

COMPUTATIONAL BIOLOGY



Group Leader
Crispin Miller

Research Scientists
Eva Freckmann
Holly Hall

Graduate Students
Britt van Abeelen
Boyu Yu

Rapid advances in technology are leading to a wealth of high-dimensional data describing the behaviour of cells in normal and tumour tissue. We are using computational approaches to interrogate and integrate these high dimensional data in order to develop a more holistic view of the altered regulatory processes that lead to the development and progression of cancer.

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?

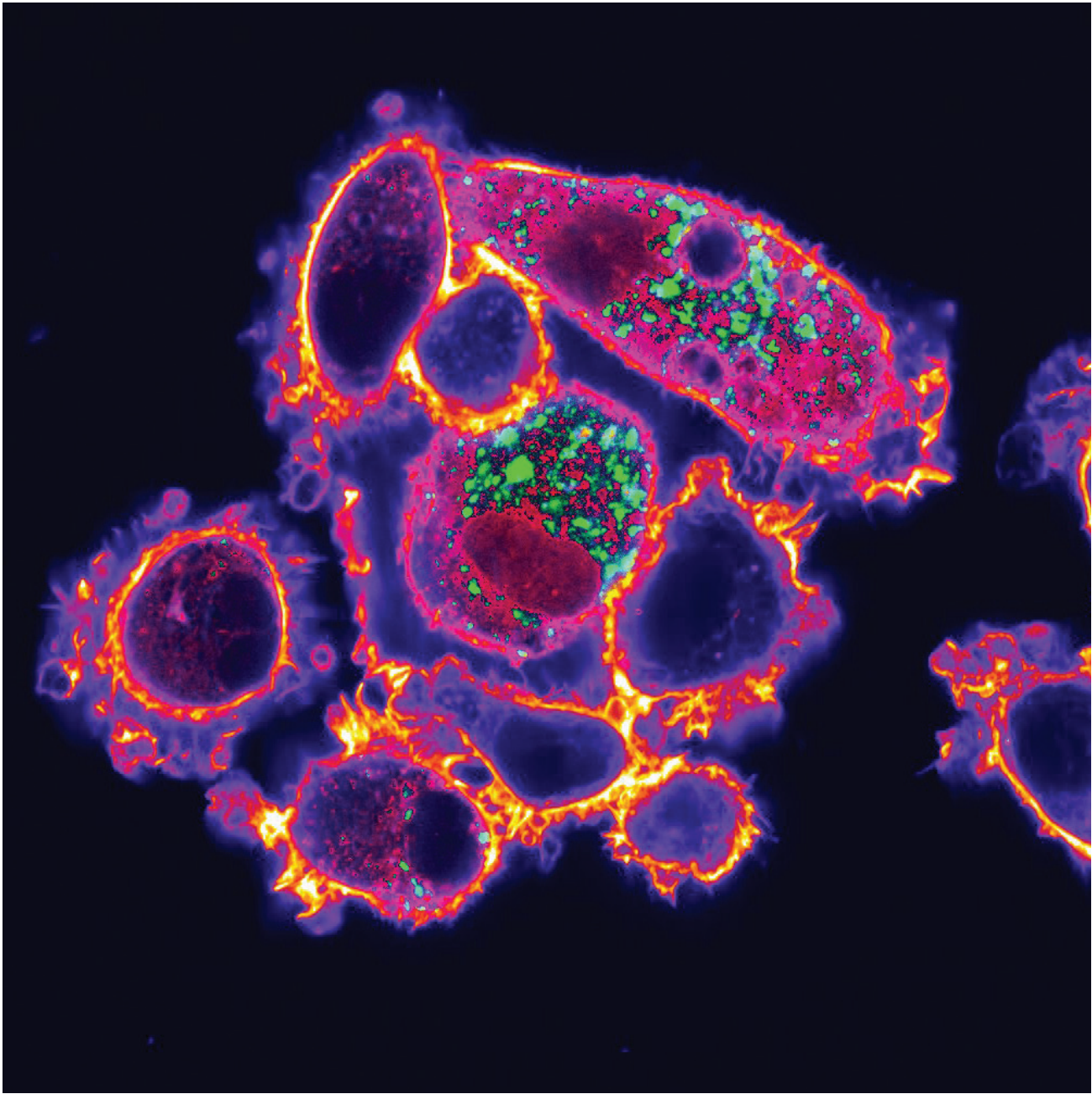
Our work exploits the availability of a large and diverse cohort of well annotated genome sequences from different species. This allows comparative genomics to be used to pursue regulatory patterns from an evolutionary perspective. In parallel, the availability of large cohorts of DNA- and RNA-sequenced 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.

Eva Freckmann a postdoc in the group is interested in regulatory sequences embedded within coding sequences, and how mutations and changes in the regulatory machinery in and around these regions can impact on protein levels. Boyu Yu, a 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 Asociación Española Contra el Cáncer (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 a postdoc in the lab is applying computational approaches to study 'omics data arising from multiple tumour types including mesothelioma, colorectal and liver cancer samples.

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 Naveed Khan to commission a High-Performance Computing system that is starting to underpin our data science efforts across the Institute.

[Publications listed on page 107](#)



PC3 prostate cancer cell line plated in 2D, fixed and then stained with Alexa Fluor 568 Phalloidin (F-actin, shown in FIRE LUT) and anti-ARF3 (ARF GTPase shown in green).

Image by Emma Sandilands