

The entire analysis has been made in R language and it's available at [LINK].

The gene counts have been filtered using a proportion test (cpm cutoff 1) as implemented in NOISeq R package [1-2].

The so filtered resulting gene counts have been normalized with a Trimmed Mean of M-values (TMM) as described in edgeR package [3-4].

In order to remove the batch effects due to samples preparation and sequencing, the so normalized counts have been processed with the RUVs method, implemented in the RUVSeq package[5].

Finally, the Differential Expression analysis has been made using the edgeR [3-4] package, as described in the section 4.2.8 of the manual.

Gene annotation has been made with org.Mm.eg.db package.

references:

[1] Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A and Conesa A (2011). "Differential expression in RNA-seq: a matter of depth." *Genome Research*, 21(12), pp. 4436.

[2] Tarazona S, Furio-Tari P, Turra D, Pietro AD, Nueda MJ, Ferrer A and Conesa A (2015). "Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package." *Nucleic Acids Research*, 43(21), pp. e140.

[3] Robinson MD, McCarthy DJ and Smyth GK (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics*, 26(1), pp. 139-140.

[4] McCarthy, J. D, Chen, Yunshun, Smyth and K. G (2012). "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation." *Nucleic Acids Research*, 40(10), pp. 4288-4297.

[5] Risso D, Ngai J, Speed T and Dudoit S (2014). "Normalization of RNA-seq data using factor analysis of control genes or samples." *Nature Biotechnology*, 32(9), pp. 896–902. In press, <http://www.nature.com/nbt/journal/v32/n9/full/nbt.2931.html>.