The example.r contains the illustration of using the HSSVD method (1) to bicluster the methylation data.

The main function of the software is “low_rank” with input:

(a) Matrix “X” : Data matrix to bicluster, if data range is [0, 1] then logit transform is recommended or [0, \infty] then log transform is recommended, otherwise no transform or standardization is needed.
(b) Vector “est” : Estimation methods for three steps can only choose “GB” or “Wold” method, default is “Wold” by (2).
(c) Vector “sparse” : Indicator that whether sparse input of left and right singular vector is needed, default is “FALSE”, see “Note” for insight.
(c) Integers “r1, r2, r3” : Rank input for three steps: rescale, mean approximation and variance approximation. Default is “NULL”, which means use the estimated rank.

Output:
(a) Matrix “mean_app” : Mean approximation of X.
(b) Matrix “std_app” : Standard deviation approximation of X.
(c) Object “rescale” : Sparse SVD approximation result of rescale, typically this type object contain singular value (d), left (u) and right (v) singular vectors.
(d) Object “result_mean” : Sparse SVD approximation result of Y.
(e) Object “result_var” : Sparse SVD approximation result of Z.
(f) Integers “k1,k2,k3” : rank input for three steps.
(g) Number ”bgmean” : Mean for null cluster.
(h) Number "bgstd" : Standard deviation for null cluster.
(i) Vector "back" : Index for null cluster.

Note: FIT-SSVD method (3) assumes that over half of the rows are approximate null rows with little or no signal and similar for the columns, their preselecting row and column steps for initial values and rank estimation will under-select informative rows/columns if all rows/columns are informative or the dimension of rows/columns are not large enough, both of which can happen in the biological research. So in practice the initial value would be the vanilla singular value decomposition’s result. For the rank approximation, the Gabriel-style “block” holdout tend to give a larger rank estimation than the Wold-style “speckled” holdout. **For our experience Wold method may be better for the large data set.** The rank estimation steps are slow, especially when using Wold-style type method, so we suggest to run in batch mode or run on clustering if rank estimation is needed.

**References**

