Bioinformatics, 33(10), 2017, 1583–1585
doi: 10.1093/bioinformatics/btx022
Advance Access Publication Date: 20 January 2017
Applications Note

OXFORD

Systems biology

NetLand: quantitative modeling and visualization of Waddington's epigenetic landscape using probabilistic potential

Jing Guo[1,2], Feng Lin[1], Xiaomeng Zhang[1], Vivek Tanavde[2] and Jie Zheng[1,3,4]*

[1]Biomedical Informatics Lab, School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore, [2]Bioinformatics Institute, Agency for Science, Technology, and Research (A*STAR), Singapore 138671, Singapore, [3]Genome Institute of Singapore, A*STAR, Singapore 138672, Singapore and [4]Complexity Institute, Nanyang Technological University, Singapore 637723, Singapore

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on June 6, 2016; revised on December 8, 2016; editorial decision on January 13, 2017; accepted on January 18, 2017

Abstract

Summary: Waddington's epigenetic landscape is a powerful metaphor for cellular dynamics driven by gene regulatory networks (GRNs). Its quantitative modeling and visualization, however, remains a challenge, especially when there are more than two genes in the network. A software tool for Waddington's landscape has not been available in the literature. We present NetLand, an open-source software tool for modeling and simulating the kinetic dynamics of GRNs, and visualizing the corresponding Waddington's epigenetic landscape in three dimensions without restriction on the number of genes in a GRN. With an interactive and graphical user interface, NetLand can facilitate the knowledge discovery and experimental design in the study of cell fate regulation (e.g. stem cell differentiation and reprogramming).

Availability and Implementation: NetLand can run under operating systems including Windows, Linux and OS X. The executive files and source code of NetLand as well as a user manual, example models etc. can be downloaded from http://netland-ntu.github.io/NetLand/.

Contact: zhengjie@ntu.edu.sg

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

1 Introduction

Waddington's epigenetic landscape was originally introduced as a theoretical model to explain the generation of distinct cell types (Waddington, 1957). Based on its essential idea, the elevation of a point on the landscape, which represents a cell state, is inversely related to the stability of the cell state (Zhou *et al.*, 2012). Thus changes of cell types are represented by transitions overcoming energy barriers between different attractors.

Recently, computational models of Waddington's epigenetic landscape have been developed based on the dynamics of gene regulatory networks (GRNs). The transcriptional regulation by GRNs is a fundamental process that orchestrates the gene activities driving the change of cell fate. Existing methods for modeling the GRN-based landscape mostly use Boolean networks (Álvarez-Buylla *et al.*, 2008; Davila-Velderrain *et al.*, 2015; Flöttmann *et al.*, 2012) or differential equations (Bhattacharya *et al.*, 2011; Li and Wang, 2013). In addition, data-driven models of Waddington's landscape based on network entropy (Banerji *et al.*, 2013) or Hopfield network (Taherian Fard *et al.*, 2016; Lang *et al.*, 2014; Maetschke and Ragan, 2014) have been proposed recently, which represent an emerging trend for integrating dynamical models with experimental data.

To visualize the Waddington's epigenetic landscape in three dimensions, typically the *x*-axis and *y*-axis correspond to the

expression levels of two marker genes that represent cell states, while the *z*-axis is reserved for the potential or energy. However, to retain global information for the whole GRN with more than two genes, dimensionality reduction techniques would be needed. For the 52-node GRN (Li and Wang, 2013), the method of root mean squared distance (RMSD) was used to reduce 52 dimensions into two new coordinates, representing distances from a state point to the stem cell attractor and differentiation state attractor, respectively. However, RMSD is applicable only to systems with exactly two attractors, which is not the case for many GRNs. In model of landscape based on the Hopfield network (Taherian Fard *et al.*, 2016; Maetschke and Ragan, 2014), principal component analysis (PCA) was used for dimensionality reduction. However, as a linear method, PCA may not be able to reflect the non-linear structure of the data. Non-linear dimensionality reduction methods, e.g. a Bayesian Gaussian process latent variable model (Bayesian GPLVM), have been used to analyze high dimensional gene expression data (Zwiessele and Lawrence, 2016). In this paper, we chose the method of Gaussian process dynamical model (GPDM) (Lawrence, 2005; Wang *et al.*, 2008). Based on a non-linear latent variable model, GPDM can take into account the dynamics of time-series gene expression data, and is capable of projection in both directions between the low-dimensional latent space and the high-dimensional observation space.

For the theoretical models of GRN and landscape to be widely used, we need user-friendly software to automate the steps of model construction, dynamic simulation and visualization of the Waddington's epigenetic landscape, etc., so that users without experience in computational modeling can also benefit from these models. Unfortunately, such a software tool has not been available in the literature. In response, we present NetLand, an open-source software for the modeling and visualization of Waddington's epigenetic landscape (Fig. 1). It provides a user-friendly interface to facilitate model construction, dynamical simulation and landscape plotting.

2 Implementation and the major functionality

2.1 Model construction and simulation

NetLand is capable of loading network structure files or computational models in the format of SBML, e.g. models extracted from BioModels database (Chelliah *et al.*, 2015). The transcriptional regulation within the network is described with ordinary differential equations (ODEs) or stochastic differential equations (SDEs). The algebraic equations depict the instantaneous changes of molecular concentrations over time in the form of Hill equations. Users can modify the network structure and kinetic model, e.g. adding/deleting genes or edges, changing parameter values, equations etc.

Three simulators are implemented based on numerical algorithms of differential equations. The Runga–Kutta method using finite difference methods is implemented for the approximate solution of a deterministic model with a set of ODEs. The stochastic model described as Langevin equations (SDEs) (Simpson *et al.*, 2004) is numerically solved using the Euler–Maruyama method (Higham, 2001). The above two methods are suitable for models with high molecular concentrations. When the copy numbers of molecules are small, leading to significant stochastic effects, users can choose Gillespie algorithm (Gillespie, 1976).

2.2 Mapping Waddington's epigenetic landscape

NetLand provides a probabilistic implementation of Waddington's epigenetic landscape based on the notion that the probability of a gene expression state determines the cellular stability. NetLand adopted the self-consistent mean field approximation method in Li and Wang (2013) and Zhou *et al.* (2012) to construct the landscape by calculating the quasi-potential. From the GRN, users can select two marker genes, of which the gene expression levels are the *x*-axis and *y*-axis for the landscape. The quasi-potential is the *z*-axis. Alternatively, users can choose GPDM to project high dimensional gene expression time-series data to the *x*–*y* plane in the landscape, using all genes to provide a global view of the network dynamics, and thereby making the landscape plot more comprehensive.

The user manual of NetLand provides a comprehensive guide for software installation, usage, running time, memory usage etc., and includes a set of case studies of multi-attractor systems.

3 Conclusion

With quantitative modeling of Waddington's epigenetic landscape, the systems dynamics of transcriptional regulation driving cellular phenotypic changes can be more easily analyzed and predicted. Therefore, by automating the modeling and visualization of Waddington's landscape,

NetLand can facilitate the studies of cell fate conversion and help design innovations in cell fate engineering and regenerative medicine.

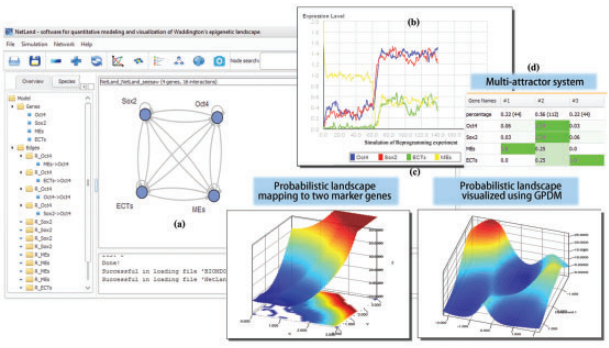

Fig. 1. A case study of a multi-attractor system by NetLand. The 4-gene regulatory network (**a**) for mouse stem cell reprogramming (Shu et al., 2013) has three attractors shown in (**d**). Based on the mathematical model, time-series data of reprogramming experiment are simulated (**b**), and the network dynamics is visualized in Waddington's epigenetic landscape (**c**)

Funding

This work has been supported by the MOE AcRF Tier 1 Seed Grant on Complexity (RGC2/13) and MOE AcRF Tier 1 Grant (2015-T1-002-094, RG120/15), Ministry of Education, Singapore.

Conflict of Interest: none declared.

References

Álvarez-Buylla,E.R. *et al.* (2008) Floral morphogenesis: stochastic explorations of a gene network epigenetic landscape. *PLoS One*, **3**, e3626.

Banerji,C.R. *et al.* (2013) Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Sci. Rep.*, **3**, 3039.

Bhattacharya,S. *et al.* (2011) A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC Syst. Biol.*, **5**, 85.

Chelliah,V. *et al.* (2015) BioModels: ten-year anniversary. *Nucleic Acids Res.*, **43**, D542–D548.

Davila-Velderrain,J. *et al*. (2015) Reshaping the epigenetic landscape during early flower development: induction of attractor transitions by relative differences in gene decay rates. *BMC Syst. Biol*., **9**, 20.

Flöttmann,M. *et al*. (2012) A stochastic model of epigenetic dynamics in somatic cell reprogramming. *Front. Physiol*., **3**, 216.

Gillespie,D.T. (1976) A general method for numerically simulating stochastic time evolution of coupled chemical-reactions. *J. Comput. Phys*., **22**, 403–434.

Higham,D.J. (2001) An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev*., **43**, 525–546.

Lang,A.H. *et al*. (2014) Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. *PLoS Comput. Biol*., **10**, e1003734.

Lawrence,N. (2005) Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *J. Mach. Learn. Res*., **6**, 1783–1816.

Li,C., and Wang,J. (2013) Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS Comput. Biol*., **9**, e1003165.

Maetschke,S.R., and Ragan,M.A. (2014) Characterizing cancer subtypes as attractors of Hopfield networks. *Bioinformatics*, **30**, 1273–1279.

Shu,J. *et al*. (2013) Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell*, **153**, 963–975.

Simpson,M.L. *et al*. (2004) Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *J. Theor. Biol*., **229**, 383–394.

Taherian Fard,A. *et al*. (2016) Not just a colourful metaphor: modelling the landscape of cellular development using Hopfield networks. *npj Syst. Biol. Appl*., **2**, 16001.

Waddington,C.H. (1957) The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser, George Allen & Unwin; 1st edition.

Wang,J.M. *et al*. (2008) Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell*., , **30**, 283–298.

Zhou,J.X. *et al*. (2012) Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface*, **9**, 3539–3553.

Zwiessele,M., and Lawrence,N.D. (2016) Topslam: Waddington landscape recovery for single cell experiments. *bioRxiv*, 057778.