The computational biology group is interested in how the processes that control gene expression are altered in tumour cells, how these changes occur, and how they drive oncogenic transformation and tumour progression. We are studying these systems by using classical- and deep- machine learning approaches to study multiomics datasets arising from clinical and in vitro studies.

While considerable attention has been directed at the regulation of transcription, many of the downstream processes such as the control of RNA processing, splicing, and mRNA stability are also under tight regulatory control. The translational machinery that governs when, and how these mature mRNAs are translated into correctly folded proteins is similarly constrained. A critical question, therefore, is how is the information that defines these systems encoded within the genome?

Our work exploits the availability of a large and diverse cohort of well annotated genome sequences from different species. This allows comparative genomics to be used to pursue regulatory patterns from an evolutionary perspective. In parallel, the availability of large cohorts of DNA- and RNA-sequenced patient tumour samples makes it possible to explore the evolutionary constraints placed upon different regions of the genome by selection pressure from within the tumour environment. In both cases, the available data are now at sufficient scale to support classical- and neural-network based machine learning algorithms, and we are applying these in combination with mathematical models that draw upon ideas from information theory.

Tamara Luck, a postdoc in the group is interested in regulatory sequences embedded within coding sequences, and how mutations in and around these regulatory sites can impact on protein levels. Boyu Yu, a new graduate student, co-supervised with the RNA and Translational Control in Cancer Group, led by Martin Bushell, is investigating the regulatory sequences embedded in the untranslated regions of protein coding genes, and how these sequences are used by cells to regulate mRNA stability and protein translation.

We are also part of PREDICT-Meso, a £5m Accelerator project funded through a partnership between CRUK, Fondazione AIRC, and Fundación Científica de la Asociacion Española Contra el Cancer (FC AECC). Mesothelioma is an incurable cancer that typically develops years after inhalation of asbestos dust and fibres. The factors that underpin the development of mesothelioma are currently poorly understood. Holly Hall recently joined the lab as a postdoc to develop computational models of mesothelioma `omics data (Tsim et al. 2021).

Underpinning all these algorithms is a requirement to perform computationally intense calculations across thousands of genome sequences with matched transcriptome and proteomics data. Over the last year, we have been working with the Information Services team to expand the High-Performance Computing infrastructure that will underpin our data science efforts across the Institute.

Publications listed on page 110

An osteosarcoma cell (U2OS) was stained with markers for the cell recycling and degradation centre, the lysosome (LAMP-2, yellow), the actin cytoskeleton (Blue, Phalloidin) and the nucleus (Grey, DAPI/DNA). The cell was imaged on the Zeiss 880 super-resolution microscope at the Beatson Institute and processed using FIJI (ImageJ) software.

Image by David McEwan