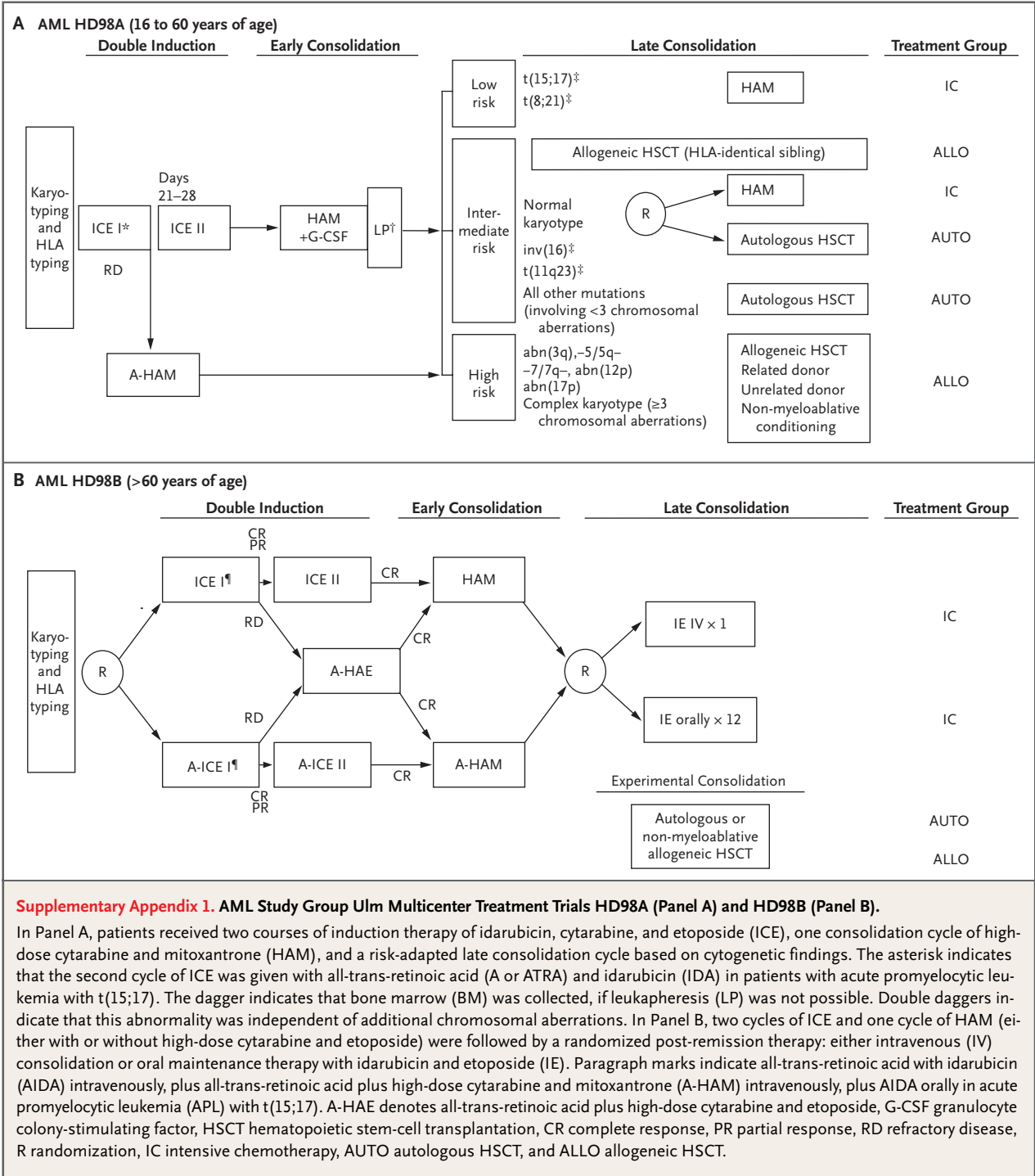


Supplementary Appendix

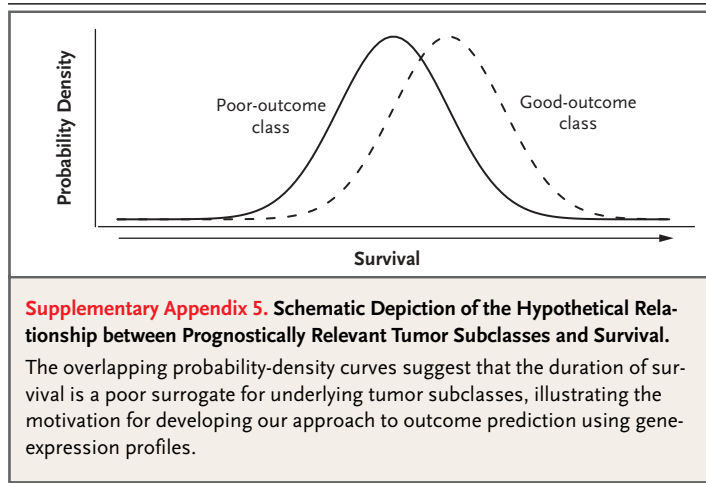
This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Bullinger L, Döhner K, Bair E, et al. Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia. *N Engl J Med* 2004;350:1605-16.



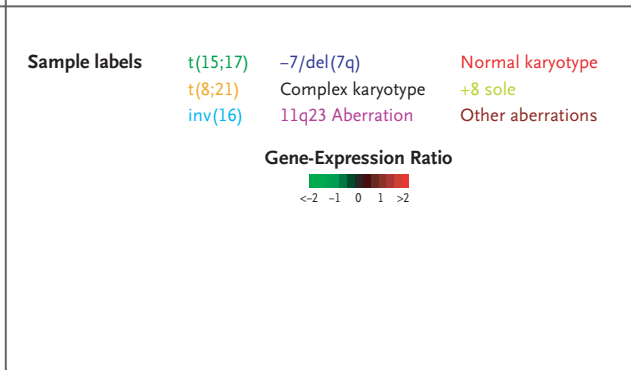
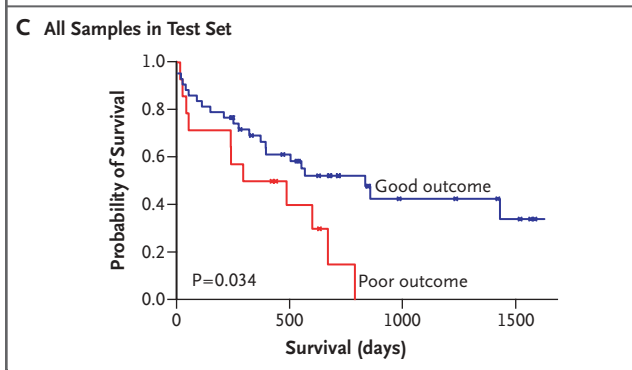
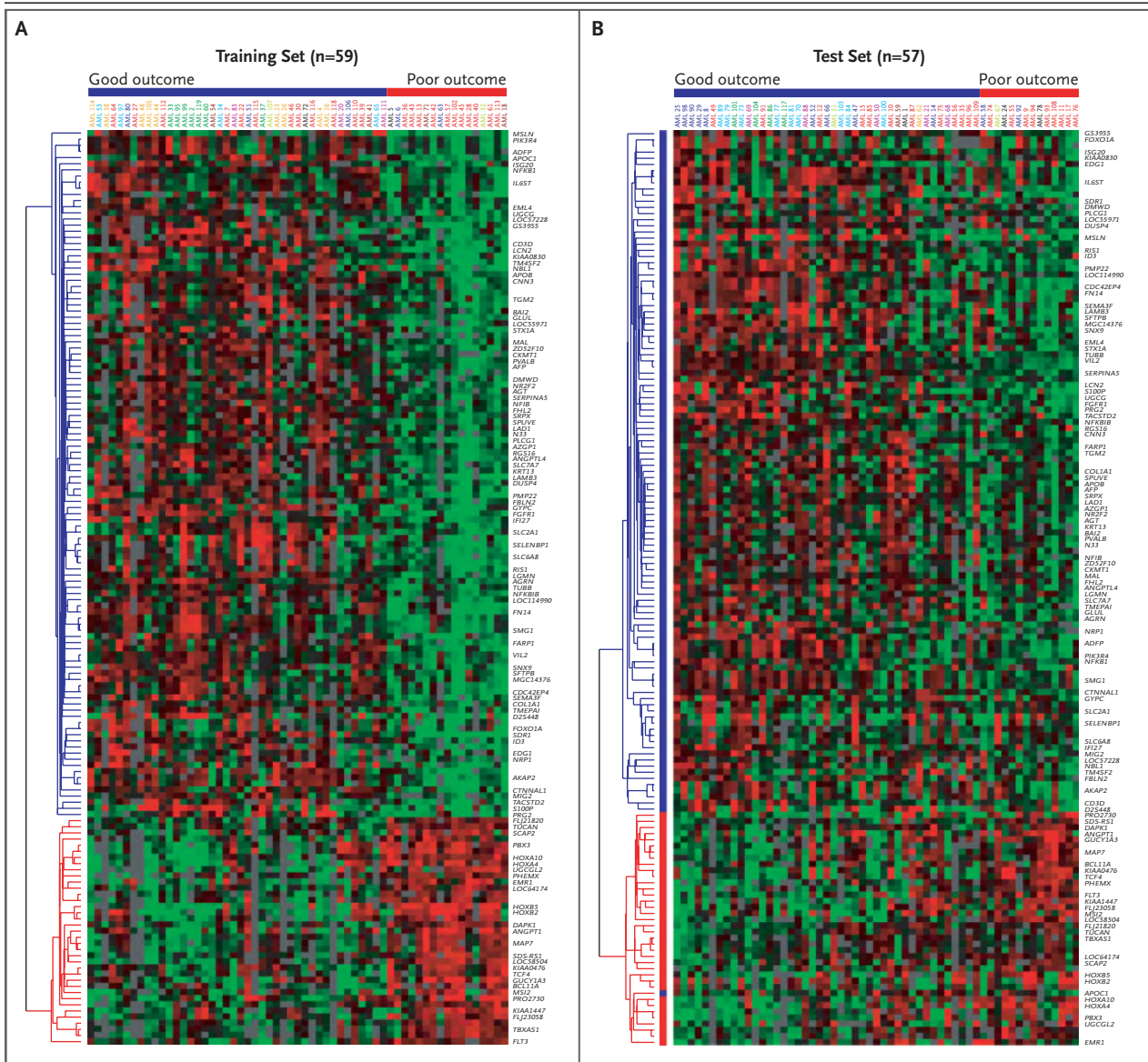
SUPPLEMENTARY APPENDIX 2.

To develop an outcome predictor in the training sample set, we used a SAM gene set (genes correlating with the survival time) to group the samples into two subgroups by means of k-means cluster analysis (which is computationally less intensive than hierarchical clustering). We next used Kaplan–Meier survival analysis to determine the prognostic relevance of the two subgroups (and to assign good-outcome and poor-outcome labels to each subgroup). We then used the PAM method to identify a 10-fold cross-validated gene-expression predictor for these cluster-defined outcome classes. The entire procedure was then iterated, with the use of fractionally fewer SAM genes each time, until a minimal Kaplan–Meier P value was obtained by means of a log-rank test (i.e., the best separation of samples into prognostic groups with the use of SAM) and a minimal cross-validation error rate (i.e., the best prediction of outcome group with the use of PAM) was achieved, represented in our analysis by a set of 1600 genes with the use of SAM and a prediction set of 2655 genes with the use of PAM. A more manageable prediction set of 133 unique genes (represented by 149 complementary DNAs) was selected near the point on the PAM shrinkage curve at which the cross-validation error started to rise steeply (but was still within 1 SE of the minimum) and was used for all subsequent analyses.



SUPPLEMENTARY APPENDIX 6.

In the training set, hierarchical cluster analysis across the 133 predictive genes (represented by 149 complementary DNAs) identified two major gene clusters, one more highly expressed in the good-outcome subgroup, and the other more highly expressed in the poor-outcome subgroup (Fig. 4A). Two major gene clusters were also identified in the test set (Fig. 4B). A chi-square test demonstrated a significant correlation between membership in the good-outcome and poor-outcome gene clusters in the training set and membership in the two major gene clusters in the test set ($P < 0.001$), confirming the detection of good-outcome and poor-outcome gene signatures in the test set. Furthermore, the average (centroid) gene-expression signature across the 133 genes for the good-outcome and for the poor-outcome subgroups in the training set was highly correlated with the average gene-expression signature across the 133 genes for each of the two test sample subgroups defined by hierarchical clustering (Pearson correlation coefficients, 0.74 and 0.72, respectively), also confirming the presence of good-outcome and poor-outcome signatures in the test set.



Supplementary Appendix 7. Outcome Prediction by Means of the Nearest Shrunken Centroid.

In Panel A, columns represent AML samples in the training set ordered according to k-means clustering; rows represent the 149 predictive complementary DNAs (cDNAs), ordered according to hierarchical clustering. Mean-centered gene-expression ratios are depicted by a log-transformed (on a base 2 scale) pseudocolor scale; gray denotes poorly measured data. Good-outcome and poor-outcome subgroups were identified by means of Kaplan–Meier analysis. In Panel B, columns represent AML samples in the test set, ordered according to the rank value of Pearson correlation to the nearest shrunken centroid of good-outcome or poor-outcome class in the training set (see text); rows represent the 149 predictive cDNAs, ordered according to hierarchical clustering. Vertical bar (left) indicates genes that were expressed in the good-outcome (blue) or poor-outcome (red) subgroup in the training set. Panel C shows Kaplan–Meier survival analysis in the two subgroups according to the defined nearest-shrunken-centroid method; P=0.034 for the difference between subgroups by the log-rank test.