

Gene set enrichment analysis (GSEA) aims to identify gene sets instead of individual genes that have a significant impact on diseases/phenotypes. Given a set of genes and two phenotypic classes, GSEA first ranks the individual genes by levels of differential expression on the two classes. Then GSEA calculates an enrichment score that reflects how the genes of a specific gene set are distributed across the ranked list. In the end, p-values are computed for estimating the significance of the enrichment scores, and adjustments for multiple hypothesis testing are conducted (FDR instead of FWER is used for the adjustment).

When the GSEA ranks the individual genes, it seems to me that the ranking makes an assumption of the independence for each gene. I feel concerned about whether it will lead to inaccurate estimations when the independence assumption does not hold.