Supplementary Table 1 Overall between-cluster differences in core variables (Kruskal–Wallis test)

Variable	H statistic	P value	Effect size (ε^2)
Age (year)	261.688003	1.94e-56	0.419
Body mass index (kg/m^2)	244.814455	8.66e-53	0.392
HbA1c (%)	200.905176	2.69e-43	0.321
HOMA-IR	371.44489	3.39e-80	0.597
HOMA-β	385.213933	3.53e-83	0.629

Overall differences across the four clusters were assessed using the Kruskal–Wallis test (non-parametric).

ε^2 (epsilon-squared) quantifies the proportion of variance explained; larger values indicate larger between-cluster effects.

Supplementary Table 2 Post-hoc pairwise comparisons of core variables among clusters (Mann-Whitney U; FDR-corrected)

Variable	Group 1	Group 2	Cliff's Δ	P value	FDR-adjusted P
Age (years)	MOD	SIR	0.057	0.443	0.443
Age (years)	MOD	SID	-0.724	<1e-16	<1e-16
Age (years)	MOD	MARD	-0.816	<1e-16	<1e-16
Age (years)	SIR	SID	-0.686	<1e-16	<1e-16
Age (years)	SIR	MARD	-0.769	<1e-16	<1e-16
Age (years)	SID	MARD	-0.147	0.0186	0.0224
Body mass index (kg/m²)	MOD	SIR	-0.758	<1e-16	<1e-16
Body mass index (kg/m²)	MOD	SID	0.530	<1e-16	<1e-16
Body mass index (kg/m²)	MOD	MARD	-0.214	0.000373	0.000373
Body mass index (kg/m²)	SIR	SID	0.951	<1e-16	<1e-16
Body mass index (kg/m²)	SIR	MARD	0.654	<1e-16	<1e-16
Body mass index (kg/m²)	SID	MARD	-0.696	<1e-16	<1e-16
HbA1c (%)	MOD	SIR	-0.567	1.31e-14	2.62e-14
HbA1c (%)	MOD	SID	-0.671	<1e-16	<1e-16
HbA1c (%)	MOD	MARD	-0.774	<1e-16	<1e-16
HbA1c (%)	SIR	SID	-0.085	0.258	0.258
HbA1c (%)	SIR	MARD	-0.203	0.00554	0.00831
HbA1c (%)	SID	MARD	-0.126	0.0419	0.049
HOMA-IR	MOD	SIR	-0.931	<1e-16	<1e-16
HOMA-IR	MOD	SID	0.579	<1e-16	<1e-16
HOMA-IR	MOD	MARD	-0.519	<1e-16	<1e-16
HOMA-IR	SIR	SID	0.999	<1e-16	<1e-16
HOMA-IR	SIR	MARD	0.679	<1e-16	<1e-16
HOMA-IR	SID	MARD	-0.896	<1e-16	<1e-16
HOMA-β	MOD	SIR	-0.856	<1e-16	<1e-16
HOMA-β	MOD	SID	0.724	<1e-16	<1e-16
HOMA-β	MOD	MARD	-0.369	8.08e-10	8.08e-10
HOMA-β	SIR	SID	0.994	<1e-16	<1e-16
HOMA-β	SIR	MARD	0.716	<1e-16	<1e-16
HOMA-β	SID	MARD	-0.923	<1e-16	<1e-16

Pairwise differences were assessed using the Mann–Whitney U test with Benjamini–Hochberg false discovery rate (FDR) correction.

Cliff's Δ is a non-parametric effect size (range −1 to 1); positive values indicate higher values in Group 1.

Supplementary Table 3 Overall between-cluster differences in OGTT-derived measures (Kruskal–Wallis test)

Variable	H statistic	P value	Effect size (ε^2)
1-hour glucose (mmol/L)	23.3052822	3.49e-05	0.033
2-hour glucose (mmol/L)	10.8423953	0.0096	0.013
1-hour insulin (mIU/L)	202.16473	1.44e-43	0.326
2-hour insulin (mIU/L)	205.139871	3.27e-44	0.328

Overall differences across clusters were tested using the Kruskal–Wallis test.

ε^2 (epsilon-squared) quantifies the magnitude of between-cluster differences.

Supplementary Table 4 Post-hoc pairwise comparisons of OGTT-derived measures among clusters (Mann–Whitney U; FDR-corrected)

Variable	Comparison	Cliff's Δ	P value	FDR-adjusted P
1-hour glucose (mmol/L)	MOD vs SIR	-0.317	2.14e-05	3.96e-05
1-hour glucose (mmol/L)	MOD vs SID	-0.178	0.005	0.00706
1-hour glucose (mmol/L)	MOD vs MARD	-0.247	4.63e-05	7.93e-05
1-hour glucose (mmol/L)	SIR vs SID	0.107	0.159	0.2
1-hour glucose (mmol/L)	SIR vs MARD	0.050	0.5	0.522
1-hour glucose (mmol/L)	SID vs MARD	-0.048	0.437	0.477
2-hour glucose (mmol/L)	MOD vs SIR	-0.219	0.00292	0.00468
2-hour glucose (mmol/L)	MOD vs SID	0.007	0.908	0.908
2-hour glucose (mmol/L)	MOD vs MARD	-0.072	0.233	0.267
2-hour glucose (mmol/L)	SIR vs SID	0.219	0.00358	0.00536
2-hour glucose (mmol/L)	SIR vs MARD	0.141	0.0531	0.0708
2-hour glucose (mmol/L)	SID vs MARD	-0.079	0.205	0.246
1-hour insulin (mIU/L)	MOD vs SIR	-0.711	<1e-16	<1e-16
1-hour insulin (mIU/L)	MOD vs SID	0.311	8.87e-07	1.77e-06
1-hour insulin (mIU/L)	MOD vs MARD	-0.411	1.26e-11	4.31e-11
1-hour insulin (mIU/L)	SIR vs SID	0.855	<1e-16	<1e-16
1-hour insulin (mIU/L)	SIR vs MARD	0.374	3.76e-07	8.2e-07
1-hour insulin (mIU/L)	SID vs MARD	-0.640	<1e-16	<1e-16
2-hour insulin (mIU/L)	MOD vs SIR	-0.728	<1e-16	<1e-16
2-hour insulin (mIU/L)	MOD vs SID	0.354	2.14e-08	5.72e-08
2-hour insulin (mIU/L)	MOD vs MARD	-0.322	9.81e-08	2.35e-07
2-hour insulin (mIU/L)	SIR vs SID	0.874	<1e-16	<1e-16
2-hour insulin (mIU/L)	SIR vs MARD	0.476	5.67e-11	1.7e-10
2-hour insulin (mIU/L)	SID vs MARD	-0.615	<1e-16	<1e-16

Pairwise differences were assessed using the Mann–Whitney U test with Benjamini–Hochberg FDR correction.

Cliff's Δ is a non-parametric effect size (range −1 to 1); positive values indicate higher values in the first group of the comparison.

Supplementary Table 5 Cross-tabulation of data-driven clusters and traditional prediabetes categories at baseline (N=621)

Cluster	IFG, n (row %)	IGT, n (row %)	IFG+IGT, n (row %)	Total, n
MARD	32 (16.8)	88 (46.3)	70 (36.8)	190
MOD	35 (19.8)	103 (58.2)	39 (22.0)	177
SID	34 (21.4)	75 (47.2)	50 (31.4)	159
SIR	12 (12.6)	50 (52.6)	33 (34.7)	95
Total	113 (18.2)	316 (50.9)	192 (30.9)	621

Values are n (row %). Total row shows n (% of N).

IFG = impaired fasting glucose; IGT = impaired glucose tolerance; IFG+IGT = combined IFG and IGT.

Clusters: MOD = mild obesity-related dysmetabolism; MARD = mild age-related dysmetabolism; SIR = severe insulin resistance; SID = severe insulin deficiency.

Association tested using Pearson chi-square test of independence; Cramér's V reported as effect size. Pearson χ^2 = 13.20, df = 6, P = 0.040; Cramér's V = 0.103

Supplementary Table 6a. Cluster stability (bootstrap Jaccard index)

Cluster	Mean	Median	Jaccard (Q1–Q3)	Jaccard (min–max)
MOD	0.847	0.869	0.816–0.913	0.411–0.979
SIR	0.779	0.800	0.677–0.916	0.235–1.000
SID	0.807	0.841	0.727–0.930	0.282–0.992
MARD	0.764	0.795	0.631–0.920	0.395–0.993

Stability was evaluated using bootstrap subsampling (80% resampling, 500 repeats). Jaccard index >0.75 indicates high cluster stability.

Random-seed sensitivity was assessed by running K-means with 50 random initializations; Adjusted Rand Index (ARI) close to 1.0 indicates highly consistent assignments.

MOD, MARD, SID, and SIR correspond to the four data-driven clusters described in the Methods section.

Supplementary Table 6b Random-seed sensitivity (Adjusted Rand Index)

	Mean	Median	Range(min–max)
Across 50 random K-means initializations	0.972	0.974	0.954–1.000

Random-seed sensitivity was assessed by running K-means with 50 random initializations; Adjusted Rand Index (ARI) close to 1.0 indicates highly consistent assignments.

MOD, MARD, SID, and SIR correspond to the four data-driven clusters described in the Methods section.

Supplementary Table 7 Gray's tests for differences in cumulative incidence functions (CIFs) among clusters

A. Overall Gray's test (4 clusters)

Outcome	Test statistic (χ^2)	df	P value
Progression to diabetes	5.023	3	0.17
Regression to normoglycemia	4.075	3	0.253
Prediabetes persistence	7.578	3	0.0556

B. Pairwise Gray's tests (Bonferroni-adjusted)

Outcome	Cluster 1	Cluster 2	Test statistic (χ^2)	P value	Bonferroni-adjusted P
Progression to diabetes	SID	MOD	1.891	0.169	1
Progression to diabetes	SID	MARD	0.360	0.548	1
Progression to diabetes	SID	SIR	0.005	0.946	1
Progression to diabetes	MOD	MARD	4.897	0.0269	0.161
Progression to diabetes	MOD	SIR	2.400	0.121	0.728
Progression to diabetes	MARD	SIR	0.273	0.601	1
Regression to normoglycemia	SID	MOD	0.115	0.734	1
Regression to normoglycemia	SID	MARD	2.544	0.111	0.664
Regression to normoglycemia	SID	SIR	0.192	0.661	1
Regression to normoglycemia	MOD	MARD	3.828	0.0504	0.302
Regression to normoglycemia	MOD	SIR	0.479	0.489	1
Regression to normoglycemia	MARD	SIR	0.861	0.353	1
Prediabetes persistence	SID	MOD	0.156	0.693	1
Prediabetes persistence	SID	MARD	1.872	0.171	1
Prediabetes persistence	SID	SIR	6.226	0.0126	0.0756
Prediabetes persistence	MOD	MARD	0.645	0.422	1
Prediabetes persistence	MOD	SIR	4.336	0.0373	0.224
Prediabetes persistence	MARD	SIR	2.475	0.116	0.694

Gray's test compares cumulative incidence functions (CIFs) under competing risks.

Overall test assesses any difference across the four clusters; pairwise tests compare each cluster pair.

Bonferroni adjustment was applied to pairwise P values.

Supplementary Table 8 Fine–Gray competing-risk models for cumulative incidence (N=367)

Variable	Diabetes (sHR, 95% CI)	P	Normoglycemia (sHR, 95% CI)	P	Prediabetes persistence (sHR, 95% CI)	P
Cluster: MARD vs MOD (ref)	1.92 (1.07–3.44)	0.0283	0.46 (0.22–0.96)	0.0395	0.84 (0.58–1.20)	0.33
Cluster: SID vs MOD (ref)	1.63 (0.89–2.98)	0.115	0.85 (0.46–1.59)	0.616	1.06 (0.75–1.50)	0.739
Cluster: SIR vs MOD (ref)	1.70 (0.87–3.33)	0.121	0.74 (0.34–1.62)	0.452	0.57 (0.35–0.95)	0.0302
Intervention: Yes vs No (ref)	0.84 (0.55–1.29)	0.428	0.57 (0.32–1.00)	0.0506	0.76 (0.57–1.02)	0.069
Sex: Male vs Female (ref)	1.19 (0.80–1.78)	0.392	0.95 (0.56–1.59)	0.84	0.87 (0.66–1.15)	0.329

sHRs were estimated using Fine–Gray subdistribution hazard models with mutually adjusted covariates (cluster, intervention, sex).

Reference cluster in these models is MOD (as coded in the analysis output). Reference intervention = No; reference sex = Female.

Abbreviations: sHR = subdistribution hazard ratio; CI = confidence interval.

Supplementary Table 9 Likelihood-ratio test for cluster × intervention interaction

Test	LR statistic (χ^2)	df	P value	Interpretation
Likelihood Ratio Test	0.023	3	0.999	No evidence of interaction (P for interaction)

Interaction was tested by comparing Cox models with vs without the cluster×intervention interaction term using a likelihood-ratio test.

A high P value indicates no evidence that the relative intervention effect differs across clusters.

Supplementary Table 10 Interaction between cluster and intervention for diabetes mellitus (Cox proportional hazards model)

covariate	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	z	p	-log2(p)
sex	0.17877922	1.19575672	0.20695928	-0.22684106	0.584399507	0.79704745	1.79391343	0.863864173	0.38766251	1.36712687
intervention_int	-0.18054 2963	0.83481 6814	0.48963 8869	-1.140217 51	0.7791315 85	0.319749465	2.17957867	-0.3687 26779	0.7123 31386	0.48937 9536
cluster_MARD	0.63233 7395	1.88200 443	0.39906 2864	-0.149811 445	1.4144862 4	0.860870282	4.1143721	1.58455 585	0.1130 67283	3.14474 657
cluster_SID	0.49541 6685	1.64118 195	0.41826 4293	-0.324366 264	1.3151996 3	0.72298539	3.72549465	1.18445 847	0.2362 31625	2.08172 598
cluster_SIR	0.54114 2055	1.71796 776	0.47559 6701	-0.391010 351	1.4732944 6	0.676373155	4.36358715	1.13781 709	0.2551 96867	1.97031 748
cluster_MARD_x_intervention	0.05046 12282	1.05175 608	0.59844 9347	-1.122477 94	1.2234003 9	0.325472295	3.39872511	0.08431 99655	0.9328 02039	0.10035 7153
cluster_SID_x_intervention	-0.02031 1706	0.97989 3187	0.62320 0066	-1.241761 39	1.2011379 8	0.288874948	3.32389729	-0.0325 925928	0.9739 99477	0.03800 70975
cluster_SIR_x_intervention	-0.01692 86561	0.98321 3828	0.67781 5935	-1.345423 48	1.3115661 6	0.260429399	3.71198274	-0.0249 752996	0.9800 74665	0.02903 6432

Model included cluster (MOD as reference), intervention, and cluster×intervention interaction terms.

Hazard ratios are presented as exp(coef) with 95% confidence intervals.

Abbreviations: MOD, mild obesity-related dysmetabolism; MARD, mild age-related dysmetabolism; SIR, severe insulin severe insulin resistance; SID, severe insulin deficiency

Supplementary Table 11 Sensitivity analysis: primary Cox vs IPCW-weighted Cox models

Outcome	Variable	N	Events	IPCW-aHR	95% CI	P value
Progression to diabetes	MARD (reference)	104	35	Ref	—	—
	MOD vs MARD	99	17	0.52	0.28–0.98	0.043
	SID vs MARD	115	28	0.83	0.50–1.38	0.472
	SIR vs MARD	49	19	0.83	0.48–1.46	0.524
	Intervention: yes vs no	—	—	0.49	0.27–0.90	0.021
	Sex: male vs female	—	—	1.27	0.85–1.90	0.235
Regression to normoglycemia	MARD (reference)	104	11	Ref	—	—
	MOD vs MARD	99	21	1.98	0.97–4.06	0.062
	SID vs MARD	115	19	1.71	0.83–3.52	0.147
	SIR vs MARD	49	9	1.47	0.61–3.52	0.386
	Intervention: yes vs no	—	—	0.63	0.31–1.29	0.206
	Sex: male vs female	—	—	0.92	0.54–1.57	0.771
Persistence of prediabetes	MARD (reference)	104	58	Ref	—	—
	MOD vs MARD	99	61	1.20	0.86–1.70	0.287
	SID vs MARD	115	68	1.28	0.90–1.81	0.172
	SIR vs MARD	49	21	0.65	0.38–1.12	0.118
	Intervention: yes vs no	—	—	0.54	0.39–0.76	0.000
	Sex: male vs female	—	—	0.93	0.70–1.22	0.584

Notes. aHR, adjusted hazard ratio. Reference categories: MARD subtype, no intervention/control group, and female sex.

Models include mutually adjusted covariates (cluster, intervention, sex)

Supplementary Table 12 Schoenfeld residual tests for the proportional hazards assumption

Covariate	P value (Diabetes)	P value (Normoglycemia)	P value (Prediabetes persistence)
Cluster: MARD (vs MOD ref)	0.846	0.144	0.33
Cluster: SID (vs MOD ref)	0.405	0.784	0.118
Cluster: SIR (vs MOD ref)	0.964	0.772	0.242
Intervention	0.409	0.506	0.00343
Sex	0.608	0.395	0.479

Schoenfeld residual tests assess whether covariate effects vary over time (violation of proportional hazards).

$P < 0.05$ suggests potential violation; results should be interpreted alongside diagnostic plots.

Supplementary Table 13 Bootstrap resampling validation of hazard ratio estimates (percentile 95% intervals)

Outcome	Covariate	Bootstrap median HR (95% interval)	Bootstrap replicates
Diabetes	Sex (Male vs Female)	1.20 (0.76–1.82)	1000
Diabetes	Intervention (Yes vs No)	0.56(0.53–1.37)	1000
Diabetes	MARD vs MOD (ref)	1.93 (1.17–3.46)	1000
Diabetes	SID vs MOD (ref)	1.62 (0.88–3.11)	1000
Diabetes	SIR vs MOD (ref)	1.72 (0.95–3.31)	1000
Normoglycemia	Sex (Male vs Female)	0.93 (0.50–1.58)	1000
Normoglycemia	Intervention (Yes vs No)	0.56 (0.29–1.01)	1000
Normoglycemia	MARD vs MOD (ref)	0.45 (0.21–0.94)	1000
Normoglycemia	SID vs MOD (ref)	0.85 (0.42–1.55)	1000
Normoglycemia	SIR vs MOD (ref)	0.72 (0.28–1.56)	1000
Prediabetes persistence	Sex (Male vs Female)	0.86 (0.65–1.16)	1000
Prediabetes persistence	Intervention (Yes vs No)	0.75 (0.54–1.03)	1000
Prediabetes persistence	MARD vs MOD (ref)	0.83 (0.59–1.14)	1000
Prediabetes persistence	SID vs MOD (ref)	1.04 (0.75–1.48)	1000
Prediabetes persistence	SIR vs MOD (ref)	0.58 (0.33–0.94)	1000

Bootstrap resampling (with replacement) was repeated 1000 times to assess the stability of hazard ratio estimates. Intervals are percentile-based (2.5th–97.5th percentiles) from the bootstrap distribution.

Supplementary Table 14 Multinomial logistic regression for final glycemic status at follow-up (N=367)

Comparison	Covariate	Relative risk ratio (RRR, 95% CI)	P value	Significance
Prediabetes persistence vs Normoglycemia	Sex (Male vs Female)	0.93 (0.52–1.68)	0.819	
Prediabetes persistence vs Normoglycemia	Intervention (Yes vs No)	1.39 (0.75–2.58)	0.295	
Prediabetes persistence vs Normoglycemia	MARD vs MOD (ref)	1.85 (0.82–4.20)	0.139	
Prediabetes persistence vs Normoglycemia	SID vs MOD (ref)	1.27 (0.62–2.61)	0.509	
Prediabetes persistence vs Normoglycemia	SIR vs MOD (ref)	0.81 (0.32–2.04)	0.649	
Diabetes vs Normoglycemia	Sex (Male vs Female)	1.27 (0.65–2.46)	0.485	
Diabetes vs Normoglycemia	Intervention (Yes vs No)	1.73 (0.87–3.44)	0.119	
Diabetes vs Normoglycemia	MARD vs MOD (ref)	4.22 (1.65–10.80)	0.00268	**
Diabetes vs Normoglycemia	SID vs MOD (ref)	1.98 (0.82–4.75)	0.126	
Diabetes vs Normoglycemia	SIR vs MOD (ref)	2.77 (0.99–7.76)	0.0525	

Multinomial logistic regression models the final follow-up glycemic status categories; Normoglycemia is the reference outcome category.

RRR = relative risk ratio. Covariates include cluster, intervention, and sex.