



**HAL**  
open science

## Analyse statistique de données radiomiques et métabolomiques : prédiction des lésions mammaires triple-négatives

Fanny Orlhac, Olivier Humbert, Thierry Pourcher, Lun Jing, Jean-Marie Guignonis, Jacques Darcourt, Nicholas Ayache, Charles Bouveyron

### ► To cite this version:

Fanny Orlhac, Olivier Humbert, Thierry Pourcher, Lun Jing, Jean-Marie Guignonis, et al.. Analyse statistique de données radiomiques et métabolomiques : prédiction des lésions mammaires triple-négatives. 12ème Conférence Francophone d'Epidémiologie Clinique (EPICLIN) et 25èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), May 2018, Nice, France. pp.S180-S181, 10.1016/j.respe.2018.03.307 . hal-01736164

**HAL Id: hal-01736164**

**<https://hal.science/hal-01736164>**

Submitted on 16 Mar 2018

**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.

## **Analyse statistique de données radiomiques et métabolomiques : prédiction des lésions mammaires triple-négatives**

Fanny Orlhac<sup>1</sup>, Olivier Humbert<sup>2,3</sup>, Thierry Pourcher<sup>3</sup>, Lun Jing<sup>3</sup>, Jean-Marie Guignonis<sup>3</sup>, Jacques Darcourt<sup>2,3</sup>, Nicholas Ayache<sup>1</sup>, Charles Bouveyron<sup>1,4</sup>

1 : Epione, Inria Sophia-Antipolis, Valbonne, France

2 : Service de médecine nucléaire, Centre Antoine Lacassagne/UCA, Nice, France

3 : Laboratoire TIRO, UMRE 4320 CEA/UCA, Nice, France

4 : Laboratoire JAD, UMR CNRS 7135/UCA, Nice, France

Thème : Apprentissage et intelligence artificielle / Essais thérapeutiques simulés

Mots clés : Analyse statistique - Analyse discriminante – Sélection de variables - Radiomique – Métabolomique

**Introduction :** La caractérisation de l'hétérogénéité tumorale à partir des images médicales (appelée aussi radiomique) et de l'extraction de données omiques est un enjeu majeur en cancérologie, notamment dans la mise en place de la médecine de précision. Or actuellement, le lien entre les variables radiomiques (VR) et les caractéristiques biologiques des lésions est encore mal connu. L'objectif de ce travail est d'étudier la corrélation entre les VR et les variables métabolomiques (VM) dans le cancer du sein, et d'analyser leur capacité à prédire le sous-type immunohistochimique des lésions.

**Méthodes :** 26 patientes porteuses d'un cancer mammaire ont été incluses dans cette étude. Chaque patiente a bénéficié d'un examen de tomographie par émission de positons (TEP). Quarante-trois VR issues des images TEP ont été calculées dans la zone tumorale. A partir de la pièce opératoire, nous avons utilisé un spectromètre de masse pour mesurer l'expression de 1500 VM référencées dans la base de données *Human metabolome*. Les coefficients de corrélation de Spearman (R) entre chaque VR et chaque VM ont été analysés. Nous avons également étudié séparément la capacité des VR et VM pour identifier les lésions mammaires triple-négatives (TN), en comparant 5 méthodes statistiques : l'analyse discriminante linéaire (LDA), l'analyse discriminante des moindres carrés partiels (PLSDA), l'analyse discriminante des structures latentes par projections orthogonales (OPLS), l'analyse discriminante en grande dimension (HDDA, Bouveyron *et al*, 2007) et l'analyse HDDA globalement parcimonieuse (gsHDDA, Bouveyron *et al*, 2016). Cette procédure a été répétée 25 fois avec une sélection aléatoire de 16 patientes pour l'apprentissage et de 10 patientes pour la validation. Les performances de chaque méthode ont été mesurées en utilisant l'index de Youden (sensibilité + spécificité -1).

**Résultats :** Sept lésions sur 26 sont TN. Le coefficient de corrélation moyen, en valeur absolue, entre les VR et VM est égal à  $0,20 \pm 0,14$  (intervalle :  $[0-0,81]$ , 3% avec  $|R| \geq 0,50$ ). Seulement 20 VR sont modérément corrélés avec au moins 50 VM ( $|R| \geq 0,50$ ). En utilisant les différentes méthodes statistiques, l'index de Youden moyen varie entre 0,18 et 0,34 à partir des VM et entre -0,12 et 0,50 pour les VR. Les meilleures performances pour la distinction des lésions TN sont obtenues pour HDDA (Youden= $0,50 \pm 0,35$ ) et gsHDDA (Youden= $0,49 \pm 0,34$ ) à partir des VR. L'un des avantages de gsHDDA est que le modèle statistique est construit à partir d'un sous-ensemble de variables sélectionnées. L'étude de ces variables a montré une corrélation modérée ( $|R| = 0,21 \pm 0,15$ , intervalle :  $[0-0,71]$ ) entre les 10 VR et les 601 VM sélectionnées (variables sélectionnées par gsHDDA dans au moins 50%

des tests), ce qui suggère que la combinaison des deux sources d'information pourrait améliorer les performances d'identification des lésions TN.

**Conclusion :** Dans le cancer du sein, nous avons montré une corrélation faible à modérée entre les VR et les VM. Cependant, les deux types de données permettent une identification des lésions TN avec des performances voisines. Le recrutement de patientes supplémentaires est en cours pour confirmer ces résultats. L'analyse conjointe des VR et VM est à l'étude afin de bénéficier de la complémentarité des informations radiomiques et métabolomiques et ainsi d'améliorer les performances de classification.