Supporting Information

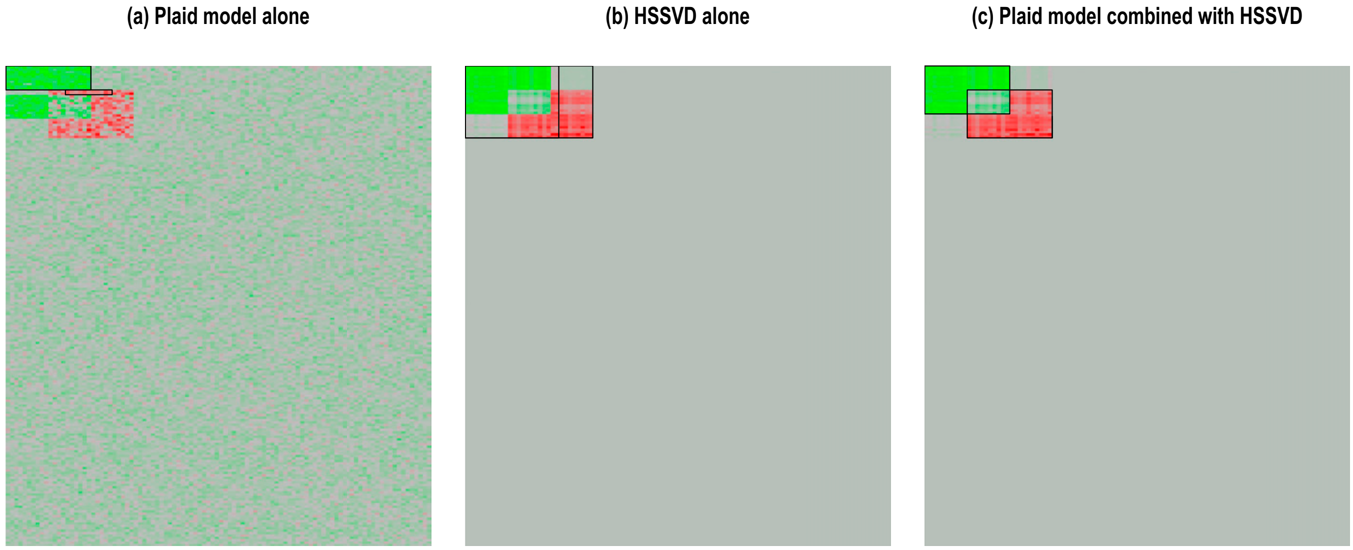
Chen et al. 10.1073/pnas.1304376110

SI Materials and Methods

Heterogeneous Sparse Singular Value Decomposition as a Denoising Procedure for Overlapped Biclusters Detection. In this section, we study overlapped mean bicluster detection through a simulated example. The plaid model is usually preferred to singular value decomposition (SVD)-based methods for overlapped mean bicluster detection, although the plaid model (1, 2) can be sensitive to variance heterogeneity. We want to show that our method as an SVD-based method is still useful for overlapped mean bicluster detection as a denoising step. First, we represent the data in a 200×100 matrix. The elements in the null cluster follow $N(0, 1^2)$. At the same time, we have two overlapping biclusters with their sizes both equal to 20×20. The elements in the two biclusters follow $N(7, 2^2)$ and $N(-5, 3^2)$, respectively, and the overlapped block size is 10×10. Hence, under the additive assumption, the elements in the overlapped block follow $N\left(2, \sqrt{13}^2\right)$. Here, we only focus on mean bicluster detection because this is the traditional purpose of biclustering methods. For comparison, we could directly apply the plaid model or the heterogeneous sparse SVD (HSSVD) method on the raw data to detect mean biclusters. Alternatively, we could first apply HSSVD to obtain a mean approximation of the raw data and then apply the plaid model to the mean approximation to detect biclusters. We use the "BCPlaid" function in the R package "biclust" (http://CRAN.R-project.org/package=biclust) as the implementation of the plaid model (1–3). The graphical results are presented in Fig. S1. The detected biclusters are highlighted by a black frame. From Fig. S1*B*, we see that although sparse SVD (SSVD) is good at mean signal approximation for nonoverlapping biclusters, it cannot recover the true overlapped bicluster structure. Meanwhile, we can see that after applying HSSVD, the plaid model (Fig. S1*C*) successfully picks out the underlying true structure, whereas applying the plaid model alone (Fig. S1*A*) was not successful. This result implies that for overlapped mean bicluster detection, the plaid model is generally better, but when there is variance heterogeneity present, the HSSVD can be quite helpful as a denoising process.

The Necessity of the Variance Detection Step in HSSVD. As we assume a low-rank structure in both mean signal and variance signal, a natural question to ask is whether such a structure can be approximated well by a higher-rank matrix for the mean structure only. In other words, can we represent the variance biclusters by using pseudomean biclusters? Our conclusion is that it is improper to use mean biclusters for variance bicluster detection. Pseudomean biclusters cannot recover the small variance biclusters at all, due to the natural shrinkage inherent in SSVD methods. Further, we will show that pseudomean biclusters can reveal some structure for the large variance biclusters; however, the approximation is rough. Consider the simulation data in example 1 (see Fig. S2 for graphical display). Here, we can compare the bicluster detection results of fast iterative thresholding for SSVD (FIT-SSVD) by ref. 4 (in Fig. S3) with input rank equal to 6 versus HSSVD with an estimated rank (in Fig. S5). We can see that there are four mean biclusters from Fig. S2. As the input rank is greater than the mean bicluster number, there will be several pseudomean biclusters (layer 5 and layer 6) in the FIT-SSVD result. For HSSVD, there will be both mean biclusters and variance biclusters. In Fig. S3, it appears that the pseudomean biclusters can detect part of the variance biclusters; however, this is because we know the "correct" order to display the graph. However, in practice, we would not know this order. Moreover, we can see that the pseudomean structure can be confounded with a type of true bicluster. For example, let $\mathbf{X}_0 = \mathbf{u}\mathbf{v}^T + \mathbf{\Phi}$, $\mathbf{u} = (\mathrm{rep}(1, 50), \mathrm{rep}(1, 50), \mathrm{rep}(0, 900))$, $\mathbf{v} = (\mathrm{rep}(1, 5), \mathrm{rep}(-1, 5), \mathrm{rep}(0, 90))$, where $\mathrm{rep}(1, 5) = (1, 1, 1, 1, 1)$ and $\mathbf{\Phi}$ is a 1000×10 matrix with entries following independent identically distributed (i.i.d.) $N(0,1)$. There is only a mean bicluster for X_0, and we can apply FIT-SSVD with input rank $= 1$. When we compare the mean bicluster result for X_0 and the pseudomean biclusters, layer 5 or layer 6 for example 1, graphically, the resulting heatmaps, given in Fig. S4 (rows and columns are reordered by hierarchical clustering) are very similar. This indicates that pseudomean biclusters can be confounded with certain true mean biclusters. This issue is probably even more complicated for real data settings. In contrast, HSSVD can provide more accurate large variance bicluster detection (layer 1 for variance) and small variance bicluster detection (layer 2 for variance), as shown in Fig. S5. Lastly, we want to emphasize that the bicross validation method (5) can be quite helpful for preventing the type of pseudomean detection which can weaken variance detection in the latter steps.

1. Lazzeroni L, Owen A (2002) Plaid models for gene expression data. *Statist Sinica* 12: 61–86.
2. Turner H, Bailey T, Krzanowski W (2003) Improved biclustering of microarray data demonstrated through systematic performance tests. *Comput Stat Data Anal* 48(2):235–254.
3. Kaiser S, Leisch F (2008) A Toolbox for Bicluster Analysis in R. *Compstat 2008–Proceedings in Computational Statistics* (Ludwig-Maximilians-Universität München, Munich).
4. Yang D, Ma Z, Buja A (2011) A sparse SVD method for high-dimensional data. arXiv: 1112.2433.
5. Owen AB, Perry PO (2009) Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Ann Appl Stat* 3(2):564–594.


(a) Plaid model alone

(b) HSSVD alone

(c) Plaid model combined with HSSVD

Fig. S1. Overlapped bicluster detection by the plaid model and HSSVD. (*A*) Original data are drawn and the plaid model detection result is highlighted with a black frame. (*B*) HSSVD detection result is highlighted with a black frame. (*C*) Mean approximation is obtained by HSSVD first and then the plaid model detection result is applied onto the mean approximation data.

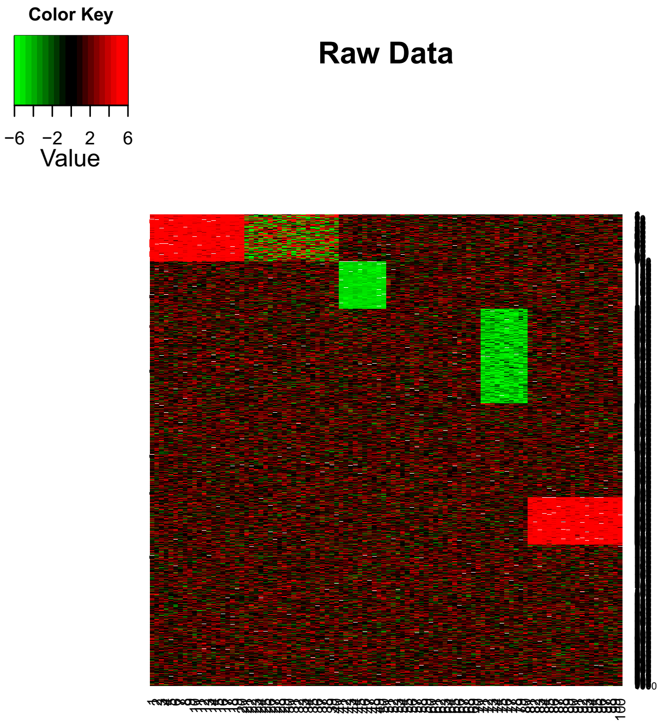


Fig. S2. Image of raw data for example 1. There are five biclusters. Red is for positive values and green is for negative values.

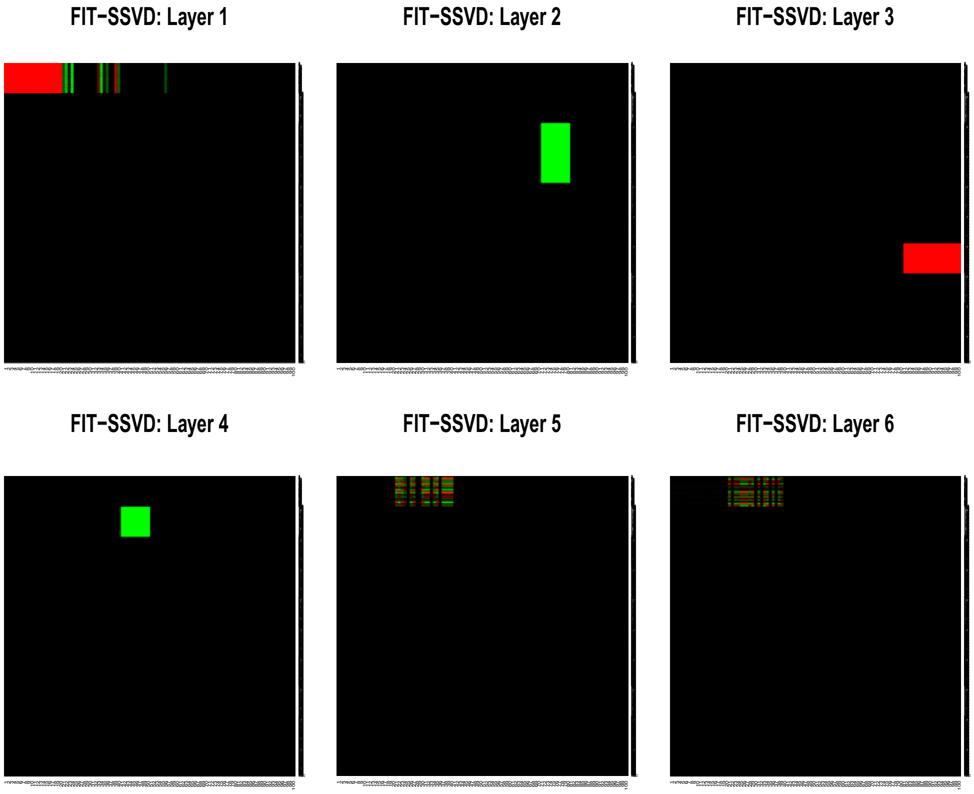


Fig. S3. FIT-SSVD for example 1. Each layer represents one bicluster. Layers 5 and 6 are pseudomean biclusters.

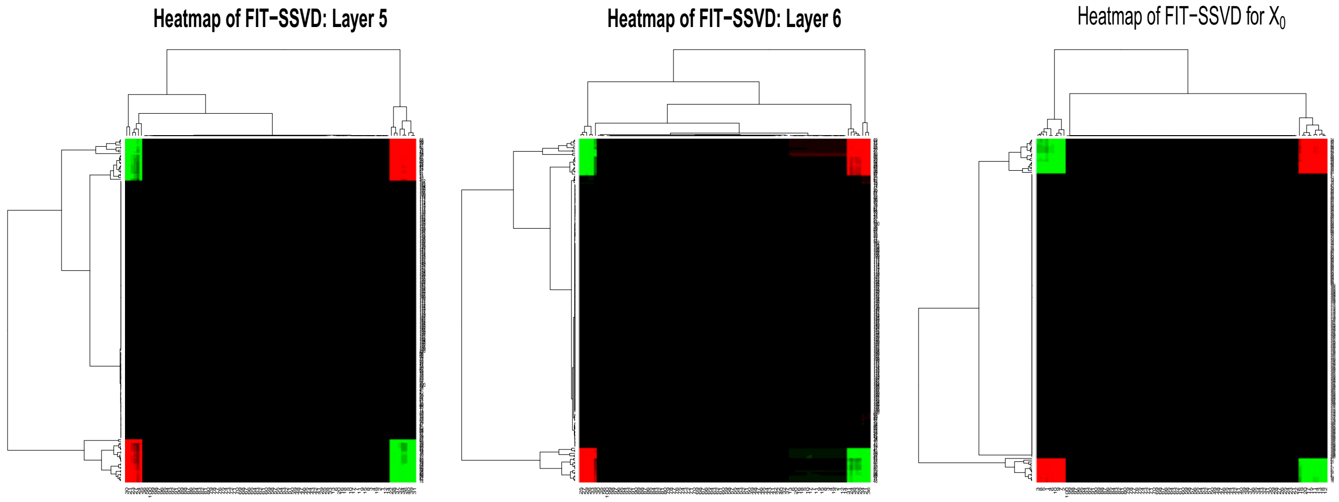


Fig. S4. Heatmaps of two pseudomean biclusters and a true mean bicluster. Rows and columns are ordered by hierarchical clustering. Only the first 200 rows (original order) are shown for better display (the remaining rows are all 0).

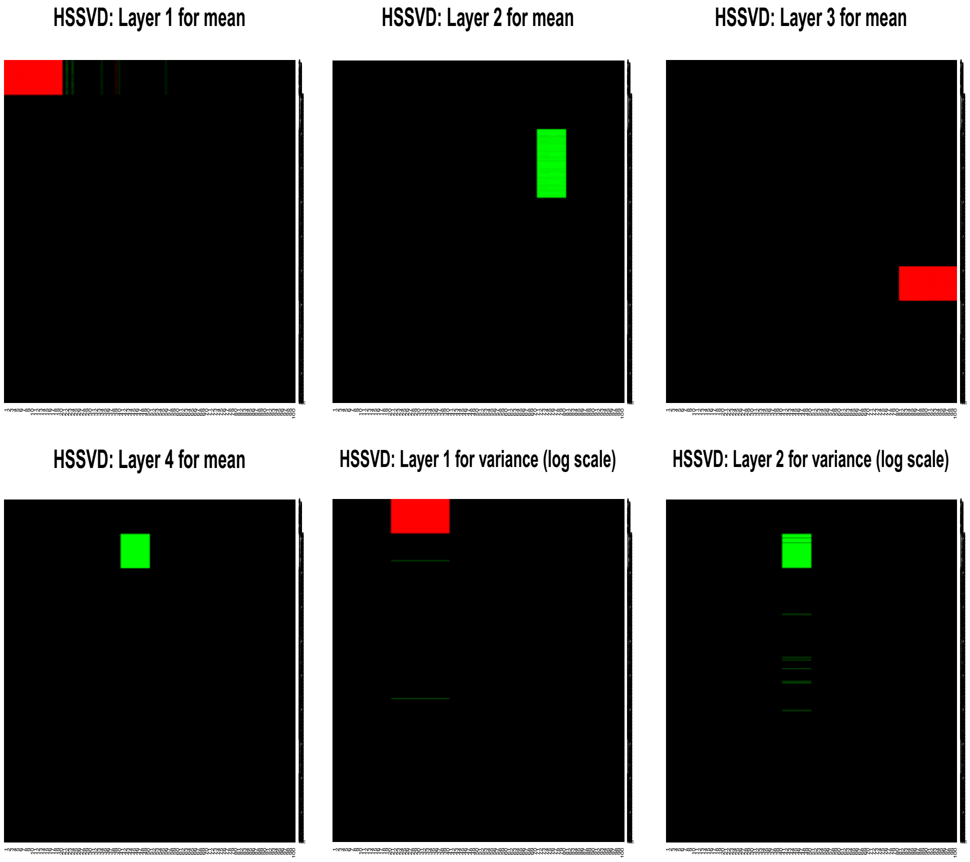


Fig. S5. HSSVD results for example 1. Each layer represents one bicluster. There are four mean biclusters and two variance biclusters.