

TF_GSEA_distance explanation

The transcription factor activity acquired from “Viper_for_TFs” for each cell line is used as input to this algorithm. For all of the cell lines, the same pipeline is followed.

At the first step, the transcription factor activity for each compound that was calculated through VIPER is transformed to a ranked matrix, with descending order (rank 1 is the lowest activity and rank 279 is the highest activity).

Then, the expression set is built for every cell line, using the above mentioned ranked matrices, and then used as input for the GSEA score computation.

At the case of the HL60 cell line, an extra step exists in order for its compounds to be in the same order as in the gene expression GSEA, since they were formed through a binding and their order had changed (see “viper_for_TFs”). For all the cell lines, the compounds have to be in the same order for the comparisons that will follow.

The result of this algorithm is a distance matrix for every cell line, containing the GSEA distances of all the pairwise combinations of the unique compounds in each cell line. The distances of these matrices are (1211x1211) for MCF7, (1078x1078) for HL60 and (1161x1161) for PC3.