Yajie Zhao

MSc (Oxon), PhD (Cantab)
MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine,
Cambridge CB2 0QQ
Email: yajie.zhao@mrc-epid.cam.ac.uk
ORCID: https://orcid.org/0000-0002-2747-0219
Website: https://www.mrc-epid.cam.ac.uk/people/yajie-zhao/
Google Scholar: https://scholar.google.com/citations?user=dSnrVrYAAAAJ&hl=en

---

EUDCATION

University of Cambridge, Cambridge, 2019 – 2022
PhD in Medical Science at MRC Epidemiology Unit, Pass without corrections
Thesis title: Detection, causes and consequences of sex chromosome mosaicism
University of Oxford, Oxford, 2018 – 2019
MSc in Mathematical Sciences, Pass
Dissertation title: The RAPIDD Ebola forecasting challenge:
epidemiological insights and statistical performance
University of Warwick, Coventry, 2015 – 2018
BSc in Mathematics, Operational Research, Statistics and Economics, 1st Honours

RESEARCH INTERESTS

My work involves two topics:

1. Mosaic loss of the X/Y chromosome – Mosaic loss of the X/Y chromosome (mLOX/Y) is the most common form of somatic mosaicism for females and males, respectively. The genetic causes and phenotypic consequences of these events are still unclear. I have developed a formula to combine different mosaic loss of the X/Y chromosome(mLOX/Y) measurement based on SNP-array and high throughput sequencing data and applied it on large trans-ethnic cohorts to reveal the genetic architecture of mLOX/Y through GWAS and Exome-wide burden analysis. Using this approach, I have found genes associated with mLOX/Y. Disruption of these genes can substantially increase the risk to have mLOY and my future work will combine dry lab work and wet lab work to investigate the mechanism behind mLOX/Y.
2. How rare variants affect human traits – The decreased cost of sequencing and its application to large research cohorts provide a chance to systematically explore how rare coding variants affect human traits. I have designed, implemented, and utilised software to identify genetic associations in a cohort of 200K humans from the UK Biobank study. I am interested in applying these tools to larger and more diverse populations to identify gene-phenotype associations and develop these findings into novel therapeutic targets.

POSITIONS

MRC Postdoctoral Fellow (2022 – Present)

MRC Epidemiology Unit, Early Life Aetiology and Mechanisms of Diabetes and Related Metabolic Disorders Group led by Prof. Ken Ong and Prof. John Perry

This post plans to utilise such data in the UK Biobank study to identify rare, protein-coding, genetic variants associated with obesity to highlight novel genes and biological mechanisms. This work will involve developing new computational pipelines and methods to perform quality control and analysis on billions of data points in half a million study participants. Genes highlighted by these analyses will then be shared with experimental collaborators, both internally in Cambridge and externally, to gain further mechanistic insight in animal and cellular models.

Graduate Student (2019 – 2022)

MRC Epidemiology Unit, Early Life Aetiology and Mechanisms of Diabetes and Related Metabolic Disorders Group led by Prof. Ken Ong and Prof. John Perry

Testing and implementing bioinformatic software to estimate somatic mosaicism; Designing pipeline to conduct exome-wide analysis; Undertaking the microarray-data based genome-wide analysis and sequencing-data based exome-wide analysis; Building collaborations with other large cohorts including China Kadoorie Biobank and Taiwan Biobank and wet labs.

Initiator and Organiser (2022 – Present)

*Human Genetics Network (*https://mp.weixin.qq.com/s/qPBU6rbpVGJ5Pp4XDDVlXA *)*

Building the communication network for the researchers interested in human genetics; Organising the weekly talks delivered by the world-leading researchers and the early-career scholars in human genetics and related areas; Managing the social media accounts to advertise talks and attract more researchers to join the network.

Part-time Assistant Editor (2016 – 2022)

The-Intellectual, founded by three Chinese scientists Yi Rao (Peking University), Bai Lu (Tsinghua University) and Yu Xie (Princeton University).

Interviewing the specialists about scientific topics; Introducing the new progress in science on social media; Managing a social media account that has 2.7 million followers; Providing reliable scientific knowledge and information for the Chinese audience.

Data Analyst Intern (2019.7 – 2019.9)

Chinese Institute for Brain Research, Beijing (CIBR), Lab of Dr. Li Zhang and Dr. Jian Chen

Identification of markers of drug sensitivity in glioblastoma stem cells based on multi-comics data; Routine data analysis.

PUBLICATIONS

Published Papers (Equal 1st Authors - *)

1. Mathieson I*, Day FR*, Barban N*, Tropf FC*, Brazel DM*, (many other authors), **Zhao Y**, et al. Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the *FADS* locus [published online ahead of print, 2023 Mar 2]. *Nat Hum Behav*. 2023;10.1038/s41562-023-01528-6. doi:10.1038/s41562-023-01528-6.
2. **Zhao Y**, Gardner EJ, Tuke MA, et al. Detection and characterization of male sex chromosome abnormalities in the UK Biobank study. *Genet Med*. 2022;24(9):1909-1919. doi:10.1016/j.gim.2022.05.011
3. Koprulu M*, **Zhao Y***, Wheeler E, et al. Identification of Rare Loss-of-Function Genetic Variation Regulating Body Fat Distribution. *J Clin Endocrinol Metab*. 2022;107(4):1065-1077. doi:10.1210/clinem/dgab877
4. **Zhao Y**, Stankovic S, Koprulu M, et al. *GIGYF1* loss of function is associated with clonal mosaicism and adverse metabolic health. *Nat Commun.* 2021;12(1):4178. Published 2021 Jul 7. doi:10.1038/s41467-021-24504-y
5. Stankovic S, Day FR, **Zhao Y** et al. Elucidating the genetic architecture underlying IGF1 levels and its impact on genomic instability and cancer risk [version 1; peer review: 2 not approved]. *Wellcome Open Res.* 2021, 6:20. doi:10.12688/wellcomeopenres.16417.1

Papers in Review/Preparation

1. Liu A*, Genovese G*, **Zhao Y***, et al. Population analyses of mosaic X chromosome loss identify genetic drivers and widespread signatures of cellular selection. In Revision. *Nature*. (2023). doi: https://doi.org/10.1101/2023.01.28.23285140
2. Brown DW*, Cato LD*, **Zhao Y***, et al. Shared and distinct genetic etiologies for different types of clonal hematopoiesis. Under Review. *Nat Commun*. (2022). doi: https://doi.org/10.1101/2022.03.14.483644
3. Stankovic S*, Shekari S*, Huang QQ*, Gardner EJ*, Owens NDL*, (5 other authors), **Zhao Y**, et al. Genetic susceptibility to earlier ovarian ageing increases de novo mutation rate in offspring. Submitted and available on medRxiv (2022). doi: https://doi.org/10.1101/2022.06.23.22276698
4. Kaisinger LR*, Kentistou KA*, Stankovic S*, Gardner EJ, Day FR, **Zhao Y**, et al. Large-scale exome sequence analysis identifies sex- and age-specific determinants of obesity. In Revision. *Cell Genomics.* (2022).

Invited Talks (presenting author - *)

Zhao Y*, Gardner EJ, Kentistou KA, et al. Detection, causes and consequences of Y chromosome mosaicism. Oct 26th, 2022. Platform - Somatic mosaicism in human health and disease. American Society of Human Genetics 2022 Annual Meeting

Liu A*, Genovese G, **Zhao Y**, et al. Genetic investigation of mosaic loss of the X chromosome in peripheral leukocytes of 918,085 women identifies germline predisposition and strong signals of haplotype selection. Oct 26th, 2022. Platform - Somatic mosaicism in human health and disease. American Society of Human Genetics 2022 Annual Meeting

Zhao Y. *GIGYF1* loss of function is associated with clonal mosaicism and adverse metabolic health. June 21st , 2022. Selected submission of UK Biobank Doctoral Student of the Year: Thesis in Three Minutes. UK Biobank Scientific Conference 2022

Liu A*, Genovese G, **Zhao Y**, et al. Genetic determinants of mosaic loss of the X chromosome in peripheral leukocytes of 800K women from 7 biobanks. June 11th – 14th , 2022. Presented in the session “PV02 Poster Viewing with Authors (Group B). ”. European Human Genetics Conference 2022

Zhao Y. Introduction about UK Biobank. January 7th , 2022. Chen Wu Group. Department of Etiology and Carcinogenesis, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Zhao Y. Overview of current human genetics analysis based on genomic data. August 13th, 2021. Yi Rao Lab. School of Life Sciences. Peking University

Zhao Y. Overview of current human genetics analysis based on genomic data. July 30th , 2021. Xiaoliang Sunney Xie Group. Beijing Advanced Innovation Center for Genomics. Peking University

TECHNICAL EXPERIENCE

Programming: Skilled in R programming language and UNIX shell use. Experience with Python programming language, the DNA Nexus Amazon Web Services interface, and Genomics England analysis environment.

Bioinformatics: Enough knowledge and experience of many common bioinformatic pipelines including genome-wide association analysis (BOLT-LMM), exome-wide association and rare variant analysis (BOLT-LMM, STAAR and SAIGE), Mendelian randomisation, LD Score regression (LDSC), polygenic risk score modelling, meta-analysis and functional annotation (COJO and MAGMA). Experience with softwares calling mosaic chromosomal alterations (MoChA) from array and whole genome sequencing data and estimating telomere length (Telomerecat) from whole genome sequencing data. Experience with advanced tools for genomic analyses (PLINK, samtools, Picard and SeqArray).

Statistics/Machine Learning: Knowledgeable in the theories and application of different traditional statistical and mathematical methods and novel machine leaning techniques.

REWARDS

2021 Chinese Government Award for Outstanding Self-financed Students Abroad (Amount: $6,000)

SOCIETY MEMBERSHIP

Team Leader (2016 – 2017)
Mandarin debating team of University of Warwick.

Team management and training arrangement; Recruitment and advertising; Organisation of the UK-wide competition.