

BIOINFORMATICS AND COMPUTATIONAL BIOLOGY



Head

Crispin Miller

Scientific Computing Specialist
Naveed Khan

Bioinformaticians
Ryan Kwan
Robin Shaw

The Bioinformatics unit provides numerical expertise across the Institute. A major aspect of our work continues to centre on the analysis of high throughput 'omics data, including a wide range of next generation sequencing, proteomics, and metabolomics datasets. A growing focus has been on the application of these techniques to imaging data.

Our team focuses on exploratory data analysis, and our ultimate goal is to provide insights that enhance our understanding of cancer biology. The need for DNA and RNA sequencing analyses has continued to grow, and this has been accompanied by continued interest in using computational and machine learning approaches to interpret imaging and proteomics data. A major aspect of our work continues to be the analysis of single cell sequencing data and we have been developing workflows that use a mixture of specific packages, such as Seurat, along with other software tools and packages from the Bioconductor project.

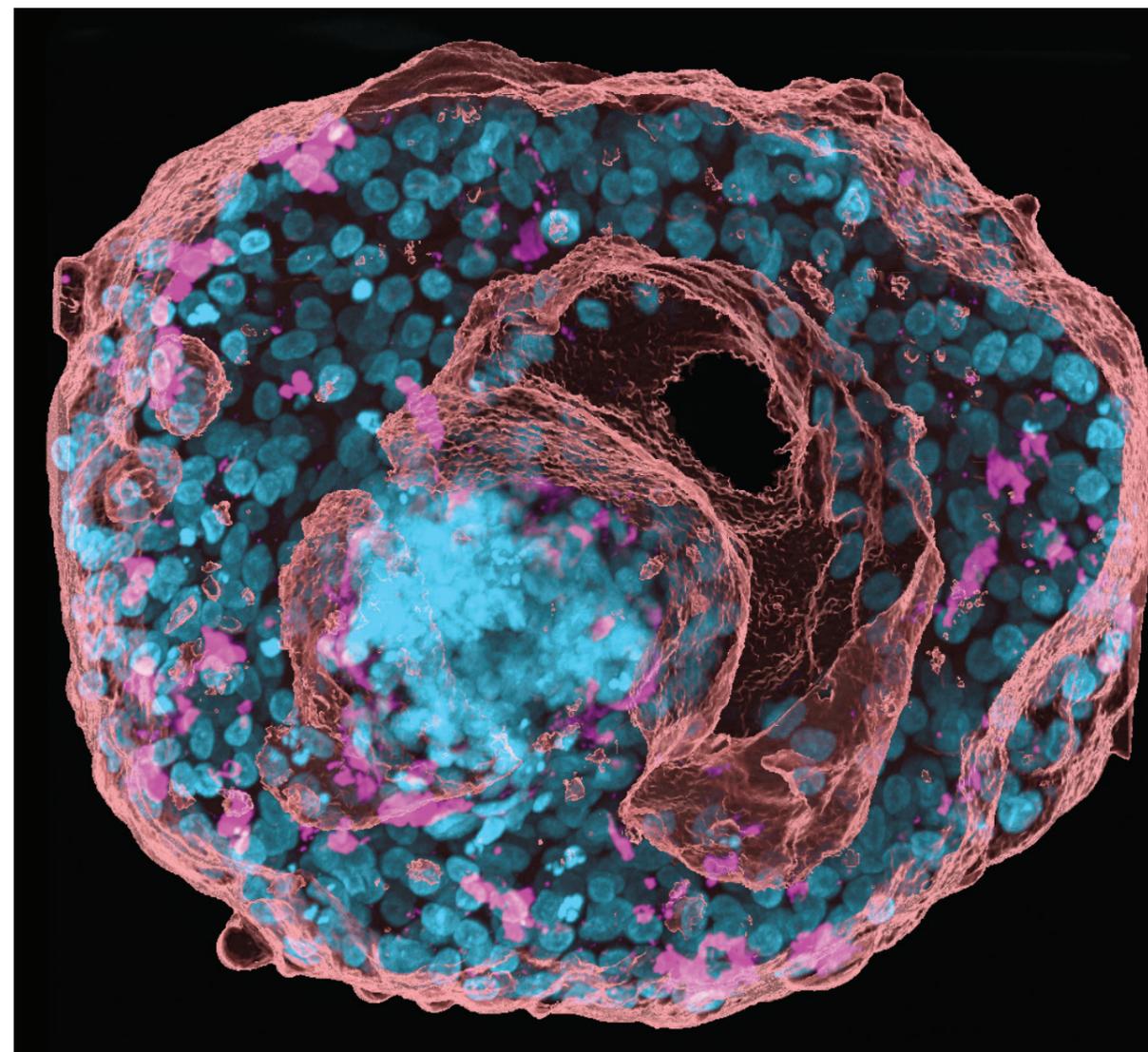
Advances in technology are leading to a rapid increase in the size of the data we are analysing, leading to significant increase in our computing requirements. Naveed Khan joined our team to lead on establishing a scientific computing platform. Naveed has begun to commission a High Performance Computing (HPC) system that combines conventional processing with GPUs and a fast filesystem in order to support our data science, AI and deep learning needs. Naveed is also working closely with IT Services on the provision of Virtual Machines (VMs) to support non-HPC tasks. Robin Shaw and Naveed are also working together with IT services to standardise data processing workflows through our computing platforms and the different filesystems within the Institute.

Data analysis and modelling is performed using a variety of open-source software environments, programming languages and scripting tools,

including Python, R, Bioconductor, Bash, PHP and Perl. We frequently make use of analytical routines that have been developed in-house, and/or in collaboration with our colleagues from the areas of mathematics, statistics, computer science and biology. We use a mixture of academic software tools for functional annotation, clustering, enrichment, ontology and pathway analysis, as well as commercial tools.

The unit also provides support and guidance to graduate students and postdocs in other research groups who are using computational approaches to analyse their data. This includes advice on R scripting (by appointment), experimental design, and data presentation. We have introduced an experimental design course and continue to operate a regular internal bioinformatics forum to provide a central point of contact to bring together bioinformaticians, researchers and students who are applying computational biology and numerical approaches to their data. Our team also participates in delivering part of the Cancer Research & Precision Oncology MSc programme at the University of Glasgow.

Publications listed on page 116



Pancreatic adenocarcinoma (PDAC) is an aggressive cancer that often metastasises to areas such as the liver. To study this metastasis, we are developing a 3D in vitro model using a co-culture system. This image shows a spheroid generated by the aggregation and culture of multiple liver cell types and PDAC cells. To study potential modelling of the extracellular matrix, the secretion of collagen 1 within the spheroid was visualised by immunofluorescence. The spheroid was also stained with DAPI and Alex Fluor 568 Phalloidin, and a z stack image was taken using a Nikon A1R confocal microscope in the BAIR facility at the CRUK Beatson Institute.

Image taken by Elaine Ma