Probabilistic Models of Splicing (Dys)Regulation in Human Disease

Abstract: Splicing, the cellular process by which "junk" intronic regions are removed from precursor messenger RNA, is tightly regulated in healthy human development but frequently dysregulated in disease. Massively parallel sequencing of RNA (RNA-seq) has become a ubiquitous technology in biology to assay the resulting "transcriptome": the collection of messenger RNA molecules expressed from the genes of an organism. However, significant computational and statistical challenges remain to translate the resulting noisy, confounded RNA-seq data into meaningful understanding of the biological system or disease state under consideration. I will describe our use of probabilistic models to address such challenges: a novel approach to quantifying alternative splicing across different tissues/diseases and a neural-network model that predicts splicing from DNA sequence, improving interpretation of rare variants from exome or whole-genome sequencing studies.

Biography: Dr. Knowles studied Natural Sciences and Information Engineering at the University of Cambridge before obtaining an MSc in Bioinformatics and Systems Biology at Imperial College London. During his PhD studies in the Cambridge University Engineering Department Machine Learning Group under Zoubin Ghahramani he worked on Bayesian nonparametric models for factor analysis, hierarchical clusterings and network analysis, as well as on (stochastic) variational inference. He is currently a post-doctoral researcher at Stanford University with Sylvia Plevritis (Center for Computational Systems Biology/Radiology) and Jonathan Pritchard (Genetics/Biology) having previously worked with Daphne Koller (Computer Science). His work involves the application of statistical machine learning in functional genomics, with the occasional foray into imaging of biological systems. As of 2017 he is an O-1 Alien of Extraordinary Ability and has a T-shirt to prove it.