

Supplementary information

An online notebook resource for reproducible inference, analysis and publication of gene regulatory networks

In the format provided by the
authors and unedited

**Supporting information for
An online notebook resource for reproducible inference, analysis, and publication of
gene regulatory networks**

Marouen Ben Guebila^{1,14}, Deborah Weighill^{1,12,14}, Camila M. Lopes-Ramos^{1,2}, Rebekka Burkholz^{1,13}, Romana T. Pop³, Kalyan Palepu⁴, Mia Shapoval⁵, Maud Fagny^{1,6}, Daniel Schlauch^{1,7}, Kimberly Glass^{1,2}, Michael Altenbuchinger⁸, Marieke L. Kuijjer^{3,9,10}, John Platig², and John Quackenbush^{1,2,11*}

¹ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

² Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

³ Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo, Norway

⁴ Harvard University, Cambridge, Massachusetts, USA

⁵ Boston University Academy, Boston, Massachusetts, USA

⁶ Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, Gif-sur-Yvette, France

⁷ Genospace, Boston, Massachusetts, USA

⁸ Institute of Medical Bioinformatics, University Medical Center Göttingen, Göttingen, Germany

⁹ Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

¹⁰ Leiden Center for Computational Oncology, Leiden University Medical Center, Leiden, The Netherlands

¹¹ Department of Data Science, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

¹² Present address: UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, North Carolina, USA

¹³ Present address: CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

¹⁴ These authors contributed equally to this work

* Corresponding author. Email: johnq@hsph.harvard.edu

Table of contents:

Supplementary Text S1 - Server configuration

Supplementary Table S1 - Case studies catalog in Netbooks and their corresponding references.

Supplementary Text S1 - Server configuration

The collection of case studies is hosted on GitHub (<https://github.com/netZoo/netbooks>), and deployed on the notebook server using Jupyter Hub (v 1.1) on an AWS EC2 instance. The data for each notebook is hosted on an AWS S3 bucket. Each Jupyter notebook instance comes with preinstalled netZooPy (v 0.7.2) on Python (v 3.8.3) and netZooR (v 0.9) on R (v 3.6.1) kernels. User access management was implemented with tmpauthenticator (v 0.6) and task scheduling on the server was implemented using systemd (v 0.14) to provide anonymous access, security, and isolation of user instances.

Supplementary Table S1 - Case studies catalog in Netbooks and their corresponding references.

Number	Case study	Network Zoo tool	Reference
1	Building PANDA and LIONESS Regulatory Networks from GTEx Gene Expression Data in R	PANDA, LIONESS	This paper
2	Comparing LIONESS Regulatory Networks using limma	LIONESS	This paper
3	Building PANDA Regulatory Networks from cell line and tissue expression data from GTEx Gene Expression Data in R	PANDA	Lopes-Ramos, Camila M., et al. "Regulatory network changes between cell lines and their tissues of origin." <i>BMC genomics</i> 18.1 (2017): 1-13.
4	Inferring Gene Regulatory Networks from GTEx Gene Expression Data in R with OTTER	OTTER	Weighill, Deborah, et al. "Gene regulatory network inference as relaxed graph matching." <i>Proceedings of the AAAI Conference on Artificial Intelligence</i> . Vol. 35. No. 11. 2021.
5	Finding drug candidates to reverse Lung Adenocarcinoma (LUAD)-induced gene regulation disruption using TCGA	PANDA, MONSTER	-
6	Estimating state transition in yeast cell cycle using MONSTER	MONSTER	Schlauch, Daniel, et al. "Estimating drivers of cell state transitions using gene regulatory network models." <i>BMC systems biology</i> 11.1 (2017): 1-10.

7	Differential gene targeting of pancreatic cancer subtypes	PANDA, LIONESS	Weighill, Deborah, et al. "Gene targeting in disease networks." <i>Frontiers in Genetics</i> 12 (2021): 501.
8	Processing TCGA gene expression data for network analysis	-	-
9	Gene regulatory network analysis in Glioblastoma	LIONESS, SAMBAR, ALPACA	Lopes-Ramos, Camila M., et al. "Regulatory network of PD1 signaling is associated with prognosis in glioblastoma multiforme." <i>Cancer research</i> 81.21 (2021): 5401-5412.
10	Generating 26 cancer gene regulatory network using TCGA datasets	PANDA	Ben Guebila, Marouen, et al. "GRAND: a database of gene regulatory network models across human conditions." <i>Nucleic acids research</i> 50.D1 (2022): D610-D621.
11	Exploring eQTL tissue regulation with CONDOR	CONDOR	Fagny, Maud, et al. "Exploring regulation in tissues with eQTL networks." <i>Proceedings of the National Academy of Sciences</i> 114.37 (2017): E7841-E7850.
12	Controlling the variance of PANDA networks	PANDA	This paper
13	Building a regulation prior network for netZoo tools	-	-
14	Benchmarking motif networks for the reconstruction of kidney cancer regulatory network	PANDA, CONDOR	-
15	Investigating potential regulatory relationships between TFs in breast cancer using DRAGON	DRAGON	Weighill, Deborah, et al. "DRAGON: Determining Regulatory Associations using Graphical models on multi-Omic Networks." <i>arXiv preprint arXiv:2104.01690</i> (2021).
16	Building a DRAGON miRNA gene regulatory network using CCLE data	DRAGON	Ben Guebila, Marouen, et al. "GRAND: a database of gene regulatory network models across human conditions." <i>Nucleic acids research</i> 50.D1 (2022): D610-D621.
17	Drug repurposing analysis in colon cancer	PANDA	Ben Guebila, Marouen, et al. "GRAND: a database of gene regulatory network models across

			human conditions." <i>Nucleic acids research</i> 50.D1 (2022): D610-D621.
18	Sex differences in lung adenocarcinoma	PANDA	This paper
19	Identification of Key Tissue-Specific, Biological Processes by Integrating Enhancer Information in Maize Gene Regulatory Networks	PANDA, LIONESS, CONDOR, ALPACA	Fagny, Maud, et al. "Identification of key tissue-specific, biological processes by integrating enhancer information in maize gene regulatory networks." <i>Frontiers in genetics</i> 11 (2021): 1703.