### Alireza Karbalayghareh

CONTACT Mortimer B. Zuckerman Research Center, Phone: (979) 676-9777
INFORMATION 417 E 68th St., E-mail: karbalaa@mske

417 E 68th St., E-mail: karbalaa@mskcc.org
New York, NY 10065 Webpage: karbalayghareh.github.io

RESEARCH INTERESTS Machine Learning, Computational Biology, Genomics

CURRENT Memorial Sloan Kettering Cancer Center, New York, NY POSITION Computational & Systems Biology Program

Postdoctoral Research Scholar (PI: Dr. Christina Leslie) Sept. 2019 - present

EDUCATION Texas A&M University, College Station, TX

Ph.D. in Electrical Engineering (PI: Dr. Edward Dougherty)

Sept. 2014 - Aug. 2019

GPA: 4.0/4.0

Sharif University of Technology, Tehran, Iran

M.Sc. in Electrical Engineering Sept. 2011 - Aug. 2014

GPA: 17.84/20.00

Iran University of Science and Technology, Tehran, Iran

B.Sc. in Electrical Engineering Sept. 2007 - Aug. 2011

GPA: 18.08/20.00

RESEARCH PROJECTS

# GraphReg: Learning Gene Regulation Mechanisms by Integrating 3D DNA Interaction and 1D Genomic/Epigenomics Data

- Developed GraphReg, a gene regulation model with graph attention networks, which integrates the chromatin interaction data (Hi-C/HiChIP/Micro-C) data and 1D genomic/epigenomic data to predict the gene expressions.
- Identified functional enhancers of the genes by interpreting the models using feature attribution methods.
- Learned the effects of distal TFs on gene regulation by motif ablation and validated by CRISPRi TF knockout experiments.

#### Epiphany: Predicting Hi-C from Epigenomic Tracks

- Used five well-known epigenomic tracks to predcit Hi-C maps using CNN, Bi-LSTM, and adversarial loss functions.
- Interpreted the model to discover the effects of epigenomic features, especially CTCF, on 3D chromatin conformation.

#### scBatch: Domain Invariant Cell Type Annotation Using scRNA-seq Data:

• Annotated 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/Time Series Classification Using 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 to the inference of the partially observable Boolean dynamical systems.
- Proposed a Bayesian classifier for the gene expression trajectories.

## SELECTED PUBLICATIONS

- 1. **A. Karbalayghareh**, M. Sahin, and C. Leslie, "Chromatin interaction aware gene regulation modeling with graph attention networks," under review in Genome Research, bioRxiv doi: https://doi.org/10.1101/2021.03.31.437978, 2021.
- 2. R. Yang, A. Das, V. Gao, A. Karbalayghareh, W. Noble, J. Bilmes, C. Leslie. "Epiphany: predicting Hi-C contact maps from 1D epigenomic signals,", under review in Genome Biology, bioRxiv doi: https://doi.org/10.1101/2021.12.02.470663, 2021.
- 3. A. Karbalayghareh, B. Lee, V. Rapolu, and C. Leslie, "Semi-supervised annotation of immune cell types in heterogeneous single-cell RNA-seq datasets with domain invariant variational autoencoders," to be submitted, 2022.
- 4. 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 32<sup>nd</sup> 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.
- 7. A. Karbalayghareh, X. Qian, and E. R. Dougherty, "Optimal Bayesian transfer learning," IEEE Transactions on Signal Processing, 2018.
- 8. 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.
- 9. **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.
- 10. 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.
- 11. 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.
- 12. **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.

Research

Leslie Lab,

EXPERIENCE

Memorial Sloan Kettering Cancer Center

Postdoctoral Research Scholar

Mentor: Dr. Christina Leslie

Sept. 2019 - present

Genomic Signal Processing (GSP) Laboratory,

Texas A&M University

Research Assistant

Sept. 2014 - Aug. 2019

Advisor: Prof. Edward R. Dougherty

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: TensorFlow, Keras, Pytorch, 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), and 3D chromatin conformation (Hi-C/HiChIP/Micro-C), and single cell multiomic (scRNA + scATAC) data.

Journal and Conference REVIEWER FOR 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

Prof. Edward R. Dougherty, Distinguished Professor,

Department of Electrical and Computer Engineering,

Texas A&M University, Email: edward@ece.tamu.edu