Using deep learning to model the hierarchical structure and function of a cell

Jianzhu Ma, Michael Ku Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak, Roded Sharan & Trey Ideker

Nature Methods 15, 290–298 (2018)

Abstract

Although artificial neural networks are powerful classifiers, their internal structures are hard to interpret. In the life sciences, extensive knowledge of cell biology provides an opportunity to design visible neural networks (VNNs) that couple the model’s inner workings to those of real systems. Here we develop DCell, a VNN embedded in the hierarchical structure of 2,526 subsystems comprising a eukaryotic cell (http://d-cell.ucsd.edu/). Trained on several million genotypes, DCell simulates cellular growth nearly as accurately as laboratory observations. During simulation, genotypes induce patterns of subsystem activities, enabling in silico investigations of the molecular mechanisms underlying genotype–phenotype associations. These mechanisms can be validated, and many are unexpected; some are governed by Boolean logic. Cumulatively, 80% of the importance for growth prediction is captured by 484 subsystems (21%), reflecting the emergence of a complex phenotype. DCell provides a foundation for decoding the genetics of disease, drug resistance and synthetic life.
References

1 Farabet, C., Couprie, C., Najman, L. & Lecun, Y. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1915–1929 (2013).

2 Mikolov, T., Deoras, A., Povey, D., Burget, L. & Černocký, J. Strategies for training large scale neural network language models. In *2011 IEEE Workshop on Automatic Speech Recognition Understanding* 196–201 (IEEE, 2011).

3 Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process. Mag.* **29**, 82–97 (2012).

4 Sainath, T.N., Mohamed, A.R., Kingsbury, B. & Ramabhadran, B. Deep convolutional neural networks for LVCSR. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 8614–8618 (IEEE, 2013).

https://www.nature.com/articles/nmeth.4627#citeas
5 Collobert, R. et al. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **12**, 2493–2537 (2011).

6 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

7 Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

8 Brosin, H.W. An introduction to cybernetics. *Br. J. Psychiatry* **104**, 590–592 (1958).

9 The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2016).

10 Dutkowski, J. et al. A gene ontology inferred from molecular networks. *Nat. Biotechnol.* **31**, 38–45 (2013).

11 Kramer, M., Dutkowski, J., Yu, M., Bafna, V. & Ideker, T. Inferring gene ontologies from pairwise similarity data. *Bioinformatics* **30**, i34–i42 (2014).

12 Carvunis, A.-R. & Ideker, T. Siri of the cell: what biology could learn from the iPhone. *Cell* **157**, 534–538 (2014).

13 Yu, M.K. et al. Translation of genotype to phenotype by a hierarchy of cell subsystems. *Cell Syst.* **2**, 77–88 (2016).
14 Copley, S.D. Moonlighting is mainstream: paradigm adjustment required. *BioEssays* **34**, 578–588 (2012).

15 Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).

16 Costanzo, M. et al. The genetic landscape of a cell. *Science* **327**, 425–431 (2010).

17 Szappanos, B. et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet.* **43**, 656–662 (2011).

18 Lee, I. et al. Predicting genetic modifier loci using functional gene networks. *Genome Res.* **20**, 1143–1153 (2010).

19 Pandey, G. et al. An integrative multi-network and multi-classifier approach to predict genetic interactions. *PLoS Comput. Biol.* **6**, e1000928 (2010).

20 Xu, C., Wang, S., Thibault, G. & Ng, D.T.W. Futile protein folding cycles in the ER are terminated by the unfolded protein O-mannosylation pathway. *Science* **340**, 978–981 (2013).

21 Free, S.J. Fungal cell wall organization and biosynthesis. *Adv. Genet.* **31**, 33–82 (2013).

22 Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081–1086 (2011).
23 Scrimale, T., Didone, L., de Mesy Bentley, K.L. & Krysan, D.J. The unfolded protein response is induced by the cell wall integrity mitogen-activated protein kinase signaling cascade and is required for cell wall integrity in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **20**, 164–175 (2009).

24 Jonikas, M.C. et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).

25 Srivas, R. et al. A UV-induced genetic network links the RSC complex to nucleotide excision repair and shows dose-dependent rewiring. *Cell Rep.* **5**, 1714–1724 (2013).

26 Cadet, J., Sage, E. & Douki, T. Ultraviolet radiation-mediated damage to cellular DNA. *Mutat. Res.* **571**, 3–17 (2005).

27 Pareto, V. *Cours d’Économie Politique* (Librairie Droz, 1964).

28 Farrugia, G. & Balzan, R. Oxidative stress and programmed cell death in yeast. *Front. Oncol.* **2**, 64 (2012).

29 Pujol-Carrion, N. & de la Torre-Ruiz, M.A. Glutaredoxins Grx4 and Grx3 of *Saccharomyces cerevisiae* play a role in actin dynamics through their Trx domains, which contributes to oxidative stress resistance. *Appl. Environ. Microbiol.* **76**, 7826–7835 (2010).

30 Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
31 Kim, H. et al. YeastNet v3: a public database of data-specific and integrated functional gene networks for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **42**, D731–D736 (2014).

32 Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

33 Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).

34 Chen, W.W., Niepel, M. & Sorger, P.K. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* **24**, 1861–1875 (2010).

35 Szappanos, B. et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet.* **43**, 656–662 (2011).

36 Karr, J.R. et al. A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).

37 Lipton, Z.C. The mythos of model interpretability. Preprint at [https://arxiv.org/abs/1606.03490](https://arxiv.org/abs/1606.03490) (2017).
38 Mahendran, A. & Vedaldi, A. Understanding deep image representations by inverting them. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 5188–5196 (IEEE, 2015).

39 Vondrick, C., Khosla, A., Malisiewicz, T. & Torralba, A. Hoggles: Visualizing object detection features. In *Proceedings of the IEEE International Conference on Computer Vision* 1–8 (IEEE, 2013).

40 Weinzaepfel, P., Jégou, H. & Pérez, P. Reconstructing an image from its local descriptors. In *CVPR 2011* 337–344 (IEEE, 2011).

41 Chakraborty, S. et al. Interpretability of deep learning models: a survey of results. Paper presented at IEEE Smart World Congress 2017 Workshop: DAIS 2017, Workshop on Distributed Analytics InfraStructure and Algorithms for Multi-Organization Federations, San Francisco, CA, USA, 7–8 August 2017.

42 Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. Preprint at https://arxiv.org/abs/1409.0473 (2016).

43 Lei, T., Barzilay, R. & Jaakkola, T. Rationalizing neural predictions. Preprint at https://arxiv.org/abs/1606.04155 (2016).

44 Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24 (2012).
45 Szegedy, C. et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition* 1–9 (IEEE, 2015).

46 Lee, C.-Y., Xie, S., Gallagher, P.W., Zhang, Z. & Tu, Z. *Deeply-Supervised Nets.* in *AISTATS* 2, 5 (2015).

47 Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Preprint at [https://arxiv.org/abs/1502.03167](https://arxiv.org/abs/1502.03167) (2015).

48 Kingma, D.P. & Ba, J. Adam: a method for stochastic optimization. Preprint at [https://arxiv.org/abs/1412.6980](https://arxiv.org/abs/1412.6980) (2017).

49 Rumelhart, D.E., Hinton, G.E. & Williams, R.J. Learning representations by back-propagating errors. *Nature* 323, 533–536 (1986).

50 Alain, G. & Bengio, Y. Understanding intermediate layers using linear classifier probes. Preprint at [https://arxiv.org/abs/1610.01644](https://arxiv.org/abs/1610.01644) (2016).

51 Franz, M. et al. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 32, 309–311 (2016).

52 Bostock, M., Ogievetsky, V. & Heer, J. D³: data-driven documents. *IEEE Trans. Vis. Comput. Graph.* 17, 2301–2309 (2011).
Acknowledgements

We gratefully acknowledge support for this work provided by grants from the National Institutes of Health to T.I. (TR002026, GM103504, CA209891, ES014811). We also wish to thank T. Sejnowski and M. Kramer for very helpful comments during development of this work.

Author information

Jianzhu Ma, Michael Ku Yu and Samson Fong: These authors contributed equally to this work.

Authors and Affiliations

Department of Medicine, University of California San Diego, La Jolla, California, USA
Jianzhu Ma, Michael Ku Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak & Trey Ideker

Program in Bioinformatics, University of California San Diego, La Jolla, California, USA
Michael Ku Yu & Trey Ideker

Department of Bioengineering, University of California San Diego, La Jolla, California, USA
Samson Fong & Trey Ideker

Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel
Roded Sharan
Contributions
J.M., M.K.Y., S.F., R.S. and T.I. designed the study and developed the conceptual ideas. J.M. implemented the main algorithm. M.K.Y. collected all the input sources. J.M. and S.F. implemented all other computational methods and conducted analysis. J.M., M.K.Y., S.F. and T.I. wrote the manuscript with suggestions from the other authors. J.M., M.K.Y., S.F., K.O., E.S. and B.D. designed and developed the server.

Corresponding author
Correspondence to Trey Ideker.

Ethics declarations
Competing interests
T.I. is co-founder of Data4Cure, Inc. and has an equity interest. T.I. has an equity interest in Ideaya BioSciences, Inc. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

Integrated supplementary information

Supplementary Figure 1 Precision-recall curves for classification of negative genetic interactions.
Performance of DCell is compared to the same methods as in Fig. 2c. Genetic interactions with scores ≤ -0.08 are labeled as negative.

Supplementary Figure 2 CliXO top subsystem states for translation of genotype to growth.

a, Ranking of all CliXO subsystems by their importance in determining genetic interactions (RLIPP score, see Methods). Inset: ten highest-scoring subsystems. b-j, Two-dimensional state maps of informative subsystems from (a), in which each subsystem’s set of neuron states is reduced to the first two Principal Components (PCs). Each point
represents the subsystem state induced by a genotype, with point color indicating the corresponding growth phenotype (genetic interaction score).

**Supplementary Figure 3 Calculating relative local improvement in predictive power (RLIPP).**

a, Two L2-regularized linear regression models are fit to predict phenotype using either the neurons of a parent subsystem (bottom) or the neurons of that subsystem’s children (top). b-c, Measured versus predicted phenotype (genetic interactions) for the children-based model (b) or the parent-based model (c). The example values are for the “DNA repair” subsystem. d, The RLIPP score is calculated from the Spearman correlation of both models.

**Supplementary information**

**Supplementary Text and Figures**
Supplementary Figures 1–3

**Life Sciences Reporting Summary**

**Supplementary Table 1**
RLIPP scores for subsystems in the Gene Ontology and CliXO

**Supplementary Table 2**
Boolean logic approximating the states of subsystems in the Gene Ontology and CliXO

**Rights and permissions**

**Reprints and permissions**

**About this article**

Cite this article

https://www.nature.com/articles/nmeth.4627#citeas
Ma, J., Yu, M., Fong, S. et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat Methods* **15**, 290–298 (2018). https://doi.org/10.1038/nmeth.4627

Received 19 July 2017  
Accepted 07 February 2018  
Published 05 March 2018

**Issue Date**  
01 April 2018

**DOI**  
https://doi.org/10.1038/nmeth.4627

**Subjects**  
Cell growth  • Computational models  • Genetic interaction

**This article is cited by**

**Prediction of anticancer drug sensitivity using an interpretable model guided by deep learning**  
Weixiong Pang, Ming Chen & Yufang Qin  
*BMC Bioinformatics* (2024)

**Applying interpretable machine learning in computational biology—pitfalls, recommendations and opportunities for new developments**  
Valerie Chen, Muyu Yang ... Jian Ma  
*Nature Methods* (2024)

**A deep learning model of tumor cell architecture elucidates response and resistance to CDK4/6 inhibitors**  
Sungjoon Park, Erica Silva ... Trey Ideker  
*Nature Cancer* (2024)

**Phenotype prediction using biologically interpretable neural networks on multi-cohort multi-omics data**

https://www.nature.com/articles/nmeth.4627#citeas
Incorporating knowledge of disease-defining hub genes and regulatory network into a machine learning-based model for predicting treatment response in lupus nephritis after the first renal flare

Ding-Jie Lee, Ping-Huang Tsai ... Yang-Hong Dai

Journal of Translational Medicine (2023)