



山东大学齐鲁医院  
QILU HOSPITAL OF SHANDONG UNIVERSITY

# Core Mathematical Function

## SlimR v1.0.6

Zhaoqing Wang

Qilu Hospital of Shandong University



## ➤ Parameter usage

- $N$ : Number of genes
- $C$ : Number of cell clusters
- $i \in \{1, \dots, C\}$ : Cluster index
- $g \in \{1, \dots, N\}$ : Gene index
- $x_{g,i}$ : Expression value of gene  $g$  in cluster  $i$
- $\mu_{g,i}$ : Average expression of gene  $g$  in cluster  $i$
- $\sigma_{g,i}$ : Standard deviation of gene  $g$  in cluster  $i$
- $f_{g,i}$ : Fraction of cells in cluster  $i$  where  $x_{g,i} > m$  (minimum expression threshold  $m$ )
- $w$ : Specificity weight parameter
- $\bar{\sigma}_i$ : Average standard deviation of all genes in cluster  $i$
- $\epsilon$ : Small constant to avoid division by zero

## Step 1: Specificity Score Calculation

For each gene  $g$  in cluster  $i$ :

$$s_{g,i} = \mu_{g,i} \cdot f_{g,i} \cdot \left( 1 + w \cdot \frac{\sigma_{g,i}}{\bar{\sigma}_i + \epsilon} \right)$$

- **Interpretation:**

- $\mu_{g,i}$ : Mean expression level.
- $f_{g,i}$ : Proportion of cells expressing  $g$ .
- $\frac{\sigma_{g,i}}{\bar{\sigma}_i}$ : Normalized variability of  $g$  compared to other genes in the cluster.
- $w$ : Amplifies the impact of high variability.

## Step 2: Normalization of Specificity Scores

Normalize  $s_{g,i}$  across genes per cluster  $i$ :

$$s'_{g,i} = \frac{s_{g,i} - \min_{g'} s_{g',i}}{\max_{g'} s_{g',i} - \min_{g'} s_{g',i}}$$

- **Purpose:** Ensures scores are comparable across genes within the same cluster.



### Step 3: Gene Weight Calculation

Compute weights for genes based on their variability-to-mean ratio:

$$g_w = \frac{\sigma_g}{\mu_g}$$

- **Where:**
  - $\sigma_g = \text{mean}(\sigma_{g,i} \text{ across all } i)$
  - $\mu_g = \text{mean}(\mu_{g,i} \text{ across all } i)$
- **Purpose:** Prioritize genes with higher variability and lower mean expression.

## Step 4: Cluster-Specific Gene Expression Score

Aggregate normalized scores  $s'_{g,i}$  into a final cluster score  $c_i$ :

$$c_i = \sum_{g=1}^N g_w \cdot s'_{g,i}$$

- **Interpretation:**

- $c_i$  reflects the weighted sum of gene-specificity scores for cluster  $i$ .
- Higher  $c_i$  indicates stronger evidence for the cluster being enriched in the target gene set.

## Final Output Matrix

For all clusters  $i \in \{1, \dots, C\}$  and genes  $g \in \{1, \dots, N\}$ , the function outputs a matrix  $R$  where:

$$R_{i,g} = s'_{g,i}$$

This matrix is transposed and row-normalized for visualization as a heatmap.



山东大学齐鲁医院  
QILU HOSPITAL OF SHANDONG UNIVERSITY

**Thanks!**

**Zhaoqing Wang**

**Qilu Hospital of Shandong University**

