



HAL
open science

ASICS : un package R pour l'identification et la quantification de métabolites dans un spectre RMN 1H

Gaëlle Lefort, Laurence Liaubet, Cécile Canlet, Nathalie Villa-Vialaneix,
Rémi Servien

► **To cite this version:**

Gaëlle Lefort, Laurence Liaubet, Cécile Canlet, Nathalie Villa-Vialaneix, Rémi Servien. ASICS : un package R pour l'identification et la quantification de métabolites dans un spectre RMN 1H. 7ème Rencontres R, Jul 2018, Rennes, France. hal-02791362

HAL Id: hal-02791362

<https://hal.inrae.fr/hal-02791362>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

ASICS : un package R pour l'identification et la quantification de métabolites dans un spectre RMN ^1H

G. Lefort^a, L. Liaubet^b, C. Canlet^c, N. Villa-Vialaneix^a et R. Servien^e

^aMIAT, Université de Toulouse, INRA, Castanet-Tolosan, France
gaelle.lefort@inra.fr, nathalie.villa-vialaneix@inra.fr

^bGenPhySE, Université de Toulouse, INRA, ENVT, Castanet Tolosan, France
laurence.liaubet@inra.fr

^cAxiom Platform, MetaToul-MetaboHUB, National Infrastructure for Metabolomics and Fluxomics, F-31027 Toulouse, France
cecile.canlet@inra.fr

^eInTheRes, Université de Toulouse, INRA, ENVT, France
remi.servien@inra.fr

Mots clés : Metabolomique, Résonance Magnétique Nucléaire (RMN), Quantification de métabolites

Les données dites *omiques* permettent de décrire le fonctionnement biologique d'un organisme à divers niveaux de l'échelle du vivant : génome (séquence d'ADN), transcriptome (ensemble des ARN présents dans une cellule ou un tissu), protéome (ensemble des protéines), métabolome (ensemble des métabolites), par exemple. L'étude du métabolome suscite un intérêt croissant car, comparativement à d'autres omiques, il est plus proche des phénotypes observés d'intérêt (généralement situés à l'échelle de l'individu) et pourrait donc avoir un potentiel important en terme de *biomarqueurs*. Il s'agit de trouver de nouvelles caractéristiques biologiques (des biomarqueurs) facilement mesurables permettant, par exemple, de détecter une maladie (dépistage médical) ou d'adapter le traitement à des caractéristiques propres à l'individu (médecine personnalisée).

Diverses approches permettent de caractériser le métabolome en mesurant simultanément un grand nombre de métabolites dans les fluides biologiques et les tissus. Parmi celles-ci, la spectrométrie par résonance magnétique nucléaire (RMN) est une technologie haut-débit qui produit des spectres caractéristiques du mélange complexe de métabolites présents dans l'échantillon d'intérêt. Parmi les technologies haut-débit, la RMN est relativement peu coûteuse et suscite donc des espoirs pour la recherche efficace de nouveaux biomarqueurs. Cependant, l'interprétation biologique des spectres obtenus est difficile car les spectres ne donnent pas une mesure explicite des différentes quantités de métabolites présentes dans l'échantillon.

Aussi, traditionnellement, l'analyse du métabolome obtenu par RMN est effectuée de la manière suivante : tous les spectres sont divisés en petits intervalles appelés *buckets*. Ensuite, l'aire sous la courbe est calculée pour chacun de ces *buckets* et les analyses sont réalisées sur ces nouvelles variables. Ceci permet d'obtenir une liste de *buckets* d'intérêt (par exemple, *buckets* significativement différents entre deux conditions d'étude). L'identification des métabolites sous-jacents nécessite l'aide d'un expert en RMN ; ce qui rend cette étape fastidieuse, longue et subjective.

Une approche alternative prometteuse consiste donc à identifier et quantifier les métabolites présents dans le mélange complexe à partir de son spectre et à réaliser l'analyse statistique sur les résultats de cette quantification. Quelques outils existant permettent de quantifier des métabolites à partir d'un spectre RMN (Autofit [4], batman [1], Bayesil [2]). Toutefois, Autofit est inclus dans un logiciel commercial, batman est très coûteux en temps de calcul et Bayesil n'est disponible qu'à travers une interface web qui n'est pas adaptée pour quantifier plusieurs spectres simultanément.

Pour améliorer la quantification du métabolome et faciliter son analyse statistique, nous avons développé un nouveau package R, **ASICS** (Automatic Statistical Identification in Complex Spectra). Ce package propose un pipeline complet d'analyse de spectres RMN. Il contient une méthode statistique d'identification et de quantification de métabolites dans un mélange complexe basé sur une librairie de spectres de références (spectres obtenus à partir de chaque métabolite pur) Tardivel et al. [3]. Une librairie de 191 spectres de métabolites purs est incluse dans le package. Elle peut être complétée, réduite ou remplacée par des spectres de référence de l'utilisateur si besoin. Les spectres à analyser peuvent être importés depuis des fichiers bruts de type Bruker grâce à une fonction dédiée. Cette importation inclut des prétraitements standards comme une correction de la ligne de base, un alignement des pics et une normalisation. À partir des spectres importés, une fonction identifie et quantifie les métabolites présents. Enfin, des analyses statistiques (ACP, OPLS-DA et tests statistiques) peuvent être réalisées directement sur les résultats de la quantification. Le package inclut enfin divers outils diagnostiques permettant d'évaluer la qualité de la quantification ou la concordance avec les résultats de l'approche traditionnelle utilisant les *buckets*.

Les performances de **ASICS** ont été comparées à celles des approches concurrentes sur un mélange complexe synthétique de 21 métabolites [3] ainsi que sur des données réelles issues du projet PORCINET (ANR-09-GENM-005) pour lequel des dosages directs ont été pratiqués en même temps que l'analyse RNM du métabolome. **ASICS** montre des performances meilleures que ses concurrents en terme d'identification, avec une bonne sensibilité et une bonne spécificité pour un temps de calcul réduit.

Les quantifications obtenues sont bien corrélées aux dosages réels et les analyses pratiquées sur les résultats de la quantification permettent de retrouver de manière simple et automatisée les résultats des analyses pratiquées sur les *buckets*. **ASICS** a été accepté sur Bioconductor et sera disponible dès le déploiement de la prochaine version de Bioconductor le 01/05/2018. Une version de développement est aussi disponible sur GitHub (<https://github.com/GaelleLefort/ASICS>).

Références

- [1] Hao, J., Astle, W., De Iorio, M., & Ebbels, T. M. D. (2012). BATMAN—an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics*, 28(15) :2088–2090.
- [2] Ravanbakhsh, S., Liu, P., Bjordahl, T. C., Mandal, R., Grant, J. R., Wilson, M., Eisner, R., Sinelnikov, I., Hu, X., Luchinat, C., Greiner, R., & Wishart, D. S. (2015). Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PLOS ONE*, 10(5) :e0124219.
- [3] Tardivel, P., Canlet, C., Lefort, G., Tremblay-Franco, M., Debrauwer, L., Concordet, D., & Servien, R. (2017). ASICS : an automatic method for identification and quantification of metabolites in complex 1D 1H NMR spectra. *Metabolomics*, 13(10) :109.
- [4] Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. (2006). Targeted Profiling : Quantitative Analysis of 1 H NMR Metabolomics Data. *Analytical Chemistry*, 78(13) :4430–4442.