

**University of California, Irvine
Statistics Seminar**

***A Hierarchical Bayesian Model for Inference of Copy Number
Variants and Their Association with Gene Expression***

**Alberto Cassese
Assistant Professor
Maastricht University**

**Thursday, February 9, 2017
4 p.m., 6011 Bren Hall
(Bldg. #314 on campus map)**

Cancer is the result of a dynamic interplay at different molecular levels (DNA, mRNA and protein). Elucidating the association between two or more of these levels would enable the identification of biological relationships and lead to improvements in cancer diagnosis and treatment. For this purpose the development of statistical methodologies able to identify these relationships is crucial. In this talk, I present a model for the integration of high-throughput data from different sources. In particular, I focus on combining transcriptomics data (gene expression profiling) with genomic data, collected on the same subjects. At DNA level I focus on measuring copy number variation (CNV) using comparative genomic hybridization (CGH) arrays. I specify a measurement error model that relates the gene expression levels to latent copy number states. Selection of relevant associations is performed employing selection priors that explicitly incorporate dependencies information across adjacent copy number states. Copy number states are related to the observed surrogate CGH measurements via a hidden Markov model, which captures their peculiar state persistence. Posterior inference is carried out through Markov chain Monte Carlo techniques. In order to tackle the computational issue, I developed an algorithm that efficiently explores the space of all possible associations. The contribution of the methodology is twofold: infer copy number variation and, simultaneously, their association with gene expression. The performance of the method is shown on simulated data and I also illustrate an application to data from a prostate cancer study.