

SOFTWARE

Open Access



# oposSOM-Browser: an interactive tool to explore omics data landscapes in health science

Henry Loeffler-Wirth<sup>1\*</sup> , Jasmin Reikowski<sup>1</sup>, Siras Hakobyan<sup>2</sup>, Jonas Wagner<sup>3</sup> and Hans Binder<sup>1</sup>

\*Correspondence:

wirth@izbi.uni-leipzig.de

<sup>1</sup> Interdisciplinary Centre for Bioinformatics, Leipzig University, Härtelstraße 16-18, 04107 Leipzig, Germany  
Full list of author information is available at the end of the article

## Abstract

**Background:** oposSOM is a comprehensive, machine learning based open-source data analysis software combining functionalities such as diversity analyses, biomarker selection, function mining, and visualization.

**Results:** These functionalities are now available as interactive web-browser application for a broader user audience interested in extracting detailed information from high-throughput omics data sets pre-processed by oposSOM. It enables interactive browsing of single-gene and gene set profiles, of molecular ‘portrait landscapes’, of associated phenotype diversity, and signalling pathway activation patterns.

**Conclusion:** The oposSOM-Browser makes available interactive data browsing for five transcriptome data sets of cancer (melanomas, B-cell lymphomas, gliomas) and of peripheral blood (sepsis and healthy individuals) at [www.izbi.uni-leipzig.de/oposom-browser](http://www.izbi.uni-leipzig.de/oposom-browser).

**Keywords:** Interactive data analysis, Transcriptomics, Results browser

## Background

Many bioinformatics tools are currently in transition from software libraries to interactive solutions designed for a broader user community including data scientists, output-oriented medical researchers and experimenters with needs in intuitive visualization and exploration for complex, multidimensional data. We here present an interactive web-tool which extends functionalities of our Bioconductor R-package ‘oposSOM’ designed to analyse transcriptome data in cancer and health research [1]. The method is based on self-organizing map (SOM) machine learning for dimension reduction, visualization, and comprehensive downstream analysis. This so-called ‘high-dimensional data portraying’ visualizes individual data landscapes, and performs function mining, modular feature selection, sample stratification, diversity analysis, and phenotype mapping [2]. It was applied to a series of data types (transcriptome, methylome, proteome, genome), diseases (cancers such as melanoma, lymphoma; autoimmune diseases) on the level of patient-cohort and cell system specimen (see, e.g., [3–5]). The method was so far used



© The Author(s) 2020. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

in more than 60 publications in a large variety of studies related to cellular development, toxicology, health studies, cell, and molecular biology. A full list of references can be found in Additional file 1: Appendix.

We further developed the analytic options of this method and here present an interactive browsing tool, which provides intuitive access to all the information generated by means of the data portraying method. The oposSOM-Browser complements and extends the functionalities of the oposSOM software package by interactive functionalities in the context of gene-expression and gene-function profiling, associations with phenotypes, and pathway activities in selected transcriptome data sets on different cancer entities and blood transcriptomes. The oposSOM-browser is hosted by the ‘Leipzig Health Atlas’, a sharing platform for publications, biomedical data, models, and software tools from the field of health research.

## Implementation

### Implementation and availability

The oposSOM-Browser is implemented using R-Shiny [6]. The Shiny web server is permanently running as Docker container hosted by the Leipzig Health Atlas app infrastructure. It can be accessed via any standard web browser under [www.izbi.uni-leipzig.de/opossom-browser](http://www.izbi.uni-leipzig.de/opossom-browser).

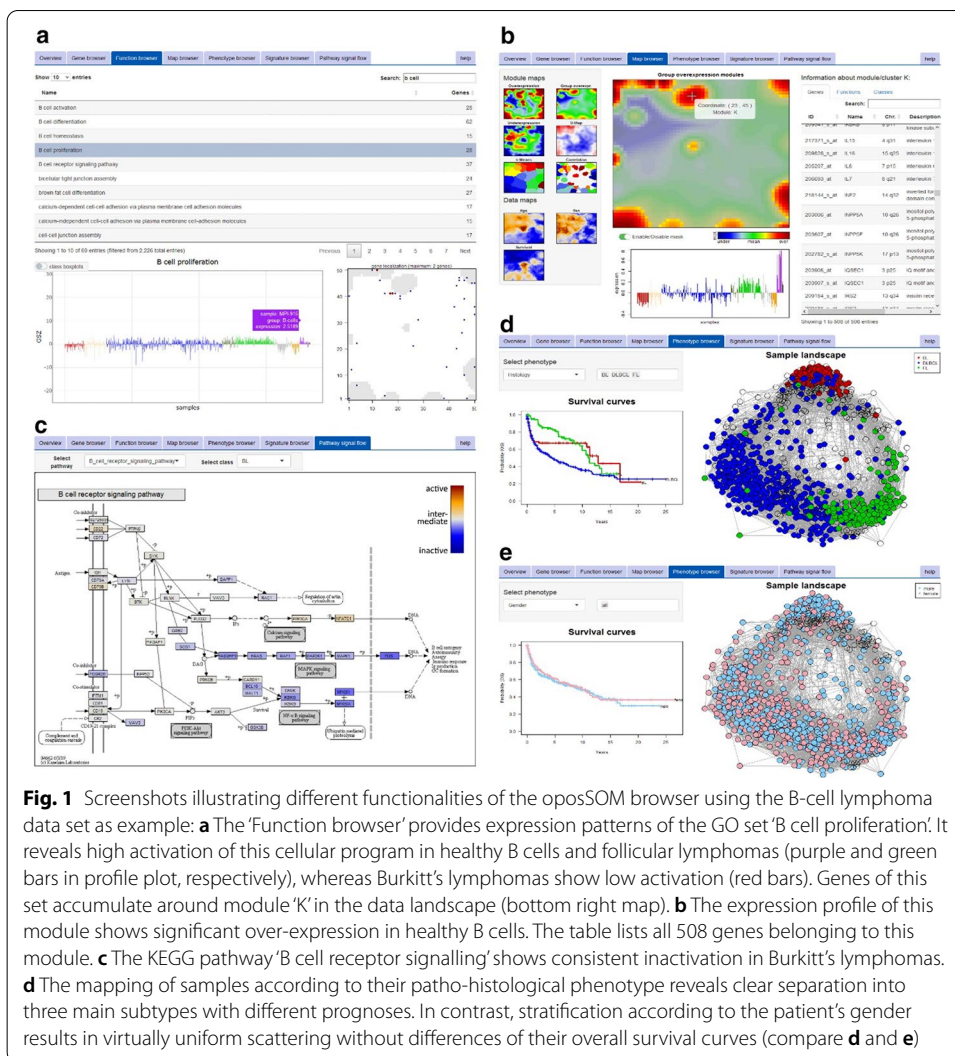
### Datasets

Five datasets are currently available in the oposSOM-Browser: (1) 917 specimen of germinal B cell lymphomas and selected healthy control samples [3], see below; (2) 80 melanoma and nevi samples [5]; (3) 137 low-grade gliomas [4]; (4) 180 blood samples of community acquired pneumonia patients [7]; and (5) 3388 blood samples collected from healthy participants of a population-based health study [8]. All data sets were pre-processed using oposSOM. Further data sets are presently in preparation for release in the browser. Interested users are invited to provide their analyses to the browser via request to the corresponding author.

### Modules

Functionalities are arranged as browser-tabs as shown in Fig. 1. Detailed descriptions are provided in Additional file 1: Appendix and in the ‘Guided tour’ tab in the Browser:

- The ‘overview’ tab provides a description of the data set selected, a link to the corresponding publication, and additional information such as the dimensionality and version of oposSOM package used for data processing.
- The ‘gene browser’ and ‘function browser’ tabs provide tables of all up to 55,000 genes in the dataset and of all up to 10,000 functional gene sets considered. It enables the visualization of feature-profiles and mapping of the selected genes into the SOM data landscape.
- The ‘map browser’ tab provides an overview about patterns of the expression landscape: Lists of co-expressed genes are given together with enriched functional gene sets (for details see [2]). Accompanying data maps are shown for age, gender and



**Fig. 1** Screenshots illustrating different functionalities of the oposSOM browser using the B-cell lymphoma data set as example: **a** The ‘Function browser’ provides expression patterns of the GO set ‘B cell proliferation’. It reveals high activation of this cellular program in healthy B cells and follicular lymphomas (purple and green bars in profile plot, respectively), whereas Burkitt’s lymphomas show low activation (red bars). Genes of this set accumulate around module ‘K’ in the data landscape (bottom right map). **b** The expression profile of this module shows significant over-expression in healthy B cells. The table lists all 508 genes belonging to this module. **c** The KEGG pathway ‘B cell receptor signalling’ shows consistent inactivation in Burkitt’s lymphomas. **d** The mapping of samples according to their patho-histological phenotype reveals clear separation into three main subtypes with different prognoses. In contrast, stratification according to the patient’s gender results in virtually uniform scattering without differences of their overall survival curves (compare **d** and **e**)

prognosis (in terms of overall survival curves) of the individuals in the respective cohort.

- The ‘phenotype’ tab provides the correlation network of sample similarities. The network can be stratified considering up to 25 different phenotypes related to patient information, clinical or molecular characteristics, along with the corresponding survival curves.
- The ‘signature’ tab enables the user to upload lists of signature genes (Ensemble-IDs or gene names). The browser delivers their mean expression profile across all samples and shows their location in the SOM data landscape. Further, the provided signature is benchmarked (ROC and AUC) with regard to the phenotype classes selected.
- The ‘pathway signal flow’ tab shows KEGG signalling pathways with genes colour-coded according to their activity level [9].

### Results: use case lymphoma browser

Our use case presents a transcriptome dataset of 917 B cell lymphoma specimen and healthy control samples [3]. The oposSOM-browser provides a holistic view on the expression landscape, the heterogeneity of activated gene-regulatory programs and their association with different lymphoma subtypes and clinical phenotypes (see Fig. 1). A first step is to examine the expression landscape in a particular data set using the map browser (Fig. 1b), assigning the series of lymphoma subtypes and healthy cell references to the corresponding expression modules together with their functional interpretation. Then, the gene browser enables investigating genes of interest, e.g. by selecting frequently mutated genes or genes previously reported as expression markers of different lymphoma subtypes, by mapping them into the expression landscape and/or by exploring their expression profile across subtypes and samples. Patterns of cellular activity can be explored using the function and PSF browser tabs (Fig. 1a, c), in order to identify subtype-specific or ubiquitous processes and signaling cascades. Finally, mapping of clinical, genetic and phenotypic subtyping schemes enables the mutual comparison of different lymphoma strata in terms of cluster structure and of survival hazard ratios (see Fig. 1d, e for stratification based on patho-histological diagnosis and by gender, respectively).

### Conclusion

oposSOM-Browser is a novel tool for the interactive exploration of high-dimensional omics data and associated phenotypes, allowing interested researchers to browse through the data by addressing specific issues and their own questions in order to generate or to validate hypotheses not or incompletely considered before.

Further extension of available data sets will build a library of annotated omics landscapes for health science.

### Availability and requirements

Project name: oposSOM-Browser

Project home page: [www.izbi.uni-leipzig.de/oposom-browser](http://www.izbi.uni-leipzig.de/oposom-browser)

Operating systems: Platform independent

Programming language: R using Shiny package

Other requirements: Internet browser

License: oposSOM-Browser is licensed under the terms of use defined for the [Leipzig Health Atlas](#)

Restrictions to use by non-academics: license needed

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12859-020-03806-w>.

**Additional file 1.** Appendix with reference list and guided tour.

### Abbreviations

AUC: Area under the ROC curve; KEGG: Kyoto Encyclopedia of genes and genomes; PSF: Pathway signal flow; ROC: Receiver operating characteristic; SOM: Self-organizing map.

### Acknowledgements

Not applicable.

### Authors' contributions

Conceptualization, supervision and draft-writing: HLW and HB. Implementation: HLW and JR. Pathway resources: HLW and SH. Server infrastructure and administration: JW. All authors have read and approved the final manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL. HLW and JW were funded by the BMBF i:DSem (collaborative) project Leipzig Health Atlas ([www.health-atlas.de](http://www.health-atlas.de)).

### Availability of data and materials

oposSOM-Browser is available under [www.izbi.uni-leipzig.de/oposom-browser](http://www.izbi.uni-leipzig.de/oposom-browser).

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The Authors declare that no competing interests exist.

### Author details

<sup>1</sup> Interdisciplinary Centre for Bioinformatics, Leipzig University, Härtelstraße 16-18, 04107 Leipzig, Germany. <sup>2</sup> Group of Bioinformatics, Institute of Molecular Biology, National Academy of Sciences, 7 Hasratyan str, 0014 Yerevan, Armenia. <sup>3</sup> LIFE, Leipzig Research Center for Civilization Diseases, Leipzig University, Philipp-Rosenthal-Straße 27, 04103 Leipzig, Germany.

Received: 16 July 2020 Accepted: 13 October 2020

Published online: 19 October 2020

### References

1. Löffler-Wirth H, Kalcher M, Binder H. oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on Bioconductor. *Bioinformatics*. 2015;31:3225–7.
2. Wirth H, Löffler M, von Bergen M, Binder H. Expression cartography of human tissues using self organizing maps. *BMC Bioinform*. 2011;12(1):306–52.
3. Loeffler-Wirth H, Kreuz M, Hopp L, Arakelyan A, Haake A, Cogliatti SB, et al. A modular transcriptome map of mature B cell lymphomas. *Genome Med*. 2019;11(1):27. <https://doi.org/10.1186/s13073-019-0637-7>.
4. Binder H, Willscher E, Loeffler-Wirth H, Hopp L, Jones DTW, Pfister SM, et al. DNA methylation, transcriptome and genetic copy number signatures of diffuse cerebral WHO grade II/III gliomas resolve cancer heterogeneity and development. *Acta Neuropathol Commun*. 2019;7(1):59. <https://doi.org/10.1186/s40478-019-0704-8>.
5. Kunz M, Löffler-Wirth H, Dannemann M, Willscher E, Doose G, Kelso J, et al. RNA-seq analysis identifies different transcriptomic types and developmental trajectories of primary melanomas. *Oncogene*. 2018;37:6136–51.
6. Chang W, Cheng J, Allaire J, Xie Y. shiny: web application framework for R. 2020.
7. Hopp L, Loeffler-Wirth H, Nersisyan L, Arakelyan A, Binder H. Footprints of sepsis framed within community acquired pneumonia in the blood transcriptome. *Front Immunol*. 2018;9:1620. <https://doi.org/10.3389/fimmu.2018.01620/full>.
8. Schmidt M, Binder H, Binder H, Hopp L, Arakelyan A, Kirsten H, et al. Portrayal of the human blood transcriptome in a population cohort and its relation to ageing and health. *Front Big Data*. 2020. <https://doi.org/10.3389/fdata.2020.548873/abstract>.
9. Nersisyan L, Löffler-Wirth H, Arakelyan A, Binder H. Gene set-and pathway-centered knowledge discovery assigns transcriptional activation patterns in brain, blood, and colon cancer: a bioinformatics perspective. *Int J Knowl Discov Bioinform*. 2016;4(2):46–69.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.