



## Bioinformatics

[www.cambridgecancer.org.uk/research/coreresources/bioinformatics](http://www.cambridgecancer.org.uk/research/coreresources/bioinformatics)

### Head **Matthew Eldridge**

#### Senior Computational Biology Analysts

Sarah Vowler

Kevin Howe

Rory Stark

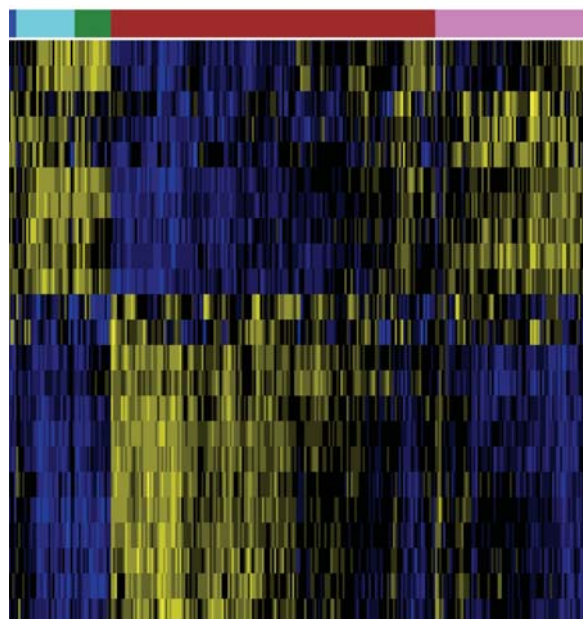
The Bioinformatics core is committed to remaining at, and pushing forward, the cutting edge of high-throughput genomic, epigenomic and proteomic analysis.

The research groups and core facilities, particularly the Genomics core, generate a tremendous amount of biological data. The Bioinformatics core facility focuses on the processing and analysis of these data, providing storage, experimental design advice, quality control, and access to standard software tools. In addition, analysts are able to conduct customised downstream analysis of specialised data, including clustering, classification, and identification of regulatory networks.

To date the Bioinformatics core has focused primarily on the analysis of data generated from different microarray technologies, including Affymetrix, Agilent, NimbleGen and Illumina. The analyses of these data has been driven by a number of research goals, including the discovery of differentially expressed genes, identification of prognostic signatures, copy number variation, splice variants, SNPs and microRNAs, as well as epigenomic analysis using ChIP-chip and other assays.

During the final quarter of the year, attention was focused on providing a processing pipeline and analysis facilities for a new sequencing-by-synthesis platform from Illumina (using Solexa technology). This high-throughput instrument generates tens of millions of short sequence reads every few days, creating a new challenge in keeping up with the terabytes of data that will be produced over the next year. The Bioinformatics core facility provides the pipeline for processing sequence reads, as well as performing downstream analysis such as sequence alignment to reference genomes. Experimental paradigms include ChIP-seq, RIP-seq, small RNA discovery, transcriptome analysis, and deep re-sequencing for SNP discovery.

During 2007, the Bioinformatics core laid the foundations that will allow a variety of services to be provided to the CRI research community by recruiting a Head and an initial team of analysts. As well as the various analyses described above, members of the team assisted with the specification and planning of a high-performance computing cluster that will be dedicated largely to the Bioinformatics Core and the Statistics and Computational Biology research group. Planning has also begun to support the integration of disparate high-throughput genomic data with other clinical information. This includes maintaining and enabling the use of a flexible database storage solution integrated with local implementations of genomic databases, and providing web-based sequence browsing tools such as Ensembl.



A heatmap showing the relative activation levels for key gene pathways in five subtypes of ER-breast cancer tumours, with bright yellow being indicating highly active pathways and bright blue indicating highly suppressed pathways.