Biomedical data science school - programme

Monday 11.03.19

9-9.30 Registration

9.30-11 Lecture, Prof Magnus Rattray, University of Manchester
Modelling gene expression dynamics with Gaussian process inference

Gaussian process (GP) inference provides a flexible non-parametric probabilistic modelling framework. We present examples of GP inference applied to time-series gene expression data and for single-cell high-dimensional "snapshot" expression data. We provide a brief overview of GP inference and show how GPs can be used to identify dynamic genes, infer degradation rates, model replicated and clustered time-series, model stochastic single-cell dynamics, and model perturbations or branching in time-series data. In the case of single-cell expression data we present a scalable implementation of the Gaussian Process Latent Variable Model, which can be used for dimensionality reduction and pseudo-time inference from single-cell RNA-Seq data. We also present a recent approach to inference of branching dynamics in single-cell data. To scale up inference in these applications we use sparse variational Bayesian inference algorithms to deal with large matrix inversions and intractable likelihood functions.

11-11.30 Coffee Break

11.30- 12.30 Talk, Prof Davide Marenduzzo, School of Physics and Astronomy, University of Edinburgh
Biophysical principles of chromosome organisation

I will describe some generic principles for DNA organisation in eukaryotic cells which were recently uncovered by biophysical models. First, I will discuss the "bridging-induced attraction", which provides a universal biophysical pathway leading to the clustering of multivalent chromatin-binding proteins; we will use this principle to explain the biogenesis of nuclear bodies such as polycomb and Cayal bodies, or transcription factories. Second, I will speak about the "loop extrusion" model, which was introduced in recent years to account for the experimentally observed bias favouring convergent CTCF loops over divergent ones in mammalian genomes. I will discuss different potential mechanisms for extrusion, which may be either unidirectional or bidirectional, in view of current experiments, both in vitro and in vivo.

12.30-14.00 Lunch

14-17 Tutorial, Prof Magnus Rattray

Tuesday 12.03.19

9.30-11 Lecture, Dr Gabriele Schweikert, University of Tubingen/ University of Dundee
Making Sense of Multivariate Epigenomic Snapshots

The cellular epigenetic machinery promotes local chromatin structure and is a major determinant of gene expression. A myriad of proteins that catalyse epigenetic modifications have been discovered and loss of function experiments for key regulators have provided valuable insights in their workings. However, experimental approaches are challenged by the global, highly dynamic and intertwined action of epigenetic proteins; On the other hand, high-throughput sequencing methods have made huge numbers of steady-state epigenomic snapshots readily available. In this case, data interpretation remains challenging as key interactions cannot easily be discerned by simple univariate analysis approaches. In this lecture we will investigate Machine learning (ML) methods to extract consistent signals from high-dimensional data with high feature redundancy.

11-11.30 Coffee break

11.30-12.30 Lecture, Dr Duncan Sproul, Institute of Genetics and Molecular Medicine
Understanding dynamics in human epigenomes

Epigenetic modifications are covalent modifications of the DNA and chromatin that regulate transcription and other DNA dependent processes. Their distribution in the genome is frequently altered in disease. My work seeks to understand the molecular mechanisms responsible for these alterations, particularly focusing on the repressive mark DNA methylation. In this talk I will describe our recent work that seeks to understand the dynamics of DNA methylation in cancer cells using a variety of experimental approaches. I will also discuss how DNA methylation patterns are altered during human aging how the analysis of longitudinal DNA methylation measurements enable us to detect the influence that DNA sequence variation has on this process.

12.30-14 Lunch

14-17 Tutorial, Dr Gabriele Schweikert

Wednesday 13.03.19

9.30-12.30 Participants' workshop (programme TBA)

Free afternoon

Thursday 14.04.19

9.30-11 Lecture, Dr Andrea Sottoriva, Institute of Cancer Research, London
Measuring and predicting cancer evolution from genomic data

Cancer genomic data is often hard to make sense of in light of tumour biology due to the extensive inter-patient variation and intra-tumour heterogeneity. However, the cancer evolution paradigm poses that tumours change over time following potentially tractable evolutionary dynamics driven by the forces of mutation, genetic drift and selection. In this seminar, I will discuss how to measure the dynamics that drive tumour evolution in each individual patient using a combination of mathematical modelling and high-throughput genomic data from human cancers. I will show how the identification of patient-specific evolutionary dynamics allows for predictions on the future course of the disease. The ultimate aim is anticipating a cancer's next step, with fundamental implications for treatment optimisation and disease management.

11-11.30 Coffee Break

11.30-12.30 Talk, Dr Diego Oyarzun, School of Informatics and School of Biological Sciences, University of Edinburgh
Systems and synthetic biology of metabolism

The broad theme of our research is the study of biomolecular networks with mathematical and computational methods. In this talk I will highlight some of our work on cellular metabolism with applications in systems biology, synthetic biology and precision medicine. Specifically, I will discuss a new approach to study metabolic alterations in disease. Metabolic alterations play a key role in a number of medical conditions, including cardiovascular disease and various forms of cancer. Yet the complex connectivity of metabolism makes it difficult to establish links between the disruption of specific pathways and the onset of disease. With a combination of network theory and stochastic diffusion processes, we have developed a new formalism to interrogate metabolic connectivity. Our results reveal structural changes in metabolic connectivity between healthy and disease genotypes, and can pinpoint genes that could be used as biomarkers or drug targets in precision medicine.

12.30-14 Lunch

14-17 Tutorial, Dr Andrea Sottoriva

Friday 15.03.19

9.30-11 Lecture, Prof John Marioni, Cancer Institute, University of Cambridge
Computational challenges in single cell biology

With recent technological developments it has become possible to characterize a single cell's genome, epigenome, transcriptome and proteome. In particular, single-cell RNA-sequencing (scRNA-seq) has been widely applied to study heterogeneity in populations of neurons, in the immune system and in early development, revealing the existence of new populations of cells and differentiation trajectories. Fully exploiting such data requires the development of novel computational methods, with many of the tools developed for bulk RNA-sequencing not being appropriate for scRNA-seq. In this presentation I will describe some of the methods we have recently developed to address these challenges, and will illustrate their application in a variety of biological contexts.

11-11.30 Coffee Break
11.30-12.30 Talk, Dr Catalina Vallejos, IGMM and Alan Turing Institute
BASiCS: Robust normalisation and variability estimation for noisy single cell gene expression data.

Single-cell RNA-sequencing (scRNA-seq) has transformed the field of transcriptomics, providing novel insights that were not accessible to bulk-level experiments. However, the promise of scRNA-seq comes at the cost of higher data complexity. In particular, a prominent feature of scRNA-seq experiments is strong measurement error, reflected by *technical dropouts* and poor correlations between technical replicates. These effects must be taken into account to reveal biological findings that are not confounded by technical variation.

In this talk, I will discuss some of the challenges that arise when *normalising* scRNA-seq data — a critical step whose aim is to adjust for unwanted biological and technical effects that can mask the signal of interest. I will also introduce BASiCS (Bayesian Analysis for Single Cell Sequencing data), a Bayesian hierarchical model in which data normalization, technical noise quantification and selected downstream analyses are simultaneously performed. To disentangle biological signal from technical artefacts, BASiCS uses: *(i)* a *vertical integration* approach, where a set of technical spike-in genes is used as a *gold-standard* and *(ii)* a *horizontal integration*framework, where technical variation is quantified by borrowing information from multiple groups of samples. Beyond highlighting changes in mean expression, BASiCS extends traditional differential expression analyses to also capture changes in transcriptional variability between cell populations (e.g. experimental conditions or cell types). In particular, I will introduce a recent extension of BASiCS which enables differential variability analyses that are not confounded by changes in mean expression. I will also illustrate the performance of our methods using control experiments and case studies in the context of immune cells. Finally, I will discuss ongoing efforts to improve the scalability of our approach.

12.30-14 Lunch

14-17 Tutorial, Prof John Marioni