Hepatocyte gene expression and DNA methylation as ancestry-dependent mechanisms in African Americans

C. S. Park, T. De, Y. Xu, Y. Zhong, E. Smithberger, C. Alarcon, E. R. Gamazon and M. A. Perera

African Americans (AAs) are an admixed population with widely varying proportion of West African ancestry (WAA). Here we report the correlation of WAA to gene expression and DNA methylation in AA-derived hepatocytes, a cell type important in disease and drug response. We perform mediation analysis to test whether methylation is a mediator of the effect of ancestry on expression. GTEx samples and a second cohort are used as validation. One hundred and thirty-one genes are associated with WAA (FDR < 0.10), 28 of which replicate and represent 220 GWAS phenotypes. Among PharmGKB pharmacogenes, VDR, PTGIS, ALDH1A1, CYP2C19, and P2RY1 nominally associate with WAA (p < 0.05). We find 1037 WAA-associated, differentially methylated regions (FDR < 0.05), with hypomethylated genes enriched in drug-response pathways. In conclusion, WAA contributes to variability in hepatocyte expression and DNA methylation with identified genes previously implicated for diseases disproportionately affecting AAs, including cardiovascular (PTGIS, PLAT) and renal (APOL1) disease, and drug response (CYP2C19).

npj Genomic Medicine (2019) 4:29 ; https://doi.org/10.1038/s41525-019-0102-y

INTRODUCTION

African Americans (AAs) are an admixed population, having varying proportions of African and European ancestry across individuals. As a consequence of their West African ancestry (WAA), AAs have more genetic variation and shorter extent of linkage disequilibrium than European ancestry populations, with the proportion of WAA varying greatly across self-identified AAs. The variability in the proportion of admixture in AAs may aid in explaining differences in hepatocyte gene expression and DNA methylation patterns that cannot be elucidated in homogeneous populations such as those of European-only ancestry. For example, WAA has been shown to predict a stronger inflammatory response to pathogens compared to those of European ancestry due to recent selective pressures specific to this population.

Furthermore, AAs suffer disproportionately from many chronic diseases and adverse drug reactions, as compared to other populations, as well as being protected from some conditions. As an example, AAs have a higher risk of cardiovascular events and negative outcomes to therapy such as the antiplatelet drug clopidogrel. They have higher incidences of death and disability from cardiovascular diseases (CVDs), thrombosis, renal dysfunction and pathologies, diabetes, cancers, and other metabolic disorders. Conversely, they have lower prevalence of disease such as testicular cancer. Differences in gene expression may help explain these observed differences.

Owing to the key role of the liver in biosynthesis, drug metabolism, and complex human diseases, genetic and epigenetic differences in the liver may be used to uncover the underlying causal genes responsible for chronic diseases that disproportionately affect AAs. Thus comprehensive mapping of liver expression quantitative trait loci (eQTLs) proposed several candidate susceptibility genes associated with type 1 diabetes, coronary artery disease and plasma cholesterol levels in a white cohort. More recently, finer-resolution mapping of liver eQTLs, combining both gene expression data with histone modification-based annotation of putative regulatory elements, identified 77 loci found to associate with at least one complex phenotype. In addition, our group has previously shown that studies specifically investigating the AA-specific genetic variants can reveal population-specific risk factors that may explain differences in drug response, such as African ancestry-specific genetic risk factors associated with a higher risk of bleeding from warfarin therapy in AAs, as well as population-specific variants associated with increased risk of thrombotic disease.

Rather than identifying disease susceptibilities through genome-wide association studies (GWAS), here we use the association of genomic ancestry to gene expression and DNA methylation to uncover potential drivers of disease and drug response that may explain differences in disease and drug response in AAs. Of the genes we identified, many are known to be dysregulated in diseases that disproportionately affect AAs, and others are associated with drug metabolism.

RESULTS

Cohorts

Sixty primary hepatocyte cultures, procured from self-identified AAs, passed all quality control (QC) steps and were used for RNA-sequencing (RNA-seq) analysis (Fig. 1). Fifty-six hepatocyte cultures were assayed for DNA methylation, with 44 used in the final methylation analysis (Fig. 1). We obtained genome-wide genotyping data on 153 subjects from the Genotype-Tissue Expression Project (GTEx) liver cohort, version 7, of which 15 AAs were used as a replication cohort (Fig. 1). All 60 of our AA hepatocyte cohort were confirmed as having WAA ranging from 41.8% to 93.7%, and 15 subjects in the GTEx liver cohort met the WAA inclusion criteria with WAA ranging from 75.6% to 99.9% (Supplementary Fig. S1). As a quantification of African ancestry,
WAA and principal component 1 (PC1) were highly correlated (Supplementary Fig. S2, Pearson correlation = −0.999). The Innocenti Liver Replication Cohort consisting of 23 AAs and 183 European Americans (EAs) was also used to validate our findings from the Hepatocyte Discovery Cohort (Fig. 1). Table 1 shows the demographics of each cohort. The GTEx Replication Cohort and the Innocenti Validation Liver Cohort are comparable to the Hepatocyte Discovery Cohort, but there are differences across groups in age (p = 0.0280, specifically between Hepatocyte Discovery Cohort and GTEx Replication Cohort, p = 0.0409), as well as sex (p = 0.0439) (Table 1). Importantly, there were no differences in proportion of WAA between the cohorts (Table 1).

Genes expressed in hepatocytes and pharmacogenes associated with African ancestry

Association analysis of RNA-seq gene expression traits in the Hepatocyte Discovery Cohort with percentage WAA identified 131 genes, for which gene expression traits were significantly associated with WAA (Supplementary Data S1; Fig. 2a, false discovery rate (FDR) < 0.10). We were able to replicate 28 of these genes in the independent Innocenti Liver Validation Cohort as differentially expressed (DE) between AAs and EAs in the liver (Table 2; Fig. 2d, p < 0.05). These 28 validated genes are associated with 220 reported and mapped disease and measurement traits that have been curated in the NCBI GWAS catalog. Of these, 118 are unique diseases or traits (Supplementary Data S2). The phenotypic disease and quantitative GWAS traits associated with WAA include blood and blood pressure measures, coronary heart and artery disease, diabetic blood measures, chronic inflammatory disease, chronic kidney disease, and various cancers.

In the GTEx Liver Replication Cohort of 15 AAs, we were able to replicate 8 of the 131 significant WAA-associated genes: DHODH (effect size [confidence interval (CI)] = 3.19 [0.03, 6.34], p = 0.048), GPI (effect size [CI] = −3.46 [−6.88, −0.04], p = 0.048), HSD17B7P2 (effect size [CI] = −9.22 [−17.0, −1.44], p = 0.027; PLCL2 (effect size [CI] = −5.76 [−11.38, −0.15], p = 0.046), SLC2A3 (effect size [CI] = −5.95 [−11.17, −0.73], p = 0.032), TRIM39 (effect size [CI] = 5.53 [0.75, 10.30], p = 0.030), VEGFA (effect size [CI] = −4.74 [−8.93, −0.56], p = 0.032), and COL26A1 (effect size [CI] = −7.13 [−13.67, −0.60], p = 0.03), where the effect size represents fold change of the corresponding gene’s expression with increased African ancestry (Supplementary Data S1).
were HYPRM (Supplementary Fig. S3d). IGR, 5 CpG sites (15.1% vs 4.8%, proportion of HYPOM CpGs located in islands compared to HYPRM

**Table 1. Demographics and clinical characteristics of hepatocyte/liver cohorts**

| Variable                      | Hepatocyte Discovery Cohort | GTEx Liver Cohort | Innocenti Validation Liver Cohort |
|-------------------------------|----------------------------|------------------|----------------------------------|
|                               | RNA-Seq cohort             | Methylation cohort|                                   |
| Percentage of AA subjects (%) | AA (n = 60)                | AA (n = 44)       | AA (n = 15)                       |
| Age, years* (mean ± SD)       | 100                        | 100              | 100                              |
| Sex (female %)**              | 48.3                       | 52.3             | 20                               |
| West African* ancestry (%) (mean ± SD) | 78.17 ± 12.86    | 79.83 ± 11.86    | 84.04 ± 7.13                     |

*AA: African Americans >40% WAA
**EA: European Americans with <1% WAA
*p = 0.0280 for age (between groups, one-way ANOVA), p = 0.0409 (between RNA-Seq and GTEx Liver cohorts)
**p = 0.0439 for sex
**p = 0.2331 for ancestry

DNA methylation patterns in hepatocytes are associated with African ancestry

To identify differentially methylated (DM) regions (DMRs) and CpGs associated with WAA, we performed linear regression on each CpG site.26 We identified 23,317 significant DM CpG sites, out of a total of 867,531 probes on the Illumina EPIC BeadChip microarray, annotated to 11,151 unique genes (Supplementary Fig. S3a; Benjamini–Hochberg (BH)-adjusted p < 0.05). These DM CpGs correspond to 1037 DMRs annotated to 1034 unique genes (Supplementary Fig. S3b; minimum FDR < 0.0001). Each DM CpG site was categorized into hypermethylated (HYPRM) and hypomethylated (HYPOM) sites: 15,404 HYPRM CpG sites constituted 7913 HYPRM DMRs, mapping to 602 unique genes; 7913 HYPOM CpG sites constituted HYPOM DMRs 602 unique genes comprised of the 602 HYPOM genes; 7913 HYPOM CpG sites constituted HYPOM DMRs 602 unique genes comprised of the 602 HYPOM genes; and 7 annotated genes had both HYPRM and HYPOM DMRs.

As compared to all CpG sites tested, the gene body (45.0% vs 40.9%, p = 0.0001, chi-squared test) and shelf (8.4% vs 6.9%, p = 0.011, chi-squared test) had significantly higher proportions of DM CpGs, while the promoter (18.4% vs 20.3%, p = 0.002, chi-squared test), intergenic regions (IGR) (24.2% vs 27.7%, p = 0.0001, chi-squared test), and shore regions (16.3% vs 18.2%, p = 0.0002, chi-squared test) had significantly lower proportions of DM CpGs. In our analysis of DM loci associated with WAA, 75.9% of DM CpG sites within islands were HYPOM, while the shore, shelf, and open sea were predominantly HYPRM (65.5%, 81.7%, and 78.9% respectively; Supplementary Fig. S3c). Within transcriptionally regulated promoter regions, 54.0% of CpG sites were HYPOM (Supplementary Fig. S3d).

Within the promoter, 71.5% of DM CpGs were HYPOM 200 kb upstream of the transcription start site (TSS), while 42.9% of DM CpGs were HYPOM 1500 kb upstream of the TSS (Supplementary Fig. S3d). DNA methylation around TSS is an established predictor of gene expression, with increased methylation leading to decreased expression.27,28 In the gene body, 73.0% of DM sites were HYPRM (Supplementary Fig. S3d), IGR, 5'-untranslated region (UTR), and 3'-UTR were predominantly HYPRM (67.5%, 64.0%, and 79.6% respectively).

Next, we characterized the locations of WAA-associated CpGs by genomic features and gene annotations. Within the 7913 HYPOM CpG sites associated with WAA, there was a greater proportion of HYPOM CpGs located in islands compared to HYPRM CpG sites (15.1% vs 4.8%, p < 0.0001, chi-squared test; Supplementary Fig. S3c). Within the 15,404 HYPRM CpGs associated with WAA, there was a greater proportion of HYPRM CpGs compared to HYPOM CpGs in the open sea regions (43.8% vs 11.7%, p < 0.0001, chi-squared test), shelf (6.9% vs 1.5%, p < 0.0001, chi-squared test), and shore (10.6% vs 5.6%, p < 0.0001, chi-squared test) (Supplementary Fig. S3c).

Gene expression trait correlation to WAA-associated DNA methylation patterns

To determine the relationship of WAA-associated DMRs with gene expression traits, we looked at the association of the 1034 unique genes corresponding to the 1037 DMRs with their respective gene expression profiles. Although there was no correlation of WAA-associated HYPRM gene regions with gene expression (Fig. 2c, Pearson’s r = 0.036, p = 0.55), HYPOM gene regions associated with WAA were negatively correlated with gene expression (Fig. 2c, Pearson’s r = −0.14, p = 0.009). In general, all genes within DMRs associated with WAA also showed negative correlation between gene expression and methylation (Pearson’s r = −0.1, p = 0.013).

From the 131 gene expression traits significantly associated with WAA, we identified an overlap of ten DM gene regions (Supplementary Data S3a). Of these ten, five genes (COL26A1, HIC1, MKNK2, RNF135, and TRIM39) showed concordant directions of effect (e.g., increased methylation leading to decreased gene expression). To determine whether methylation mediated the association of WAA to gene expression in these five genes, we performed mediation analysis. We found that the expression of two of these genes was mediated by methylation: COL26A1 (mediated effect = 2.336, 95% CI = 0.563–4.39, p < 2E–16) and MKNK2 (mediated effect = 1.095, 95% CI = 0.016–2.61, p = 0.046) (Supplementary Data S3b and S3c). A comprehensive Circos plot summarizes these findings: 23,317 HYPOM and HYPRM CpG sites, 1037 DMRs, 131 genes significantly associated with WAA, and 220 GWAS disease traits from the GWAS Catalog that are associated with the 28 validated genes (Fig. 3, Supplementary Data S1).

Functional representation of WAA-associated genes and DM genes associated with African ancestry

To understand the biological relevance of differentially HYPRM and HYPOM genes associated with WAA, we performed Gene Ontology (GO) analysis of the 432 unique genes comprised of the 435 HYPRM DMRs and 602 unique genes comprised of the 602 HYPOM DMRs using a gene panel of all annotated genes in the GO database. HYPRM genes are enriched for “cell development” (BH-adjusted p = 0.0016) and “apoptotic process” (BH-adjusted p = 0.032), and increased expression.27,28 In the gene body, 73.0% of DM sites showed concordant directions of effect (e.g., increased methylation leading to decreased gene expression).

As compared to all CpG sites tested, the gene body (45.0% vs 40.9%, p = 0.0001, chi-squared test) and shelf (8.4% vs 6.9%, p = 0.011, chi-squared test) had significantly higher proportions of DM CpGs, while the promoter (18.4% vs 20.3%, p = 0.002, chi-squared test), intergenic regions (IGR) (24.2% vs 27.7%, p = 0.0001, chi-squared test), and shore regions (16.3% vs 18.2%, p = 0.0002, chi-squared test) had significantly lower proportions of DM CpGs. In our analysis of DM loci associated with WAA, 75.9% of DM CpG sites within islands were HYPOM, while the shore, shelf, and open sea were predominantly HYPRM (65.5%, 81.7%, and 78.9% respectively; Supplementary Fig. S3c). Within transcriptionally regulated promoter regions, 54.0% of CpG sites were HYPOM (Supplementary Fig. S3d).

Within the promoter, 71.5% of DM CpGs were HYPOM 200 kb upstream of the transcription start site (TSS), while 42.9% of DM CpGs were HYPOM 1500 kb upstream of the TSS (Supplementary Fig. S3d). DNA methylation around TSS is an established predictor of gene expression, with increased methylation leading to decreased expression.27,28 In the gene body, 73.0% of DM sites were HYPRM (Supplementary Fig. S3d), IGR, 5'-untranslated region (UTR), and 3'-UTR were predominantly HYPRM (67.5%, 64.0%, and 79.6% respectively).

Next, we characterized the locations of WAA-associated CpGs by genomic features and gene annotations. Within the 7913 HYPOM CpG sites associated with WAA, there was a greater proportion of HYPOM CpGs located in islands compared to HYPRM CpG sites (15.1% vs 4.8%, p < 0.0001, chi-squared test; Supplementary Fig. S3c). Within the 15,404 HYPRM CpGs associated with WAA, there was a greater proportion of HYPRM CpGs compared to HYPOM CpGs in the open sea regions (43.8% vs 11.7%, p < 0.0001, chi-squared test), shelf (6.9% vs 1.5%, p < 0.0001, chi-squared test), and shore (10.6% vs 5.6%, p < 0.0001, chi-squared test) (Supplementary Fig. S3c). By genomic annotation, the promoter had greater proportion of HYPOM CpG sites (46.0% vs 54.0%, p < 0.0001, chi-squared test), while the gene body (72.9% vs 27.1%, p < 0.0001, chi-squared test), IGR (67.5% vs 32.5%, p < 0.0001, chi-squared test), 5'-UTR (63.9% vs 36.1%, p < 0.0001, chi-squared test), and 3'-UTR (79.5% vs 20.5%, p < 0.0001, chi-squared test) had greater proportions of HYPRM CpG sites (Supplementary Fig. S3d).
0.0076) within the ontology of biological processes (Fig. 4a). HypoM genes were enriched for “system development” (BH-adjusted $p = 4.0 \times 10^{-6}$), “response to drug” (BH-adjusted $p = 1.7 \times 10^{-4}$), and “response to hypoxia” (BH-adjusted $p = 0.0068$) within the ontology of biological processes (Fig. 4b). In addition, HypoM genes are enriched for “sequence-specific DNA binding” (BH-adjusted $p = 4.4 \times 10^{-5}$) and “RNA polymerase II transcription factor activity” (BH-adjusted $p = 4.4 \times 10^{-5}$) within the ontology of molecular functions (Fig. 4b).

With respect to GO analysis of the 131 WAA-associated gene expression traits in hepatocytes (FDR < 0.10), “angiogenesis” (BH-adjusted $p = 4.5 \times 10^{-3}$), “leukocyte activation involved in immune response” (BH-adjusted $p = 4.5 \times 10^{-3}$), “acute-phase response” (BH-adjusted $p = 0.0011$), “positive regulation of endothelial cell proliferation” (BH-adjusted $p = 0.0055$), “zymogen activation” (BH-adjusted $p = 0.0059$), and “cell proliferation” (BH-adjusted $p = 0.0085$) were biological processes that showed significant enrichment (Fig. 4c). Molecular functions, including “oxidoreductase activity” (BH-adjusted $p = 9.6 \times 10^{-5}$), “signaling receptor binding” (BH-adjusted $p = 0.0023$), “glucocorticoid receptor binding” (BH-adjusted $p = 0.0045$), and the KEGG pathway “HIF-1 signaling pathway” (BH-adjusted $p = 8.2 \times 10^{-4}$), were also enriched (Fig. 4c).

**DISCUSSION**

Several studies have shown that the first several PCs of methylation data can capture population structure in cohorts composed of European and African ancestry individuals. Recently, it was shown that genetic ancestry can be used as a proxy for uncovering unknown covariates contributing to epistatic and gene–environment interactions in both gene expression and DNA methylation data. A recent study reported that approximately 75% of variation in DNA methylation was attributable to shared genomic ancestry. This variability in epigenetic processes, accounted for by ancestry, however, may include a genetic component, demographic history (e.g., Transatlantic migration), environmental exposures, and a disease-altered component. Moreover, clinically meaningful measures can be associated with the proportion of WAA, as has been shown for lung function prediction.
In our study, we investigated population-specific gene expression and DNA methylation in AAs, an admixed population. We found HypoM genes, which indicate increased gene expression, are enriched for “system development” and “response to drug.” HypoM, or demethylation, may be a better predictor of gene expression than methylation, which, in contrast to demethylation, may or may not affect gene expression depending on the gene region methylated or the pattern of methylation. For example, methylation within the TSS of the promoter is well known to repress gene expression while methylation within the gene body may or may not affect gene expression depending on the gene. HypoM, or demethylation, may be a better predictor of gene expression than methylation, which, in contrast to demethylation, is involved in beneficial immune response and VEGFA, which has been linked to renal disease and microvascular complications of diabetes. Other genes associated with WAA include ALDH1A1, which is involved in alcohol and aldehyde metabolism disorders and cancer risk; IL-33, which is involved in beneficial immune response, and VEGFA, which has been linked to renal disease and microvascular complications of diabetes.

Among the PharmGKB VIP genes, we identified five genes associated with WAA: ALDH1A1, CYP2C19, P2RY1, PTGIS, and VDR. Of particular importance, we found that, for every 1% increase in African ancestry, there was a corresponding 1.36% decrease in CYP2C19 expression and 1.61% increase in P2RY1 expression. CYP2C19 is involved in the metabolism of many commonly prescribed drugs, including clopidogrel, an antplatelet therapy widely used for thrombo-prophylaxis of CVDs and linked to substantial inter-patient differences in drug response. It is also known to inhibit the P2RY family of receptor proteins on the surface of platelets. By inhibiting P2RY2 function, clopidogrel indirectly suppresses platelet clustering and clot formation and prevents clots contributing to heart attack, stroke, and deep vein thrombosis. Clopidogrel indirectly suppresses platelet clustering and clot formation and prevents clots contributing to heart attack, stroke, and deep vein thrombosis. Clotting requires CYP2C19-mediated conversion to its active form. It has been shown that different populations have different levels of CYP2C19 activity. The underlying mechanism is multifactorial and genetic polymorphisms contribute to variable drug response within an individual and across populations. A study conducted across 24 U.S. hospitals showed 1-year mortality rates of 7.2% in clopidogrel-treated AAs, compared to 3.6% for Caucasians on clopidogrel. This study also found that AAs were at a higher risk of cardiovascular events and mortality from poor antiplatelet response to clopidogrel. Our finding that CYP2C19 expression is reduced with WAA while P2RY1 expression is inversely increased suggests that clopidogrel resistance and susceptibility of AAs to thrombotic disease may be due to ancestry-associated gene expression as an underlying mechanism.

Another WAA-associated gene was PLAT, for which gene expression is decreased with increased WAA. The PLAT gene, which is involved in plasminogen activation and encodes tissue plasminogen activator (t-PA), is linked to thrombosis and increased risk of CVD. In AAs, polymorphisms in PLAT have been linked to CVD and higher levels of t-PA antigen have been observed in both myocardial infarction and venous thromboembolism. In addition, increased plasma fibrinogen level, which is involved in the fibrinolytic pathway and regulated by t-PA, has been linked to increased venous thrombosis risk in AAs.

VDR is also important in health and disease because Vitamin D and its active metabolite 1,25(OH)2D are exogenous hormones...
created by sun exposure or through diet, with higher risk for deficiencies in both Vitamin D and its bioactive metabolites in AAs. Those of African ancestry are known to have lower plasma 1,25(OH)₂D levels. One potential mechanism may be the upregulation of VDR with increased WAA to compensate for these lower levels. However, focused studies would need to confirm this hypothesis. In addition, VDR single-nucleotide polymorphisms (SNPs) have been implicated in CVDs and various cancer susceptibilities in AAs. SNPs within VDR may contribute to an already deficient Vitamin D environment in AAs.

Several limitations exist in this study. First, we were limited by the 60 primary hepatocyte cultures used in this analysis, which reduced our power to detect small changes in gene expression associated with ancestry. Second, the GTEx replication liver cohort, with 15 AA livers, also lacked power to replicate our findings (although we supplement this with a much larger Innocenti Validation Liver dataset for differential expression analysis.)
between AAs and EAs). Third, while most of the genes found in the PharmGKB VIP subset analysis were not statistically significant in the comprehensive genome-wide analysis, we found ALDH1A1 and PTGIS to be significantly associated with WAA in both the genome-wide and in the 64 PharmGKB VIP analysis, and we replicated 8 of the 131 significant gene expression traits associated with WAA (FDR < 0.10). p Values are BH-adjusted p values obtained from gProfiler.

Another differentiation, though not considered a limitation, was that we used cultured hepatocytes as opposed to frozen liver tissues, as was the case with GTEx. Primary cultures may show differences in gene expression profiles from those seen in the intact organ, which consist of approximately 60% hepatocytes. However, our study design has the advantage of assaying only the gene expression of a single-cell type as opposed to the multiple-cell types found in the liver. Previous studies have shown that primary human hepatocytes show similar gene expression levels for both Phase I (CYPs) and Phase II (e.g., UGTs) drug-metabolizing enzymes to those obtained from frozen liver tissue. Also, the use of primary hepatocyte cultures reduces the effect of environmental confounders inherent in liver tissue (i.e., there is less effect of previous drug/disease exposure) due to controlled tissue culture processes following hepatocyte isolation. The artifact of previous drug/disease exposures is typically present in all transcriptome studies conducted in postmortem human liver tissue. However, as with other studies, we too derived our cultures from non-transplantable livers, which may bias our study to less healthy individuals. Because of the anonymous nature of these samples, we also lack clinical and environmental factors that may affect gene expression.

Currently, genome-wide genetic, epigenetic, and multi-omics datasets of AAs are generally lacking in both the scientific literature and in public databases and repositories. Since genomics studies should be inclusive of all populations to comprehensively unravel disease etiology, the dearth of genetic data in diverse and underrepresented populations poses a major scientific and medical dilemma in new drug development, precision medicine, and public health policies. An archetypal example is the rs12777823 SNP in CYP2C9, which was found to associate with lower warfarin dose requirement, but only in AAs.

Fig. 4 Enrichment of biological processes and molecular functions of differentially methylated genes associated with West African ancestry and corresponding gene expression traits. a Gene ontology terms that are enriched for biological processes (BP) and molecular functions (MF) for 432 genes annotated to differentially hypermethylated regions, b 602 genes annotated to differentially hypomethylated regions, and c 131 genes with gene expression traits associated with WAA (FDR < 0.10). p Values are BH-adjusted p values obtained from gProfiler.
Findings, such as these, were only made possible due to focused studies of an underrepresented minority patient population.

In conclusion, our study has important implications in the use of genetic ancestry in understanding phenotypic differences and health disparities in AAs. Our study also has major implications for future investigations of genetic factors and potential for drug-response outcomes in admixed populations. As evidenced by the limited genome-wide data in AAs in public databases and biobanks, our study further illustrates the need for inclusion of diverse populations.

METHODS

Cohorts
A total of 68 AA primary hepatocyte cultures were used for this study. Cells were either purchased from commercial companies (BioIVT, TRL, Life technologies, Corning, and Xenotech) or isolated from cadaveric livers using a modified two-step collagenase perfusion procedure. Liver specimens were obtained through collaborations with Gift of Hope, which supplies non-transplantable organs to researchers. In addition, we used specimens were obtained through collaborations with Gift of Hope, which for approval as this study used human samples obtained from deceased individuals and was thus not considered human subjects’ research.

Primary hepatocyte isolation
Cadaveric livers obtained from Gift of Hope were transferred to the perfusion vessel Büchner funnel (Carl Roth) and the edge was carefully examined to locate the various vein and artery entries that were used for perfusion. Curved irrigation cannulae with olive tips (Kent Scientific) were inserted into the larger blood vessels on the cut surface of the liver piece. The liver was washed by perfusion of 1 L Solution 1 (HEPES buffer, Sigma-Aldrich), flow rate 100–400 mL/min, with no recirculation, followed by perfusion with 1 L of Solution 2 (EGTA buffer, Sigma-Aldrich), flow rate 100–400 mL/min, with no recirculation. The tissue was washed to remove the EGTA compound by perfusion of 1 L Solution 1, flow rate 100–400 mL/min, with no recirculation. The liver was digested by perfusion with Solution 3 (collagenase buffer, Sigma-Aldrich), flow rate 100–400 mL/min, with recirculation. Following perfusion, liver section was placed in a crystallizing dish (Omnimag) containing 100–200 mL of Solution 4 (Bovine Serum Albumin, Sigma-Aldrich). The Glisson’s capsule was carefully removed and the tissue was gently shaken to release hepatocytes. The cell suspension was then filtered through a 70-μm nylon mesh (Fisher Scientific), and centrifuged at 72 x g for 5 min at 4 °C. The pellets contained hepatocytes that were washed twice with Solution 4 and resuspended in plating medium (Fisher Scientific).

For primary hepatocyte cultures, cell viability was determined by trypan blue (Lonza) exclusion using a hemocytometer (Fisher Scientific). If viability was low, Percoll gradient (Sigma-Aldrich) centrifugation of cell suspensions was carried out to improve the yield of viable cells. Cell were plated at a density of 0.6 x 10^6 cells/well in 500 μL InVitroGro-CP media (BioIVT, Baltimore, MD) in collagen-coated plates with matrigel (Corning, Bedford, MA) overlay and incubated overnight at 37°C. Cultures were maintained in InVitroGro-HI media (BioIVT) supplemented with Torpedo antibiotic mix (BioIVT) per the manufacturer’s instructions. The media was replaced every 24 h for 3 days. RNA was extracted after 3 days using the RNAeasy Plus Mini-Kit (Qiagen) per the manufacturer’s instructions.

Genotyping and QC
DNA was extracted from 1–2 million cells from each hepatocyte line using the Gentra Puregene Blood Kit (Qiagen) as per the manufacturer’s instructions. All DNA samples were then bar-coded for genotyping. SNP genotyping was conducted on the Illumina Multi-Ethnic Genotyping array (MEGA) at the University of Chicago’s Functional Genomics Core using standard protocols.

QC steps were applied, as previously described, with imputation info metric threshold of >0.4. Briefly, a sex check was performed on genotypes using PLINK (version 1.9) to identify individuals with discordant sex information. Duplicated or related individuals were identified using identity-by-descent method with a cutoff score of 0.125, which indicates third-degree relatedness. A total of five individuals were excluded after genotyping QC analysis. SNPs located on the X and Y chromosomes and mitochondrial SNPs were excluded. SNPs with a missing rate of >5% or those that failed Hardy–Weinberg equilibrium tests (p < .00001) were also excluded.

African ancestry measurement
The genotypes of 68 primary hepatocytes and 153 GTEx subjects were merged with HapMap phase 3 reference data from four global populations: Yoruba in Ibadan, Nigeria (YRI); Utah residents with Northern and Western European ancestry (CEU); Han Chinese in Beijing, China (CHB); and Japanese in Tokyo, Japan (JPT). Population structure of the merged data was inferred by the Bayesian clustering algorithm STRUCTURE deployed within fastStructure v1.0 and performed without any prior population assignment. We employed the admixture model, and the burn-in-period and number of Markov Chain Monte Carlo repetitions were set to 20,000 and 100,000, respectively. The number of parental populations (K) was set to 3, purporting three main continental groups (African, European, or Asian). WAA percentages of the primary hepatocytes and GTEx subjects were calculated as the probability of being grouped as Yoruba African, Caucasian, and East Asian, respectively. All individuals in our 60 AA cohort had WAA >40%.

GTEx replication liver cohort
Of the 153 GTEx liver IDs extracted from “GTEx_Sample_Attributes” file, five were replicates and therefore removed from the analysis. Among the remaining 139 unique liver IDs, 118 had available genotype information for ancestry determination. Individuals with WAA >40% were included in the GTEx AA Liver cohort. After WAA estimation, normalized gene expression reads for 15 subjects meeting the ancestry inclusion criteria were extracted from GTEx expression file (GTEx_Liver.v7.normalized_expression.bed). Age and sex information of these subjects were extracted from the subject phenotype file on the GTEx Portal site (GTEx_v7_Annotations_SubjectPhenotypesDS.txt). Genes were deemed replicated if they showed both significant association with WAA (p < 0.05) and concordant direction of effect.

Validation analysis in an independent liver transcriptome dataset
Gene expression and genotype information for the Innocenti Liver Validation Cohort was obtained from the GEO database (Accession: GSE124076). The gene expression profiling in the liver had previously been conducted in 206 samples using Agilent-014850 4 x 44 k arrays (GPL4133). These samples had come from donor livers not intended for organ transplantation. Genotyping on these samples had been done using the Illumina Human 610–Quad beadchip platform (GPL8887) at the Northwestern University Center for Genetic Medicine Genomics Core Facility and imputation was subsequently performed using BIMBAM. Principal component analysis (PCA) was used to quantify ancestry using the Human Genome Diversity Panel with African and European populations as reference, as previously described.

We conducted differential expression analysis between the European samples and the AA samples using Linear Models for Microarray Data (limma) as implemented in the Bioconductor package. This Bayesian methodology uses a “moderated” t-statistic from the posterior variance assuming an inverse chi-squared prior for the unknown variance for a gene. We used Bonferroni-adjusted p < 0.05 based on the total number of genes that were tested for replication. Genes were deemed validated if they were significantly DE between populations (p < 0.05) with a concordant direction of effect (i.e., increased expression in AAs and increased expression with higher WAA).

RNA-sequencing and QC
Total RNA was extracted from each primary hepatocyte culture after 3 days in culture using the Qiagen RNeasy Plus Mini-Kit per the manufacturer’s instructions. RNA-QC was performed using an Agilent Bio-analyzer and samples with RNA integrity number scores >8 were used in subsequent sequencing. RNA-seq libraries were prepared for sequencing using the Illumina mRNA TruSeq RNA Sample Prep Kit, Set A (Illumina catalog # FC-122-1001) according to the manufacturer’s instructions. The cDNA libraries of 38 and 22 samples were prepared and sequenced using Illumina HiSeq 2500 and HiSeq 4000 machines, respectively, by the University of Chicago’s C.S. Park et al. npj Genomic Medicine (2019) 29 Published in partnership with CEGMR, King Abdulaziz University

npj Genomic Medicine (2019) 29

Published in partnership with CEGMR, King Abdulaziz University
Functional Genomics Core to produce single-end 50 bp reads with approximately 50 million reads per sample. Batch effects were corrected for in QC below.

Quality of the raw reads was assessed by FastQC (version 0.11.2). Fastq files with a per base sequence quality score >20 across all bases were included in downstream analysis. Reads were aligned to human Genome sequence GRCh38 and comprehensive gene annotation (GENCODE version 25) by STAR 2.5. Only uniquely mapped reads were retained and indexed by SAMTools (version 1.2). Nucleotide composition bias, GC content distribution, and coverage skewness of the mapped reads were further assessed using read_NVC.py, read_GC.py, and geneBody_coverage.py scripts, respectively, from RSeQC (version 2.6.4). Samples without nucleotide composition bias or coverage skewness and with normally distributed GC content were reserved. Lastly, Picard CollectRnaSeqMetrics (version 2.1.1) was applied to evaluate the distribution of bases within transcripts. Fractions of nucleotides within specific genomic regions were measured for QC and samples with >80% of bases aligned to exons and UTRs were considered for subsequent analysis.

RNA-seq data analysis
Post alignment and QC, reads were mapped to genes referenced with comprehensive gene annotation (GENCODE version 25) by HTseq (version 0.6.1p1) with union mode and minqual = 20. HTseq raw counts were supplied for gene expression analysis using Bioconductor package DESeq2 (version 1.20.0). 

Counts were normalized using regularized log transformation and PCA was performed in DESeq2. PC1 and PC2 were plotted to visualize samples' expression pattern. Three samples with distinct expression patterns were excluded as outliers resulting in 60 samples used in RNA-seq analysis. We calculated TPM (transcript per million) by first normalizing the counts by gene length and then normalizing by read depth. Gene expression values were filtered based on the expression thresholds >0.1 TPM in at least 20% of samples and ≥6 reads in at least 20% of samples as performed by the GTEx consortium (https://gtexportal.org). Analysis Methods, V7, updated 09/05/2017).

WAA percentage, gender, age, platform, and batch were used as covariates for downstream analysis. Probabilistic estimation of expression residuals (PEER) method v1.3 was used to identify PEER factors. Linear regression was run on inverse normal transformed expression data with five PEER factors as covariates, on the basis of GTEx's determination of the number of factors as a function of sample size. WAA-associated genes were identified from a genome-wide list of 18,854 genes for our hepatocyte-derived AA cohort samples and replicated in 21,730 genes with the eight young subjects, were excluded owing to known differences associated with African ancestry. We filtered for significance at a BH-adjusted p value <0.05. We set hierarchical sorting by dependencies between terms on the strongest setting, “best per parent group,” and only annotated genes were used as the statistical domain size parameter to determine hypergeometrical probability of enrichment. We considered ontology terms to be statistically significant at a BH-adjusted p value <0.01. Correlation of log fold change in DNA methylation of DM CpG sites by genomic feature was performed on significantly DM probes identified by limma and plotted using ggplot2 v3.1.0. Genomic features and transcriptionally regulated regions were subdivided for terms “island,” “open sea,” “shell,” “shore,” and “3’UTR” of TSS 1500 “TSS 200,” and “Body.” Respectively, where “island” refers to 1 kB regions of high CpG density, “shore” refers to 2 kB regions flanking “islands,” “shell” refers to 2 kB regions flanking “shores,” and “open sea” refers to IGRs. Location of CpGs in relation to gene annotation was defined as within the 5’-UTRs, within 1500 bp of the TSS (TSS 1500), within 200 bp of the TSS (TSS 200), within the gene body (Body), or within the 3’-UTR. A chi-squared test was used for all categorical comparisons. Correlation with methylation and gene expression was performed using the ggsomct function in ggpdbur v0.2, with the correlation method set to Pearson, C. set at 95%, and regression calculations included for the following subsets: HypoM, HyprM, and comprehensive CpG sites. Gene name conversions and annotated gene attributes were determined by BioMart v2.38.0. The Circos plot was created using the Circos tool v0.69-6.

Mediation analysis was performed using the R package “mediation” v4.4.7. We first performed linear regression to test whether African ancestry affects methylation for each of the five genes we discovered to be associated with WAA. For each non-zero effect, we subsequently performed causal mediation analysis to determine the proportion of methylation that affects gene expression compared to that of WAA. The Average Causal Mediation Effects score represents the mediation effect of methylation. We also report the 95% C.I., its upper and lower bounds, and the statistical significance of the mediation effect (p value).
Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY
The datasets supporting the conclusion of this article are included with the article and in additional files. All raw gene expression data has been submitted to the Gene Expression Omnibus (GEO accession number GSE124076).

Received: 11 March 2019; Accepted: 27 September 2019; Published online: 25 November 2019

REFERENCES
1. Baharian, S. et al. The Great Migration and African-American genomic diversity. PLoS Genet. 12, e1006059 (2016).
2. Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genomics Hum. Genet. 9, 403–433 (2008).
3. Nedelec, Y. et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell 167, 657–669 e621 (2016).
4. De, T., Park, C. S. & Perera, M. A. Cardiovascular pharmacogenomics: does it matter if you’re Black or White? Annu. Rev. Pharmacol. Toxicol. https://doi.org/10.1146/annurev-pharmtox-010818-012154 (2018).
5. Dau, J. Contribution of four comorbid conditions to racial/ethnic disparities in mortality risk. Am. J. Prev. Med. 52, 595–5102 (2017).
6. Cresci, S. et al. Cytochrome p450 gene variants, race, and mortality among clodipogol-treated patients after acute myocardial infarction. Circ. Cardiovasc. Genet. 7, 277–286 (2014).
7. Sivasankaradajah, G. A. et al. Vegfa protects the glomerular microvasculature in diabetes. Diabetes 61, 2958–2966 (2012).
8. Sarkissyan, M. et al. Vitamin D receptor FokI gene polymorphisms may be associated with colorectal cancer among African American and Hispanic participants. Cancer 120, 1387–1393 (2014).
9. Mishra, D. K. et al. Vitamin D receptor gene polymorphisms and prognosis of breast cancer among African-American and Hispanic women. PLoS ONE 8, e57967 (2013).
10. Boulter, A. C. et al. Interaction of Alu polymorphisms and novel measures of demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genomics Hum. Genet. 9, 403–433 (2008).
11. Scott, D. M. & Taylor, R. E. Health-related effects of genetic variations of alcohol-metabolizing enzymes in African Americans. Alcohol. Res. Health 30, 18–21 (2007).
12. De, T. et al. Association of genetic variants with warfarin-associated bleeding among patients of African descent. JAMA 320, 1670–1677 (2018).
13. Kruzel-Davalia, E., Wasser, W. G. & Skorecki, K. APO1 nephropathy: a population genetics and evolutionary medicine detective story. Semin. Nephrol. 37, 490–507 (2017).
14. Liu, C. T. et al. Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. PLoS Genet. 7, e1002264 (2011).
15. Li, Y. C. Vitamin D receptor signaling in renal and cardiovascular protection. Semin. Nephrol. 33, 443–447 (2013).
16. Ghazarian, A. A. et al. Incidence of testicular germ cell tumors among US men by race and ethnicity. Cancer 121, 4181–4189 (2015).
17. Wang, F. S., Fan, J. G., Zhang, Z., Gao, B. & Wang, H. Y. The global burden of liver cancer in 2017. Nat. Rev. Gastroenterol. Hepatol. 14, 167–176 (2017).
18. Wang, Y. et al. Associations between RNA splicing regulatory variants of stemness-related genes and racial disparities in susceptibility to prostate cancer. Int. J. Cancer 141, 731–743 (2017).
19. Vasiiliou, V. & Pappa, A. Polymorphisms of human aldehyde dehydrogenases. Consequences for drug metabolism and disease. Pharmacology 61, 192–198 (2000).
20. Caliskan, M. et al. Genetic and epigenetic interactions of Alu polymorphisms and novel measures of demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genomics Hum. Genet. 9, 403–433 (2008).
21. Schadt, E. E. et al. Mapping the genetic architecture of gene expression in primary human liver tissue. PLoS Genet. 7, e1002078 (2011).
22. Ghazarian, A. A. et al. Incidence of testicular germ cell tumors among US men by race and ethnicity. Cancer 121, 4181–4189 (2015).
23. Innocenti, F. et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 7, e1002078 (2011).
24. McArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 45, D896–D901 (2017).
56. Hetherington, S. L. et al. Dimorphism in the P2Y1 ADP receptor gene is associated with increased platelet activation response to ADP. Arterioscler Thromb. Vasc. Biol. 25, 252–257 (2005).
57. Maloney, J. P. et al. The ENTPD1 promoter polymorphism -860 A > G (rs3814159) is associated with increased gene transcription, protein expression, CD39/NTPDase1 enzymatic activity, and thromboembolism risk. FASEB J. 31, 2771–2784 (2017).
58. Zhang, H. et al. Impact of CYP2C19 gene polymorphism on warfarin maintenance doses in patients with non-valvular atrial fibrillation. Gene 591, 80–84 (2016).
59. Simon, T. et al. Genetic determinants of response to clopidogrel and cardiovascular events. N. Engl. J. Med. 360, 363–375 (2009).
60. Shuldiner, A. R. et al. Association of cytochrome P450 2C19 genotype with the risk of coronary artery disease in African Americans. Circulation 104, 2419–2424 (2001).
61. Lane, S. et al. The population pharmacokinetics of R- and S-warfarin: effect of genetic and clinical factors. Br. J. Clin. Pharmac. 73, 66–72 (2016).
62. Pendyala, L. K. et al. Racial disparity with on-treatment platelet reactivity in African-American adults with a diagnosis of myocardial infarction or venous thromboembolism. Thromb. Res. 172, 113–119 (2018).
63. Nguyen, T. A., Diodati, J. G. & Pharand, C. Resistance to clopidogrel: a review of the evidence. J. Am. Coll. Cardiol. 45, 1157–1164 (2005).
64. Kathiresan, S. et al. Comprehensive survey of common genetic variation at the brinogen and factor VII genes associated with low vitamin D levels in African descent populations: a meta-analysis of genome-wide association studies. J. Intern. Med. 271, 1048–1101 (2012).
65. Mondul, A. M., Weinstein, S. J., Layne, T. M. & Albanes, D. Vitamin D and cancer risk and mortality: state of the science, gaps, and challenges. Epidemiol. Rev. 39, 28–48 (2017).
66. O'Connor, M. N., Thoresen, C. K., Ramsey, N. L., Ricks, M. & Sumner, A. E. The uncertain significance of low vitamin D levels in African descent populations: a review of the bone and cardiometabolic literature. Prog. Cardiovasc. Dis. 56, 261–269 (2013).
67. Baysal, M. A. et al. Vitamin D receptor and megalin gene polymorphisms are associated with longitudinal cognitive change among African-American urban adults. J. Nutr. 147, 1048–1062 (2017).
68. Beydoun, M. A. et al. Vitamin D metabolism-related gene haplotypes and their association with metabolic disturbances among African-American urban adults. Sci. Rep. 8, 8035 (2018).
69. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature 542, 352–356 (2017).
70. Rodriguez-Antonio, C., Jover, R., Gomez-Lechon, M. J. & Castell, J. V. Quantitative RT-PCR measurement of human coagulation factor F12 using real-time detection systems. Arch. Biochem. Biophys. 376, 109–116 (2000).
71. Andersen, M. R., Farin, F. M. & Omiecinski, C. J. Quantification of multiple human coagulation factor F12 mRNA molecules using competitive reverse transcriptase-PCR. DNA Cell Biol. 17, 231–238 (1998).
72. Perera, M. A. et al. Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. Lancet 382, 790–796 (2013).
73. Lee, S. M., Schelcher, C., Demmel, M., Hauner, M. & Thasler, W. E. Isolation of human hepatocytes by a two-step collagenase perfusion procedure. J. Vis. Exp. https://doi.org/10.3791/50615 (2013).
74. Keemink, J., Oorts, M. & Annaert, P. Primary hepatocytes in sandwich culture. Methods Mol. Biol. 1250, 175–188 (2015).
75. Anderson, C. A. et al. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573 (2010).
76. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197, 573–589 (2014).
77. Gamaon, E. R. et al. A genome-wide integrative study of microRNAs in human liver. BMC Genomics 14, 395 (2013).
78. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genetics data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78, 629–644 (2006).
79. Ritchie, M. E. et al. Imma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).
80. Stegle, O., Faris, L., Pipari, M., Wenn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. 7, 500–507 (2012).
81. Bibikova, M. et al. High density DNA methylation array with single CpG site resolution. Genomics 98, 288–295 (2011).
82. Lizio, M. et al. Gateways to the FANTOMS promoter level mammalian expression atlas. Genome Biol. 16, 22 (2015).
83. Siggens, L. & Ekwall, K. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. J. Intern. Med. 276, 201–214 (2014).
84. Teschendorff, A. E. et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 K DNA methylation data. Bioinformatics 29, 189–196 (2013).
85. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127 (2007).
86. Leek, J. T., Johnson, W. E., Park, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882–883 (2012).
87. Teschendorff, A. E. et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS ONE 4, e8274 (2009).
88. We would like to acknowledge Dr. Heather Wheeler for her invaluable scientific input.
89. Cancer Genome Atlas Research, N. et al. Comprehensive molecular characterization of papillary renal-cell carcinoma. N. Engl. J. Med. 374, 135–145 (2016).
90. Nordlund, J. et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol. 14, 1 (2013).
91. Sliker, R. C. et al. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin 11, 25 (2018).
92. Reimand, J. et al. gProfiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 44, W83–W89 (2016).
93. Sandqvist, J. et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6, 692–702 (2011).
94. Edgar, R., Tan, P. P., Portales-Casamar, E. & Pavlidis, P. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. Epigenetics Chromatin 7, 28 (2014).

ACKNOWLEDGEMENTS
We would like to acknowledge Dr. Heather Wheeler for her invaluable scientific input. E.R.G. wishes to thank the President and Fellows of Clare Hall, University of Cambridge for providing a stimulating intellectual home and for the generous support during the Academic Terms of Cambridge University. E.R.G. is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R35HG010718. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was supported by NIH National Institute on Minority Health and Health Disparities (NMHD) Research Project 1R01MD009217-01 (R01).

Published in partnership with CEGMR, King Abdulaziz University
npj Genomic Medicine (2019) 29
AUTHOR CONTRIBUTIONS
C.S.P. and M.A.P.: conception and design, analysis and interpretation of data, drafting or revising the article; Y.X.: design, analysis and interpretation of data; T.D.: design and acquisition of data; Y.Z.: analysis; E.R.G.: analysis and interpretation of data; E.S. and C.A.: acquisition of data.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information is available for this paper at https://doi.org/10.1038/s41525-019-0102-y.

Correspondence and requests for materials should be addressed to M.A.P.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019