

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 in the lab of Dr. Pinello at the Massachusetts General Hospital/Harvard Medical School. My research projects at the lab involve 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
 - **Haystack Pipeline:** 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.
 - **Histone Code:** Deep generative modeling of chromatin signal data across multiple cell types and histone marks with the goal of learning a latent representation of chromatin state dynamics (i.e. a continuous histone code). The modeling efforts involve applying a Variational Autoencoder (VAE) approach and its extensions (e.g. recurrent VAE) to sequential data.
 - **Statistical Modeling of CRISPR Double Knockout Data:** A hierarchical negative binomial Bayesian framework for modeling CRISPR-Cas9 Double-Knockout screen count data.

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.

Independent Tutor.....

- **Test Preparation Instructor** **Moscow, Russia**
Instructor of the Graduate Management Admission Test (GMAT) **2009–2012**
- **Teacher of English as a Foreign Language** **Moscow, Russia**
Teacher of General and Academic English **2001–2009**

Education

Academic Qualifications.....

- **PhD in Quantitative Psychology** **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 **2012–2014**
Course Projects
 - *Retinotopic Mapping of the Human Visual Cortex Using Independent Component Analysis* (STAT 7560: Multivariate Statistics)
 - *Bayesian Analysis of Noisy Images Using Markov Random Fields* (STAT 7730: Advanced Computational Statistics)
 - *Modeling Categorical Perception of Speech Sounds using Beta Regression* (STAT 6750: Statistical Consulting)
- **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**

Conference Presentations

- 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

- Graduate Student Conference Presentation Award 2015
 - The Ohio State University
- The Center for Cognitive and Brain Sciences Summer Graduate Fellowship 2015
 - The Ohio State University

Project Proposal: 'Decoding the Pixels of the Face Image from the Voxels of fMRI BOLD Activity Patterns'
- The Social and Behavioral Sciences Summer Fellowship 2014
 - The Ohio State University
- University Fellowship 2012
 - The Ohio State University

Publications and Software

Journal Papers.....

- Baskin, E., Farouni, R., and Mathè, E. (2016). ALTRE: workflow for defining ALTerred 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. (accepted pending revision, first co-author)

Preprints.....

- 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.....

- 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

- 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.....

- Histone Code VAE: Deep generative modeling of chromatin signal data across multiple cell types. GitHub Repo (private): https://github.com/pinellolab/histone_code_vae.
- Haystack: systematic analysis of the variation of epigenetic states and cell-type specific regulatory elements. GitHub Repo: https://github.com/pinellolab/haystack_bio.
- ALTRE: A Workflow for Identifying ALTered Regulatory Elements using Chromatin Accessibility Data. GitHub Repo: <https://github.com/Mathelab/ALTRE>.

Journal Review Service.....

- **Psychometrika**
Ad Hoc Reviewer 2015
- **Psychological Methods**
Ad Hoc Reviewer 2016

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
- International Society for Computational Biology

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)