# Effects of genetic ancestry and socioeconomic deprivation on ethnic differences in serum creatinine

Leonardo Mariño-Ramírez [a,b,1,*], Shivam Sharma [a,c,1], Lavanya Rishishwar [a,b,d], Andrew B. Conley [a,b,d], Shashwat Deepali Nagar [b,c], I. King Jordan [b,c,d,*]

[a] National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, USA
[b] PanAmerican Bioinformatics Institute, Valle del Cauca, Cali, Colombia
[c] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[d] IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA, USA

## ARTICLE INFO

## ABSTRACT

The inclusion of ethnicity in equations for estimating the glomerular filtration rate (eGFR) from serum creatinine levels has been challenged since ethnicity is socially defined and therefore a poor proxy for biological differences. We hypothesized that genetic ancestry (GA) would be more strongly associated with creatinine levels among healthy individuals than self-identified ethnicity. We studied a diverse cohort of 35,590 participants characterized as part of the UK Biobank, grouped by self-reported ethnicity: Black, East Asian, Mixed, Other, South Asian, and White. We used multivariable modeling to test for associations between ethnicity, GA, socioeconomic deprivation, and serum creatinine levels, including covariates for age, sex, height, and body mass index. Model fit comparisons and relative importance analysis were used to compare the effects of ethnicity and GA on creatinine levels. Black ethnicity shows a positive effect on participant serum creatinine levels ($\beta = 9.36 \pm 0.38$), whereas East Asian ($\beta = -1.80 \pm 0.66$) and South Asian ($\beta = -0.28 \pm 0.36$) ethnicity show negative effects on creatinine. Male sex ($\beta = 17.69 \pm 0.34$) and height ($\beta = 0.13 \pm 0.02$) also show high positive associations with creatinine levels, while socioeconomic deprivation ($\beta = -0.04 \pm 0.04$) shows no significant association. African ancestry has the highest association ($\beta = 13.81 \pm 0.52$) with creatinine levels. Overall, GA (9.06%) explains significantly more of the variation in creatinine levels than ethnicity (4.96%), with African ancestry (6.36%) alone explaining more of the variation than ethnicity. We found that GA explains more of the variation in serum creatinine levels than socioeconomic deprivation, suggesting the possibility that ethnic differences in creatinine are shaped by genetic rather than social factors.

## 1. Introduction

The glomerular filtration rate (GFR) is the total filtration rate of the nephrons in the kidney, and it is considered to be the best overall index of kidney function (Levey et al., 2020a). GFR is used to diagnose kidney malfunction or disease, and to guide decisions about prognosis and treatment. Since GFR is impracticable to quantify in clinical practice, serum creatinine levels are used as a surrogate measure for GFR in clinical settings(Stevens and Levey, 2009). The National Institute of

Diabetes and Digestive Kidney Disease (NIDDK), the National Kidney Foundation (NKF), and the American Society of Nephrology (ASN) all recommend estimating GFR from serum creatinine.

The equations used to estimate GFR (eGFR) from serum creatinine levels include adjustments based on patients' age, sex, and race (ethnicity), specifically Black or non-Black; in principle, these variables are surrogates for the non-GFR determinants of serum creatinine, namely generation of creatinine from muscle and diet, tubular secretion of creatinine, and extra-renal elimination in the GI tract (Stevens and

Levey, 2009; Levey et al., 2020a; Levey et al., 2020b). These eGFR equation adjustments account for the observations in the study population used for development of the equations that younger patients had higher measured GFR (mGFR) than older patients at the same level of serum creatinine, and males had higher mGFR than females at the same level of serum creatinine, and Black patients had higher mGFR than members of other ethnicities at the same level of serum creatinine (Levey et al., 1999; Levey et al., 2006; Levey et al., 2009).

The eGFR equation adjustments for age and sex are non-controversial; however, the adjustment for ethnicity is highly contentious and has been recently challenged (Eneanya et al., 2019; Vyas et al., 2020). Opponents of the binary (Black or non-Black) ethnicity adjustment to the eGFR equation hold that ethnicity is socially defined and therefore a poor proxy for biological differences. There is concern that the ethnicity adjustment in the eGFR equation leads to higher eGFR values for Black patients, thereby jeopardizing their access to potentially life-saving technologies, such as dialysis and kidney transplant. These authors hold that the reliance on ethnicity is only justified if it confers a substantial benefit and cannot be achieved by other means.

This controversy persists even though it has recently been confirmed that use of ethnicity improves the estimation of mGFR from serum creatinine levels for Black patients and that ethnicity adds more predictive information than height and weight (Levey et al., 2020b; Levey et al., 2020c). Accordingly, proponents of the use of ethnicity for eGFR calculation assert that removing race from the equation will lead to misdiagnosis of Black patients and potentially exacerbate disparities in kidney disease. In support of this notion, they show how underestimation of eGFR for Black patients can lead to false detection of kidney disease, lower doses or contraindication for a number of antibiotics, and rejection of kidney donor candidates.

This study was motivated by a desire to explore the empirical basis for the consideration of ethnicity when calculating eGFR from patient serum creatinine levels. The specific objective of the study was to investigate the effects of ethnicity, genetic ancestry (GA), and socio-economic deprivation (SED) on serum creatinine levels in a diverse, cosmopolitan population. We compared Black, East Asian, Mixed, Other, and South Asian minority ethnic groups from the United Kingdom to the majority White ethnic group, which served as the reference group for all comparisons. We used multivariable modeling to test for associations between ethnicity and creatinine levels, including covariates for SED, age, sex, height, and body mass index (BMI). We hypothesized that, in addition to the previously observed association between Black ethnicity and creatinine levels, Asian ethnic groups would also show significant associations with serum creatinine levels.

GA is a characteristic of the genome and therefore a better proxy for genetic diversity than socially defined ethnic groups (Yudell et al., 2016; Yudell et al., 2020), and serum creatinine levels are known to be heritable, i.e. influenced by genetic factors (Arpegård et al., 2015; Sinnott-Armstrong et al., 2021). African GA was previously shown to be associated with serum creatinine levels in African American and Hispanic/Latino American patients (Udler et al., 2015). Given the heritability of serum creatinine levels, we hypothesized that GA would explain more of the variance in creatinine levels than socially defined ethnicity. Model fit comparisons and relative importance analysis were used to compare the magnitude of the effects of ethnicity and genetic ancestry on serum creatinine levels.

Levels of SED vary significantly among socially constructed ethnic groups, and SED affects a number of health outcomes, including kidney disease (Young et al., 1994; Drey et al., 2003; Nagar et al., 2021). Previous studies on the relationship between patient ethnicity and serum creatinine levels have not adjusted for SED (Udler et al., 2015; Levey et al., 2020b; Levey et al., 2020c). We hypothesized that controlling for SED could attenuate ethnic differences in serum creatinine levels.

## 2. Materials and methods

### 2.1. UK Biobank study cohort

Study participants and data were taken from the UK Biobank, a prospective cohort study of more than 500,000 participants between 40 and 70 years of age, enrolled between 2006 and 2010 (Bycroft et al., 2018). UK Biobank participants provided biological samples, completed questionnaires, and underwent medical assessments. Ethics approval for the UK Biobank was obtained from the North West Multi-centre Research Ethics Committee (MREC) for the United Kingdom, the Patient Information Advisory Group (PIAG) for England and Wales, and the Community Health Index Advisory Group (CHIAG) for Scotland (see https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics). This study was conducted using the UK Biobank resource under application number 65,206 granting access to LMR and IKJ to the corresponding UK Biobank biomarkers, and phenotype data. UK Biobank data is publicly available upon application on the UK biobank website (https://www.ukbiobank.ac.uk/register-apply/).

## 3. Model outcome, predictors, and covariates

Multivariable linear regression was used to model participant serum creatinine levels (outcome) by participant ethnicity, GA, and SED (predictors). Age, sex, height, and BMI were included as model covariates.

*Serum creatinine levels:* Serum creatinine levels were modeled as continuous variable in μmol/L units. Participant serum creatinine level data were taken from UK Biobank Field 30700: Creatinine. UK Biobank participant serum samples were derived from whole blood samples collected using silica clot accelerator tubes (SST) and stored for a variety of biochemical assays as previously described (Elliott et al., 2008). Participant serum creatinine levels were measured by enzymatic analysis on a Beckman Coulter AU5800 machine using a kinetic modification of the Jaffe procedure, calibrated to an isotope dilution mass spectrometry (IDMS) reference method (Jaffe, 1886; Cook, 1971).

*Ethnicity:* Ethnicity was modeled as a categorical variable according to participants' ethnic group self-identification. Participant ethnic group information was taken from the UK Biobank Field 21000: Ethnic background. UK Biobank participants self-identified as belonging to one of six possible ethnic groups via questionnaire at the time of enrollment: Asian, Black, Chinese, Mixed, White, or Other. The UK Biobank ethnic group choices are based on the ethnicity categories used by the UK National Health Service (NHS) (Gov.uk, 2020). It should be noted that the UK Biobank participant ethnic group labels make no distinction between the related concepts of race and ethnicity and therefore the ethnic groups used here may correspond to racial groups used other countries, such as the United States (US) (Kertzer and Arel, 2002). The term ethnicity, instead of race, is used throughout this study in keeping with the convention for data collected from the UK. The UK Biobank ethnic groups reflect the history of immigration to the UK, and the resulting make-up of the modern population, in a way that may not be obvious to readers from the US or other countries. For example, the UK Biobank makes a distinction between Asian and Chinese ethnic groups, where the UK Asian ethnic group corresponds to South Asians from Bangladesh, India, and Pakistan. Black and White ethnicity in the UK correspond to similarly labeled racial groups in the US. Ethnic group labels that correspond approximately to US racial group labels are used for this study: Black, East Asian, South Asian, or White. Supplementary Table 1 shows the correspondence between UK Biobank ethnic group designations, the ethnic group labels used for this study, and racial group designations used in the US.

*Genetic ancestry (GA):* GA was modeled as a continuous variable in percent ancestry values for each of six regional ancestry groups: African, Central Asian, East Asian, European, South Asian, and West Asian. UK Biobank participant whole blood samples were used to derive buffy coat

**Table 1**

**Characteristics of the study cohort.** Numbers of participants are shown across ethnic groups, and the percentage of males is shown across groups. For all other measures, average values are shown with standard deviations in parentheses. Units of measurement are shown for height, weight, and BMI measures and serum creatinine levels. SED is socioeconomic deprivation as measure by the Townsend deprivation index.

| | | All | Black | East Asian | South Asian | White | Mixed | Other |
|---|---|---|---|---|---|---|---|---|
| **n** | | 35,590 | 7,045 | 1,412 | 8,721 | 10,000 | 2,706 | 5,706 |
| **Creatinine (μmol/L)** | | 73.23 (24.68) | 79.98 (30.14) | 66.37 (18.25) | 71.74 (24.91) | 72.17 (16.32) | 71.15 (26.71) | 71.69 (27.66) |
| **% Genetic Ancestry** | **African** | 21.96 (38.44) | 90.64 (15.68) | 0.21 (0.95) | 0.30 (1.77) | 16.88 (21.30) | 16.54 (33.91) | 0.01 (0.20) |
| | **East Asian** | 7.39 (24.62) | 0.18 (5.94) | 98.39 (8.06) | 4.76 (18.58) | 5.82 (14.94) | 11.41 (29.81) | 0.03 (0.77) |
| | **South Asian** | 16.67 (30.58) | 1.09 (7.78) | 0.14 (1.21) | 57.38 (31.21) | 11.37 (19.12) | 9.44 (25.34) | 0.05 (0.94) |
| | **European** | 40.76 (44.75) | 6.45 (10.93) | 1.02 (6.81) | 3.68 (8.12) | 58.40 (22.76) | 38.53 (42.62) | 99.37 (5.30) |
| **SED (Townsend index)** | | 0.38 (3.61) | 2.65 (3.44) | −0.49 (3.35) | 0.32 (3.18) | −1.43 (3.00) | 0.47 (3.58) | 0.99 (3.73) |
| **Age (years)** | | 53.95 (8.37) | 51.99 (8.08) | 52.47 (7.66) | 53.29 (8.47) | 56.82 (7.92) | 51.78 (8.12) | 53.73 (8.26) |
| **% Male** | | 46.35 | 42.97 | 37.60 | 53.57 | 45.57 | 37.77 | 47.09 |
| **Height (cm)** | | 166.40 (9.36) | 167.33 (8.70) | 161.77 (7.88) | 163.95 (9.23) | 168.60 (9.33) | 166.86 (9.14) | 165.89 (9.55) |
| **Weight (kg)** | | 76.89 (16.01) | 82.61 (16.06) | 63.35 (11.60) | 73.16 (13.92) | 78.18 (15.87) | 76.41 (16.52) | 76.84 (16.44) |
| **BMI (kg/m²)** | | 27.70 (4.98) | 29.51 (5.40) | 24.11 (3.43) | 27.16 (4.38) | 27.43 (4.76) | 27.38 (5.23) | 27.83 (5.08) |

aliquots for DNA extraction, and whole genome genotypes (WGG) were characterized from extracted DNA using the UK Biobank Axiom Array or the UK BiLEVE Axiom Array as previously described (Welsh et al., 2017). Participant WGG were merged and harmonized with whole genome sequence (WGS) data from global reference populations characterized by the 1000 Genomes Project (1KGP) and the Human Genome Diversity Project (HGDP) (Li et al., 2008; Genomes Project et al., 2015; Bergström et al., 2019). WGG and WGS variant data were merged to include variants present in all three datasets with variant strand flips and identifier inconsistencies corrected as needed. Variant sample missingness < 5% and a minor allele frequency greater than 1% filters were used for merging. The merged genome variant data set was pruned for linkage disequilibrium (LD) using PLINK v2 with '–indep pairwise 100 10 0.05′ (Chang et al., 2015).

Principal component analysis (PCA) of the merged and LD pruned genome variant dataset was performed using the FastPCA program implemented in PLINK v2 (Galinsky et al., 2016). Visual inspection of the PCA results was used together with information on the 1KGP and HGDP samples' geographic origins to characterize a set of genetically and geographically distinct reference populations corresponding to six regions: African (sub-Saharan), Central Asian, East Asian, European, South Asian, and West Asian (Supplementary Table 2 and Supplementary Fig. 1). We developed and applied a fast and efficient method for genome-wide GA inference that analyzes the PCA data from global reference populations and non-reference individuals with non-negative least squares (NNLS) to assign GA fractions for each of the broad geographic regions to each individual UK Biobank participant. To do so, mean PC vectors for each individual reference population are first calculated. Then for each non-reference individual, NNLS is used to assign non-negative coefficients corresponding to each individual

reference population, where all of the reference population coefficients together best approximate that non-reference individuals PC vector. Finally, the individual reference population coefficients are summed for each region to yield regional GA percentages for each non-reference individual. This NNLS GA inference method is adopted from previous studies from our group (Conley et al., 2017; Jordan et al., 2019).

*Socioeconomic deprivation (SED):* SED was modeled as a continuous variable using the Townsend deprivation index, which is a widely used measure of SED that is known to be associated with a variety of health outcomes (Foster et al., 2018). Participant Townsend deprivation index data were taken from UK Biobank Field 189: Townsend deprivation index at recruitment. The Townsend Index is a composite metric of SED that includes: (1) unemployment, (2) non-car ownership, (3) non-home ownership, and (4) household overcrowding in a given area (Townsend et al., 1988). Townsend index values range from −6.23 to 10.59; higher (positive) values of the Townsend index indicate high material deprivation, and lower (negative) values indicate less deprivation (i.e. relative affluence). SED was chosen as a covariate, since ethnic groups experience different levels of SED, which are linked to differences in health outcomes among individuals and groups (Williams et al., 2016).

*Age and sex:* Age was modeled as a continuous variable in years, and sex was modeled as a categorical variable for male or female. Participant age data were taken from the UK Biobank Field 21003: Age when attended assessment center, and participant sex data were taken from the UK Biobank Field 31: Sex. Age and sex were chosen as covariates, since they have been previously associated with serum creatinine levels (Levey et al., 2009).

*Height, weight, and body mass index (BMI):* Height was modeled as a continuous variable in centimeters (cm), and weight was modeled as a continuous variable in kilograms (kg). BMI modeled as a continuous variable calculated as weight divided by the square of height in meters (m): BMI = $kg/m^2$. Participant data on height were taken from the UK Biobank Field 12144: Height, and participant data on weight were taken from the Field 21002: Weight. These body size measures were chosen as covariates, given the relationship between muscle mass and serum creatinine levels (Levey et al., 2020c).

**Table 2**

**Log likelihood ratio test comparing the creatinine model with all ethnic groups considered (Model 5) and the creatinine model with only Black ethnicity considered (Model 8).** The White ethnic group is taken as the reference group in the all ethnic group model, and all non-Black ethnic groups are taken as reference in the Black-only model.

| Models compared | Df[1] | R² | LogLik[2] | ΔDf[3] | χ² | P-value |
|---|---|---|---|---|---|---|
| Model 5: Creatinine ~ Ethnicity + SED + Age + Sex + Height + BMI | 12 | 0.1797 | −161,082 | 4 | 29.52 | 6.1 × 10⁻⁶ |
| Model 8: Creatinine ~ Black ethnicity + SED + Age + Sex + Height + BMI | 8 | 0.1790 | −161,097 | | | |

[1] Degrees of freedom for each model.
[2] Log likelihood for the goodness fit of each model.
[3] Difference in degrees of freedom.

*3.1. Statistical analysis methods and software*

The R statistical language v3.6.1 was used for all statistical analyses (R Development Core Team, 2020). Multivariable linear regression analysis was performed with the *lm* function in base R. Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) values were computed using the *AIC* and *BIC* functions in base R. Multicollinearity of model predictor variables was assessed using the variance inflation factor *vif* function implemented in the *car* R package (Fox and Weisberg, 2019). Bidirectional stepwise regression analysis was performed using the *MASS* R package (Ripley, 2002). Model forest plots were generated using the *sjPlot* R package (Lüdecke, 2021). Model
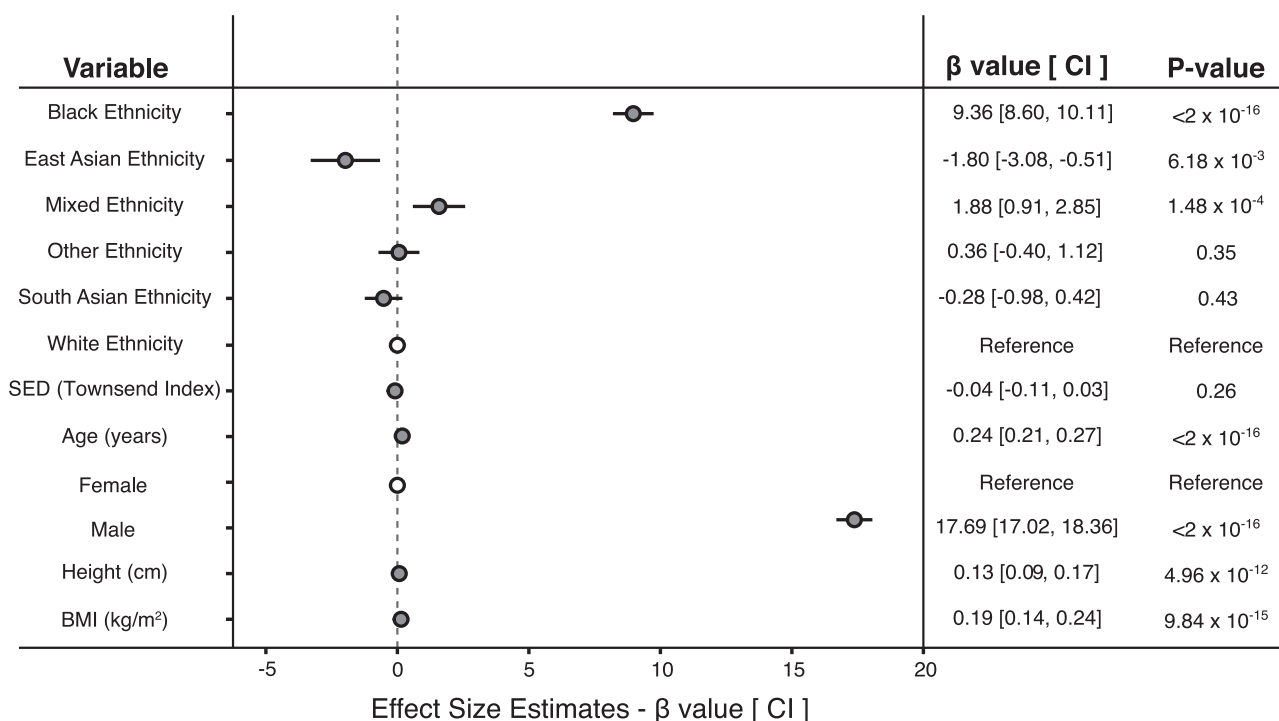
## Creatinine (μmol/L) ~ Ethnicity + SED + Age + Sex + Height + BMI



| Variable | β value [ CI ] | P-value |
|---|---|---|
| Black Ethnicity | 9.36 [8.60, 10.11] | $<2 \times 10^{-16}$ |
| East Asian Ethnicity | -1.80 [-3.08, -0.51] | $6.18 \times 10^{-3}$ |
| Mixed Ethnicity | 1.88 [0.91, 2.85] | $1.48 \times 10^{-4}$ |
| Other Ethnicity | 0.36 [-0.40, 1.12] | 0.35 |
| South Asian Ethnicity | -0.28 [-0.98, 0.42] | 0.43 |
| White Ethnicity | Reference | Reference |
| SED (Townsend Index) | -0.04 [-0.11, 0.03] | 0.26 |
| Age (years) | 0.24 [0.21, 0.27] | $<2 \times 10^{-16}$ |
| Female | Reference | Reference |
| Male | 17.69 [17.02, 18.36] | $<2 \times 10^{-16}$ |
| Height (cm) | 0.13 [0.09, 0.17] | $4.96 \times 10^{-12}$ |
| BMI (kg/m²) | 0.19 [0.14, 0.24] | $9.84 \times 10^{-15}$ |

Effect Size Estimates - β value [ CI ]

**Fig. 1. Ethnic disparities in serum creatinine levels.** Forest plot showing the results of multivariable linear regression modeling serum creatinine levels with ethnicity, SED, age, sex, height, and BMI. Effect size estimates (β-values), 95% confidence intervals (CIs), and association significance levels (P-values) are shown for each independent (predictor) variable in the model. The model is specified as shown. Serum creatinine levels are modeled as μmol/L units; ethnicity and sex are modeled as categorical variables, with White ethnicity and female sex as reference values; age, height, BMI, and SED are modeled as continuous variables with units as shown in the Figure.

comparisons were performed using the *lmtest* R package (Zeileis and Hothorn, 2002). Nested models where all ethnic groups are included compared to when only Black ethnicity is considered were compared using the likelihood ratio test with the *lrtest* function. Accuracies of analogous models for serum creatinine using SED, age, sex, height, and BMI, and either ethnicity or the GA measures substituted for them, were compared using the Cox and Davidson-MacKinnon J tests, implemented respectively by the *coxtest* and *jtest* R functions. Relative importance analysis was performed using the *relaimpo* R package (Grömping, 2006). The specifications for all statistical models used here are shown in Supplementary Table 3.

### 4. Results

#### 4.1. Study cohort and participant characteristics

A total of 502,493 participants from the UK Biobank progressive cohort were used to build the cohort for this study (Supplementary Fig. 2). First, participants with missing genetic data were excluded from
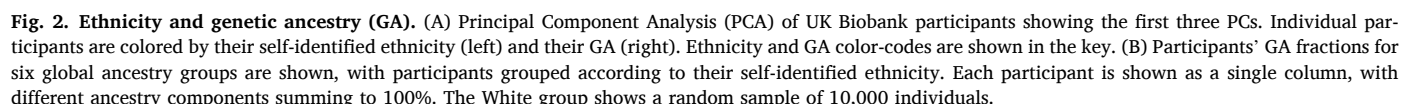
the study cohort. Next, participants with missing demographic data – ethnicity, age, or sex – and missing data on creatinine levels were excluded from the study cohort. Next, individuals with missing genetic, height, and weight data were excluded from the study cohort. All remaining individuals from the Black, East Asian, South Asian, Mixed, and Other ethnic groups were included in the study cohort, along with a random subsample of 10,000 individuals from the White ethnic group. Subsampling of the White ethnic group was performed to create a cohort with relatively balanced numbers of participants across ethnic groups, given that the majority of UK Biobank participants identify as White.

The final study cohort consists of 35,590 individuals from the Black, East Asian, South Asian, White, Mixed, and Other ethnic groups (Table 1). The Black ethnic group shows the highest average serum creatinine level (79.98), the South Asian (71.74), White (72.17), Mixed (71.15), and Other (71.69) groups show intermediate levels, and the East Asian group (66.37) shows the lowest average creatinine level. Ethnic group differences in serum creatinine levels stratified by sex are shown in Supplementary Table 4. Differences in genetic ancestry (GA) percentages across ethnic groups are shown for the four regional ancestry groups that correspond to the ethnic groups studied here: African, East Asian, South Asian, and European. SED is measured by the Townsend deprivation index, where lower (negative) values indicate less deprivation and higher (positive) values indicate more deprivation. The Black ethnic group has the highest average SED, by far (2.65), followed by the Other (0.99), Mixed (0.47), and South Asian (0.32) groups, all of which show positive average SED values. Both the East Asian (-0.49) and White (-1.43) ethnic groups show negative average SED values, indicative of less deprivation and relative affluence. The average age of the cohort is 54 years, with the White group showing the highest age. The cohort is 46% male on average; the East Asian and Mixed groups have the lowest percentage of males, and the South Asian group

**Table 3**
**J test comparing the creatinine model with ethnicity versus the creatinine model with ancestry.** Model 5: Creatinine ~ All ethnic groups + SED + Age + Sex + Height + BMI. Model 13: Creatinine ~ Genetic ancestry + SED + Age + Sex + Height + BMI.

| Models compared | Estimate (β) | Std. Error | t-value | P-value |
|---|---|---|---|---|
| Model 5 + fitted (Model 13) | 0.91 | 0.06 | 15.77 | $<2 \times 10^{-16}$ |
| Model 13 + fitted (Model 5) | 0.15 | 0.07 | 2.15 | 0.03 |

**Fig. 2. Ethnicity and genetic ancestry (GA).** (A) Principal Component Analysis (PCA) of UK Biobank participants showing the first three PCs. Individual participants are colored by their self-identified ethnicity (left) and their GA (right). Ethnicity and GA color-codes are shown in the key. (B) Participants' GA fractions for six global ancestry groups are shown, with participants grouped according to their self-identified ethnicity. Each participant is shown as a single column, with different ancestry components summing to 100%. The White group shows a random sample of 10,000 individuals.

**Table 4**

**Relative importance of ethnicity versus genetic ancestry for explaining variation in serum creatinine levels.** Model x:

| Predictor | Relative Importance[1] |
|---|---|
| Ethnicity | 4.96 (±0.59) |
| Genetic Ancestry (all) | 9.06 |
| African Ancestry | 6.36 (±1.29) |
| East Asian Ancestry | 1.29 (±0.32) |
| European Ancestry | 0.86 (±0.16) |
| South Asian Ancestry | 0.59 (±0.11) |

[1] Percent of variation (±95% CI) in serum creatinine levels explained by each predictor. Note that genetic ancestry (all) is the sum of variation across all four ancestry components.

is the only group with more males than females. There are differences in body size across ethnic groups – as measured by height, weight, and BMI – where Black and White group participants tend to be larger than participants from the other groups.

### 4.2. Ethnic differences in creatinine levels

We measured the association of ethnicity with serum creatinine levels using multivariable linear regression, controlling for SED, age, sex, and BMI. SED was included as a covariate since ethnic groups experience different levels of deprivation, which is known to be associated with health outcomes. Age and sex were included as covariates since they are known to be associated with creatinine levels. Height, weight, and BMI were all considered as potential covariates given the relationship between muscle mass and creatinine levels. Since these three body size measures are highly correlated, model fit comparisons and stepwise regression were used to choose the best set of size covariates to include in the final model, resulting in inclusion of height and BMI and the exclusion of weight (Models 1–4; Supplementary Table 5).

Taking the White ethnic group as the reference, the Black group shows the highest effect on serum creatinine levels in the multivariable model (Model 5; Fig. 1 and Supplementary Table 6). Controlling for all other model predictors, participants from the Black ethnic group have 9.36 μmol/L higher creatinine levels, on average, than participants from the White ethnic group. The Black and Mixed groups are the only ethnic groups that show significant positive effects on creatinine levels. The East Asian ethnic group shows a significant negative effect on creatinine levels, whereas the South Asian and Other ethnic groups do not show significant effects on creatinine levels. Age shows a slightly positive association with creatinine levels, and male sex shows the highest overall and positive effect on creatinine. Height and BMI are both slightly positively associated with creatinine levels. The positive effect of the Black ethnic group on creatinine levels, and the negative effect of the East Asian group remain in the unadjusted creatinine model, which includes ethnicity variables alone (Model 6; Supplementary Table 7). The South Asian, Mixed, and Other ethnic groups do not show significant effects on creatinine levels in the unadjusted model.

SED is not significantly associated with creatinine levels in the fully adjusted model (Model 5; Fig. 1 and Supplementary Table 6). When the cohort is stratified by sex and SED terciles (low, medium, high), there are virtually no differences in serum creatinine levels observed among

SED terciles and the same ethnic differences in creatinine can be seen within each tercile (Supplementary Fig. 3). Comparison of the fully adjusted model with a model of serum creatinine levels (outcome) without SED predictor variable shows that inclusion of SED does not attenuate the observed ethnic associations with creatinine (Model 7; Supplementary Table 8).

In the US, equations for estimating eGFR from serum creatinine levels have included a binary ethnicity correction factor for Black or non-Black patients; ethnic differences among non-Black patients are not considered. We assessed the utility of including information on other ethnic groups for predicting creatinine levels by comparing the fits for the creatinine model that includes all four ethnic groups versus a model that only includes Black and non-Black groups, using a likelihood ratio test (Models 5 & 8; Table 2). The full model shows a significantly better fit than the binary Black or not model, underscoring the contribution of other UK ethnic groups to the observed variation in creatinine levels. Nevertheless, the overall difference in the amount of variation explained by both models is small, consistent with the major effect of Black ethnicity on creatinine.

### 4.3. Ethnicity and genetic ancestry (GA)

GA refers to genetic similarities derived from common ancestors and reflects distinct allele frequency patterns found in ancestral populations that were reproductively isolated. In contrast to self-identified ethnicity, GA is a characteristic of the genome that can be objectively characterized via comparison with diverse global reference populations. Genomes of individuals from modern cosmopolitan populations, such as the UK, often show GA contributions from multiple ancestral source populations. GA was characterized for study cohort participants as a continuous variable reflecting percent ancestry contributions from each of six regional ancestry groups: African, Central Asian, East Asian, European, South Asian, and West Asian.

We characterized the GA of the cohort participants and compared the patterns of GA to their ethnic self-identity. Overall similarities and differences between ethnic self-identity and GA can be seen when individuals are mapped to the first three genomic principal components (PCs) and color-coded according to either ethnicity or GA (Fig. 2A). While there is broad concordance among participants' ethnic groups and their GA, there are also numerous individuals that show discordant patterns of ethnicity and GA. Moreover, participant genetic diversity is distributed along a continuum of PC values as opposed to forming discrete, coherent groups. More detailed views of the PCA results can be seen for both ethnicity (Supplementary Figure 4) and GA (Supplementary Figure 5).

Participant GA fractions were inferred from PCA data for six broad geographic regions: African, Central Asian, East Asian, European, South Asian, and West Asian (Fig. 2B). Overall, UK Biobank ethnic groups show clear differences in GA fractions, consistent with shared historical origins among group members. Nevertheless, there is substantial variation seen within ethnic groups, which is particularly true for the Mixed and Other ethnic groups. The East Asian and White ethnic groups show the most coherent within-group patterns of GA. The South Asian ethnic group shows predominantly South Asian ancestry with a substantial Central Asian component as well as smaller East Asian and European fractions. The Black ethnic group shows primarily African ancestry with a secondary European component and a minor West Asian component.
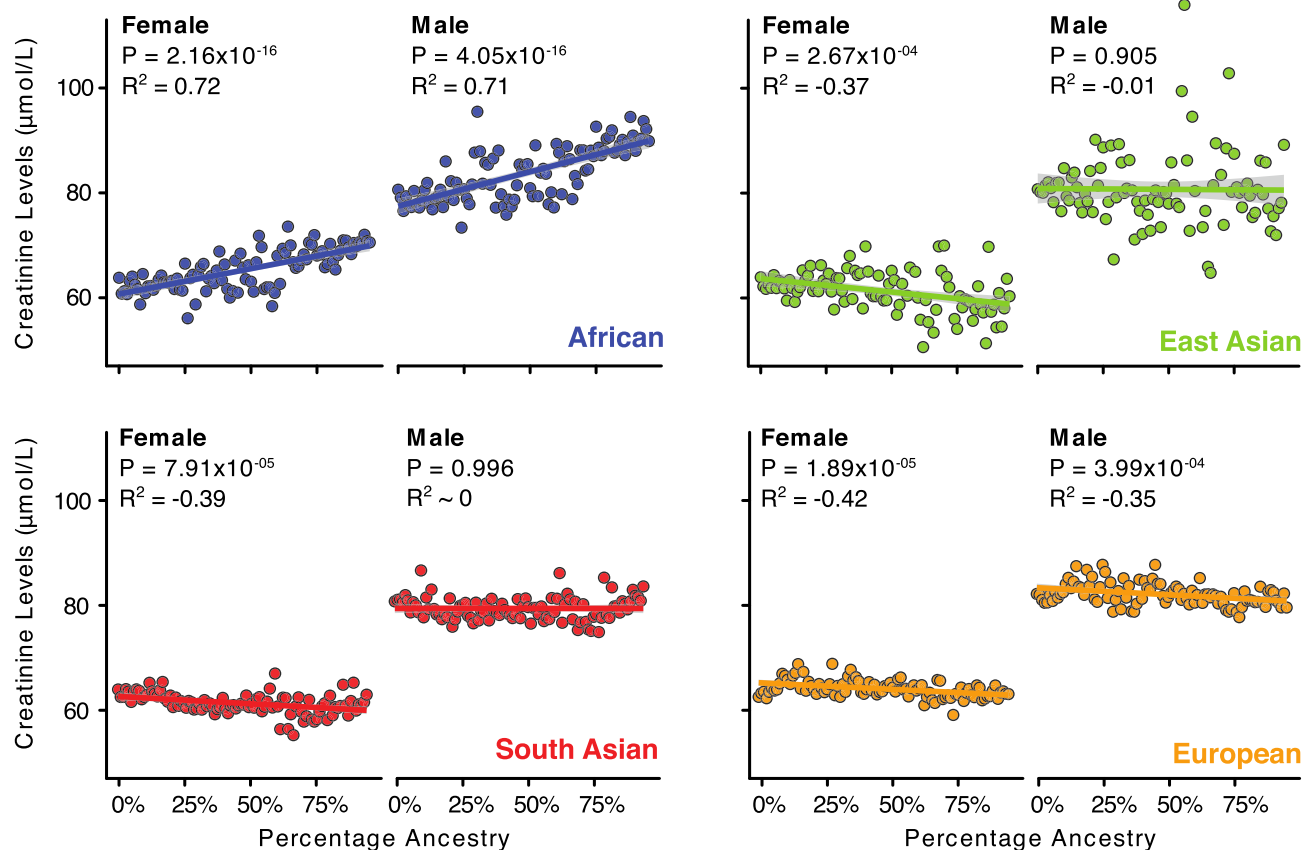


**Fig. 3. Genetic ancestry (GA) and serum creatinine levels.** Serum creatinine levels (y-axis) are regressed against percent GA (x-axis) for four ancestry groups: African (blue), East Asian (green), South Asian (red), and European (orange). Results for each ancestry group are stratified by sex. Participants are grouped into percentiles and linear trend lines with standard errors were fit with linear regression. The $R^2$ and P-values for Pearson correlation are shown for each ancestry-sex combination.

## 4.4. Genetic ancestry (GA) and creatinine levels

Participants' serum creatine levels were regressed against GA fractions for the four most prominent ancestry components observed in the study cohort: African, East Asian, South Asian, and European. African ancestry is the only ancestry component that shows a positive correlation with creatinine levels, and African ancestry shows by far the highest magnitude correlation with creatinine levels (Fig. 3). The remaining ancestry components show either negative correlations or no correlation with creatinine levels. Overall, the European ancestry component shows the next highest correlation with creatinine followed by the South Asian and East Asian ancestry components. These patterns hold for both males and females across all four ancestry components.

We further evaluated the associations between ancestry and serum creatinine levels by modeling the association of ancestry fractions for each ancestry separately adjusting for SED, age, sex, height, and BMI (Models 9–12; Supplementary Table 9). The results of the ancestry-specific models are consistent with the results of the regression analyses. African ancestry shows the largest effect size and the only positive effect size for all four ancestries, and the African ancestry model shows the best fit to creatinine levels for any of the models.

We evaluated the amount of variation in serum creatinine levels explained by ethnicity versus the amount explained by GA by comparing the specification of the creatinine model that includes ethnicity and covariates (Model 5; Supplementary Table 6) to the specification of the model that includes ancestry and covariates (Model 13; Supplementary Table 10). Serum creatinine levels have a slightly skewed distribution, and ancestry fractions are non-normally distributed, as most ethnic groups show a single majority ancestry component with much smaller contributions from other ancestries. We checked the residual distribution for the creatinine-ancestry model to ensure that the model estimates are reliable (Supplementary Figure 6). The ancestry model shows better specification, i.e. a more correct set of regressors, compared to the ethnicity model, using both the J test and the Cox test for non-nested model comparisons. For the J test, adding the fitted values of the ancestry model to the ethnicity model improves the model specification far more than adding the fitted values of the ethnicity model to the ancestry model (Model 5 & Model 13; Table 3). For the Cox test, fitting the regressors from the ancestry model to the fitted values of the ethnicity model has high explanatory value, whereas fitting the regressors from the ethnicity model to the fitted values of the ancestry model does not (Model 5 & Model 13; Supplementary Table 11).

We further compared the amount of variation in serum creatinine levels explained by ethnicity versus GA using relative importance analysis of a creatinine model that included both ethnicity and GA as predictors, along with SED, age, sex, height, and BMI covariates (Model 14; Table 4 & Supplementary Table 12). GA explains more of the variation in serum creatinine levels than ethnicity, and African ancestry alone explains more of the variation in creatinine levels than ethnicity. African ancestry explains 4.9 times more variation than the next highest East Asian ancestry and an order of magnitude more variation than European or South Asian ancestry.

## 5. Discussion

We modeled the association of ethnicity, GA, and SED with serum creatinine levels controlling for age, sex, height, and BMI. Black ethnicity shows a positive effect on serum creatinine levels, whereas East Asian and South Asian ethnicity have negative effects on creatinine levels. While Black ethnicity has the highest magnitude effect on creatinine levels, inclusion of additional ethnic groups contributes significantly to the observed variation in creatinine levels and slightly improves the model fit. African ancestry has the highest and the only positive GA association with serum creatinine levels. Overall, GA explains more of the variation in creatinine levels than ethnicity, yielding a significantly better model fit, with African ancestry alone explaining

more of the variation in creatinine levels than ethnicity.

It is possible that the association between serum creatinine levels and ethnicity is related to social determinants of health. Although GA is a characteristic of the genome, it could also co-vary with social factors, thereby confounding our analysis. We controlled for SED using the Townsend deprivation index and did not see any attenuation of the effects of ethnicity or GA. In fact, SED has no significant effect on serum creatinine levels, in contrast to the expectations of the social determinants hypothesis. Furthermore, when the cohort is stratified into low, medium, and high SED terciles, there is no difference in serum creatinine levels among terciles and the ethnic differences are the same within all terciles. These results suggest that social and environmental factors, at least as measured by the Townsend Index, have a negligible effect on serum creatinine levels. Nevertheless, the inclusion of additional, more detailed measures of the social determinants of health could shed light on the genetic versus environmental contributions to ethnic differences in creatinine levels.

### 5.1. Limitations and future directions

Our study was motivated by a desire to explore the empirical basis of ethnicity corrections for equations that calculate estimated GFR (eGFR) from patient serum creatinine levels. One limitation of the study is the absence of measured GFR (mGFR) data for our cohort participants. Without mGFR, we cannot determine whether the differences they observed in serum creatinine among groups defined by ethnicity or GA represent differences in mGFR or differences in the non-GFR determinants of serum creatinine, which has implications beyond the use of GFR estimating equations as a tool for assessment of GFR. Previous findings on the association of ethnicity on mGFR have been recently reviewed (Levey et al., 2020a; Levey et al., 2020b). These data do not suggest a large difference in mGFR between US Black and White participants enrolled in MESA, a community-based study, in which differences in serum creatinine were observed. Lower GFR has been observed healthy South Asians and Whites compared with historical data in Whites, but differences in mGFR methods and in diet may have contributed to the observed differences. Further study is required to understand these observations along with the findings presented here.

It must be stressed that this study was conducted on a cohort from the UK using ethnic group designations specific to the country. The extent to which our findings may be transferable to other countries like the US is an open question, particularly as it relates to the social dimension of ethnicity. While the UK ethnic groups studied here correspond approximately to US racial groups, the social experiences of these groups and the health implications of ethnicity and race may be distinct in different countries. Furthermore, country-specific patterns of migration could yield distinct patterns of GA and differences in the nature of ethnicity-ancestry covariation.

For this study, GA was calculated at the level of whole genomes with respect to broad geographical population groups. Future studies on the relationship between GA and serum creatinine levels could benefit from more fine-scale ancestry inference. For example, there are likely to be high levels of genetic diversity with the regional ancestry groups studied here, particularly for African ancestry. Delineation of African ancestry into more highly resolved ancestry components could shed additional light on the link between GA and creatinine. Local ancestry inference, i. e. the assignment of ancestral origins to specific haplotypes across the genome, could be used to further interrogate the genetic basis of our observations. In particular, the association of local ancestry segments with serum creatinine levels via admixture mapping could help to uncover specific genes (loci) that influence ethnic differences in creatinine levels.

### 5.2. The clinical utility of genetic ancestry: Pros and cons

UK Biobank participants self-identified their ethnicity based on

shared history, culture, and family ties, i.e. ethnic groups in the UK Biobank and elsewhere are socially determined. The social construction of these groups is underscored by the use of the term 'ethnicity' as opposed to 'race' in the UK census and health system, whereas similar groups are referred to as races in the US and elsewhere. Nevertheless, socially defined ethnic groups are considered as important epidemiologic proxies, since they co-vary with both genetic and non-genetic factors that can impact health outcomes (Borrell et al., 2021; Oni-Orisan et al., 2021). In contrast to socially defined ethnic groups, GA can be used to more accurately capture patterns of genetic diversity that are associated with patients' biogeographical origins (Yudell et al., 2016; Yudell et al., 2020).

GA provides a number of advantages for biomedical research and clinical practice. GA can be inferred directly from genetic data, independently of the social dimensions of ethnicity, and can therefore be calculated objectively and with precision. GA can be defined at different levels of relatedness, *e.g.* for broad geographical regions or for more closely related populations within regions. Finally, GA can be calculated at the level of whole genomes or at the local level with respect to the origins of individual haplotypes. Indeed, GA has been associated with a number of different health-related outcomes, including creatinine levels (Borrell et al., 2021).

GA more accurately reflects human genetic diversity, compared to racial and ethnic categories, and therefore allows for more precise stratification of patient populations, as seen here for the case of eGFR calculation from serum creatinine levels. Thus, the utility of using GA instead of race or ethnicity in clinical decision making is promising. Nevertheless, accurate calculation of patients' GA requires genome-wide genotype data of the kind analyzed here, or at the very least panels of scores or hundreds of ancestry informative markers (Ding et al., 2011; Schraiber and Akey, 2015). Racial and ethnic minorities who bear a disproportionate burden of disease are vastly underrepresented among clinical genetic study cohorts and far less likely to have access to their own genetic information (Bustamante et al., 2011; Petrovski and Goldstein, 2016; Popejoy and Fullerton, 2016). On the other hand, race and ethnicity, are useful proxies in their own right and readily available to health care providers, and prior to elimination of race and ethnicity in eGFR calculation, the impact of doing so on the exacerbation of existing health disparities must be carefully assessed.

While our manuscript was under review, a study of the Chronic Renal Insufficiency Cohort (CRIC) reported lower accuracy of eGFR for Black patients when race was eliminated from consideration in the estimation model (Hsu et al., 2021). The use of GA instead of race resulted in similar, and more accurate, eGFR estimates, consistent with the results reported here. They also show miscalculation of eGFR for Black patients could not be corrected when non-GFR determinants of serum creatinine were accounted for, similar to what we found. Taken together, the results of this recent study underscore the reliability of race as a proxy for GA and serum creatinine levels among Black patients. However, this study also showed that cystatin C could be used to accurately estimate GFR for Black patients without the use of race in the model. The authors stress that the use of cystatin C for GFR estimation, instead of creatinine, could mitigate potential negative consequences of race-based approaches.

## 6. Conclusion

We found that GA explains more of the variation in serum creatinine levels than socioeconomic deprivation or anthropometric factors, suggesting that ethnic differences in creatinine are shaped by genetic rather than social factors.

Author contributions

LMR and IKJ conceptualized the study and oversaw project administration. Funding acquisition was done by LMR and IKJ. The methodology was defined by LMR, SS, LR, ABC, SDN, and IKJ. SS, LR, and SDN curated the data, which LMR and IKJ verified. Data analysis was performed by SS, LR, ABC, and SDN. SS, LR, and SDN were responsible for data visualization. LMR, SS, and IKJ helped write the manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.146709.

## References

Arpegård, J., Viktorin, A., Chang, Z., de Faire, U., Magnusson, P.K.E., Svensson, P., 2015. Comparison of heritability of Cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. J. Am. Heart Assoc. 4 (1) https://doi.org/10.1161/JAHA.114.001467.

Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J.-F., Cann, H., Mallick, S., Reich, D., Sandhu, M.S., Skoglund, P., Scally, A., Xue, Y., Durbin, R., Tyler-Smith, C., 2019. Insights into human genetic variation and population history from 929 diverse genomes. bioRxiv, 674986.

Borrell, L.N., Elhawary, J.R., Fuentes-Afflick, E., Witonsky, J., Bhakta, N., Wu, A.H.B., Bibbins-Domingo, K., Rodríguez-Santana, J.R., Lenoir, M.A., Gavin, J.R., Kittles, R. A., Zaitlen, N.A., Wilkes, D.S., Powe, N.R., Ziv, E., Burchard, E.G., 2021. Race and genetic ancestry in medicine - a time for reckoning with racism. N. Engl. J. Med. 384 (5), 474–480.

Bustamante, C.D., De La Vega, F.M., Burchard, E.G., 2011. Genomics for the world. Nature 475 (7355), 163–165.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562 (7726), 203–209.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7.

Conley, A.B., Rishishwar, L., Norris, E.T., Valderrama-Aguirre, A., Marino-Ramirez, L., Medina-Rivas, M.A., Jordan, I.K., 2017. A Comparative Analysis of Genetic Ancestry and Admixture in the Colombian Populations of Choco and Medellin. G3 (Bethesda) 7, 3435–3447.

Cook, J.G.H., 1971. Creatinine assay in the presence of protein. Clin Chim Acta 32 (3), 485–486.

Ding, L., Wiener, H., Abebe, T., Altaye, M., Go, R.C., Kercsmar, C., Grabowski, G., Martin, L.J., Khurana Hershey, G.K., Chakorborty, R., Baye, T.M., 2011. Comparison of measures of marker informativeness for ancestry and admixture mapping. BMC Genomics 12, 622.

Drey, N., Roderick, P., Mullee, M., Rogerson, M., 2003. A population-based study of the incidence and outcomes of diagnosed chronic kidney disease. Am. J. Kidney Dis. 42 (4), 677–684.

Elliott, P., Peakman, T.C., Biobank, U.K., 2008. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int. J. Epidemiol. 37, 234–244.

Eneanya, N.D., Yang, W., Reese, P.P., 2019. Reconsidering the consequences of using race to estimate kidney function. JAMA 322 (2), 113. https://doi.org/10.1001/jama.2019.5774.

Foster, H.M.E., Celis-Morales, C.A., Nicholl, B.I., Petermann-Rocha, F., Pell, J.P., Gill, J. M.R., O'Donnell, C.A., Mair, F.S., 2018. The effect of socioeconomic deprivation on the association between an extended measurement of unhealthy lifestyle factors and health outcomes: a prospective analysis of the UK Biobank cohort. Lancet Public Health 3 (12), e576–e585.

Fox, J., Weisberg, S., 2019. An R companion to applied regression. Sage Publications, Thousand Oaks, CA.

Galinsky, K., Bhatia, G., Loh, P.-R., Georgiev, S., Mukherjee, S., Patterson, N., Price, A., 2016. Fast principal-component analysis reveals convergent evolution of ADH1B in europe and east asia. Am. J. Hum. Genet. 98 (3), 456–472.

Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. and Abecasis, G.R., 2015. A global reference for human genetic variation. Nature 526, 68-74.

Gov.uk, 2020. Ethnicity facts and figures.

Grömping, U., 2006. Relative imortance for linear regression in R: the package relaimpo. Journal of Statistical Software 17, 1-27.

Hsu, C.Y., Yang, W., Parikh, R.V., Anderson, A.H., Chen, T.K., Cohen, D.L., He, J., Mohanty, M.J., Lash, J.P., Mills, K.T., Muiru, A.N., Parsa, A., Saunders, M.R., Shafi, T., Townsend, R.R., Waikar, S.S., Wang, J., Wolf, M., Tan, T.C., Feldman, H.I., Go, A.S., Investigators, C.S., 2021. Race, genetic ancestry, and estimating kidney function in CKD. N. Engl. J. Med. 385, 1750–1760.

Jaffe, M., 1886. Method for measurement of creatinine in serum. Physiol. Chem. 10, 31.

Jordan, I.K., Rishishwar, L., Conley, A.B., Tishkoff, S.A., 2019. Native American admixture recapitulates population-specific migration and settlement of the continental United States. PLoS Genet 15 (9), e1008225.

Kertzer, D.I., Arel, D., 2002. Census, identity formation, and the struggle for political power. In: Kertzer, D.I., Arel, D. (Eds.), Census and identity: the politics of race, ethnicity, and language in national censuses. Cambridge University Press, Cambridge, pp. 1–42.

Levey, A.S., Bosch, J.P., Lewis, J.B., Greene, T., Rogers, N., Roth, D., 1999. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. modification of diet in renal disease study group. Ann. Intern. Med. 130, 461–470.

Levey, A.S., Coresh, J., Greene, T., Stevens, L.A., Zhang, Y.L., Hendriksen, S., Kusek, J. W., Van Lente, F., Epidemiology, C.K.D., C.,, 2006. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann. Intern. Med. 145, 247–254.

Levey, A.S., Coresh, J., Tighiouart, H., Greene, T., Inker, L.A., 2020a. Measured and estimated glomerular filtration rate: current status and future directions. Nat. Rev. Nephrol. 16 (1), 51–64.

Levey, A.S., Stevens, L.A., Schmid, C.H., Zhang, Y.L., Castro 3rd, A.F., Feldman, H.I., Kusek, J.W., Eggers, P., Van Lente, F., Greene, T., Coresh, J., Ckd, E.P.I., 2009. A new equation to estimate glomerular filtration rate. Ann. Intern. Med. 150, 604–612.

Levey, A.S., Tighiouart, H., Titan, S.M., Inker, L.A., 2020b. Estimation of Glomerular Filtration Rate With vs Without Including Patient Race. JAMA Intern. Med. 180 (5), 793. https://doi.org/10.1001/jamainternmed.2020.0045.

Levey, A.S., Titan, S.M., Powe, N.R., Coresh, J., Inker, L.A., 2020c. Kidney Disease, Race, and GFR Estimation. Clin. J. Am. Soc. Nephrol. 15 (8), 1203–1212.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319 (5866), 1100–1104.

Lüdecke, D., 2021. sjPlot: Data visualization for statistics in social science. R Foundation for Statistical Computing, Vienna, Austria.

Nagar, S.D., Nápoles, A.M., Jordan, I.K., Mariño-Ramírez, L., 2021. Socioeconomic deprivation and genetic ancestry interact to modify type 2 diabetes ethnic disparities in the United Kingdom. EClinicalMedicine 37, 100960. https://doi.org/10.1016/j.eclinm.2021.100960.

Oni-Orisan, A., Mavura, Y., Banda, Y., Thornton, T.A., Sebro, R., 2021. Embracing genetic diversity to improve black health. N. Engl. J. Med. 384 (12), 1163–1167.

Petrovski, S., Goldstein, D.B., 2016. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. Genome Biol 17, 157.

Popejoy, A.B., Fullerton, S.M., 2016. Genomics is failing on diversity. Nature 538 (7624), 161–164.

R Development Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ripley, B.D., 2002. Modern applied statistics with S. Springer, New York, NY.

Schraiber, J.G., Akey, J.M., 2015. Methods and models for unravelling human evolutionary history. Nat Rev Genet 16 (12), 727–740.

Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G.R., Wainberg, M., Ollila, H.M., Kiiskinen, T., Havulinna, A.S., Pirruccello, J.P., Qian, J., Shcherbina, A., Rodriguez, F., Assimes, T.L., Agarwala, V., Tibshirani, R., Hastie, T., Ripatti, S., Pritchard, J.K., Daly, M.J., Rivas, M.A., 2021. Genetics of 35 blood and urine biomarkers in the UK Biobank. Nat Genet 53 (2), 185–194.

Stevens, L.A., Levey, A.S., 2009. Measured GFR as a confirmatory test for estimated GFR. J. Am. Soc. Nephrol. 20 (11), 2305–2313.

Townsend, P., Phillimore, P., Beattie, A., 1988. Health and deprivation: inequality and the North. Routledge.

Udler, M.S., Nadkarni, G.N., Belbin, G., Lotay, V., Wyatt, C., Gottesman, O., Bottinger, E. P., Kenny, E.E., Peter, I., 2015. Effect of Genetic African Ancestry on eGFR and Kidney Disease. J. Am. Soc. Nephrol. 26 (7), 1682–1692.

Vyas, D.A., Eisenstein, L.G., Jones, D.S., 2020. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. N. Engl. J. Med. 383 (9), 874–882.

Welsh, S., Peakman, T., Sheard, S., Almond, R., 2017. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. BMC Genomics 18, 26.

Williams, D.R., Priest, N., Anderson, N.B., 2016. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. Health Psychol. 35 (4), 407–411.

Young, E.W., Mauger, E.A., Jiang, K.-H., Port, F.K., Wolfe, R.A., 1994. Socioeconomic status and end-stage renal disease in the United States. Kidney Int. 45 (3), 907–911.

Yudell, M., Roberts, D., DeSalle, R., Tishkoff, S., 2016. SCIENCE AND SOCIETY. taking race out of human genetics. Science 351 (6273), 564–565.

Yudell, M., Roberts, D., DeSalle, R., Tishkoff, S., 2020. NIH must confront the use of race in science. Science 369 (6509), 1313–1314.

Zeileis, A., Hothorn, T., 2002. Diagnostic checking in regression relationships. R Foundation for Statistical Computing, Vienna Austria.