Supplementary Material for

Association between Mitochondrial DNA Copy Number in Peripheral Blood and Incident Chronic Kidney Disease in the Atherosclerosis Risk in Communities (ARIC) Study

Adrienne Tin, Morgan E. Grams, Foram N. Ashar, John A. Lane, Avi Z. Rosenberg, Megan L. Grove, Eric Boerwinkle, Elizabeth Selvin, Josef Coresh, Nathan Pankratz, Dan E. Arking

Supplementary Table 1. Participant Characteristics by Prevalent CKD status, defined as eGFR < 60 mL/min/1.73m^2 (n=9244)

	CKD	no CKD	P-Value
N	183	9061	
Age, mean (SD)	61.1 (4.8)	56.6 (5.7)	<0.001
Female, % (n)	53.6 (98)	56.7 (5141)	0.43
African Americans, % (n)	37.7 (69)	20.6 (1867)	<0.001
eGFR, mL/min/1.73m2, mean (SD)	45.6 (14.7)	97.4 (13.8)	<0.001
Prevalent diabetes, % (n)	33.3 (61)	14 (1269)	<0.001
Prevalent hypertension, % (n)	73.8 (135)	33.7 (3055)	<0.001
Smoking status, % (n)			0.003
Current smoker	16.4 (30)	23.8 (2156)	
Former smoker	49.2 (90)	37.4 (3387)	
Never smoked	34.4 (63)	38.8 (3518)	
Prevalent coronary heart disease, % (n)	16.9 (31)	4.8 (438)	<0.001
BMI, kg/m2, mean (SD)	29.1 (5.7)	27.9 (5.4)	0.003
White blood cell count, 10^3/uL, median (1st & 3rd quartile)	6.7 (5.4, 8.1)	5.8 (4.9, 7)	<0.001
HbA1c, median (1st & 3rd quartile)*	5.7 (5.5, 6.4)	5.4 (5.2, 5.8)	<0.001
Serum albumin, mean (SD)*	4.04 (0.46)	4.18 (0.28)	<0.001
hsCRP, median (1st & 3rd quartile)*	5.0 (2.6, 10.3)	2.2 (1.0, 4.7)	<0.001

*Variables are available at visit 2 only (n=8781 for HbA1c, 8475 for serum albumin, and 8548 for hsCRP

Supplementary Table 2. Cross-sectional association between prevalent CKD defined as eGFR <60mL/min/1.73m2 and mtDNA copy number

	N	Change in mtDNA copy number in SD (95% confidence interval)
Model 1		
European Americans	7308	-0.47 (-0.66, -0.29)
African Americans	1936	-0.64 (-0.89, -0.40)
Combined	9244	-0.54 (-0.68, -0.39)
Model 2	9244	-0.43 (-0.57, -0.28)
Model 3	9244	-0.40 (-0.54, -0.25)
Model 4	8288	-0.37 (-0.53, -0.22)

Model 1 adjusted for age, gender, center. Self-reported race was included in the combined model. P-value for interaction between self-reported race and mtDNA copy number in the combined model: 0.28
Model 2 added log-transformed WBC
Model 3 added prevalant coronary heart disease, prevalent diabetes, hypertension, smoking status, and BMI
Model 4 added serum albumin, hemoglobin A1c, log-transformed hsCRP at visit 2

Supplementary Table 3. ICD-9 or ICD-10 Codes Used for Capturing Hospitalization or Death Events Related to CKD

ICD-9-code	ICD-10-code	Description
582	N03	Chronic glomerulonephritis
583	--	Nephritis and nephropathy
585, 585.x where x≥3	N18, N18.x where x≥3	Chronic kidney disease
586	N19	Renal failure
587	N26	Renal sclerosis
588	N25	Disorders resulting from impaired renal function
403	I12	Hypertensive chronic kidney disease
404	I13	Hypertensive heart and kidney disease
593.9	--	Unspecified disorder of the kidney and ureter
250.4	E10.2, E11.2, E13.2	Diabetes with renal complications
V42.0	Z94.0	Kidney replaced by transplant
55.6	--	Transplant of kidney
996.81	--	Complications of transplanted kidney
V45.1*	Z99.2*	Renal dialysis status
V56*	Z49*	Admission for dialysis treatment or session
39.95*	--	Hemodialysis
54.98*	--	Peritoneal dialysis
--	Z45.2*	Encounter for adjustment and management of vascular access device

*These codes are counted as incident CKD only if a concomitant AKI code (ICD-9: 584.x, ICD-10: N17) is not present.

Supplementary Methods

Affymetrix Genome-wide Array 6.0 Data Quality Control in the ARIC study

Genomic DNA was isolated from buffy coat samples and hybridized to Affymetrix 6.0 microarrays according to the manufacturer's protocol. Genotypes were called using Birdseed (version 2). The genotyping in the European American cohort was supported by the National Institute of Health Gene Environment Association Studies (GENEVA) project.[1] The initial quality filters included a call rate of 0.95, filter for gender mismatch, and blind duplicates in each plate. Of the 910,030 single nucleotide polymorphisms (SNPs) genotyped, 839,048 passed the initial quality filters. Of the SNPs that failed the initial quality filters, 96% were due to call rates. After the genotype data was received from GENEVA (9,713 participants), additional quality filters were applied to exclude closely related individuals (identity by state distance [DST] >0.86 calculated by PLINK[2]) or genetic outliers (more than 8 SD away from any of the 10 principal components generated from autosomal SNPs with MAF >10%, call rate >95% and Hardy-Weinberg equilibrium [HWE] p-value >0.001). Altogether 669 individuals were excluded resulting in a genotype dataset of 9044 individuals.

The genotyping platform and quality filters for the African American cohort were the same as those for European Americans, except for the criteria for relatedness exclusion. For the African American cohort, individuals were excluded based on DST >0.8 or 6 SD away from any of the first 10 principal components generated from autosomal SNPs with MAF >10%, call rate >95% and HWE p-value >0.001. The DST threshold was selected based on the distribution within the population. Of 3,207 African Americans participants with genotype data, the data from 2874 participants passed quality control. Altogether, the genotype data of 11,918 participants from both European and African Americans passed the quality control for genotyping.

mtDNA Copy Number Estimation and Validation

Mitochondrial SNPs were selected by hand-curating a list of high-quality SNPs using the Genvisis software package, which can efficiently process, assess the quality of, analyze, and visualize intensity data from genome-wide arrays.[3] To reduce the risk of selecting SNPs with probe sequences that cross-hybridize to the nuclear genome, we used BLAST[4] to remove SNPs without a perfect match to the annotated mitochondrial location and SNPs with off-target matches longer than 20bp. A total of 25 mitochondrial SNPs passed our selection criteria. The probe

intensity of SNPs was determined using quantile sketch normalization (apt-probeset-summarize) as implemented in the Affymetrix Power Tools software.[5] The median of the normalized intensity, log R ratio (LRR), for homozygous calls of the 25 SNPs was corrected for GC content and used as the initial estimates of mtDNA copy number for each sample. The LRR values were computed by Genvisis based on the PennCNV-Affy pipeline.[3, 6, 7]

To correct for batch effects, DNA quality, and differing amounts of DNA quantity, we generated principal components (PCs) using the GC content-corrected LRR of 43,316 autosomal SNPs. These autosomal SNPs were selected using the following quality filters: call rate >98%, HWE p-value >0.00001, PLINK mishap p-value >0.0001, association with sex p-value >0.00001, linkage disequilibrium pruning (r^2<0.30), BLAST filtering (number of alignments with a score greater than 20 must equal exactly one), and maximum autosomal spacing (median distance between markers = 41.7 kb). Samples where the standard deviation of all LRR values >0.5 or sample call rate <95% were excluded from the principal component analysis. The resulting sample size was 11,480. From the initial pool of 400 PCs generated, we performed forward stepwise linear regression to select 15 PCs that together explained 67.6% of the variance of the initial estimates of mtDNA copy number. The standardized residuals from the linear model adjusted for these 15 PCs, age, sex, and study center were used as the estimates of relative mtDNA copy number in the present analyses.

Previously we reported an association of mtDNA copy number with frailty and all-cause mortality in the ARIC study.[8] The current estimates of mtDNA copy number represented an improvement over the previous estimates based on the correlation with an assay using multiplexed real-time quantitative polymerase chain reaction (qPCR) as described below. The drivers of this improvement came from the selection of 25 high quality SNPs using BLAST, the use of LRR to estimate intensity, the use of GC content correction, a larger curated list of autosomal markers for the principal component analysis (manuscript in preparation).

To validate the microarray-based estimates of mtDNA copy number, we estimated mtDNA copy number for 96 samples using a Taqman qPCR assay with *ND1* (Assay ID Hs02596873_s1) as the mitochondrial gene and *RPPH1* (Assay ID Hs03297761_s1) as the nuclear control gene, which was selected for being non-repetitive with no known alternative splicing events. Each sample was run in triplicate on a single 384-well plate in a 10-μL reaction containing 20 ng of DNA. The cycle threshold (Ct) value was determined from the amplification

curve for each target by the ABI Viia7 software. A ΔCt value was computed as the difference between the Ct for the *RPPH1* target and the Ct for the *ND1* target and used as a measure of mtDNA copy number relative to nuclear DNA copy number. The microarray-based and the qPCR-based estimates have a Spearman r of 0.51, supporting the validity of the microarray-based measures. The appropriate laboratory methods for quantifying mtDNA copy number is an area of ongoing research.[9-11] While qPCR is an established laboratory method, its use in accurately and reliability quantifying mtDNA copy number can be limited by factors such as sequence similarity between mitochondrial and nuclear genes and the size difference between the mitochondrial and nuclear genome.[9] Work is ongoing to determine which method provides a more accurate and precise quantification.

References

1. Cornelis, MC, Agrawal, A, Cole, JW, Hansel, NN, Barnes, KC, Beaty, TH, Bennett, SN, Bierut, LJ, Boerwinkle, E, Doheny, KF, Feenstra, B, Feingold, E, Fornage, M, Haiman, CA, Harris, EL, Hayes, MG, Heit, JA, Hu, FB, Kang, JH, Laurie, CC, Ling, H, Manolio, TA, Marazita, ML, Mathias, RA, Mirel, DB, Paschall, J, Pasquale, LR, Pugh, EW, Rice, JP, Udren, J, van Dam, RM, Wang, X, Wiggs, JL, Williams, K, Yu, K, Consortium, G: The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol,* 34**:** 364-372, 2010.
2. Purcell, S, Neale, B, Todd-Brown, K, Thomas, L, Ferreira, MA, Bender, D, Maller, J, Sklar, P, de Bakker, PI, Daly, MJ, Sham, PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**:** 559-575, 2007.
3. Pankratz, N: Genvisis is a robust software package that efficiently process, assess quality of, analyze, and visualize the intensity data from GWAS arrays (http://genvisis.org).
4. Altschul, SF, Gish, W, Miller, W, Myers, EW, Lipman, DJ: Basic local alignment search tool. *Journal of molecular biology,* 215**:** 403-410, 1990.
5. Korn, JM, Kuruvilla, FG, McCarroll, SA, Wysoker, A, Nemesh, J, Cawley, S, Hubbell, E, Veitch, J, Collins, PJ, Darvishi, K, Lee, C, Nizzari, MM, Gabriel, SB, Purcell, S, Daly, MJ, Altshuler, D: Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet,* 40**:** 1253-1260, 2008.
6. Diskin, SJ, Li, M, Hou, C, Yang, S, Glessner, J, Hakonarson, H, Bucan, M, Maris, JM, Wang, K: Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res,* 36**:** e126, 2008.
7. PennCNV-Affy (http://penncnv.openbioinformatics.org/en/latest/user-guide/affy).
8. Ashar, FN, Moes, A, Moore, AZ, Grove, ML, Chaves, PH, Coresh, J, Newman, AB, Matteini, AM, Bandeen-Roche, K, Boerwinkle, E, Walston, JD, Arking, DE: Association of mitochondrial DNA levels with frailty and all-cause mortality. *J Mol Med (Berl),* 93**:** 177-186, 2015.
9. Malik, AN, Shahni, R, Rodriguez-de-Ledesma, A, Laftah, A, Cunningham, P: Mitochondrial DNA as a non-invasive biomarker: accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias. *Biochem Biophys Res Commun,* 412**:** 1-7, 2011.

10. Grady, JP, Murphy, JL, Blakely, EL, Haller, RG, Taylor, RW, Turnbull, DM, Tuppen, HA: Accurate measurement of mitochondrial DNA deletion level and copy number differences in human skeletal muscle. *PLoS One,* 9**:** e114462, 2014.
11. Phillips, NR, Sprouse, ML, Roby, RK: Simultaneous quantification of mitochondrial DNA copy number and deletion ratio: a multiplex real-time PCR assay. *Sci Rep,* 4**:** 3887, 2014.