

Soyeon Kim, Ph.D.

Contact Information

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

Research Interests

Education

Applications: High-throughput genomics, epigenetics, transcriptomics, multi-omics data analyses
Methods: Statistical machine learning; Multivariate analysis; Regression analysis; Deep learning

Rice University, Houston, Texas, USA

University of Texas MD Anderson Cancer Center, Houston, TX, USA (Joint Program)

Ph.D., Statistics, 2015

Advisor: J.Jack Lee, Ph.D. and Veerabhadran Baladandayuthapani, Ph.D.

Dissertation title: Prediction Oriented Marker Selection (PROMISE) for High Dimensional Regression with Application to Personalized Medicine

Rice University, Houston, TX

Master of Arts, Statistics, 2013

Yonsei University, Seoul, South Korea

Bachelor of Arts with Honors, Applied Statistics, 2008

Publications Zhenjiang Fan, Soyeon Kim, Yulong Bai, Brenda Diergaarde, Ste Oesterreich, Hyun Jung Park, "3-UTR shortening contributes to subtype-specific cancer growth by breaking stable ceRNA crosstalk of housekeeping genes." *Frontiers Bioengineering* In press [doi:10.1101/601526](https://doi.org/10.1101/601526)

Soyeon Kim, Erick Forno, Rong Zhang, Qi Yan, Nadia Boutaoui, Edna Acosta-Perez, Glorisa Canino, Wei Chen, Juan C. Celeda "Expression quantitative trait methylation analysis reveals methylomic associations with gene expression in childhood asthma" *bioRxiv* (2020) 2020.02.13.937391; [doi:10.1101/2020.02.13.937391](https://doi.org/10.1101/2020.02.13.937391)

Erick Forno[^], Rong Zhang[^], Yale Jiang, Soyeon Kim, Qi Yan, Zhao Ren, Yueh-Ying Han, Nadia Boutaoui, Franziska Rosser, Daniel E. Weeks, Edna Acosta-Perez, Angel Colao-Semidey, Mara Alvarez, Glorisa Canino, Wei Chen, Juan C. Celeda \Transcriptome-wide and differential expression network analyses of childhood asthma in nasal epithelium" The Journal of Allergy and Clinical Immunology In press [doi:10.1016/j.jaci.2020.02.005](https://doi.org/10.1016/j.jaci.2020.02.005)

Soyeon Kim, Hyun Jung Park, Xiangqin Cui, Degui Zhi \Collective effects of long-range DNA methylations predict gene expressions and estimate phenotypes in cancer" Scientific Reports 10, 3920 (2020) [doi:10.1038/s41598-020-60845-2](https://doi.org/10.1038/s41598-020-60845-2)

Soyeon Kim[^], Erick Forno[^], Qi Yan, Yale Jiang, Rong Zhang, Nadia Boutaoui, Edna Acosta-Perez, Glorisa Canino, Wei Chen, Juan C. Celeda \SNPs identified by GWAS affect asthma risk through DNA methylation and expression of cis-genes in airway epithelium" European Respiratory Journal (2019), 1902079; [doi:10.1038/10.1183/13993003.02079-2019](https://doi.org/10.1038/10.1183/13993003.02079-2019)

Soyeon Kim, Yulong Bai, Zhenjiang Fan, Brenda Diergaarde, George C. Tseng, Hyun Jung

Soyeon Kim

2

Park, \Alternative Polyadenylation Regulates Patient-specific Tumor Growth by Individualizing the MicroRNA Target Site Landscape." bioRxiv (2019) 601518; doi.org/10.1101/601518

Huan Zhong, Soyeon Kim, Degui Zhi, Xiangqin Cui, \Predicting Gene Expression Using DNA Methylation in Two Human Populations" PeerJ (2019): e6757. doi.org/10.7717/peerj.6757

Hyun Jung Park[^], Ping Ji[^], Soyeon Kim, Zheng Xia, Benjamin Rodriguez, Chioniso P. Masamha, Ann-Bin Shyu, Joel R. Neilson, Eric J. Wagner, Wei Li, \3'-UTR Shortening Represses Tumor Suppressors in trans by Disrupting ceRNA Crosstalk" in Nature Genetics volume 50, pages 783-789 (2018) .[doi:10.1038/s41588-018-0118-8](https://doi.org/10.1038/s41588-018-0118-8)

Hyun Jung Park, Soyeon Kim, Wei Li, \ Model-based analysis of

competing-endogenous pathways (MACPath) in human cancers" PLoS Computational Biology 14.3 (2018): e1006074. [doi:10.1371/journal.pcbi.1006074](https://doi.org/10.1371/journal.pcbi.1006074)

Soyeon Kim, Veerabhadran Baladandayuthapani, J. Jack Lee, "Prediction-Oriented Marker Selection (PROMISE) with Application to High-Dimensional Regression" Statistics in Biosciences 9.1 (2017): 217-245. [doi:10.1007/s12561-016-9169-5](https://doi.org/10.1007/s12561-016-9169-5)
(ASA Biopharmaceutical Section Student Paper Award)

Professional appointments

Division of Pulmonary Medicine, Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Postdoctoral Research Fellow

July 2018-

Mentor: Juan C. Celedon, MD, DrPH and Wei Chen, Ph.D.

Identified associations between genotype, methylation, and gene expression data in childhood asthma

Conducted mediation analyses to identify pathogenesis of asthma using multi-omics data

Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, USA

Postdoctoral Research Fellow
2016 - December 2017

December

Mentor: Degui Zhi, Ph.D.

Developed statistical machine learning models to predict gene expression using long-range regulatory elements using epigenetics data in cancer
Predicted 5 years survival of breast cancer patients using the predicted gene expression

Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, USA
Research Intern
January 2016 - June 2016

Built penalized regression methods to identify predictive markers that are associated with treatment response for personalized medicine.

Soyeon Kim

3

Department of Biostatistics, University of Texas MD
Anderson Cancer Center , Houston, Texas, USA
Graduate Research Assistant
2010 - December 2015

June

Developed a novel penalty selection method, Prediction Oriented Marker Selection (PROMISE), to identify predictive and prognostic biomarkers, and predict outcomes of treatments in high dimensional data for personalized medicine.

Integrated BATTLE (Biomarkers-Integrated Approaches of Targeted Therapy for Lung Cancer Elimination) trial data and patients' gene expression data for data analysis.

Incorporated KEGG pathway information to effectively identify predictive markers.

Molecular Neuro-Oncology Laboratory, Texas Children's Hospital ,
Houston, Texas, USA Graduate Research Assistant
May 2009 - May 2010

Built regression models to quantify heritability of Glioma using parents-offspring trio genotype data (CNV) in collaboration with medical doctors and biologists.

Cancer Biostatistics Branch, National Cancer Center , Ilsan,
Gyeonggi, South Korea Student Researcher
June 2007 - August 2007

Assessed risk factors of colon cancer using Cox proportional-hazards models.

Collaboration
Experience

Computational Cancer Epigenomics and Transcriptomics Laboratory,
Dan L. Duncan Cancer Center, Baylor College of Medicine ,
Houston, Texas, USA
2017

2014 -

Built penalized regression models to identify a mechanism of tumor growth in breast cancer using RNA sequencing data.
Helped to write an NIH grant proposal in a statistical method section.

University of Texas Southwestern Medical Center , Dallas, Texas, USA

2014

Identified metabolites that affect lupus status and predicted lupus status using the selected metabolites

Neuroscience, Baylor College of Medicine, Houston, Texas, USA

2013

Consulted statistical analysis of human cell line experiments to find out structure-activity relationship of human protein.

Awards and Grants

Presentations

and Posters

T32 Grant, NIH, 2018-

Honorable Mention, Student Paper Award Competition, ASA Biopharmaceutical Section,

2016 Korean Honor Scholarship, the Embassy of the Republic of Korea, 2011

Third place, SAS Data Mining Championship, SAS Korea,

2007 Honors student, Yonsei University, 2005-2007

Invited Seminars

"Big Data Analysis for Precision Medicine from Omics Data to Electronic Medical Records", Division of Medical Genetics, Department of Pediatrics, University of Pittsburgh School of Medicine, 2018
"Precision Medicine using Statistical Learning in High-Dimensional Data", the Department of Human Genetics, University of Pittsburgh, 2017

Soyeon Kim

4

Presentations

"SNPs Identified By GWAS Affect Asthma Risk Through DNA Methylation And Expression Of Cis-genes In Nasal (airway) Epithelium", The Annual Meeting of the American Thoracic Society (ATS), Philadelphia, PA, 2020 (Accepted)

"Expression quantitative trait methylation (eQTM) analyses in nasal epithelium tissue reveal methylomic associations with gene expression in childhood asthma", Pulmonary Medicine Research Conference, UPMC

Children's hospital of Pittsburgh, 2019
"Gene regulation through SNPs and methylation in childhood asthma",
Pulmonary Medicine Research Conference, UPMC Children's hospital of
Pittsburgh, 2019
"Prediction-Oriented Marker Selection (PROMISE) with Application to High-
Dimensional Regression", Center for Precision Health Seminar in
Precision Medicine, University of Texas Health Science Center
at Houston, 2017
"Prediction-Oriented Marker Selection (PROMISE) with Application to High-
Dimensional Regression", Joint Statistical Meeting, 2016
"Biomarker selection for personalized medicine via statistical learning method
s", Graduate Communications Seminar, Rice University, 2013
"Selecting biomarkers for personalized medicine in randomized
clinical trials via statistical learning methods", Joint Statistical
Meeting, 2012
"Lasso vs. Bayesian Additive Regression Trees (BART) for
identifying significant biomarkers", Mathematical Science Virge
Seminar, Rice University, 2011
Posters
"A Genome-wide eQTM study of childhood asthma in nasal
epithelium." The Annual Meeting of the American Thoracic Society
(ATS), Dallas, TX, 2019
"Prediction-Oriented Marker Selection (PROMISE) with Application to High-
Dimensional Regression", Precision Medicine Day, University of Texas
Health Science Center at Houston, 2017
"Randomized Phase II Clinical Trials", National Cancer Center, South
Korea, 2007

Professional Services

Mentoring

Teaching Experience

Skills

Paper review

Human Molecular Genetics

(2019) Biometrika (2018)

2017 IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences

Ph.D. Student

Rong Zhang, University of Pittsburgh Huan Zhong, Hong Kong Baptist

University Medical Student

Zhongli Xi, Tsinghua University

Teaching Assistant

Statistics 310: \Probability and Statistics"(Undergraduate Level), Rice University,

Spring 2009 Statistics 431: \Overview of Mathematical Statistics"(Masters Level), Rice

University, Fall 2008 Tutor

Trigonometry, Statistics (Community college student),

2014-2015 Allegra, Calculus, Trigonometry (High school

students), 2006-2008

Statistical Software: R, SAS, STATA, MATLAB, Minitab,

SPSS Computer Languages: Python, C/C++

System: high-performance cluster in PBS

Soyeon Kim

Memberships

American Statistical Association (2010 - current)

American Statistical Association Pittsburgh Chapter (2018-

current) American Thoracic Society (2018 - current)

The International Biometric Society (2014 - 2017)