

ARTICLE

Development and Validation of a Novel RNA Sequencing–Based Prognostic Score for Acute Myeloid Leukemia

Mei Wang, Johan Lindberg, Daniel Klevebring, Christer Nilsson, Sören Lehmann, Henrik Grönberg, Mattias Rantalainen

See the Notes section for the full list of authors' affiliations.

Correspondence to: Mattias Rantalainen, PhD, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, PO Box 281, SE-171 77, Stockholm, Sweden (e-mail: mattias.rantalainen@ki.se).

Abstract

Background: Recent progress in sequencing technologies allows us to explore comprehensive genomic and transcriptomic information to improve the current European LeukemiaNet (ELN) system of acute myeloid leukemia (AML).

Methods: We compared the prognostic value of traditional demographic and cytogenetic risk factors, genomic data in the form of somatic aberrations of 25 AML-relevant genes, and whole-transcriptome expression profiling (RNA sequencing) in 267 intensively treated AML patients (Clinseq-AML). Multivariable penalized Cox models (overall survival [OS]) were developed for each data modality (clinical, genomic, transcriptomic), together with an associated prognostic risk score.

Results: Of the three data modalities, transcriptomic data provided the best prognostic value, with an integrated area under the curve (iAUC) of a time-dependent receiver operating characteristic (ROC) curve of 0.73. We developed a prognostic risk score (Clinseq-G) from transcriptomic data, which was validated in the independent The Cancer Genome Atlas AML cohort (RNA sequencing, $n = 142$, iAUC = 0.73, comparing the high-risk group with the low-risk group, hazard ratio [HR]_{OS} = 2.42, 95% confidence interval [CI] = 1.51 to 3.88). Comparison between Clinseq-G and ELN score iAUC estimates indicated strong evidence in favor of the Clinseq-G model (Bayes factor = 26.78). The proposed model remained statistically significant in multivariable analysis including the ELN and other well-known risk factors (HR_{OS} = 2.34, 95% CI = 1.30 to 4.22). We further validated the Clinseq-G model in a second independent data set ($n = 458$, iAUC = 0.66, adjusted HR_{OS} = 2.02, 95% CI = 1.33 to 3.08; adjusted HR_{EFS} = 2.10, 95% CI = 1.42 to 3.12).

Conclusions: Our results indicate that the Clinseq-G prediction model, based on transcriptomic data from RNA sequencing, outperforms traditional clinical parameters and previously reported models based on genomic biomarkers.

Risk stratification of acute myeloid leukemia (AML) guides the postremission strategy after standard induction chemotherapy. The current risk stratification of AML patients, the European LeukemiaNet (ELN) system, is based mainly on cytogenetics and common somatic aberrations (*NPM1*, *FLT3-ITD*, *CEBPA*, *RUNX1*, *ASXL1*, and *TP53*) (1). However, there is heterogeneity in the outcome for individuals in each risk group, especially in the intermediate group (2). In addition to the molecular and clinical characteristics included the ELN risk classification system, alternative biomarker panels and somatic aberrations have been

proposed to refine the risk classification (3–7). We have recently summarized and independently evaluated several of these proposed AML risk classification systems (8).

With the rapid ongoing development of molecular profiling technologies using sequencing, comprehensive molecular phenotyping, including gene expression profiling and somatic mutation profiling, is now feasible and will most likely become routine in the clinical setting in the near future (8). Both Li's model, a 24-gene expression signature (3), and Patel's 18-gene mutation panel (5) showed improved risk stratification

Received: August 28, 2017; Revised: December 21, 2017; Accepted: January 26, 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

compared with the ELN system for all AML patients. In this study, we investigate the prognostic value of different types of predictors, including traditional demographic parameters, cytogenetic risk factors, somatic mutations, and whole-transcriptomic gene expression profiling by RNA sequencing, and develop a prediction model based on the most informative data set to predict and stratify the outcome of AML.

Methods

Patients

Three AML cohorts were used in this study: the Clinseq cohort (8), The Cancer Genome Atlas (TCGA)-AML cohort (9), and the GSE6891 data set (10). The Clinseq cohort, which contains 267 AML patients, served as a training cohort to develop the prediction model. The TCGA cohort, comprised of 142 AML patients, and the GSE6891 data set, with 458 patients, were used as validation cohorts.

Patients in the Clinseq cohort were diagnosed in Sweden between February 1997 and August 2014. Bone marrow or peripheral blood samples were obtained at the time of diagnosis. All patients were treated with intensive induction regimens, including anthracyclines and cytosine arabinoside, according to national guidelines (11). Clinical data were retrieved from the Swedish Acute Leukemia Registry (SALR) (12) or from patient records. The regional ethical review board in Stockholm, Sweden, approved the study. The samples were used in accordance to informed consent given by included patients.

The first validation cohort (TCGA-AML) includes 142 AML patients with intensive induction treatment. Clinical and mutational data were retrieved from the data portal of TCGA (<https://gdc.cancer.gov>) and [Supplementary Table 1](#) (available online) of the publication of the TCGA-AML study (9).

The second validation data set (GSE6891) includes 458 AML patients (age ≤ 60 years). Detailed clinical, cytogenetic, and molecular information is available in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo; accession No. GSE6891) and in the original study (10).

Outcome

In this study, we investigated overall survival (OS) as the primary outcome. Overall survival was measured as from the date of diagnosis to the date of death from any cause. Patients alive at last follow-up were censored. Patients subjected to allogeneic hematopoietic stem cell transplantation were censored at the date of transplantation. The median follow-up time was 263 days in the Clinseq cohort and 455 days in the TCGA cohort. Patients in the Clinseq cohort who had stem cell transplantation were censored at the date of transplantation; however, the TCGA doesn't provide this information. If not censored at the date of transplantation, the median follow-up time was 404 days in the Clinseq. The definition of OS and event-free survival (EFS) in the GSE6891 is described in their original publication (13).

Predictor Sets

We consider four classes of predictors: demographic parameters (hereafter referred to as D), karyotypes (K), somatic mutations (M), and whole-transcriptome gene expression profile (G) (Figure 1). Demographic parameters (D) include age, sex, etiology of AML (whether the AML is de novo or secondary or therapy related). Karyotypes (K) are common AML-related cytogenetic

aberrations. In the Clinseq cohort, the transcriptomic profile was acquired by RNA sequencing, and somatic mutations were characterized by targeted sequencing of a panel of 25 genes. The details about sample and library preparation, sequencing, and bioinformatics processing of mutation calling and RNA-seq are described in the [Supplementary Methods](#) (available online).

Prognostic Model

The overall design of this study is outlined in [Figure 1](#). We developed three in-house models (Clinseq-DK, M, and G) based on the four sets of predictors (predictor sets D and K were combined into a single model). The other three models that were evaluated are used as reference models: 1) the ELN risk classification system (1); 2) Li's model (3) based on a 24-gene expression signature; 3) Patel's model (5) based on an 18-gene panel of somatic mutations.

The Clinseq-DK, M, and G models were developed using an elastic net penalized Cox regression model (14). The prognostic score is predicted from the fitted regression model as a linear model (sum of the normalized RNA-seq counts weighted by their coefficients derived from the elastic net penalized Cox regression model).

The Clinseq-DK, M, and G and the three reference models (the ELN, Li's model, and Patel's model) were evaluated in the TCGA cohort. The details of the implementation of Li's and Patel's models are to be found in the [Supplementary Methods](#) (available online). After evaluation of model performance in the TCGA cohort, the best model, Clinseq-G, was further tested in the GSE6891 data set to further assess the generalizability of the model.

To evaluate the prognostic value of different types of data (D, K, M, and G) in the training cohort (Clinseq-AML), cross-validation was conducted 100 times in each data set. Each time, samples were randomly split into training (80%) and test sets (20%). The frequency of events in the training and test sets was kept even when sampling. In each cross-validation iteration, the model was fitted in the training set. The relative survival risk of samples in the test set was predicted using the fitted model.

The Clinseq-G model was further evaluated in the microarray-based GSE6891 data set; see the [Supplementary Methods](#) (available online) for details.

Assessment of Prognostic Model Performance

To assess the performance of the prognostic prediction models, we used four measurements: 1) concordance index (C-index) (15,16), 2) integrated area under the curve (iAUC) of the time-dependent receiver operating characteristic (ROC) curve, 3) integrated Brier score (iBS), and 4) hazard ratio (HR) estimates in the univariate Cox regression model.

The iAUC of the time-dependent ROC curve (15) from six to 36 months was calculated as a measurement of discrimination. A larger iAUC indicates a better predictability of survival. C-index is another measurement of discrimination, which is the probability that a randomly selected person with the event will have a higher predicted risk than a randomly selected person without the event (15). The iBS is an integrated measurement of the mean squared error over time (16,17). In contrast to iAUC and C-index, the lower the iBS is, the better the average predictability of OS will be. The iAUC, C-index, and iBS were calculated using R package `survcomp` (18). Under each model, the patients were further dichotomized to high- and low-risk groups at the median. Univariate and multivariable Cox proportional hazards regression models were fitted at a time-on scale. Variables included in the multivariable Cox regression model were age, sex,

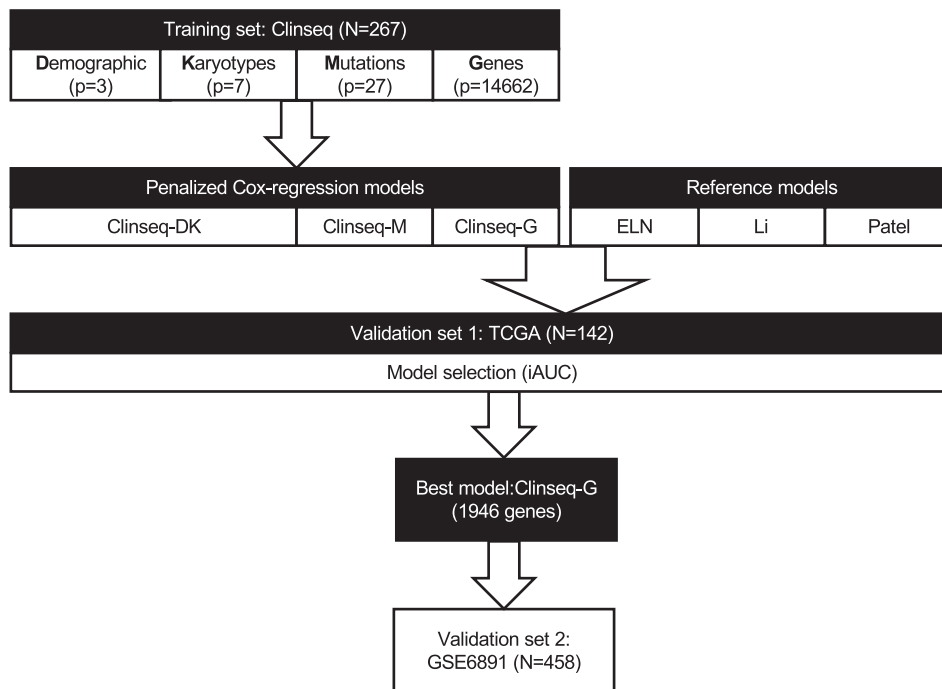


Figure 1. Overall design of the study. The prediction model was developed in the training set, the Clinseq, with 267 acute myeloid leukemia (AML) patients. There are four types of predictors: D, the demographic parameters (age and sex); K, karyotypes (common cytogenetic aberrations in AML); M, mutations (27 somatic mutations); and G, genes (the normalized RNA-seq counts of 14 662 genes). Penalized (elastic-net) Cox regression models were fitted 1) Clinseq-DK, in the predictor sets D and K; Clinseq-M, in the predictor set M; Clinseq-G, in the predictor set G. Predictor sets D and K were combined to build one model. The prognostic scores of The Cancer Genome Atlas cohort were predicted by each model (Clinseq-DK, Clinseq-M, Clinseq-G). There are three models used as reference models: 1) the European LeukemiaNet risk classification system; 2) Li's model based on a 24-gene expression signature; 3) Patel's model based on an 18-gene panel of somatic mutations. The performance of the models was evaluated by the integrated area under the curve, integrated Brier score, concordance index, and hazard ratio of risk groups (dichotomized at the median prognostic score). The Clinseq-G model provided the best performance. In order to further investigate the generalizability, the Clinseq-G model was tested in the GSE6891 data set. ELN = European LeukemiaNet; iAUC = integrated area under the curve; TCGA = The Cancer Genome Atlas.

etiology (de novo, secondary, or therapy-related AML), stem cell transplantation, percentage of bone marrow blast, white blood cell count, peripheral blood blast count, and mutational status of *NPM1*, *FLT3-ITD*, and *CEBPA*. Proportional hazards assumptions were checked using Schoenfeld residuals. The survival analysis was conducted by the R package survival (19).

Bayes factor (20,21) was calculated to compare the prognostic performance between models, following the method described by Guinney et al. (22). According to Kass et al.'s guidance (21), a Bayes factor from 1 to 3 indicates information hardly worth mentioning; from 3 to 20 indicates positive evidence; 20 to 150 indicates strong evidence; and greater than 150 indicates very strong evidence.

Additional Statistical Analyses

The Student *t* test or Kruskal-Wallis test was conducted for continuous variables, and the χ^2 test was conducted for categorical variables to compare baseline characteristics of participants between cohorts. All tests were two-sided, with *P* values of less than .05 considered statistically significant.

Results

Study Population

The basic characteristics of patients, including demographics, clinical parameters, cytogenetic aberrations, and somatic mutations in the Clinseq, TCGA, and GSE6891 cohorts, are listed in

Supplementary Table 1 (available online). To investigate whether the Clinseq is representative of the AML population in general, we compared the Clinseq in terms of clinical characteristics and survival outcomes with patients diagnosed as AML in Sweden (SALR) during the recruiting period (1997–2014) (Supplementary Table 2, Supplementary Figure 1, available online). There were no statistical differences found in any of the measurements. This indicates that the Clinseq cohort can be seen as a representative sample from the Swedish AML patient population in general.

Evaluation of the Prognostic Value of Demographic, Cytogenetic, Genetic, and Transcriptomic Data

In order to evaluate the relative prognostic value of different types of predictors, including demographic variables (age, sex, and performance status) (D), cytogenetic (K), somatic mutations (M), and transcriptomics (G), cross-validation was conducted in the training set (Clinseq-AML). The predicted risk score of overall survival in each cross-validation round was evaluated by several methods assessing the performance of prognostic scores such as iAUC, iBS, and C-index (Figure 2A). Predictor set G (transcriptomic profile) provided the best performance, especially in respect to iAUC (of note: for iAUC and C-index, the higher the better, while for iBS, the lower the better). Supplementary Table 3 (available online) presents the mean of iAUC, iBS, and C-index for the prediction of overall survival.

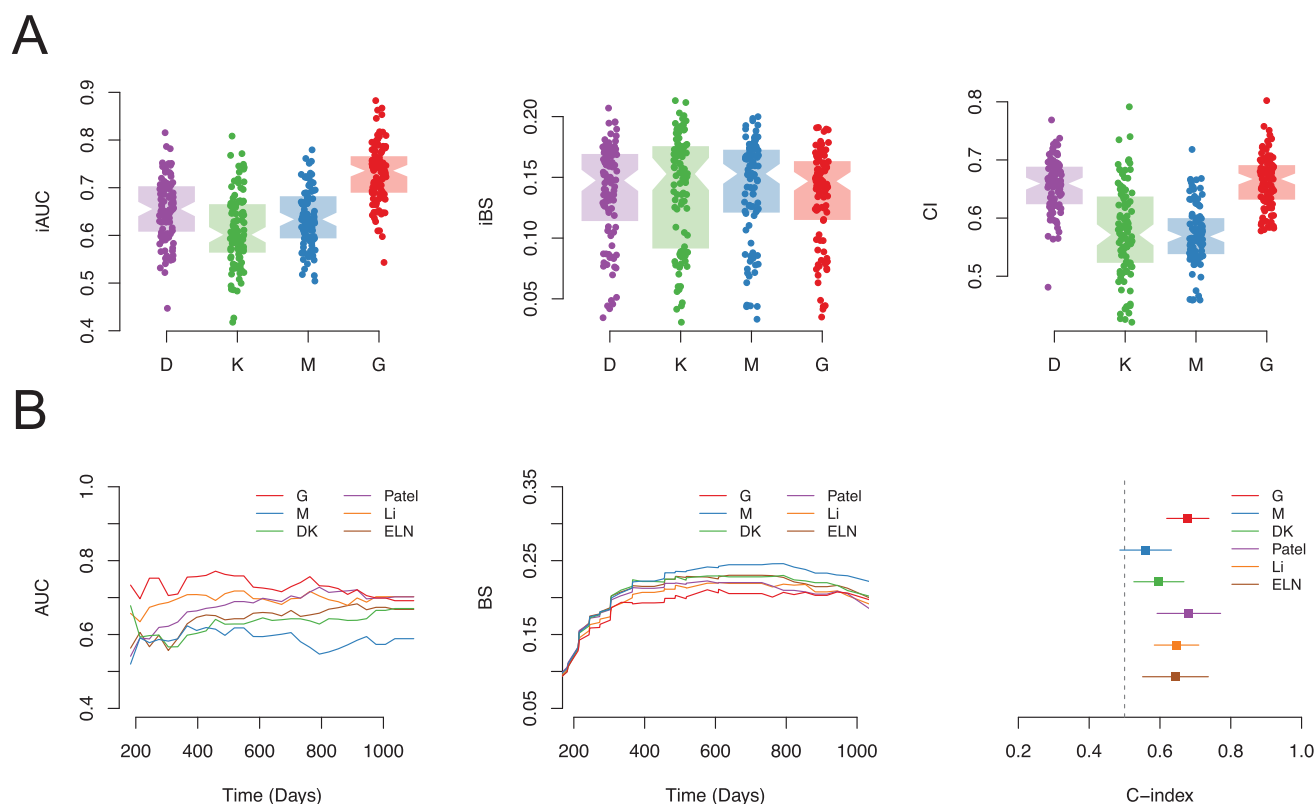


Figure 2. Evaluation of the prognostic value of different data types. **A)** Prognostic value of different types of data (D, demographic; K, karyotypes; M, mutations; and G, genes) in the Clinseq data set, evaluated by 100-round cross-validation (CV), measured by integrated area under the curve (iAUC), integrated Brier score (iBS), and concordance index (C-index). In each round of CV, 80% of patients were randomly selected as the training set. The iAUC, iBS, and C-index were assessed in the rest of the patients predicted by the fitted model. Each dot in the boxplot represents the iAUC/iBS/C-index measured in one round of CV. **B)** Prognostic score validation in the The Cancer Genome Atlas-AML cohort, predicted by the Clinseq-DK, M, G models, compared with the reference model (the European LeukemiaNet, Patel's model, and Li's model) in terms of the AUC of time-dependent receiver operating characteristic curve, Brier score, and C-index. AUC = area under the curve; BS = Brier score; CI = confidence interval; ELN = European LeukemiaNet; iAUC = integrated area under the curve; iBS = integrated Brier score.

RNA Sequencing-Based Overall Survival Prediction Model

We fitted penalized Cox regression models (Clinseq-DK, M, and G models) on overall survival in the Clinseq cohort based on the different types of data separately. Demographic parameters (age and sex; D) and cytogenetic aberrations (K) were combined into one feature set. There were 1946 genes included in the RNA sequencing-based prediction model. The predictors and corresponding coefficients of the Clinseq-DK, M, and G models are listed in [Supplementary Tables 4–6](#) (available online).

External Model Evaluation in the TCGA Cohort

The prediction models (Clinseq-DK, M, G) were subsequently tested in the TCGA-AML cohort. The prognostic scores predicted by the models, listed in [Supplementary Table 7](#) (available online), were compared in terms of iAUC, iBS, C-index, and hazard ratio of dichotomized risk groups ([Figure 2B](#) and [Table 1](#)). The ELN risk classification and prognostic risk score/groups developed by two previous publications, Li (3) and Patel (5), were also compared with our models. These models were the top two best models based on a previous systematic evaluation of prognostic performance in AML (8).

Model Clinseq-G, which is the penalized Cox regression model based on the RNA-seq data set, shows the highest time-dependent AUC under the ROC curve (0.73; C-index = 0.68, 95%

confidence interval [CI] = 0.62 to 0.74) and second lowest iBS (0.17, the lowest is 0.16), which means a good discrimination and overall prediction accuracy ([Table 1](#)). The iAUC of Clinseq-G is 0.73, with a Bayes factor greater than 4 compared with all the other models, indicating positive support for the Clinseq-G model. The Bayes factor of Clinseq-G compared with the ELN (reference model) is 26.78, indicating strong evidence in favor of the Clinseq-G model. Besides the C-index presented here, we also investigated the time-dependent C-index proposed by Uno et al. (23). The distribution of different prediction models is similar to the AUC of time-dependent ROC curves and Brier scores ([Supplementary Figure 2](#), available online). Dichotomizing patients in the TCGA cohort based on the prognostic score generated by the Clinseq-G model, the hazard ratio comparing the high-risk with the low-risk group was 2.42 (95% CI = 1.51 to 3.88). To compare categorized risk groups with the ELN and Patel's model, which have three levels, the patients were divided into tertiles based on the prognostic score of Li's model and the Clinseq-DK, M, and G models, respectively ([Supplementary Figure 3](#), available online). The Kaplan-Meier curves of three risk groups defined by the Clinseq-G model show statistically significant differences in terms of the probability of overall survival without overlapping ([Supplementary Figure 3](#), bottom right, available online). In multivariable analysis ([Table 2](#)), after adjusting for age, sex, percentage of bone marrow blast, white blood cell count, peripheral blood blast count, and the ELN risk classification, the Clinseq-G is still

Table 1. Validation of the prognostic value of Clinseq-DK, M, G models in the TCGA cohort

Model	C-index (95% CI)	Integrated Brier score	iAUC*	HR (95% CI)	Bayes_factor (x vs ELN3 as reference)	Bayes_factor (Clinseq-G vs other model)
ELN	0.64 (0.55 to 0.73)	0.19	0.65	NA	NA	26.78
Li	0.65 (0.59 to 0.72)	0.16	0.70	2.87 (1.77 to 4.67)	6.09	4.41
Patel	0.68 (0.59 to 0.77)	0.17	0.68	NA	8.90	5.94
Clinseq-DK	0.60 (0.53 to 0.67)	0.18	0.63	1.80 (1.12 to 2.87)	0.63	34.71
Clinseq-M	0.56 (0.49 to 0.64)	0.20	0.59	0.93 (0.59 to 1.47)	0.13	124.00
Clinseq-G	0.68 (0.62 to 0.74)	0.17	0.73	2.42 (1.51 to 3.88)	26.78	NA

*Integrated area under the curve of time-dependent receiver operating characteristic. C-index = concordance index; CI = confidence interval; ELN = the European LeukemiaNet risk stratification system; HR = hazard ratio; iAUC = integrated area under the curve; NA = not applicable; TCGA = The Cancer Genome Atlas.

Table 2. Multivariable overall survival analysis in the TCGA cohort, Clinseq-G model

Factor	HR (95% CI)	P*
Prognostic score†	2.34 (1.30 to 4.22)	.005
Age	1.01 (0.99 to 1.03)	.20
Male sex	0.76 (0.46 to 1.26)	.29
Percentage bone marrow blast	1.01 (1.00 to 1.03)	.08
White blood cell	1.01 (1.00 to 1.01)	.009
Peripheral blood blast count	1.00 (0.99 to 1.01)	.46
ELN (intermediate)	1.48 (0.77 to 2.83)	.24
ELN (high)	3.05 (1.42 to 6.58)	.004
Stem cell transplant	0.44 (0.25 to 0.76)	.003

*Two-sided likelihood ratio test. CI = confidence interval; ELN = the European LeukemiaNet risk stratification system; HR = hazard ratio; TCGA = The Cancer Genome Atlas.

†Prognostic score predicted by the Clinseq-G model.

statistically significantly associated with overall survival (HR = 2.34, 95% CI = 1.30 to 4.22, $P = .005$) (Table 2).

External Validation of the Clinseq-G Model in a Microarray-Based AML Data Set

To further validate the prognostic value of the Clinseq-G model, we applied it to a microarray-based gene expression data set, GSE6891. There are 458 AML patients in the GSE6891 data set. The median OS for the total cohort was 1.67 years with a standard deviation of 4.67, and the median follow-up for survivors was 10.13 years with an SD of 3.39. The predicted scores are listed in Supplementary Table 8 (available online). The iAUC of the prognostic score predicted in the GSE6891 is 0.66, the iBS is 0.21, and the C-index is 0.61 (95% CI = 0.58 to 0.65). The prognostic score is statistically associated with OS and EFS (HR_{OS} = 3.56, 95% CI = 2.51 to 5.06; HR_{EFS} = 3.41, 95% CI = 2.45 to 4.75). Equally dividing the patients according to the prognostic score into three groups, Kaplan-Meier curves of OS and EFS show statistically significantly different trends, without overlap (Figure 3). After adjusting for age, sex, cytogenetic risk stratification, NPM1, FLT3-ITD, and double mutation of CEBPA, the prognostic score is still statistically associated with OS and EFS (HR_{OS} = 2.02, 95% CI = 1.33 to 3.08; adjusted HR_{EFS} = 2.10, 95% CI = 1.42 to 3.12 (Table 3).

Evaluation of the Prognostic Value of the Clinseq-G Model in the Cytogenetic Normal Subgroup

Additionally, the prognostic score could further stratify the risk in cytogenetic normal AML (CN-AML) patients. In the GSE6891

cohort, there are 187 patients with CN-AML. Within the CN-AML subgroup, the prognostic value of the score is consistent with the whole cohort. The iAUC of the prognostic score predicted in the GSE6891 is 0.66, the iBS is 0.21, and the C-index is 0.62 (95% CI = 0.57 to 0.67). The prognostic score is statistically associated with OS and EFS (HR_{OS} = 3.62, 95% CI = 2.04 to 6.40; HR_{EFS} = 2.95, 95% CI = 1.72 to 5.05). Illustrated by Kaplan-Meier curves, the OS and EFS trends of risk groups were highly consistent within the whole cohort (Supplementary Figure 4, available online). However, the association between the prognostic score and OS in the TCGA CN-AML subgroup was not statistically significant, which might be due to lack of power as the number of patients is limited ($n = 64$).

Gene Enrichment Analysis of Clinseq-G Model Predictors

To determine which genes were individually prognostic, we performed Cox regression on OS with each gene. We found that 306 out of 14 662 genes (the whole Clinseq-G predictor set) were associated with overall survival (false discovery rate-adjusted $P < .05$); 262 of them (85.6%) are included in the 1946 predictors selected in the Clinseq-G prediction model (data not shown).

To investigate potential enrichment of particular molecular mechanisms in the 1946-gene panel in the Clinseq-G prediction model, we conducted gene ontology (GO) enrichment analysis (24). Genes were enriched in ribosome, hemoglobin complex, T cell receptor complex, and neuron projection membrane (Supplementary Figure 5 and Supplementary Table 9, available online). Many of these genes are annotated in both the hemoglobin gene families and the ribosomal protein gene families (Supplementary Figure 6, available online). It has been previously reported that ribosomal dysfunctions, for example, Schwachman-Diamond syndrome, Diamond-Blackfan anemia, dyskeratosis, and congenita, predispose to an increased lifetime risk of myelodysplastic syndromes and/or AML (25,26).

Discussion

Prognostic prediction is a challenging task because AML is a heterogeneous disease, especially in respect to clonal diversification and complexity. As a consequence, the ability to predict survival on the basis of routinely available clinical and cytogenetics variables, even with the addition of an increasing number of mutations, is relatively limited. In this study, we investigated the prognostic value relating to AML outcomes using multiple data modalities, including demographic, cytogenetic, somatic mutations, and transcriptomic gene expression profiles. The transcriptomic gene expression data were found to provide the strongest prognostic value. We propose a new prognostic

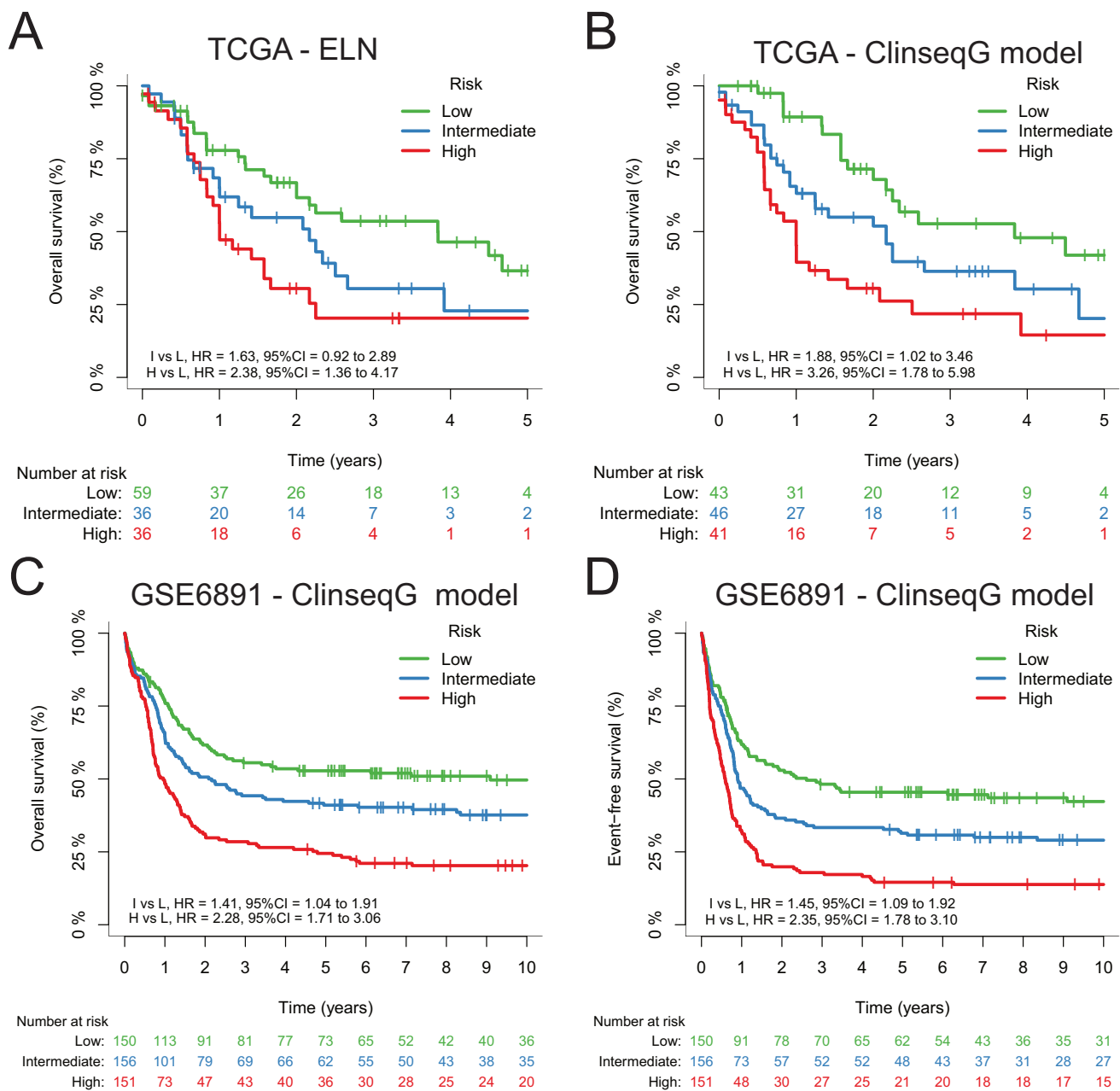


Figure 3. Kaplan-Meier curves of overall survival (OS) and/or event-free survival (EFS) in The Cancer Genome Atlas and GSE6891 data sets. **A)** Kaplan-Meier curves of OS by the European LeukemiaNet risk groups in The Cancer Genome Atlas (TCGA) cohort. **B)** Kaplan-Meier curves of OS by the Clinseq-G model risk groups in the TCGA cohort. **C and D)** Kaplan-Meier curves of OS and EFS by the Clinseq-G model risk groups in the GSE6891 data set. CI = confidence interval; ELN = European LeukemiaNet; HR = hazard ratio; TCGA = The Cancer Genome Atlas.

score, Clinseq-G, based on transcriptomic gene expression data, to predict and stratify AML patients' survival after diagnosis. The Clinseq-G prognostic score was validated in two independent cohorts, which provided evidence of generalizability and independent prognostic value of this risk score.

Multiple AML risk stratification models have been proposed previously (3–7). However, there are several advantages of the present study. First, the performance of the models was evaluated in a more comprehensive and systematic manner. We assessed the discriminability by multiple metrics, including iAUC and C-index, the overall performance by iBS, the generalizability by validation in independent data sets. Second, the

Clinseq-G prognostic score performed better than previously proposed models. Third, the Clinseq-G model could be successfully validated in two independent AML cohorts. We also believe that it is a strength that the prognostic score was developed in a population-based cohort consisting of older patients (median age [range] = 64 [18–84] years) using RNA sequencing, while the risk score could be validated in a younger (≤ 60 years) and microarray-based data set. Furthermore, the risk score was associated with both overall and event-free survival.

There are also some limitations that have to be acknowledged. First, not all potential clinical factors were included in this study, because of the incompleteness of the data on

Table 3. Validation of the Clinseq-G model in the GSE6891 cohort, multivariable survival analysis conducted by Cox regression models

Factor	Overall survival		Event-free survival	
	HR (95% CI)	P*	HR (95% CI)	P*
Prognostic score†	2.02 (1.33 to 3.08)	.001	2.10 (1.42 to 3.12)	<.001
Age	1.01 (1.00 to 1.02)	.06	1.00 (0.99 to 1.01)	.39
Male sex	0.94 (0.74 to 1.20)	.64	1.04 (0.83 to 1.31)	.72
Cytogenetic risk (intermediate)	2.49 (1.64 to 3.79)	<.001	2.15 (1.46 to 3.16)	<.001
Cytogenetic risk (poor)	3.73 (2.36 to 5.88)	<.001	3.40 (2.23 to 5.17)	<.001
NPM1 (positive)	0.53 (0.38 to 0.73)	<.001	0.52 (0.38 to 0.71)	<.001
FLT3-ITD (positive)	1.63 (1.23 to 2.17)	.001	1.58 (1.19 to 2.10)	.001
CEBPA (double mutated)	0.36 (0.18 to 0.72)	.004	0.42 (0.23 to 0.77)	.005
Stem cell transplantation	1.69 (1.28 to 2.22)	<.001	1.38 (1.07 to 1.79)	.01
White blood cell count	1.0019 (0.9998 to 1.0040)	.07	1.0024 (1.0005 to 1.0044)	.01
Percentage of bone marrow blast	0.9947 (0.9896 to 0.9999)	.048	0.9926 (0.9877 to 0.9975)	.003

*Two-sided likelihood ratio test. CI = confidence interval; HR = hazard ratio.

†Prognostic score was predicted by the Clinseq-G model.

variables across different cohorts. For instance, the minimal residual disease, which needs to be sampled at the time of complete remission, was not available in this study. Second, we applied the updated ELN (version 2017), except for one of the criteria, the *FLT3*-ITD allelic ratio. The status of *FLT3*-ITD was used instead of the allelic ratio. The *FLT3*-ITD ratio is derived from a particular genotyping method (1). However, no corresponding sequencing-based estimation method of the allelic ratio has been validated yet. Compared with the 2010 version (27), the updated ELN system successfully reclassified patients with intermediate risk to the high-risk category in the Clinseq and TCGA cohorts (Supplementary Figure 7, available online). Third, the prediction model was developed in the Clinseq cohort, which is representative of the overall AML population in Sweden. However, both the TCGA cohort and the GSE6891 cohort represent subpopulations of AML patients. In the case of TCGA, only de novo AML patients are included, and in the case of the GSE6891 cohort, only patients younger than age 60 years are included. Fourth, the follow-up time in the training set, the Clinseq cohort, is relatively short (median = 263 days). However, the prognostic score could be validated in the GSE6891 cohort, which had a longer follow-up time (median = 1.7 years). Last, we acknowledge that currently the relative cost of sequencing-based diagnostics does not motivate immediate clinical implementation. However, we anticipate that sequencing cost in just a few years will be manageable. Sequencing-based diagnostics also provide detailed molecular information that is likely to become increasingly important as precision medicine and targeted therapies become gradually more available, including molecular subtyping, targeted drug identification, and transplantation donor matching.

In conclusion, we present the first RNA sequencing-based prognostic score to predict survival after diagnosis of AML. The proposed risk score was validated in two independent cohorts and improves patient stratification compared with the risk classification system used in the clinical setting. Prospective clinical studies are required for further validation of the proposed prognostic score.

Funding

We acknowledge funding from Swedish Cancer Society (Cancerfonden), Swedish e-Science Research Centre (SERC;

e-Science for Cancer Prevention and Control), the Swedish Research Council (Vetenskapsrådet), the Strategic Research Programme in Cancer (StratCan) at the Karolinska Institutet, and the Stockholm County Council.

Notes

Affiliations of authors: Department of Medical Epidemiology and Biostatistics (MW, DK, HG, MR) and Department of Medical Epidemiology and Biostatistics, Science for Life Laboratory (JL), Karolinska Institutet, Stockholm, Sweden; Hematology Centre, Karolinska University Hospital, Stockholm, Sweden (CN, SL); Department of Medicine, Karolinska Institutet, Huddinge, Stockholm, Sweden (CN, SL); Department of Medical Sciences, Uppsala University, Uppsala, Sweden (SL).

The funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

We thank Dr. Peter J. M. Valk (Department of Hematology, Erasmus University Medical Center, Rotterdam) for providing the survival information for the GSE6891 data set.

The authors declare no conflict of interest.

References

- Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424–447.
- Komanduri KV, Levine RL. Diagnosis and therapy of acute myeloid leukemia in the era of molecular risk stratification. *Annu Rev Med*. 2016;67:59–72.
- Li Z, Herold T, He C, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: An international collaborative study. *J Clin Oncol*. 2013;31(9):1172–1181.
- Marcucci G, Yan P, Maharry K, et al. Epigenetics meets genetics in acute myeloid leukemia: Clinical impact of a novel seven-gene score. *J Clin Oncol*. 2014;32(6):548–556.
- Patel JP, Gonen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1079–1089.
- Metzeler KH, Hummel M, Bloomfield CD, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood*. 2008;112(10):4193–4201.
- Valk PJ, Verhaak RG, Beijnen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004;350(16):1617–1628.
- Wang M, Lindberg J, Klevebring D, et al. Validation of risk stratification models in acute myeloid leukemia using sequencing-based molecular profiling. *Leukemia*. 2017;31(10):2029–2036.

9. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059–2074.
10. Verhaak RG, Wouters BJ, Erpelinck CA, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94(1):131–134.
11. Wahlin A, Billstrom R, Bjor O, et al. Results of risk-adapted therapy in acute myeloid leukaemia. A long-term population-based follow-up study. *Eur J Haematol*. 2009;83(2):99–107.
12. Lazarevic V, Horstedt AS, Johansson B, et al. Incidence and prognostic significance of karyotypic subgroups in older patients with acute myeloid leukemia: The Swedish population-based experience. *Blood Cancer J*. 2014;4:e188.
13. de Jonge HJ, Valk PJ, Veeger NJ, et al. High VEGFC expression is associated with unique gene expression profiles and predicts adverse prognosis in pediatric and adult acute myeloid leukemia. *Blood*. 2010;116(10):1747–1754.
14. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Royal Stat Soc Ser B Stat Methodol*. 2005;67(2):301–320.
15. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56(2):337–344.
16. Graf E, Schmoor C, Sauerbrei W, et al. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17–18):2529–2545.
17. Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biom J*. 2006;48(6):1029–1040.
18. Schroder MS, Culhane AC, Quackenbush J, et al. survcomp: An R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*. 2011;27(22):3206–3208.
19. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Berlin/Heidelberg, Germany: Springer Science & Business Media; 2000.
20. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med*. 1999;130(12):1005–1013.
21. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90(430):773–795.
22. Guinney J, Wang T, Laajala TD, et al. Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: Development of a prognostic model through a crowdsourced challenge with open clinical trial data. *Lancet Oncol*. 2017;18(1):132–142.
23. Uno H, Cai T, Pencina MJ, et al. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105–1117.
24. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
25. Narla A, Ebert BL. Ribosomopathies: Human disorders of ribosome dysfunction. *Blood*. 2010;115(16):3196–3205.
26. Ruggero D, Shimamura A. Marrow failure: A window into ribosome biology. *Blood*. 2014;124(18):2784–2792.
27. Dohner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: Recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010;115(3):453–474.