SURCOMED User’s Guide: Quick Start

SURCOMED is a method to identify cancer-driving mechanisms based on the search of gene expression survival signatures using tumor transcriptomic and clinical data from cancer patients. The usage of the SURCOMED web-based interface can be described in 5 steps:

1. Select the search space (see Figure 1)
2. Add constraints (see Figure 2)
3. Configure optimization (see Figure 3)
4. Run the search (see Figure 4)
5. Identify the underlying biological mechanisms (see Figure 5)

**Figure 1. Select the search space.** The search space is the list of genes eligible for the combinations. They can be copied and pasted in the entry box or loaded from a file. For a genome-wide search, click the radio button with the label "genome". Each gene symbol should be specified in a different line.
Figure 2. Add constraints. These constraints refer to gene states that will be avoided or included by the algorithm in every combination. Constraints can be manually added individually through the interface or loaded as a set from a file. In the latter case, the state can be either "high" or "low", and the syntax should be as follows: \texttt{gene\_name::state}

Example:
TNFSF9::low
IL2::high
CER1::low
Figure 3. Configure optimization. There are 5 optimization parameters that can be changed, namely, population size, selection number, elite size, iteration number, and combination size.

**Population size.** This parameter refers to the number of combinations that are going to be evaluated at each iteration of the algorithm. Large populations help to prevent the optimization from becoming trapped in a local optimum, but they are more time consuming.

**Selection number.** This parameter refers to the number of combinations that are going to be selected from the population after the evaluation of these combinations in survival terms. We suggest to use a selection number of half the population size.

**Elite size.** This parameter refers to the number of the best historical combinations across iterations that will be directly transferred from one generation to the next to prevent loss of combinations with the best score. The elite number should be smaller than the selection number. We suggest an elite number not higher than half of the selection number in order to provide a certain freedom to the optimization process.

**Iteration number.** This parameter refers to the number of rounds of combinations that are going to be generated, evaluated, and selected by the algorithm. We suggest to start with some exploratory runs with 3-5 iterations, changing the population size before performing an actual analysis with 20 or more iterations. The greater the number of iterations, the longer it will take the algorithm to finish the search.

**Combination size.** This parameter refers to the number of gene states included in each combination. On the website, this number is limited to a maximum value of 5.
4. Run the search
If an email is entered here, the user will receive a notification when the search is done with a link to the results.

Figure 4. Run the search. Once the optimization parameters are set up, the user only has to click on the ‘Search’ button and wait. We strongly recommend to complete the dedicated box with an email address, so the search will be running on our server and the user will receive a notification when the search is done with a link to the results. By doing so, the user can leave the website and follow the link later on to access the results.
Figure 5. Identify the underlying biological mechanisms. Once the search is complete, the best combinations are displayed on the website with their corresponding survival curves. Just below the survival curves, there is a link to perform the Gene Set Enrichment Analysis (GSEA), or, alternatively, to generate and download the required files in order to run the GSEA locally with personalized settings.

Once the GSEA is done, the results can be explored online using the Internet browser or may be downloaded (see Figure 5). The main result is the list of up- and down-regulated biological mechanisms, which can be visualized online in the browser or downloaded.
Figure 5. GSEA results main page. Results from GSEA can be easily explored following the links from the main page, which can be accessed online or downloaded and visualized using your Internet browser. These results include the list of up- and down-regulated biological mechanisms associated with predefined gene sets, and other complementary information concerning the GSEA such as enrichment plots and statistical scores (p-values, normalized enrichment scores, or normalized enrichment score (NES)).