

# SARTools : un pipeline complet pour l'analyse différentielle de données RNA-Seq

H. Varet<sup>a,b</sup>, J.-Y. Coppée<sup>b</sup> et M.-A. Dillies<sup>a,b</sup>

<sup>a</sup>Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI)

Institut Pasteur

28 rue du Docteur Roux, 75015 Paris

{hugo.varet}, {marie-agnes.dillies} @pasteur.fr

<sup>b</sup>Plate-forme Transcriptome et Epigénome

Institut Pasteur

28 rue du Docteur Roux, 75015 Paris

jean-yves.coppee@pasteur.fr

**Mots clefs** : RNA-Seq, analyse différentielle, recherche reproductible, reporting.

SARTools est un package R dédié à l'analyse différentielle de données RNA-Seq dans le cadre de plans d'expériences simples, i.e. de plans d'expériences comparant deux ou plusieurs conditions biologiques d'un même facteur. SARTools fournit des outils pour importer les données de RNA-Seq, pour générer des graphiques descriptifs et de diagnostic, pour réaliser l'analyse différentielle avec DESeq2 [1] ou edgeR [2] et pour exporter les listes de gènes différentiels dans des fichiers textes tabulés compatibles avec tout tableur. En s'appuyant sur le package knitr [3], SARTools génère également un rapport final au format HTML qui reprend toutes les figures produites, explique les méthodes statistiques et donne les résultats de l'analyse différentielle. Dans un souci de recherche reproductible, le rapport affiche également les paramètres choisis pour l'analyse ainsi que les versions des packages utilisés. SARTools ne remplace pas DESeq2 ou edgeR mais fournit un environnement complet pour les mettre en oeuvre. De plus, la vignette du package fournit une aide à l'utilisation du workflow et donne des conseils pour détecter divers problèmes telles que la présence d'effets batch dans les données, ou une inversion déchantillons.

SARTools est disponible sur GitHub et est distribué avec deux scripts R génériques : un pour DESeq2 et un pour edgeR. Une fois le package et ses dépendances installés, il suffit de définir les paramètres de l'analyse dans le préambule d'un des scripts et de l'exécuter entièrement pour générer tous les fichiers de résultats décrits ci-dessus.

## Références

- [1] Love MI, Huber W and Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, pp. 550.
- [2] Robinson MD, McCarthy DJ and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, pp. -1.
- [3] Xie Y (2015). "knitr: A General-Purpose Package for Dynamic Report Generation in R". R package version 1.9.