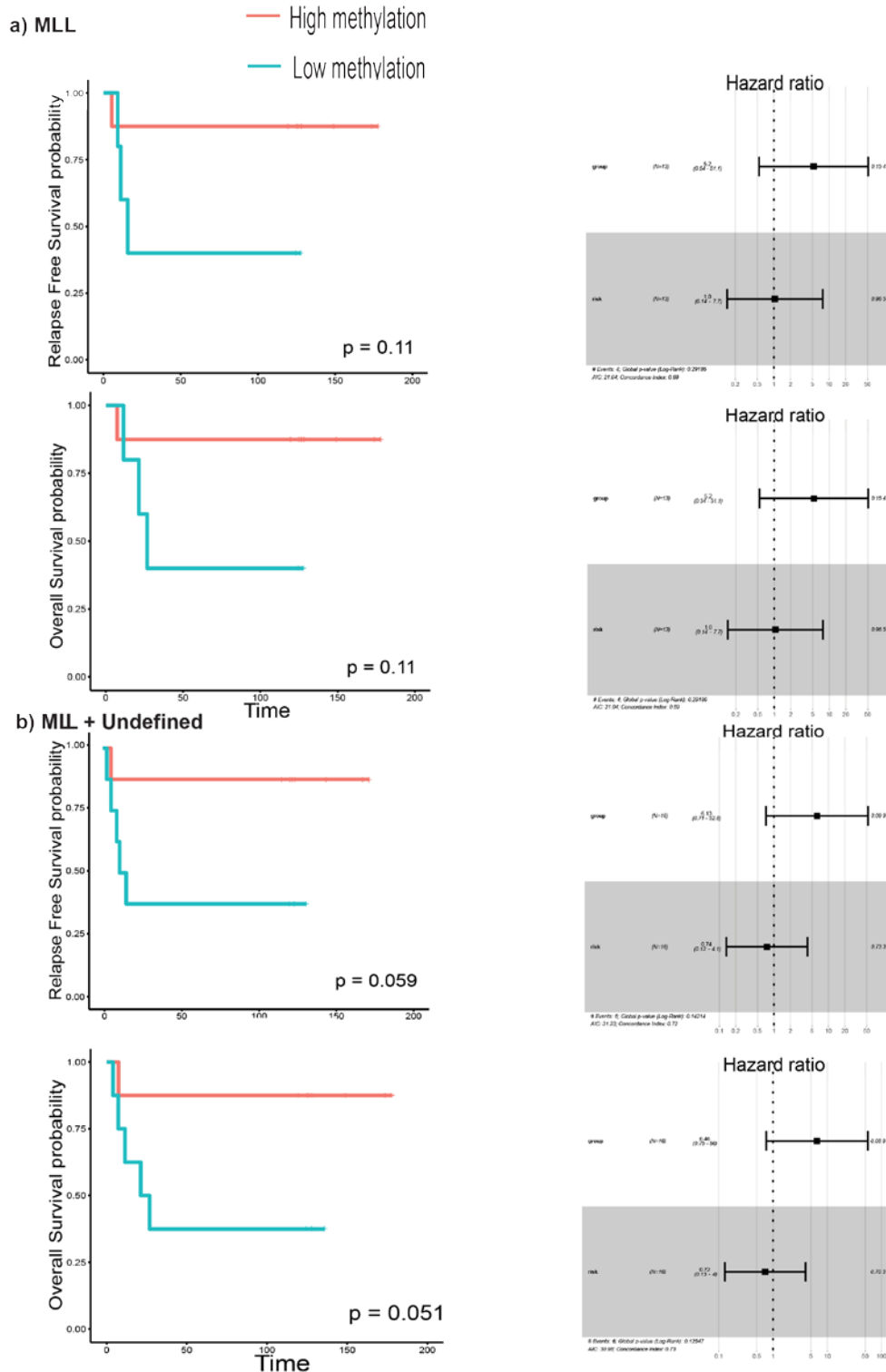


**Figure S7.** Survival analysis of the NK subtype patients treated on the NOPHO-93 or NOPHO-2004 protocols. A) Kaplan-Meier curves and cox regression analysis for relapse-free and overall survival based on 25 NK of confirmed diagnostic subtype (diagnostic-known) and B) in the diagnostic-known NK samples in addition to the seven DNA-methylation predicted samples of undefined cytogenetics. Results from the Cox Regression analysis is plotted to the right. Methylation group (group) and treatment protocol (NOPHO-93 vs NOPHO-2004, i.e. protocol) are included as covariates.





**Figure S8.** Survival analysis of the *MLL/KMT2A*-rearranged patients treated on the NOPHO-93 or NOPHO-2004 protocols. A) Kaplan-Meier curves and cox regression analysis for relapse-free and overall survival based on 19 *MLL/KMT2A*-rearranged cases with confirmed diagnostic subtype (diagnostic-known) and B) in the diagnostic-known *MLL/KMT2A*-rearranged cases in addition to the five DNA-methylation predicted samples of undefined cytogenetics. Results from the Cox Regression analysis is plotted to the right. Methylation group (group) and treatment protocol (NOPHO-93 vs NOPHO-2004, i.e. protocol) are included as covariates.



**Figure S9.** Survival analysis of the *MLL/KMT2A*-rearranged patients treated on the NOPHO-2004 protocol. A) Kaplan-Meier curves and cox regression analysis for relapse-free and overall survival based on 13 *MLL*-rearranged cases with confirmed diagnostic subtype (diagnostic-known) and B) in the diagnostic-known *MLL/KMT2A*-rearranged cases in addition to the three DNA-methylation predicted samples of undefined cytogenetics. Results from the Cox Regression analysis is plotted to the right. Methylation group (group) and risk group (standard vs high risk, i.e. risk) are included as covariates.