

DAVID VAN DIJK

Machine Learning; Big Biomedical Data; Computational Biology

Yale Genetics & Computer Science

(+1) 917 325 3940 ◊ david.vandijk@yale.edu ◊ www.davidvandijk.com

EDUCATION

University of Amsterdam

2008 - 2013

Department of Computer Science

Ph.D. Computer Science

Dissertation title: The Symphony of Gene Regulation: Predicting gene expression and its cell-to-cell variability from promoter DNA sequence

Advisors: Jaap Kaandorp (UoA), Peter Sloot (UoA), Eran Segal (Weizmann Institute of Science)

Free University Amsterdam

2002 - 2008

Department of Computer Science

M.Sc. Computer Science

B.Sc. Computer Science

RESEARCH SUMMARY

My research focuses on the development of new machine learning algorithms for unsupervised learning on high-dimensional biological data, with an emphasis on single-cell RNA-sequencing data. Among other tools, I developed MAGIC, a data-diffusion method for denoising and imputing sparse and noisy single-cell data, PHATE, a dimensionality reduction and visualization method that emphasizes progression structure in single-cell and other high-dimensional data, and SAUCIE, a deep learning method for combined clustering, batch-correction, imputation and embedding of single-cell data.

The mission for my lab will be to develop new unsupervised machine learning methods, based on deep learning and spectral graph theory, that model biological data manifolds, in order to 1) map cellular state spaces, 2) predict state changes as a result of perturbations, 3) infer regulatory logic that generates the state space, and 4) integrate diverse data modalities into a unified 'view' of biology.

PUBLICATIONS

1. van Dijk, David, S. Gigante, A. Strzalkowski, G. Wolf, and S. Krishnaswamy. Modeling dynamics of biological systems with deep generative neural networks. *arXiv preprint arXiv:1802.03497 [Accepted at SampTA2019]*, 2018
2. S. Gigante, J. S. S. III, N. Vu, David van Dijk, K. Moon, G. Wolf, and S. Krishnaswamy. Compressed diffusion. *arXiv preprint arXiv:1902.00033 [Accepted at SampTA2019]*, 2019
3. A. Tong*, van Dijk*, David, J. S. S. III, M. Amodio, G. Wolf, and S. Krishnaswamy. Graph spectral regularization for neural network interpretability. *arXiv preprint arXiv:1810.00424 [Accepted at ICLR2019 workshop]*, 2018
4. van Dijk, David, R. Sharma, J. Nainys, K. Yim, P. Kathail, A. Carr, C. Burdziak, K. R. Moon, C. L. Chaffer, D. Pattabiraman, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 2018
5. van Dijk, David, E. Sharon, M. Lotan-Pompan, A. Weinberger, E. Segal, and L. B. Carey. Large-scale mapping of gene regulatory logic reveals context-dependent repression by transcriptional activators. *Genome research*, 27(1):87–94, 2017
6. K. R. Moon, J. Stanley, D. Burkhardt, van Dijk, David, G. Wolf, and S. Krishnaswamy. Manifold learning-based methods for analyzing single-cell rna-sequencing data. *Current Opinion in Systems Biology*, 2017
7. D. E. Lowther, B. A. Goods, L. E. Lucca, B. A. Lerner, K. Raddassi, van Dijk, David, A. L. Hernandez, X. Duan, M. Gunel, V. Coric, et al. Pd-1 marks dysfunctional regulatory t cells in malignant gliomas. *JCI insight*, 1(5), 2016
8. van Dijk, David, R. Dhar, A. M. Missarova, L. Espinar, W. R. Blevins, B. Lehner, and L. B. Carey. Slow-growing cells within isogenic populations have increased rna polymerase error rates and dna damage. *Nature communications*, 6:7972, 2015
9. L. Keren*, van Dijk*, David, S. Weingarten-Gabbay, D. Davidi, G. Jona, A. Weinberger, R. Milo, and E. Segal. Noise in gene expression is coupled to growth rate. *Genome research*, pages gr–191635, 2015
10. E. Sharon*, van Dijk*, David, Y. Kalma, L. Keren, O. Manor, Z. Yakhini, and E. Segal. Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome research*, pages gr–168773, 2014
11. van Dijk, David, O. Manor, and L. B. Carey. Publication metrics and success on the academic job market. *Current Biology*, 24(11):R516–R517, 2014
12. L. B. Carey*, van Dijk*, David, P. M. Slood, J. A. Kaandorp, and E. Segal. Promoter sequence determines the relationship between expression level and noise. *PLoS biology*, 11(4):e1001528, 2013
13. M. Dadiani, van Dijk, David, B. Segal, Y. Field, G. Ben-Artzi, T. Raveh-Sadka, M. Levo, I. Kaplow, A. Weinberger, and E. Segal. Two dna-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise. *Genome Research*, 2013
14. van Dijk, David, G. Ertaylan, C. A. Boucher, and P. M. Slood. Identifying potential survival strategies of hiv-1 through virus-host protein interaction networks. *BMC systems biology*, 4(1):96, 2010
15. van Dijk, David, P. M. Slood, J. Tay, M. C. Schut, et al. Individual-based simulation of sexual selection: A quantitative genetic approach. In *ICCS*, number 1, pages 2003–2011, 2010
16. S. Jaeger, G. Ertaylan, van Dijk, David, U. Leser, and P. Slood. Inference of surface membrane factors of hiv-1 infection through functional interaction networks. *PLoS One*, 5(10):e13139, 2010

Preprint / In Review:

17. K. R. Moon*, van Dijk*, David, Z. Wang*, D. Burkhardt, W. Chen, A. van den Elzen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf, et al. Visualizing transitions and structure for high dimensional data exploration. *bioRxiv [in revision at Nature Biotech]*, page 120378, 2017
18. M. Amodio*, van Dijk*, David, K. Srinivasan*, W. S. Chen, H. Mohsen, K. R. Moon, A. Campbell, Y. Zhao, X. Wang, M. Venkataswamy, A. Desai, R. V., P. Kumar, R. Montgomery, G. Wolf, and S. Krishnaswamy.

Exploring single-cell data with deep multitasking neural networks. *bioRxiv [in revision at Nature Methods]*, 2018

19. M. Amodio, van Dijk, David, R. Montgomery, G. Wolf, and S. Krishnaswamy. Out-of-sample extrapolation with neuron editing. *arXiv preprint arXiv:1805.12198 [submitted to ICML2019]*, 2018
20. W. S. Chen, N. Zivanovic, van Dijk, David, G. Wolf, B. Bodenmiller, and S. Krishnaswamy. Embedding the single-cell sample manifold to reveal insights into cancer pathogenesis and disease heterogeneity. *bioRxiv [In revision at Nature Methods]*, 2018
21. D. B. Burkhardt, J. S. Stanley, A. L. Perdigoto, S. A. Gigante, K. C. Herold, G. Wolf, A. Giraldez, D. van Dijk^{*†}, and S. Krishnaswamy^{*†}. Enhancing experimental signals in single-cell rna-sequencing data using graph signal processing. *bioRxiv:532846*, 2019
22. David van Dijk, D. Burkhardt, M. Amodio, A. Tong, G. Wolf, and S. Krishnaswamy. Finding archetypal spaces for data using neural networks. *arXiv:1901.09078 [Submitted to ICML2019]*, 2019

MANUSCRIPTS IN PREPARATION

1. S. Gigante, J. Stanley III, D. B. Burkhardt, D. van Dijk^{*†}, and S. Krishnaswamy^{*†}. Pseudo single-cell imputation using compressive sensing
2. J. Pappalardo, D. van Dijk^{*}, S. Krishnaswamy, and D. Hafler. Characterizing the immune state space in the brain using single-cell analysis

* equal contribution, † co-corresponding

TEACHING

My goal in teaching is to motivate students to drive their own learning by becoming curious about science, conduct investigations to follow their curiosity and even get involved in research. This, combined with my passion for interdisciplinary science, drives my style of teaching - I strive to motivate and inspire confidence in my students to step out of their comfort zone and learn about other disciplines. This is essential in learning how to be an interdisciplinary scientist, such as in my field of computational biology. I believe in creating a maximally diverse and inclusive environment in terms of education and personal background, and an environment where students are allowed and even expected to be creative. I have found that in such an environment students discover and develop their interest for science and even become excited about participating in research.

Yale University

2018 - present

Department of Computer Science

Course: Machine Learning for Biology

Duties: taught algorithms for machine learning

Yale University

2017 - 2018

Department of Computer Science

Course: Advanced Topics in Machine Learning & Data Mining

Duties: Advising student presentations, grading and guiding discussions

University of Amsterdam

2008 - 2010

Department of Computer Science

Course: Scientific Computing

Duties: taught algorithms and programming techniques for parallel computing

University of Amsterdam

2008 - 2010

Department of Computer Science

Course: Complex Systems Simulation

Duties: taught algorithms and simulation methods for deterministic and stochastic simulation of complex systems

University of Amsterdam

2004 - 2007

Department of Humanities

Course: Academic Research Skills

Duties: taught students from 20 different undergraduate programs computer science related to doing academic research. This involved, among other activities, teaching web-development (HTML and JavaScript)

RESEARCH POSITIONS

Yale University <i>Associate Research Scientist</i> Lab of Prof. Smita Krishnaswamy Dept. of Genetics, Dept. of Computer Science Machine learning; computational biology; single-cell data	<i>2016 - present</i>
Columbia University <i>Rubicon Fellow</i> , working with the Pe'er lab and Prof. Smita Krishnaswamy (Yale)	<i>2015 - 2016</i>
Weizmann Institute of Science <i>Postdoctoral Fellow</i> Computational biology; Single-cell gene expression Group of Prof. Eran Segal - Faculty of Mathematics and Computer science	<i>2013 - 2015</i>
Weizmann Institute of Science <i>Visiting Graduate Student Research Assistant</i> Computational biology; Quantitative modeling of gene regulation Group of Prof. Eran Segal - Faculty of Mathematics and Computer science	<i>2010 - 2013</i>
University of Amsterdam <i>Graduate Student Research Assistant</i> Computational Biology; Machine learning Groups of Prof. Jaap Kaandorp and Prof. Peter Sloot - Dept. of Computer Science	<i>2008 - 2013</i>
Nanyang Technological University, Singapore <i>Visiting Researcher for MSc thesis research project</i> Quantitative modeling of Stalk-eyed fly evolution EvoCom research group, School of Computer Engineering	<i>2007 - 2008</i>
Free University of Amsterdam <i>Research Assistant</i> Modeling and simulation of self-organizing systems Computational Intelligence Group, Section Artificial Intelligence, Dept. of Computer Science	<i>2007</i>

PROFESSIONAL ACTIVITIES

Invited program committee member for ECCB 2014
Reviewer for:
Nature Communications
Journal of Computational Science
International Conference of Computational Science (ICCS) 2008
RECOMB 2016, 2018
NIPS 2016, 2018

FELLOWSHIPS AND AWARDS

NWO Rubicon postdoctoral fellowship (2 years, €130k)	<i>2015</i>
Azrieli fellowship	<i>2013</i>
Weizmann postdoctoral fellowship	<i>2012</i>
EMBO short term fellowship	<i>2010</i>
Netherlands Bioinformatics Centre travel fellowship	<i>2010</i>
VU University travel fellowship	<i>2008</i>
1 st place DECOI 2007	<i>2007</i>
1 st place VU data mining competition	<i>2007</i>

INVITED TALKS

Princeton CS and Lewis-Sigler Institute	2019
EPFL Institute of Bioengineering	2019
University of Washington CS	2019
UCLA Computational Medicine	2019
UTSW Bioinformatics	2019
MIT Whitehead Institute	2019
Broad institute of MIT Harvard	2019
Crick Institute	2019
Hubrecht Institute	2019
Yale Cardiology	2019
Columbia University Mathematical Genomics	2019
Broad Institute - Methods, Inference and Algorithms (https://t.co/BFGSKRv31D)	2018
Yale Applied Math Seminar	2018
Yale Immunology department seminar	2018
Yale Genetics, lab of Prof. Sherman Weismann	2018
The Jackson Laboratory Genomic Medicine institute seminar	2018
IBM Zurich, group of Prof. Maria Rodriguez Martinez	2018
Memorial Sloan Kettering CBio seminar	2017
KAIST Korea, Bio-IT Healthcare Initiative Workshop on Single Cell Systems Biology	2017
Single-cell Genomics, Hinxton UK	2016
CRG Barcelona, genetics department seminar	2013
Hebrew University, lab of Prof. Nir Friedman	2012
Amsterdam Symposium on Networks in Biology	2008,2009
ICCS Conference Amsterdam	2008

PRESS COVERAGE

Coverage of the PI predictor (pipredictor.com) by Science and Nature

<http://www.sciencemag.org/careers/2014/06/career-moneyball>

<http://www.nature.com/news/computer-model-predicts-academic-success-1.15337>

Nature Reviews Genetics Research Highlight - Gene expression: Influences on noise

<http://www.nature.com/nrg/journal/v14/n4/full/nrg3448.html>

REFERENCES

Prof. Smita Krishnaswamy
Yale University, Depts. of Genetics and Computer Science
smita.krishnaswamy@yale.edu

Prof. Eran Segal
Weizmann Institute of Science, Faculty of Mathematics and Computer Science
eran.segal@weizmann.ac.il

Prof. Jaap Kaandorp
University of Amsterdam, Dept. of Computer Science
j.a.kaandorp@uva.nl

Prof. David Hafler
Yale University, Depts. of Neurology and Immunobiology, Chair of Neurology
david.hafler@yale.edu

Prof. James Noonan
Yale University, Dept. of Genetics
james.noonan@yale.edu

Prof. Lucas Carey
Universitat Pompeu Fabra, Dept. of Experimental and Health Sciences
lucas.carey@upf.edu

Prof. Antonio Giraldez
Yale University, Dept. of Genetics
antonio.giraldez@yale.edu

Prof. Peter Sliot
University of Amsterdam, Dept. of Computer Science
alice.matthews@yale.edu