Bioinformatics, 2019, 1–3
doi: 10.1093/bioinformatics/btz827
Advance Access Publication Date: 7 November 2019
Applications Note

Systems biology

SigHotSpotter: scRNA-seq-based computational tool to control cell subpopulation phenotypes for cellular rejuvenation strategies

Srikanth Ravichandran[1], András Hartmann[1,2] and Antonio del Sol[1,3,4,5,*]

[1]Computational Biology Group, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg, [2]Institute for Globally Distributed Open Research and Education (IGDORE), [3]IKERBASQUE, Basque Foundation for Science, Bilbao 48013, Spain, [4]CIC bioGUNE, Bizkaia Technology Park, Derio 48160, Spain and [5]Moscow Institute of Physics and Technology, Dolgoprudny 141701, Russia

*To whom correspondence should be addressed.

Associate Editor: Jan Gorodkin

Received on June 5, 2019; revised on October 8, 2019; editorial decision on November 2, 2019; accepted on November 5, 2019

Abstract

Summary: Single-cell RNA-sequencing is increasingly employed to characterize disease or ageing cell subpopulation phenotypes. Despite exponential increase in data generation, systematic identification of key regulatory factors for controlling cellular phenotype to enable cell rejuvenation in disease or ageing remains a challenge. Here, we present SigHotSpotter, a computational tool to predict hotspots of signaling pathways responsible for the stable maintenance of cell subpopulation phenotypes, by integrating signaling and transcriptional networks. Targeted perturbation of these signaling hotspots can enable precise control of cell subpopulation phenotypes. SigHotSpotter correctly predicts the signaling hotspots with known experimental validations in different cellular systems. The tool is simple, user-friendly and is available as web-server or as stand-alone software. We believe SigHotSpotter will serve as a general purpose tool for the systematic prediction of signaling hotspots based on single-cell RNA-seq data, and potentiate novel cell rejuvenation strategies in the context of disease and ageing.

Availability and implementation: SigHotSpotter is at https://SigHotSpotter.lcsb.uni.lu as a web tool. Source code, example datasets and other information are available at https://gitlab.com/srikanth.ravichandran/sighotspotter.

Contact: antonio.delsol@uni.lu

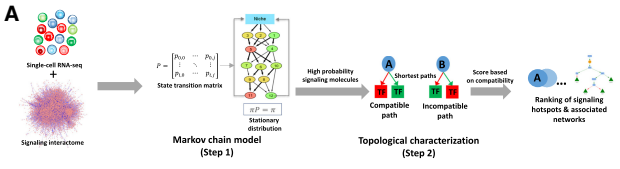
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The ability to control cellular phenotypes offers a great potential for developing novel regenerative medicine strategies. In particular, rejuvenation strategies for counteracting the detrimental effect of the aged or diseased niche that impairs normal cellular functioning are essential (Cheung and Rando, 2013; Del Sol *et al.*, 2019; Lane *et al.*, 2014; Neves *et al.*, 2017). Advances in single-cell RNA-seq that allows for profiling of distinct cell subpopulations could aid in this endeavor. However, despite increasing amount of data generation, there is a lack of computational approaches that leverages single-cell omics data to identify specific factors that can enable cell rejuvenation. Here, we present a general computational tool, SigHotSpotter, which relies on a probabilistic Markov chain model of signal transduction, previously developed by our lab, for the prediction of hotspots (key molecules) of signaling pathways that are constantly activated/inhibited by the niche that maintain neural stem cells in a quiescence state (Kalamakis *et al.*, 2019). We define signaling hotspots as those specific molecules that are involved in the sustained transmission of the external niche signals for the stable maintenance of the cell subpopulation phenotypes. Importantly, the tool aims at predicting hotspots, that exhibit highest signal flux through them in a sustained manner, rather than inferring the whole signaling pathways. Functionally, such hotspots are more likely to transmit the constitutive signals from the niche for phenotype maintenance, in contrast to strong but transient signals which are usually associated with a change in cellular response or phenotype (Wang and Wagers, 2011). With the increasing amount of single-cell RNA-seq data being generated, especially in the context of ageing and disease, SigHotSpotter can be of general utility for predicting signaling hotspots that that maintain cell subpopulation phenotypes. Further, this could aid the development of novel cell rejuvenation strategies that aim to counteract the detrimental effect of the diseased or aged niche.

2 Materials and methods

The key steps involved in SigHotSpotter are represented in Figure 1A. Detailed description of the method is provided in the Supplementary



Cell Type	Signaling molecule	SPIA	PATHIFIER	Enrichr	GSEA	SigHotSpotter
ESCs	Gsk3b/Wnt signaling pathway	No	N/A	No	No	Yes
	Mek1/MAPK pathway	Yes	N/A	No	No	Yes
ST-HSCs	Akt1/Akt mTOR signaling	No	No	Yes[1]	No	Yes
	Mek1/MAPK pathway	No	Yes	No	No	Yes
	Gsk3b/Wnt signaling pathway	No	No	No	No	Yes
LT-HSCs	Akt1/Akt mTOR signaling	No	Yes	Yes[1]	No	Yes
	Mek1/MAPK pathway	No	Yes	No	No	Yes
	Gsk3b/Wnt signaling pathway	No	Yes	No	No	Yes
Bulge HFSCs	Bmpr1a/BMP signaling pathway	Yes[2]	No	No	No	Yes
	Gsk3b/Wnt signaling pathway	No	No	No	No	Yes
Telogen HG	Bmpr1a/BMP signaling pathway	Yes[2]	No	No	No	Yes
	Gsk3b/Wnt signaling pathway	No	No	No	No	Yes
OPCs	Notch1/Notch signaling pathway	No	No	No	Yes[3]	Yes
	Gsk3b/Wnt signaling pathway	No	No	No	No	Yes
	Fyn/Src signaling pathway	No	No	No	No	Yes

Fig. 1. SigHotSpotter steps involved and case study application. (**A**) Series of steps involved in SigHotSpotter. (**B**) Model of a toy-signaling network in two different phenotypes. The node colors in the network represent the steady state probability of the signal to be in a specific node (signaling molecule). Inhibitory edges are shown in red and activation edges in green. The edge thickness represents the interaction probability of the two molecules inferred from single-cell RNA-seq. The inverted triangle nodes represent receptors/ligands, circles represent intermediate signaling molecules and the squares represent TFs. Nodes that exhibit both high steady state probability and compatibility with the downstream TF expression are identified as signaling hotspots. The black dotted circle indicates the higher compatibility of Node 12 in niche condition 2. (**C**) Results of SigHotSpotter for the case study applications, a comparison with other signaling pathways/network inference and enrichment methods. Comparison of SigHotSpotter with SPIA, Pathifier and differential expression based enrichment with Enrichr. Yes/No denote if the pathway was found significant, pathway inhibition and activation are represented by red and green color, respectively. If the pathways were not available for certain methods, we looked up similar pathways from KEGG, these are marked in superscript with notation: [1]PI3K-Akt signaling pathway, [2]TGF-beta signaling pathway and [3]Notch4 signaling pathway. (Color version of this figure is available at *Bioinformatics* online.)

Material. As an input, SigHotSpotter requires single-cell RNA-seq data. Further, it relies on a signaling interactome network, constructed by combining Reactome, Omnipath databases and transcriptional interactions from Metacore (Clarivate Analytics) obtained from Turei *et al.*, (2016), Wu *et al.*, (2010) and Zaffaroni *et al.*, (2019). In the first step, a state transition matrix is constructed based on the signaling interactome and the input single-cell RNA-seq data. The signal transduction process from the niche to intracellular signaling pathways is modeled as a finite discrete time-homogenous Markov chain (Kalamakis *et al.*, 2019). The stationary distribution of this Markov chain enables shortlisting those signaling molecules (defined as receptors, ligands, kinases and phosphatases) that exhibit high steady state probability, which reflects the high signal flux through them for a given phenotype (Fig. 1). This information alone is not sufficient to infer whether these molecules are in an active or inactive state to maintain the phenotype. In the second step, SigHotSpotter attempts to delineate the potential regulatory activity status of the high probability signaling molecules by employing a topological characterization of their compatibility with differential expression status of the downstream transcription factors (TFs; Fig. 1 and Supplementary Figs S1 and S2). A compatibility score is calculated for each high probability signaling molecule based on their net effect on the differentially expressed downstream TFs, via all the shortest paths in the network. This score relies on the steady state probabilities from the Markov chain model and serves to both classify and rank; the active and inactive signaling hotspots by their importance to maintain the corresponding phenotype (Fig. 1). Finally, we use Igraph (Csardi and Nepusz, 2006) implementation of Dijkstra's shortest paths algorithm (Dijkstra, 1959) to extract the subnetwork controlled by the predicted hotspots. A screenshot of SigHotSpotter web tool is shown in Supplementary Figure S3.

3 Results

We demonstrate SigHotSpotter as a case study in four different cellular systems based on single-cell RNA-seq data from embryonic stem cells (ESCs) (Kolodziejczyk *et al.*, 2015), hematopoietic stem cells (Kowalczyk *et al.*, 2015), hair-follicle stem cells (Yang *et al.*, 2017) and oligodendrocyte progenitor cells (Marques *et al.*, 2016). The list of computational predictions and associated literature support is listed in Supplementary Table S1. ESCs maintained under *in vitro* culture conditions serve as good model system to initially assess the performance of SigHotSpotter, since these cells can be stably maintained in different defined culture conditions such as 2i (inhibition of Gsk3b and Mek) or leukaemia inhibitory factor, and also exhibit condition dependent differences in their phenotypes (Ying *et al.*, 2008). The culture conditions employing 2i (inhibition of Gsk3b and Mek) is known to maintain the mESCs in a naive pluripotency state, whereas, the LIF alone maintain the mESCs in a relatively primed/metastable pluripotency state (Ying *et al.*, 2008). Importantly, sustained inhibition of Gsk3b and Mek is required for stable maintenance of naive pluripotency. In the example of mESCs (Kolodziejczyk *et al.*, 2015), where exact signaling molecules that are constantly inhibited by the niche (culture conditions in this case) are clearly known, SigHotSpotter could correctly predict the inhibition (i.e. as inactive) of Gsk3b and Mek (Map2k1) under 2i conditions (Supplementary Table S1 and Supplementary Fig. S1). Although, to our knowledge no general method currently exists for the task of identifying signaling hotspots that control cellular phenotypes, we compared the performance of SigHotSpotter with other general methods for signaling pathway/network inference and enrichment analysis that rely on only differential expression or network topology characterization (Drier *et al.*, 2013; Kuleshov *et al.*, 2016; Tarca *et al.*, 2009). Notably, these methods were not able identify either Wnt signaling or Map kinase signaling along with their deregulation status for mESCs phenotype control (Fig. 1). Other case study applications (Fig. 1 and Supplementary Table S2) and the comparison of SigHotSpotter with four other pathways/network inference and enrichment methods, namely, SPIA (Marques *et al.*, 2016), GSEA (Subramanian *et al.*, 2005), EnrichR (Wu *et al.*, 2010) and Pathifier (Drier *et al.*, 2013) are described in the Supplementary Material.

4 Discussion and conclusion

Computational methods that combine molecular interaction databases and genomics data have been very useful for the inference of dysregulated signaling pathways and networks, especially in the context of cancer (Leiserson *et al.*, 2015). However, these methods are not specifically built for the prediction of key molecules or hotspots that constantly mediate cell-extrinsic niche cues for the stable maintenance of the cellular phenotype. Furthermore, this is a challenge, since signal transduction involves several post-translational modifications, and is not a deterministic linear cascade of biochemical interactions (as often depicted in pathway diagrams), but rather a probabilistic process involving multiple protein–protein interactions (Ladbury and Arold, 2012). In this regard, although SigHotSpotter is based on transcriptomics data, it benefits from the heterogeneity of single-cell gene expression, and attempts to overcome some of these challenges by relying on a probabilistic model to infer signaling hotspots that most likely transmit the sustained niche-induced signals, rather than inferring the entire dysregulated signaling pathways. Hence, SigHotSpotter is qualitatively different from the plethora of methods for pathway enrichment or inference of functional signaling networks (Amadoz *et al.*, 2018), as it predicts specific signaling molecules and their regulatory effect on the cellular phenotype. In addition, the ranking of the hotspots along with their associated network controlling the downstream TFs will serve as a guide experimentalists to prioritize the predicted targets for further study.

In summary, SigHotSpotter employs single-cell RNA-seq data to serve as a general purpose tool for predicting signaling hotspots that control cell subpopulation phenotypes. Importantly, this can enable the development of cell rejuvenation strategies for counteracting the detrimental effect of the niche due to disease or ageing, where endogenous stem cells lose their activation potential.

Acknowledgement

We thank Sascha Jung for stimulating discussions on this topic.

Funding

This work was supported by University of Luxembourg IRP Grant [R-AGR-3227-11 to S.R.].

Conflict of Interest: none declared.

References

Amadoz,A. *et al*. (2018) A comparison of mechanistic signaling pathway activity analysis methods. *Brief. Bioinformatics*, doi: 10.1093/bib/bby040.

Cheung,T.H. and Rando,T.A. (2013) Molecular regulation of stem cell quiescence. *Nat. Rev. Mol. Cell Biol.*, **14**, 329–340.

Del Sol,A. *et al*. (2019) Computational strategies for niche-dependent cell conversion to assist stem cell therapy. *Trends Biotechnol.*, **37**, 687.

Dijkstra,E.W. (1959) A note on two problems in connexion with graphs. *Numer. Math.*, **1**, 269–271.

Drier,Y. *et al*. (2013) Pathway-based personalized analysis of cancer. *Proc. Natl. Acad. Sci. USA*, **110**, 6388–6393.

Csardi G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJ. Complex Syst.*, **1695**, 1–9.

Kalamakis,G. *et al*. (2019) Quiescence modulates stem cell maintenance and regenerative capacity in the aging brain. *Cell*, **176**, 1407–1419.e1414.

Kolodziejczyk,A.A. *et al*. (2015) Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*, **17**, 471–485.

Kowalczyk,M.S. *et al*. (2015) Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.*, **25**, 1860–1872.

Kuleshov,M.V. *et al*. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–97.

Ladbury,J.E. and Arold,S.T. (2012) Noise in cellular signaling pathways: causes and effects. *Trends Biochem. Sci.*, **37**, 173–178.

Lane,S.W. *et al*. (2014) Modulating the stem cell niche for tissue regeneration. *Nat. Biotechnol.*, **32**, 795–803.

Leiserson,M.D. *et al*. (2015) Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.*, **47**, 106–114.

Marques,S. *et al*. (2016) Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science*, **352**, 1326–1329.

Neves,J. *et al*. (2017) Rejuvenating strategies for stem cell-based therapies in aging. *Cell Stem Cell*, **20**, 161–175.

Subramanian,A. *et al*. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.

Tarca,A.L. *et al*. (2009) A novel signaling pathway impact analysis. *Bioinformatics*, **25**, 75–82.

Turei,D. *et al*. (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods*, **13**, 966–967.

Wang,L.D. and Wagers,A.J. (2011) Dynamic niches in the origination and differentiation of haematopoietic stem cells. *Nat. Rev. Mol. Cell Biol.*, **12**, 643–655.

Wu,G. *et al*. (2010) A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.*, **11**, R53.

Yang,H. *et al*. (2017) Epithelial-mesenchymal micro-niches govern stem cell lineage choices. *Cell*, **169**, 483–496. e413.

Ying,Q.L. *et al*. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.

Zaffaroni,G. *et al*. (2019) An integrative method to predict signalling perturbations for cellular transitions. *Nucleic Acids Res.*, **47**, e72.