Qi Yan

(205) 396-6942 | qy2253@cumc.columbia.edu
545 1st Ave. Apt 10R, New York, NY 10016
https://qiyanpitt.github.io/

Immigration Status

U.S. Permanent Resident

Work Experience

- *Assistant Professor (tenure-track)*, Obstetrics and Gynecology (OBGYN), Columbia University Irving Medical Center (CUIMC) 2020-present
  - Lead whole genome sequencing (WGS) analyses of one TOPMed project, nuMoM2b-Heart Health Study.
  - Lead genome- and phenome-wide association studies (GWAS and PheWAS) in pregnancy research.
  - Use polygenic risk score (PRS) coupled with clinical variables for predictions of preterm birth and preeclampsia.
  - Examine pQTLs during pregnancy using proteomics data obtained from the SomaLogic platform.
  - Develop statistical methods for Mendelian randomization (MR) analysis utilizing both common and rare variants.
  - Provide statistics service for pre-clinical and clinical trials design and analysis in the department of OBGYN.
  - Collaborate with *Natera, Inc* to predict the risk of adverse pregnancy outcomes using multi-omics data.
- *Research Assistant Professor*, Pediatrics, University of Pittsburgh (Pitt) 2018-2020
  - Led GWAS and post-GWAS analyses of childhood asthma using rich human genetic datasets (e.g., UK Biobank).
  - Led multi-omics data analyses (e.g., epigenome-wide association study [EWAS] and RNA-seq analysis).
  - Utilized machine learning/deep learning techniques to predict age-related macular degeneration (AMD) status using genetic markers and clinical images.
  - Developed statistical methods and open-source tools (e.g., GitHub and CRAN) for analysis of multi-omics data.
- *Research Instructor*, Pediatrics, University of Pittsburgh 2017-2018
  - Led GWAS of AMD, childhood asthma and Alzheimer's disease.
  - Developed statistical methods for testing effects of rare genetic variants.
  - Built RNA-seq pipeline.

Education

- **Department of Pediatrics, University of Pittsburgh**, Pittsburgh, PA
  Post-doctoral fellow (mentor: Wei Chen) 2014-2016
- **Department of Biostatistics, University of Alabama at Birmingham**, Birmingham, AL
  Ph.D. and Master of science 2009-2014
- **Department of Biomedical Engineering, University of Alabama at Birmingham**, Birmingham, AL
  Master of Science 2007-2009
- **Department of Biomedical Engineering, Beijing Institute of Technology**, Beijing, China
  Bachelor of Science 2003-2007

Technical Skills

- ***Proficient*** in UNIX/LINUX, Bash, R, R-shiny, Python, TensorFlow/Keras, cloud-computing (e.g., AWS)
- ***Extensive experience*** in analyses (e.g., GWAS, PheWAS, EWAS, TWAS, PRS, rare-variant analysis, eQTL/pQTL); tools (e.g., PLINK, REGENIE, GENESIS, STAAR, VCFtools/BCFtools); and datasets (e.g., UK Biobank, All of Us, TOPMed, GTEx)

Teaching Experience

- *Invited Lecturer*, OBGYN and Genetic Counseling Education (Clinical data analysis, RCT) 2022-2024
- *Invited Lecturer*, Statistical Genetics, *Human Genetics at Pitt* 2020
- *Invited Lecturer*, Foundations of Translational Bioinformatics, *Biomedical Informatics at Pitt* 2019
- *Invited Lecturer*, Applied mixed model analysis, *Biostatistics at Pitt* 2016
- *Invited Lecturer*, Introductory high-throughput genomic data analysis I: data mining and applications, *Biostatistics at Pitt* 2015,2016,2017

SELECTED PEER REVIEWED PUBLICATIONS

GWAS and EWAS:

- **Yan Q**, Blue NR, Truong B, Zhang Y, Guerrero RF, Liu N, Honigberg MC, Parry S, McNeil RB, Simhan HN, Chung J, Mercer BM, Grobman WA, Silver R, Greenland P, Saade GR, Reddy UM, Wapner RJ, Haas DM. Genetic associations with dynamic placental proteins identify causal biomarkers for hypertension in pregnancy. Under revision in *Am J Obstet Gynecol. MedRxiv 2023;* 10.1101/2023.05.25.23290460
- **Yan Q**, Guerrero RF, Khan RR, Surujnarine AA, Wapner RJ, Hahn MW, Raja A, SallebAouissi A, Grobman WA, Simhan H, Blue NR, Silver R, Chung JH, Reddy UM, Radivojac P, Pe'er I, Haas DM. Searching and visualizing genetic associations of pregnancy traits by using GnuMoM2b. *Genetics 2023;* 225.
  - *R Shiny app: https://gnumom2b.cumcobgyn.org/*
- **Yan Q**, Forno E, Cardenas A, Qi C, Han YY, et al. Exposure to violence, chronic stress, nasal DNA methylation, and atopic asthma in children. *Pediatr Pulmonol* 2021; 56: 1896-1905.
- **Yan Q**, Forno E, Yang G, Herrera-Luis E, Pino-Yanes M, et al. A genome-wide association study of asthma hospitalizations in adults. *J Allergy Clin Immunol* 2020. *Chosen by the Editors to be highlighted in the AAAAI website.*
- **Yan Q**, Forno E, Herrera-Luis E, Pino-Yanes M, Qi C, et al. A genome-wide association study of severe asthma exacerbations in Latino children and adolescents. *Eur Respir J* 2020.
- **Yan Q**, Nho K, Del-Aguila JL, Wang X, Risacher SL, et al. Genome-wide association study of brain amyloid deposition as measured by Pittsburgh Compound-B (PiB)-PET imaging. *Mol Psychiatry* 2018.
- **Yan Q**, Ding Y, Liu Y, Sun T, Fritsche LG, et al. Genome-wide analysis of disease progression in age-related macular degeneration. *Hum Mol Genet* 2018; 27: 929-940.
- **Yan Q**, Brehm J, Pino-Yanes M, Forno E, Lin J, et al. A meta-analysis of genome-wide association studies of asthma in Puerto Ricans. *Eur Respir J* 2017; 49. *With an accompanying Editorial.*

Prediction Using Genetic Markers and Clinical Images:

- **Yan Q**, Jiang Y, Huang H, Swaroop A, Chew EY, et al. Genome-Wide Association Studies-based Machine Learning for Prediction of Age-Related Macular Degeneration Risk. *Transl Vis Sci Technol*. 2021; 10: 29.
  - *R Shiny app: https://yanq.shinyapps.io/no_vs_amd_NN/*
- **Yan Q**†, Weeks DE, Xin H, Swaroop A, Chew EY, et al. Deep-learning-based Prediction of Late Age-Related Macular Degeneration Progression. *Nat Mach Intell* 2020; 2: 141-150.
  - *Website: http://52.90.194.108/calculator/*
  - *Python source code: https://github.com/QiYanPitt/AMDprogressCNN*

Statistical Methods in Genetics and Multi-omics:

- **Yan Q**†, Forno E, Celedón JC, Chen W, Weeks DE. Allele-specific method for testing the association between molecular quantitative traits and phenotype-genotype interaction. *Bioinformatics* 2021; doi: 10.1093/bioinformatics/btab554.
  - *Python source code: https://github.com/QiYanPitt/CHIT*
- **Yan Q**†, Forno E, Celedón JC, Chen W. A region-based method for causal mediation analysis of DNA methylation data. *Epigenetics* 2021; 17, 286-296.
  - *R package: https://cran.r-project.org/web/packages/MRmediation/index.html*
- **Yan Q**†, Liu N, Forno E, Canino G, Celedón JC, et al. An integrative association method for omics data based on a modified Fisher's method with application to childhood asthma. *PLoS Genet* 2019; 15: e1008142.
  - *R package: https://cran.r-project.org/web/packages/OmnibusFisher/index.html*
- **Yan Q**†, Fang Z, Chen W. KMgene: a unified R package for gene-based association analysis for complex traits. *Bioinformatics* 2018; 34: 2144-2146.

- *R package: https://cran.r-project.org/web/packages/KMgene/index.html*

- **Yan Q**, Chen R, Sutcliffe JS, Cook EH, Weeks DE, et al. The impact of genotype calling errors on family-based studies. *Sci Rep* 2016; 6: 28323.
- **Yan Q**, Weeks DE, Tiwari HK, Yi N, Zhang K, et al. Rare-Variant Kernel Machine Test for Longitudinal Data from Population and Family Samples. *Hum Hered* 2015; 80: 126-138.
- **Yan Q**, Weeks DE, Celedón JC, Tiwari HK, Li B, et al. Associating Multivariate Quantitative Phenotypes with Genetic Variants in Family Samples with a Novel Kernel Machine Regression Method. *Genetics* 2015; 201: 1329-1339. *Chosen by the GENETICS Editors as one of the December 2015 Highlights.*
- **Yan Q**, Tiwari HK, Yi N, Gao G, Zhang K, et al. A Sequence Kernel Association Test for Dichotomous Traits in Family Samples under a Generalized Linear Mixed Model. *Hum Hered* 2015; 79: 60-68.
- **Yan Q**, Tiwari HK, Yi N, Lin WY, Gao G, et al. Kernel-machine testing coupled with a rank-truncation method for genetic pathway analysis. *Genet Epidemiol* 2014; 38: 447-456. *2014 Best Paper Award - UAB from the Science Unbound Foundation.*

Book Chapters in Statistical Genetics:

- **Yan Q**, Ding Y, Weeks DE, Chen W. AMD Genetics: Methods and Analyses for Association, Progression, and Prediction. *Adv Exp Med Biol* 2021; 1256: 191-200.
- **Yan Q**. Novel Methods for Family-Based Genetic Studies (pages 135-144), *Genetic Epidemiology (Methods and Protocols). Springer*, 2018.

- In addition to above listed leading author papers, I published papers in *Science*, *Lancet*, and *Cell* journals as a biostatistician and/or bioinformatician.
- Complete list of published work in My Bibliography: https://www.ncbi.nlm.nih.gov/myncbi/qi.yan.1/bibliography/public/

Invited Presentations

1. Multi-omics data analysis in complex human diseases. Workshop of Statistical Methods in Genetic/Genomic Studies. Institute for Mathematical Sciences at National University of Singapore. Jan 14 2022
2. Statistical method and application in omics data analysis. MFM Research Meeting. CUIMC. Aug 20, 2021
3. An "incomplete" overview of omics data analysis. Seminar Series. UMD-BRIGHT at the University of Maryland. May 5, 2021.
4. A region-based method for causal mediation analysis of DNA methylation data. Seminar Series. Biostatistics at Pitt. October 22, 2020.
5. Convolutional neural network with application in AMD progression prediction. Seminar Series. Preventive Medicine at University of Tennessee Health Science Center. July 13, 2020.
6. Multi-omics data analysis in asthma in Puerto Rican children. Lecture Series. Biomedical Informatics at Pitt. March 9, 2019.

Contributed Presentations

1. Genetic associations with dynamic placental proteins identify biomarkers for hypertension in pregnancy. STATGEN. Platform talk. May 2, 2024.
2. The ongoing nuMoM2b-HHS WGS analysis. NHLBI TOPMed annual meeting. Platform talk. February 14, 2024.
3. GWAS and Mendelian randomization analysis for placental proteins in early pregnancy. ICSA Conference. Platform talk. June 12, 2023.
4. Placenta fraction of maternal blood cell-free RNA associates with pre-eclampsia during pregnancy. ASHG Conference. Poster. October 25-29, 2022.
5. Allele-specific method for testing the association between gene expression and phenotype-genotype interaction. ASHG Virtual Conference. Poster. October 27-30, 2020.
6. Deep-learning-based Prediction of Late Age-Related Macular Degeneration Progression. ASHG Conference. Poster. October 17, 2019.
7. KMgene: a unified R package for gene-based association analysis for complex traits. ASHG Conference. Poster. October 19, 2018.

8. Genome-wide analysis of age-related macular degeneration progression. ASHG Conference. Poster. October 19, 2017.
9. An Omnibus Test for Gene-Level Effects of Multi-Omics Data with Application to Childhood Asthma. ICSA Conference (Shanghai, China). Platform talk. December 22, 2016.
10. Associating Multivariate Quantitative Phenotypes with Genetic Variants in Family Samples with a Novel Kernel Machine Regression Method. ICSA Conference (Atlanta, GA). Platform talk. June 13, 2016.
11. Set-based Methods for DNA Methylation Analysis. ASHG Conference. Poster. October 8, 2015.
12. Rare-Variant Kernel Machine Test for Longitudinal Data for Population and Family Samples. JSM Conference. Platform talk. August 12, 2015.
13. Sequence kernel association test for multivariate quantitative phenotype in family samples. ASHG Conference. Platform talk. October 19, 2014.
14. Kernel Machine Testing Coupled with Rank Truncation Method for Genetic Pathway Analysis. JSM Conference. Platform talk. August 4, 2013.

AWARDS

- R03 OD034501 (NIH) Principal Investigator 2022-2023
  Integration of GTEx and Hubmap Data to Gain Population-Level Cell-Type-Specific Insights
  Total award amount: $314,739
- K01 HL138098 (NIH/NHLBI) Principal Investigator 2018-2023
  Novel Methods for Analysis of Genetic and Epigenetic Studies of Childhood Asthma
  Total award amount: $714,630
- SIRS (Columbia University) co-Principal Investigator 2021-2023
  Artificial Intelligence Based Spontaneous Preterm Birth Prediction Using Ultrasound and EMR Data
  Total award amount: $170,000
- RAC (UPMC/University of Pittsburgh) Principal Investigator 2016
  Novel Methods for Analysis of Genetic and Epigenetic Studies of Childhood Asthma
  Total award amount: $45,100
- R01 NS122449 (NIH/NINDS) Biostatistician 2021-2026
  Residential Radon Exposure and Stroke Risk: the REGARDS Study
  Total award amount: $3,129,742
- R01 AG077255 (NIH/NIA) Biostatistician 2023-2027
  Radon Exposure in Relation to the Risk of Cognitive Impairment and Mitochondrial Function
  Total award amount: $4,045,009
- R01 HL171376 (NIH/NHLBI) Biostatistician 2024-2028
  Predicting Post-Covid Pulmonary Fibrosis with Explainable Deep Learning and Optimal Biomarker Discovery
  Total award amount: $1,233,749

PROFESSIONAL SERVICES

- **Topic Editor**: statistical approaches in omics data association studies, *Frontiers in Genetics* 2020-2022
- **Editorial Board**: *Frontiers in Genetics* 2014-
- **Guest Editor**: the supplement of *Big Data Analytics for Health* 2015
- **Reviewer**: *Pediatric Pulmonology* (2023, 2021), *Nature Genetics* (2022), *eLife* (2022), *Briefings in Bioinformatics* (2021), *Patterns* (2020), *European Respiratory Journal* (2019, 2022), *Chest* (2019, 2020, 2021), *Frontiers in Genetics* (2019), *Plos One* (2015, 2018), *Meta Gene* (2017), *Scientific Reports* (2016, 2018), *Human Heredity* (2016), *Bioinformatics* (2015), *Human Genetics* (2015, 2017, 2018), *Genetic Epidemiology* (2015, 2018), *Statistics and Its Interface* (2015), *International Journal of Cancer* (2015), *Mediators of Inflammation* (2015), *Biometrics & Biostatistics International Journal* (2015), *Annals of Nutrition and Metabolism* (2014), *Annals of Human Genetics* (2014), *Colombian Journal of Statistics* (2014).