[BIOINFO 2023]

Cellular abundance-based prognostic model associated with deregulated gene expression of leukemic stem cells in acute myeloid leukemia

Dong-Jin Han[1,2] and Tae-Min Kim[1,2,3,*]

[1]*Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea*

[2]*Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea*

[3]*Department of Biomedicine & Health Sciences, Graduate School, The Catholic University of Korea, Seoul, Republic of Korea*

**Corresponding author: tmkim@catholic.ac.kr*

While it has been previously reported that genes highly expressed in leukemic stem cell (LSC) may dictate the patients' survival probability and the expression-based cellular deconvolution may be informative in forecasting prognosis, it is still of debate whether the prognosis of acute myeloid leukemia (AML) patients can be predicted based on gene expression. Nine different cell type abundances of a training set, GSE37642 composed of 422 AML patients, were used to build a model for predicting prognosis by least absolute shrinkage and selection operator Cox regression. This model was validated in two different validation sets, TCGA-LAML and Beat AML (n = 179 and 451, respectively). We introduced a new prognosis predicting model for AML called LSC fraction (LSCF) score, which incorporates the abundance of five cell types, granulocyte-monocyte progenitor (GMP), common myeloid progenitor (CMP), CD45RA+ cell (RApos), megakaryocyte-erythrocyte progenitor (MEP), and multipotent progenitor (MPP). Overall survival probabilities between high and low LSCF score groups were significantly different both in TCGA-LAML and Beat AML cohorts (log-rank p-value = 3.3×10^{-4} and 4.3×10^{-3}, respectively). Also, multivariate Cox regression analysis on these two validation sets shows that LSCF scores are independent prognostic factors considering age, gender, and cytogenetic risk [hazard ratio (HR) = 2.17; 95% CI 1.40-3.34; $p < 0.001$ and HR = 1.20; 95% CI 1.02-1.43; $p < 0.03$, respectively]. LSCF scores showed comparable performance compared to other prognostic models of LSC17, APS, and CTC score, as estimated by area under the curves (AUCs). Gene set variation analysis with six LSC related functional gene sets indicated that high and low LSCF score is associated with upregulated and downregulated genes in LSCs. In short, we

developed a new AML patients' prognosis prediction scoring system, LSCF score, leveraging the deconvoluted cell type abundance with potential clinical relevance.