

Alireza Karbalayghareh

CONTACT INFORMATION	Mortimer B. Zuckerman Research Center, 417 E 68th St., New York, NY 10065	<i>E-mail:</i> karbalaa@mskcc.org <i>Webpage:</i> karbalayghareh.github.io
RESEARCH INTERESTS	Computational biology, regulatory genomics, causal gene regulation, single-cell multiomics	
CURRENT POSITION	Memorial Sloan Kettering Cancer Center , New York, NY Computational & Systems Biology Program Postdoctoral Research Associate Supervisor: Dr. Christina Leslie	Sept. 2019 - present
EDUCATION	Texas A&M University , College Station, TX Ph.D. in Electrical Engineering, GPA: 4.0/4.0 Lab: Genomic Signal Processing (GSP) Advisor: Dr. Edward Dougherty Dissertation title: Optimal Bayesian Transfer Learning for Classification and Regression	Sept. 2014 - Aug. 2019
	Sharif University of Technology , Tehran, Iran M.Sc. in Electrical Engineering, GPA: 17.84/20.00	Sept. 2011 - Aug. 2014
	Iran University of Science and Technology , Tehran, Iran B.Sc. in Electrical Engineering, with honors, GPA: 18.08/20.00	Sept. 2007 - Aug. 2011
RESEARCH PROJECTS	Learning single-cell multiomic trajectories in germinal center B cells <ul style="list-style-type: none">• Developed an algorithm that jointly learns the dynamics and latent time of cells in both transcriptomic (RNA) and epigenomic (ATAC) spaces and captures cyclic and branching dynamics of germinal center B cells.• Predicted RNA velocities and recovered the missing ones.• Learned the transcription factor (TF) motif velocities showing the rate of change in motif accessibility across cells.• Learned regulation patterns within and across RNA expression and motif accessibility.• Predicted in-silico gene perturbation effects on RNA expression and motif accessibility. ENCODE-E2G: genome-wide prediction of enhancer-gene (E-G) interactions in various human cell types <ul style="list-style-type: none">• Developed a supervised method of E-G prediction that uses important epigenomic data, distance, 3D contacts, and other relevant features from unsupervised methods, and is trained on combined CRISPRi datasets of K562 cells.• Ablated the model and concluded which feature categories have the most significant impact on E-G predictions.• Predicted the genomewide E-G predictions across 352 ENCODE cell lines to be used for downstream eQTL and GWAS analyses. GraphReg: learning gene regulation mechanisms by integrating 3D chromatin interaction and 1D genomic/epigenomic data <ul style="list-style-type: none">• Developed GraphReg, a gene regulation model that uses graph attention networks to integrate chromatin interaction data (Hi-C/HiChIP/Micro-C) data and 1D genomic/epigenomic data to predict gene expression.• Identified functional enhancers of genes by interpreting GraphReg models using feature attribution methods.	

- Learned the effects of distal TFs on gene regulation by motif ablation and validated by CRISPRi TF knockout experiments.

scGraphReg: modeling gene regulation in single cells using Single-cell Multiomic and chromatin interaction data

- Developed scGraphReg to predict scRNA from scATAC, DNA sequence, and 3D interaction data and to uncover regulatory programs in single cells.

Epiphany: predicting Hi-C from epigenomic tracks

- Helped with designing Epiphany, an algorithm that uses widely available epigenomic tracks to predict Hi-C maps using CNN, Bi-LSTM, and adversarial loss functions.

scDIVA: domain invariant cell type annotation of immune cells using scRNA-seq data

- Developed an algorithm for semi-supervised annotation of tumor-infiltrating immune cell types from heterogeneous scRNA-seq data of different patients/cancers by correcting the batch effects using domain-invariant VAEs.

Optimal Bayesian transfer learning

- Proposed a Bayesian transfer learning framework, where the source and target domains are related through the joint prior density of the model parameters.
- Derived the posteriors and posterior predictive densities in closed forms in terms of generalized hypergeometric functions of matrix argument.
- Proposed the Optimal Bayesian Transfer Learning (OBTL) classifier, resulting in much better performance than the rival transfer learning and domain adaptation methods.
- Proposed Optimal Bayesian Transfer Regression (OBTR) operator, transferring optimally from the source domain to regression in the target domain.

Bayesian multi-domain learning

- Generalized Bayesian transfer learning to a multi-domain scenario with count observations using negative binomial factor analysis.
- Provided a Gibbs sampling inference using state-of-the-art data augmentation techniques.
- Proposed the Bayesian Multi-Domain Learning (BMDL) algorithm, outperforming the rival methods.

Trajectory inference in gene regulatory networks

- Proposed the optimal classifier for the gene expression trajectories using the partial knowledge of gene regulatory networks.
- Proposed an Expectation-Maximization (EM) method for the inference of partially observable Boolean dynamical systems.
- Proposed a Bayesian classifier for the gene expression trajectories.

MANUSCRIPTS IN
PREPARATION

- A.R. Gschwind*, K.S. Mualim*, **A. Karbalayghareh***, M.U. Sheth*, K.K. Dey*, E. Jagoda*, R.N. Nurtdinov*, W. Xi*, ..., C.S. Leslie, L.M. Steinmetz, A. Kundaje, J.M. Engreitz, “ENCODE4: Regulatory enhancer-gene interactions in the human genome,” in preparation.
- **A. Karbalayghareh**, C. Chin, S.Z. Josefowicz, A.M. Melnick, and C.S. Leslie, “Learning multi-omic trajectories of germinal center B cells,” in preparation.
- **A. Karbalayghareh**, J. Li, D. Koyyalagunta, and C.S. Leslie, “scGraphReg: gene regulation model for single cells using multiomics and chromatin interactions data,” in preparation.
- **A. Karbalayghareh**, V. Rapolu, B. Lee, and C.S. Leslie, “Semi-supervised and batch-correcting annotation of immune cell types in tumor single-cell RNA-seq data with domain invariant variational autoencoders,” in preparation.

SELECTED
PUBLICATIONS

- **A. Karbalayghareh**, M. Sahin, and C.S. Leslie, “Chromatin interaction aware gene regulation modeling with graph attention networks,” *Genome Research*, doi: 10.1101/gr.275870.121, 2022.
- R. Yang, A. Das, V. Gao, **A. Karbalayghareh**, W.S. Noble, J. Bilmes, C.S. Leslie. “Epiphany: predicting Hi-C contact maps from 1D epigenomic signals,” *Genome Biology*, doi: 10.1186/s13059-023-02934-9, 2023.
- **A. Karbalayghareh**, X. Qian, and E.R. Dougherty, “Optimal Bayesian transfer learning for count data,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2019.
- E. Hajiramezani, S. Zamani, **A. Karbalayghareh**, M. Zhou, X. Qian, “Bayesian multi-domain learning for cancer subtype discovery,” *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, Montreal, CA, 2018.
- **A. Karbalayghareh**, X. Qian, and E.R. Dougherty, “Optimal Bayesian transfer regression,” *IEEE Signal Processing Letter*, 2018.
- **A. Karbalayghareh**, X. Qian, and E.R. Dougherty, “Optimal Bayesian transfer learning,” *IEEE Transactions on Signal Processing*, 2018.
- **A. Karbalayghareh**, U. Braga-Neto, and E.R. Dougherty, “Intrinsically Bayesian robust classifier for single-cell gene expression trajectories in gene regulatory networks,” *BMC Systems Biology*, 2018.
- **A. Karbalayghareh**, U. Braga-Neto, E.R. Dougherty, “Classification of single-cell gene expression trajectories from incomplete and noisy Data,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2017.
- **A. Karbalayghareh**, U. Braga-Neto, and E.R. Dougherty, “Classification of Gaussian trajectories with missing data in Boolean gene regulatory networks,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, LA, 2017.
- **A. Karbalayghareh**, U. Braga-Neto, J. Hua, and E.R. Dougherty, “Classification of state trajectories in gene regulatory networks,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2016.
- **A. Karbalayghareh** and T. Hu, “Inference of sparse gene regulatory network from RNA-seq time series data,” *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, 2015.

TALKS,
CONFERENCES

- May 2023, Biology of Genomes - Cold Spring Harbor Laboratory, talk.
- November 2022, Guest Lecture in *Introduction in Computational Biology* invited by Dr. Michael Love, Biostatistics and Genetics Department, UNC-Chapel Hill, virtual.
- November 2022, Machine Learning in Computational Biology, virtual.
- May 2022, AI&A Journal Club, Astrazeneca, invited by Dr. Ehsan Ramezani, talk, virtual.
- April 2022, 4D Nucleome Scientific Webinar, talk, virtual.
- December 2021, 4D Nucleome Consortium Meeting, poster, virtual.
- November 2021, Genome Informatics - Cold Spring Harbor Laboratory, talk, virtual.
- November 2021, Machine Learning in Computational Biology, poster, virtual.
- May 2021, Biology of Genomes - Cold Spring Harbor Laboratory, poster, virtual.
- January 2021, ENCODE Consortium Meeting, talk, virtual.
- November 2020, Machine Learning in Computational Biology, poster, virtual.
- December 2019, ENCODE Consortium Meeting, Pacific Grove, CA.
- August 2017, ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB), Boston, MA, talk.

- March 2017, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, poster.
- December 2015, IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, talk.

MENTORSHIP
EXPERIENCE

- Viraj Rapolu, Tri-I PhD student in Computational Biology and Medicine (CBM), 2021 - now
- Jiaxin Li, Tri-I PhD student in CBM, summer and fall 2022
- Divya Koyyalagunta, Tri-I PhD student in CBM, summer 2021
- Brennan Lee, Tri-I MSc student in CBM, 2020 - 2021
- Changlin Wan, PhD summer intern from Purdue University, summer 2021

KEY GRADUATE
COURSES

Probabilistic Graphical Models, Pattern Recognition, Bayesian Pattern Recognition, Stochastic Systems, Bayesian Learning, Large-Scale Biological Data Analysis, Real Analysis (Math), Statistical Bioinformatics, Optimal Signal Processing Under Uncertainty, Convex Optimization, Coding Theory, Estimation Theory, Stochastic Processes.

TECHNICAL SKILLS

Programming Languages: Python, R, MATLAB.

Deep Learning: Pytorch, TensorFlow, Keras, Variational Autoencoders (VAE), Generative Adversarial Networks (GAN), Probabilistic Generative Models (PGM), Normalizing Flows, Graph Neural Networks (GNN), Domain Adaptation, Neural Ordinary Differential Equations (ODE), Neural Stochastic Differential Equations (SDE).

Computational Biology and Genomics: Experience with genomic (DNA sequence), epigenomic ([sc]ATAC, DNase, HistMod-ChIP, TF-ChIP), transcriptomic ([sc]RNA, CAGE), 3D chromatin conformation (Hi-C/HiChIP/Micro-C), and single cell multiomic (scRNA + scATAC) data.

REVIEW SERVICE

Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), International Conference on Learning Representations (ICLR), Genome Research, IEEE Transactions on Signal Processing (TSP), IEEE Signal Processing Letters (SPL), IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), Association for the Advancement of Artificial Intelligence (AAAI), Pattern Recognition Letter (PRL), BMC Bioinformatics, IEEE Access, International Joint Conference on Artificial Intelligence (IJCAI), Journal of Applied Statistics.

LANGUAGES

English (Fluent), Persian (Native-Bilingual), Turkish (Native-Bilingual)

REFERENCES

Dr. Christina Leslie,
Member
Computational & Systems Biology Program
Memorial Sloan Kettering Cancer Center
Email: cleslie@cbio.mskcc.org

Dr. Edward R. Dougherty
Distinguished Professor
Department of Electrical and Computer Engineering
Texas A&M University
Email: edward@ece.tamu.edu