

 Open access • Posted Content • DOI:10.1101/2020.12.03.400754

webMCP-counter: a web interface for transcriptomics-based quantification of immune and stromal cells in heterogeneous human or murine samples — [Source link](#)

Maxime Meylan, Etienne Becht, Catherine Sautès-Fridman, Aurélien de Reyniès ...+2 more authors

Institutions: University of Paris, Fred Hutchinson Cancer Research Center, University of Edinburgh

Published on: 03 Dec 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Related papers:

- [Cerebro: Interactive visualization of scRNA-seq data](#)
- [epiRomics: a multi-omics R package to identify and visualize enhancers](#)
- [The winning methods for predicting cellular position in the DREAM single cell transcriptomics challenge](#)
- [CelltrackR: an R package for fast and flexible analysis of immune cell migration data](#)
- [OEFinder: a user interface to identify and visualize ordering effects in single-cell RNA-seq data](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/webmcp-counter-a-web-interface-for-transcriptomics-based-f9mnhw322p>

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

Maxime Meylan^{1,2}, Etienne Becht^{1,3}, Catherine Sautès-Fridman², Aurélien de Reyniès¹, Wolf H. Fridman² and Florent Petitprez^{1,4,*}

¹ Programme Cartes d'Identité des Tumeurs, Ligue Nationale Contre le Cancer, Paris, France

² 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.

Availability and Implementation: Freely available from <http://134.157.229.105:3838/webMCP/>. Website developed with the R package shiny. Source code available from GitHub: <https://github.com/FPetitprez/webMCP-counter>.

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 (Euclidian 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.

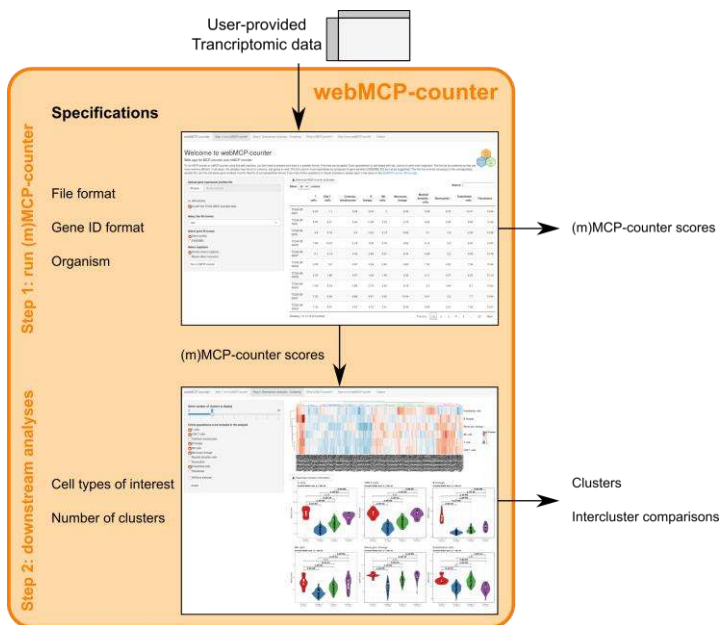


Fig. 1. Diagram explaining the main steps of a typical analysis using webMCP-counter. In the first step, the user uploads transcriptomic data and sets the parameters. (m)MCP-counter is then run on the user-provided data and can be accessed as output and passed on to the second step. Here, a clustering is performed based on the user-selected cell populations and number of classes.

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 cohort⁸. 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 step 1 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 also 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^{9–11}. This table can be downloaded as comma-separated values table (.csv) for use in other software, such as Microsoft Excel (Figure 2, marker 3).

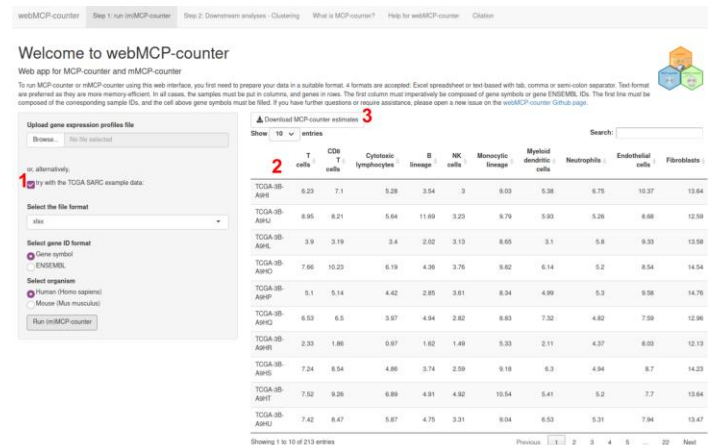


Fig. 2. Step 1 of the analysis of the TCGA SARC example dataset. Marker 1 indicates the box to tick to access the example dataset instead of a user-provided dataset. 2 indicates the MCP-counter scores table, accessible after MCP-counter has been run. Finally, 3 indicates the download link for the .csv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

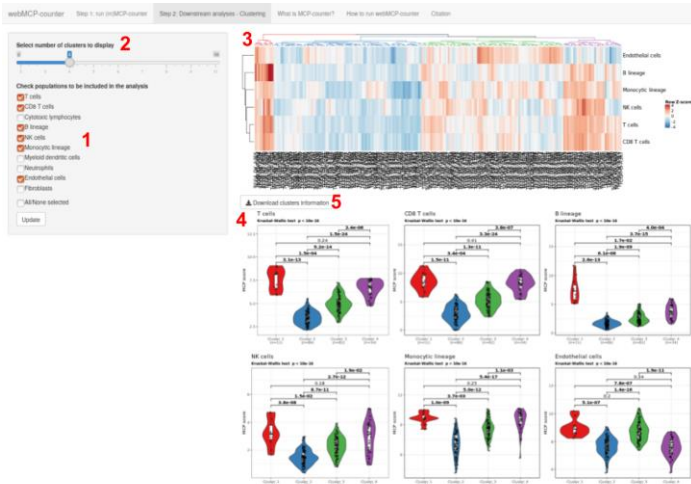


Fig. 3. Step 2 of the analysis of the TCGA SARC example dataset. 1 marks the selection of the populations of interest. 2 indicates the selection of the number of clusters. 3 indicates the heatmap output. 4 marks the violinplots/boxplots output. 5 marks the download link for the clusters information file.

The heatmap (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 methods¹², 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 methods⁷. 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 R¹³. The heatmap generation is carried using the ComplexHeatmap package¹⁴. Other used packages include dendextend¹⁵, ggplot2¹⁶ 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 Medecine 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

1. Fridman, W. H., Pagès, F., Sautès-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306 (2012).
2. Fridman, W. H., Zitvogel, L., Sautès-Fridman, C. & Kroemer, G. The immune contexture in cancer prognosis and treatment. *Nature Reviews Clinical Oncology* **14**, 717–734 (2017).
3. Petitprez, F., Meylan, M., de Reyniès, A., Sautès-Fridman, C. & Fridman, W. H. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. *Front. Immunol.* **11**, (2020).
4. Petitprez, F. *et al.* Quantitative Analyses of the Tumor Microenvironment Composition and Orientation in the Era of Precision Medicine. *Front Oncol* **8**, 390 (2018).
5. Finotello, F. & Trajanoski, Z. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol Immunother* **67**, 1031–1040 (2018).
6. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology* **17**, 218 (2016).
7. Petitprez, F. *et al.* The murine Microenvironment Cell Population counter method to estimate abundance of tissue-infiltrating immune and stromal cell populations in murine samples using gene expression. *Genome Medicine* **12**, 86 (2020).
8. The Cancer Genome Atlas Research Network *et al.* Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* **171**, 950–965.e28 (2017).
9. Becht, E. *et al.* Prognostic and theranostic impact of molecular subtypes and immune classifications in renal cell cancer (RCC) and colorectal cancer (CRC). *Oncoimmunology* **4**, e1049804 (2015).
10. Petitprez, F. *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, 556–560 (2020).
11. Helmink, B. A. *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577**, 549–555 (2020).
12. Sturm, G. *et al.* Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, i436–i445 (2019).
13. Chang, W., Cheng, J., Allaire, J. J., Xie, Y. & McPherson, J. *shiny: Web Application Framework for R.* (2020).
14. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
15. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
16. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* (Springer-Verlag New York, 2016).

User-provided
Transcriptomic data



webMCP-counter

Specifications

File format

Gene ID format

Organism

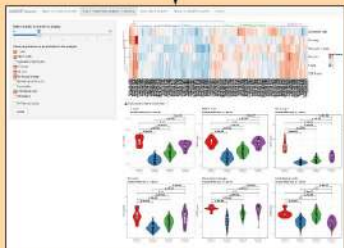


(m)MCP-counter scores

(m)MCP-counter scores

Cell types of interest

Number of clusters



Clusters

Intercluster comparisons

Step 1: run (m)MCP-counter

Step 2: downstream analyses



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](#).

Upload gene expression profiles file

Browse...

or, alternatively,

try with the TCGA SARC example data.

Select the file format

xls

Select gene ID format

Gene symbol

ENSEMBL

Select organism

Human (Homo sapiens)

Mouse (Mus musculus)

3

Show 10 entries

Search:

	2 T cells	CD8 T cells	Cytotoxic lymphocytes	B lineage	NK cells	Monocytic lineage	Myeloid dendritic cells	Neutrophils	Endothelial cells	Fibroblasts
TCGA-3B-ASHI	6.23	7.1	5.28	3.54	3	9.93	5.26	6.75	10.37	13.84
TCGA-3B-ASHJ	9.95	8.21	5.64	11.69	2.23	8.79	5.92	5.26	8.68	12.59
TCGA-3B-ASHL	3.3	3.19	3.4	2.02	3.13	8.65	3.1	5.8	8.33	13.58
TCGA-3B-ASHO	7.96	10.23	6.19	4.36	3.76	9.82	6.14	5.2	8.54	14.54
TCGA-3B-ASHP	5.1	5.14	4.42	2.85	3.91	6.34	4.99	5.3	9.56	14.76
TCGA-3B-ASHQ	6.53	6.5	3.97	4.94	2.82	6.83	7.52	4.82	7.59	12.96
TCGA-3B-ASHR	2.31	1.96	0.87	1.62	1.49	5.32	3.11	4.37	8.03	12.13
TCGA-3B-ASHS	7.24	8.54	4.85	3.74	2.59	9.18	6.3	4.94	8.7	14.23
TCGA-3B-ASHT	7.52	8.28	6.89	4.91	4.92	10.54	8.41	5.2	7.7	13.84
TCGA-3B-ASHI	7.42	6.47	5.87	4.75	3.31	9.94	6.53	6.31	7.94	13.47

Showing 1 to 10 of 210 entries

Previous ... Next

