AI4Bi

One Health Genomics

THE UNIVERSITY OF EDINBURGH

# The Edinburgh Symposium on AI for Genomics

3 June 2025

# Event agenda

| | |
|---|---|
| 9:00-9:30 | **Registration** |
| 9:30-9:40 | **Welcome** |
| 9:40-10:40 | **Session 1**<br>James Prendegast (Personal Chair of Bioinformatics, Roslin Institute)<br>Linus Schumacher (Reader, Centre for Regenerative Medicine)<br>Christopher Aldous Oldnall (Doctoral Researcher, School of Mathematics/IGC) |
| 10:40-11:15 | **Keynote Speaker:** Sarah Ennis, University of Southampton |
| 11:15-11:45 | **Break** |
| 11:45-13:00 | **Session 2**<br>Athina Spiliopoulou (Chancellor's Fellow, Centre for Population Health Sciences, Usher Institute)<br>Mario Herrero Gonzalez (Doctoral Researcher, School of Informatics)<br>Lucas Guirardel (Doctoral Researcher, School of Informatics)<br>Jano van Hemert (Director of Data Services, EPCC) |
| 13:00-14:00 | **Lunch & Poster Session** |

| | |
|---|---|
| 14:00-14:35 | **Keynote Speaker**: James Dickinson, NHS Genomic AI Network |
| 14:35-15:15 | **Session 3**<br>Andrii Iakovliev (Cross-Disciplinary Post-Doctoral Fellow, MRC University Unit for Human Genetics)<br>Ekaterina Noskova (Postdoctoral Research Associate, Institute of Ecology and Evolution) |
| 15:15-15:45 | **Break** |
| 15:45-16:25 | **Session 4**<br>Ajitha Rajan (Personal Chair of Software Testing and Verification, School of Informatics)<br>Rachael Harkness (Machine Learning Engineer, Centre for Inflammation Research) |
| 16:25-17:00 | **Keynote Speaker:** Sarah Cunningham-Burley, Centre for Biomedicine, Self & Society, University of Edinburgh |
| 17:00-17:30 | **Expert Panel & Closing Remarks** |
| 17:30-18:30 | **Drinks Reception** |

# Keynote Speakers

## Using algorithms and AI to make genomic data work better for patients

Genomic data is vast and increasingly sparse. This problem grows as we sequence more and more patients. Herein, Sarah will describe an algorithmic tool that reduces the sparsity of genomic data to provide a gene-level pathogenic score (GenePy) - for all individuals for all genes. Sarah will describe how this tool has been effective in: 1) identifying missed rare disease diagnoses and 2) detecting critical genes impacting quantitative traits.

Data will further demonstrate how integration of GenePy scores with rich, standardised clinical data (digitally extracted using large language models), can be used to extract monogenic diagnoses in patients diagnosed with common diseases at scale.

**Sarah Ennis**

Professor of Genomics,

University of Southampton

Genomic Informatics

Research Laboratory

With ~200 published research papers, Sarah's research focuses on the analysis of genomic data for the purposes of (novel) disease gene detection, genomic diagnostics, prognostics and prediction; and the development, testing and application of methods to optimise genomic data modelling. She is Chief Investigator of NIHR studies recruiting patients and extracting high-integrity longitudinal clinical data from electronic patient records. Research projects use these local data in addition to Genomics England and UKBiobank datasets. Sarah is a Research Director within the Central & South Genomics Medicine Service and a project-lead within the Genomic AI Network of Excellence. Across these roles, she works closely with BRCs and SDEs to reduce barriers to increase the value of genomic data to patients and public.

## Applying AI in Genomics across the clinical pathway

As healthcare systems come under increased pressure to deliver 'more with less', AI is increasingly seen as a magic bullet that can help support cost savings while improving the patient experience. The NHS's Genomic AI Network has been set up to accelerate the adoption of AI technologies across Genomics in England, and in this talk we will explore some of the practical use cases we are supporting, as well as barriers to adoption and the current limitations of AI in clinical settings.

**James Dickinson**

Director of the NHS Genomic AI Network (GAIN)

GAIN is a national community working to accelerate the safe and responsible adoption of Artificial Intelligence within Genomic Medicine in England. James has extensive experience in the delivery of AI and other technologies in support of healthcare improvement. From working on one of the earliest examples of automated breast cancer detection while at University in the early 2000s, through an extended spell working for Palantir, where James set up the company's first health & social care project in 2014, and latterly working on the Federated Data Platform with NHS England. Most recently prior to GAIN James was contracted to lead work on Generative AI deployment at the NHS Business Service Authority.

# AI and Genomics: persistent and emergent social and ethical issues

AI and genomics are becoming more entwined, generating new social and ethical issues as well as amplifying more familiar ones, such as privacy, agency and power. This presentation will examine what we know from work on data ethics, genomics and ethics, and AI ethics, and then consider some of the most pressing issues that we need to attend to as these technologies develop and impact on society. This will involve reflecting from the perspective of individuals and systems, as well as the wider political economy, particularly with respect to health and health care. Each level of consideration is required to ensure ethically robust and socially beneficial research and innovation. Wider publics have a role to play in these debates and in shaping the future of AI and genomics and related governance and regulation.

**Sarah Cunningham-Burley**
Professor of Medical and Family Sociology & Co-Head of the Centre for Biomedicine, Self and Society at the Usher Institute, University of Edinburgh

The Centre for Biomedicine, Self and Society is an interdisciplinary social science and humanities centre focusing on social and ethical issues of developments in health-related biomedical science and health care. Sarah's own sociological research interests span personalized (genomic) medicine, AI and data science, and the integration of patient and public perspectives in research and policy.

Sarah is also Chair of the Nuffield Council on Bioethics, a leading independent policy and research centre, and the foremost bioethics body in the UK. The NCOB aims to put ethics at the centre of decisions about biomedicine and health. Recent reports include 'Predicting: The Future of Health' - a joint project with the Ada Lovelace Institute.

# Speakers

**Lucas Guirardel**

Doctoral Researcher, School of Informatics, University of Edinburgh

## Representation learning for variant effect prediction: learning functional representations from single-cell transcriptomics

High-throughput single-cell experiments can be used to predict the phenotypic effects of disease gene variants. We develop a method based on deep supervized representation learning to extract phenotypic information from scRNA-seq experiments.

**Rachael Harkness**

Machine Learning Engineer, Centre for Inflammation Research, University of Edinburgh

## Learning interpretable biological features by decomposing DNA language models

DNA language models are allowing us to read the genome in entirely new ways, combining the power of AI with the fundamental code of life. By systematically decomposing these models and analysing their internal representations, we aim to uncover interpretable biological features that could reveal both fundamental mechanisms of gene regulation and clinically relevant patterns in disease-associated variants.

**Andrii Iakovliev**

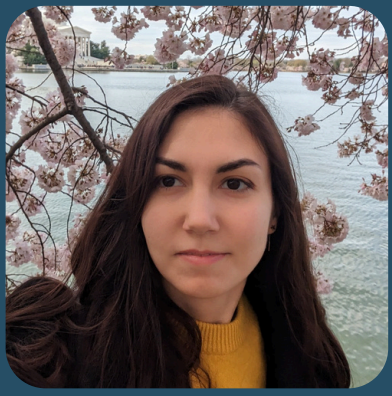
Cross-Disciplinary Post-Doctoral Fellow, MRC University Unit for Human Genetics

## Discovery and prioritisation of core genes as putative drug targets for polygenic diseases.

Approximately 90% of drugs targeting polygenic diseases fail in clinical trials, largely due to challenges in identifying causal and biologically relevant targets. In contrast, rare diseases often have clear genetic causes, making drug target validation more successful. To bridge this gap, we developed a method called GATE (Genome-wide Aggregated Trans Effects). Unlike traditional approaches, GATE models the influence of common genetic variants on distant gene expression or protein levels—known as trans effects. Without reconstructing entire gene networks, GATE identifies a sparse set of "core" genes that are critical in disease pathways. When applied to various complex diseases, GATE reveals 10–20 core genes per condition. Remarkably, many of these are also Mendelian genes in rare monogenic forms of the same diseases, providing strong evidence of causality and promising new drug targets.

**Ekaterina Noskova**

Postdoctoral Research Associate, Institute of Ecology and Evolution, University of Edinburgh



## Uncovering Non-Identifiable Demographic Histories with Generative AI

Demographic history describes the past population size changes, splits, and migrations, shaping genetic variation observed today. Demographic inference is often complicated by model non-identifiability, where different histories produce indistinguishable genetic patterns. In this talk, I will present how generative AI can be used to address this problem by learning the space of plausible histories and distinguishing between models that appear identical under existing methods. I will outline my approach and demonstrate how generative models can improve the resolution of demographic inference.

**Ajitha Rajan**

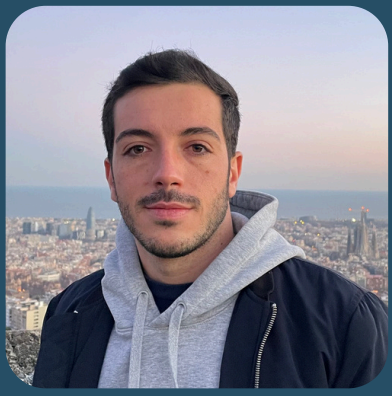
Personal Chair of Software Testing and Verification, University of Edinburgh

## Navigating the challenge of Accuracy versus Interpretability for Deep Learning-based Predictors over Omics sequences.

Over the past five years, deep learning has had a dramatic impact in predicting protein functions, protein design, immune responses, understanding complex relationships in DNA and RNA sequence data with foundation models. While these models demonstrate impressive accuracy, their complexity often reduces interpretability—crucial for clinical trust and application. This talk explores these challenges and potential solutions, focusing on predicting immune responses through antigen presentation on MHC Class I molecules and analyzing cancer survival using RNA sequence data.

**Mario Herrero Gonzalez**

Doctoral Researcher, School of Informatics



## Reconstructing Single-Cell Expression Distributions: Exploring Quantum Advantage

Small quantum processors can learn patterns in single-cell RNA sequencing data, focusing on cell-cycle genes. Our method aligns a quantum circuit's output distribution with the single-cell distribution and samples activation states. With more genes and correlations, we measure how resource requirements scale for quantum versus classical sampling. Although quantum devices exploit superposition and entanglement, they remain constrained by coherence times and noise. This raises two questions: can quantum circuits represent high-dimensional gene distributions more compactly than classical models, and how challenging is parameter optimization to uncover biological structure? Moreover, many problems are efficiently simulable on classical hardware, obscuring the threshold of quantum advantage. By benchmarking resource use, we pinpoint where quantum generative models outperform classical approaches, highlighting both potential benefits and practical limitations for transcriptomic analysis.

**Christopher Aldous Oldnall**

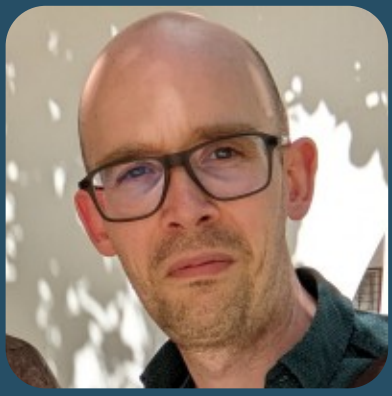
Doctoral Researcher, School of Mathematics

## Identifying causal proteomic targets for cardiovascular disease via robust Mendelian randomisation

Mendelian randomisation (MR), a biological application of the instrumental variable framework, is a powerful tool for identifying causal relationships in biomedical research. By using genetic variants such as single nucleotide polymorphisms, MR bridges genome-wide association studies and translational medicine, enabling the discovery of causal proteins in diseases like cardiovascular disease (CVD), however pleiotropy—violations of the exclusion restriction criterion—remains a challenge. Newer novel pleiotropically robust estimators leverage potentially invalid instruments against valid ones, offering a systematic, assumption-light approach to handling pleiotropy. Using proteomic data from the UK Biobank, we identify causal proteins across multiple CVD risk traits, including hypertension and diabetes. We evaluate the overlap and differences between proteins identified by more traditional MR methods and the pleiotropy robust estimators, critically assessing their consistency and robustness.

**James Prendergast**

Personal Chair of Bioinformatics, Roslin Institute, University of Edinburgh



## AI-Driven Prediction of Mammalian Functional Regulatory Variation in Non-Model Species

Pinpointing functional genomic variation is essential to understand how genetic differences drive important traits and diseases. While AI-based methods show promise in predicting functional regulatory variants in humans, their application to non-model mammals has been limited by limited training datasets. To address this in livestock, we generated powerful datasets for training AI models. This includes genome-wide massively parallel reporter assays (MPRA) testing the regulatory impact of over 13 million variants, as well as PRO-Cap data mapping transcription initiation at base-pair resolution. Using these data, we are training novel AI models to predict the functional impact of genomic variants in cattle. Our results show how these tools can fine-map GWAS loci, improve breeding values, and identify targets for genome editing—ultimately enhancing productivity and resilience in livestock.

**Linus Schumacher**

Reader, Centre for Regenerative Medicine, University of Edinburgh

## Quantifying tissue state dynamics from spatial protein data

The dynamics of a biological tissue arises from the behaviour of its constituent cells and their interactions. Phenomena such as regeneration, over-proliferation, patterning, inflammation, and scarring are defined at the level of cell populations, not individual cells. In this talk I will present data-driven approaches to quantify and model dynamic tissue states applied to triple-negative breast cancer biopsies as well as in vitro models of embryo development.

**Athina Spiliopoulou**

Chancellor's Fellow, Centre for Population Health Sciences, Usher Institute

## Mendelian randomization with multi-SNP genetic instruments and multiple exposures

Mendelian randomization (MR) analyses study the causal effects of modifiable exposures on health outcomes using genetic variants (typically single nucleotide polymorphisms; SNPs) as instrumental variables. Current methods for Mendelian randomization (MR) are restricted to using a single SNP from each genetic region associated with the exposure. I will describe our recent work to overcome this limitation and construct a scalar instrumental variable for each region using all SNPs associated with the exposure. I will also describe methods for joint modelling of the effects of these multi-SNP instruments on multiple exposures to distinguish causal from direct (pleiotropic) effects on the outcome by allowing for sharing of pleiotropy between overlapping instruments.

**Jano van Hemert**

Director of Data Services,

Edinburgh Parallel

Computing Centre (EPCC)

## Edinburgh International Data Facility: data and compute services for large data science and AI research

We will provide an overview of the Edinburgh International Data Facility's services. All these services are optimised for running large data science and AI training tasks. For instance, EIDF has hundreds of GPU compute nodes and 40 petabytes of storage. We explain the access model for researchers and PhD students.

# Organising Committee

**Tim Aitman**
Director, Centre for Genomic & Experimental Medicine

**Maria Delgado Ortet**
Cross-disciplinary Fellow, Institute of Genetics and Cancer

**Ava Khamseh**
Lecturer in Biomedical Artificial Intelligence

**Rashi Krishna**
PhD Student, Institute of Genetics and Cancer

**Daniel Toddie-Moore**
Scientific Research and Funding Manager, Institute of Genetics and Cancer

**Ian Simpson**
Director, AI Centre for Doctoral Training in Biomedical Innovation

**Ekaterina Churkina**
Manager, AI Centre for Doctoral Training in Biomedical Innovation

**Lola Springbett**
Administrator, AI Centre for Doctoral Training in Biomedical Innovation

# Connect with us

ai4bicdt@ed.ac.uk

genomics_network@ed.ac.uk

@UKRI AI Centre for Doctoral Training in Biomedical Innovation

@One Health Genomics Edinburgh

AI4Bi

One Health Genomics

THE UNIVERSITY of EDINBURGH