SPPLEMENTARY METHODS S1 – DETAILED CLUSTERING METRICS

Determination of the Optimal Number of Clusters (K)

The optimal number of clusters was determined through an integrative framework combining multiple statistical criteria and clinical interpretability, thereby avoiding solutions that are statistically optimal but clinically implausible.

Statistical evaluation included the following complementary metrics, computed across a range of candidate values (K = 2 to 10):

Elbow method: The within-cluster sum of squares (WCSS) was plotted against K. The "elbow" point—where the marginal reduction in WCSS markedly diminishes—was considered a primary candidate, indicating diminishing returns in within-cluster homogeneity with further increases in K.

Silhouette score: The average silhouette coefficient (range: −1 to 1) was calculated for each K. Higher values indicate greater cohesion within clusters and better separation between clusters, with values closer to 1 reflecting superior partitioning.

Calinski–Harabasz index (CHI): This ratio of between-cluster dispersion to within-cluster dispersion was computed; higher values signify more distinct and compact clusters.

Davies–Bouldin index (DBI): This measure of average similarity between each cluster and its most similar counterpart was evaluated; lower values (approaching 0) indicate better-defined clustering.

Composite scoring: To synthesize evidence across metrics, a composite score was derived by: (i) min–max normalizing the Silhouette score and CHI (both positively oriented), (ii) min–max normalizing the DBI and reversing its direction (so that higher = better), and (iii) averaging the three normalized components. The K value yielding the highest composite score was designated the statistically optimal solution.

Convergence criteria: The K-means algorithm converged when the change in within-

cluster sum of squares (WCSS) fell below the default scikit-learn threshold (1e−4).

Clinical interpretability filtering: In cases where multiple K values (e.g., K=3 vs. K=4) showed comparable statistical support, clinical interpretability was prioritized. Specifically, any cluster whose central tendency or distribution violated the defining criterion of advanced chronic kidney disease (i.e., eGFR ⩾ 45 mL/min/1.73m^2 in a majority of its members) was deemed inconsistent with the target population and excluded from consideration. The final K was selected only if all resulting subphenotypes exhibited coherent metabolic profiles and alignment with known pathophysiological mechanisms of late-stage CKD.

Supplementary Table S1 Missing values after outlier detection

Variable	Missing (n)	Missing (%)
eGFR	0	0.0
Uric Acid	2	0.5
Phosphorus	10	2.5
PTH	84	21.0
Albumin	16	4.0
Hemoglobin	16	4.0
Glucose	26	6.5
D-dimer	30	7.5
Total Cholesterol	14	3.5

Supplementary Table S2. Clustering Results in the uACR-Available Subcohort (n=304)

Endotype	n	eGFR (mL/min/1.73m²)	Uric acid (μmol/L)	Glucose (mmol/L)	Albumin (g/L)	Hemoglobin (g/L)	D-dimer (mg/L)	Dialysis rate
1	2	61.0	297.0	7.0	17.1	115.0	20.07	0%
2	2	65.7	640.0	5.0	42.0	160.0	0.13	0%
3	288	37.2	383.5	6.0	38.5	119.5	0.87	34.7%
4	12	26.3	397.7	21.1	36.7	103.3	0.75	66.7%

Note: Data are presented as median values for each endotype cluster center. Endotypes were derived from K-means clustering using the same nine variables as the primary analysis (eGFR, uric acid, glucose, phosphate, PTH, albumin, hemoglobin, D-dimer, and total cholesterol) in patients with available uACR measurements. Dialysis rate represents the proportion of patients who initiated dialysis during follow-up.

Supplementary Table S3. Clustering Results with 10 Variables Including Imputed uACR (n=400)

Endotype	n	eGFR (mL/min/1.73m²)	Uric acid (μmol/L)	Glucose (mmol/L)	Albumin (g/L)	Hemoglobin (g/L)	D-dimer (mg/L)	uACR (mg/g)	Dialysis rate
1	2	65.7	640.0	5.0	42.0	160.0	0.13	771.6	0%
2	2	61.0	297.0	7.0	17.1	115.0	20.07	73.9	0%
3	372	37.4	382.2	6.0	37.7	117.5	0.94	318.1	34.4%
4	12	26.3	397.7	21.1	36.7	103.3	0.75	135.1	66.7%

Supplementary Table S4 Clinical management recommendations for metabolic endotypes in CKD 2-4

Inflammatory-hypercoagulable	Thrombosis + rapid disease progression	Every 3 months	Inflammation + coagulation function	Anti-inflammatory therapy (e.g., low-dose aspirin) + nutritional support
Severe CKD-MBD-anemia	High dialysis risk + bone metabolism disorders	Every 1-2 months	PTH + phosphate + hemoglobin	Phosphate binders + erythropoietin + vitamin D analogs
Metabolically favorable	Low progression risk	Every 6 months	Maintenance of metabolic homeostasis	Conservative monitoring + lifestyle interventions (low-salt diet + regular exercise)
Glycolipid-uric acid dysregulation	Metabolic syndrome + aggravated renal injury	Every 3 months	Blood glucose + uric acid + phosphate	SGLT2 inhibitors + urate-lowering agents + low-purine diet