# Keren Xu

*Los Angeles, CA*

✉ xukeren418@gmail.com | 🏠 xukeren.github.io | 🐙 XUKEREN | 🔗 kerenxu | 🐦 kerenxuepi

## Employment

**Tempus AI, Inc.** *Los Angeles, CA*
Senior Bioinformatics Scientist, Bioinformatics Scientist *Mar. 2024-Present, June 2022-Mar. 2024*
- Develop bayesian probability models and loss functions for tumor purity estimation from next generation sequencing data
- Develop circular binary segmentation algorithm to identify copy number variations
- Evaluate and implement new software to improve short variant / structural variation calling and classification
- Design and conduct validation studies for Tempus first FDA-approved tumor assay (xT CDx) as part of its FDA submission
- Produce high quality and detailed documentation for all projects using git version control system
- Present project outcomes to interdisciplinary groups of scientists and engineers to translate research into clinically actionable insights

**University of Southern California, Center for Genetic Epidemiology** *Los Angeles, CA*
Predoctoral Fellow *Aug. 2018-May 2022*
- Built bioinformatics pipelines for germline/somatic variants (short variants/structural variations/copy number alterations) discovery and mutational signature analysis of over 10-terabyte human whole-genome/exome sequencing data on high-performance computing cluster
- Conducted epigenome-wide association studies, methylation quantitative trait loci analyses, causal mediation analyses, and meta-analyses to identify genetic risks that lead to leukemia by altering DNA methylation
- Conducted genome-wide association studies and construct polygenetic risk scores to assess the associations of blood cell traits with childhood ALL patient characteristics
- Detected accelerated aging in Down syndrome newborns using epigenetic age clocks derived from elastic net regression
- Assessed the impact of prenatal smoking on somatic gene deletion burden in childhood ALL patients using the polyepigenetic DNA methylation smoking score constructed by logistic lasso regression
- Held weekly lectures and office hours as a teaching assistant for 5 graduate-level courses in Biostatistics, Data analysis, and Epidemiology

**Dartmouth-Hitchcock Medical Center** *Lebanon, NH*
Statistician *July 2017-June 2018*
- Analyzed patient survey data in multiple projects using R programming
- In charge of cluster randomized trial design, survey design, IRB submission, and patient recruitment
- Collaborated with medicine fellow in writing study proposal which received $15,000 Cardiovascular Medicine Fellowship Research Award
- Organized stakeholder group meeting including 4 clinicians and 8 patients to collect data for grant application

**Columbia University Medical Center** *New York, NY*
Graduate Research Assistant *May 2016-July 2017*
- Analyzed questionnaire data using structural equation modeling, multilevel modeling, and multinomial logistic regression in SAS and R
- Created and maintained three-year survey database in SPSS to efficiently track project outcomes

## Education

**University of Southern California, Keck School of Medicine** *Los Angeles, CA*
Doctor of Philosophy in Epidemiology, GPA: 3.92/4.00 *2018-2022*
- Dissertation: Genetic Epidemiological Approaches in the Study of Risk Factors for Hematologic Malignancies

**Columbia University, Mailman School of Public Health** *New York, NY*
Master of Public Health in Epidemiology, Certificate in Advanced Epidemiology, GPA: 3.81/4.00 *2015-2017*
- Thesis: Modeling the Impacts of Numbers of Affected Relatives and Perceived Chance of Having Epilepsy-Related Mutation

**East China University of Science and Technology, School of Pharmacy** *Shanghai, China*
Bachelor of Science in Pharmaceutical Sciences, Minor in English, GPA: 3.73/4.00 *2011-2015*
- Thesis: Pluripotent Stem Cells Culture and ICOSL Gene Knockout

## Relevant Skills

**Data Science:** Prediction, Inference, Machine Learning, Visualization, Data Cleaning, Modelling.
**Programming:** Python, R, SQL, Linux/Unix, Bash, Git, Docker, Singularity, AWS, GCP, SAS, STATA, SPSS.
**Epidemiology & Biostatistics:** Clinical trial, Study design, Quantitative analysis, Systematic review, Meta analysis.
**Bioinformatics:** GATK, PLINK, Conda, Bioconductor, WDL, EWAS, GWAS, PRS.
**Languages:** Mandarin (Native), English (Full professional working proficiency).

## Certifications

| 2023 | Python 3 Programming Specialization by University of Michigan on Coursera |
|------|--------------------------------------------------------------------------|
| 2023 | Applied Data Science with Python Specialization by University of Michigan on Coursera |
| 2021 | Machine Learning by Stanford University on Coursera |
| 2021 | Genomic Data Science Specialization by Johns Hopkins University on Coursera |
| 2020 | Association Mapping: GWAS and Sequencing Data, 25th Summer Institute in Statistical Genetics, University of Washington, Seattle, WA |
| 2020 | Computational Pipeline for WGS Data, 25th Summer Institute in Statistical Genetics, University of Washington, Seattle, WA |
| 2020 | The Unix Workbench by Johns Hopkins University on Coursera |
| 2019 | American Language Institute Certified International Teaching Assistant |
| 2019 | Data Scientist with R Track on DataCamp |
| 2017 | R Programming by Johns Hopkins University on Coursera |
| 2017 | SAS Certified Advanced Programmer for SAS 9 |
| 2012 | National Computer Rank Examination C Programming Level 2 |

## Honors & Awards

| 2020-2022 | USC Graduate Research Assistantship |
|-----------|-------------------------------------|
| 2019-2020 | USC Graduate Teaching Assistantship |
| 2018-2019 | USC Graduate Student Fellowship |
| 2015 | Shanghai City Outstanding Graduate |
| 2014 | Jiang Hualiang (Dean) Scholarship First Prize |
| 2013 | Shanghai City Undergraduate Innovation Project |
| 2013 | ECUST Social Work Prize Level A |
| 2013 | ECUST Outstanding Student Leader |
| 2012-2013 | ECUST Outstanding Student |
| 2012-2013 | ECUST Integrated Curriculum First Prize |
| 2012 | ECUST Social Work Prize Level B |
| 2012 | Jiang Hualiang (Dean) Scholarship Third Prize |
| 2012 | The Fourth National Undergraduate Mathematical Competition Third Prize |

## Publications & Presentations

13 peer-reviewed journal publications (4 first author), 3 conference presentations, 3 workshops teaching R programming, and 10 accepted conference abstracts. Full list available on `Google Scholar`.

### PUBLICATIONS

- Liu, T., **Xu, K.**, Pardeshi, A., Myint, S. S., Kang, A. Y., Morimoto, L. M., Lieber, M. R., Wiemels, J. L., Kogan, S. C., Metayer, C., de Smith, A. J. Early-life tobacco exposure is causally implicated in aberrant RAG-mediated recombination in childhood acute lymphoblastic leukemia. *Leukemia*, 2024.
- Feng, Q., **Xu, K.**, Shah, M., Li, S., Leavitt, A. D., Godley, L. A., de Smith, A. J., Wiemels, J. L.. Evaluation of the genetic basis of familial-associated early-onset hematologic cancers in an ancestral/ethnically diverse population. *Haematologica*, 2024.
- Wang, A., Xu, Y., Yu, Y., Nead, K. T., Kim, T., **Xu, K.**, Dadaev, T., Saunders, E., Sheng, X., Wan, P., …, Clonal hematopoiesis and risk of prostate cancer in large samples of European ancestry men. *Human Molecular Genetics*, 2023.
- Sanogo, F., **Xu, K.**, Cortessis, V., Weigensberg, M. J., Watanabe, R. M., Mind and Body-based Interventions Improve Glycemic Control in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis. *Journal of Integrative and Complementary Medicine*, 2023.
- **Xu, K.**, Li, S., Pandey, P., Kang, A. Y., Morimoto, L. M., Mancuso, N., Ma, X., Metayer, C., Wiemels, J. L., de Smith, A. J., Investigating DNA methylation as a mediator of genetic risk in childhood acute lymphoblastic Leukemia. *Human Molecular Genetics*, 2022.
- **Xu, K.**, Li, S., Muskens, I. S., Elliott, N., Myint, S. S., Pandey, P., Hansen, H. M., Morimoto, L. M., Kang, A. Y., Ma, X., Metayer, C., Mueller, B. A., Roberts, I., Walsh, K. W., Horvath, S., Wiemels, J. L., de Smith. A. J., Accelerated Epigenetic Aging in Newborns with Down Syndrome. *Aging Cell*, 2022.
- Li, S., Sok, P., **Xu, K.**, Muskens, I. S., Elliott, N., Myint, S. S., Pandey, P., Hansen, H. M., Morimoto, L. M., Kang, A. Y., Metayer, C., Ma, X., Mueller, B. A., Roy, A., Roberts, I., Rabin, K. R., Brown, A. L., Lupo, P. J., Wiemels, J. L., de Smith, A. J., Epigenome-Wide Association Study of Acute Lymphoblastic Leukemia in Children with Down Syndrome. *Blood Adv*. 2022.
- Yu, F., Cato, L. D., Weng, C., Liggett, L. A., Jeon, S., **Xu, K.**, Chiang, C. W., Wiemels, J. L., Weissman, J. S., de Smith, A. J., Sankaran, V. G., Variant to function mapping at single-cell resolution through network propagation. *Nature Biotechnology*, 2022.
- **Xu, K.**, Feng, Q., Wiemels, J. L., de Smith, A. J., Disparities in acute lymphoblastic leukemia risk and survival across the lifespan in the United States of America. *Journal of Translational Genetics and Genomics*, 2021.
- **Xu, K.**, Li, S., Whitehead, T. P., Pandey, P., Kang, A. Y., Morimoto, L. M., Kogan, S., Metayer, C., Wiemels, J. L., de Smith, A. J., Epigenetic Biomarkers of Prenatal Tobacco Smoke Exposure are Associated with Gene Deletions in Childhood Acute

Lymphoblastic Leukemia. *Cancer Epidemiology and Prevention Biomarkers*, 2021. **Featured in "Cancer Epidemiology, Biomarkers & Prevention Highlights: Selected Articles from This Issue."**

- Schott, S. L., Berkowitz, J., Dodged, S. E., Petersen, C. L., Saunders, C. H., Sobti, N. K., **Xu, K.**, Coylewright, M., Personalized, Electronic Health Record–Integrated Decision Aid for Stroke Prevention in Atrial Fibrillation: A Small Cluster Randomized Trial and Qualitative Analysis of Efficacy and Acceptability. *Circulation: Cardiovascular Quality and Outcomes*, 2021.
- Coylewright, M., O'Neill, E., Sherman, A., Gerling, M., Adam, K., **Xu, K.**, Grande, S. W., Dauerman, H. L., Dodged, S. E., Sobti, N. K., Saunders, C. H., Schott, S. L., Elwyn, G., Durand, M., The Learning Curve for Shared Decision-making in Symptomatic Aortic Stenosis. *JAMA cardiology*, 2020.
- Coylewright, M., Keevil, J. G., **Xu, K.**, Dodge, S., Frosch, D., Field, M. E., Pragmatic Study of Clinician Use of a Personalized Patient Decision Aid Integrated into the Electronic Health Record: An Eight-year Experience. *Telemedicine and e-Health*, 2019.

## Abstracts

- Sanogo, F., **Xu, K.**, Cortessis, V., Weigensberg, M. J., Watanabe, R. M., Meta-analysis Shows Mind- and Body-Based Interventions Improve Glycemic Control in Patients with Type 2 Diabetes. Poster presentation at the American Diabetes Association's 82nd Scientific Sessions, 2022.
- **Xu, K.**, Wadé, N. B., Hwang, A., Conti, D. V., Salehi, M., Mack, T. M., Cortessis, V., van den Berg, D., de Smith, A. J., Cozen, W., Whole-exome sequencing in multiplex families to identify novel AYA classical Hodgkin lymphoma predisposition genes. Poster presentation at the 63rd American Society of Hematology (ASH) Annual Meeting and Exposition, 2021.
- Li, S., Sok, P., **Xu, K.**, Muskens, I. S, Elliott, N., Myint, S. S., Pandey, P., Hansen, H. M., Morimoto, L. M., Kang, A. Y., Metayer, C., Ma, X., Mueller, B. A., Roberts, I., Rabin, K. R., Brown, A. L., Lupo, P. J., Wiemels, J. L., de Smith, A. J., Epigenome-wide association study of acute lymphoblastic leukemia in children with Down syndrome. Oral presentation at the 63rd American Society of Hematology (ASH) Annual Meeting and Exposition, 2021.
- **Xu, K.**, Wadé, N. B., Hwang, A., Conti, D. V., Salehi, M., Mack, T. M., Cortessis, V., van den Berg, D., de Smith, A. J., Cozen, W., Whole-exome sequencing in multiplex families to identify novel AYA classical Hodgkin lymphoma predisposition genes. 15 minutes talk at the 2021 International Lymphoma Epidemiology Consortium Annual Meeting, 2021.
- **Xu, K.**, Li, S., Muskens, I. S, Elliott, N., Myint, S. S., Pandey, P., Ma, X., Metayer, C., Mueller, B. A., Walsh, K. W., Roberts, I., Horvath, S., Wiemels, J. L., de Smith. A. J., Accelerated epigenetic aging in newborns with Down syndrome. Poster presentation at the American Society of Human Genetics (ASHG) Virtual Meeting, 2021.
- Schott, S. L., **Xu, K.**, Berkowitz, J., Petersen, C., Saunders, C., Sobti, N., Coylewright, M., Timing of Electronic Health Record Integrated Decision Aid (IDEA) for Stroke Prevention in Atrial Fibrillation Matters. American College of Cardiology Scientific Sessions, 2019.
- Coylewright, M., Sherman, A., Grande, S. W., **Xu, K.**, Kirk, J., Dillon, G., O'Neill, E., Elwyn, G., Does a Telehealth "Virtual Consult" Including Referring Physicians, Specialist Physicians and Patients Increase Shared Decision Making for Patients with Heart Disease? Quality of Care and Outcomes Research Scientific Sessions, 2018.
- Coylewright, M., O'Neill, E., Sherman, A., Gerling, M., Adam, K., **Xu, K.**, Grande, S., Dauerman, H. L., Elwyn, G., A Learning Curve for Shared Decision Making: The Impact of Physician Experience on Decision Aid Efficacy in Severe Aortic Stenosis. American College of Cardiology Scientific Sessions, 2018.
- Castaner, M. M., **Xu, K.**, Repka K., Gennetian L., Kennedy, J., Duch, H., A Multi-Dimensional Approach to Parent Engagement. Results from An Integrated School Readiness Intervention. Society for Research in Child Development, 2017 Biennial Meeting.
- **Xu, K.**, Castaner, M. M., Duch, H., A Multi-Dimensional Approach to Parent Engagement. Results from an Integrated School Readiness Intervention. Mailman School of Public Health. EPI Masters Student Day, 2016. Poster Presentation.

## Conference Presentations

- Whole-exome sequencing in multiplex families to identify novel AYA classical Hodgkin lymphoma predisposition genes (Poster presentation). the 63rd American Society of Hematology (ASH) Annual Meeting and Exposition, Virtual. Dec. 2021.
- Accelerated epigenetic aging in newborns with Down syndrome (Poster presentation). the American Society of Human Genetics (ASHG) 2021 Meeting, Virtual. Oct. 2021.
- Whole-exome sequencing in multiplex families to identify novel AYA classical Hodgkin lymphoma predisposition genes (15-minute talk). the 2021 International Lymphoma Epidemiology Consortium Annual Meeting, Virtual. Sept. 2021.

## Workshops

- Automatic Reports with RMarkdown. One hour workshop on RMarkdown for MPH/MS/PhD students at the Keck School of Medicine of USC, University of Southern California. Aug. 2020.
- Draw fractals from root finding iteration in R. Half hour workshop on drawing fractals from root finding iteration in R, LA R Users Group. April 2020.
- Building a Personal Website Using Blogdown. Two hour workshop on creating personal website using R package Blogdown, R-Ladies Pasadena. Nov. 2019.