Welcome to SurfaceGenie!

***Integrating predictive and empirical data for rational marker prioritization***

SurfaceGenie is a tool for analyzing proteomic datasets to identify proteins of interest for immunophenotyping, immunotherapy, drug targeting, and other applications. It works by prioritizing the likelihood that a protein is informative for distinguishing among sample groups (*i.e.* cell types, experimental conditions). SurfaceGenie generates a score for each protein based on how likely it will be found on the cell surface, the number of samples it is observed in within a comparison set, and the magnitude of the measurement variable (*i.e.* relative abundance). While a major benefit of SurfaceGenie is the ability to prioritize molecules that are localized to the cell surface, it is also possible to analyze data without this parameter to find proteins of interest that reside in other subcellular localizations. SurfaceGenie works well with approaches that specifically identify cell surface proteins (*e.g.* Cell Surface Capture) and more generic approaches (*e.g.* analyses of whole cell lysate). The SurfaceGenie score is context-dependent, meaning that the tool will consider all data within a single dataset input (which may contain multiple experiments and/or cell types). If a user performs a comparison and subsequently determines additional data should be considered, a new file containing all data for the new comparison is required.

**If you use SurfaceGenie in your research, please cite the article:**

{ADD Reference and PUBMED LINK HERE}

SurfaceGenie Website Tools

SurfaceGenie

*Input:*SurfaceGenie accepts a .csv file containing a list of proteins (UniProt Accession) and a surrogate value representative of abundance (*e.g.* number of peptide spectrum matches, peak area) identified within a set of samples. There is no limit to the number of samples that can be analyzed in a single file.

*Data Processing:* SurfaceGenie calculates the dot product of three independent scores:

1. Surface Protein Consensus (SPC) score. A predictive measure of the likelihood that a particular protein can be present at the cell surface. This value is a sum of the number of predictive datasets for which a protein has been predicted to be localized to the cell surface. Scores range 0-4. For more details on the predictive datasets used, click here.
2. Distribution Score. A measure of how evenly or unevenly distributed a protein is among multiple samples within a comparison dataset. It is based on the Gini coefficient for calculating statistical dispersion of values. Scores range 0 – 1/(1-N).
3. Signal Strength. An approximate measure of protein abundance for cell types in which a protein is observed. Proteins at the lower limit of detection are of lower priority than those with more observations, because it is expected that those of higher abundance will practically serve as more accessible markers for downstream technologies. Scores typically range 0 - 4 .

*Output:*

* SurfaceGenie Plot: SurfaceGenie scores plotted in order of priority for all proteins in a dataset.
* SPC Histogram: Shows the distribution of SPC scores
* Clustered mthe relationship among samples within a dataset based on the relative abundance measurement contained in the .csv file.
* CSV Download: Columns of selected data types (*e.g.* SPC score, SurfaceGenie score, etc) are appended to the original input file.

*Customization:*

1. In the default mode, SurfaceGenie will incorporate the SPC Score. Hmay be ignored for analyses where surface localization is not required
2. can be retrieved without analyzing data through the SurfaceGenie tool. See

SPC Score Lookup

This feature enables users to obtain Surface Protein Consensus (SPC) score for proteins of interest without analyzing data through SurfaceGenie. Users may perform a batch retrieval by uploading a .csv file containing UniProt Accession numbers or may search individual UniProt accession numbers.

*Publication Info [Gundry Lab 2018]*

)