# Gene Program Modelling to Decode Preleukemic Stem Cell Fitness

**Professor Bertie Gottgens** https://www.haem.cam.ac.uk/staff/bertie_gottgens

**Professor Nicola Wilson** https://www.stemcells.cam.ac.uk/directory/nicola-wilson

Single-cell transcriptomics has transformed our understanding of cellular diversity, but translating this complexity into mechanistic insight remains a major challenge. Traditional approaches often reduce rich cellular states to single-vector representations, limiting interpretability and obscuring key regulatory mechanisms.

This project builds on Tripso, a machine learning framework developed by the Lotfollahi group in collaboration with the Gottgens lab. Tripso uses gene program (GP)-specific transformers to learn multiple interpretable representations of gene activity at the single-cell level. Applied to over 500,000 haematopoietic cells, Tripso has revealed GP usage patterns linked to cell fate decisions and stem cell fitness.

Preleukaemic mutations are thought to modulate the fitness of blood stem cells. The student will work closely with computational researchers to identify candidate gene programs and regulatory targets associated with preleukemic stem cells. These predictions will be tested using CRISPR-based perturbations, pharmacological interventions, and flow cytometry. Molecular characterisation will assess functional outcomes and validate the biological relevance of computational findings.

This interdisciplinary project bridges computational modelling and experimental haematology. The student will receive training in single-cell transcriptomics, machine learning (including transformer architectures and optimal transport theory), and experimental techniques. They will learn to generate hypotheses from large-scale data and design experiments to test gene program function across developmental and preleukemic contexts.

Ultimately, the project aims to uncover mechanisms driving stem cell fitness and lineage commitment in preleukemia, with implications for early detection and therapeutic intervention.