

Habil Zare

Email: zare@u.washington.edu
Lab homepage: <http://oncinfo.org>
Phone: (+1) 206 612-7830
Canadian citizen, and US green card holder

Department of Cell Systems & Anatomy
University of Texas Health Science Center
San Antonio, TX 78229, USA

RESEARCH INTERESTS

Computational Biology, Cancer Research, Genetics, Epigenomics, Translational Bioinformatics, Data Mining, and Statistical Machine Learning.

CURRENT POSITION

Assistant Professor in Department of Computer Science, Texas State University, USA, Sept. 2014–current.

Responsibilities: Research, teaching, and academic service.

Principal Investigator, established Oncinfo Lab (<http://oncinfo.org>), which is focused on research in computational biology and translational bioinformatics, Sept. 2014–current.

Adjunct Assistant Professor in Department of Cell Systems & Anatomy, University of Texas Health Science Center in San Antonio, USA, May 2015–current.

EDUCATION

Ph.D. in Computer Science, University of British Columbia, Canada, January 2010¹–December 2011.

Dissertation: Automatic Analysis of Flow Cytometry Data and its Application in Lymphoma Diagnosis.

Using sophisticated machine learning techniques, developed novel bioinformatics tools to improve data analysis, and discovered a criterion useful for lymphoma diagnosis.

Supervisors: Arvind Gupta and Ryan Brinkman.

Committee members: Andrew Weng, Gabor Tardosh, and Valentine Kabanets.

M.Sc. in Computer Science, Sharif University of Technology, Iran, 2002–2004.

Dissertation: Markov Chain (Random Walk) on Graphs.

Performed a comprehensive study on different definitions for Laplacian of a graph and proposed a unifying viewpoint to spectral graph theory from various aspects of mathematics, such as algebra, geometry, stochastic processes, and mathematical analysis. Application to randomized algorithms were discussed.

B.Sc. in Mathematics, Sharif University of Technology, Iran, 1998–2002.

POSTDOCTORAL

Senior Fellow in Noble Laboratory, Department of Genome Sciences, The University of Washington, USA, January 2012–2014.

CURRENT RESEARCH

Department of Computer Sciences, Texas State University, September 2014–current.

Conceiving and leading collaborative studies in the areas of computational biology and translational bioinformatics including:

¹transferred from Simon Fraser University with my senior supervisor.

- Identifying molecular mechanisms of Alzheimer's disease based on high-throughput data of mRNAs, piRNAs, miRNAs, esiRNAs, and transposons in a transgenic *Drosophila* model.
Collaborator: Dr. Bess Frost from the University of Texas Health Science Center in San Antonio.
- Performing integrative network analysis using gene expression and DNA methylation data to identify prognostic biomarkers for acute myeloid leukemia (AML).
Collaborator: Prof. Aly Karsan from British Columbia Cancer Agency.
- Investigating response to treatment and prognosis of hepatocellular carcinoma using gene expression RNA-Seq data of a mouse model and also a database of Hispanic samples.
Collaborators: Prof. Christi Walter and Prof. LuZhe Sun from the University of Texas Health Science Center in San Antonio.
- Developing a bioinformatics pipeline for the analysis of data produced using unconventional targeted sequencing approaches including RNA-mediated oligonucleotide Annealing, Selection, and Ligation with next-generation Sequencing (RASL-Seq). Using this approach to identify the role of miRNAs in immune response in *Arabidopsis*.
Collaborator: Dr. HongGu Kang from Biology Department at Texas State University.

PREVIOUS RESEARCH

Department of Genome Sciences, School of Medicine, The University of Washington, January 2012–June 2014.

Developing and applying advanced machine learning techniques to analyze genomic and epigenomic data in Noble Laboratory including:

- Studied the relation between histone modification and gene expression; Modeled the epigenomic, ChIP-Seq data from ENCODE using a dynamic Bayesian network to predict de novo genes and their tissue-specific expression. Also, designed a deep learner to analyze biological sequential data.
- Inferred the clonal structure of a breast tumor using high-throughput exome sequencing; Designed and implemented Clomial, an algorithm that models next generation sequencing data from several samples of a single tumor by binomial distributions. The model can be used to identify the clonal structure of tumor, such as the number of clones, their genotypes, and the proportion of each clone in the tumor.

British Columbia Cancer Agency, Terry Fox Laboratory, April 2009–December 2011.

Developing tools and techniques for analyzing flow cytometry data in Brinkman Laboratory including:

- Designed and implemented a novel data reduction scheme to reduce size of flow cytometry data without significant loss in biological information. My approach resulted in enhancing spectral clustering technique such that it could be efficiently applied on flow cytometry data for the first time.
- Designed and implemented a novel feature scoring scheme based on the LASSO, useful for feature selection.
- Applied supervised machine learning techniques to build a classifier for differential diagnosis between subtypes of lymphoma in 860 patients.
- Applied my methodology to data of 110 patients diagnosed in BC Cancer Agency and discovered three novel flow cytometry features that discriminate between Mantel Cell Lymphoma (MCL) and Small Lymphocytic Lymphoma (SLL) with high accuracy. The novel features were shown to be clinically useful.

PUBLICATIONS

■ As of May 2018, the most cited scholar in the department, total citations: 632, H-index: 7.

Published Journal Articles

- **Zare H.**, Shooshtari, P., Gupta, A., Brinkman R.R., Data reduction for spectral clustering to analyze high throughput flow cytometry data. *BMC Bioinformatics* 2010, **11**:403–413. (“**Highly accessed**” according to the journal

website.) Impact factor: 2.7, Citations: 109.

Available at: <http://www.biomedcentral.com/1471-2105/11/40>

Contribution: Spectral clustering cannot be applied directly to large size data because it has time complexity $O(n^3)$. I designed and implemented the novel "faithful sampling" algorithm (95%), the methods for computing similarity matrix (95%), estimating the number of cluster (80%), and post-processing steps (70%). I also performed the experiments (90%) and wrote the manuscript (70%).

- **Zare H.**, Bashashati A., Kridel R., Aghaeepour N., Haffari G., Connors J., Gupta A., Gascoyne R., Brinkman R.R., Weng A., Automated analysis of multidimensional flow cytometry data improves diagnostic accuracy between mantle cell lymphoma and small lymphocytic lymphoma, *American Journal of Clinical Pathology* 2012 **137**:75–85. Impact factor: 3.0, Citations: 25.

Available at: <http://ajcp.ascpjournals.org/content/137/1/75.long>

Contribution: The novel features improve the accuracy of diagnosis from 70% to 98%. I designed the pipeline (80%), implemented the algorithms (95%), analyzed data (70%), and discovered novel phenotypes (95%).

- Bashashati A., Nathalie A. J., Hadjkhodabakhshi A., Whiteside M., **Zare H.**, Scott D.W., Lo K., Gottardo R., Brinkman F., Connors J., Slack G., Randy D.G., Weng A., Brinkman R.R., B-cells with high side scatter parameter by flow cytometry correlate with inferior survival in diffuse large B cell lymphoma, *American Journal of Clinical Pathology* 2012, **137**:805–814. Impact factor: 3.0, Citations: 10.

Available at: <http://ajcp.ascpjournals.org/content/137/5/805.long>

Contribution: I contributed in data analysis (5%).

- Aghaeepour N., Chattopadhyay P.K., Ganesan A., O'Neill K., **Zare H.**, Jalali A., Hoos H.H., Roederer M., Brinkman R.R., Early immunologic correlates of HIV protection can be identified from computational analysis of complex multivariate T-cell flow cytometry assays, *Bioinformatics* 2012, **28**:1009–1016. Impact factor: 4.6, Citations: 58.

Available at: <http://bioinformatics.oxfordjournals.org/content/28/7/1009.long>

Contribution: I contributed in data analysis (15%), preparing figures, presenting results and writing the manuscripts (30%), and biological interpretation (5%).

- **Zare H.**, Haffari G., Weng A., Gascoyne R., Gupta A., Brinkman R.R., Scoring relevancy of features based on combinatorial analysis of Lasso with application to lymphoma diagnosis, *BMC Genomics* 2013, **14**:S14. Impact factor: 4.0, Citations: 21.

Available at: <http://www.biomedcentral.com/1471-2164/14/S1/S14>

Contribution: designed a novel feature scoring scheme useful for feature selection (90%), implemented the algorithm (95%), developed a mathematical framework to explain the performance of the method (70%), theoretically proved its efficiency (80%), performed experiments (95%) and wrote the manuscript (70%).

- Aghaeepour N., Finak G., Hoos H., Mosmann T, Brinkman R., Gottardo et al., Scheuermann R., **Zare H.**, et al. Critical assessment of automated flow cytometry data analysis techniques, *Nature Methods* 2013, **10**:228–238. Impact factor: 26.0, Citations: 317.

Available at: <http://www.nature.com/nmeth/journal/v10/n3/full/nmeth.2365.html>

Contribution: This is a comparison between different computational approaches for analyzing flow cytometry data. I contributed in data analysis; running SamSPECTRAL on data (15%), preparing figures (5%), presenting results (5%) and writing the manuscripts (5%).

- **Zare H.**, Wang J., Hu A., Weber K., Smith J., Nickerson D., Witten D., Blau A.C., and Noble W.S., Inferring clonal composition from multiple sections of a breast cancer, *PLOS Computational Biology* 2014, **10**(7). Impact factor: 4.9, Citations: 64.

Available at: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003703>

Contribution: I proposed a scheme to model Next Gen Sequencing data of a tumor with binomial distributions, and developed an algorithm to infer the parameters using EM (80%), derived the formulas (90%), ran experiments (95%), prepared figures (60%), and wrote the manuscripts (40%).

- Zainulabadeen A., Yao P, **Zare H.**, Underexpression of specific interferon genes is associated with poor prognosis of melanoma, *PLOS ONE* 2017, **11**(1). Impact factor: 3.2, Citations: 2

Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170025>

Contribution: I designed the experiments (95%), administrated the project (95%), identified the appropriate datasets (90%), ran experiments (20%), prepared figures (30%), and wrote the manuscripts (70%).

- Foroushani A., Agrahari R., Docking R., Chang L., Duns G., Hudoba M., Karsan A., **Zare H.**, Large-scale gene

network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications *BMC Medical Genomics* 2017, **10**(16). Impact factor: 2.7.

Available at: <https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-017-0253-6>

Contribution: I designed the experiments (70%), administrated the project (80%), identified the appropriate approaches and developed the methodology (70%), ran experiments (10%), prepared figures (50%), and wrote the manuscripts (70%).

- Agrahari R., Foroushani A., Docking R., Chang L., Duns G., Hudoba M., Karsan A., **Zare H.**, Applications of Bayesian Network Models in Predicting Types of Hematological Malignancies *Scientific Reports* 2018, **8**(1). Impact factor: 4.3.
Contribution: I designed the experiments (90%), administrated the project (90%), identified the appropriate approaches and developed the methodology (90%), ran experiments (10%), prepared figures (30%), and wrote the manuscripts (80%).
- Sun W., Beckmann A., Samimi H., **Zare H.**, Frost B., Transposable element activation in Alzheimer's disease and related neurodegenerative tauopathies *Nature Neuroscience* 2018, In Press. Impact factor: 18.
Contribution: I conceived and performed the statistical analysis (80%), administrated the RNA-Seq data analysis (95%), identified the piRNA and miRNA targets (90%), and wrote the manuscripts (10%).

Articles Under Review

- Zavadil JZ., Herzig M., Hildreth K., Foroushani A., Boswell W., Walter R., Reddick R., White H., **Zare* H.**, Walter* C., A Murine Model (C3HeB/FeJ) of Hepatocellular Carcinoma Displays Similarities to Human Including Histologic, Radiologic, and Gene Expression Features *Molecular Carcinogenesis* 2017. Impact factor: 4.
Contribution: () My contribution was equal to the senior last author. I performed the statistical analysis (40%), administrated the RNA-Seq data analysis (80%), ran in silico experiments (30%), and wrote the manuscripts (20%).*
- Wu y., Polican A., Hildreth K., Yarbrough B., **Zare H.**, Ngu A., Real-Time Smartwatch based Fall Detection Accessor *International Conference for Smart Health 2017 (accepted)*. *Contribution: I commented on the appropriate machine learning approaches (5%), and wrote the manuscripts (20%).*

Journal Articles In Preparation

- Samimi H., Docking R., Zainulabadeen A., Karsan A., **Zare H.**, DNA methylation analysis improves the prognostication of AML *Blood* 2018. Impact factor: 13.
Contribution: I designed the experiments (90%), administrated the project (70%), identified the appropriate approaches and developed the methodology (80%), ran experiments (20%), prepared figures (30%), and wrote the manuscripts (70%).
- Samimi H., Docking R., Chang L., Duns G., Hudoba M., Karsan A., **Zare H.**, iNETgrate: How to leverage epigenome data when they are missing? *PLOS Computational Biology* 2017. Impact factor: 4.9.
Contribution: I designed the experiments (80%), administrated the project (90%), identified the appropriate approaches and developed the methodology (80%), ran experiments (10%), prepared figures (10%), and wrote the manuscripts (60%).
- Bordiya Y., Hurtado G., Samimi Y., **Zare H.**, Kang H., Small RNAs regulate the transcriptional dynamics of defense genes under biotic stress in Arabidopsis *Plant Cell* 2018. Impact factor: 8.5.
Contribution: I administrated the statistical analysis (90%), and the RNA-Seq and RASL-Seq data analyses (80%), ran in silico experiments (50%), and wrote the manuscripts (20%).

Invited Talks

- "Prediction of melanoma progression using large-scale gene network analysis" University of Texas MD Anderson Cancer Center, Houston, TX, April 2017.
- "Advancements and applications of gene network analysis in biomarker discovery", The Center for Computational Visualization, The University of Texas in Austin, TX, February 2017.

- “Large-scale gene network analysis and its application in biomarker discovery: an introduction to the Pigengene package”, Dahshu: Data Science & Computational Precision Health, San Francisco, CA, February 2017.
- “Computational approaches for gene network analysis with applications in studying AML and other cancers”, The University of Texas Health Science Center at San Antonio, TX, USA, March 2015.
- “SamSPECTRAL: Efficient spectral clustering on flow cytometry data”, FlowCAP summit, American National Institutes of Health Campus (NIH), M.D., USA, September 2010.
- “Automatic analysis of flow cytometry data”, British Columbia CLL Day, BC, Canada, February 2010.

Presentations

- “Inferring clonal composition from multiple sections of a breast cancer”, Seattle Genetic Instability and Cancer Symposium, Fred Hutchinson Cancer Research Center, Seattle, April 2014.
- “How to infer clonal composition of a tumor from multiple tissue samples?” Post-Doctoral Presentation, University of Washington, April 2014.
- “Scoring relevancy of features based on combinatorial analysis of Lasso with application to lymphoma” Asia Pacific Bioinformatics Conference, Vancouver, January 2013.
- “Automated analysis of multidimensional flow cytometry data improves diagnostic accuracy between mantle cell lymphoma and small lymphocytic lymphoma”, ImmunoVancouver meeting, University of British Columbia, Vancouver, June 2011.
- “From graph theory to cancer diagnosis”, Bioinformatics Reading Group, University of British Columbia, Vancouver, June 2011.
- “Discrimination between Mantle Cell Lymphoma and Small Lymphocytic Lymphoma can be improved by automated analysis of multidimensional flow cytometry data”, Work in Progress (WIP) seminars, BC Cancer Research Center, Vancouver, February 2011.
- “Is it possible to apply spectral clustering method to analyze high throughput biological data?”, Honors Seminar, Computer Science Dept., University of British Columbia, October 2010.
- “Lymphoma Diagnosis Based on Automated Analysis of Flow Cytometry Data”, Cancer Bioinformatics Workshop, Cambridge Cancer Research, September 2010.
- “Automatic analysis of flow cytometry data for lymphoma diagnosis”, BC Cancer Research Center, Trainee Seminars, May 2010.
- “Automated identification of cell populations in flow cytometry data”, Work in Progress (WIP) seminars, BC Cancer Research Center, Vancouver, November 2009.

PATENTS

- Blau A. C., Noble W., Wang J., Witten D., **Zare H.**, “Inferring clonal composition from multiple sections of a cancer”, US Provisional Patent Application 61/955,118, March 18, 2014.

PEER REVIEWING

- National Science Foundation.
- ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB'17), Program Committee Member.
- Bioinformatics.
- Intelligent Systems for Molecular Biology (ISMB).
- Neural Information Processing Systems (NIPS).
- PLOS ONE.

- Nature Partner Journals (NPJ) Genomic Medicine.
- BMC Genomics.
- IEEE/ACM Transactions on Computational Biology and Bioinformatics.
- Journal of Bioinformatics and Computational Biology.
- Cancer Chemotherapy and Pharmacology
- International Conference on Intelligent Biology and Medicine (ICIBM 2018), Program Committee Member.
- Bioinformatics and Biology Insights.
- PeerJ.
- Cancer Informatics.
- Cytometry: Part A.
- Therapeutics and Clinical Risk Management
- Pacific Symposium on Biocomputing.
- International Conference on Genome Informatics.
- Uncertainty in Artificial Intelligence (UAI) conference.
- OMICS: A Journal of Integrative Biology published by Mary Ann Liebert, Inc.
- Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

SCIENTIFIC SOFTWARE

- SamSPECTRAL (approved by Bioconductor)
Developed and maintain this R package, an implementation of my enhancement to spectral clustering method that is capable of clustering flow cytometry data with 100K events in minutes.
- FeaLect (approved by CRAN)
Developed and maintain this R package that implements my novel scoring scheme useful for feature selection.
- Clomial (approved by Bioconductor)
Developed and maintain this R package that fits binomial distributions to counts obtained from Next-Generation Sequencing data of multiple samples of the same tumor. The trained parameters can be interpreted to infer the clonal structure of the tumor.
- Pigengene (approved by Bioconductor)
Maintain this R package that was developed in the Oncinfo Lab. Infers biological signatures (eigengenes) from gene expression profiles and uses them to train a classifier.

Programming Skills and Bioinformatics Tools

Research experience in analyzing omics data including: RNA-Seq, exome sequencing, DNA methylation, microarray, ChIP-Seq, RASL-Seq, flow cytometry, etc.

Expert: R Statistical Programming Language, Maple, Mathematica, Python, Prolog, ~~TeX~~TeX, UCSS Genome Browser, FastQC, BWA, SAMtools, IGV, Salmon, etc.

Familiar: Matlab, C, Java, Perl, HTML, GTAK, bedtools, SVN, git, MEME, etc.

TEACHING

The University of Texas Health Science Center, Department of Cell Systems & Anatomy

Bioinformatics for biologist workshops, November 2015 and March 2016.

Texas State University, Department of Computer Science

Instructor, Fall 2014–current.

Formal Languages and Foundations of Computer Science II, each twice per year.

Simon Fraser University, School of Computer Science

Sessional Instructor, Discrete Mathematics 1, Spring 2008.

Mentoring

- Supervising Hanie Samimi, a graduate student in the Oncinfo Lab, 2017–present.
- Supervising Bryan Shaw, a graduate student in the Oncinfo Lab, 2016–present.
- Supervising Gabriel Hurtado, an undergraduate student in the Oncinfo Lab, 2017–present.
- Supervising three high school students in the Summer Math Camp, 2017.
- Supervising two students per year in the NSF-supported Research Experiences for Undergraduates Program
Summer 2016: led to a publication in PLOS ONE. Summer 2017: in progress.
- Mentoring Amir Foroushani, postdoc in the Oncinfo Lab, 2014–2016. Joined the Institute of Allergy and Infectious Diseases at NIH as a Computer Scientist.
- Supervising Rupesh Agrahari, a graduate student in the Oncinfo Lab, 2014–2016. Joined the Becton Dickinson's Technical Support team.
- Mentoring Abirama Ganesan, a graduate biology student in an independent research course, 2015.
- Mentoring a rotating student in the Noble Lab at University of Washington, leading to a publication in PLOS Computational Biology, 2013–2014.

PROFESSIONAL SERVICE

- Member of Graduate Committee, Computer Science Department, Texas State University, 2014–current.
- Coordinator of Sarenakh Mathematical Seminar Series, Sharif University of Technology, 2002–2003.
- Secretary of Student Mathematical Association at Sharif University of Technology, 2002–2003.

HONORS and AWARDS

- **1st Rank** in Iranian National Entrance Exam for Master Program in Mathematics among 3000 students, 2002.
- Member of Graduate Faculty, Texas State University, 2014–current.
- Research Experience Program, Texas State University Internal Grant, \$8,000, 2016.
- National Science Foundation, Directorate for Computer and Information Science and Engineering Travel Award, \$800, 2015.
- International Society for Advances of Cytometry Travel Award, \$650, 2011 and 2012.
- University of British Columbia Faculty of Science Graduate Award, \$2,800, 2009, 2010, and 2011.
- University of British Columbia Travel Award, \$500, 2011.
- Simon Fraser University Graduate Fellowship Award, \$3,000, 2006, 2007, 2008.

- Institute for Research in Fundamental Sciences Fellowship Award, \$300, 2000-2002.

External Grant Funding

- NIH, NLM Career Development Award in Biomedical Informatics and Data Science (K01).
Title: The diagnostic and prognostic value of eigengenes as robust biological signatures.
Role: PI.
Budget: \$500K.
Impact Score: 36 (Payline: 33). Will resubmit in 2018.
- NIH, National Institute on Aging (NIA) Understanding Alzheimer's Disease in the Context of the Aging Brain (R01).
Title: Mechanisms of tau- and aging-induced neurological dysfunction: Focus on the nucleus.
Role: Co-Investigator.
Budget: \$840K.
Status: Under review.

REFERENCES

Dr. Aly Karsan, MD

Medical Director
Centre for Clinical Genomics
British Columbia Cancer Agency
Vancouver, BC, Canada.
Phone: (+1) 604.8776086
Email: akarsan@bcgsc.ca

Dr. Anthony Blau, MD

Professor of Medicine/Hematology
The University of Washington
850 Republican Street, Room N535
Seattle, WA, USA.
Phone: (+1) 206.685.6873
Email: tblau@u.washington.edu

Dr. Christi Walter

Professor and Chair
Department of Cell Systems & Anatomy
The University of Texas Health Science
Center, San Antonio, TX, USA
Phone: (+1) 210.567.3800
Email: Walter@uthscsa.edu

Dr. Andrew P. Weng, MD, PhD

Director, Hematopathology,
British Columbia Cancer Agency
Vancouver, BC, Canada.
Phone: (+1) 604.675.8136
Email: aweng@bccrc.ca

Dr. William Noble

(my postdoc mentor)
Professor of Genome Sciences
The University of Washington
Seattle, WA, USA.
Phone: (+1) 206.221.4973
Email: noble@gs.washington.edu

Dr. Ronald Walter

Professor and University Chair
in Cancer Research
Department of Chemistry & Biochemistry
Texas State University,
San Marcos, TX, USA.
Phone: (+1) 512.245.0357
Email: RWalter@txstate.edu

Last updated: May 31, 2018