The Bioinformatics and Computational Biology 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. Our remit also includes mathematical modelling, and 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 analysis has continued to grow, and this has been accompanied by an increasing interest in using computational and machine learning approaches to interpret imaging and proteomics data. In order to meet these demands, another bioinformatician (Robin Shaw) joined our team in July 2018. The last year has also seen a growth in the use of machine learning approaches to help characterise our data, and we have seen an increasing demand to apply quantitative approaches to the analysis of microscopy and imaging data generated by the Beatson Advanced Imaging Resource (BAIR; pp 66).

Data analysis and modelling is performed using a variety of open-source software environments, programming languages and scripting tools, including R, Bioconductor, KNIME, Fortran, 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, including Oncomine Research Premium Edition, Ingenuity Pathway Analysis and GeneGo MetaCore.

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 established a bi-weekly 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 postgraduate Cancer Sciences MSc programme.

The proteomics team has an outstanding expertise in high-resolution Orbitrap-based mass spectrometry (MS) proteomics, accurate quantification approaches and MS data analysis. We work in collaboration with research groups at the Institute and outside, and we actively develop MS-based proteomic platforms to address a variety of questions to help scientists better understand the mechanisms that regulate various aspects of cancer.

To achieve this, we are well equipped with three nano-liquid chromatography (nLC)-MS systems. This year we have installed the newest generation Orbitrap instrumentation, Fusion-Lumos; we also have a Q-Exactive HF and an Orbitrap Elite. All our instruments are coupled online to EASY-nLC systems, and high-resolution chromatography is achieved by packing our nano-columns in-house.

We house a number of dedicated pieces of software, of which MaxQuant is the most used, for highly accurate label-free or label-based quantitative analysis. Moreover, we use Skyline for the analysis of RM data. Finally, we use Perseus and Scaffold for data analysis and dissemination.

This year we have recruited Kelly Hodge as a new senior scientific officer, who is expert in sample preparation for proteomic analysis and MS data analysis. We have also recruited our first bioinformatician, Greg Koulouras, who is expert in web and Android programming and in working with big data and sensitive personal information. With Greg, we are currently working at an ambitious project of developing a cancer-centric database for the navigation through and analysis of MS–proteomic data.

During 2018, we have worked with many of the groups at the Institute and significantly contributed to the success of their research. Following the development of our novel stable isotope-based method, SICyLIA, for the global measurement of cysteine oxidation we are now further developing this approach to enable dynamic measurement of cysteine oxidation and to achieve higher depth of our analyses in a shorter time.

We are continuously striving to develop methods to answer more complex biological questions using proteomics and improve the methods currently in place to enrich the quality of the data that the facility can provide. We have now set up tandem mass tag (TMT)-based labeling approaches to multiplex up to ten samples in one. Moreover, we have been working to improve the depth of the proteomes and sub-proteomes achieved starting with small amounts of sample, e.g. FACS-sorted immune cells, using high-pH reverse-phase LC Fractionation. This enables us to perform improved global proteomic, sub-proteomic and post-translational modification analyses of primary cells, circulating blood cells and 3D organotypic cell cultures.

Publications listed on page 95