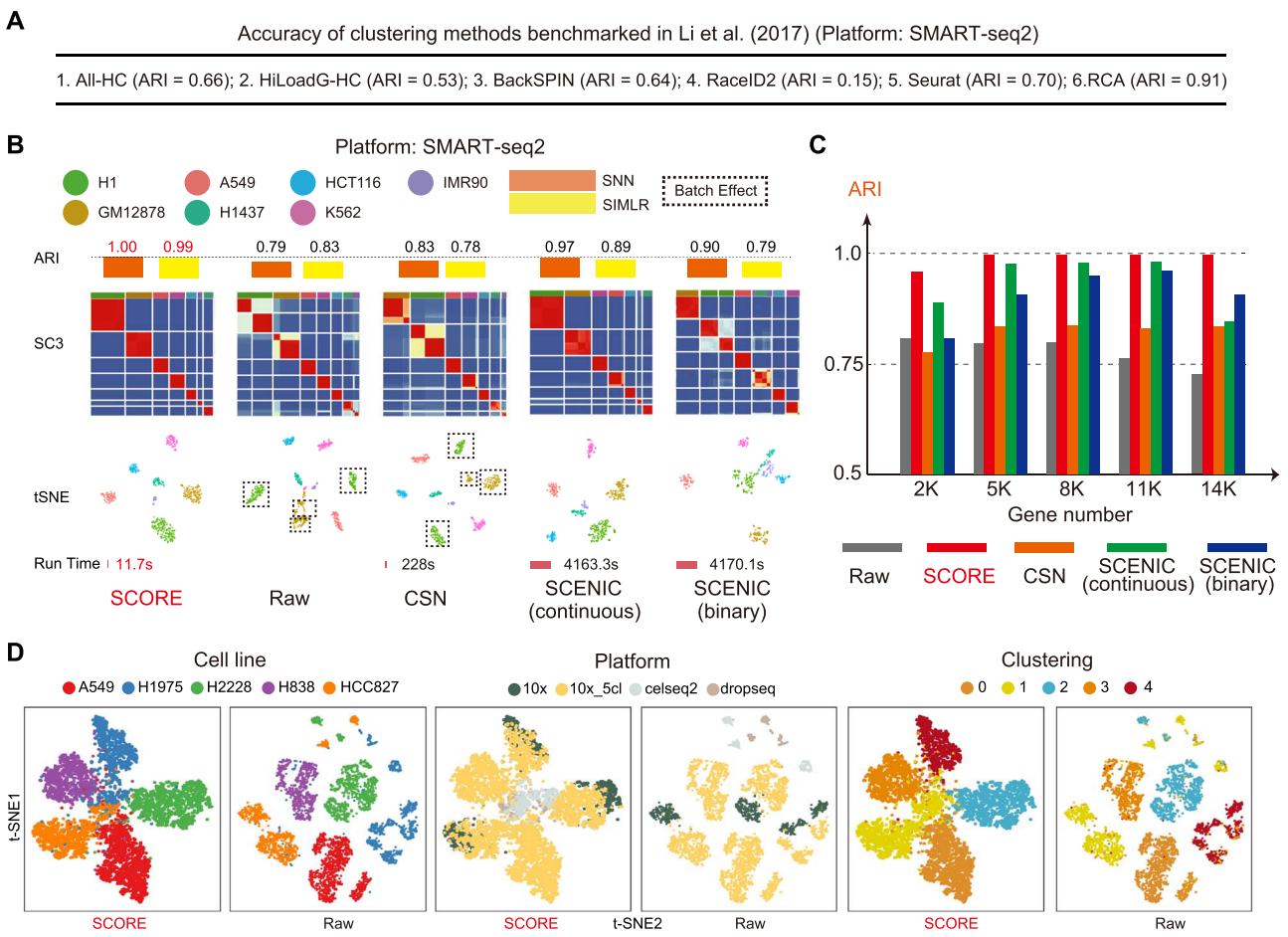

Figure 3. Integration of multi-platform datasets. (**A**) Accuracy of clustering methods evaluated on the gold-standard cell-line benchmarking dataset. (**B**) The performance of SCORE on the Smart-seq2 dataset composed of seven cell lines compared with the direct analysis on raw expression matrix within Seurat pipeline (denoted as Raw) and other GRN-based methods (i.e. CSN and SCENIC with continuous/binary features). SCORE eliminates the batch effects (denoted by dotted box) and outperforms other methods in terms of both clustering accuracy indicated by ARI and running time. (**C**) SCORE outperforms other methods in terms of clustering accuracy indicated by ARI regardless of the number of selected genes. (**D**) The performance of SCORE on the dataset with multiple platforms compared with the direct analysis within Seurat pipeline. The cells are colored by cell line information (the left column), platform information (middle) and clustering results based on integrated data (right), respectively.

with other state-of-the-art feature extraction methods based on gene regulatory network (GRN). The performance of SCORE was robust under the selection of input variable genes for analysis, and SCORE consistently outperformed other methods for the selected gene numbers ranging from 2000 to 14 000 (Figure 3B and C, Supplementary Figure 4, see Supplementary Data available online at http://bib.oxfordjournals.org/). Both the t-Distributed Stochastic Neighbor Embedding (t-SNE) plot and SC3 [31] similarity matrix demonstrated that the batch discrepancy was removed for the same cell lines (Figure 3B). In comparison, while SCENIC achieved satisfactory clustering results with 5000–11 000 genes, the clustering ARI was significantly reduced for the case of 14 000 genes due to the severe batch effects.

For the dataset with multiple platforms, there are 5319 cells derived from five human cell lines (Figure 3D) [30]. The direct clustering based on expression matrix within Raw pipeline confronted with limitations, by blurring the distinct cell identities and discriminating the unwanted experimental batches. In comparison, SCORE yielded a more reasonable result, in which the t-SNE plot showed that the batch effect was eliminated for the same cell lines, while the sharp distinction among different cell lines was still retained (Figure 3D). Overcall, application of SCORE to the gold-standard dataset validates our rationale: the incorporation of molecular interaction network in scRNA-seq analysis can facilitate the accurate cell identity dissection and reduce the technical artifacts in experiments.

Importantly, the implementation of SCORE is also efficient: with 14 000 top variable genes, SCORE finalized the analysis within 2 min, while the implementation of SCENIC lasted for 6 h in R (Figure 4A). Meanwhile, SCORE is also able to deal with the zero counts which are very common in scRNA-seq datasets (Figure 4B). We randomly replaced the expression value by zero with proportions from 10 to 90% in two scRNA-seq datasets [11, 32] and found that the ARI scores were still relatively high even if the zero count ratios were higher than 50%, suggesting the robustness of SCORE. Besides, with SCORE, we could obtain not only the differentially expressed genes (DEGs) but also the differentially activated modules (DAMs), and their dinstictive activities in different cell states supported the clustering accuracy of SCORE (Supplementary Figure 5A, see Supplementary Data available online at http://bib.oxfordjournals.org/). We could then construct the CMIN of a certain cluster to further explore the relationships and interactions of crucial genes (Figure 4C