Rick Farouni | Curriculum Vitae

Massachusetts General Hospital-Harvard Medical School

✉ rfarouni@gmail.com   •   🌐 http://rfarouni.github.io/

I am a statistical modeler and applied machine learning researcher with expertise in probabilistic, Bayesian, and latent variable modeling. Currently, I am a post-doctoral research fellow in computational biology, working on the application of Bayesian statistics, machine learning, and deep learning to epigenomics and CRISPR data. I received a PhD in Quantitative Psychology (statistics applied to modeling psychological and neuroimaging data), a masters degree in Mathematical Statistics, and a masters degree in Psychometrics from the Ohio State University.

Experience

Postdoctoral Research Fellow

Massachusetts General Hospital-Harvard Medical School   **Massachusetts, USA**
Research Lab of Professor Luca Pinello, Molecular Pathology Unit   *06/2017-Present*

Projects in Progress

- **Histone Code**: Deep generative modeling (i.e. a Variational Autoencoder) of chromatin signal data across multiple cell types and histone marks with the goal of learning a latent representation of a continuous histone code. Slides: `https://docs.google.com/presentation/d/1MTAGBvQtS_LI17pGrsFeF-PtpEeiB3wl5E3J5ELNqck/edit#slide=id.p3`
- **Statistical Modeling of CRISPR Double Knockout Data**: A hierarchical negative binomial Bayesian framework for modeling CRISPR-Cas9 Double-Knockout screen count data. Slides: `https://drive.google.com/file/d/1AIdOuiQBlxVobVsyWsd5Sb12Q4vvo8e5/view`

Research Intern

Department of Biomedical Informatics Summer Internship Program   **Ohio, USA**
Research Lab of Professor Ewy Mathè, The Ohio State University   *05/2016-08/2016*

Projects

- **ALTRE**: An R package and a Shiny app for the analysis of data generated from genome-wide chromatin accessibility assays such as ATAC-seq and DNase-seq with the goal of identifying regulatory elements involved in the cancer epigenetic landscape.

University Teaching Assistant

The Ohio State University   **Ohio, USA**
Graduate Teaching Associate (Statistics)   *2013–2017*

Served as a Teaching Assistant for three courses: Repeated Measures Models, Covariance Structure Models, and Data Analysis in Psychology.

Test Preparation Instructor

Independent Tutor ****Moscow, Russia****
Teacher of General English, TOEFL, and GMAT *2001-2012*

Education

Academic Qualifications

PhD in Quantitative Psychology (i.e. *Applied Statistics*) ****Ohio, USA****
The Ohio State University *2015–2017*

Dissertation Topic:*'Application of Deep Latent Generative Models to the Unsupervised Learning of Chromatin States'*

Master of Science in Statistics ****Ohio, USA****
The Ohio State University *2012-2015*

Master's Degree in Psychometrics ****Ohio, USA****
The Ohio State University *2012-2014*

Thesis Project: *'Latent Variable Modeling of Categorical Item Responses in a Hierarchical Bayesian Framework'*

Bachelor's Degree in Psychology ****Pennsylvania, USA****
The Pennsylvania State University *2011-2012*

Publications and Software

Journal Papers

- Baskin, E.[†], Farouni, R.[†], and Mathè, E. (2016). ALTRE: workflow for defining ALTered Regulatory Elements using chromatin accessibility data. ***Bioinformatics*** doi: 10.1093/bioinformatics/btw688. Preprint available at `http://www.biorxiv.org/content/early/2016/10/14/080564.full.pdf+html` (**first co-author**)

- Pinello, L.[†], Farouni, R.[†], and Yuan, G-C. (2018). Haystack: systematic analysis of the variation of epigenetic states and cell-type specific regulatory elements. ***Bioinformatics*** `https://doi.org/10.1093/bioinformatics/bty031`. Preprint available at `https://doi.org/10.1101/199067` (**first co-author**)

- Clement, K.[†], Farouni, R.[†], Bauer, D. E., and Pinello, L. (2018). Design and analysis of unique molecular identifiers for deep amplicon sequencing. ***Bioinformatics*** (accepted). Preprint available at `https://www.biorxiv.org/content/biorxiv/early/2018/03/23/288118.full.pdf` (**first co-author**)

- Grüning, B., Dale, R., Sjödin, A., Rowe, J., Chapman, B. A., Tomkins-Tinch, C. H., The Bioconda Team, and Köster, J.(2018) Bioconda: A sustainable and comprehensive software distribution for the life sciences. ***Nature Methods***. (accepted), Preprint available at `https://www.biorxiv.org/content/early/2017/10/27/207092` (**Bioconda team member contributor**)

Preprints

o Farouni, R. (2017). A Contemporary Overview of Probabilistic Latent Variable Models. **_arXiv preprint_**. Preprint available at `https://arxiv.org/abs/1706.08137`

Dissertation and Thesis

o Farouni, R. (2017). Application of Deep Latent Generative Models to the Unsupervised Learning of Chromatin States. _PhD Dissertation_. Manuscript available at `https://etd.ohiolink.edu/!etd.send_file?accession=osu1492189894812539&disposition=inline`

o Farouni, R. (2014). Latent Variable Models of Categorical Responses in the Bayesian and Frequentist Frameworks. _Masters Thesis_. Manuscript available at `https://etd.ohiolink.edu/!etd.send_file?accession=osu1412374136&disposition=inline`

Software Development

o Histone Code VAE: Deep generative modeling of chromatin signal data across multiple cell types. GitHub Repo (private): `https://github.com/pinellolab/histone_code_vae`.

o Haystack: A Python bioinformatics pipeline for the identification of genomic regions of epigenetic variability across different cell-types, cell-type specific cis-regulatory elements, and their associated transcription factors. GitHub Repo: `https://github.com/pinellolab/haystack_bio`.

o ALTRE: A Workflow for Identifying ALTered Regulatory Elements using Chromatin Accessibility Data. GitHub Repo: `https://github.com/Mathelab/ALTRE`.

Journal Review Service

o **Psychometrika**
Ad Hoc Reviewer 2015

o **Psychological Methods**
Ad Hoc Reviewer 2016

Conference Presentations

o **Joint Statistical Meetings** **Seattle**
Poster Presentation 2015

Poster Title: Across-Subject Predictive Modeling of fMRI BOLD Responses to Faces using a sparse Bayesian Group Factor Analysis Model (available at `https://rfarouni.github.io/assets/posters/jsm2015.pdf`).

Awards and Fellowships

o **Graduate Student Conference Presentation Award**
The Ohio State University 2015

o **The Center for Cognitive and Brain Sciences Summer Graduate Fellowship**
The Ohio State University 2015

Project Proposal: _'Decoding the Pixels of the Face Image from the Voxels of fMRI BOLD Activity Patterns'_

○ **The Social and Behavioral Sciences Summer Fellowship**
The Ohio State University *2014*

○ **University Fellowship**
The Ohio State University *2012*

Technical Skill Set

○ **Statistics and Machine Learning**

- **Scientific Programming Languages:** Proficient in and comfortable transitioning between *R*, *Python* depending on computing goals. Familiar with *Julia*.
- **Deep Learning Frameworks:** Experienced in using Tensorflow and Keras.
- **Probabilistic Programming Languages:** Proficient in Stan, familiar with Edward.

○ **Bioinformatics and Neuroimaging Software**

- **Neuroimaging data analysis:** Nipype, PyMVPA, FreeSurfer, FSL.
- **Next Generation Sequencing (NGS) data analysis:** R's Bioconductor core packages, Bowtie2, MACS2, and bedtools.

○ **Cluster and High-Performance Computing**
- Good knowledge of running neuroimaging and bioinformatics analysis pipelines on super-computing clusters.

○ **Web and Software Development Tools**

- LATEX, Linux OS, Git, Docker, and Bash. Basic knowledge in website development tools such as HTML, CSS, and Jekyll.

Affiliations

○ Broad Institute of MIT and Harvard

Personal Details

○ **Country of Current Residence:** USA
○ **Country of Previous Residence:** Russia (12 years)
○ **Marital Status:** Married
○ **Interests and Hobbies:** Theoretical Linguistics, Evolutionary Biology, Experimental Music
○ **Languages Spoken:** English, Arabic, Russian, Spanish (intermediate), French (elementary)