

## Gene set enrichment analysis (GSEA) in single-cell RNAseq data using output from the Self Assembling Manifolds (SAM)

Alexander J. Tarashansky, Bo Wang

Department of Bioengineering, Stanford University

Contact: [tarashan@stanford.edu](mailto:tarashan@stanford.edu)

Github link: <https://github.com/atarashansky/self-assembling-manifold>

*This protocol was generated using SAMv0.7.4 and the Windows version of GSEAv4.1*

1. Run SAM analysis.
2. Save the resulting SAM weights (sorted in descending order) to a tabular '.rnk' file:  
``sam.adata.var['weights'].sort_values(ascending=False).to_csv('filename.rnk',sep='\t')``,  
where `sam` is the SAM object.
3. Open the '.rnk' file in a text editor and add a "#" to the first header row to comment it out.
4. Download and install GSEA v4.1 from [here](#).
5. Click on "Load data" in the GSEA interface.
6. Load the ".rnk" file generated previously.
7. Click on "RUN GSEAPreranked" in the GSEA interface.
8. Choose the Gene sets database of interest (e.g. "h.all.v7.1.symbols.gmt [Hallmarks]").
9. In the SAM study, we did not remap or collapse the gene symbols ("No\_Collapse"). Other datasets with different gene IDs may want to remap their IDs to standardized gene symbols. In this case, select "Remap\_Only" and choose the associated chip file (a table that translates each of your identifiers to a standardized gene symbol).
10. Choose parameters such as number of permutations, maximum set size, minimum set size, enrichment statistic, etc. We chose the "weighted" enrichment statistic, set the minimum set size to be 5, and used default parameters for everything else.
11. Click the 'Run' button.
12. The job should show up in the 'GSEA Reports Processes' panel. When it completes, click on the green 'Success', and an HTML report should show up.
13. From here, the upregulated and downregulated enrichment results can be downloaded in TSV format.