JDI Journal of Diabetes Investigation Open access
Official Journal of the Asian Association for the Study of Diabetes
ORIGINAL ARTICLE

Identification of two novel subgroups in patients with diabetes mellitus and their association with clinical outcomes: A two-step cluster analysis

Xiao-fen Xiong, Yuan Yang, Ling Wei, Ying Xiao, Li Li, Lin Sun*

Department of Nephrology, The Second Xiangya Hospital, Central South University, Changsha, China

Keywords
Classification of diabetes, Cluster analysis, Type 2 diabetes

***Correspondence**
Lin Sun
Tel.: +86-135-7410-0216
Fax: +86-731-8553-3525
E-mail address:
sunlin@csu.edu.cn

J Diabetes Investig 2021; 12: 1346–1358

doi: 10.1111/jdi.13494

ABSTRACT

Aims/Introduction: The aim of this study was to determine whether distinct subphenotypes of patients with type 2 diabetes in the European classification exist in Chinese populations, and to further establish novel subphenotypes more suitable for Chinese populations.

Material and Methods: The research retrospectively analyzed 5414 patients with type 2 diabetes from the National Clinical Research Center for Metabolic Diseases Diabetes Center in China, and a two-step cluster analysis was carried out. First, we confirmed the European classification in Chinese populations by six parameters, including age at disease onset, body mass index, glycosylated hemoglobin, homeostatic model assessment 2 to estimate β-cell function and insulin resistance, and glutamate decarboxylase antibodies. Furthermore, triglycerides and uric acid were added to refine the cluster analysis, and Cox regression was used to evaluate the risk of diabetic complications.

Results: Just three clusters were replicated in our cohort according to Emma Ahlqvist's European classification. When other variables were added to the cluster analysis, seven subgroups were identified, including five clusters of the European classification and two novel subgroups, namely, uric acid-related diabetes and inheritance-related diabetes. Compared with patients with inheritance-related diabetes, patients with severe insulin-resistant diabetes showed a higher risk of diabetic peripheral neuropathy, hypertension and chronic kidney disease, and the uric acid-related diabetes subgroup showed a higher risk of coronary heart disease, cerebral vascular disease and end-stage renal disease. Patients with severe insulin-deficient diabetes showed a higher risk of diabetic retinopathy and diabetic foot than those with inheritance-related diabetes. Furthermore, there were sex-specific associations between subgroups and clinical outcomes. No significant difference was observed in the prevalence of cancer in each subgroup.

Conclusions: Seven subgroups of type 2 diabetes were identified in Chinese populations, with distinct characteristics and disparate clinical outcomes. This etiology-based stratification might contribute to the diagnosis and management of type 2 diabetes.

INTRODUCTION

Globally, approximately one in 11 adults has diabetes mellitus; this prevalence has quadrupled over the past 30 years[1]. Preventive care for diabetes patients, such as effective lifestyle modification, potent social support and satisfactory medication adherence, has substantially improved[2], and the incidence of diabetic complications has significantly reduced, whereas the overall numbers of patients with end-stage renal disease (ESRD), stroke and amputation are still persistently elevated[3]. One explanation is that diabetes mellitus, which is characterized

Received 10 September 2020; revised 18 December 2020; accepted 1 January 2021

by heterogeneity, varies significantly in clinical manifestations and disease progression[4].

Currently, the widely accepted classification of diabetes mellitus was first proposed in 1936[5]. A large proportion of patients with diabetes mellitus (~90%) are distinguished by relatively deficient insulin resulting from insulin resistance and dysfunctional pancreatic β-cells, defined as type 2 diabetes mellitus (type 2 diabetes)[1]. Approximately 10% of patients with diabetes mellitus, characterized by absolute insulin deficiency and resultant hyperglycemia due to autoimmunity, are classified as having type 1 diabetes mellitus. As described in the data, the occurrence of type 1 diabetes in adults is as high as 50%, and 50% of adulthood cases might be misclassified as type 2 diabetes[6]. Furthermore, latent autoimmune diabetes in adults is attributed to autoimmunity, with mild metabolic dysfunction and no requirement of insulin treatment at disease onset, which leads to a misdiagnosis of type 2 diabetes due to its heterogeneous pathophysiology and clinical manifestations[7]. Furthermore, other rare and specified types are monogenic diabetes mellitus, such as maturity-onset diabetes of the young, which is mainly caused by the mutation of HNF4A in p.R114W[8] and neonatal diabetes.

The complexity and heterogeneity of diabetes mellitus, especially of type 2 diabetes, impede precise diagnosis and treatment[9]. Furthermore, the traditional classification defined the phenotypes by age at disease onset and single metabolic measurement, which were far less effective clinical indicators. Thus, the identification of novel subgroups based on pathophysiological, genetic and other risk factors is critical. Recently, Ahlqvist *et al.*[10] established five subclassifications in type 2 diabetes with distinct patient characteristics and heterogeneous diabetes complications based on six parameters (age at disease onset, body mass index [BMI], glycosylated hemoglobin [HbA_1c], homeostatic model assessment 2 to estimate β-cell function [HOMA2-B] and insulin resistance [HOMA2-IR], and glutamate decarboxylase antibodies [GADA]). That study was a notable attempt, and provided considerable information on the risk of disease progression and potential therapeutic strategies for type 2 diabetes[11].

China is the largest country in the diabetes epidemic, and has an estimated prevalence of diabetes and prediabetes of 11.6% and 35.7%, respectively[12]. Due to differences in ethnicities, regions, genetic backgrounds, lifestyles and natural environment, it is unclear whether the European subclassification is applicable to Chinese populations. In addition to the aforementioned variables, other risk factors, such as hypertriglyceridemia, hyperuricemia, smoking and genetic factors, also contribute to the progression of type 2 diabetes[2]. Furthermore, studies from China have implied that among individuals with type 2 diabetes, 63.9% had dyslipidemia[13] and 32.6% had hyperuricemia[14], showing the need to include dyslipidemia and hyperuricemia in a cluster analysis of the Chinese population. In the present study, we initially validated the availability of the European classification in Chinese populations and then remodeled the cluster analysis based on additional parameters to produce a more adjustable classification in Chinese adult-onset diabetes.

MATERIALS AND METHODS

Study population

The study included a retrospective cohort of 5,414 patients from the National Clinical Research Center for Metabolic Diseases Diabetes Center, China, including the Department of Endocrinology and Nephrology of the Second Xiangya Hospital of Central South University, Changsha, China, from January 2010 to November 2018. In this cohort, the average duration of diabetes was 8.6 ± 6.3 years, and cluster analysis based on six parameters (age at disease onset, BMI, HbA_1c, HOMA2-B and HOMA2-IR, and GADA) was used to confirm the utility of the European classification in Chinese populations. Then, parameters including triglycerides (TG) and uric acid (UA) were added to refine the cluster analysis of the European classification with the data of 5011 patients due to missing information for 403 patients. We have evaluated the differences between eligible 5,011 patients and 403 patients in baseline demographic characteristics and laboratory parameters. The results showed that there was no statistical difference between the eligible 5,011 and excluded 403 (data not shown). Cox regression was used to evaluate the risk of disease progression among 4,899 patients (112 patients excluded). The present study was approved by the Hunan Research Ethics Committees in the Second Xiangya Hospital of Central South University, China.

The inclusion criteria were as follows: (i) patients diagnosed with type 2 diabetes; and (ii) patients whose age of disease onset was older than 18 years. The exclusion criteria were as follows: (i) patients diagnosed with type 1 diabetes (n = 556); (ii) latent autoimmune diabetes in adults (n = 118); (iii) patients with newly diagnosed diabetes mellitus (n = 578); and (iv) patients with missing data for clustered variables (n = 403).

Measurements

An ADVIA 2120 automated hematology analyzer (Siemens Healthcare Diagnostics, Erlangen, Germany) was used to measure the level of hemoglobin. The blood biochemical indexes were carried out with standard automated enzymatic methods (Hitachi 912 automated analyzer, Boehringer Mannheim, Mannheim, Germany), including TG, total cholesterol (TC), high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, albumin (Alb), creatinine (Cr), fasting glucose and UA. Furthermore, HbA_1c was analyzed by high-performance liquid chromatography (VARIANT-II Hemoglobin Testing System; Bio-Rad, Hercules, CA, USA). Proteinuria evaluation, diagnosed as the urine albumin excretion rate ≥30 mg/24 h, was carried out by the immunoturbidimetric method. The Modular Cobas E601 Analyzer or ADVIA Centaur XP Immunoassay System was applied to measure the level of 25-hydroxyvitamin D concentrations. Fasting C-peptide was

analyzed by standard methods, and GADA was measured by radioligand assay.

Definitions

Type 2 diabetes was diagnosed by the following criteria: fasting plasma glucose level ≥7.0 mmol/L (126 mg/dL); random plasma glucose ≥11.1 mmol/L (200 mg/dL); or 2-h plasma glucose level during an oral glucose tolerance test ≥11.1 mmol/L; or $HbA_1c \geq 6.5\%$. BMI was calculated by weight (kg) / height (m^2). The calculation of HOMA2 was carried out by fasting plasma glucose and c-peptide. The estimated glomerular filtration rate (eGFR) was obtained by the formula of the Modification of Diet in Renal Disease study. Drinkers or smokers were referred to as those who engaged in drinking or smoking daily for more than 1 year. Waist-to-hip ratio was measured by the formula of waist circumference / hip circumference. A family history of diabetes was referred to as a first-degree family member (children, parents and siblings) who had or did not have diabetes. Drug use was defined as taking medications regularly for 3 months.

Hypertension (HTN) was defined as having a systolic blood pressure (SBP) ≥130 mmHg or a diastolic blood pressure ≥80 mmHg for two or more readings on two or more occasions or currently receiving antihypertensive treatment. Coronary heart disease (CHD) was referred to as having a history of angina and/or myocardial infarction. The definition of cerebral vascular disease (CVD) was cerebral dysfunction caused by cerebrovascular disease, such as cerebral hemorrhage and cerebral infarction. The criteria for chronic kidney disease (CKD) were chronic structural and functional impairment of the kidney for more than 3 months. The criterion for ESRD was an eGFR ≤15 mL/min per 1.73 m^2.

Diabetic retinopathy (DR) was characterized by fundus photographs of lesions of varying degrees, such as microaneurysms, hemorrhages and new vessel formation, eventually transforming into retinal thickening. Diabetic foot was defined as lower limb infection, ulceration and/or destruction of deep tissue in diabetes patients caused by the combination of neuropathy and various degrees of peripheral vascular lesions. The diagnosis of diabetic peripheral neuropathy (DPN) mainly relied on related clinical symptoms, and neurological and electrophysiological investigations.

Statistical analysis

Two-step clustering is an intelligent clustering method in which categorical and continuous variables can be simultaneously addressed, and in which the optimal clustering number is automatically determined. It identifies clusters by two processes: first, preclustering, followed by hierarchical clustering. Hierarchical algorithms were used to estimate the optimal clustering number based on the silhouette width, the calculation of the distance using the log-likelihood and clustering in accordance with Schwarz's Bayesian criterion.

Cox regression was applied to calculate the risk of complications after adjustments were made for age at diagnosis, sex, SBP, smoking habit, drinking habit, Alb, eGFR and BMI. The time at which the patient was first diagnosed with type 2 diabetes was defined as the starting time, and the time at which the patient was diagnosed with complications was defined as the ending time. The interval was considered to be the duration of diabetes. The timing of diabetic complications and comorbidities was confirmed by a review of the electronic medical records. SPSS version 24.0 was used, and a *P*-value of <0.05 was considered statistically significant.

RESULTS

Validating the applicability of the European classification in Chinese populations with type 2 diabetes

Two-step cluster analysis was carried out in patients with type 2 diabetes, including 3,087 men and 2,327 women, with a focus on six variables (GADA, age at diagnosis, BMI, HbA_{1c}, HOMA2-B and HOMA2-IR). Compared with European subgroups, three clusters were identified in both men and women (Figure 1, Figure 2), which were cluster 1 (severe autoimmune diabetes [SAID]), cluster 2 (severe insulin-deficient diabetes [SIDD]) and cluster 3 (mild age-related diabetes [MARD]). The cluster centers are shown in Table S2. Interestingly, the proportions of SIDD and MARD were 1,506 (48.8%) and 1,474 (47.7%), respectively, in men (Figure 1a), and 726 (31.2%) and 1,510 (64.9%), respectively, in women (Figure 2a). Additionally, the characteristics of SAID in men were the same as those in women, except for β-cell function (Figure 1e). Characteristics, including age, BMI, HbA_{1c}, HOMA2-B and HOMA-IR, are shown in Figure 1b–f and Figure 2b–f. In addition, severe insulin-resistant diabetes (SIRD) and mild obesity-related diabetes (MOD) were not observed.

Identification of seven clusters by remodeling the cluster analysis based on eight variables

The clustered results based on eight variables were shown as 7 subgroups (Figure 3). Cluster centers are described in Table S2. In subgroup 1 (SIRD), 95/5011 (1.9%) patients had insulin resistance (Figure 3a), poor metabolic control with higher TG (Figure 3g) and UA (Figure 3h). In subgroup 2 (SIDD), 999/5011 (19.9%) patients were characterized as having severe deficiency of insulin (Figure 3a), poor control of HbA_1c levels (Figure 3d) but normal levels of lipid profiles (Figure 3g) and UA (Figure 3h). In subgroup 3 (MOD), 859 of 5,011 (17.1%) patients showed obesity (Figure 3a), early-onset disease (Figure 3c) and relatively poor control of HbA_1c (Figure 3d), TG (Figure 3g) and UA levels (Figure 3h), whereas they did not have insulin resistance (Figure 3f). In subgroup 4 (SAID), 128 of 5,011 (2.6%) patients showed the presence of GADA and lower BMI (Figure 3b). In subgroup 5 (named UA-related diabetes [UARD]), which was a distinct subgroup, 618 of 5,011 (12.3%) patients were shown as having the highest UA level (Figure 3g), accompanied by mild insulin

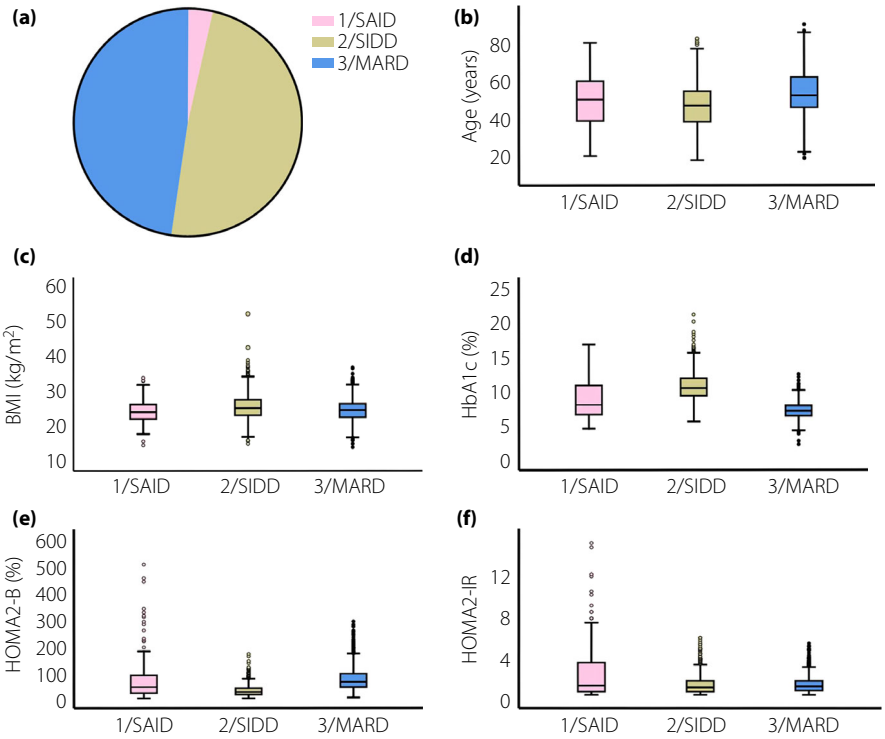

Figure 1 | Participant distributions and characteristics in Chinese men with type 2 diabetes according to the European classification. (a) Distributions of Chinese men (n = 3087) according to the two-step cluster analysis. (b) Distributions of age at diagnosis for each cluster. (c) Distributions of body mass index (BMI) for each cluster. (d) Distributions of glycosylated hemoglobin (HbA_1c) for each cluster. (e) Distributions of homeostatic model assessment 2 to estimate β-cell function (HOMA2-B) for each cluster. (f) Distributions of homeostatic model assessment 2 to estimate insulin resistance (HOMA2-IR) for each cluster. HOMA2-B and HOMA2-IR were applied to estimate the function of β-cell and insulin resistance, respectively, which were calculated using fasting plasma glucose and c-peptide. BMI was calculated as weight (kg) / height (m^2). MARD, mild age-related diabetes; SAID, severe autoimmune diabetes; SIDD, severe insulin-deficient diabetes.

resistance (Figure 3f) and good control of HbA_1c (Figure 3d), superior β-cell function (Figure 3e) and older-onset disease (Figure 3c). In subgroup 6 (MARD), 1,236 of 5,011 (24.7%) patients were labeled as having senile-onset disease (Figure 3c) and mild disturbance in levels of TG (Figure 3g). In subgroup 7 (called inheritance-related diabetes [IRD]), 1,076 of 5,011 (21.5%) patients were shown to have the highest proportion of family history of diabetes (Table 1), and moderate levels of TG, UA and insulin resistance compared with patients in other subgroups.

Other traits of seven classifications are shown in Table 1. Cluster 1 (severe insulin-resistant diabetes [SIRD]) was shown to have severe lipid metabolic dysregulation accompanied by a substantial elevation of TC (7.1 ± 3.1 mmol/L), but a significant reduction in high-density lipoprotein cholesterol (0.9 ± 0.7 mmol/L) and a major decrease in vitamin D (23.8 ± 9.5 nmol/L). The proportions of men and drinking were highest in cluster 3, accounting for 65.7% and 26%, respectively. Cluster 5 (UARD) was shown to have higher SBP, lower hemoglobin and Alb, and worse renal function, with higher levels of Cr and proteinuria. The proportion of family history was lowest in cluster 6 (MARD; 24.7%) and highest in cluster 7 (IRD; 46.5%).

Risk evaluation of diabetic complications in seven clusters

Compared with the patients in cluster 7 (IRD), the patients in cluster 1 (SIRD) showed a higher risk of DR (HR 1.3390, 95% confidence interval [CI] 0.919–1.95, P = 0.128; Figure 4a, Table S3), DPN (HR 1.711, 95% CI 1.282–2.283, $P = 2 \times 10^{-4}$; Figure 4b, Table S3), HTN (HR 1.835, 95% CI 1.388–2.425, $P = 2 \times 10^{-5}$; Figure 4f, Table S4) and CKD (HR 2.37, 95% CI 1.683–3.336, $P = 7.6 \times 10^{-7}$; Figure 4d, Table S6) after adjustments were made for the confounding factors of age at diagnosis, sex, SBP, smoking habit, drinking habit, Alb, eGFR and BMI. For CHD (Figure 4g, Table S4), CVD (Figure 4h, Table S4) and ESRD (Figure 4e, Table S6), the adjusted risk in cluster 5 (UARD) increased to, respectively, 1.526- (95% CI 1.184–1.966, P = 0.001), 1.487- (95% CI 1.089–2.03, P = 0.013) and 5.002-fold (95% CI 1.886–6.894, $P = 7.4 \times 10^{-5}$) higher than that in cluster 7 (IRD). Patients in cluster 2 (SIDD) showed a much higher risk of DR (HR 1.229, 95% CI

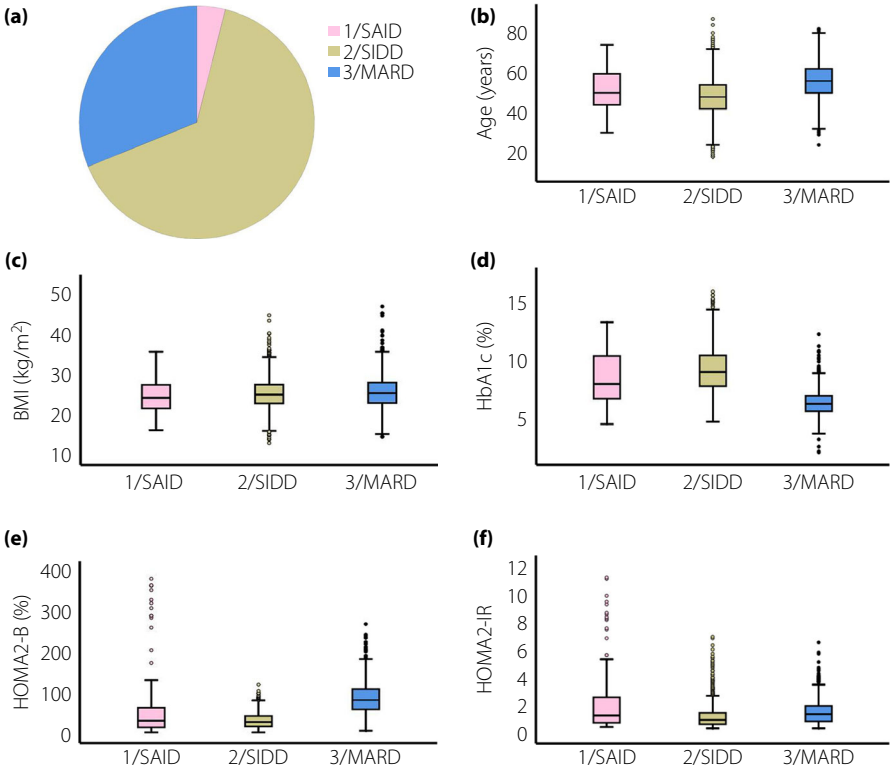

Figure 2 | Participant distributions and characteristics in Chinese women with type 2 diabetes according to the European classification. (a) Distributions of Chinese women (n = 2327) according to the two-step cluster analysis. (b) Distributions of age at diagnosis for each cluster. (c) Distributions of body mass index (BMI) for each cluster. (d) Distributions of glycated hemoglobin (HbA_1c) for each cluster. (e) Distributions of homeostatic model assessment 2 to estimate β-cell function (HOMA2-B) for each cluster. (f) Distributions of homeostatic model assessment 2 to estimate insulin resistance (HOMA2-IR) for each cluster.

1.045–1.446, P = 0.013) and diabetic foot (HR 2.051, 95% CI 1.405–2.994, $P = 1.96 \times 10^{-4}$) than those in cluster 7 (IRD; Figure 4c, Table S3). Furthermore, the prevalence of cancer (Figure 4i, Table S5) in each cluster was not significantly different ($P > 0.05$).

There were sex-specific associations between subgroups and clinical outcomes. The men with SIRD showed relatively higher risk factors in DR (HR 1.733, 95% CI 1.016–2.957, P = 0.044; Figure 5a, Table S7), CKD (HR 3.225, 95% CI 2.074-5.016, $P = 2.03 \times 10^{-7}$; Figure 5d, Table S7), whereas women with SIDD showed higher risk factors in DR (HR 1.293, 95% CI 1.037–1.613, P = 0.022; Figure 6a, Table S7), DPN (HR 1.331, 95% CI 1.108–1.598, P = 0.002; Figure 6b, Table S7) and CKD (HR 1.517, 95% CI 1.171–1.966, P = 0.002; Figure 6d, Table S7), compared with those in cluster 7 (IRD). Additionally, compared with the patients in cluster 7 (IRD), men in cluster 5 (UARD) had a higher risk of DR (HR 1.258, 95% CI 1.007–1.641, P = 0.044; Figure 5a, Table S7), DPN (HR 1.237, 95% CI 1.009–1.517, P = 0.041; Figure 5b, Table S7), CHD (HR 1.584, 95% CI 1.108–2.266, P = 0.012; Figure 5g, Table S7) and CKD (HR 1.412, 95% CI 1.104–1.805, P = 0.006; Figure 5g, Table S7).

DISCUSSION

In the present study, we found that just three subphenotypes of diabetes (SAID, SIDD, MARD) described previously by Ahlqvist *et al.* were identified in Chinese populations when identical clinical parameters were used. Two other classifications (SIRD and MOD) were not identified. These findings suggest that the European classification is relatively heterogeneous among various ethnicities, regions, genetic backgrounds, lifestyles and natural environments. When TG and UA were added to refine our cluster analysis, seven subgroups were stratified, five of which were duplicated (SAID, SIDD, MARD, SIRD and MOD) and two of which were more distinct classifications (UARD and IRD). Ultimately, patients in the seven clusters responded differently to disease progression; for example, patients in cluster 5 (UARD) had a higher risk of CHD, CVD and ESRD, whereas patients in cluster 1 (SIRD) had increasing hazards of DPN, HTN and CKD, and there were sex-specific associations between subgroups and clinical outcomes Those findings indicated that the selection of different variables might reflect disparate risk stratification.

Ahlqvist *et al.* proposed a notable classification to predict the risk of related complications and paved the way for

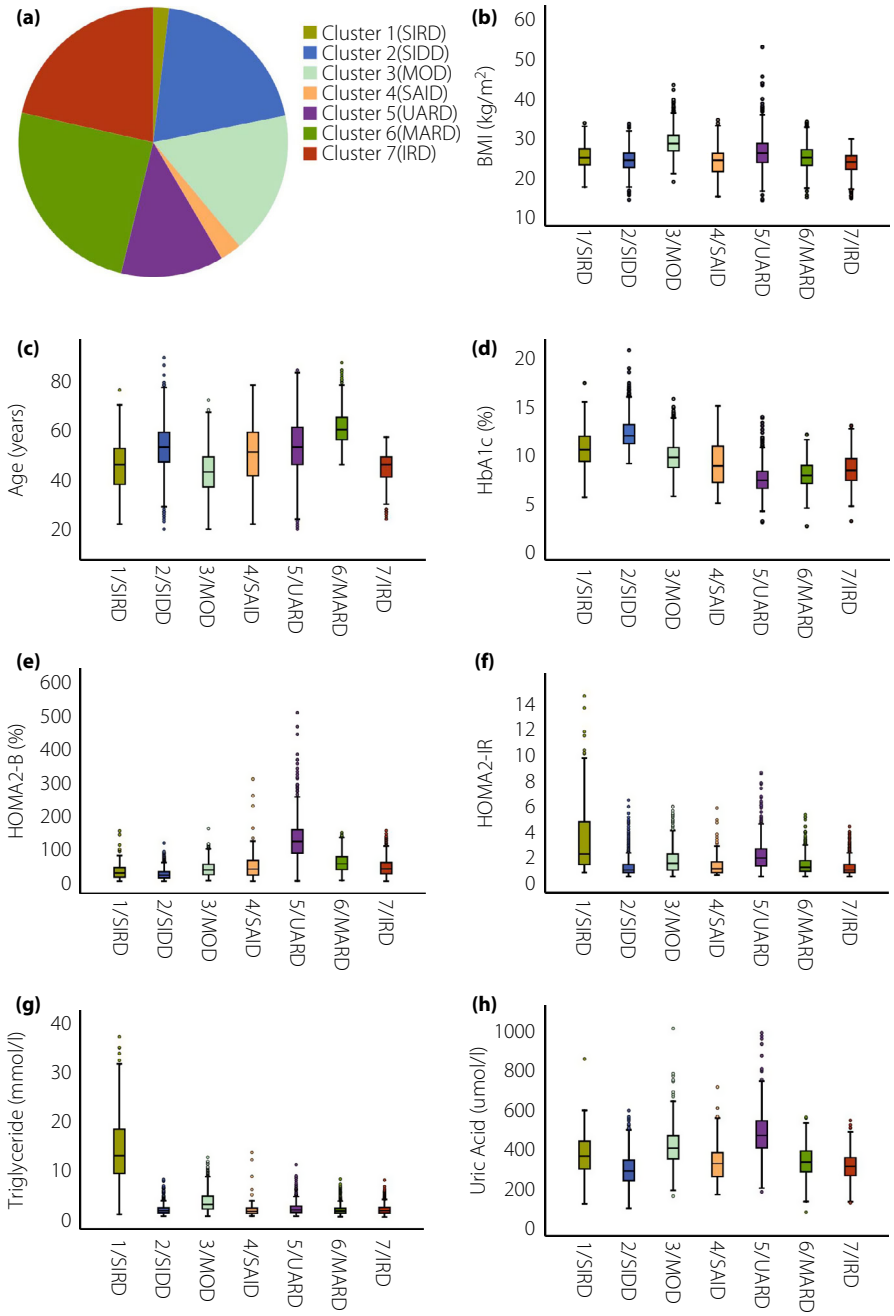

Figure 3 | Participant distributions and characteristics of seven classifications in the Chinese population with type 2 diabetes. (a) Distributions of the Chinese population (n = 5011) according to the two-step cluster analysis. (b) Distributions of body mass index (BMI) for each cluster. (c) Distributions of age at diagnosis for each cluster. (d) Distributions of glycosylated hemoglobin (HbA_1c) for each cluster. (e) Distributions of homeostatic model assessment 2 to estimate β-cell function (HOMA2-B) for each cluster. (f) Distributions of homeostatic model assessment 2 to estimate insulin resistance (HOMA2-IR) for each cluster. (g) Distributions of triglycerides for each cluster. (h) Distributions of uric acid for each cluster. IRD, inheritance-related diabetes; MOD, mild obesity-related diabetes; SIRD, severe insulin-resistant diabetes; UARD, UA-related diabetes.

personalized medicine in different subgroups of diabetes[11]. Although their classification was identified as limited to European populations, other risk factors, such as history of drinking or smoking, blood pressure, lipid profiles, UA, inflammatory biomarkers and genetic factors, have not been investigated[15]. The addition of those risk factors to the cluster analysis might be more beneficial to the classification and management of diabetes.

Table 1 | Participant characteristics of seven classifications in Chinese population with type 2 diabetes

	1/SIRD (*n* = 95)	2/SIDD (*n* = 999)	3/MOD (*n* = 859)	4/SAID (*n* = 128)	5/UARD (*n* = 618)	6/MARD (*n* = 1236)	7/IRD (*n* = 1076)
Men (%)	52 (54.7%)	521 (52.2%)	564 (65.7%)	70 (54.7%)	378 (61.2%)	622 (50.3%)	562 (52.2%)
Smoking (%)	33 (34.7%)	271 (27.1%)	278 (32.4%)	38 (29.7%)	165 (26.7%)	251 (20.3%)	262 (24.3%)
Drinking (%)	23 (24.2%)	199 (19.9%)	223 (26%)	22 (17.2%)	117 (18.9%)	189 (15.3%)	202 (18.8%)
Family history of diabetes (%)	31 (32.6%)	337 (33.7%)	338 (39.4%)	42 (32.8%)	176 (28.5%)	306 (24.7%)	500 (46.5%)
GADA (%)	0	0	0	128 (100%)	0	0	0
Duration (years)	6.8 ± 5.3	8.2 ± 5.9	9.5 ± 6.4	8.7 ± 6	9.3 ± 6.2	8.2 ± 5.5	10.3 ± 8.8
Age (years)	44.2 ± 11.6	50.7 ± 10.0	40.8 ± 9.1	48.4 ± 11.8	51.4 ± 10.8	59 ± 6.8	42.6 ± 6.1
SBP (mmHg)	138.1 ± 21.2	135.3 ± 20.5	138.4 ± 19.1	134.4 ± 18.4	143.3 ± 22.8	140 ± 19.8	136.0 ± 21.3
DBP (mmHg)	83.5 ± 13.8	80.3 ± 11.5	84.7 ± 11.9	80.4 ± 12.3	81.0 ± 12.6	78.8 ± 11.4	81.3 ± 11.6
BMI (kg/m^2)	24.7 ± 3.4	23.7 ± 2.8	28.2 ± 3.1	23.7 ± 3.8	25.9 ± 4.3	24.5 ± 2.9	23.0 ± 2.5
WHR	1.0 ± 0.06	0.9 ± 0.07	1.0 ± 0.06	0.9 ± 0.07	1.0 ± 0.07	0.9 ± 0.07	0.9 ± 0.06
Hb (g/L)	130.6 ± 26.1	131.4 ± 18.6	134.2 ± 21.5	122.8 ± 20.2	115.5 ± 24.5	124.1 ± 18.9	127.1 ± 20.5
TG (mmol/L)	14.6 ± 10.9	1.7 ± 1.0	3.5 ± 2.2	1.8 ± 1.7	2.1 ± 1.3	1.7 ± 1.0	1.8 ± 1.0
TC (mmol/L)	7.1 ± 3.1	4.5 ± 1.1	4.8 ± 1.2	4.2 ± 1.1	4.3 ± 1.3	4.2 ± 1.0	4.3 ± 1.1
HDL (mmol/L)	0.9 ± 0.7	1.0 ± 0.3	0.9 ± 0.3	1.0 ± 0.3	1.0 ± 0.3	1.0 ± 0.3	1.1 ± 0.3
LDL (mmol/L)	2.4 ± 1.6	2.9 ± 1.0	3.0 ± 1.0	2.5 ± 0.9	2.7 ± 1.0	2.6 ± 0.9	2.7 ± 1.0
Alb (g/L)	37.3 ± 5.6	35.4 ± 4.8	37.1 ± 5.1	36.6 ± 4.6	34.9 ± 5.7	36.6 ± 4.4	36.8 ± 5.2
HbA_{1C}(%)	10.1 ± 2.2	11.7 ± 1.5	9.3 ± 1.6	8.7 ± 2.3	7.1 ± 1.4	7.5 ± 1.2	8.0 ± 1.4
HOMA2-B (%)	35.3 ± 28.8	25.1 ± 15.9	40.7 ± 22.5	49.6 ± 45	129.0 ± 66.6	58.7 ± 27.7	44.8 ± 25.3
HOMA2-IR	5.5 ± 10.7	1.1 ± 0.7	1.6 ± 0.9	1.2 ± 0.9	2.1 ± 1.2	1.3 ± 0.7	1.1 ± 0.6
Cr (umol/L)	84.0 ± 59.3	66.0 ± 26.6	88.0 ± 61.0	77.9 ± 53.6	173.0 ± 152.2	80.8 ± 49.1	77.8 ± 60.6
eGFR (mL/min/1.73 m^2)	138.1 ± 76.6	140.7 ± 49.1	124.8 ± 56.1	134.4 ± 57.5	72.8 ± 46.5	113.6 ± 39.9	135.4 ± 53.2
UA (umol/L)	341.3 ± 111.0	263.3 ± 73.9	372.5 ± 86.4	300.6 ± 96.7	436.7 ± 108.0	303.6 ± 71.9	280.5 ± 65.2
Proteinuria (mg/day)	172.3	130.7	180.2	125.6	321.3	120.5	126.4
Vitamin D (nmol/L)	23.8 ± 9.5	39.5 ± 18.3	35.3 ± 15.2	43.6 ± 18.9	38.5 ± 19.6	41.4 ± 18.6	41.5 ± 18.3
ACEI (%)	17 (17.9%)	174 (17.4%)	175 (20.4%)	12 (9.4%)	101 (16.3%)	236 (19.1%)	168 (15.6%)
ARB (%)	26 (27.4%)	228 (22.8%)	261 (30.4%)	29 (22.7%)	173 (28.0%)	339 (27.4%)	222 (20.6%)
Lipid-lowering drugs	83 (87.4%)	766 (76.6%)	703 (81.8%)	89 (69.5%)	453 (73.3%)	908 (73.5%)	727 (67.6%)
OHA	71 (74.7%)	756 (75.6%)	724 (84.3%)	83 (64.8%)	377 (61.0%)	934 (75.6%)	786 (73.0%)
Insulin	83 (87.4%)	881 (88.1%)	637 (74.2%)	92 (71.9%)	326 (52.8%)	697 (56.4%)	741 (68.9%)

Homoeostatic model assessment 2-B (HOMA2-B) and homoeostatic model assessment 2-insulin resistance (HOMA2-IR) were applied to estimate the function of β-cell and insulin resistance, respectively, which were calculated using fasting plasma glucose and c-peptide. ACEI, angiotensin-converting enzyme inhibitor; Alb, albumin; ARB, angiotensin receptor blocker; BMI, body mass index; DBP, diastolic blood pressure; Cr, creatinine; eGFR, estimated glomerular filtration rate; GADA, glutamate decarboxylase antibodies; Hb, hemoglobin; HbA_{1C}, glycosylated hemoglobin; HDL-C, high-density lipoprotein-cholesterol; LDL-C, low-density lipoprotein-cholesterol; OHA, oral hypoglycemic agents; PLT, platelet; SBP, systolic blood pressure; TC, total cholesterol; TG, triglyceride; WHR, waist-to-hip ratio; UA, uric acid.

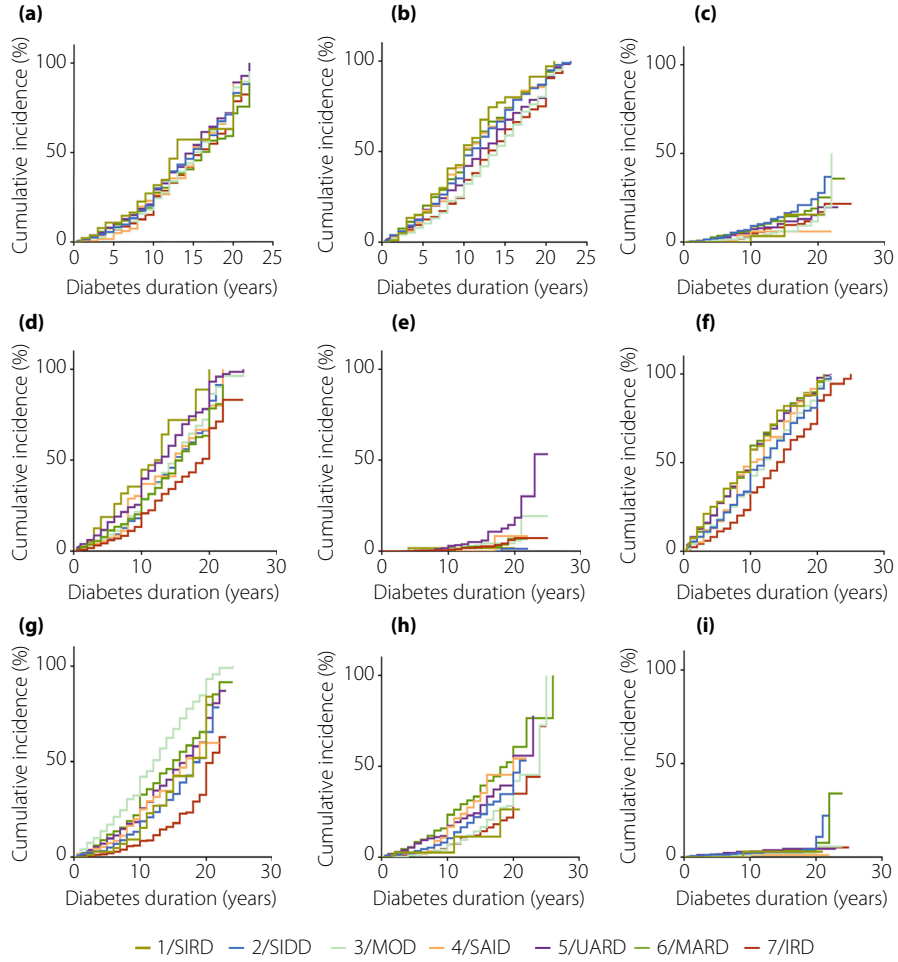

Figure 4 | Cox regression analysis of disease progression over time by seven clusters. (a) Time to diabetic retinopathy. (b) Time to diabetic peripheral neuropathy. (c) Time to diabetic foot. (d) Time to chronic kidney disease. (e) Time to end-stage renal disease. (f) Time to hypertension. (g) Time to coronary heart disease. (h) Time to cerebral vascular disease. (i) Time to cancer. IRD, inheritance-related diabetes; MOD, mild obesity-related diabetes; SIRD, severe insulin-resistant diabetes; UARD, UA-related diabetes.

Recently, Zou *et al.*[16] applied Ahlqvist's classification to newly diagnosed cases of diabetes in China and the USA, and four clusters were identified as SAID, SIDD, MOD and MARD. However, the index of GADA was absent, and surrogate parameters, such as HbA_{1c} (or, alternatively, mean plasma glucose), were used. Additionally, only newly diagnosed diabetes patients were incorporated, whereas patients with a long duration were not enrolled. Here, the same variables as Ahlqvist's classification were used, and patients in China with a long duration of diabetes were incorporated in our cluster analysis. The present study showed that just three subphenotypes were duplicated. Two possible reasons might be responsible for this. On the one hand, the characteristics between East Asians and Europeans with type 2 diabetes were different[17]. For example, a previous study suggested that East Asian populations developing type 2 diabetes had a relatively lower mean BMI in comparison with those of European populations; at any specific BMI, East Asians showed different body fat and visceral adiposity; Asian patients had diabetes at a younger age and early β-cell dysfunction, with early requirement of insulin treatment[17]. Thus, the patients in the present study might show different clinical characteristics to European populations in BMI and insulin resistance. Therefore, SIRD and MOD were not identified in the Chinese populations with the European classification, even though both HOMA2-IR and BMI were all used as indicators in the two-step cluster analysis. On the other hand, racial difference might be an explanation for this. As we know, genetic factors and lifestyle plays an important role in type 2 diabetes[18], which might affect the clinical characteristics. Thus, the European classification might not be completely consistent with the Chinese population, and different variables, even though the same variables, might have different results. Furthermore, the present study showed different characteristics. For example, patients in Chinese populations with SAID were

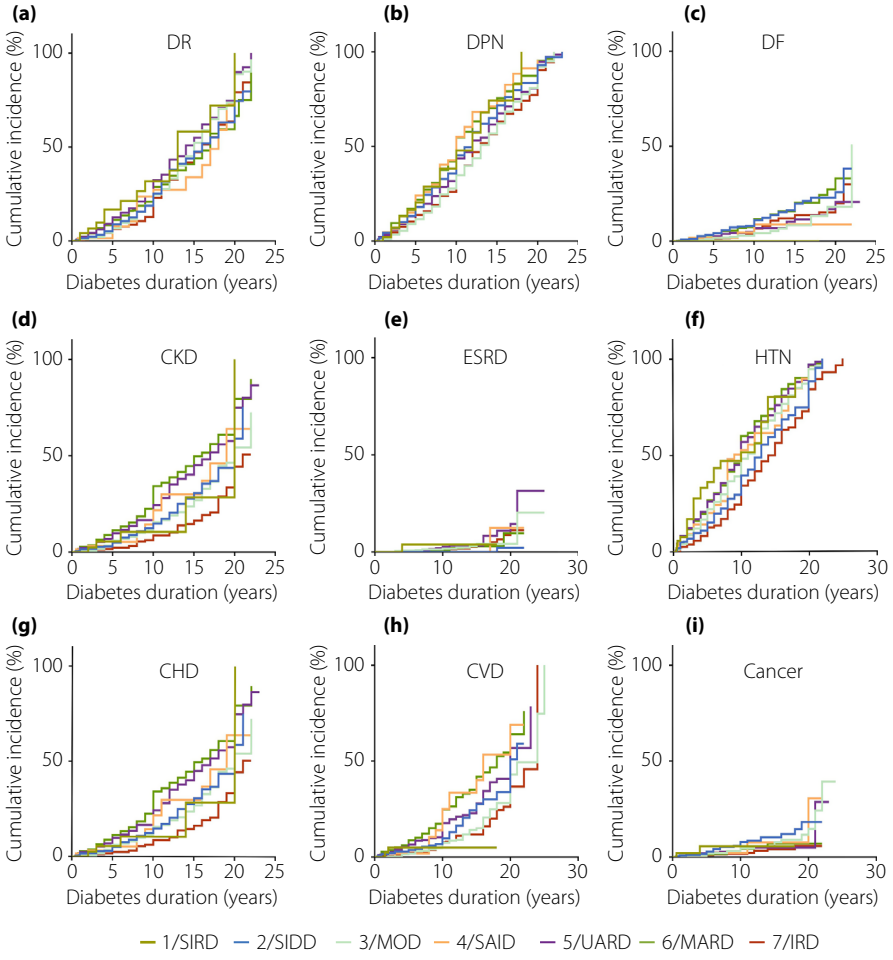

Figure 5 | Cox regression analysis of disease progression over time by seven clusters in men. (a) Time to diabetic retinopathy (DR). (b) Time to diabetic peripheral neuropathy (DPN). (c) Time to diabetic foot (DF). (d) Time to chronic kidney disease (CKD). (e) Time to end-stage renal disease (ESRD). (f) Time to hypertension (HTN). (g) Time to coronary heart disease (CHD). (h) Time to cerebral vascular disease (CVD). (i) Time to cancer. IRD, inheritance-related diabetes; MOD, mild obesity-related diabetes; SIRD, severe insulin-resistant diabetes; UARD, UA-related diabetes.

relatively older and had severe insulin resistance, whereas corresponding patients in European populations were younger and had slight or scarce insulin resistance. In addition to ethnic variability, other possible mechanisms require further study for detection.

In the present study, in contrast to the European classification, we refined the cluster analysis by adding other risk factors that showed a close association with the progression of diabetes. We also used TC instead of TG to do the clustering analysis. It showed similar results. Furthermore, plasma triglyceride concentration is also important in diabetes. Many studies have noted that diabetic dyslipidemia was featured as a high concentration of plasma TG, reduced concentration of high-density lipoprotein cholesterol and increased concentration of low-density lipoprotein cholesterol[19,20]. A previous study showed that out of the 291 diabetes patients enrolled, 22.3% had hypercholesterolemia (TC ≥200) and 61.9% had hypertriglyceridemia[21], indicating a higher percentage of hypertriglyceridemia than hypercholesterolemia. Based on this evidence, we selected serum TG to do the cluster analysis. Cox regression showed that different subgroups displayed different clinical risks. In the European classification, a higher risk of CKD and ESRD was found in the SIRD subgroup than in the MARD subgroup. The present study showed not only a higher risk of CKD (Figure 4d), but also a higher risk of DPN (Figure 4b) and HTN (Figure 4f) in the SIRD subgroup than in the IRD subgroup. Patients in the UARD subgroup and not in the SIRD subgroup were more susceptible to ESRD than those in the IRD subgroup. Furthermore, CHD (Figure 4g; $P = 0.001$) and CVD (Figure 4h; $P = 0.013$) were much more likely to occur among patients in the UARD cluster; these findings differed from those for European populations, in which no significant difference in the risk of coronary events and stroke was found among each cluster.

SIRD had been shown to be closely associated with disease progression not only in the European population, but also in

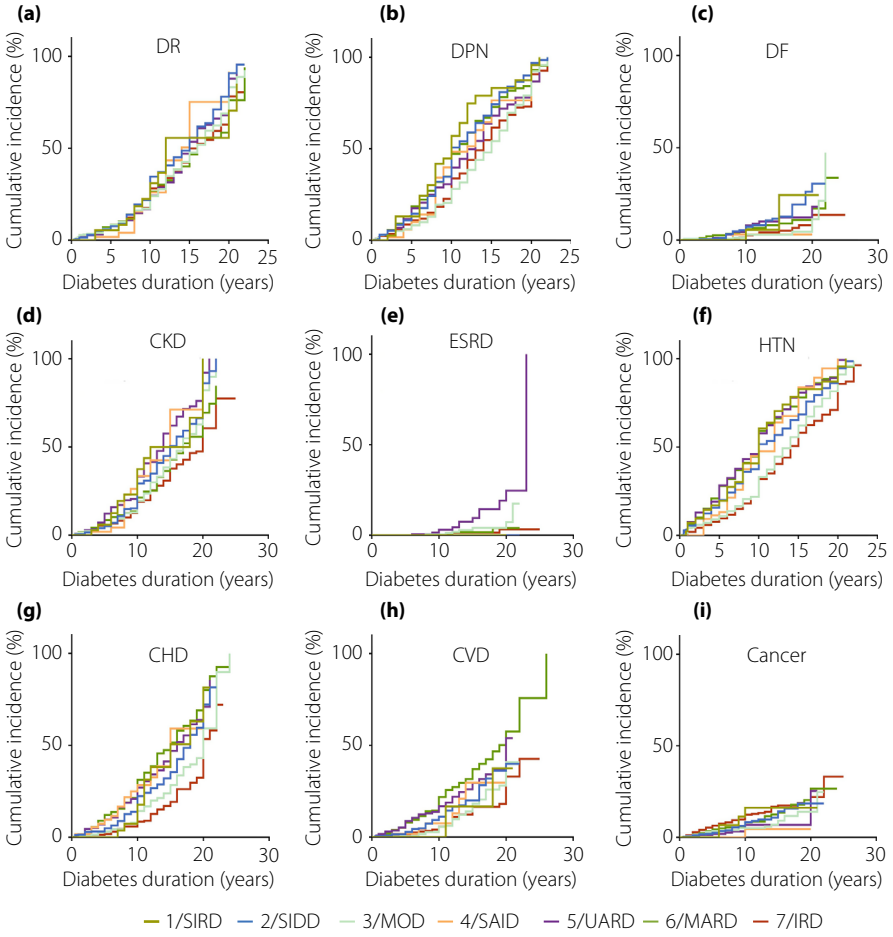

Figure 6 | Cox regression analysis of disease progression over time by seven clusters in women. (a) Time to diabetic retinopathy (DR). (b) Time to diabetic peripheral neuropathy (DPN). (c) Time to diabetic foot (DF). (d) Time to chronic kidney disease (CKD). (e) Time to end-stage renal disease (ESRD). (f) Time to hypertension (HTN). G) Time to coronary heart disease (CHD). (h) Time to cerebral vascular disease (CVD). (i) Time to cancer. IRD, inheritance-related diabetes; MOD, mild obesity-related diabetes; SIRD, severe insulin-resistant diabetes; UARD, UA-related diabetes.

the Chinese cohort. The elevated risk of CKD and HTN in SIRD patients substantially validated the relationship between insulin resistance and HTN and kidney complications[22]. Insulin resistance, which is involved in typical hallmarks of diabetic nephropathy, such as renal glomerular hypertension, hyperfiltration and high salt sensitivity, leads to renal damage and blood pressure elevation[23]. Furthermore, insulin resistance was also correlated with DPN in the present study. As suggested in the study by Callaghan *et al.*[24], in addition to hyperglycemia, insulin resistance attributed to cellular dysfunction, such as mitochondrial damage, oxidative stress, DNA lesions and ultimately cell apoptosis, also contributes to DPN.

In the present study, a novel subgroup of UARD was identified and presented with dissimilarity to other groups, especially with respect to the risk of disease complications. UA, originating from purine nucleotide metabolism, plays a critical role in various human diseases, such as heart failure[25], HTN[26] and CKD[27]. A prospective study showed that UA might be a biomarker of early coronary atherosclerosis in postmenopausal women[28]. In addition, a close relationship was also found between UA and diabetes[29], especially in diabetes complications, such as DPN[30], CHD[31], DKD[32] and DR[33]. Marcus *et al.*[34] found that UA was a risk factor for sudden cardiac death (HR 2.41, 95% CI 1.16–5.00) and cardiovascular death (HR 1.77, 95% CI 1.12–2.81), which was partially similar to the findings for the UARD subgroup. This subgroup featured a relatively high level of UA and a higher risk of CHD. Additionally, the concentrations of serum UA showed a close relationship with CVD[35-37]. A meta-analysis carried out by Du *et al.*[35] suggested that higher serum UA levels might contribute to cerebral infraction in type 2 diabetes patients, which was consistent with the present study. Furthermore, the concentrations of serum UA were related to the progression of CKD

patients with eGFR ≥45 mL/min/1.73 m^2[38] and the development of DKD[32]. Additionally, the Reduction on Endpoints in NIDDM with the Angiotensin II Antagonist Losartan (RENAAL) trial[39] showed a reduction in UA per 0.5 mg/dL after 6 months of losartan treatment, leading to a decrease in renal risk events, such as doubled Cr values and a decreased risk of ESRD by 6% in patients with type 2 diabetes. This was further confirmed by the present findings that participants in the UARD cluster showed a risk of ESRD that was almost fourfold higher than the risk of those in the IRD cluster. This might be attributable to oxidative stress[40], endothelial dysfunction[41] and insulin resistance[40] induced by high levels of urate concentrations.

Another novel subgroup was the IRD cluster, which was marked by atypical characteristics of each variable, such as relatively normal metabolic control, no insulin resistance and lower BMI. Interestingly, the IRD cluster had the highest percentage of family history of diabetes, elucidating that genetic factors were crucial in the pathogenesis of diabetes. Thus, the genetic association should be evaluated by a genome-wide association study and prospective strategies in the future.

Several limitations need to be mentioned. First, patients with a long duration of diabetes were included in the present study, and seven clusters should also be shown in newly diagnosed patients and heterogeneous populations. Second, genetic associations in this sample were not shown. Furthermore, the duration of diabetes was determined from the natural course of disease, and prospective studies need to be carried out in the future. Furthermore, there was a measurement bias of the duration of diabetes mellitus that many patients develop clinical manifestations of diabetes mellitus before diagnosis of diabetes mellitus. Additionally, parametric survival models might be more appropriate than Cox regression to examine the associations. However, parametric survival models might be not suitable for the present study, as the patients collected in our cohort were all alive. We will do the parametric survival models in future. Finally, the present study targeted personalized medicine based on the clinical characteristics of the patients through clustering analysis. Thus, defined cut-off values or combination panels of eight variables should be identified to differentiate these phenotypes when different patients are diagnosed with type 2 diabetes mellitus by increasing the number of patients and multicenter studies, which might be more practical in clinical application.

Taken together, three subgroups of the European classification were identified in Chinese participants with type 2 diabetes. Furthermore, seven subgroups were identified with different disease progression when more parameters were included, which might be more applicable to Chinese populations. This discovery could provide important evidence for the etiology-based stratification and personalized management of various subgroups in type 2 diabetes.

ACKNOWLEDGMENTS

The National Key R&D Program of China (2018YFC1314002 and 2016YFC1305501) and the National Natural Science Foundation of China (81730018) assisted with this study. We express our sincere acknowledgement for the help and support of the National Clinical Research Center for Metabolic Diseases Diabetes Center.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet* 2017; 389: 2239–2251.
2. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018; 14: 88–98.
3. Gregg EW, Li Y, Wang J, *et al*. Changes in diabetes-related complications in the United States, 1990–2010. *N Engl J Med* 2014; 370: 1514–1523.
4. Tuomi T, Santoro N, Caprio S, *et al*. The many faces of diabetes: a disease with increasing heterogeneity. *Lancet* 2014; 383: 1084–1094.
5. Kim SH. Measurement of insulin action: a tribute to Sir Harold Himsworth. *Diabet Med* 2011; 28: 1487–1493.
6. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet* 2018; 391: 2449–2462.
7. Buzzetti R, Zampetti S, Maddaloni E. Adult-onset autoimmune diabetes: current knowledge and implications for management. *Nat Rev Endocrinol* 2017; 13: 674–686.
8. Laver TW, Colclough K, Shepherd M, *et al*. The Common p. R114W HNF4A Mutation Causes a Distinct Clinical Subtype of Monogenic Diabetes. *Diabetes* 2016; 65: 3212–3217.
9. The L. Diabetes: a dynamic disease. *Lancet* 2017; 389: 2163.
10. Ahlqvist E, Storm P, Karajamaki A, *et al*. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018; 6: 361–369.
11. Rossing P. Subclassification of diabetes based on quantitative traits. *Nat Rev Nephrol* 2018; 14: 355–356.
12. Wang L, Gao P, Zhang M, *et al*. Prevalence and Ethnic Pattern of Diabetes and Prediabetes in China in 2013. *JAMA* 2017; 317: 2515–2523.
13. Ji L, Hu D, Pan C, *et al*. Primacy of the 3B approach to control risk factors for cardiovascular disease in type 2 diabetes patients. *Am J Med* 2013; 126: e911–922.
14. Wang J, Chen RP, Lei L, *et al*. Prevalence and determinants of hyperuricemia in type 2 diabetes mellitus patients with central obesity in Guangdong Province in China. *Asia Pac J Clin Nutr* 2013; 22: 590–598.
15. Chatterjee S, Davies MJ. Accurate diagnosis of diabetes mellitus and new paradigms of classification. *Nat Rev Endocrinol* 2018; 14: 386–387.

16. Zou X, Zhou X, Zhu Z, *et al*. Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol* 2019; 7: 9–11.
17. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Ann N Y Acad Sci* 2013; 1281: 64–91.
18. Lawlor N, Khetan S, Ucar D, *et al*. Genomics of Islet (Dys)function and Type 2 Diabetes. *Trends Genet* 2017; 33: 244–255.
19. Chehade JM, Gladysz M, Mooradian AD. Dyslipidemia in type 2 diabetes: prevalence, pathophysiology, and management. *Drugs* 2013; 73: 327–339.
20. Gadi R, Samaha FF. Dyslipidemia in type 2 diabetes mellitus. *Curr Diab Rep* 2007; 7: 228–234.
21. Shahwan MJ, Jairoun AA, Farajallah A, *et al*. Prevalence of dyslipidemia and factors affecting lipid profile in patients with type 2 diabetes. *Diabetes Metab Syndr* 2019; 13: 2387–2392.
22. Groop L, Ekstrand A, Forsblom C, *et al*. Insulin resistance, hypertension and microalbuminuria in patients with type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia* 1993; 36: 642–647.
23. Gnudi L, Coward RJM, Long DA. Diabetic Nephropathy: Perspective on Novel Molecular Mechanisms. *Trends Endocrinol Metab* 2016; 27: 820–830.
24. Callaghan BC, Cheng HT, Stables CL, *et al*. Diabetic neuropathy: clinical manifestations and current treatments. *Lancet Neurol* 2012; 11: 521–534.
25. Huang H, Huang B, Li Y, *et al*. Uric acid and risk of heart failure: a systematic review and meta-analysis. *Eur J Heart Fail* 2014; 16: 15–24.
26. Kuwabara M, Hisatome I, Niwa K, *et al*. Uric Acid Is a Strong Risk Marker for Developing Hypertension From Prehypertension: A 5-Year Japanese Cohort Study. *Hypertension* 2018; 71: 78–86.
27. Jalal DI, Chonchol M, Chen W, *et al*. Uric acid as a target of therapy in CKD. *Am J Kidney Dis* 2013; 61: 134–146.
28. Prasad M, Matteson EL, Herrmann J, *et al*. Uric Acid Is Associated With Inflammation, Coronary Microvascular Dysfunction, and Adverse Outcomes in Postmenopausal Women. *Hypertension* 2017; 69: 236–242.
29. Li LX, Wang AP, Zhang R, *et al*. Decreased urine uric acid excretion is an independent risk factor for chronic kidney disease but not for carotid atherosclerosis in hospital-based patients with type 2 diabetes: a cross-sectional study. *Cardiovasc Diabetol* 2015; 14: 36.
30. Yu S, Chen Y, Hou X, *et al*. Serum Uric Acid Levels and Diabetic Peripheral Neuropathy in Type 2 Diabetes: a Systematic Review and Meta-analysis. *Mol Neurobiol* 2016; 53: 1045–1051.
31. Verdoia M, Barbieri L, Schaffer A, *et al*. Impact of diabetes on uric acid and its relationship with the extent of coronary artery disease and platelet aggregation: a single-centre cohort study. *Metabolism* 2014; 63: 640–646.
32. Macisaac RJ, Ekinci EI, Jerums G. Markers of and risk factors for the development and progression of diabetic kidney disease. *Am J Kidney Dis* 2014; 63: S39–62.
33. Kuwata H, Okamura S, Hayashino Y, *et al*. Serum uric acid levels are associated with increased risk of newly developed diabetic retinopathy among Japanese male patients with type 2 diabetes: A prospective cohort study (diabetes distress and care registry at Tenri [DDCRT 13]). *Diabetes Metab Res Rev* 2017; 33: e2905.
34. Kleber ME, Delgado G, Grammer TB, *et al*. Uric Acid and Cardiovascular Events: A Mendelian Randomization Study. *J Am Soc Nephrol* 2015; 26: 2831–2838.
35. Du L, Ma J, Zhang X. Higher Serum Uric Acid May Contribute to Cerebral Infarction in Patients with Type 2 Diabetes Mellitus: a Meta-Analysis. *J Mol Neurosci* 2017; 61: 25–31.
36. Milionis HJ, Kalantzi KJ, Goudevenos JA, *et al*. Serum uric acid levels and risk for acute ischaemic non-embolic stroke in elderly subjects. *J Intern Med* 2005; 258: 435–441.
37. Chiquete E, Ruiz-Sandoval JL, Murillo-Bonilla LM, *et al*. Serum uric acid and outcome after acute ischemic stroke: PREMIER study. *Cerebrovasc Dis* 2013; 35: 168–174.
38. Srivastava A, Kaze AD, McMullan CJ, *et al*. Uric Acid and the Risks of Kidney Failure and Death in Individuals With CKD. *Am J Kidney Dis* 2018; 71: 362–370.
39. Miao Y, Ottenbros SA, Laverman GD, *et al*. Effect of a reduction in uric acid on renal outcomes during losartan treatment: a post hoc analysis of the reduction of endpoints in non-insulin-dependent diabetes mellitus with the Angiotensin II Antagonist Losartan Trial. *Hypertension* 2011; 58: 2–7.
40. Fabbrini E, Serafini M, Colic Baric I, *et al*. Effect of plasma uric acid on antioxidant capacity, oxidative stress, and insulin sensitivity in obese subjects. *Diabetes* 2014; 63: 976–981.
41. Borgi L, McMullan C, Wohlhueter A, *et al*. Effect of Uric Acid-Lowering Agents on Endothelial Function: A Randomized, Double-Blind. *Placebo-Controlled Trial. Hypertension* 2017; 69: 243–248.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 | Cluster centers in Chinese populations with type 2 diabetes according to the European classification.

Table S2 | Cluster centers of seven classifications in Chinese population with type 2 diabetes.

Table S3 | Cox regression analysis of diabetic complications risk in seven clusters.

Table S4 | Cox regression analysis of cardiovascular events risk in seven clusters.

Table S5 | Cox regression analysis of cancer events risk in seven clusters.

Table S6 | Cox regression analysis of kidney events risk in seven clusters.

Table S7 | Cox regression analysis of complications risks in males and females in seven clusters.