Tom Willis

tom.willis@live.com   twillis.xyz   twillis1993   twillis209   Thomas W. Willis

I am a final-year PhD student in biostatistics at the University of Cambridge and will defend in November 2024. I have worked in bioinformatics and statistics in academia and industry, predominantly on GWAS and RNA-seq analysis. My PhD work relates to the role of common variants in rare disease, in particular in the antibody deficiencies selective IgA deficiency and common variable immunodeficiency. I have an interest in reproducibility and have leveraged my background in computer science to develop well-maintained, easily-deployed scientific software and pipelines to support my work.

Skills

Programming

- Daily: `Python`, `R`, `bash` in a Linux environment
- Infrequently: `C++` (alone and integrated with `R` through `Rcpp`
- Past experience: `Java`, `C`, `SQL`, `Javascript` with Vue 3 and Cloud services, use of `MPI` and `OpenCL` for GPU programming

Ancillaries

- `git`, `CMake`, `docker`, `conda`
- Unix-style core utilities like `awk`, `sed`, etc.
- Use of `slurm` and the SGE/UGE system on HPC facilities, parallelism with `openmp`, `parallel`, etc.

Bioinformatics

- Routine *seq utilities such as `plink`, `samtools`, etc.

Bioinformatics (cont.)

- `snakemake` workflows
- Contributions to `R` packages `locuszoomr` and `fcfdr`, use of `devtools`, `conda`, and `testthat` in package development
- GWAS with `fastGWA-GLMM`
- Past RNA-seq experience with `salmon` and `DESeq2`

Education

University of Cambridge *PhD Biostatistics* Oct 2020 – Nov 2024

- Thesis on the role of common variants in antibody deficiencies.
- End-to-end generation and analysis of GWAS data from genotype imputation and QC to regression with `fastGWA-GLMM`, visualisation, hit work-up, and heritability and genetic correlation estimation.
- Improvement of an existing nonparametric method, the ‘GPS test’, for the detection of genetic similarity in the small-sample context, implemented in `C++`.

University of Leeds *MSc Statistics, Distinction* Oct 2019 – Sep 2020

- Taught Master’s with dissertation on the phenomenon of zero inflation in scRNA-seq data.

University of Leeds *BSc Computer science, 1st class* Oct 2015 – Jun 2019

Experience

Bioinformatics intern *Novartis Institutes for BioMedical Research, Emeryville, CA* Jul 2017 – Jun 2018

- Bulk RNA-seq in bacteria to assist antibiotic development and interrogate mechanisms of resistance, culminating in coauthorship of a paper characterising regulation of efflux pump expression in *Pseudomonas aeruginosa* (Ranjitkar et al.)
- Development of a Python 2 library for querying internal genomic data and metadata for use cases common across disease areas. The library’s API was exposed to wet-lab staff through a web portal. Development with bioinformatic libraries like `biopython` and `faidx`.

Amgen Scholar *University of Cambridge* Jul 2018 – Sep 2018

- Bulk *de novo* RNA-seq in two species of African cichlid fish to identify differential expression of orthologous transcripts in testes and liver and novel ORFs contributing to speciation, published as Puntambekar et al. (see Publications).

Publications

Leveraging pleiotropy identifies common-variant associations with selective IgA deficiency — Sep 2024
Thomas W. Willis, Effrossyni Gkrania-Klotsas, Nicholas J. Wareham, Eoin McKinney, Paul A. Lyons, Kenneth G. C. Smith, and Chris Wallace
10.1016/j.clim.2024.110356

Accurate detection of shared genetic architecture from GWAS summary statistics in the small-sample context — Aug 2023
Thomas W. Willis and Chris Wallace
10.1371/journal.pgen.1010852

Leveraging auxiliary data from arbitrary distributions to boost GWAS discovery with Flexible cFDR — Oct 2021
Anna Hutchinson, Guillermo Reales, ***Thomas W. Willis***, and Chris Wallace
10.1371/journal.pgen.1009853

Evolutionary divergence of novel open reading frames in cichlids speciation — Dec 2020
Shraddha Puntambekar *et al.*
10.1038/s41598-020-78555-0

Target (MexB)- and Efflux-Based Mechanisms Decreasing the Effectiveness of the Efflux Pump Inhibitor D13-9001 in *Pseudomonas aeruginosa* PAO1 (abbrv.) — Jan 2019
Srijan Ranjitkar *et al.*
10.1128/aac.01718-18

Conference proceedings

Talks:

Human Immunology: Genes and Environment (Wellcome) — May 2024
Leveraging pleiotropy to interrogate the common-variant architecture of the most common inborn error of immunity

Young Statisticians Meeting (Royal Statistical Society) — Aug 2022
Accurate detection of genetic sharing between rare and common diseases enables more powerful association discovery in the rare disease context

Symposium on Biological and Life Sciences (Babraham Institute, Cambridge, UK) — Nov 2021
Unravelling the genetics of primary immunodeficiency through genomic pleiotropy

Posters:

Genomics of Rare Diseases (Wellcome) — Apr 2023
GWAS meta-analysis and the conditional false discovery rate overcome limited power in a common variant study of rare immunodeficiency

31st Annual Meeting of the International Genetic Epidemiology Society — Sep 2022
Accurate detection of genetic sharing between rare and common diseases enables more powerful association discovery in the rare disease context

50th European Mathematical Genetics Meeting — Apr 2022
Leveraging genetic pleiotropy for more powerful association discovery in the rare disease context

Scholarships and awards

Amgen Scholarship Funding for summer research internship at the University of Cambridge as part of the European Amgen Scholars programme — 2018

The Wren, Hutchinson, and Cook Prizes Awarded for achieving the best academic performance in each year of my BSc — 2016, 2017, 2019

Faculty of Engineering Excellence Scholarship Awarded for ongoing academic performance — 2015-2019