# Session Change-point detection

Vincent Brault

Title: ***Fast bi-dimensional segmentation for Hi-C data***

Summary:
In order to deal with new kind of genomic data such as Hic data (Rao et al., 2014), biologists aim at forming groups in rows and columns of a matrix without permutations to obtain a grid panel. However, this is a challenging issue for different reasons: the dynamic programming algorithm, typically used in one dimension, cannot be applied and the size of the studied matrix enforces to use powerful algorithms.
In fact, we show that this problem is equivalent to a linear parsimonious high dimensional problem with a fast and efficient method of least squares.
In this talk, we will show how our method provides a grid for a matrix in high dimension (10 000x10 000). We will then explain how the bi-dimensional segmentation enables to have a consistent estimator of the change-point under some conditions. Finally, our method is applied on simulated and real data.

Paul Fearnhead, Guillem Rigaill, Rob Maidstone and Adam Letchford

Title: ***Efficient Algorithms for Changepoint Detection***

Summary:
A common statistical approach to detecting changepoints is to find the best segmentation of data that minimise a penalised cost (such as residual sum of squares). In situations where the cost can be written as a sum over segments, then standard dynamic programming (DP) approaches can find the optimal segmentation with a cost that is either quadratic or cubic in the amount of data. Recently DP solutions to this problem that can have a linear computational cost have been developed, for example with the PDPA algorithm of Rigaill (2015). Here we extend the ideas of PDPA to a wider class of changepoint models. These include applications where the cost is robust to outliers, and situations where there is dependence across segments, and where standard DP algorithms currently do not apply. Finding the best segmentation for changepoint models with dependence has been claimed to be an NP-hard problem, yet we will describe a novel DP algorithm that (empirically) has an average computational cost that scales linearly or quadratically with the amount of data.
These changepoint algorithms are motivated by applications such as detecting copy-number-variation and analysing the motion of bacteria.

Merle Behr, Chris Holmes, Axel Munk

Title: ***Multiscale Inference for Blind Demixing with Applications in Cancer Genetics***

Summary:
*We discuss a new methodology for statistical recovery of single linear mixtures of piecewise constant signals (sources) with unknown mixing weights and change points in a multiscale fashion. Exact recovery within an epsilon-neighborhood of the mixture is obtained when the sources take only values in a known nite alphabet. Based on this we provide the SESAME (SEparateS-nite-Alphabet-MixturEs) estimators for the mixing weights and sources for gaussian error. We obtain uniform condence sets and optimal rates (up to log-factors) for all quantities.*
*SESAME is motivated from cancer genetics where one aims to assign copy-number variations from genetic sequencing data to dierent tumor-clones and their corresponding proportions in the tumor. We analyze such data using the proposed method in order to estimate the number of clones, their proportion in the tumor, and the corresponding copy number variations.*