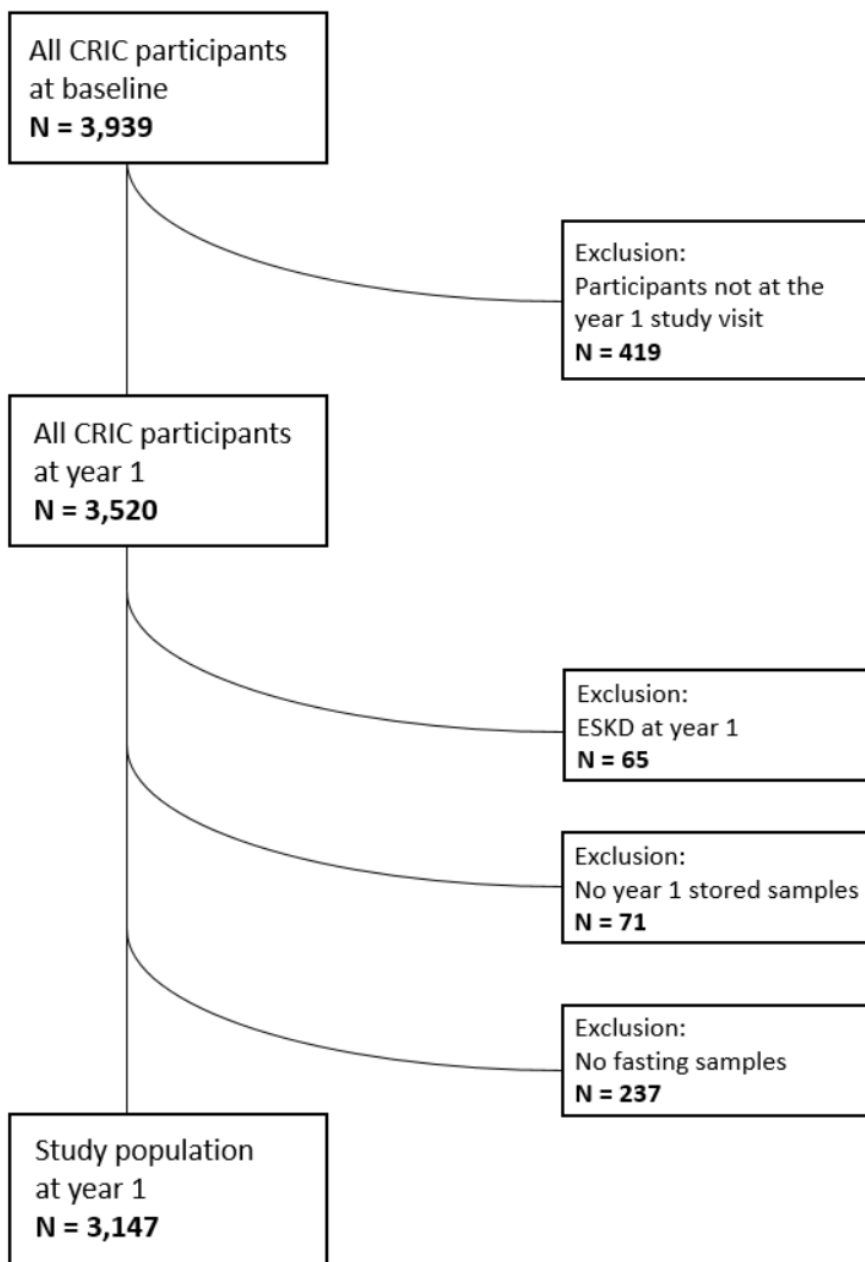**Figure S1.** Study population derived from the total CRIC Study population

From the total CRIC Study population of 3,939 participants who completed the baseline visit, we

excluded 419 participants who did not attend the year 1 CRIC Study visit because they died,

were lost to follow-up, withdrew from the study, or missed the year 1 study visit. From the 3,520

people at the year 1 study visit, we excluded those who had progressed to ESKD, did not have a

stored sample available at the year 1 visit, or did not have a fasting sample. These exclusions

yielded the final study population of 3,147 participants for our analysis.

Abbreviations: CRIC, Chronic Renal Insufficiency Cohort; DCA, deoxycholic acid; ESKD, end-

stage kidney disease.

Item S1. Supplementary Methods

Body weight, height, and blood pressure (BP) were measured using standard protocols.[S1] Diabetes was defined as fasting glucose ≥126 mg/dl, non-fasting glucose ≥200 mg/dl, or the use of antidiabetic medications. Prevalent CVD was defined as self-reported coronary artery disease, heart failure, stroke, or peripheral artery disease. Hemoglobin, cholesterol, phosphate, calcium, magnesium, and serum albumin were measured using standard laboratory methods. Twenty-four-hour urinary protein was measured using the turbidometric method with benzethonium chloride. Total parathyroid hormone (PTH) was measured using the total PTH assay, which includes the 1-84 PTH molecule and 7-84 fragments assay (Scantibodies, Santee, CA). Fibroblast growth factor 23 (FGF23) was measured by a second-generation C-terminal assay (Immutopics). High-sensitivity C-reactive protein (CRP) and interleukin-6 (IL-6) were measured using the particle-enhanced immunonephelometry method. Estimated glomerular filtration rate (eGFR) was calculated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation.[S2]

Since DCA levels were measured from stored serum samples obtained at the year one follow up visit, most covariates were from this time point. IL-6 and CRP values were only available at the CRIC Study baseline visit. Markers of mineral metabolism were only available in subset of participants at the year one follow up visit, but were available in nearly all participants at the CRIC Study baseline visit. To handle missing data, when values were not available at the year one follow up visit, we used values from the CRIC Study baseline visit. Missing data from all covariates was <2% in the final analyses.

To examine the associations between dietary variables and DCA, we used data from the self-reported Dietary Health Questionnaire (DHQ) collected during the baseline study visit. We examined the bivariate associations between log-DCA and animal protein, plant protein, total fat, and saturated fat in the diet using Pearson correlations. We also investigated the effect of the sample age on our adjusted Cox models that examined DCA with the outcome variables. The age of the DCA sample was included in Model 5, which adjusted for factors in model 4 plus sample age.

Table S1. Characteristics of all CRIC Study participants who attended the year 1 visit and of the participants included in the study population

Characteristics	All CRIC participants at year 1 N = 3520	Study population at year 1 N = 3147	Excluded population at year 1 N = 373
Age, years	59 ± 11	59 ± 11	58 ± 12
Female, N (%)	1594 (45.3)	1427 (45.3)	167 (44.8)
African American, N (%)	1469 (41.7)	1277 (40.6)	192 (51.5)
Hispanic, N (%)	405 (11.5)	361 (11.5)	44 (11.8)
Current smoking, N (%)	428 (12.2)	379 (12.0)	49 (13.1)
BMI, kg/m^2	32.1 ± 7.8	32.1 ± 7.8	31.6 ± 7.9
Systolic BP, mmHg	127 ± 22	127 ± 22	131 ± 23
Diabetes, N (%)	1749 (49.7)	1510 (48.0)	239 (64.1)
History of CVD, N (%)	1271 (36.1)	1099 (34.9)	172 (46.1)
Total cholesterol, mg/dl	182.0 ± 44.2	182.3 ± 43.9	179.4 ± 47.3
Statin use, N (%)	2082 (59.2)	1844 (58.6)	238 (63.8)

Number of BP medications	2.5 ± 1.3	2.5 ± 1.3	2.8 ± 1.2
eGFR, ml/min/1.73m2	42.2 ± 16.0	42.5 ± 16.0	40.1 ± 15.8
Urinary protein, g/24h	0.18 (0.07 – 0.95)	0.17 (0.07 – 0.89)	0.34 (0.09 – 1.51)
Serum albumin, g/dl	4.04 ± 0.44	4.05 ± 0.44	3.95 ± 0.47
Calcium, mg/dl	9.29 ± 0.53	9.30 ± 0.51	9.20 ± 0.63
IL-6, pg/ml	1.85 (1.14 – 3.03)	1.83 (1.12 – 2.97)	2.03 (1.33 – 3.51)
CRP, mg/dl	2.51 (1.02 – 6.18)	2.48 (1.02 – 6.11)	2.73 (1.21 – 6.56)
Phosphate, mg/dl	3.88 ± 1.03	3.87 ± 1.04	3.97 ± 0.93
PTH, pg/ml	62.0 (40.0 – 100.0)	61.0 (40.0 – 97.0)	68.3 (41.0 – 124.0)
FGF23, RU/ml	150.5 (95.4 – 281.3)	147.0 (93.8 – 272.2)	187.3 (116.7 – 358.0)

Continuous variables are presented as mean ± standard deviation for normally distributed data or as median and interquartile range for skewed data. Categorical variables are presented as total number and proportions.

Abbreviations: BMI, body mass index; SBP, systolic blood pressure; CVD, cardiovascular disease; SBP, systolic blood pressure; eGFR, estimated glomerular filtration rate; CRP, C-reactive protein; IL-6, Interleukin 6; PTH, parathyroid hormone; FGF23, fibroblast growth factor 23.

Table S2. Associations of dietary protein and fat with log-DCA

Variable	N	Pearson correlation coefficient	P value
Percent calories from fat	2424	0.026	0.202
Percent calories from saturated fat	2424	0.029	0.155
Percent animal protein	2411	-0.01878	0.357
Percent plant protein	2411	0.01878	0.357

Bivariate linear correlations between dietary variables and log-DCA.

Table S3. Associations of fasting DCA level with clinical outcomes, adjusted for age of sample[a]

	Atherosclerotic events	
	Below DCA median, < 68.45 ng/ml	**Above DCA median, > 68.45 ng/ml**
Events / Total Number	261/1574	251/1573
Mean follow-up time (year ± standard deviation)	6.9 ± 3.3	6.7 ± 3.3
Hazard ratio (95% CI) per SD of log-transformed DCA		
Unadjusted	0.88 (0.56 – 1.37)	1.83 (0.89 – 3.74)
Model 1[b]	0.82 (0.53 – 1.28)	1.49 (0.72 – 3.08)
Model 2[c]	0.91 (0.58 – 1.43)	1.48 (0.72 – 3.01)
Model 3[d]	0.90 (0.57 – 1.42)	1.58 (0.77 – 3.22)
Model 4[e]	0.88 (0.56 – 1.40)	1.52 (0.74 – 3.12)
Model 5[f]	0.88 (0.55 – 1.39)	1.52 (0.74 – 3.12)
	Heart failure events	
	Below DCA median, < 68.45 ng/ml	**Above DCA median, > 68.45 ng/ml**
Events / Total Number	303/1574	272/1573
Mean follow-up time (year ± standard deviation)	7.1 ± 3.2	7.0 ± 3.2
Hazard ratio (95% CI) per SD of log-transformed DCA		
Unadjusted	0.75 (0.50 – 1.13)	1.58 (0.80 – 3.14)
Model 1[b]	0.74 (0.49 – 1.12)	1.40 (0.70 – 2.80)
Model 2[c]	0.86 (0.56 – 1.31)	1.40 (0.71 – 2.73)
Model 3[d]	0.81 (0.53 – 1.24)	1.29 (0.66 – 2.53)

	Below DCA median, < 68.45 ng/ml	Above DCA median, > 68.45 ng/ml
Model 4[e]	0.82 (0.54 – 1.27)	1.22 (0.63 – 2.38)
Model 5[f]	0.82 (0.53 – 1.26)	1.22 (0.63 – 2.37)
ESKD events		
	Below DCA median, < 68.45 ng/ml	**Above DCA median, > 68.45 ng/ml**
Events / Total Number	456/1574	373/1573
Mean follow-up time (year ± standard deviation)	6.9 ± 3.3	7.0 ± 3.1
Hazard ratio (95% CI) per SD of log-transformed DCA		
Unadjusted	0.58 (0.41 – 0.81)	1.24 (0.69 – 2.23)
Model 1[b]	0.62 (0.44 – 0.88)	1.33 (0.74 – 2.41)
Model 2[c]	0.76 (0.54 – 1.08)	2.12 (1.18 – 3.82)
Model 3[d]	0.78 (0.55 – 1.12)	2.21 (1.23 – 3.99)
Model 4[e]	0.98 (0.68 – 1.40)	2.67 (1.51 – 4.74)
Model 5[f]	1.02 (0.71 – 1.46)	2.76 (1.56 – 4.90)
All-cause mortality events		
	Below DCA median, < 68.45 ng/ml	**Above DCA median, > 68.45 ng/ml**
Events / Total Number	408/1574	411/1573
Mean follow-up time (year ± standard deviation)	8.1 ± 2.6	7.9 ± 2.6
Hazard ratio (95% CI) per SD of log-transformed DCA		
Unadjusted	0.88 (0.63 – 1.24)	2.36 (1.37 – 4.07)
Model 1[b]	0.81 (0.58 – 1.15)	2.05 (1.17 – 3.56)
Model 2[c]	1.03 (0.72 – 1.46)	2.12 (1.24 – 3.64)

Model 3[d]	0.98 (0.69 – 1.40)	2.11 (1.23 – 3.64)
Model 4[e]	1.00 (0.70 – 1.43)	2.13 (1.25 – 3.64)
Model 5[f]	1.03 (0.72 – 1.48)	2.15 (1.26 – 3.66)

[a]Abbreviations: DCA, deoxycholic acid; CI, confidence interval; SD, standard deviation; CVD, cardiovascular disease; ESKD, end-stage kidney disease.
[b]Model 1 stratified by study site and adjusted for age, sex, African American race and Hispanic ethnicity.
[c]Model 2 adjusted for: Model 1 + eGFR, log urinary protein, diabetes, SBP, number of antihypertensive medications, current smoking, history of cardiovascular disease, total cholesterol, and statin use.
[d]Model 3 adjusted for: Model 2 + log IL-6 and log CRP.
[e]Model 4 adjusted for: Model 3 + log FGF23, log PTH, phosphate, calcium and albumin.
[f]Model 5 adjusted for: Model 4 + sample age.

Table S4. Associations between log-DCA and DCA predictors

Variable	N	Unit/category	Estimate (95% CI)	Wald Chi-Square	*p-value*
Age	3147	Per 1 unit increase of age in years	0.009 (0.004 – 0.014)	14.71	0.0001
Gender	3147	Male	-0.084 (-0.178 – 0.009)	3.16	0.0755
		Female	Reference		
Race	3147	African American	-0.089 (-0.186 – 0.008)	3.25	0.0713
		Other	Reference		
Ethnicity	3147	Hispanic	0.038 (-0.125 – 0.201)	0.21	0.6463
		Other	Reference		
eGFR	3147	Per 1 unit increase of eGFR in ml/min/1.73m2	0.005 (0.001 – 0.008)	6.28	0.0122
BMI	3144	Per 1 unit increase of BMI in kg/m2	0.005 (-0.001 – 0.011)	2.49	0.1143
Diabetes	3147	Yes	0.097 (0.002 – 0.191)	4.04	0.0444
		No	Reference		
Systolic BP	3147	Per 1 unit increase of Systolic BP in mmHg	0.001 (-0.002 – 0.003)	0.26	0.6131
Number of BP medications	3147	Per 1 unit increase of BP medications	-0.034 (-0.074 – 0.005)	2.86	0.0908
Current smoking	3147	Yes	-0.065 (-0.197 – 0.067)	0.92	0.3366
		No	Reference		
History of CVD	3147	Yes	0.034 (-0.061 – 0.129)	0.49	0.4852
		No	Reference		

Total cholesterol	3146	Per 1 unit increase of total cholesterol in mg/dl	-0.001 (-0.002 – 0.0002)	2.6	0.1070
Statin use	3147	Yes	-0.157 (-0.250 – -0.064)	11.04	0.0009
		No	Reference		
Log urinary protein, g/24h	3123	Per 1 unit increase of ln_urinary protein in g/24h	0.019 (-0.016 – 0.054)	1.12	0.2898
Log CRP, mg/dl	3143	Per 1 unit increase of ln_CRP in mg/dl	-0.041 (-0.080 – -0.003)	4.46	0.0347
Log IL-6, pg/ml	3104	Per 1 unit increase of ln_IL6 in pg/ml	0.084 (0.028 – 0.140)	8.69	0.0032
Log FGF23, RU/ml	3124	Per 1 unit increase of ln_FGF23 in pg/ml	-0.008 (-0.067 – 0.052)	0.06	0.8016
Log PTH, pg/ml	3128	Per 1 unit increase of ln_IL6 in pg/ml	-0.010 (-0.083 – 0.063)	0.08	0.7841
Phosphate	3117	Per 1 unit increase of phosphate in mg/dl	-0.023 (-0.067 – 0.021)	1.06	0.3023
Calcium	3147	Per 1 unit increase of calcium in mg/dl	0.017 (-0.080 – 0.113)	0.11	0.7370
Serum albumin	3147	Per 1 unit increase of serum albumin in g/dl	0.102 (-0.020 – 0.224)	2.68	0.1015

Tobit regression between log-DCA and independent variables. The Wald Chi-square value and p-value are from the Type III Analysis of Effects. We inferred strength of association between DCA and the variables based on the Chi-Square values.

Abbreviations: BMI, body mass index; SBP, systolic blood pressure; CVD, cardiovascular disease; BP, blood pressure; eGFR, estimated glomerular filtration rate; CRP, C-reactive protein; IL-6, Interleukin 6; PTH, parathyroid hormone; FGF23, fibroblast growth factor 23.

References:

S1. Lash JP, Go AS, Appel LJ, et al. Chronic Renal Insufficiency Cohort (CRIC) Study: baseline characteristics and associations with kidney function. *Clinical journal of the American Society of Nephrology : CJASN.* 2009;4(8):1302-1311.

S2. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150(9):604-612.