Clinical PhD Project 2024

# Combining single-cell genomics and machine learning to understand adaptive immune responses in early cancer.

Principal supervisor's name: Dr **Jamie Blundell**
Principal supervisor's email address: jrb75@cam.ac.uk

Principal supervisor's CRUK CC theme: Early Cancer Institute

Department for student registration: Oncology
Department or institute where research will take place: Early Cancer Institute

Co-supervisor's name (if applicable): Caroline Watson
Co-supervisor's email address (if applicable): cw672@cam.ac.uk

Postgraduate scheme: **Clinical Research Training Fellow (3-year PhD)**

## PhD project outline:

In the PhD component of the project the student will perform ultra deep targeted DNA sequencing of serial blood samples collected every 6 weeks from 20 healthy volunteers over a period of 18 months, as part of the Legacy study (https://www.legacy-study.org/). Using these data clones harbouring somatically acquired driver mutations that have expanded to >1% variant allele frequency in the blood will be identified and computational tools (netMHC) will be . To determine whether there are adaptive immune responses to these mutant clones, the student will use single-cell RNA sequencing of T-cells combined with a barcoded tetramer technology (BEAM-T, 10x genomics) to determine the which TCR sequences are responsible for recognising the neoantigens harboured by the mutant clones. Using this data the student will then be encouraged to build machine learning models that are capable of extracting the common features of VDJ sequences that indicate whether they are specific for a particular precancerous neoantigen and determine whether immune recognition of pre-cancerous neoantigens influences the growth of the mutant clone over time (measured via the trajectory of the mutant clone in blood).

## PhD experimental plan:

Duplex sequencing of serial blood samples from 20-healthy volunteers to identify mutant clones
Computational prediction of neoantigens from mutations using netMHC and associated software.
scRNA-sequencing of T-cells that have been strained with barcoded tetramers to map TCR-pMHC interactions for the neoantigens identified in (2)
Assessment of mutant clone trajectories from serial bloods.

## Main techniques:

Single-cell RNA sequencing
Tetramer straining and T-cell flow sorting
Duplex DNA sequencing of serial blood samples
Computational peptide prediction software

Computational identification of somatic variants
TCR feature extraction and machine learning.

**Key references:**

Watson, C.J. et al. (2020) 'The evolutionary dynamics and fitness landscape of clonal hematopoiesis', Science, 367(6485), pp. 1449–1454. Available at: https://doi.org/10.1126/science.aay9333.

Poon, G.Y.P. et al. (2021) 'Synonymous mutations reveal genome-wide levels of positive selection in healthy tissues', Nature Genetics, 53(11), pp. 1597–1605. Available at: https://doi.org/10.1038/s41588-021-00957-1.
https://www.10xgenomics.com/blog/introducing-beam-barcode-enabled-antigen-mapping-benefits-of-rapid-antigen-specific-b-and-t-cell-discovery

Ott, P.A. et al. (2017) 'An immunogenic personal neoantigen vaccine for patients with melanoma', Nature, 547(7662), pp. 217–221. Available at: https://doi.org/10.1038/nature22991.

Saini, S.K. et al. (2022) 'Neoantigen reactive T cells correlate with the low mutational burden in hematological malignancies', Leukemia, 36(11), pp. 2734–2738. Available at: https://doi.org/10.1038/s41375-022-01705-y.

Rojas, L.A. et al. (2023) 'Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer', Nature, 618(7963), pp. 144–150. Available at: https://doi.org/10.1038/s41586-023-06063-y.