Electronic Supplementary Material (ESI) for Molecular Omics.

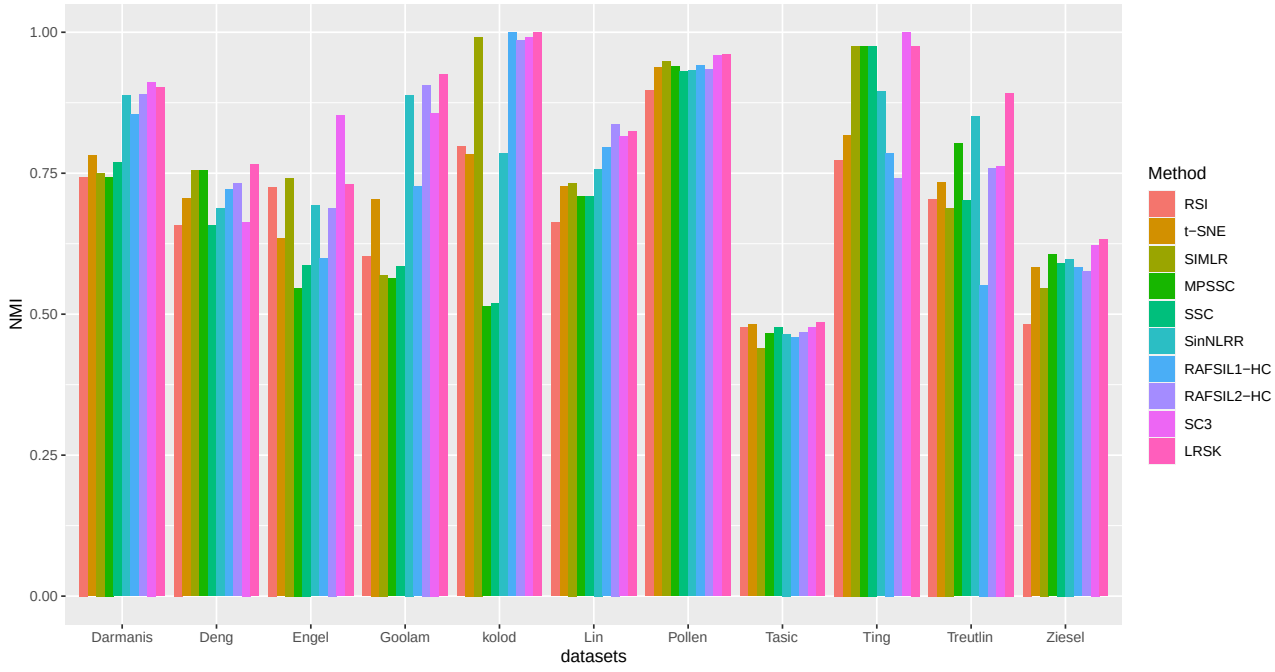

Figure 1: Comparison results of ten methods based on evaluation standard NMI under eleven simple cell RNA-sequencing data

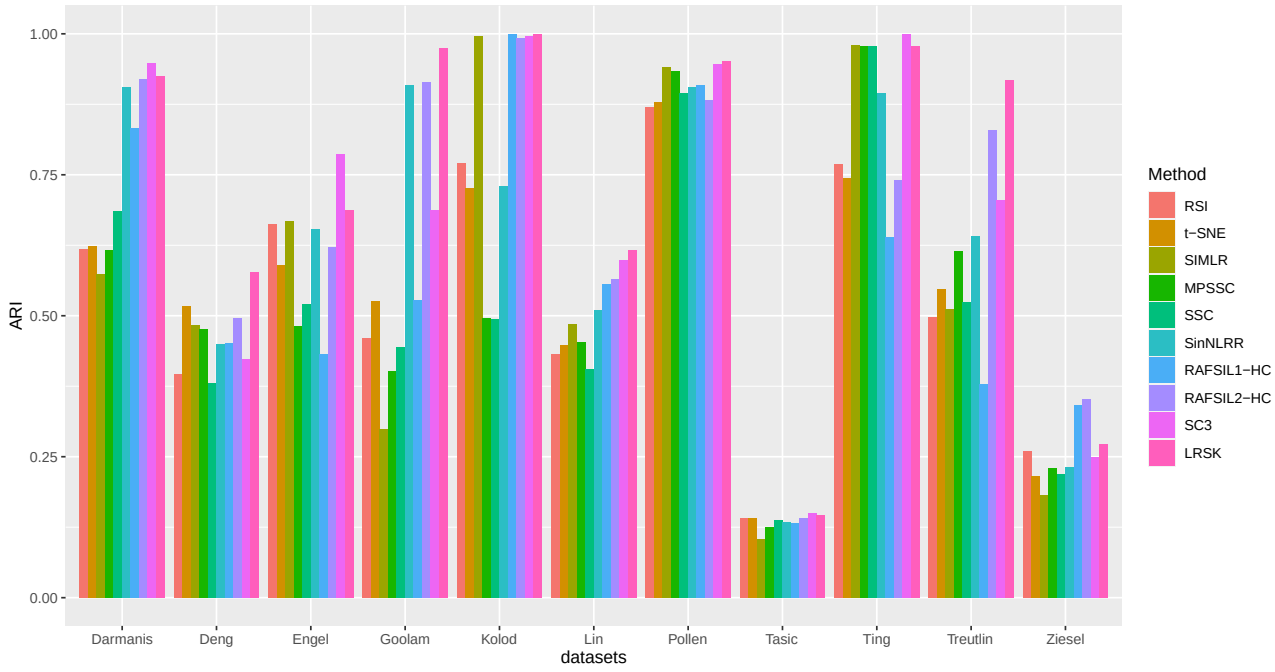

Figure 2: Comparison results of ten methods based on evaluation standard ARI under eleven simple cell RNA-sequencing data

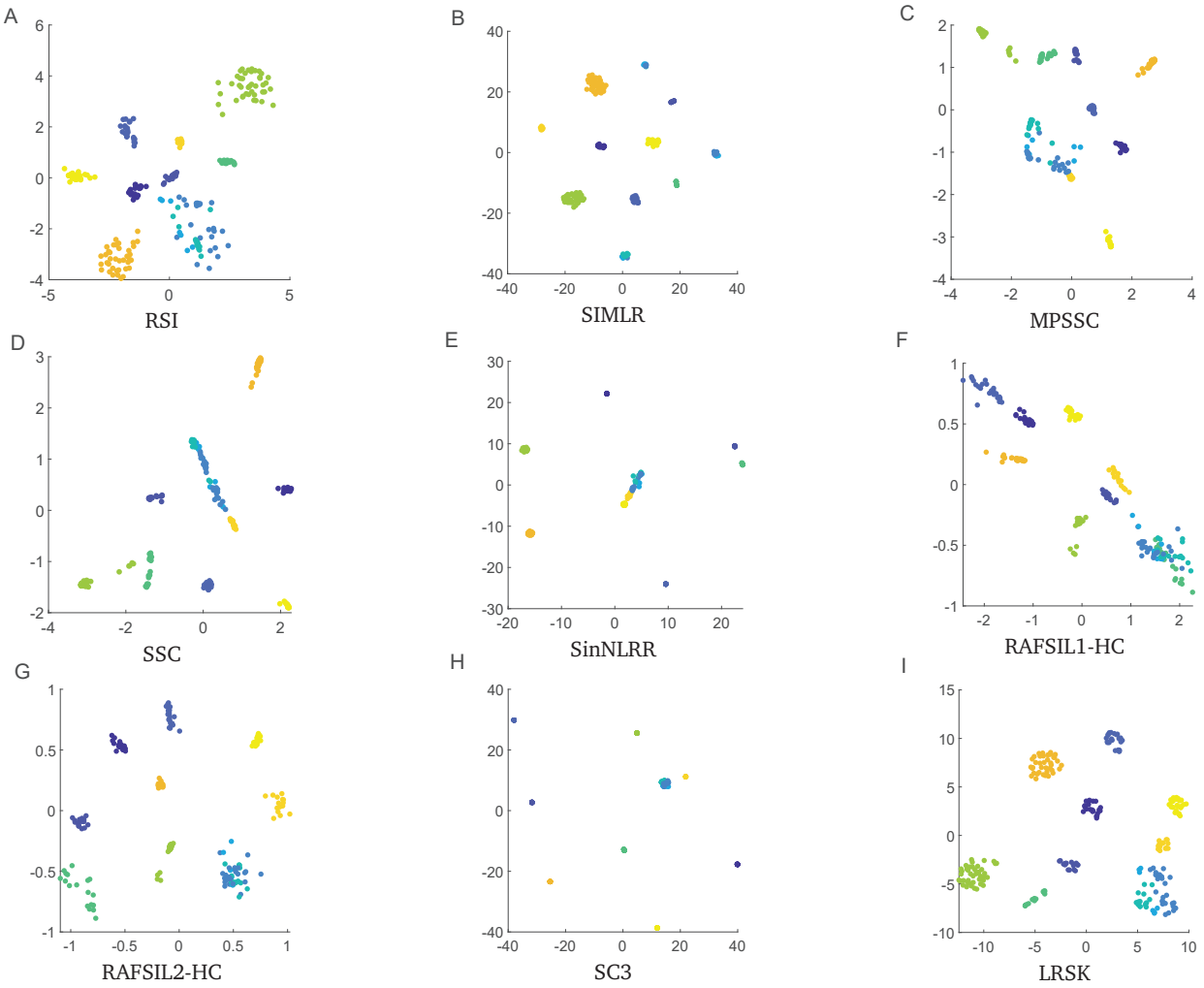

Figure 3: Visualization comparison of cells with the clusters located in low-dimensional space on case data Pollen. The graphs(A-I) of data points show the clustering performance of putting similar cell populations together for RSI, SIMLR, MPSSC, SSC, SinNLRR, RAFSIL1-HC, RAFSIL2-HC, SC3 and LRSK through improved t-SNE, where each point represents a cell and eleven distinct colors represent eleven different cell types.

Table 1: Comparison of the NMI, Purity and ARI achieved by LRSK and other clustering methods on a large scRNA-seq dataset of human livers

	RSI	t-SNE	SIMLR	MPSSC	SSC	RAFSIL1-HC	SC3	LRSK
NMI	0.7196	0.8010	0.7366	0.7122	0.7197	0.7184	0.7964	**0.8310**
Purity	0.7557	0.8202	0.7266	0.7425	0.7536	0.4516	0.7231	**0.8611**
ARI	0.5314	0.6348	0.5768	0.5380	0.5368	0.3618	0.6095	**0.7229**

[a]Normalized Mutual Information (NMI) (Strehl and Ghosh, 2003), Purity (Wagner and Wagner, 2007a), and Adjusted Rand Index (ARI) (Wagner and Wagner, 2007b).

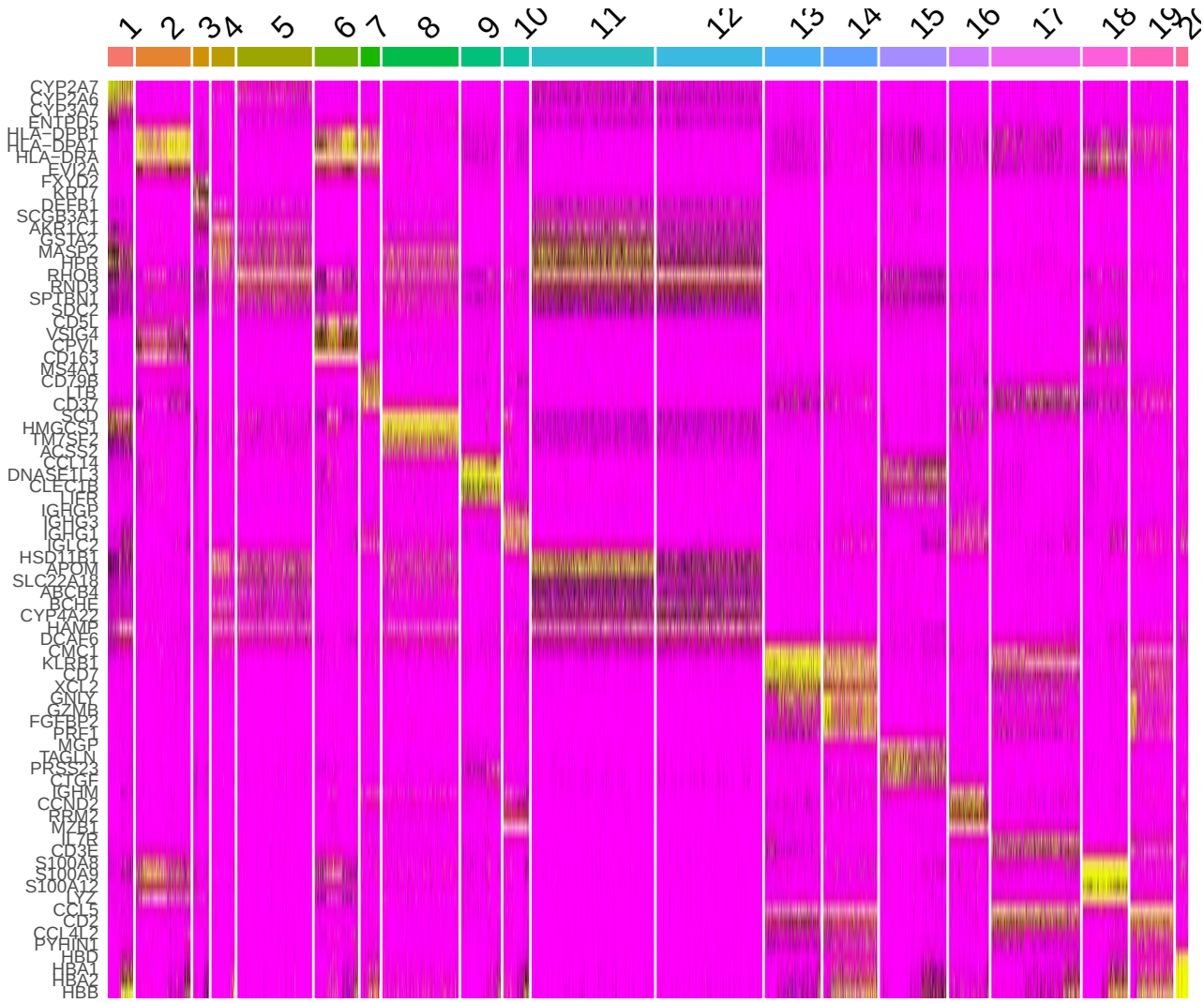

Figure 4: Heat map analysis using known marker genes expression in terms of the identified key differential genes

References

[1] Strehl, A. and Ghosh, J. (2003). Cluster ensemblesa knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, 583 – 617.

[2] Wagner, S. and Wagner, D. (2007a). Comparing clusterings- an overview. *Universit at Karlsruhe, Technical Report*.

[3] Wagner, S. and Wagner, D. (2007b). Comparing clusterings: an overview. *Universit at Karlsruhe, Universität Karlsruhe, Fakultät für Informatik Karlsruhe.*