

Rick Farouni | Curriculum Vitae

ETH Zurich - Department of Biosystems Science and Engineering (D-BSSE)
Basel – Switzerland

📞 +41 79 723 9098 • ✉ rfarouni@gmail.com
🌐 <http://rfarouni.github.io/>

I am an applied statistician and data scientist with expertise in Bayesian statistics, generalized linear mixed models, machine and deep learning, and the bioinformatics analysis of genomics, AAV evolution, and CRISPR editing data.

Experience

Bioinformatics Scientist.....

- **The Swiss Federal Institute of Technology in Zurich (ETH Zurich)** **Basel, Switzerland**
Research Lab of Dr. Platt, Department of Biosystems Science and Engineering 10/2020-Present
 - Statistical consulting, advising, and collaboration with graduate and post-doctoral students.
 - Analysis of transcriptional recording data.
 - Analysis of single nucleus RNA-seq in vivo CRISPR perturbation screens.
 - Development of a sequencing databases mining pipeline to discover uncharacterized Cas1-RT fusion proteins.
 - Design and analysis of base and prime editing CRISPR screens.
 - Statistical modeling of in vivo and invitro CRISPRi screens.
 - Development of target sequence extraction pipeline that matches the sequencing read into a given template and partitions it into specified parts. https://github.com/plattlab/quantify_target_sequence (private repo currently)
 - Analysis and modeling of AAV directed evolution experimental data for gene therapy applications.
 - Analysis of Deep Mutational Scans (DMS) screen data.

Postdoctoral Researcher.....

- **McGill University-Génomique Québec Innovation Centre** **Quebec, Canada**
Research Lab of Dr. Najafabadi, Department of Human Genetics 06/2018-09/2020

Published two first author/coauthor papers and one middle author paper.

 - **PhantomPurgeR**: A probabilistic model for the estimation, and remediation of sample index hopping in multiplexed droplet-based single-cell RNA-seq data.
 - **Transcriptional reprogramming of muscle cells**: A collaboration in which I was responsible for the computational analysis of the data generated by bulk RNA-seq and the single-cell RNA-seq assays.

- Massachusetts General Hospital-Harvard Medical School**

Research Lab of Dr. Pinello, Molecular Pathology Unit

-The Broad Institute of MIT and Harvard (Affiliated Member)

Massachusetts, USA

06/2017-05/2018

Published two first author/coauthor papers and three middle author papers.

 - **AmpUMI**: Mathematical derivation of a closed-form analytical solution for determining the collision probabilities of biological molecules in deep amplicon sequencing.
 - **haystack_bio**: A python pipeline to determine epigenetic variability, cross-cell-type plasticity of chromatin states, and transcription factors (TFs) motifs aimed at providing mechanistic insights into chromatin structure, cellular identity and gene regulation.
 - **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-PtpEeiB3w15E3J5ELNqck/edit#slide=id.p3

Research Intern.....

- Department of Biomedical Informatics Summer Internship Program**

Research Lab of Dr. Ewy Mathè, The Ohio State University

Ohio, USA

05/2016-08/2016

Published one first author/coauthor paper.

 - **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**

Graduate Teaching Associate (Statistics)

Ohio, USA

2013-2017

Served as a Teaching Assistant for three courses:

 - Repeated Measures Models
 - Covariance Structure Models
 - Data Analysis in Psychology

Education

Academic Qualifications.....

- PhD in Quantitative Psychology (i.e. Applied Statistics)**

The Ohio State University

Ohio, USA

2015-2017

Dissertation Topic: 'Application of Deep Latent Generative Models to the Unsupervised Learning of Chromatin States'
- Master of Science in Mathematical Statistics**

The Ohio State University

Ohio, USA

2012-2015
- Master's Degree in Psychometrics**

The Ohio State University

Ohio, USA

2012-2014

Publications

Journal Papers (first author/co-author).....

- Lazure, F.[†], **Farouni, R.[†]**, Sahinyan, K., Blackburn, D.M., Hernández-Corchado, A., Perron, G., Lu, T., Osakwe, A., Ragoussis, J., Crist, C. and Perkins, T.J., Jahani-Asl, A., Najafabadi, H.S., & Soleimani, V.D. (2023). Transcriptional reprogramming of skeletal muscle stem cells by the niche environment. *Nature Communications*, 14(1), 535. Paper available at <https://www.nature.com/articles/s41467-023-36265-x>. Website: https://csglab.github.io/transcriptional_reprogramming_muscle_cells/.
- **Farouni, R.**, Djambazian, H., Ferri, L. E., Ragoussis, J., & Najafabadi, H. S. (2020). Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. *Nature Communications*. Paper available at <https://www.nature.com/articles/s41467-020-16522-z>. Website: <https://csglab.github.io/PhantomPurgeR/>.
- Clement, K.[†], **Farouni, R.[†]**, Bauer, D. E., & Pinello, L. (2018). Design and analysis of unique molecular identifiers for deep amplicon sequencing. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty264>. Preprint available at <https://www.biorxiv.org/content/biorxiv/early/2018/03/23/288118.full.pdf>
- Pinello, L.[†], **Farouni, R.[†]**, & 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>
- Baskin, E.[†], **Farouni, R.[†]**, & Mathè, E. (2017). ALTRE: workflow for defining ALTERed Regulatory Elements using chromatin accessibility data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btx386>. Preprint available at <http://www.biorxiv.org/content/early/2016/10/14/080564.full.pdf+html>

Journal Papers (middle author).....

- Belli, O., Karava, K., **Farouni, R.**, & Platt, R. J. (2024). Multimodal mutational scanning with base and prime editing. *Under review*.
- Santinha, A. J., Klingler, E., Kuhn, M., **Farouni, R.**, Lagler, S., Kalamakis, G., Lischetti, U., Jabaudon, D., & Platt, R. J. (2023). Transcriptional linkage analysis with in vivo AAV-Perturb-seq. *Nature*, 622(7982), 367-375. <https://www.nature.com/articles/s41586-023-06570-y>
- Schmidt, F., Zimmermann, J., Tanna, T., **Farouni, R.**, Conway, T., Macpherson, A.J., & Platt, R.J. (2022). Noninvasive assessment of gut function using transcriptional recording sentinel cells. *Science*, 376(6594), p.eabm6038. <https://www.science.org/doi/abs/10.1126/science>.

abm6038

- Anderson-Trocme, L., **Farouni, R.**, Bourgey, M., Kamatani, Y., Higasa, K., Seo, J., Kim, C., Matsuda, F., & Gravel, S. (2019). Legacy Data Confounds Genomics Studies. ***Molecular Biology and Evolution***. <https://doi.org/10.1093/molbev/msz201>
- Seruggia, D., Oti, M., Tripathi, P., Canver, M. C., LeBlanc, L., Di Giammartino, D. C., Bullen, M.J., Nefzger C.M., Sun, Y.B.Y., **Farouni, R.**, Polo, J.M., Pinello, L., Apostolou, E., Kim J., Orkin, S.H., & Das, P.P. (2019). TAF5L and TAF6L Maintain Self-Renewal of Embryonic Stem Cells via the MYC Regulatory Network. ***Molecular Cell***.
- Clement K., Rees H., Canver, M.C., Gehrke J. M., **Farouni, R.**, Hsu, J.Y., Cole, M., Liu D.R., Joung J.K., Bauer, D.E., & Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. ***Nature Biotechnology***.
- Hsu, J. Y., Fulco, C.P., Cole, M., Canver, M.C., Pellin D., Sher, F, **Farouni, R.**, Clement K., Biasco L., Engreitz, J.M., Lander, E.S., Joung J.K., Bauer, D.E., & Pinello, L. (2018). CRISPR-SURF: Exploratory and interactive software for analyzing CRISPR-based tiling screens. ***Nature Methods***.

Consortium Journal Papers (contributor).....

- Grüning, B., Dale, R., Sjödin, A., Rowe, J., Chapman, B. A., Tomkins-Tinch, C. H., The Bioconda Team, & Köster, J.(2018). Bioconda: A sustainable and comprehensive software distribution for the life sciences. ***Nature Methods***. (Bioconda team member contributor, ranked 94 in the consortium author list). .

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

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>).

Software Development

- PhantomPurgeR: The optimal purging of phantom molecules by the robust estimation of the sample index hopping rate in multiplexed droplet-based single-cell RNA-seq data. GitHub Repo: <https://csglab.github.io/PhantomPurgeR/>.
- 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.
- ALTRE: A Workflow for Identifying ALTered Regulatory Elements using Chromatin Accessibility Data. GitHub Repo: <https://github.com/Mathelab/ALTRE>.

Awards and Fellowships

- **Graduate Student Conference Presentation Award**
The Ohio State University 2015
- **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* and *Python* depending on computing goals. Familiar with *Julia*.
 - **Deep Learning Frameworks:** Experienced in using Pytorch, Tensorflow, and Keras.
 - **Probabilistic Programming Languages:** Proficient in Stan, familiar with Edward.
- **Bioinformatics:**
 - **Analysis of sequencing data:** scRNAseq, RNAseq, ChIPseq, ATACseq, Deep Mutational Scans (DMS), CRISPR screens, Alternative Splicing, Amplicon Sequencing, PacBio sequencing.
 - **Workflow development:** Development of custom workflows using R's Bioconductor core package, large database protein mining pipelines. Good knowledge of running bioinformatics analysis pipelines on super-computing clusters.
 - **Established tools:** STAR, Bowtie2, bedtools, samtools, PEAR, Prokka, etc.
- **Cloud, Web, and Software Development Tools:** \LaTeX , Linux OS, Git, Docker, and Bash. Basic knowledge in website development tools such as HTML, CSS, and Jekyll. Familiar with cloud computing platforms such as GCP.

Personal Details

- **Country of Current Residence:** Switzerland
- **Countries of Previous Residence:** Canada, USA, Russia, Jordan
- **Languages Spoken:** English, Arabic, Russian, French (B1), Spanish (B1), German (A1)