中文題目：人工智慧解密器官溝通：找尋糖尿病腎臟病變中的跨組學生物標記與分子機轉

英文題目：Artificial intelligence-assisted discovery of trans-omics biomarker signature that reveals molecular mechanism involving inter-organ communication in diabetic kidney disease

作　　者：吳逸文[1]，蔡宗憲[2]，駱啟仁[3]，蔡亭芬[4]， 司徒惠康[4]

服務單位：[1]基隆長庚紀念醫院腎臟科，[2]宏碁股份有限公司，[3]長庚大學代謝體學實驗室，[4]國家衛生研究院

Background: Diabetic kidney disease (DKD) is the leading cause of end-stage kidney disease worldwide; however, the integration of high-dimensional trans-omics data to predict DKD is rare.

Method: We developed artificial intelligence (AI)-assisted models using machine learning (ML) algorithms to identify a biomarker signature that predisposes high risk patients with diabetes mellitus (DM) to DKD based on 71 clinical indices, 13231 untargeted metabolomes, 147 lipidomes and 392885 genome-wide single nucleotide polymorphisms genotyping from 618 subjects. The models were further validated in an independent external cohort of 178 subjects.

Results: Participants were spit into training (557 subjects) and testing cohort (61 subjects), respectively. Of the subjects, 338 subjects (54.7%) were controls, 106 (17.2%) had type 2 DM, 73 (11.8%) had non-diabetic CKD, and 101 subjects (16.3%) had DKD. The mean age of the study population was 63.8±12.9 years old and included 287 males (46.4%). The median eGFR was 83.0 mL/min/1.73 m^2. The external validation cohort consisted of control (100), DM (26), non-diabetic CKD (22), and DKD (30patients) , respectively; with a mean age of 60.6 years old and a mean eGFR of 84.1 mL/min/1.73 m^2. Firstly, we conducted analysis of the performance and feature selection of the trans-omics biomarker associated with four groups of subjects. Three models were developed. In model 1, the top 20 features selected by AI gave an accuracy rate of 0.83 and an area under curve (AUC) of 0.89 when differentiating DM and non-DM individuals. In model 2, among DM patients, a biomarker signature of 10 AI-selected features gave an accuracy rate of 0.70 and an AUC of 0.76 when identifying subjects at high risk of renal impairment. In model 3, among non-DM patients, a biomarker signature of 25 AI-selected features gave an accuracy rate of 0.82 and an AUC of 0.76 when pinpointing subjects at high risk of chronic kidney disease (CKD). In addition, the performance of the three models was

rigorously verified using an independent validation cohort obtaining similar results. Second, we performed combination of statistical significant interaction features selected from previous step and also “type 2 diabetes” as a label, then ranked them by summation of the selected counts using 100-time bootstrapped random samples and the three ML methods (Random Forest, SVM and LASSO). Subsequently, 33 features were needed to give an AUC of performance of 0.792 and accuracy rate of 0.721. Third, we used 10-fold cross validation of this interaction model using the following ML-algorithms: Extra-Tree, Random Forest, SVM, Logistic Regression and Extreme Gradient Boosting. The Extra-Tree yield the best accuracy rate (0.76) and AUC (0.81). Intriguingly, analysis of the protein-protein interaction network of the genes containing the identified SNPs (RPTOR, CLPTM1L, ALDH1L1, LY6D, PCDH9, B3GNTL1, CDS1, ADCYAP and FAM53A) revealed that, at the molecular level, there seems to be interconnected factors that have an effect on the progression of renal impairment among DM patients.

Finally, multiplication of two of the specific interaction features, originated from different organs, enhances the effectiveness when distinguishing CKD and non-CKD patients. For example, the representative plots for interactions of Kynurenine*Alanine, asymmetric dimethylarginine (ADMA)*Age, Citrulline*Kynurenine, and Serine*LysoPC a C28:1 result in a more dramatic difference than either one of the above when used separately.

Conclusion: Our findings reveal the potential of employing ML algorithms to augment traditional methods and suggest a possible organ-crosstalk that is potentially involved in the interaction features that may have an impact on the interconnected pathways and/or functional networks linking type 2 diabetes with renal progression.