Review of: "Visual Clustering of Transcriptomic Data from Primary and Metastatic Tumors: Dependencies and Novel Pitfalls"

Jia xingang

Potential competing interests: The author(s) declared that no potential competing interests exist.

In this work, the authors used t-SNE and UMAP methods to show that there was no visual association between resection site and the transcriptome for three considered metastatic datasets. Moreover, to avoid pitfalls in analyzing visual clustering and to enhance reproducibility, they recommend using standardized nomenclature, such as the UTMC-legend.However, there are some problems in the paper:
1.In fact,   these two dimension reduction methods were able to help us to select efficiently features. That is, for the different types of samples, these constructed features is reliable if their t-SNE projections have clear boundaries. Moreover, this work also verified that t-SNE (Figure 4 B+C) and UMAP (Figure 4 E+F) approaches based on log10 or the log10+1 transformed data(KIPAN) yielded a separation of samples matching the histopathologic expectation.

However, it needed to specify why these clusters were not exclusively subgroup-specific and displayed certain outliers, such as noise or distribution of these samples.

2.Furthermore, these two dimension reduction methods were also used to visualize the clustering results were not robust that were generated by selected features. That is, for the different types of samples, these constructed features was not reliable if their t-SNE and UMAP projections have mixed boundaries. In fact, this work also verified that visualization techniques (t-SNE vs. UMAP) and data transformation (unprocessed vs. log10 vs. log10+1 transformed data) heavily affected clustering results, and so on.

However, it needed to specify why these visualization techniques and data transformation affected clustering results, such as high dimension or distribution of these samples.In fact, we may choose part efficient features by some methods, such as F-scores.