
Systems Biology

Mining single-cell time-series datasets with Time Course Inspector

Maciej Dobrzyński^{1,*}, Marc-Antoine Jacques¹ and Olivier Pertz¹

¹Institute of Cell Biology, University of Bern, Bern, 3012, Switzerland

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Thanks to recent advances in live cell imaging of biosensors, microscopy experiments can generate thousands of single-cell time-series. To identify sub-populations with distinct temporal behaviours that correspond to different cell fates, we developed Time Course Inspector (TCI) – a unique tool written in R/Shiny to combine time-series analysis with clustering. With TCI it is convenient to inspect time-series, plot different data views and remove outliers. TCI facilitates interactive exploration of various hierarchical clustering and cluster validation methods. We showcase TCI by analysing a single-cell signalling time-series dataset acquired using a fluorescent biosensor.

Availability and Implementation: <https://github.com/pertzlal/shiny-timecourse-inspector>

Contact: maciej.dobrzynski@izb.unibe.ch

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Modern time-lapse microscopy can yield thousands of time-resolved measurements of protein activities in individual cells. A novel emerging concept is that single-cell measurements are heterogeneous (Fig. S1A) and thus more complex than population averages obtained with classic biochemical methods (Lee and Covert, 2010; Lahav and Purvis, 2013). For example, growth factor stimulation of an isogenic population of PC-12 cells, a paradigm for the study of neuronal differentiation, can lead to transient or sustained MAPK ERK activity profiles across cell population (Ryu *et al.*, 2015). These different dynamics relate to distinct cell fates (Fig. S1BC). Thus, understanding complexity of signalling in a cell population requires extraction of dynamic patterns from the multitude of time-series via clustering. The workflow typically involves scripting in Python or R, and subsequent iterations to obtain best partitioning. To ease this process we created an interactive R/Shiny application for hierarchical clustering, cluster validation, and easy plotting of various data views.

2 Features and implementation

TCI accepts CSV files with time-series in long or wide formats. Data views include population averages, individual time-series, or mean power

spectral density. The area under the curve or measurements at chosen time points can be displayed as box-, violin-, and/or dot-plots. Scatter plots visualise correlations between two time points. Other modules offer hierarchical clustering with various distance measures, e.g. Euclidean, dynamic time warping, and sparse hierarchical clustering (Witten and Tibshirani, 2010). Validation module assesses clustering tendency and quality of clustering. At various stages of the analysis plots can be viewed in the interactive mode and saved in PDF/PNG formats or as an R object for further editing. The results, e.g. assignments of time-series to clusters, can be exported as CSV files. For more information about the modules and input/export formats see supplementary information (SI).

3 Test case

To showcase TCI's ability to identify dynamic signalling states in single-cell time-lapse data, we study ERK activity dynamics in PC-12 cells (Blum *et al.*, 2019). We used microfluidics to challenge cells with 3' pulses of fibroblast growth factor 2 (FGF2) at 20' intervals, and measured single-cell ERK activity using a fluorescence resonance energy transfer (FRET)-based biosensor (Fritz *et al.*, 2013). Image segmentation, cell tracking, and the computation of single-cell FRET ratios yielded hundreds of single-cell ERK activity trajectories. The INPUT (FGF2 pulses) – OUTPUT (single-cell ERK activity dynamics) relationship contains considerable

1

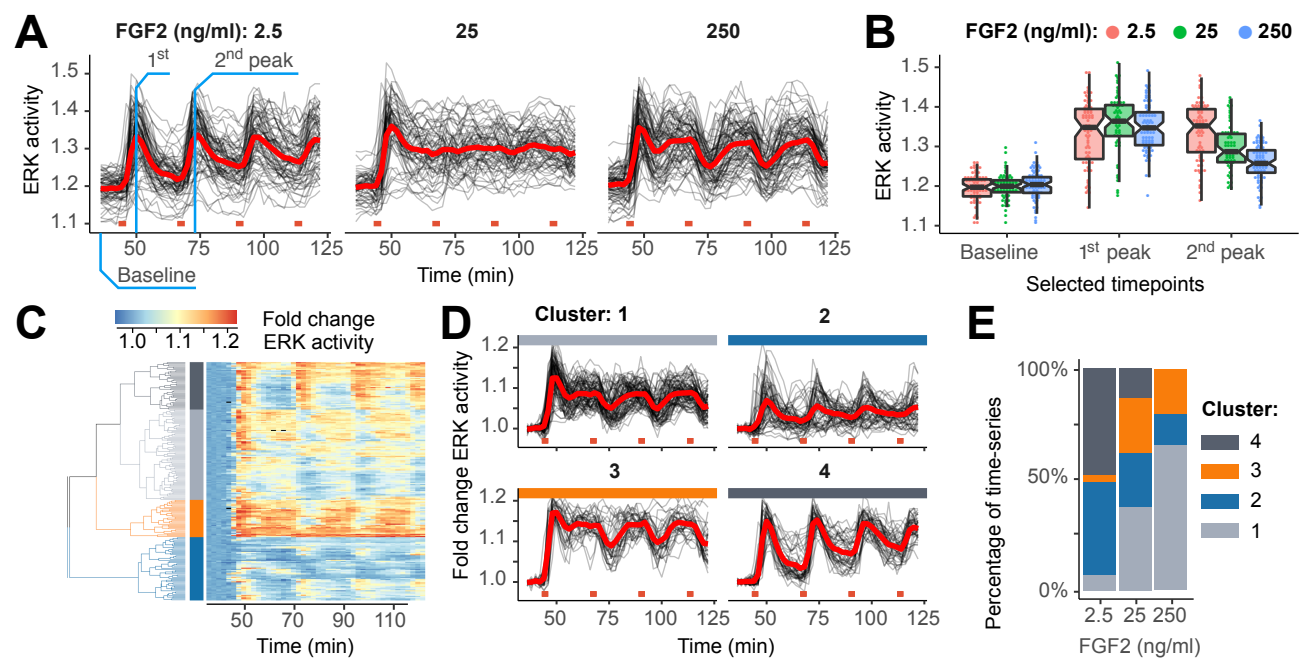


Fig. 1. (A) ERK activity in single cells stimulated with 3' pulses of FGF2 (red segments). Solid red lines – the population mean. Outliers removed manually; see Fig. S4. (B) Single-cell ERK activity before FGF2 application, and at the 1st and 2nd peak response. (C) Clustering of normalised and clustered ERK activity using Manhattan distance and complete linkage. Each row is a single time-series; black points – missing data. Dendrogram manually cut at 4 clusters to highlight main branches. (D) Time-series in each of the 4 clusters. (E) Contribution of clusters to experimental treatments. 25 ng/ml FGF2 is a mix of in- and anti-phasic responses, even though the mean of this population in A is flat. All plots generated with TCI.

information about the underlying structure of the MAPK network. Using TCI we identified important population dynamics.

At 2.5 ng/ml FGF2, peaks of ERK activity follow the stimulus, while at 250 ng/ml FGF2, the response is in anti-phase with the stimulus indicated with red segments in Fig. 1A. ERK activity at selected time points (Fig. 1B) shows that the anti-phasic response to 250 ng/ml FGF2 ensues after the 1st peak when ERK activity remains high and only transiently decreases in response to subsequent pulses. We discuss the biological mechanism behind this effect in Blum *et al.*, 2019.

At 2.5 and 250 ng/ml FGF2, the majority of cells in the population are synchronous and have the characteristic in- and anti-phasic behaviours reflected by the population mean (Fig. 1A). An intermediate dose of 25 ng/ml FGF2 evokes high variability in ERK activity at the single-cell level, leading to a flat population mean. To dissect this population we perform clustering with TCI (Fig. 1C-E). The screencast, including the removal of outliers, normalisation, and plotting different data views, is available online at <https://youtu.be/pwRqUzf5HMs>.

We cut the dendrogram from hierarchical clustering into four branches (Fig. 1C). We then inspected the composition of clusters (Fig. 1D) and calculated the contribution of each cluster to the experimental condition (Fig. 1E). Clusters 2 and 4 include trajectories where ERK is activated transiently right after the stimulation. In contrast, trajectories that transiently decrease after the pulse of FGF2 comprise clusters 1 and 3 (Fig. 1D). In line with an earlier observation, the 2.5 ng/ml FGF2 stimulation consists of *in-phase* clusters 2 and 4, while 250 ng/ml FGF2 consists of *anti-phase* clusters 1 and 3. The intermediate 25 ng/ml FGF2, for which the mean ERK response was flat, consists of a mixture of in- and anti-phasic responses (Fig. 1E).

4 Conclusion

Single-cell, time-lapse fluorescence microscopy can generate hundreds of time-series from a single experiment. Successive rounds of clustering

with different parameters are required to identify dynamic patterns in such complex datasets. TCI's interactivity enables rapid exploration and therefore greatly accelerates time-series analysis, even for users without coding experience. The mining of single-cell dynamic patterns can radically enhance our understanding of signalling as demonstrated in the case study. Clustering of single-cell responses of signalling pathway activation revealed a complex behaviour within the population. This allowed us to discriminate between mechanistic models of the FGF2/MAPK signalling pathway (Blum *et al.*, 2019). TCI can also aid the discovery of behavioural patterns in longitudinal studies such as those conducted in social sciences. Owing to a highly modular structure of the app it is easy to incorporate other analytic approaches.

Funding

This work was supported by the Swiss National Science Foundation (grants 31003A-163061 and 51PHPO-163583), and the Swiss Cancer League.

References

- Blum, Y. *et al.* (2019) Temporal perturbation of ERK dynamics reveals network architecture of FGF2-MAPK signaling, *BioRxiv* 629287 [Preprint], doi:10.1101/629287.
- Fritz, R.D. *et al.* (2013) A versatile toolkit to produce sensitive FRET biosensors to visualize signaling in time and space, *Sci Signal.*, **6**(285).
- Lahav, G., and Purvis, J.E. (2013) Encoding and decoding cellular information through signaling dynamics, *Cell*, **152**(5), 945-956.
- Lee, T.K., and Covert, M.W. (2010) High-throughput, single-cell NF- κ B dynamics, *Curr. Opin. Genet. Dev.*, **20**(6), 677-683.
- Ryu, H. *et al.* (2015) Frequency modulation of ERK activation dynamics rewires cell fate, *Mol Syst Biol.*, **11**(838).
- Witten, D.M., and Tibshirani, R. (2010) A framework for feature selection in clustering, *J Am Stat Assoc.*, **105**(490), 713-726.