

A07 : Comparaison de méthodes d'apprentissage non supervisées chez des patientes atteintes d'un cancer du sein localisé à l'aide de données de métabolomique

Titre

Français : Comparaison de méthodes d'apprentissage non supervisées chez des patientes atteintes d'un cancer du sein localisé à l'aide de données de métabolomique

Anglais : Comparison of unsupervised machine learning methods in localised breast cancer patients based on metabolomic signatures.

Auteurs

J Gal (1), C Bailleux (2), D Chardin (3), T Pourcher (4), JM Guignonis (4), E Barranger (5), JM Ferrero (6), R Schiappa (7), O Humbert (8), E Chamorey (7)
(1) Département d'Epidemiologie, de Biostatistiques et des Données de Santé, Centre Antoine Lacassagne, 33 avenue de Valombrose, 06189, Nice, France
(2) Département d'oncologie médicale, Centre Antoine Lacassagne, 33 avenue de Valombrose, 06189, Nice, France
(3) Département de médecine nucléaire, Antoine Lacassagne, avenue de Valombrose, 06189, Nice, France
(4) Transporteurs, Imagerie et Radiothérapie en Oncologie UMR E-4320, Faculté de médecine, 28 Avenue de Valombrose, 06107, Nice, France
(5) Départment de Chirurgie, Centre Antoine Lacassagne, avenue de Valombrose, 06189, Nice, France
(6) d'oncologie médicale, Centre Antoine Lacassagne, avenue de Valombrose, 06189, Nice, France
(7) Données de Santé, Centre Antoine Lacassagne, avenue de Valombrose, 06189, Nice, France
(8) médecine nucléaire, Centre Antoine Lacassagne, avenue de Valombrose, 06189, Nice, France

Responsable de la présentation

Nom : Gal
Prénom : Jocelyn
Adresse professionnelle : 33 avenue de Valombrose
Code postal : 06189
Ville : Nice
Pays : France
Newsletter :

Mots clés

Français : Cancer du sein, Métabolomique, Méthodes de machine learning non supervisée
Anglais : Breast cancer, Metabolomic, Unsupervised machine learning methods

Spécialité

Principale : Autres
Secondaire : Biologie

Texte

Introduction
Le traitement des cancers du sein est guidé par deux classifications: la classification traditionnelle reposant sur le TNM, les facteurs histologiques ainsi que certains autres marqueurs et la classification moléculaire issue de la transcriptomique qui introduit la notion de cancer du sein luminal A/B, HER2 et basal-like. Cette dernière a été obtenue à l'aide d'analyses non supervisées. La métabolomique étudie l'ensemble des métabolites primaires, elle permet de mieux comprendre les perturbations des voies du métabolisme de la cellule tumorale et ainsi de mieux comprendre l'oncogénèse. Tout comme la transcriptomique, la métabolomique génère des données de grandes dimensions. L'objectif de ce travail est de mettre en évidence des signatures métabolomiques de cancer du sein obtenues par différentes méthodes d'apprentissage non supervisées.
Méthodes
Rétrospectivement, 52 patientes atteintes d'un cancer du sein et traitées par chimiothérapie adjuvante entre 2013 et 2016 ont été incluses. Les échantillons tumoraux ont été analysés par chromatographie liquide haute pression et spectromètre de masse permettant d'identifier 1271 métabolites. Cinq méthodes d'apprentissage non supervisées ont été comparées: PCA-Kmeans, Sparcl, SIMLR, Spectral clustering et K-sparse. Le nombre de cluster optimal a été évalué par la méthode gap statistic et l'algorithme t-sne a permis leur représentation visuelle. La séparabilité et l'homogénéité des clusters ont été évaluées à l'aide de l'indice de silhouette (IS). Nous avons analysé les différences clinico-biologiques retrouvées entre les clusters identifiés par chaque méthode. Puis, nous avons comparé les taux de métabolites retrouvés dans chaque cluster pour les 5

méthodes.

Résultats

Pour chacune des 5 méthodes, 3 clusters ont été identifiés. Les méthodes K-sparse et SIMLR sont les plus discriminantes avec des IS de 0,85 et 0,91 respectivement. Pour ces 2 méthodes, des différences significatives inter-cluster ont été retrouvées pour les variables clinico-biologiques suivantes: stade et phénotype tumoral, grade histologique et ki-67. De plus, les patientes associées au cluster avec les facteurs histopronostiques les plus défavorables présentaient une variation de concentration significative des acides aminés fondamentaux tels que L-Méthionine, L-Phénylalanine, L-Isoleucine, L-Proline, L-Glutamate, L-Glutamine mais aussi un taux élevé en UDP-Glucose et en acide Pyruvique. Cette concentration élevée en acides aminés protéinogènes pourrait être corrélée à un mauvais pronostic.

Conclusion

Nos résultats ont montré qu'il était possible d'utiliser les méthodes d'apprentissage non supervisées sur des données de métabolomique afin de mettre en évidence des clusters de patients identifiés sur des acides aminés protéinogènes. K-sparse et SIMLR semblent être les méthodes les plus discriminantes, elles mettent en évidence 3 populations dont les caractéristiques clinico-biologiques standards sont distinctes. La métabolomique semble être un outil pertinent et prometteur dans la classification des cancers du sein, cependant, la pertinence clinique de cette signature ainsi que son utilité dans la stratégie thérapeutique devront être évaluées précisément.