webMCP-counter: a web interface for transcriptomics-based quantification of immune and stromal cells in heterogeneous human or murine samples

Maxime Meylan¹,², Etienne Becht¹,³, Catherine Sautès-Fridman², Aurélien de Reyniès¹, Wolf H. Fridman² and Florent Petitprez¹,⁴,*

¹ Programme Cartes d’Identité des Tumeurs, Ligue Nationale Contre le Cancer, Paris, France
² Centre Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, Université de Paris, Equipe inflammation, complément et cancer, Paris, France
³ Fred Hutchinson Cancer Center, Seattle, USA
⁴ MRC Centre for Reproductive Health, The University of Edinburgh, The Queen's Medical Research Institute, Edinburgh, UK
* Correspondence: florent.petitprez@ed.ac.uk

Abstract

Summary: We previously reported MCP-counter and mMCP-counter, methods that allow precise estimation of the immune and stromal composition of human and murine samples from bulk transcriptomic data, but they were only distributed as R packages. Here, we report webMCP-counter, a user-friendly web interface to allow all users to use these methods, regardless of their proficiency in the R programming language.


Keywords: Deconvolution, web application, MCP-counter

Introduction

In many areas of biology, it is crucial to quantify the composition of heterogeneous samples at the cell type level. For instance, tumors are highly heterogeneous systems harboring not solely cancer cells, but also a wide variety of immune and stromal cells¹. The extent and composition of this immune infiltrate has been repeatedly shown to be associated with patients’ survival² and response to treatment³. Bulk transcriptomics capture signals from all cell types in a sample and can therefore be used to estimate its composition through deconvolution methods⁴. Several methods have been reported to estimate the immune and/or stromal composition of human samples, and only a few for murine samples⁵.

Our group developed MCP-counter⁶ and its murine counterpart mMCP-counter⁷, which were distributed as R packages. However, using these methods requires proficiency in R programming. To open the use of these methods to all biologists, we report here the development of webMCP-counter, a user-friendly web interface for both MCP-counter and mMCP-counter. This tool can be accessed using this url: http://134.157.229.105:3838/webMCP/.

webMCP-counter

Our web interface, webMCP-counter, is composed of two main steps (Figure 1). The first step aims at running (m)MCP-counter per se. This step requires the user to upload a transcriptomics dataset that can be presented under various formats. The user also needs to specify several details regarding the dataset (file
format, gene ID format and organism). Either HUGO gene symbol or ENSEMBL ID are valid formats for genes. Depending on the organism, either MCP-counter (human data) or mMCP-counter (mouse data) is run, the scores are presented as a table that can be visualized within the tool or downloaded for other purposes.

The second step consists in a typical downstream analysis. The user is prompted to choose a number of clusters and the cell populations to include in the analysis among the populations available, which depends on the organism and the data provided. Then, a hierarchical clustering with standard parameters (Euclidean distance and Ward linkage criterion) is run on the (m)MCP-counter Z-scores limited to the requested populations is run. The output is displayed as a heatmap and individual violin plots detailing the differences between clusters for each selected population, and tested with a Kruskal-Wallis test followed by Benjamini-Hochberg-corrected Dunn test for pairwise comparisons, or a Mann-Whitney test for two clusters. The assignment of samples to clusters can be downloaded as a tabular file. Our tool is, to the best of our knowledge, the only one to offer to perform easily this downstream analysis from the MCP-counter output for users who are not proficient in the R programming language.

**Example on the TCGA SARC dataset**

For users who simply wish to test the possibilities offered by webMCP-counter, we integrated an example dataset from the TCGA SARC cohort. This dataset is composed of n=213 tumors of soft-tissue sarcoma patients analyzed by RNA-seq. By selecting the example dataset on the webMCP-counter (Figure 2, marker 1), all other parameters are internally set to the correct values, and the user-provided parameters are ignored.

On the first tab, users obtain a table with the MCP-counter scores for each sample in the TCGA SARC cohort (Figure 2, marker 2). This table is interactive and users can sort samples on increasing or decreasing order for each population. It is possible to search for one particular sample by entering its sample name in the search bar. It is important to note that scores should not be interpreted as absolute proportions of cell populations in the samples, but are instead expressed in arbitrary units, similar to gene expression values. However, they can robustly be used to compare the abundances of the cell populations between different samples, such as what has been done in several studies.

This table can be downloaded as comma-separated values table (.csv) for use in other software, such as Microsoft Excel (Figure 2, marker 3).

**Fig. 2.** Step 1 of the analysis of the TCGA SARC example dataset. Marker 1 indicates the box to tick to access the example dataset instead of a user-provided dataset. 2 indicates the MCP-counter scores table, accessible after MCP-counter has been run. Finally, 3 indicates the download link for the .csv file containing the full table.

On the Step 2 tab (Figure 3), users can perform a classical downstream analysis from the MCP-counter scores table, sample clustering based on their immune and stromal composition. To use this tab, the first step must be completed beforehand. In this tab, the user is prompted to select the populations of interest on which to perform hierarchical clustering (marker 1) and the number of clusters (marker 2). In the example of Figure 3, the clustering was performed on T cells, CD8+ T cells, B lineage, NK cells, monocytic lineage and endothelial cells, and 4 clusters have been required.
Two graphical outputs are provided when the “update” button is selected: a heatmap (marker 3) and boxplots for all selected populations (marker 4). Both can easily be downloaded via a right click in the image. The clusters information, detailing the cluster of belonging of each sample, can be downloaded by clicking on the appropriate link (marker 5).

Fig. 3. Step 2 of the analysis of the TCGA SARC example dataset. 1 marks the selection of the populations of interest. 2 indicates the selection of the number of clusters. 3 indicates the heatmap output. 4 marks the violinplots/boxplots output. 5 marks the download link for the clusters information file.

The heat map (Figure 3, marker 3) represents the cell population Z-scores for each samples. The clustering is computed using the Euclidian metric and Ward linkage criterion. Both samples and cell populations are clustered.

The violin plots (marker 4) output details the MCP-counter scores for all cell population of interest depending on the clusters presented in the heatmap. If one or more clusters has less than 3 samples, violin plots are replaced by boxplots. For each cell population, the relationship between the amount of this population and the clusters is estimated with a Kruskal-Wallis test. All pairwise comparisons are estimated by a post-hoc Dunn test and all p-values are indicated in the plots. P-values below the classical 0.05 threshold for significance are indicated in bold.

Comparison with other methods

Several other methods to deconvolute the immune and stromal composition from transcriptomic data of human tissues have been previously reported. In a benchmark study published last year comparing human-based methods, MCP-counter was found to be among the top-scoring methods for both accuracy (measured through correlation between the scores and randomized mixtures compositions) and specificity (measured through “spillover effect”). Murine counterparts for these methods are scarcer. In the article reporting the development of mMCP-counter, we demonstrated on an independent dataset that it outperformed pre-existing methods. Importantly, our set of methods is the only one to allow a similar analysis of human and murine samples, which can prove extremely valuable for comprehensive studies that include murine models and human samples. webMCP-counter includes methods that are fast-computing, as they rely on the computation of the mean expression of gene signatures and do not require more complex data processing.

Software implementation and availability

webMCP-counter was developed as a Shiny application in R. The heatmap generation is carried using the ComplexHeatmap package. Other used packages include dendextend, ggplot2 and dunn.test. All source code for webMCP-counter, placed under copyleft GNU license, is available from GitHub: https://github.com/FPetitprez/webMCP-counter. Users wishing to run webMCP-counter locally can download it from GitHub, install dependencies and run it as a regular shiny application.

Conclusion

In many areas of immunology, it is crucial to accurately estimate immune and stromal composition of heterogeneous samples, such as the tumor microenvironment in cancer immunology. To this end, we previously reported MCP-counter and mMCP-counter to quantify cell populations in human and murine samples, respectively, using bulk transcriptomic data. However, these tools were only accessible to R users. By developing webMCP-counter, we offer the possibility to all users to access our methods and apply them on their datasets.

Acknowledgements

We thank our colleagues who helped us test webMCP-counter.

Funding

This work was supported by the Programme Cartes d’Identité des Tumeurs of the Ligue Nationale contre le cancer, the Institut National de la Santé et de la Recherche Médicale, Université de Paris, Sorbonne Université and the Cancer Research for Personalized Medicine programme (CARPEM). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 754923. The material presented and views expressed here are the responsibility of the authors only. The EU Commission takes no responsibility for any use made of the information set out.

Conflict of Interest

The authors have no conflict of interest to declare.
References


Welcome to webMCP-counter

Web app for MCP-counter and mMCP-counter

To run MCP-counter or mMCP-counter using this web interface, you first need to prepare your data in a suitable format. 4 formats are accepted: Excel spreadsheet or text-based with tab, comma or semi-colon separator. Text format are preferred as they are more memory-efficient. In all cases, the samples must be put in columns, and genes in rows. The first column must imperatively be composed of gene symbols or gene ENSEMBL IDs. The first line must be composed of the corresponding sample IDs, and the cell above gene symbols must be filled. If you have further questions or require assistance, please open a new issue on the webMCP-counter Github page.

1. Upload gene expression profiles file
   Browse... No file selected

2. or, alternatively,
   (1) try with the TCGA SARC example data:

3. Select the file format
   xls

4. Select gene ID format
   - Gene symbol
   - ENSEMBL

5. Select organism
   - Human (Homo sapiens)
   - Mouse (Mus musculus)

6. Run (m)MCP-counter

<table>
<thead>
<tr>
<th></th>
<th>T cells</th>
<th>CD8+ T cells</th>
<th>Cytotoxic lymphocytes</th>
<th>NK cells</th>
<th>Monocytic lineage</th>
<th>Myeloid dendritic cells</th>
<th>Neutrophils</th>
<th>Endothelial cells</th>
<th>Fibroblasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA-3B-AshH</td>
<td>6.23</td>
<td>7.1</td>
<td>5.26</td>
<td>3.54</td>
<td>3</td>
<td>9.03</td>
<td>5.38</td>
<td>6.75</td>
<td>10.37</td>
</tr>
<tr>
<td>TCGA-3B-AshH1</td>
<td>8.95</td>
<td>8.21</td>
<td>5.64</td>
<td>11.69</td>
<td>3.23</td>
<td>9.79</td>
<td>5.59</td>
<td>5.26</td>
<td>8.68</td>
</tr>
<tr>
<td>TCGA-3B-AsH</td>
<td>3.9</td>
<td>3.19</td>
<td>3.4</td>
<td>2.02</td>
<td>3.13</td>
<td>8.65</td>
<td>3.1</td>
<td>5.8</td>
<td>9.33</td>
</tr>
<tr>
<td>TCGA-3B-AsH0</td>
<td>7.06</td>
<td>10.23</td>
<td>6.19</td>
<td>4.36</td>
<td>3.76</td>
<td>9.82</td>
<td>6.14</td>
<td>5.2</td>
<td>8.54</td>
</tr>
<tr>
<td>TCGA-3B-AsH0P</td>
<td>5.1</td>
<td>5.14</td>
<td>4.42</td>
<td>2.85</td>
<td>3.61</td>
<td>8.34</td>
<td>4.99</td>
<td>5.3</td>
<td>9.58</td>
</tr>
<tr>
<td>TCGA-3B-AsH0Q</td>
<td>6.53</td>
<td>6.5</td>
<td>3.97</td>
<td>4.94</td>
<td>2.82</td>
<td>8.83</td>
<td>7.32</td>
<td>4.82</td>
<td>7.59</td>
</tr>
<tr>
<td>TCGA-3B-AsH1P</td>
<td>2.33</td>
<td>1.66</td>
<td>0.97</td>
<td>1.62</td>
<td>1.49</td>
<td>5.33</td>
<td>2.11</td>
<td>4.37</td>
<td>8.03</td>
</tr>
<tr>
<td>TCGA-3B-AsH1Q</td>
<td>7.24</td>
<td>8.54</td>
<td>4.86</td>
<td>3.74</td>
<td>2.59</td>
<td>9.18</td>
<td>5.3</td>
<td>4.94</td>
<td>8.7</td>
</tr>
<tr>
<td>TCGA-3B-AsH2P</td>
<td>7.52</td>
<td>9.26</td>
<td>6.89</td>
<td>4.91</td>
<td>4.92</td>
<td>10.54</td>
<td>5.41</td>
<td>5.2</td>
<td>7.7</td>
</tr>
<tr>
<td>TCGA-3B-AsH2Q</td>
<td>7.42</td>
<td>8.47</td>
<td>5.87</td>
<td>4.75</td>
<td>3.31</td>
<td>9.04</td>
<td>6.53</td>
<td>5.31</td>
<td>7.04</td>
</tr>
</tbody>
</table>