



Universiteit
Leiden
The Netherlands

Multi-omics in research: epidemiology, methodology, and advanced data analysis

Faquih, T.O.

Citation

Faquih, T. O. (2023, March 28). *Multi-omics in research: epidemiology, methodology, and advanced data analysis*. Retrieved from <https://hdl.handle.net/1887/3589838>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3589838>

Note: To cite this publication please use the final published version (if applicable).

ملخص النتائج

توفر علوم تحليل منتجات الأيض (الميتابولوميكس / Metabolomics) وتحليل البروتينات (البروتيوميكس / proteomics) وتحليل الجينات (الجينوميكس / Genomics) رؤى عميقة في علم الأحياء البشري وعلم وظائف الأعضاء المتعلقة بالأمراض وأسبابها في الفيزيولوجيا المرضية البيولوجية (Pathophysiology) على نطاق غير مسبق. وقد استكشفنا في هذه الأطروحة التحديات البحثية التي تواجه تطبيق تقنيات الـ OMICS¹ (أوميكس) بنظرة تكاملية عميقة وأوضحنا تطبيقاتها العملية في الدراسات الوبائية (Epidemiological Studies).

وفي الجزء الأول من الأطروحة المكون من ثلاثة فصول، ركزنا على بعض تحديات طرق البحث المنهجية التي تواجه أبحاث الاوميكس. ويتفحص **الفصل الأول من الجزء الأول (Chapter 2)** من هذه الرسالة العلمية العلاقة ما بين نتائج قياسات البروتينات باستخدام "الأبتامير" (aptamer-based proteomics) للمؤشرات الحيوية للتخثر الوريدي (VTE)² مع طرق القياس المعتمدة التي تتم عادة في المراكز الصحية والممارسات الاكلينيكية المتعارف عليها. وقد وجدت دراستنا البحثية أن المؤشرات الحيوية التسعة التي تم قياسها للتخثر الوريدي أظهرت توافقا ضعيفا مع نتائج القياسات المعتمدة. وكانت العلاقة التوافقية ضعيفة بشكل خاص بالنسبة لـ "ثنائيات د" (D-dimer)، وهو مؤشر يستخدم لتشخيص داء التخثر الوريدي.

في الفصل الثاني من الجزء الأول (Chapter 3) من هذه الرسالة، تم دراسة التحديات المتمثلة في التعامل مع البيانات المفقودة في تحليل منتجات الأيض، لا سيما من منصات بيانات تحليل منتجات الأيض الـ "غير مستهدفة". حيث قمنا بتقييم أداء طريقتين لاحتمال القيم المفقودة في بيانات تحليل منتجات الأيض، وهما طريقة "K-nearest neighbor" (kNN) وطريقة "Multiple imputation using chained equations" (التضمين المتعدد)، ومن خلال محاكاة الأنماط المختلفه لفقدان البيانات من الدراسة البحثية الهولندية عن السمعة³ (NEO). أظهرت نتائجنا أن طريقة "التضمين المتعدد" كان لها تحيز أقل من طريقة kNN في معظم الحالات، إلا في الحالات ذات حجم عينة صغيره مع نسبه عاليه من البيانات المفقوده. كما قدمنا أيضا برنامجاً نصي حاسوبي بلغة برمجة R متاح للعامة لتبسيط عملية استخدام الطريقتين لاحتمال القيم المفقودة في بيانات قياسات منتجات الأيض.

بإمكان منتجات الأيض (الميتابولايكس / Metabolites) توفير نظرة ثاقبة وفريدة في الفيزيولوجيا المرضية البيولوجية لأنها تشمل تأثير العوامل الوراثية و تأثير نمط الحياة و تأثير العوامل البيئية. فلذلك تم استخدام علم تحليل منتجات الأيض لدراسة العملية الحيوية للتقدم بالعمر و الهرم. وقد استخدمت بعض الدراسات منتجات الأيض لتقدير "العمر الأيضي"، وهو انعكاس للتقدم بالعمر على مستوى منتجات الأيض، باستخدام طرق نمذجة التنبؤ الإحصائي. **في الفصل الثالث من الجزء الأول (Chapter 4)**، تناولنا التحديات الشائعة لتطوير نماذج التنبؤ بالعمر القائمة على منتجات الأيض

وطورنا نموذجاً للتنبؤ بالعمر المبني على ذلك باستخدام أجهزة منصة Metabolon لتحليل منتجات الأيض. بعد ذلك قمنا بتطوير نموذج في الدراسة البحثية INTERVAL⁴ باستخدام أساليب الإحصاء "Ridge regression" و "Bootstrapping". وقد أظهرت المجموعة المتضمنة في دراسة الـ INTERVAL ان عينة كبيرة من المجتمع كانت في صحة جيدة نسبياً ($n = 11,977$) ومع نطاق عمري واسع ولذلك فإن هذه المجموعة كانت مناسبة لتطوير نموذج تنبؤ العمر الأيضي. وبعد تطوير النموذج والتحقق الداخلي من صحته، أظهرت النتائج ان أداء التنبؤ الإحصائي للعمر الأيضي كان جيداً جداً (R^2 المعدل = 0.83).

اما الجزء الثاني من الأطروحة ، والمكون من ثلاثة فصول ايضاً، فيتناول دراسة عدة أسئلة بحثية في علم epidemiology من خلال استخدام البيانات الجينومية و بيانات تحليل الأيض (باستخدام منصتين Metabolon و Nightingale) بالإضافة إلى طرق تحليل البيانات المتقدمة.

في الفصل الأول من الجزء الثاني (Chapter 5)، قمنا بفحص ارتباطات أكثر من 1300 منتج أبيض، تم قياسها بواسطة منصة Metabolon، و محتوى الدهون الثلاثية الكبدية (HTGC). وقد تضمنت هذه الارتباطات مجموعات من منتجات الأيض تابعه لمسارات حيوية مختلفة (خاصة من مجموعات الأحماض الأمينية والدهون) ، بالإضافة إلى منتجات أبيض غير معهوده. بعد ذلك، قمنا ببناء شبكتين شموليتين متكاملتين، واحدة قائمة على أساس حيوي والأخرى قائمة على أساس إحصائي ، لتعيين وربط منتجات الأيض مع المسارات الحيوية المناسبة. وقد كشفت هذه الشبكات التفاعلية عن مسارات حيوية غير مألوفة يمكن ان تكون مرتبطة بالـ HTGC وربما أمراض الكبد الدهنية ككل هذا وقد تم توفير هذه الشبكات على صفحة الويب المتاحة على الإنترنت للمزيد من الاستكشاف من قبل المجتمع البحثي العلمي.

اما في الفصل الثاني من الجزء الثاني (Chapter 6)، فاستكشفنا الارتباطات المحتملة لأحجام تباين تكرار CAG الأقل حجماً من عتبة تسبب مرض هنتنغتون (اقل من 36 تكرار) في جين *HTT* مع مستويات الأيض المقاسة بواسطة منصة Nightingale. وتبين أن أحد أنواع الخلل الجينية في الجينوم هو خلل ناتج عن اختلاف عدد تسلسل نيوكليوتيدات (Nucleotide) السيتوزين-الأدينين-الجوانين (تكرار تسلسل *CAG*⁵). وعندما يتجاوز حجم هذا التكرار العتبة الغير مرضية في جينات معينة (أو بالقرب من هذه الجينات) ، فإنها تؤدي إلى العديد من الأمراض التنكسية العصبية الشديدة. إحدى هذه الحالات هي حجم تكرار تسلسل *CAG* الكبيرة في جين هنتنغتين⁶ (*HTT*) مما يؤدي إلى مرض هنتنغتون⁷. من والمثير للاهتمام ، أن الدراسات الحديثة وجدت أن حجم تكرار *CAG* حتى اذا كان داخل النطاق الغير مرضي فان له تأثير على وظائف صحية مثل الوظيفة الإدراكية والاكتئاب وانخفاض وزن الجسم وبناء على ذلك تم إجراء تحليل البيانات بطريقه تحليل "التأثيرات المختلطة"⁸ و تحليل "الوساطة"⁹ في بيانات تم جمعها من ثلاث دراسات أوروبية ضخمة ($n = 10,275$). حيث أظهرت نتائجنا أن أحجام تكرار *CAG* الكبيرة الغير مسببة للأمراض كانت مرتبطة بنمط أبيض غير ملائم للصحة، على غرار النمط الأبيض المتصل بزيادة خطر الإصابة بأمراض القلب والأوعية الدموية.

وتتمثل إحدى مزايا تحليل منتجات الأيض في قدرة الغطاء على قياس منتجات الأيض "الخارجية الغريبة"، والتي يشار إليها غالباً باسم Xenobitoics. وهذه المنتجات الأيضية يكون منشأها من استخدام الأدوية والنظام الغذائي والتعرض للتلوث البيئي. وإحدى هذه الكائنات الخارجية المقاسة هي مواد "Per- and Polyfluoroalkyl substances (PFAS)" (بيفاس) الصناعية - والمعروفة باسم "الكيمائيات الأبدية" بسبب تشبثها في البيئة وجسم الكائنات الحية، و بالإضافة إلى طبيعتها التشبثية، فقد اوضحت الدراسات ان التركيزات العالية من PFAS مرتبطة بمشاكل صحية شديدة في الانسان. وانه على الرغم من الجهود المبذولة لتنظيم إنتاجها، لا تزال مستويات PFAS قابلة للاكتشاف في التربة ومياه الشرب وعينات الدم البشري في كل من هولندا وألمانيا.

فلذلك قمنا في الفصل الثالث من الجزء الثاني (Chapter 7)، بدراسة العلاقة ما بين مستويات PFAS في الدم، التي تم قياسها بواسطة منصة Metabolon، و مستويات المنتجات الأيضية و البروتين الدهني التي تم قياسها بواسطة منصة Nightingale في مجموعة من عامة سكان هولندا (دراسة NEO) وألمانيا (دراسة راينلاند¹⁰). ووجدنا ان مستوى المنتجات الأيضية للـ PFAS قابل للقياس في جميع الافراد تقريباً في كلتا الدراستين، علاوة على ذلك وكما أكد تحليل الميتا¹¹ فقد ارتبطت مستويات PFAS بنمط أيضي غير صحي يتميز بزيادة مستوى "الأحماض الدهنية" و "البروتينات الدهنية منخفضة الكثافة" ومستويات "البروتين الشحمي (ب)"¹². وكانت هذه الارتباطات متشابهة بالنمط الأيضي الذي يزيد خطر الإصابة بأمراض القلب والأوعية الدموية. ووجدنا أيضاً أن هذا النمط شديد بشكل خاص لدى الأفراد الأصغر سناً. هذه النتائج تشير إلى أن PFAS يمثل خطراً صحياً حتى في الأفراد الذين لم يتعرضوا على ما يبدو لمستويات عالية من PFAS.

¹ "OMICS" هو مجال في العلم الحديث الذي يتطلع إلى الفهم التام للحياة والعمليات الحيوية عبر التحليل الجزيئي للعناصر المختلفة التي تشكل الحياة، مثل الجينات والبروتينات والأملاح والمركبات الكيميائية الأخرى حيث يتم تحليل هذه العناصر عبر عدة أنماط "أوميكس" مختلفة، مثل "جينوميكس" (تحليل الجينات) و "بروتيوميكس" (تحليل البروتينات). يستخدم هذا المجال في التحقيقات العلمية لفهم أفضل للعمليات الحيوية وللغثور على حلول للمشاكل الصحية المختلفة.

² Venous Thrombosis

³ The Netherlands Epidemiology of Obesity Study (NEO)

⁴ تم إنشاء دراسة INTERVAL بالتعاون بين جامعي كامبريدج وأوكسفورد والتعاون أيضاً مع دائرة الدم والتبرعات الصحية في المملكة المتحدة

⁵ Cytosine-adenine-guanine nucleotide sequences

⁶ Huntingtin gene

⁷ Huntington's disease

⁸ Multilevel mixed-effects model

⁹ Mediation analysis

¹⁰ Rhineland study

¹¹ Meta analysis

¹² Apolipoprotein B

