

Sample-Set Enrichment Analysis

We proposed sample-set enrichment analysis (SSEA) to identify genes highly expressed in specific TME subtypes. We first sorted the samples by gene A expression to gain Ranked Sample List (RSL) of gene A. Similar to gene set enrichment analysis (GSEA), we calculated enrichment score (ES) for RSL in each TME subtype based on clusterProfiler [1]:

$$ES(i) = \sum_{a \in S \text{ \& } a \leq i} \frac{|R_a|}{\sum_{j \in S} R_j} - \sum_{a \notin S \text{ \& } a \leq i} \frac{1}{(N - N_S)}$$

$$ES = ES(x) \text{ when } |ES(x)| = \max(|ES(i)|)$$

Where R_a is expression value of sample a in RSL, S is sample set of one of the TME subtype, N represents number of samples in dataset and N_S represents number of samples in the sample set of one of the TME subtype. The p value was calculated by permutation test.

[1] G Yu, LG Wang, Y Han, QY He. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology 2012, 16(5):284-287. doi:[10.1089/omi.2011.0118](http://dx.doi.org/10.1089/omi.2011.0118)