BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	Algorithms in Bioinformatics
Title	BANDITS: Bayesian ANalysis of DIfferenTial Splicing A Bayesian hierarchical model for differential splicing accounting for sample-to-sample variability and mapping uncertainty
All Authors	Simone Tiberi (1,2) and Mark D Robinson (1,2)
Affiliation	

1) Institute of Molecular Life Sciences, University of Zurich, Zurich
2) Swiss Institute of Bioinformatics, University of Zurich, Zurich

Motivation

Alternative splicing plays a fundamental role in the biodiversity of proteins as it allows a single gene to generate several transcripts and, hence, to code for multiple proteins. However, variations in splicing patterns can be involved in diseases. When investigating differential splicing (DS) between conditions, typically healthy vs disease, scientists are increasingly focusing on differential transcript usage (DTU), i.e. in changes in the proportion of transcripts.

A big challenge in DTU analyses is that, unlike gene level studies, the counts at the transcript level, which are of primary interest, are not observed because most reads map to multiple transcripts. Tools such as salmon or kallisto allow, via expectation maximization (EM) algorithms, to estimate the expected number of fragments originating from each transcript. Most DTU methods (e.g., DRIMSeq, BayesDRIMSeq and SUPPA2) follow a plug-in approach and take the estimated counts as input by treating them as real transcript counts, thus neglecting the uncertainty in the estimates. In order to overcome this issue, some methods, such as cjBitSeq and casper, consider what transcripts each read is compatible with (also called equivalence class); nevertheless, none of these tools allows for sample-specific proportions (i.e., they assume all samples to share the same transcript relative abundance).

Methods

To overcome the limitations of current methods for DTU, we present BANDITS, a Bioconductor R package* to perform DTU based on RNA-seq data. BANDITS uses a Bayesian hierarchical model, with a Dirichlet-multinomial structure, to explicitly model the variability between samples. Our tool inputs the equivalence class of each read, by treating the allocations of reads to the transcripts as latent variables. When a read is compatible with more than one gene, the gene allocation is also treated as a latent variable. The parameters of the model are inferred via Markov chain Monte Carlo (MCMC) techniques where, via a data augmentation procedure, we alternately sample the Dirichlet-multinomial parameters and the latent variables.

To ensure that the MCMC posterior chains have converged, we assess the stationarity of the full log-posterior density via Heidelberg and Welch's convergence diagnostic. Despite the computational complexity of full MCMC algorithms, the core of our method is coded in C++, which makes BANDITS highly efficient and feasible to run on a laptop, even for complex model organisms.

Results

In order to test for DTU, at both transcript and gene level, we approximate the posterior densities of the parameters by a multivariate normal distribution and apply a multivariate Wald test. Our method tests for DTU at both transcript and gene level, allowing scientists to investigate what specific transcripts are differentially used in selected genes. Furthermore, our tool is not limited to two group comparisons and also allows to test for DTU when samples belong to more than two groups.

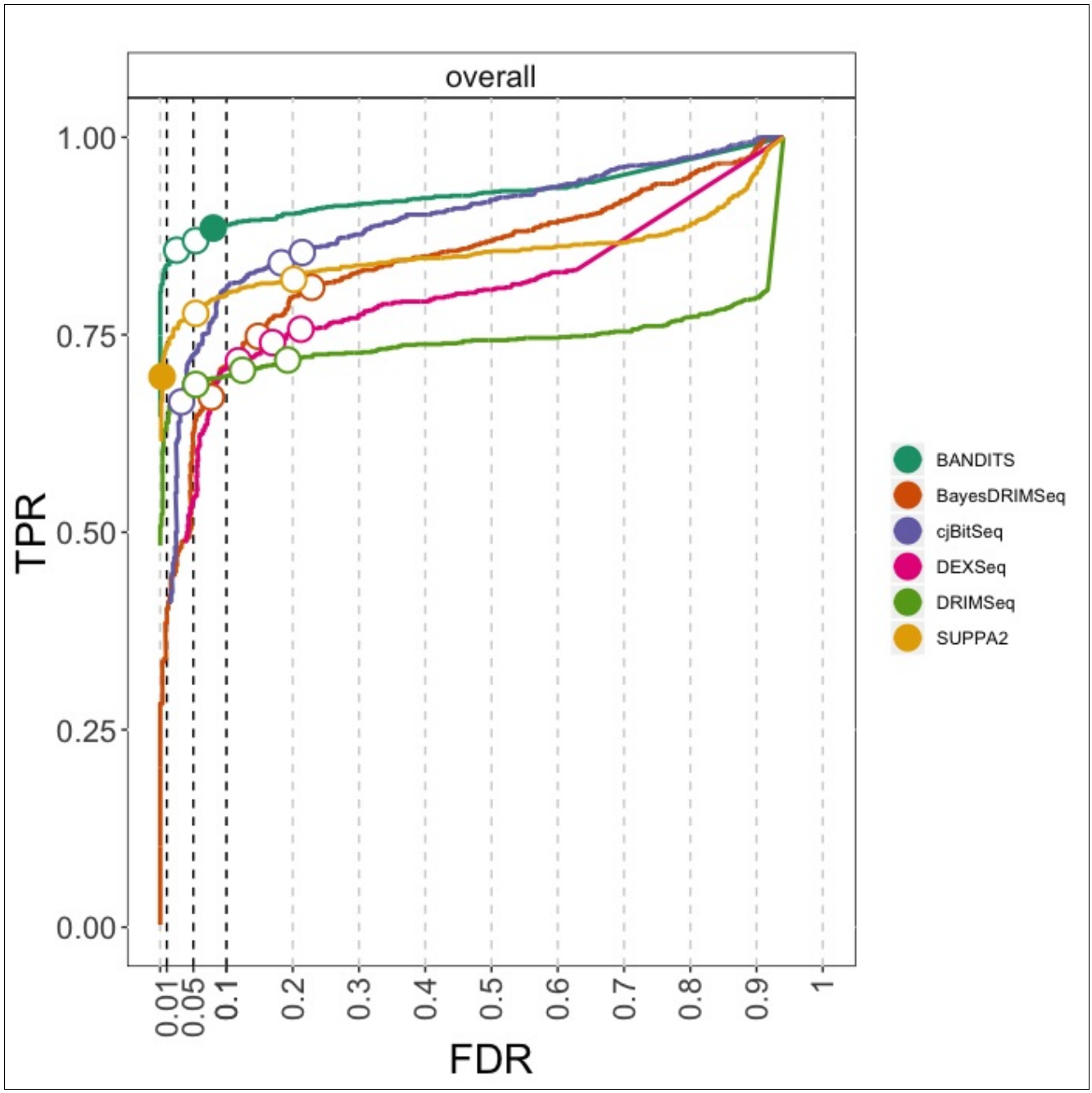
We will show how, both in simulation studies and experimental data analyses, the proposed methodology outperforms existing methods (e.g., see Figure 1).

*BANDITS is currently available on Github (https://github.com/SimoneTiberi/BANDITS) and will be included in Bioconductor release 3.9 at the end of April 2019.

Caption for Figure 1: True positive rate (TPR) vs. false discovery rate (FDR) computed for several methods for DTU in a 6 vs 6 RNA-seq simulation study from a human genome. For any given FDR threshold, Bspliced provides a significantly higher TPR than any other method considered. We obtained similar results in all simulation and experimental data analyses we performed.

Info
-

Figure



Availability	-

Corresponding Author

Name, Surname	Simone, Tiberi
Email	Simone.Tiberi@uzh.ch

Submitted on	04.04.2019