Discovery of a Biomarker Signature That Reveals a Molecular Mechanism Underlying Diabetic Kidney Disease via Organ Cross Talk

Diabetic kidney disease (DKD) is the leading cause of end-stage kidney disease, and it creates tremendous medical care costs. Accurate prediction of DKD and its potential molecular implications remain incompletely understood. Here, we apply artificial intelligence (AI) algorithms to build up an interaction model that tackles the complex interconnections between diabetes and chronic kidney disease (CKD) and to identify a biomarker signature that predisposes high-risk type 2 diabetes patients to progression to DKD. The cohort in this study contains 618 subjects, and these can be split into training (557 subjects) and testing (61 subjects) cohorts. Their mean age was 63.8 ± 12.9 years, and they included 287 males (46.4%). The median estimated glomerular filtration rate was 83.0 mL/min/1.73 m². Of the subjects, 338 (54.7%) were control subjects, 112 (18.1%) had type 2 diabetes, 73 (11.8%) had nondiabetic CKD, and 95 (15.4%) had DKD (Fig. 1A). The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Institutional Review Board of Chang Gung Medical Foundation (Institutional Review Board no. 201800802B0, 202000077B0A3, 201800273B0C602, and 202002535B0). Informed consent was obtained from all subjects involved in the study.

The interaction model uses the diabetes label as a feature, together with a combination of statistically significant features; this was done by integrating high-dimensional data. This information was collected from 71 clinical indices, using untargeted metabolomics (13,231 metabolites), using lipidomics (P180 metabolites), and by genome-wide single nucleotide polymorphism analysis (392,885 single nucleotide polymorphisms) data sets. The features were ranked by summation of the selected counts using 100-times-bootstrapped random samples and three machine learning methods (random forest, support vector machine, and least absolute shrinkage and selection operator) (1). Subsequently, the minimum features needed to give the highest area under the curve performances and accuracy rates were extracted (Fig. 1B). Finally, we performed 10-fold cross-validation of this model. The top 33 features (Fig. 1C) yield a good accuracy rate (0.76) and area under curve (0.81) when differentiating CKD and non-CKD among patients with diabetes. Intriguingly, multiplication of two of the specific interaction features enhances the effectiveness when distinguishing CKD and non-CKD patients. For example, the representative plots for the interactions of kynurenine (KYN)*alanine, asymmetric dimethylarginine (ADMA)*age, citrulline*KYN, and serine*lysophosphatidylcholine acyl C28:1 (LysoPC a C28:1) result in a more drastic difference than any one of the above when used separately (Fig. 1D).

Figure 1E depicts the interorgan communication that is potentially involved in the interaction features identified by our AI-based methods. In healthy individuals (Fig. 1E, left), the metabolites associated with the interaction features are processed mainly in the liver and kidney. Briefly, in the liver, tryptophan is...
Figure 1—An interaction model was built to tackle the complex interconnections between diabetes and CKD and to identify a biomarker signature that predisposes high-risk diabetes patients to DKD. A: The workflow to predict the occurrence of DKD among patients with diabetes. B: The numbers of features were determined by area under curve and accuracy rate. C: The top 33 features selected by the interaction model for predicting DKD. D: Representative interaction feature plots for CKD and non-CKD. Ranking of the interaction features: KYN*alanine (rank 1), ADMA*age (rank 4), citrulline*kynurenine (rank 6), and serine*LysoPC a C28:1 (rank 8). E: Graphic summary illustrating the interaction features of metabolites in healthy individuals and DKD patients. In healthy individuals, the metabolites (citrulline, KYN, and lysoPC a C28:1) are processed in the liver and excreted by the kidney. In DKD patients, all the AI-identified interaction features of metabolites are dysregulated in the liver, the blood, and the kidney, leading to an elevated level of reactive oxygen species and an increase of inflammatory response. Together, these abnormalities accelerate
the progression of renal impairment in patients with diabetes. AUC, area under curve; DG, diglyceride; DM, diabetes mellitus; EC-SOD, extracellular superoxide dismutase; IFN-γ, interferon-γ; IL-6, interleukin-6; KA, KYN acid; PC, diacyl-phosphatidylcholines; PC ae, acyl-alkyl-phosphatidylcholines; ROS, reactive oxygen species; SNP, single nucleotide polymorphism; TNF-α, tumor necrosis factor alpha; TRP, tryptophan. The figure was created with BioRender.com.

converted into KYN and delivered to the kidney via the circulation; 2) arginine is converted into citrulline via the urea cycle or, alternatively, is processed into ADMA, with both citrulline and ADMA being delivered to the kidneys; 3) phosphatidylcholine is converted into LysoPC a C28:1, which can stimulate the expression of extracellular superoxide dismutase in hepatocytes and in the endothelial cells of blood vesicles, thereby enhancing the antioxidant system of these cells; and 4) all of the above metabolites (KYN, citrulline, and ADMA) are excreted from the kidney.

In contrast, the metabolites associated with the interaction features are dysregulated in DKD patients (Fig. 1E, right) as follows. 1) An elevated level of KYN accumulates in the blood; this is probably caused, at least in part, by impaired excretion by the damaged kidneys. Additionally, the proinflammatory milieu of diabetes can activate immune cells that also can convert tryptophan into KYN, further increasing the levels of KYN (2) and creating a vicious cycle.

2) Elevated levels of ADMA and citrulline can be detected in the blood of DKD patients; this may be caused by an impairment of the urea cycle and arginine metabolism in the liver. Accumulation of ADMA can in turn cause renal dysfunction when there is hyperglycemia, while a high level of citrulline may correlate with cardiac dysfunction (3).

3) The decline in LysoPC is likely to result in decreased expression of extracellular superoxide dismutase (4); this may compromise the antioxidant system and lead to elevated levels of reactive oxygen species aggravating renal damage. The increase in alanine and cysteine in the blood may be attributable to the protein degradation that accompanies muscle atrophy, which is a consequence of type 2 diabetes (5).

It will be of great interest to validate our findings using larger-scale ethnically diverse cohorts. Additionally, longitudinal studies are warranted to validate the usefulness of our model. Finally, the cause-and-effect relationships and the sources of the circulating biomarkers need to be further investigated.

In conclusion, AI-assisted discovery of the biomarker signature reveals a potential molecular mechanism underlying the complex interorgan communication occurring during DKD pathogenesis. This signature, which consists of risk (or predictive) biomarkers, may provide novel diagnostic or therapeutic insights (clinical trial reg. no. NCT04839796, ClinicalTrials.gov).

**Acknowledgments.** The authors thank Professor Kung-Yee Liang (Institute of Population Health Science, National Health Research Institutes) for his conceptual input. The authors thank Chun-Hsien Li (Advanced Tech BU, Acer Inc.), Yun-Hsuan Chan (Advanced Tech BU, Acer Inc.), and Jun-Hong Chen (Advanced Tech BU, Acer Inc.) for their technical assistance with the AI analyses. The authors thank the following physicians and scientists from Keelung Chang Gung Memorial Hospital for their contributions in subject recruitment: Dr. Heng-Chih Pan, Dr. Heng-Jung Hsu, Dr. Chun-Yu Chen, Dr. Chin-Chan Lee, Dr. Yu-Chiau Shyu, and Dr. Chih-Lang Lin. We also acknowledge the participant and sample preservation carried out as part of the Northeastern Taiwan Community Medicine Research Cohort Study.

**Funding.** This research was funded by grants from the Ministry of Health and Welfare (Smart Healthcare for Obesity Therapeutics, PD-109-GP-02 and MG-110-GP-03, to H.-K.S.) and from Chang Gung Memorial Hospital (CMRPG2H0121-124 to I.-W.W., CORPG2H0041-0043 and CMRPG2H000091-0093 to C.-H.Y., and CMRPG2K0141-142 to C.-C.L.).

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Author Contributions.** All authors contributed to manuscript preparation. C.-C.L., H.-K.S., and T.-F.T. co-designed the topics and research framework. I.-W.W. recruited subjects and defined the clinical stages. I.-W.W., Y.-J.C., and C.-H.Y. prepared the figures and drafted the manuscript. T.-H.T. contributed to AI analyses. C.-J.L. and M.-L.C. contributed to metabolomics analyses. T.-F.T. wrote the final version of the manuscript. All authors have read and agreed to the published version of the manuscript. C.-C.L., H.-K.S., and T.-F.T. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**References**
1. Yang NI, Yeh CH, Tsai TH, et al. Artificial intelligence-assisted identification of genetic factors predisposing high-risk individuals to asymptomatic heart failure. Cells 2021;10:10
2. Westbrook R, Chung T, Lovett J, et al. Kynurenines link chronic inflammation to functional decline and physical frailty. JCI Insight 2020;5:5
3. Finkelman BS, Putt M, Wang T, et al. Arginine-nitric oxide metabolites and cardiovascular disease in patients with breast cancer. J Am Coll Cardiol 2017;70:152–162
4. Law SH, Chan ML, Marathe GK, Parveen F, Chen CH, Ke LY. An updated review of lysophosphatidylcholine metabolism in human diseases. Int J Mol Sci 2019;20:20
5. Perry BD, Caldow MK, Brennen-Speranza TC, et al. Muscle atrophy in patients with type 2 diabetes mellitus: roles of inflammatory pathways, physical activity and exercise. Exerc Immunol Rev 2016;22:94–109