

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **24 juin 2024**

Nom de famille et prénom de l'auteur : **Madame BREEUR Marie**

Titre de la thèse : « *Profils Métabolomiques et Risque de Cancer : Applications de l'Apprentissage Automatique dans la Cohorte EPIC* »



Résumé

Les recherches examinant la relation complexe entre le cancer et les altérations métaboliques sont cruciales au sein de la cohorte EPIC (European Prospective Investigation into Cancer and Nutrition). Les données métabolomiques accumulées au fil de ces études représentent une ressource inestimable pour approfondir diverses questions, telles que l'impact des facteurs environnementaux et du mode de vie sur le métabolisme, ou encore les études englobant plusieurs types de cancers. Toutefois, malgré ce potentiel, les initiatives visant à analyser conjointement les données métabolomiques issus de plusieurs études nichées au sein d'EPIC demeurent rares.

Les données métabolomiques non ciblées sont particulièrement difficiles à combiner, dans la mesure où un même métabolite mesuré dans deux études différentes sera identifié différemment dans chaque étude, compliquant l'identification des métabolites communs. Pour pallier cette limitation, nous avons élaboré GromovMatcher, une méthode non supervisée pour aligner les données métabolomiques non ciblées provenant de différentes études, reposant sur le transport optimal. Les performances de GromovMatcher ont surpassé celles des méthodes existantes, tant sur des données simulées que sur des données issues d'EPIC, ouvrant ainsi une voie prometteuse.

Par ailleurs, nous avons mené une étude exhaustive sur les liens entre les métabolites et divers types de cancer. Grâce à une approche multi-tâches permettant de tirer parti d'un échantillon plus vaste tout en conservant la spécificité de chaque type de cancer, nous avons mené une analyse pan-cancer utilisant des données métabolomiques ciblées provenant de plusieurs cohortes de l'étude EPIC. Cette analyse a mis en lumière des métabolites associés de manière cohérente au risque de cancer

dans différents types de cancer, suggérant des voies métaboliques communes dans la cancérogenèse ainsi que des associations spécifiques à certains types de cancer.

Enfin, nous avons exploré des techniques de réduction de dimension non linéaires pour découvrir des variables métaboliques latentes. En utilisant des autoencodeurs et une analyse en composantes principales (ACP) sur les données métabolomiques ciblées, nous avons identifié des variables métaboliques latentes et évalué leur association avec le risque de cancer ainsi qu'avec des facteurs génétiques. Nos résultats ont notamment mis en évidence un composant fortement lié à la fois au risque de cancer et à l'activité des acides gras poly-insaturés, soulignant ainsi l'efficacité des méthodes non linéaires dans la compréhension du métabolisme du cancer.