

# CURRICULUM VITAE

## Personal information

Jing Hua Zhao

Department of Public Health and Primary Care, Cardiovascular Epidemiology Unit

University of Cambridge

Victor Phillip Dahdaleh Heart & Lung Research Institute

Papworth Road, Cambridge Biomedical Campus

Cambridge CB2 0BB

Email: [jinghuazhao@hotmail.com](mailto:jinghuazhao@hotmail.com)

Web: <https://jinghuazhao.github.io/>, <https://cambridge-ceu.github.io/>

## Education

- PhD, Statistical Genetics (supervisor: Prof Pak Sham), King's College London, UK

## Professional positions

2005.9-2018.7 Investigator scientist in Genetics, MRC Epidemiology Unit

2018.8- Genetic Analyst / Senior Research Associate, Cardiovascular Epidemiology Unit,  
Department of Public Health and Primary Care, University of Cambridge

## Research interests

My work is human health-related research which over years includes familial aggregation, segregation, linkage, candidate genes and genomewide association studies (GWASs). The most recent is proteogenomics within the SCALLOP consortium using the Olink and mass spectrometry (MS) panels measured for the INTERVAL samples. I have also led collaborative analysis to the SCALLOP-Seq(ence, both WES and NGS) consortium and Host Genetics Initiative.

I have promoted reproducible research through CRAN (<https://cran.r-project.org>), GitHub (<https://github.com>) and websites. I developed genetic analysis package (gap), protein quantitative trait tools (pQTLtools) and curated <https://jinghuazhao.github.io/Computational-Statistics/>, <https://jinghuazhao.github.io/software-notes/> and <https://jinghuazhao.github.io/Omics-analysis/>. By closely following up developments in computational statistics, machine learning and artificial intelligence, I have made computing and omics analysis tools available from the University HPC, <https://cambridge-ceu.github.io/csd3/>, whose components include AI with BitNet, GeminiCLI, Ollama, llama.cpp, llm, featuring AI for MS data with InstaNovo & DIA-NN, molecule optimization with DrugAssist, single-cell omics with Seurat, scp, scanpy, scvi-tools, scGPT, C2S-Scale, mtDNA analysis with MToolBox, fNUMT, haplogrep as well as long-read sequencing analysis with SVanalyzer, hap.py, sniffles, truvari. Explorations in population genetics include selscan, angsd, relate, clues2 and fastsimcoal2. These are coupled with extensive experiments on various models, architectures, platforms as with libraries such as PyTorch, scikit-llm, LangChain.

## Key publications

**Zhao JH**, et al. Mapping pQTLs of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and novel therapeutic targets. *Nat Immunol* 2023, **24**(9):1540-1551, 10.1038/s41590-023-01588-w, <https://www.nature.com/articles/s41590-023-01588-w>.

COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* **600**:472–477 (2021); Pathak, G.A. et al. A first update on mapping the human genetic architecture of COVID-19. *Nature* **608**:E1-E10 (2022); The Host Genetics Initiative. A second update on mapping the human genetic architecture of COVID-19. *Nature* **621**:E7–E26 (2023).



**Zhao JH**, Luan JA, Congdon P. Bayesian linear mixed model of polygenic effects. *J Stat Soft*. 2018, **85**(6):1-27. doi: 10.18637/jss.v085.i06

**Zhao JH**, Luan JA. Mixed modeling with whole genome data. *J Prob Stat*. 2012. doi: 10.1155/2012.485174.

Xue F, Li S, Luan J, Yuan Z, Luben RN, Khaw K-T, Wareham NJ, Loos RJF, **Zhao JH**. A latent variable partial least squares path modeling approach to regional association and polygenic effect with applications to a human obesity study. *PLoS ONE* 2012, **7**(2): e31927

Loos RJ, *et al*. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008; **40**(6):768-75

**Zhao JH**. gap: genetic analysis package. *J Stat Soft* 2007, **23** (8):1-18. doi: 10.18637/jss.v023.i08.

## Publications since 2018

188. Demenais F, *et al*. Multi-ancestry genome-wide association study identifies new asthma susceptibility loci that co-localize with immune cell enhancer histone marks. *Nat Genet* 2018; **50**(1):42-53
189. Medina-Gomez C, *et al*. Life-course genome-wide association study meta-analysis of total body BMD and assessment of age-specific effects. *Am J Hum Genet* 2018; **102**(1):88-102
190. Sung YJ, *et al*. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am J Hum Genet* 2018, **102**(3):375-400. doi: 10.1016/j.ajhg.2018.01.015.
191. Turcot V, *et al*. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure underpinning obesity. *Nat Genet* **50**:26-41
192. **Zhao JH**, Luan JA, Congdon P. Bayesian linear mixed model of polygenic effects. *J Stat Soft* 2018, **85**(6):1-27. doi: 10.18637/jss.v085.i06
193. Feitosa MF, *et al*. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570K individuals across multiple ancestries. *PLoS One* 2018, **13**(6):e0198166. doi: 10.1371/journal.pone.0198166.
194. Lee JJ, *et al*. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018, **50**:1112–1121, <https://doi.org/10.1038/s41588-018-0147-3>.
195. Ligthart S, *et al*. Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *Am J Hum Genet* 2018, **103**(5):691-706. doi: 10.1016/j.ajhg.2018.09.009.
196. Evangelou E, *et al*. Genetic analysis of over 1 million people identifies 535 new loci for blood pressure traits. *Nat Genet* 2018, **50**(10):1412-1425. doi: 10.1038/s41588-018-0205-x.
197. Merino J, *et al*. Genome-wide meta-analysis of macronutrient intake of 91,114 European ancestry participants from the cohorts for heart and aging research in genomic epidemiology consortium. *Mol Psychiatr* 2018 Jul 9. doi: 10.1038/s41380-018-0079-4.
198. Kilpeläinen TO, *et al*. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Comm* 2019, **10**, 376, <https://doi.org/10.1038/s41467-018-08008-w>.
199. Giri A, *et al*. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* 2019, **51**:51-62
200. Karasik D, *et al*. Disentangling the genetics of lean mass. *Am J Clin Nutr* 2019, **109**(2): 276-287, <https://doi.org/10.1093/ajcn/nqy272>.

□

201. de Vries PS, et al. Multi-ancestry genome-wide association study of lipid levels incorporating gene-alcohol interactions. *Am J Epidemiol* 2019, 188(6):1033-1054, <https://doi.org/10.1093/aje/kwz005>.
202. Justice AE, et al. Protein-coding variants highlight the importance of lipolysis in adipocytes for body fat distribution. *Nat Genet* 2019, <https://doi.org/10.1038/s41588-018-0334-2>.
203. Shrine N, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries, *Nat Genet* 2019, <https://doi.org/10.1038/s41588-018-0321-7>.
204. Zhao J, et al. Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. *Nat Comm* 2019, 10, Article number: 1534 (2019), <https://www.nature.com/articles/s41467-019-09462-w>
205. Sung YJ, et al. A multi-ancestry genome-wide study incorporating gene–smoking interactions identifies multiple new loci for pulse pressure and mean arterial pressure. *Hum Mol Genet* 2019, doi: 10.1093/hmg/ddz070
206. Clark DW, et al. Associations of autozygosity with a broad range of human phenotypes. *Nat Comm* 2019, NCOMMS-18-33232A-Z
207. Bentley A.R., et al. Multi-ancestry genome-wide gene–smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat Genet* **51**, 636–648 (2019). <https://doi.org/10.1038/s41588-019-0378-y>
208. Shah S, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Comm* 2020, 11(1):163. Published 2020 Jan 9. doi:10.1038/s41467-019-13690-5
209. Surendran P, et al. Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. *Nat Genet* 2020, 52(12):1314-1332, doi: 10.1038/s41588-020-00713-x.
210. Lin W, Ji J, Zhu Y, Li M, Zhao J, Xue F, Yuan Z. PMINR: pointwise mutual information-based network regression – with application to studies of lung cancer and Alzheimer’s disease. *Front Genet* 2020, 11:556259, <https://doi.org/10.3389/fgene.2020.556259>
211. Cuellar-Partide G, et al. Genome-wide association study identifies 48 common genetic variants associated with handedness. *Nat Hum Behaviour* 2021, 5: 59–70.
212. Gaziano L, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat Med* 2021, 27:668–676.
213. Chen J, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* 2021, 53:840–860.
214. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* 2021.
215. Zhang Y, et al. Mendelian randomisation highlights hypothyroidism as a causal determinant of idiopathic pulmonary fibrosis. *EBiomed*. 2021, <https://doi.org/10.1016/j.ebiom.2021.103669>
216. Graham SE, et al. The power of genetically diverse individuals in genome-wide association studies of blood lipid levels. *Nature* 2021, **600**: 675–679, DOI:10.1038/s41586-021-04064-3
217. Jin X, Zhang L, Ji J, Ju T, Zhao JH, Yuan Z. NeRiT -- Network regression in transcriptome-wide association studies. *BMC Genomics* (2022) 23:562, <https://doi.org/10.1186/s12864-022-08809-w>
218. Pathak, G.A. et al. A first update on mapping the human genetic architecture of COVID-19. *Nature* **608**, E1-E10 (2022).
219. Ramdas S, et al. A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids. *Am J Hum Genet*. 2022 Aug 4;**109**(8):1366-1387. doi: 10.1016/j.ajhg.2022.06.012.
220. Wang Z, et al. Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention *Nat Genet*. 2022, **54**:1332–1344

□

221. Yengo L, et al. A saturated map of common genetic variants associated with human height from 5.4 million individuals of diverse ancestries. *Nature* 2022, **610**:704–712.
222. Kanoni, S., et al. Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis. *Genome Biology* 2022, **23**(1): 268 (medRxiv, DOI:10.1101/2021.12.15.21267852).
223. Shrine N, et al. Multi-ancestry genome-wide association study improves resolution of genes, pathways and pleiotropy for lung function and chronic obstructive pulmonary disease. *Nat Genet* 2023, **55**:410–422
224. van de Vegte JY, et al. Genetic insights into resting heart rate and its role in cardiovascular disease. *Nat Comm* **14**:4646 (2023), <https://www.nature.com/articles/s41467-023-39521-2>
225. **Zhao JH**, et al. Mapping pQTLs of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and novel therapeutic targets. *Nat Immunol* 2023, **24**(9):1540-1551, 10.1038/s41590-023-01588-w, <https://www.nature.com/articles/s41590-023-01588-w>.
226. Lagou V, et al. GWAS of random glucose in 476,326 individuals provide insights into diabetes pathophysiology, complications and treatment stratification. *Nat Genet* **55**:1448–1461 (2023), <https://www.nature.com/articles/s41588-023-01462-3>, *medRxiv*, DOI: 10.1101/2021.04.17.21255471.
227. The Host Genetics Initiative. A second update on mapping the human genetic architecture of COVID-19. *Nature* **621**:E7–E26 (2023).
228. Macdonald-Dunlop E, et al. Mapping genetic determinants of 184 circulating proteins in 26,494 individuals to connect proteins and diseases. *medRxiv*, DOI:10.1101/2021.08.03.21261494.
229. Klaric L, et al. Mendelian randomisation identifies alternative splicing of the FAS death receptor as a mediator of severe COVID-19. *medRxiv*, doi: <https://doi.org/10.1101/2021.04.01.21254789>.
230. Gaziano L, et al. Transcriptome- and proteome-wide Mendelian randomization to prioritize therapeutic targets for coronary heart disease. *medRxiv*, DOI: 10.1101/2024.06.27.24309406v1.
231. **Zhao JH**. Genetic association analysis with R (II). *The Biomedical & Life Sciences Collection* **2025**, <https://github.com/jinghuazhao/gaawr2>, (2025).
232. Kelemen M, et al. Performance of deep-learning based approaches to improve polygenic scores. *Nat Comm* **16**, 5122 (2025). <https://doi.org/10.1038/s41467-025-60056-1>
233. Smit RAJ, et al. Polygenic scores to predict body mass index and obesity across populations and through the life course powered by data from 5.1 million individuals, *Nat Med* (2025). <https://www.nature.com/articles/s41591-025-03827-z>.
234. Koprulu M, et al. Large-scale proteogenomic meta-analyses highlight N-linked glycosylation as important regulator of the circulating proteome.
235. Bo C, Li R, Zhao JH, Feng S, Ji X, Xue F. Agentic T-STROBE: A Front-Shifted STROBE paradigm for observational research.