**Analysis Summary – Systematic identification of TFs associated with AML patient prognosis:**

In this analysis, we systematically calculated TF activity in 6 different AML datasets using the ENCODE ChIP-seq data. 292 ChIP-seq experiments from K562 (231 TFs) and 120 ChIP-seq experiments from GM12878 (101 TFs) were used to generate TF binding weight profiles from the TIP output. Specifically, the binding score of a TF to each gene (outputted by the TIP algorithm) was z-transformed and a one-sided z-test was carried out to generate p-values corresponding to each TF-gene binding interaction. P-values were -log10-transformed and trimmed at -10 or 10. Weight profiles were re-scaled by subtracting each value in a TF weight profile by the minimum and dividing by the range so that all values fell between 0 and 1. These weight profiles were used as input into the BASE algorithm to calculate TF activity scores for AML patient samples derived from the following gene expression datasets:

GEO -- GSE37642 (GPL\_96) (Herold, n=422)

NCI caArray -- willm-0019 (Wilson n=170)

GEO -- GSE14468 (Wouters, n=526)

Survival analysis was performed for each TF to identify those that were significantly associated with AML patient mortality. Namely, the TF’s iRASs (activity scores) across patient samples were used as the independent variable in a Cox proportional hazards model. A hazard ratio <1 indicates that a TF’s activity is associated with favorable prognosis and a hazard ratio of >1 indicates that a TF’s activity is associated with unfavorable prognosis in AML patient samples. Since a separate model was fit to each TF’s iRASs, p-values corresponding to the hazard ratios were adjusted for multiple hypothesis testing by using the Benjamini-Hochberg correction procedure.

In the results, we report the HR, P-value, and Adjusted P-value for each TF and their association with patient survival in each of the 3 AML gene expression datasets. The column labeled “number\_datasets\_significant\_P005” indicates the number of datasets in which the TF’s activity was observed to be significantly associated with AML patient prognosis at P<0.05. In particular, the EZH2, STAT1, and NR2C2 TFs were found to be significantly associated with prognosis in all 3 datasets. 15 other TFs were found to be significant in 2 datasets.