

SCENERY: a Web-Based Application for Network Reconstruction and Visualization of Cytometry Data

Giorgos Athineou¹, Giorgos Papoutsoglou¹, Sofia Triantafillou¹, Ioannis Basdekis², Vincenzo Lagani¹, and Ioannis Tsamardinos¹

¹ University of Crete - Department of Computer Science
Voutes Campus, 700 13 Heraklion, Greece

² Foundation for Research and Technology, Hellas - Institute of Computer Science
Vassilika Vouton, 700 13 Heraklion, Greece
`tsamard@csd.uoc.gr`

Abstract. Cytometry techniques allow to quantify morphological characteristics and protein abundances at a single-cell level. Data collected with these techniques can be used for addressing the fascinating, yet challenging problem of reconstructing the network of protein interactions forming signaling pathways and governing cell biological mechanisms. Network reconstruction is an established and well studied problem in the machine learning and data mining fields, with several algorithms already available. In this paper, we present the first web-oriented application, SCENERY, that allows scientists to rapidly apply state-of-the-art network-reconstruction methods on cytometry data. SCENERY comes with an easy-to-use user interface, a modular architecture, and advanced visualization functions. The functionalities of the application are illustrated on data from a publicly available immunology experiment.

Keywords: Cytometry, CyTOF, Network Reconstruction, Web Application, Signaling Pathway

1 Introduction

Signaling networks are well-organized chains of complex molecular events [1]. Molecular stimuli trigger these events by orderly changing the state of specific proteins ultimately perturbing the cell's metabolism, shape, gene expression, or ability to divide. Flow cytometry is a robust and broadly accessible method nowadays, able to provide quantitative measurements on such sensitive macromolecular interactions [2]. Still, reconstruction of signaling networks from flow cytometry measurements has not become popular, primarily due to the limited molecular quantities the method can measure. Recently, a novel technique called Mass Cytometry was introduced revolutionizing the state of the art [2]. Its inherent ability to investigate more than 30 quantities simultaneously, offers, now, the opportunity to delve deeper into cell signaling networks.

Inducing signaling networks from data can be thought as a Network Reconstruction (NR) problem. NR methods have become increasingly popular in biology, especially for inferring gene-gene interaction networks, with numerous scientific works currently published on this subject [3]. The first successful case of signaling NR in the cytometry field was achieved by Sachs and co-authors [4], followed by several applications [5], [6]. However, NR methods are not yet routinely used on single-cell cytometry data. Arguably, this is mainly due to the intrinsic complexity of the task. Attempting to reconstruct signaling pathways requires knowing in detail the semantics of the data, the peculiarities of the cytometry technology, and all available information on the specific pathway and its components. On the other hand, successfully applying NR algorithms requires mastering all the technicalities of these methods, since inaccuracies in the analysis pipeline are potentially able to invalidate all results [7].

In addition to those intrinsic problems, a horizontal factor that limits broadcasting, sharing and reusing scientific results in this domain is the reluctance towards social media [8] and the limited use of online services that support sharing of credible and accurate results. These tools promote openness, hiding at the same time sensible, core aspects of an experiment of this kind (in order to prevent copyright infringement).

In this work, we present SCENERY (Single CELL Network Reconstruction sYstem), a web-based application specifically devised to allow researchers to apply NR methods on single-cell cytometry data, even with limited knowledge of the technical details of these algorithms. SCENERY interface guides the user through a set of easy steps; from data loading and study design specification, to the set-up of the analysis and results visualization and sharing. Its core is built on R and its modularity grants to easily add extensions, particularly additional NR methods. To the best of our knowledge, SCENERY is the first available software of its kind. Several other applications exist for cytometry data analysis, both as stand-alone software (FlowJo, www.flowjo.com), web-service (CytoBank, [9]) and libraries [10]; however none of these tools provide the user with NR functionalities.

The rest of the paper is structured as follows. First, we provide an overview of SCENERY functionalities and internal architecture. A use case on real, publicly available data is then presented for better illustrating SCENERY capabilities. Future extensions of the application are then discussed in the conclusions.

2 The SCENERY Application

2.1 Software Functionality

SCENERY's users are guided by a wizard through a specific sequence of analysis steps [11], as shown in Fig. 1. The user supplies the data (step 1), defines the computational experiment (step 2) and then sets the execution parameters (step 3). The user may run the analysis, and redefine the execution parameters until the desired outcome is achieved. This sequence of steps also serves as

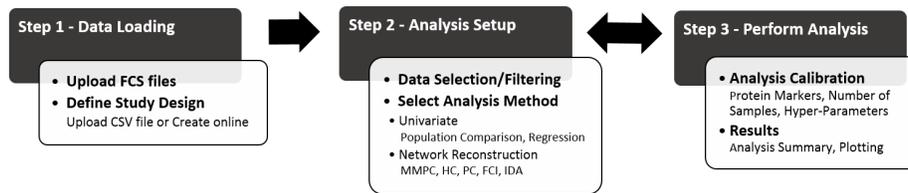


Fig. 1. Flowchart of typical user-application interaction. In step 1 users upload data and define the study design. In step 2 they setup a computational experiment by selecting datasets and the analysis method. In step 3, users calibrate the input parameters and execute the analysis. The analysis can be reconfigured and repeated multiple times.

an educational path for less experienced users who are interested in exploring any aspect of the available analysis methods. Analysis output can be exported in various ways, mainly publication-quality figures and standard formats for graph-representation (i.e., Graph Exchange XML Format, GEXF). In future versions of SCENERY the user will be able to share the output of the analysis privately (via email or a repository) or publicly (via social media or blogs accounts) to a group of colleagues for further analysis and discussion. Recent social-media citation practices indicate that scientific content is becoming more and more part of every day conversation, thus increasing chances of citation [12].

Data Loading. As first step, the user uploads one or more data files in Flow Cytometry Standard (FCS) format. This format is adopted universally by cytometry analysis software, allowing SCENERY to import and analyze datasets pre-processed by other applications. FCS files may correspond to different samples (e.g. patients, cell types) or conditions (e.g., stimuli, inhibitor dosages). SCENERY goes beyond traditional study design declaration, and it allows users to assert any type of metadata knowledge concerning variables, quantities, attributes, or characteristics of the samples (e.g. gender, age, etc.). Hence, any type and number of factors can be defined in a custom study design, both qualitative (e.g. cell type) and quantitative (e.g., drug dosage). *This flexibility permits to accommodate virtually all possible study designs, both current and future ones.*

Analysis Setup. In the second step, the user sets up the desired *computational experiment*. This essentially involves the selection of (a subset of) the uploaded data files on the basis of factors of the study design, and the application of a single data analysis method. Analysis methods included in the current version of SCENERY are subdivided into standard statistical methods and NR algorithms. Statistical analysis methods include the t-test and the analysis of variance for comparisons across different levels of the same factor. Univariate linear and logistic regression methods are also available for modeling the relationship between markers and study design’s factors.

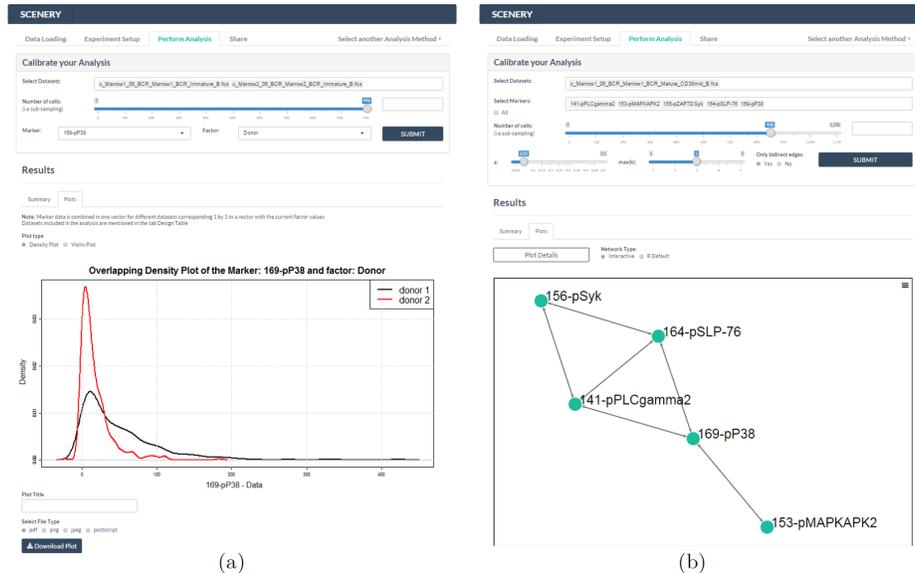


Fig. 2. Visualizing results in SCENERY. (a) Overlapping density plots for the marker p38 on 2 donors. (b) The retrieved reconstructed network after applying MMPC on selected mass cytometry data (see text for details). In both screen-shots the analysis calibration panel is also displayed on top of the graphs, as indicative of the UI.

SCENERY is the first free software to support a number of NR algorithms for single-cell data. All NR methods represent statistical relationships in the data as networks composed by nodes and edges. Nodes always stand for measurements (e.g., protein abundances). Edges, on the contrary, have different semantics, depending on the type of network the method outputs. Association Networks (AN) connect two nodes with an undirected edge if the corresponding measurements are found statistically associated. Conditional Association Network (CAN) are similar to AN, but associations between nodes are computed conditioning on all (or part) of the remaining measurements. Bayes Networks (BNs) use Directed Acyclic Graphs (DAGs) for representing the multivariate distribution of the data. A common misconception is interpreting a directed edge in a BN as an indication of causal interaction. This is possible only under the standard causal discovery assumptions (Causal Markov Condition, Faithfulness). Even then, not all causal relationships are identifiable by data alone. Therefore, some algorithms output Partial DAGs (PDAGs), that use directed edges for representing causal relations, and undirected edges to represent edges whose causal direction is unclear. If hidden common confounders are also a possibility, Maximal Ancestral Graphs (MAGs) are typically used instead of BNs. MAGs use directed edges to represent causal relationships, and bi-directed edges to represent confounded relationships. Again, since some causal directions are not identifiable, the algo-

rithms usually output Partial Ancestral Graphs (PAGs) that use circle endpoints to indicate ambiguous orientations.

In this version we included five state-of-the-art NR algorithms. We use the MMPC algorithm [13] for deriving CANs and the HC algorithm [14] for reconstructing BNs. Algorithms [15] PC and FCI are used for reconstructing PAG and MAG models, respectively. Finally, the IDA algorithm [15] allows the estimation of a lower bound for the effect size of the causal relation between two variables.

Perform Analysis. Once an analysis method is selected the wizard redirects to the third step where the user is provided with analysis calibration options. Common options for all methods is deciding which markers and the number of cells to employ for the analysis. Next, the user defines the method-specific hyper-parameters and submits the analysis to the system. The analysis output is presented in a separate results panel. This panel consists of two sections/tabs; namely, Summary and Plots. The Summary tab recapitulates the performed analysis reporting metadata information and a textual overview of the results. In the Plots tab and depending on the selected analysis method a separate graphical (downloadable) representation of the results is included. Regarding the population comparison methods, results are graphically displayed using overlapping density plots (Fig. 2-a) and violin plots while, scatter-plots with fitted regression lines are used for regression analyses. For the NR analysis two network visualizations are available (Fig. 2-b), a static and an interactive one. Interactive visualization includes features such as zooming, node re-positioning and several others rendering a user-centric layout.

Finally, data visualization functionalities are available for exploring the data by histograms on stand-alone markers or by scatter-plots for multiple markers.

2.2 Application Architecture

SCENERY is a platform-independent web application of Client-Server architecture; built on R and PHP running on an Apache web server (www.apache.org/). The interface on the Client side is implemented using HTML5, CSS3 for structuring and presenting the content and JavaScript for light-weight tasks such as validation of forms, effects on moving elements, asynchronous communication and more. In order to alleviate the overhead associated with common tasks in web development, the Bootstrap web framework (getbootstrap.com/) is used, while the R Shiny web framework is used (shiny.rstudio.com/) to allow R functions communicate among Client and Server. All analysis methods are implemented in R and run on the Server. Additionally, PHP was used as an application skeleton/controller, for managing most of the Client-Server interaction and database operations. A MySQL database is used for storing user information and history (www.mysql.com/).

While the current architecture is a work in-progress, future version(s) will be extended with operational information built upon Design Strategy for Device Independence [16] enabling the web application to be utilized in various screen dimensions and environments of use.

Modularity. One of the main features of SCENERY is its modularity. Each analysis method is provided by a single R function with a standardized signature (dataset, method’s options) and results’ type (summary, visualization). This ensures that further analysis methods can be easily integrated within the step-wise SCENERY structure. We are planning to allow users submitting their own NR methods as R code in future versions of SCENERY.

3 Application on Immunology

3.1 Definition of the Problem

To a computer scientist, signaling networks are informal causal models for which proteins are key members. Their role is to relay the signal by switching between active and inactive states, thereby altering their function. Information is passed on sequentially from one protein to the other until the response is produced. We demonstrate an application of SCENERY by trying to reconstruct a part of a signaling pathway, using public mass cytometry data published in [17]. In the original study, the authors use mass-cytometry to measure 31 proteins related to the human hematopoietic system in two healthy bone marrow donors. Cells were stimulated with several activators to uncover distinct signaling mechanisms. Here, we use data from B-cell populations. Particularly, cells treated with stimulus of the B cell antigen-receptor (BCR). BCR signaling is known to trigger several signaling cascades simultaneously permitting many distinct outcomes [18]. Hence, this dataset provides an excellent showcase for the features and applicability of SCENERY in the research of signaling pathway networks.

3.2 Analyzing Cytometry Data with SCENERY

In the following examples we employ a subset of proteins, known to be involved in BCR signaling, namely SYK, BLNK, PLC γ 2, p38 and MAPKAPK2. Fig. 2-a illustrates how SCENERY would visualize a population comparison result (Univariate Analysis). For this graph we employed data from 2 donors for the protein marker p38. At the upper half, the screen-shot shows the configuration of the calibration options. At the bottom half, the two overlapping density plots for this specific analysis are shown.

In the same spirit, Fig. 2-b displays the NR results as they were retrieved by running the MMPC algorithm in SCENERY. Bi-directed edges denote correlation between the respective protein markers in a sense that both bi-connected nodes have been selected in the Parent-Children set of each other. Starting from the top left corner, then, the reconstructed network indicates that SYK, BLNK and PLC γ 2 are inter-correlated. This is true biologically because the stimulated BCR attracts and activates SYK which, in turn, attracts, interacts and phosphorylates both BLNK and PLC γ 2 [18],[19]. This process is part of a complex stimulation process that ultimately activates several proteins. One of them is p38 which interacts with both PLC γ 2 and BLNK in order to be activated [20].

This process is captured in SCENERY's output and is shown in the reconstructed network by the respective correlation edges. After p38 and further downstream, the reconstructed network extends to MAPKAPK2. This edge is also consistent with the literature, where MAPKAPK2 is found to be directly phosphorylated by p38 [21].

4 Discussion and Conclusions

In this work we introduced SCENERY, the first freely available web-based application for single-cell NR analysis. SCENERY packages advanced machine-learning methods in a user-friendly environment: a wizard guides users through all phases of the complex NR effort delivering a simple-to-use interface. This allows biology researchers unfamiliar with the technical details to exploit NR methods in discovering novel signaling pathways. It also allows SCENERY to serve as an educational tool for exploring the features of NR methods.

We showcase some of SCENERY's features using published mass cytometry B-cell data. We illustrate in Fig. 2-a the simplicity with which the software can represent standard statistical analysis results. In Fig. 2-b we show how users can easily assess results from a NR analysis.

More features will be implemented in feature releases. These include: implementation of an online archive for the users sessions; establish connection with other online services for directly loading public data; standard cytometry pre-processing functionalities (e.g., gating, data compensation); sharing analysis results via email or social media; and enriching SCENERY with even more analysis methods. Particularly, we envision that in the future users will be able to submit their own NR analysis methods as R source code for being added to the functionalities of SCENERY. These custom methods may be re-utilized by colleagues of members of a group specified by the end-user.

Our efforts towards this open-source approach hold the promise to transform SCENERY in an essential tool for the cytometry community for understanding the organization of complex cellular processes such as signaling networks.

Acknowledgments. This work was funded by European Research Council (ERC) and is part of the CAUSALPATH - Next Generation Causal Analysis project, No 617393. We sincerely thank Karen Sachs, David Gomez-Cabrero, Angelika Schmidt and Jesper Tegner for their invaluable comments, suggestions and encouragement on the start of this project.

Availability. Instructions on how to access and use SCENERY are available at <http://mensxmachina.org/en/software/>.

References

1. Ullrich, A., Schlessinger, J.: [Signal transduction by receptors with tyrosine kinase activity. Cell, 61\(2\):203-212 \(1990\)](#)

2. Bendall, S., Nolan, G., Roederer, M., Chattopadhyay, P.: A Deep Profilers Guide to Cytometry. *Trends Immunol.*, 33(7):323-332 (2012)
3. Marbach, D., Costello, J.C., Kffner, R., et. al.: Wisdom of crowds for robust gene network inference. *Nat. Methods.* 2012 Jul 15; 9(8): 796804.
4. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, DA., Nolan, GP.: [Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data.](#) *Science*, 308: 523-529 (2005)
5. Itani, S., Ohannessian, M., Sachs, K., et. al.: Structure learning in causal cyclic networks. *JMLR Workshop and Conference Proceedings* (2010)
6. Qiu, P., Simonds, E.F., Bendall, S.C., et. al.: Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotech.* 29, 886891 (2011)
7. Lagani, V., Triantafyllou, S., Ball, B., Tegner, J., and Tsamardinos, I.: Probabilistic Computational Causal Discovery for Systems Biology. Book chapter in *Uncertainty in Biology, A Computational Modeling Approach.* Springer, (2015)
8. Bik H.M., Goldstein M.C.: [An Introduction to Social Media for Scientists.](#) *PLoS Biol* 11(4): e1001535. doi:10.1371/journal.pbio.1001535 (2013)
9. Kotecha, N., Krutzik, PO., Irish, JM.: [Web-based Analysis and Publication of Flow Cytometry Experiments.](#) *Curr. Prot. Cyt.*, Chapter 10, Unit10.17. (2010)
10. Levine, JH., Simonds, EF., Bendall, SC., et. al.: Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* (2015)
11. Kelkar, S. A.: *Usability and Human-Computer Interaction: A Concise Study*, pp 306, (2009).
12. Priem, J. and Costello, K.L.: [How and why scholars cite on Twitter.](#) *Proceedings of ASIST*, 47(1), 104 (2010)
13. Lagani, V., Athineou, G., Borboudakis, G. and Tsamardinos, I.: MXM: Discovering Multiple, Statistically-Equivalent Signatures. R package version 0.4.3. <http://CRAN.R-project.org/package=MXM> (2015).
14. Scutari, M.: [Learning Bayesian Networks with the bnlearn R Package.](#) *JSS*, 35(3), 1-22. <http://www.jstatsoft.org/v35/i03/> (2010)
15. Kalisch, M., Maechler, M., Colombo, D., Maathuis, MH. and Buehlmann, P.: [Causal Inference Using Graphical Models with the R Package pcalg.](#) *JSS*, 47(11), 1-26. <http://www.jstatsoft.org/v47/i11/> (2012)
16. Karampelas, P., Basdekis, I. and Stephanidis, C.: [Web user interface design strategy: Designing for device independence.](#) In: C. Stephanidis (ed.). *Proceedings of the 13th International Conference on Human-Computer Interaction HCI International 2009 July 19-24, 2009, San Diego, CA, USA.* pp. 515-524. (2009)
17. [Bendall, S.C., Simonds, E.F., et. al.: Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum.](#) *Science*. 332 (6030): 687696. (2011)
18. Dal Porto, J., Gauld, S., Merrell, et.al.: B cell antigen receptor signaling 101. *Mol. Imm.* 41:599613 (2004)
19. Ishiai, M., Kurosaki, M., Pappu, R., et.al.: BLNK Required for Coupling Syk to PLC γ 2 and Rac1-JNK in B Cells. *Imm.* 10: 117125 (1999)
20. Guinamard, R., Signoret, N., Ishiai, et.al.: B cell antigen receptor engagement inhibits stromal cell-derived factor (SDF)-1 α chemotaxis and promotes protein kinase C (PKC)-induced internalization of CXCR4. *J.Exp.Med.* 189(9):1461-6 (1999)
21. [Cargnello, M., Roux, P.: Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases.](#) *Microbiol. Mol. Biol. Rev.* 75(1):5083 (2011)