

Unlocking RNA-seq tools for zero inflation and single cell applications using ZINB-WaVE observation weights

Koen Van den Berge, Ghent University

BioC Europe 2017, Cambridge

The team



Koen Van den Berge*



Fanny Perraudeau*



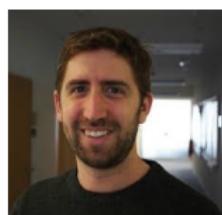
Davide Risso



Jean-Philippe Vert



Charlotte Soneson



Michael Love



Mark Robinson

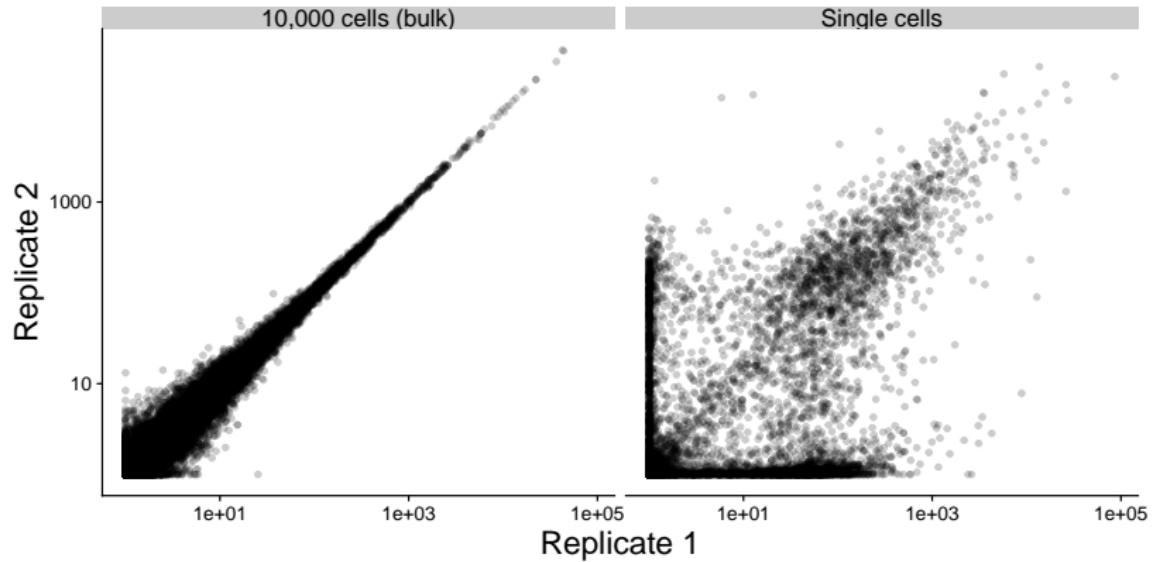


Sandrine Dudoit



Lieven Clement

single-cell RNA-sequencing (scRNA-seq) is noisier than bulk RNA-seq



data from [Shalek et al. 2013]

Observation weights unlock Bioconductor RNA-seq tools towards zero inflation

Excess zeros observed → zero inflation

We propose to model counts with a zero inflated negative binomial (ZINB) distribution

$$f_{ZINB}(y_{ij}; \mu_{ij}, \theta_j, \pi_{ij}) = \pi_{ij}\delta + (1 - \pi_{ij})f_{NB}(y_{ij}; \mu_{ij}, \theta_j). \quad (1)$$

Observation weights unlock Bioconductor RNA-seq tools towards zero inflation

Excess zeros observed → zero inflation

We propose to model counts with a zero inflated negative binomial (ZINB) distribution

$$f_{ZINB}(y_{ij}; \mu_{ij}, \theta_j, \pi_{ij}) = \pi_{ij}\delta + (1 - \pi_{ij})f_{NB}(y_{ij}; \mu_{ij}, \theta_j). \quad (1)$$

A ZINB model corresponds to a weighted NB where **observation weights** are posterior probabilities

$$w_{ij} = \frac{(1 - \pi_{ij})f_{NB}(y_{ij}; \mu_{ij}, \theta_j)}{f_{ZINB}(y_{ij}; \mu_{ij}, \theta_j, \pi_{ij})} \quad (2)$$

Observation weights unlock Bioconductor RNA-seq tools towards zero inflation

Excess zeros observed → zero inflation

We propose to model counts with a zero inflated negative binomial (ZINB) distribution

$$f_{ZINB}(y_{ij}; \mu_{ij}, \theta_j, \pi_{ij}) = \pi_{ij}\delta + (1 - \pi_{ij})f_{NB}(y_{ij}; \mu_{ij}, \theta_j). \quad (1)$$

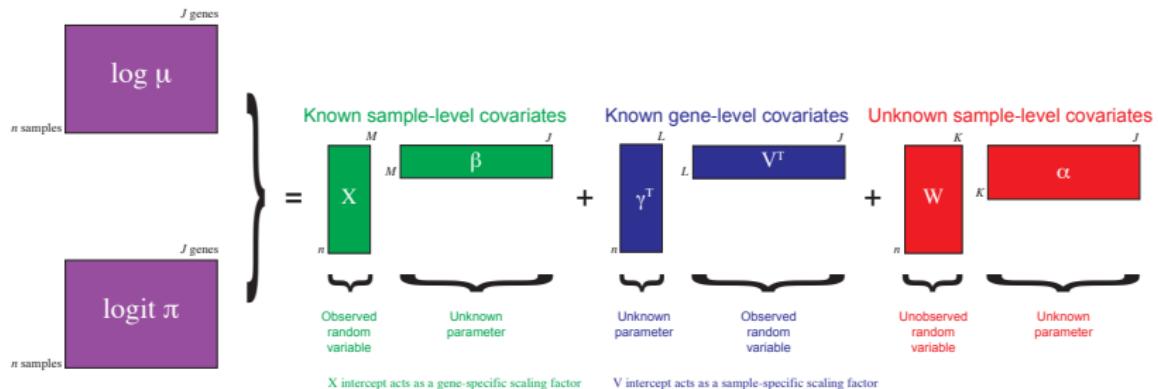
A ZINB model corresponds to a weighted NB where **observation weights** are posterior probabilities

$$w_{ij} = \frac{(1 - \pi_{ij})f_{NB}(y_{ij}; \mu_{ij}, \theta_j)}{f_{ZINB}(y_{ij}; \mu_{ij}, \theta_j, \pi_{ij})} \quad (2)$$

Weights are used to unlock RNA-seq NB models (edgeR, DESeq2) for zero inflation [Van den Berge et al. 2017].

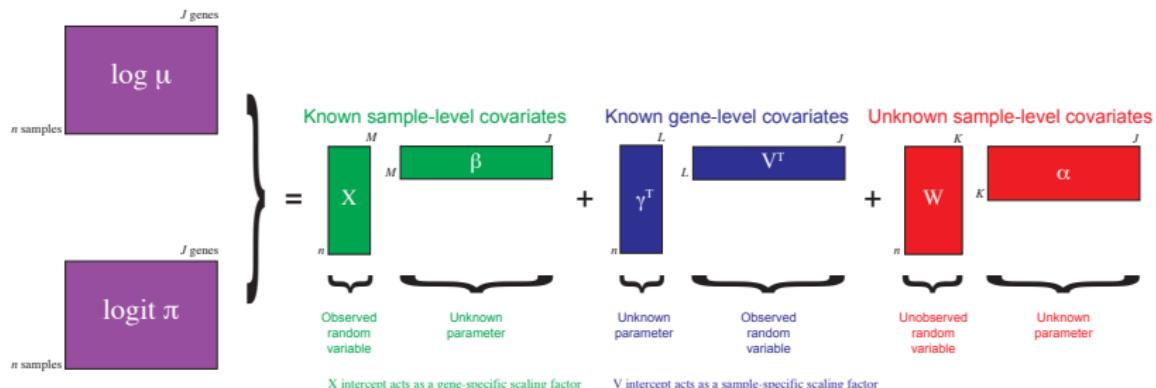
zinbwave can be used to fit ZINB models in scRNA-seq

Estimation of the ZINB parameters using penalized likelihood
implemented in the ZINB-WaVE model [Risso et al. 2017]
Bioconductor: <http://bioconductor.org/packages/zinbwave/>



zinbwave can be used to fit ZINB models in scRNA-seq

Estimation of the ZINB parameters using penalized likelihood
implemented in the ZINB-WaVE model [Risso et al. 2017]
Bioconductor: <http://bioconductor.org/packages/zinbwave/>



Alternatively: EM-algorithm (but slow, and not as extensive)

zingeR: unlocking RNA-seq tools for zero-inflation and single cell applications

Koen Van den Berge, Charlotte Soneson, Michael I. Love, Mark D. Robinson, Lieven Clement

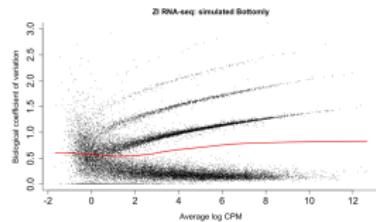
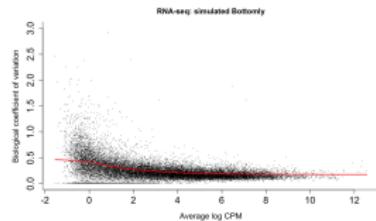
doi: <https://doi.org/10.1101/157982>

This article is a preprint and has not been peer-reviewed [what does this mean?].



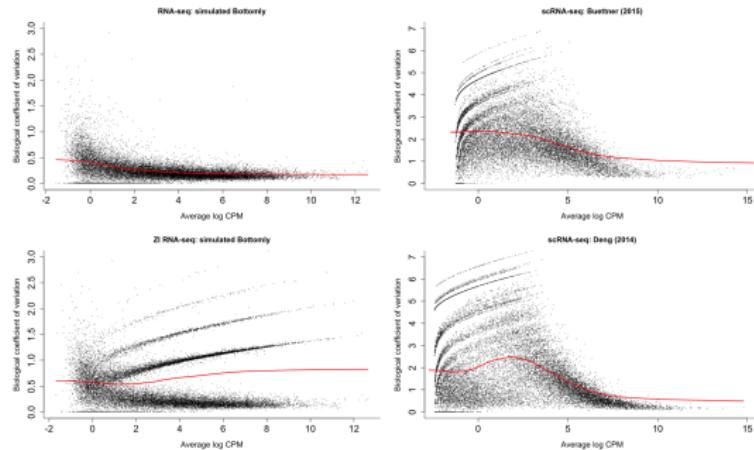
Downweighting excess zeros recovers mean-variance trend, resulting in high power

Simulated (ZI-)bulk RNA-seq data using [Zhou et al. 2014] framework



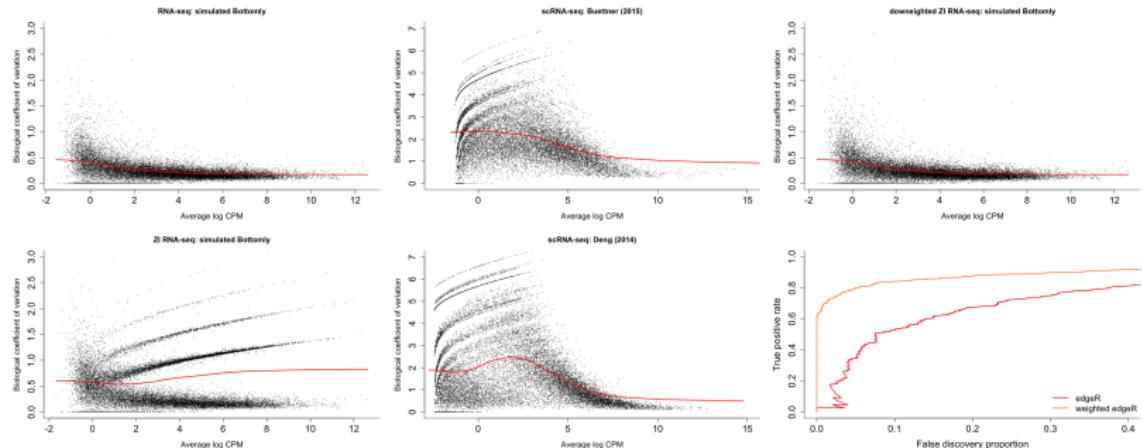
Downweighting excess zeros recovers mean-variance trend, resulting in high power

Simulated (ZI-)bulk RNA-seq data using [Zhou et al. 2014] framework



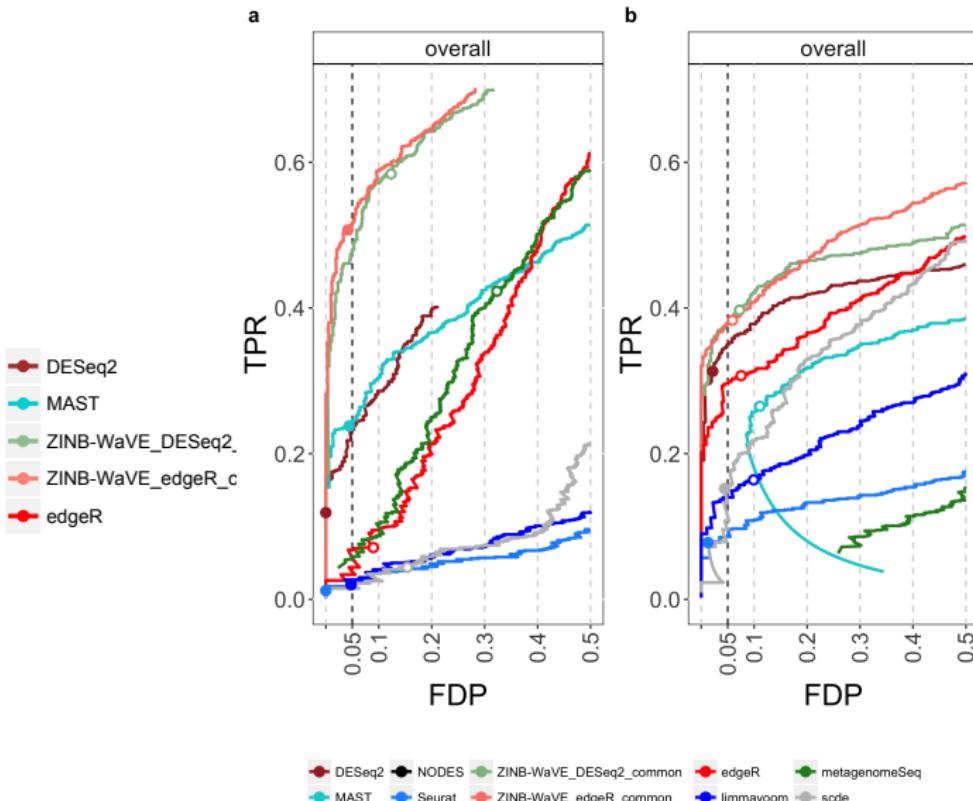
Downweighting excess zeros recovers mean-variance trend, resulting in high power

Simulated (ZI)-bulk RNA-seq data using [Zhou et al. 2014] framework



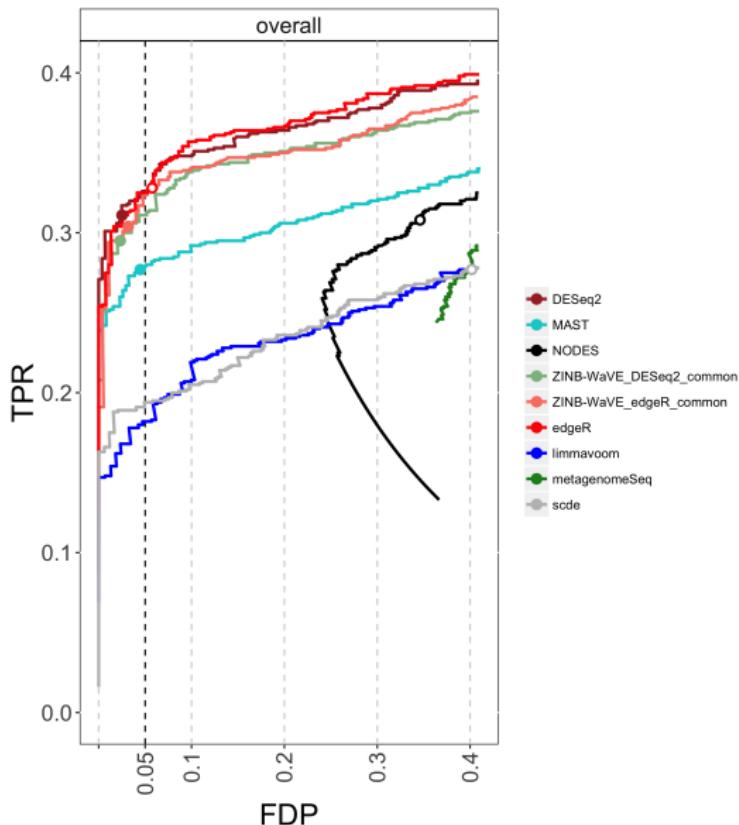
High power, good FDR control in scRNA-seq simulations

Full-length protocols

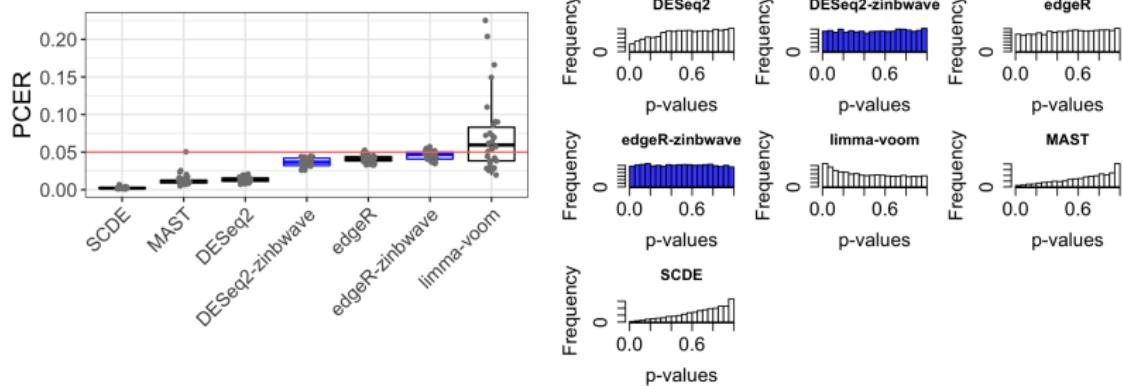


High power, good FDR control in scRNA-seq simulations

3' tag UMI counting protocols, e.g. 10X Genomics Chromium



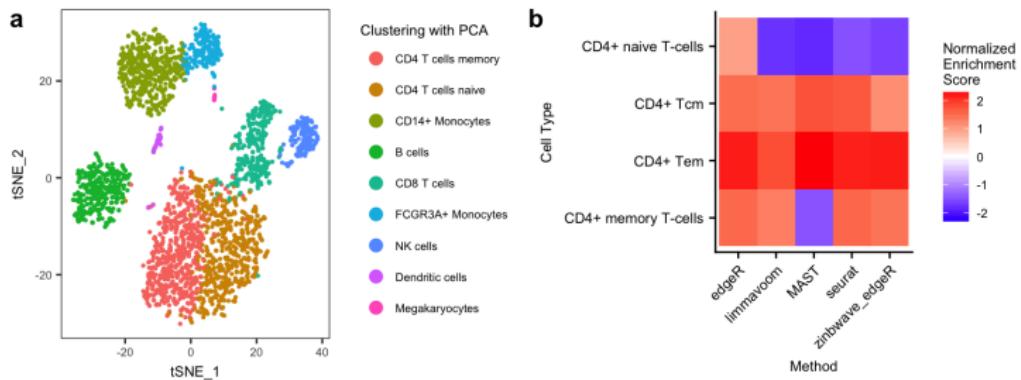
Mock comparisons on real data show good FPR control



Non-UMI dataset on 622 neuronal cells from [Usoskin et al. 2015].
45 vs. 45 mock comparisons.

Downweighting leads to biologically meaningful results

10X Genomics PBMC dataset, preprocessed using tutorial from Seurat.



Method is implemented in zinbwave Bioc package

- ▶ `computeObservationalWeights` for weights calculation
- ▶ `edgeR:glmWeightedF` for ZI-adjusted inference
- ▶ `DESeq2:nbinomWaldTest` and `nbinomLRT` for ZI-adjusted inference
- ▶ Tutorial available in `zinbwave` vignette very soon!