Bayesian models of expression in the transcriptome for clinical RNA-seq

Principal Investigators

Dr J J Hensman

Institution

University of Sheffield

Contact information of lead PI
Country

United Kingdom

Title of project or programme

Bayesian models of expression in the transcriptome for clinical RNA-seq

Source of funding information

MRC

Total sum awarded (Euro)

€ 511,086

Start date of award

01/09/2013

Total duration of award in years

4.0

The project/programme is most relevant to:

Motor neurone diseases

Keywords
Research Abstract
Background — RNA-Seq technology is enabling investigation of gene expression at the transcript level, including the identification of alternatively spliced isoforms. In Motor Neurone Disease, alternative splicing has been strongly implicated as a pathogenic mechanism. RNA-Seq for clinical data and MND in particular requires the development of new statistical methodologies to tackle challenges specific to such data. Bayesian statistical methods for RNA-

seq are desirable to deal with the uncertainty in quantifying transcript expression, but existing approaches are prohibitively slow for big data. Aims & Objectives — 1) To develop practical algorithms for transcript quantification from RNA-Seq in the Bayesian statistical framework. 2) To build statistical models *around* the transcript quantification problem, addressing problems specific to clinical data. 3) To use the developed algorithms to investigate the effects of splicing in Motor Neurone Disease Methodology — The Bayesian statistical framework will be the cornerstone of the project. Whilst Bayesian methods are often computationally demanding, I shall make use of approximate posterior inference. I'll build on recent work in this area to make fast algorithms for the analysis of RNA-Seq data. I'll collaborate closely with clinical and wet-lab staff in the SITraN neuroscience facility, giving my work immediate impact on research into MND. Scientific opportunities — The quantification of transcripts in RNA-Seq bears a close resemblance to Latent Dirichlet Allocation (LDA), a statistical model used for the analysis of text corpora. Investigation of this link will enable the transfer of knowledge from this field to enable statistical advances for processing RNA-Seq.

Lay Summary
Further information available at:

Types:
Investments > €500k

Member States:
United Kingdom

Diseases:
Motor neurone diseases

Years:
2016

Database Categories:
N/A

Database Tags:
N/A