Uncovering the role admixture in health disparities: Assocication of hepatocyte gene expression and DNA methylation to African Ancestry in African-Americans.

Park CS[1], De T[1], Xu Y[1,2], Zhong Y[1], Smithberger E[1,3], Alarcon C[1], Perera MA[1*]

1. Department of Pharmacology, Center for Pharmacogenomics, Feinberg School of Medicine, Northwestern University, Chicago IL
2. Center for Translational Data Science, University of Chicago, Chicago IL
3. Department of Pathology and laboratory medicine, Univeristy of North Caroline School of Medcine, Chapel Hill, NC.

* corresponding author: minoli.perera@northwestern.edu

ABSTRACT

In African Americans (AA), the proportion of West African ancestry (WAA) may explain the genetic drivers of health disparities in disease. Analysis of RNA sequencing data from sixty AA-derived primary hepatocytes identified 32 gene expression profiles associated with WAA (FDR <0.05) with enrichment in angiogenesis and inflammatory pathways (FDR <0.1). Association of DNA methylation to WAA identified 1037 differentially methylated regions (FDR <0.05), with hypomethylated genes enriched for drug response pathways. Within the PharmGKB pharmacogene, *VDR*, *PTGIS*, *ALDH1A1*, *CYP2C19* and *P2RY1* were associated with WAA (p <0.05) with replication of *CYP2C19* and *VDR* in the GTEx liver cohort. For every 1% increment in WAA, *P2RY1* gene expression increased by 1.6% and *CYP2C19* gene expression decreased by 1.4%, suggesting effects on clopidogrel response and platelet aggregation. We conclude that WAA contributes to variablity in hepatic gene expression and DNA methylation with identified genes indicative of health disparities prevalent in AAs.

Keywords: African-American, pharmacogenomics, liver, hepatocytes, RNA-Seq, methylation, health disparity

Introduction

African Americans (AAs) are an admixed population, having varying proportions of African and European ancestry between individuals and within sub-populations (Baharian et al., 2016). As a consequence of their African Ancestry, AAs have more genetic variation and less linkage disequilibrium (LD) than other global populations, with the proportion of African Ancestry varying greatly between self-identified AAs (Campbell & Tishkoff, 2008). It is this genetically driven variability that may aid in explaining differences in hepatic gene expression and DNA methylation patterns, which cannot be elucidated in homogenous populations such as those of European-only ancestry. For example, African ancestry 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 populations (Nedelec et al., 2016).

However, it must also be taken into account that AAs also suffer disproportionately from many chronic diseases as well as adverse drug reactions compared to other populations (Daw, 2017; De, Park, & Perera, 2018). AAs have a higher risk of cardiovascular events and negative outcomes from adverse antiplatelet drug response (Cresci et al., 2014). They have higher incidences of death and disability from cardiovascular diseases (CVDs), thrombosis, renal dysfunction and pathologies, diabetes, cancers, alcoholism and other metabolic disorders (Boulter et al., 2015; De, Alarcon, et al., 2018; Kruzel-Davila, Wasser, & Skorecki, 2017; Y. C. Li, 2013; C. T. Liu et al., 2011; Mishra et al., 2013; Sarkissyan et al., 2014; Scott & Taylor, 2007; Sivaskandarajah et al., 2012).

Due to the key role of the liver in biosynthesis, drug metabolism and complex human diseases, genetic and epigenetic differences in the liver may uncover the underlying causal genes responsible for health disparities in AAs (Ongen et al., 2017; F. S. Wang, Fan, Zhang, Gao, & Wang, 2014). With the first comprehensive mapping of liver eQTLs, several candidate susceptibility genes were identified to associate with type I diabetes, coronary artery disease and plasma cholesterol levels in a Caucasian cohort (Schadt et al., 2008). More recently, even finer-detailed 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 (Caliskan et al., 2018). In addition, our group has previously shown that studies specifically exploring the unique genetic composition of AAs can reveal population-specific risk factors that may explain health disparities in drug response – we recently identified several African ancestry-specific genetic risk factors that exposes AAs to a higher risk of bleeding from warfarin therapy (De, Alarcon, et al., 2018).

Rather than pursuing the well-established approach of identifying disease susceptibilities through genome-wide association studies (GWAS), here we use variability in genetic ancestry to uncover potential drivers of disease or drug response that may explain AA health disparities. We characterized gene expression and differentially methylated regions with percent West African ancestry (WAA). Through this approach, we show that the percentage of WAA associates with hepatic gene expression profiles that are known to be dysregulated in diseases that affect AAs disproportionately and genes known to be responsible for adverse drug responses in AA populations.

Results

Cohorts

60 primary hepatocyte cultures, procured from self-identified AAs, passed all quality control steps and were used for RNA-sequencing (RNA-seq) analysis. Fifty-six hepatocyte cultures were assayed for DNA methylation, with 44 used in the final methylation analysis. We obtained genome-wide genotyping data of 446 subjects from Genotype-Tissue Expression Project (GTEx), release version 7 (Consortium, 2013). All 60 of our AA hepatocyte cohort were confirmed as having WAA ranging from 41.8% to 93.7%, and 12 subjects in the GTEx liver cohort met the WAA inclusion criteria ranging from 12.5% to 91.4% (**Supplementary Figure 1**). Table 1 shows the demographics of each cohort. In general, the GTEx liver cohort were older and had a lower percentage of females than the primary hepatocyte cohort.

Hepatic genes and pharmacogenes associated with African ancestry

Testing of RNA-seq gene expression levels in the AA hepatocyte cohort with percentage WAA identified 32 genes in which expression levels are significantly associated with WAA (**Table 3**; **Figure 1A**, FDR < 0.05). No gene in the genome-wide analysis was replicated within the GTEx liver cohort (**Table 3**; FDR < 0.05). Due to the importance of the liver in pharmacologic drug response, we also tested the association of gene expression levels with percentage African ancestry in a subset of genes belonging to the very important pharmacogenes (VIP) in PharmGKB, consisting of 64 genes known to be expressed in hepatocytes. These represent drug-metabolizing enzyme (DME) genes that are well-established for their role in drug response as well as transporters and drug targets (Whirl-Carrillo et al., 2012). Testing the association between African ancestry and gene expression in the VIP genes identified five genes: *VDR* (effect size = 0.35, p = 0.003), *PTGIS* (effect size = -0.57, p = 0.002), *ALDH1A1* (effect size = 1.03, p = 0.032), *CYP2C19* (effect size = -1.36, p = 0.032) and *P2RY1* (effect size = 1.61, p = 0.002) such that for every one percent change in African ancestry there is a corresponding one percent change in gene expression **(Figure 1B)**. We were able to replicate both *VDR* (effect size = 2.23, p = 0.019) and *CYP2C19* (effect size = -2.48, p = 0.004) expression levels in the GTEx liver cohort (**Table 2**).

DNA methylation patterns are associated with African ancestry in human liver

To identify differentially methylated (DM) regions (DMRs) associated with WAA, we conducted linear regression on each CpG site to find African ancestry-related differential methylation (Morris et al., 20). We identified 23,317 significant DM CpG sites, out of a total of approximately 850,000 probes on the Illumina EPIC BeadChip microarray, annotated to 21,878 unique genes (**Supplementary Figure 2A**; BH-adjusted p < 0.05). These DM CpGs correspond to 1037 DMRs annotated to 1034 unique genes (**Supplementary Figure 2B**; minimum FDR < 0.0001). Each DM CpG sites were categorized into hyper- (HyprM) and hypomethylated (HypoM) sites such that, 15,404 HyprM CpG sites constituted 435 HyprM DMRs, mapping to 432 unique genes, 7913 HypoM CpG sites constituted HypoM 602 DMRs, mapping to 602 unique genes, and seven annotated genes had both HyprM and HypoM DMRs.

In our analysis of DM loci associated with WAA, HypoM CpG sites are enriched in islands (75.9% vs 24.1% HypoM), while the shore, shelf and open sea are predominantly HyperM (65.5% vs 34.5%, 81.7% vs 18.3% and 78.9% vs 21.1% HyprM, respectively) **(Supplementary Figure 2C)**. Within transcriptionally-regulated promoter regions, HypoM CpG sites are approximately equal to HyprM CpG sites (53.9% versus 46.1%). However, CpGs in the subsite of the promoter – the 200 kb upstream region of the transcription start site (TSS) – were significantly more HyperM (75.9% versus 24.1%) **(Supplementary Figure 2D)**. DNA methylation around TSS is an established predictor of gene expression, with increased methylation leading to decreased expression (Doi et al., 2009; Jiao, Widschwendter, & Teschendorff, 2014). Similarly in the gene body, a higher percentage of HyprM CpG sites were found (73.0% versus 27.0%) **(Supplementary Figure 2D)**. Intergenic, 5'-UTR and 3'-UTR regions were predominantly HyperM (67.5% versus 32.5%, 64.0% versus 36.0%, 79.6% versus 20.4% 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 CpGs located in islands. Within the 15,404 HyperM CpGs associated with WAA, there was a greater proportion in the shelf and open sea regions **(Supplementary Figure 2E),** 5'-UTR, promoter regions and gene body **(Supplementary Figure 2F)**.

DNA methylation patterns with respect to gene expression profiles in African ancestry

To determine the relationship of DMRs to gene expression signatures associated with WAA, we examined the association of 1034 unique genes corresponding to the 1037 DMRs with their respective gene expression profiles. Although there was no correlation of African ancestry-associated HyprM genes with gene expression (**Figure 1C**, Pearson's r = 0.036, p = 0.55), HypoM genes associated with WAA were negatively correlated with gene expression (**Figure 1C**, Pearson's r = -0.14, p = 0.009). In general, all DM genes associated with WAA also showed negative correlation between gene expression and methylation (Pearson's r = -0.1, p = 0.013). From the 32 genes with gene expression significantly associated with WAA, we identified an overlap of only two DM genes, *HIC1* (minFDR = 5.35×10^{-8}, mean Beta = 0.0011, consisting of 35

CpGs) and *RNF135* (minFDR = 3.17 x 10^{-84}, mean Beta = -0.2504, consisting of 18 CpGs) which had concordant directions of effect (e.g. increased methylation leading to decrease gene expression). A multi-layered overview of the 23,317 HypoM and HyprM CpG sites that constitute the 1037 DMRs, and the corresponding gene expression found to be positively or negatively expressed with increasing African ancestry with a fold change greater than 1.5, is represented in a Circos plot (**Figure 1D**).

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 African ancestry, we performed gene ontology analysis of the 432 unique genes comprising the 435 HyprM DMRs and 602 unique genes comprising 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 = 0.0076) within the category of biological processes **(Figure 2A)**. HypoM genes were enriched for "system development" (BH-adjusted p = 4.0 x 10^{-6}), "response to drug" (BH-adjusted p = 1.7 x 10^{-4}), and "response to hypoxia" (BH-adjusted p = 0.0068) within the category of biological processes (**Figure 2B**). In addition, HypoM genes are enriched for "sequence-specific DNA binding" (BH-adjusted p = 4.4 x 10^{-5}) and "RNA polymerase II transcription factor activity" (BH-adjusted p = 4.4 x 10^{-5}) within the category of molecular function **(Figure 2B)**.

With respect to gene ontology analysis of the top 131 gene expression profiles associated with African ancestry (FDR < 0.10), "angiogenesis" (BH-adjusted p = 4.5 x 10^{-5}), "leukocyte activation involved in immune response" (BH-adjusted p = 4.5 x 10^{-4}), "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) are inflammatory biological processes that are enriched in hepatocytes **(Figure 2C)**. Molecular functions, such as "oxidoreductase activity" (BH-adjusted p = 9.6 x 10^{-4}), "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 x 10^{-4}) are also enriched **(Figure 2C)**.

Discussion

Several studies have shown that the first several principal components of methylation data can capture population structure in cohorts composed of European and African individuals (Barfield et al., 2014). A recent paper by Park DS et al. found that genetic ancestry can be used as a proxy, not only for unknown covariates contributing to epistatic and gene-environment interactions from gene expression data, but also from DNA methylation data (D. S. Park et al., 2018). Indeed, approximately 75% of differential methylation between global populations were attributable to genetic ancestry in addition to ethnically distinct DMRs, which are known to correlate with environmental factors, such as maternal smoking during pregnancy (Galanter et al., 2017).

In our study, we have taken these findings a step further by investigating population-specific gene expression and DNA methylation in an admixed population – African-Americans. We found HypoM hepatic genes are enriched for "system development" and "response to drug" ontology terms in AAs. Interestingly, we found that the DM drug response genes, *ALDH1A1* and *CYP2C19*, are also represented as significantly associated with African Ancestry in the PharmGKB subset analysis. Thus, 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. Methylation within the TSS of the promoter is well known to repress gene expression while methylation within the gene body results in more variable expression (Jiao et al., 2014; Yang et al., 2014).

Gene expression associated with percentage of African ancestry identified hepatic genes that are enriched for "angiogenesis" as the primary ontology term, followed by inflammatory response categories including "leukocyte activation in immune response", "acute-phase response", "positive regulation of vascular endothelial proliferation", "zymogen activation" and

"cell proliferation". Angiogenesis and inflammatory response terms are highly biologically relevant to health disparities that affect AAs, as AAs are more susceptible to CVD and other chronic inflammatory disease. In particular, *APOL1*, *PTGIS* and *PLAT* expression, which we show to be associated with percentage of WAA, have been shown to increase CVD risk and renal disease in AAs (Kruzel-Davila et al., 2017; Nagareddy & Smyth, 2013; Notarangelo, Bontardelli, & Merlini, 2013). Other interesting associations include *ALDH1A1*, which is involved in alcohol and aldehyde metabolism disorders and cancer (Scott & Taylor, 2007; Vasiliou & Pappa, 2000; Y. Wang et al., 2017), *IL-33*, which is involved in beneficial immune response (Chalubinski, Luczak, Wojdan, Gorzelak-Pabis, & Broncel, 2016; Liew, Girard, & Turnquist, 2016; Ndaw et al., 2017), and *VEGFA*, which has been linked to renal disease and microvascular complications of diabetes in AAs (Boger & Heid, 2011; C. T. Liu et al., 2011; Majumder & Advani, 2017; Sivaskandarajah et al., 2012).

Of the 32 genome-wide gene expression profiles associated with WAAs, we also identified an overlap with two DM genes, *HIC1* and *RNF135*. *HIC1* is a potential tumor suppressor that has been linked to poorer outcomes in laryngeal cancer in AAs (Rathi et al., 2003; Stephen et al., 2010). *RNF135* is a ring finger protein that is regulated by several population-specific variants (L. Park, 2016) . RNF135, itself, then regulates other genes at distant loci and has been implicated in glioblastomas and autism (Y. Liu et al., 2016; L. Park, 2016; Tastet et al., 2015). Relevant for African ancestry populations, *RNF135* has been found to be under selective pressure specifically in African populations (Barreiro, Laval, Quach, Patin, & Quintana-Murci, 2008; Grossman et al., 2013).

In PharmGKB VIP genes, we identified five genes and replicated two, *CYP2C19* and *VDR*, in the GTEx liver cohort (Consortium, 2013; Whirl-Carrillo et al., 2012). Increased expression of *VDR* is of particular importance because of its role in health and disease. Vitamin D and its active metabolite 1,25$(OH)_2$D are exogenous hormones created by sun exposure or through diet, with established deficiencies in both Vitamin D and its bioactive metabolites in AAs (Mondul, Weinstein, Layne, & Albanes, 2017; O'Connor, Thoreson, Ramsey, Ricks, & Sumner, 2013).

People of African ancestry have lower plasma 1,25$(OH)_2$D levels and our findings suggest that this may lead to an upregulation of *VDR* with increased African ancestry to compensate for lower levels of ligand. Additionally, *VDR* single nucleotide polymorphisms (SNPs) have been implicated in CVDs and various cancer susceptibilities in AAs (Beydoun et al., 2018; Beydoun et al., 2017; Y. C. Li, 2013; Mishra et al., 2013; Sarkissyan et al., 2014). Our findings suggest that *VDR* upregulation may ameliorate some of the effect of lower plasma 1,25$(OH)_2$D levels, while SNPs within *VDR* have profound effects because they exacerbate an already deficient vitamin D environment in AAs.

Within our PharmGKB VIP gene set, we found that for every 1% increment in African ancestry, there was a corresponding 1.36% decrement in *CYP2C19* expression and 1.61% increment in *P2RY1* expression. *CYP2C19* is involved in the metabolism of many commonly prescribed drugs including clopidogrel, an antiplatelet therapy widely used for thromboprophylaxis of CVDs and linked to substantial inter-patient differences in drug response. Clopidogrel is known to inhibit the *P2RY* family of receptor proteins on the surface of platelets (Dean, 2012; Jarrar et al., 2016; Nagareddy & Smyth, 2013; Notarangelo et al., 2013). By inhibiting *P2RY12* function, platelet clustering and clot formation are suppressed by clopidogrel, which mitigates clots contributing to heart attack, stroke and deep vein thrombosis (Janicki et al., 2017; J. L. Li et al., 2018; Timur et al., 2012). It has been shown that decreased activity of *CYP2C19* results in reduced efficacy of clopidogrel which is a pro-drug that requires conversion to its active form via this enzyme (De, Alarcon, et al., 2018; Lane et al., 2012; Shuldiner et al., 2009; Simon et al., 2009; Zhang et al., 2016). The underlying mechanism is multifactorial and genetic polymorphisms are one of the main causes of variable drug response within an individual and across populations (Cresci et al., 2014; Nguyen, Diodati, & Pharand, 2005; Pendyala et al., 2013). A study conducted across 24 U.S. hospitals showed one-year mortality rate of 7.2% in clopidogrel-treated AAs, compared to 3.6% for Caucasians on clopidogrel (Cresci et al., 2014). 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 African ancestry

while *P2RY12* expression is inversely increased, suggests that clopidogrel resistance of AAs and the susceptibility of AAs to CVDs, may be due to ancestry-associated genetic regulation.

In the genome-wide analysis of gene expression, we found that *PLAT* expression decreases 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 (Huang et al., 2014; Kathiresan et al., 2005). In AAs, polymorphisms in *PLAT* have been linked to CVD and higher levels of t-PA antigen seen in both myocardial infarction and venous thromboembolism (Hooper et al., 2000). In addition, increased plasma fibrinogen levels, which is involved in the fibrinolytic pathway and regulated by t-PA, was linked to increased venous thrombosis risk in AAs (Austin et al., 2000; Mannucci et al., 1997).

Several limitations exist in this study. First, we only had 60 primary hepatocyte cultures included in this analysis, which limits our power to detect small changes in gene expression associated with ancestry. Second, we were only able to replicate two genes associated to ancestry in the GTEx liver cohort. Given that only 12 AA livers were available for the replication analysis, this dataset lacked the power to detect AA-specific differences in gene expression with a high degree of confidence. Third, while most of the genes found to have association in the PharmGKB VIP gene set were not statistically significant in the larger genome-wide analysis, both *ALDH1A1* and *PTGIS* were found in the genome-wide analysis that identified 32 genes associated with African ancestry, and *CYP2C19* and *VDR* were replicated in the 12 AAs of the GTEx liver cohort. Lastly, 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. However, our study design has the advantage of only assaying the gene expression of a single cell type as opposed to the multiple cell types found in 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 (Andersen, Farin, & Omiecinski, 1998; Rodriguez-Antona, Jover, Gomez-Lechon, & Castell, 2000). Also, the use of primary hepatocyte cultures reduce the effect of the

environmental confounders inherent in liver tissue (i.e. there is less effect of previous drug/disease exposure) through the controlled tissue culture processes following hepatocyte isolation. The artifact of previous disease/drug exposures are present in all transcriptome studies conducted in post-mortem human liver tissue.

Currently, genome-wide genotypes, sequences and multi-omics datasets of AAs are lacking in both the scientific literature and in public databases and repositories. With the understanding that genomics studies should be inclusive of all populations to comprehensively unravel disease etiology, the dearth of genetic data in minorities 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 decreased warfarin dose requirement, but only in AAs (Perera et al., 2013). Findings, such as these, were only possible with focused studies of the AA 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 implications in estimating inter-individual genetic factors and drug response outcomes in admixed populations for the implementation of precision medicine. As evidenced by limited genome-wide data in AAs within public databases and biobanks, such as in GTEx, our study further stresses the need for genome-wide studies in minority and under-represented populations.

Materials and methods

Cohorts

A total of 68 African American (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 (Lee, Schelcher, Demmel, Hauner, & Thasler, 2013). Liver specimens were obtained

through collaborations with Gift of Hope, which supplies non-transplantable organs to researchers. In addition, we used GWAS data for 446 subjects from Genotype-Tissue Expression Project (GTEx) release version 7, of which 12 samples were used in our replication analyses.

Primary Hepatocyte Isolation

Briefly, 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 100-400mL/min, with recirculation. Following perfusion, liver section was placed in a crystallizing dish (Omnilab) containing 100-200 ml of Solution 4 (BSA, Sigma-Aldrich). The Glisson's capsule was carefully removed and the tissue was gently shaken to release hepatocytes. The cell suspension was then filtered by a 70 µm nylon mesh (Fisher Scientific), and centrifuged at 72 x g for 5 min at 4 °C. The pellets contained hepatocytes that was 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) (Keemink, Oorts, & Annaert, 2015). If viability was low, Percoll gradient (Sigma-Aldrich) centrifugation of cell suspensions was carried out to improve yield of viable cells. Cell were plated at a density of 0.6 x 106 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 hours for three days. RNA was extracted after three days using the RNeasy Plus mini-kit (Qiagen) per manufacturer's instructions.

Genotyping and quality control

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

Quality control (QC) steps were applied as previously described but with imputation info > 0.4 (Hernandez et al., 2016). 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 (IBD) method with a cutoff score of 0.125 indicating 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 (HWE) tests ($p < .00001$) were also excluded.

African ancestry measurement

The genotypes of 68 primary hepatocytes and 446 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) (Anderson et al., 2010). 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 (MCMC) repetition was set to 20,000 and 100,000, respectively (Raj, Stephens, & Pritchard, 2014). The number of parental populations (K) was set to 3, purporting three main continental groups (AFR, EUR and EAS). WAA percentages of the primary hepatocytes and GTEx subjects were calculated

as the probability of being grouped as Yoruba African, Caucasian and Asian, respectively (Raj et al., 2014).

GTEx liver cohort

Of the 144 GTEx liver IDs were extracted from "GTEX_Sample_Attributes" file, five were replicates and were removed from the analysis. Among the remaining 139 unique liver IDs, 118 had available genotype information for ancestry determination. Individuals with African ancestry percentages > 10% were considered for the GTEx AA Liver cohort. Age and sex information of these subjects were extracted from the subject phenotype file on the GTEx Portal site (GTEx_v7_Annotations_SubjectPhenotypesDS.txt). After WAA estimation, normalized gene expression reads for 12 subjects meeting the ancestry inclusion criteria were extracted from GTEx expression file (GTEx_Liver.v7.normalized_expression.bed).

RNA-sequencing and quality control

Total RNA was extracted from each primary hepatocyte culture after three days in culture using the Qiagen RNeasy Plus mini-kit per manufacturer's instructions. Samples with RNA integrity number (RIN) scores above 8 were used in subsequent sequencing. RNA was prepared for sequencing using the TruSeq RNA Sample Prep Kit, Set A (Illumina catalog # FC-122-1001) according to manufacturer's instructions. The cDNA libraries were prepared and sequenced using both Illumina HiSeq 2500 and HiSeq 4000 machines by the University of Chicago's Functional Genomics Core to produce single-end 50 bp reads with approximately 50 million reads per sample. Batch effects were corrected for in quality control 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 GRCh37 and comprehensive gene annotation (GENCODE version 25) was performed using 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 regions 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 minaqual = 20 (Anders, Pyl, & Huber, 2015). HTSeq raw counts were supplied for gene expression analysis using Bioconductor package DESeq2 (version1.20.0) (Love, Huber, & Anders, 2014). Counts were normalized using regularized log transformation and principal component analysis (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 (Wagner, Kin, & Lynch, 2012). 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 done in the GTEx consortium (gtexportal.org, Analysis Methods, V7, updated 09/05/2017). We normalized expression values across samples by TMM (trimmed mean of M-values) implemented in edgeR (Robinson & Oshlack, 2010) and then normalized for each gene by inverse normal transformation.

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 and linear regression was run on inverse normal transformed expression data using five PEER factors, based on GTEx's determination of number of factors as a function of sample size (Consortium et al., 2017; Stegle, Parts, Piipari, Winn, & Durbin, 2012). African ancestry-associated genes were identified from a genome-wide list of 18,854 genes for

our hepatocyte-derived AA cohort and replicated in 21,730 genes for the GTEx-derived AA cohort at FDR cutoff of 0.05. FDR calculations from linear regressions on gene expression were conducted with the "p.adjust" function in R and set to method "fdr". In addition, we also conducted a subset analysis with 64 PharmGKB "very important pharmacogenes". These genes have extensive literature support for association to drug responses. We analyzed this subset for association to WAA using the same linear regression method at a *p*-value cutoff of 0.05 due to the smaller number of genes being tested, 64, and our small sample size of 60.

Methylation sample preparation and data analysis

DNA was isolated from hepatocytes or liver tissue. Liver tissue was homogenized in a bead mill (Fisher Scientific) using 2.8 mm ceramic beads. Then, DNA from hepatocytes or liver was extracted with the Gentra Puregene Blood kit (Qiagen) as per manufacturer instructions. One microgram of DNA was bisulfite converted at the University of Chicago Functional Genomics Core using standard protocols and methylation and analyzed on an Illumina HiScan. Of the 60 available hepatocyte cultures only 56 produced sufficient bisulfite converted DNA for analysis.

Illumina MethylationEPIC BeadChip microarray (San Diego, California, USA), consisting of approximately 850,000 probes and containing 90% of CpGs on the HumanMethylation450 chip and with more than 350,000 CpGs regions identified as potential enhancers by FANTOM5 (Lizio et al., 2015) and the ENCODE project (Siggens & Ekwall, 2014), was used for methylation profiling of DNA extracted from 53 AA hepatocytes which overlap with the samples used for gene expression analysis. Raw probe data was analyzed using the ChAMP R package for loading and base workflow (Morris et al., 2014), and included the following R packages: BMIQ for type-2 probe correction method (Teschendorff et al., 2013); ComBat for correction of multiple batch effects including Sentrix ID, gender, age, slide and array (Johnson, Li, & Rabinovic, 2007; Leek, Johnson, Parker, Jaffe, & Storey, 2012); svd for singular value decomposition analysis after correction (Teschendorff et al., 2009); limma to calculate differential methylation using linear regression on each CpG site with WAA as a numeric, continuous variable (Smyth, 2004); DMRcate for identification of DMRs, using default parameters, and the corresponding number of CpGs,

minimum FDR (minFDR); Stouffer scores, and maximum and mean Beta fold change values (Peters et al., 2015); minfi for loading and normalization (Aryee et al., 2014); missMethyl for gometh function for GSEA analysis (Geeleher et al., 2013); and FEM for detecting differentially methylated gene modules (Jiao et al., 2014).

Methylation data quality control in ChAMP's *champ.load()* and *champ.filter()* function removed the following probes: 9204 probes for any sample which did not have *p*-value < 0.01, and thus considered as a failed probe, 1043 probes with a bead count < 3 in at least 5% of samples, 2975 probes with no CG start sites, 78,753 probes containing SNPs (Cancer Genome Atlas Research et al., 2016), 49 probes that align to multiple locations as identified in by Nordlund et al. (Nordlund et al., 2013), and 17,235 probes located on X and Y chromosomes. Threshold for significantly differentially methylated probes was set at an adjusted *p*-value of 0.05 using Benjamini-Hochberg correction for multiple testing in the limma package. Significance for DMRs was set at adjusted FDR < 0.05 in the DMRcate package using default parameters. Analysis was performed with R statistical software (version 3.4.3 and version 3.5 for ChAMP (version 2.10.1)). Three outliers and nine samples from young subjects, less than five years old were excluded due to known differences in methylation profiles associated with age (Slieker, Relton, Gaunt, Slagboom, & Heijmans, 2018) leaving 44 samples in the analysis.

Gene ontology analysis was performed using g:Profiler (biit.cs.ut.ee/gprofiler) using g:GOSt to provide statistical enrichment analysis of our significantly HypoM and HyperM genes, and significantly expressed genes associated with African ancestry (Reimand et al., 2016). We filtered for significance at a Benjamini-Hochberg 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 significance 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 feature and transcriptionally-regulated regions were subset for terms "island", "opensea", "shelf", "shore" and “"5'UTR", "TSS1500", "Body", "3'UTR", respectively. Correlation with methylation and gene expression was performed using the ggscatter function in ggpubr v0.2, with the correlation method set to Pearson, confidence interval set at 95% and regression calculations included for the following subsets: HypoM, HyperM and cumulative CpG sites. Gene name conversions and annotated gene attributes were determined by BioMart v2.38.0. The Circos plot was created using Circos tool v0.69-6.

Funding

This work was supported by NIH National Institute on Minority Health and Health Disparities (NIMHD) Research Project 1R01MD009217-01 (R01).

Acknowledgements

We would like to acknowledge Dr. Heather Wheeler for her invaluable scientific input.

Competing Interests

The authors declare no competing interests.

Author Contributions

CSP and MAP Conception and Design, Analysis and interpretation of data, Drafting or revising the article; YX Design, Analysis and interpretation of data; YZ Analysis; TD Design and Acquisition of data; ES, CA Acquisition of data.

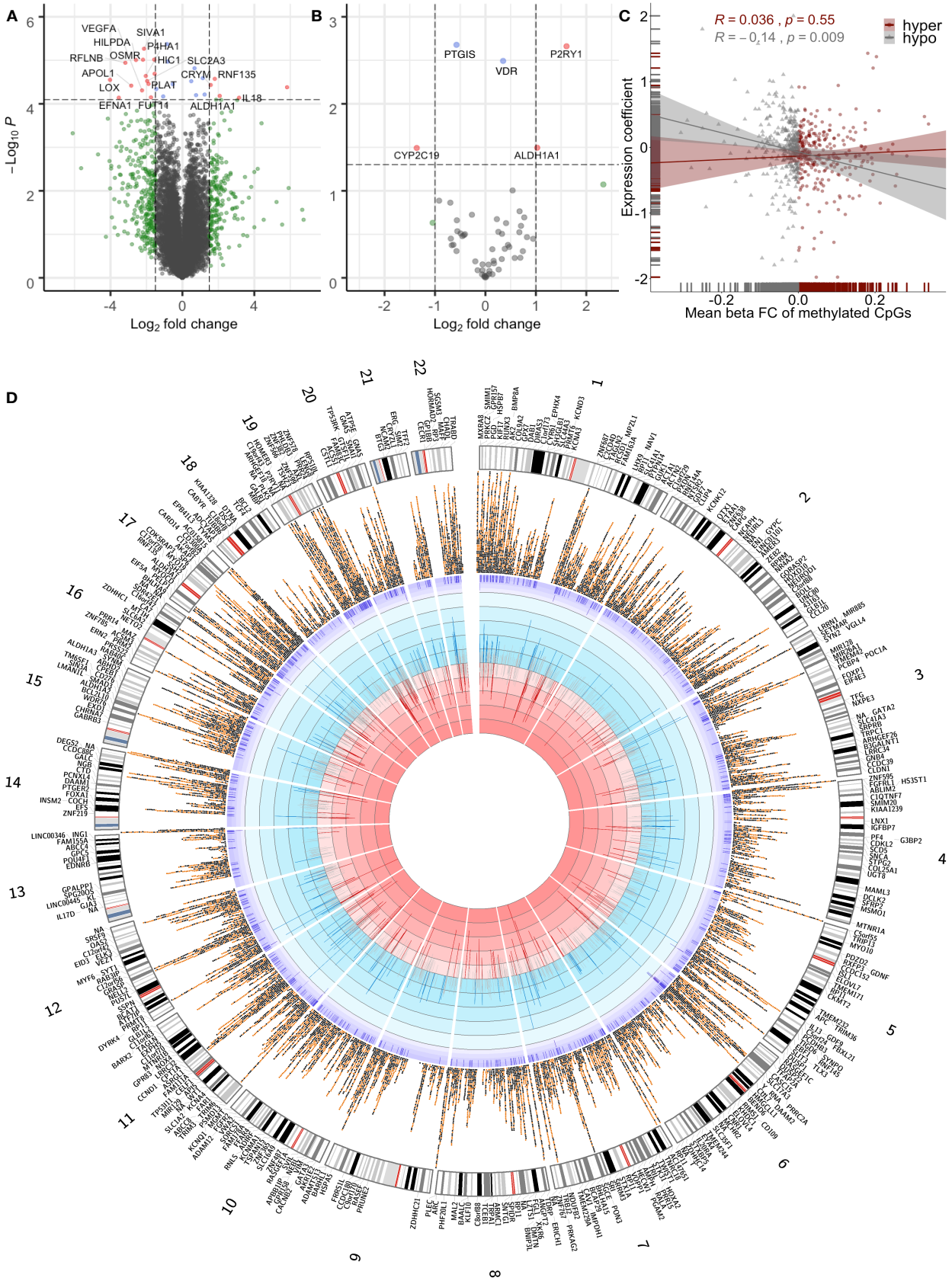

A
VEGFA
SIVA1
HILPDA
P4HA1
RFLNB
OSMR
HIC1
SLC2A3
APOL1
CRYM
RNF135
LOX
PLAT
IL18
EFNA1
FUT11
ALDH1A1
−Log10 P
Log2 fold change
B
PTGIS
P2RY1
VDR
CYP2C19
ALDH1A1
Log2 fold change
C
R = 0.036 , p = 0.55
R = −0.14 , p = 0.009
hyper
hypo
Expression coefficient
Mean beta FC of methylated CpGs
D

Figure 1. Genes and methylation patterns associated with West African ancestry in hepatocytes. Enhanced volcano plot of gene expression associated with increasing WAA plotted against -log10 *p*-values of **(A)** all 18,854 genes expressed in hepatocytes resulting in 32 genes significantly associated with WAA (red circles: FDR < 0.05 and logFC > 1.5; blue circles: FDR < 0.05 and logFC < 1.5; green circles: FDR > 0.05 and logFC > 1.5, grey circles: FDR > 0.05 and logFC < 1.5) and **(B)** within subset of 64 PharmGKB "very important genes" (red circles: $p < 0.05$ and logFC > 1.0; blue circles: *p*-value < 0.05 and logFC < 1.0; grey circles: *p*-value > 0.05). **(C)** Correlation of 1034 unique genes containing DMRs significantly associated with WAA (mean Beta fold change from DMRcate) with coefficient of gene expression at each gene (indicating the direction of association to ancestry). Each point represents a gene with grey triangles representing hypomethylated genes (Pearson's r = -0.014, $p = 0.009$); maroon red circles representing hypermethylated genes (Pearson's r = 0.036, $p = 0.55$). Grey and maroon hash marks on the x- and y-axis axis represent genes plotted with both expression and methylation values. Grey and maroon shading around each regression line represents the 95% confidence interval. **(D)** Circos plot summarizing significantly associated CpGs and DMRs, and correlated gene expression signatures associated with WAA. The innermost ring represents 2112 negatively correlated gene expression signatures associated with WAA (pink shading shows the expression coefficient ranging from -0.5 to -6.1; red bars represent fold change < -1.5; grey bars represent fold change > -1.5). The second ring represents 1895 positively correlated gene expression signatures associated with WAA (blue shading shows the expression coefficient ranging 0.5 to 6.7; blue bars represent fold change > 1.5; grey bars represent fold change < 1.5). The third ring represents 1037 DMRs significantly associated to WAA (shaded purple; purple tiles represents stacked overlapping DMRs). The fourth ring represents 23,317 significantly differentially methylated CpGs that constitute the 1037 DMRs in the third circle (black squares represent 7913 hypomethylated CpG sites; orange circles represent 15,404 hypermethylated CpG sites). The next ring represents the karyotype of the human genome (reference hg38) and the outer-most ring corresponds to the gene names of the 604 unique genes with the DMRs found to have either < -1.5 or > 1.5 fold change associated with WAA.

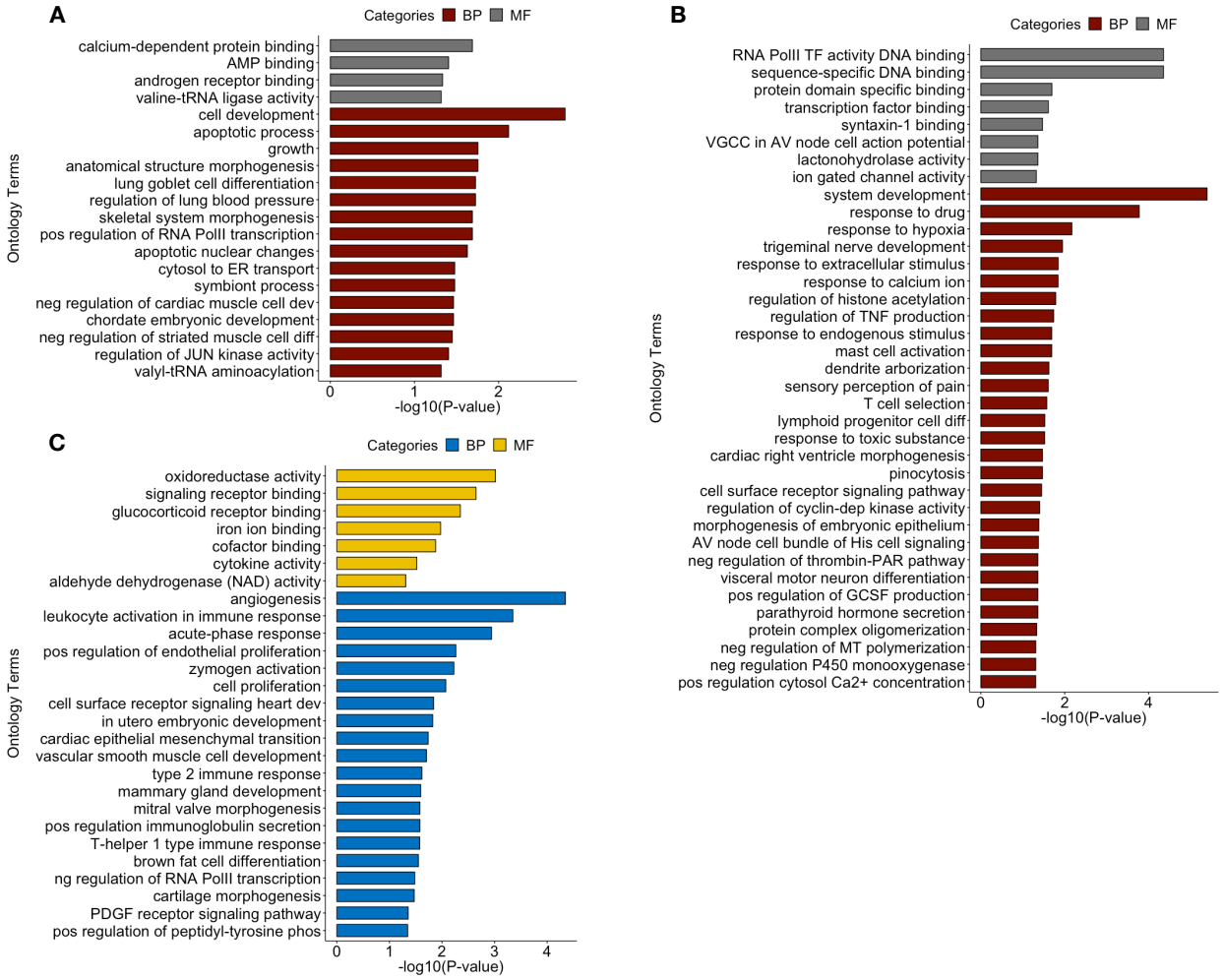


Figure 2. Enrichment of biological processes and molecular functions of differentially methylated genes and genes expression signatures associated with West African ancestry. Gene ontology terms that are enriched for biological processes (BP) and molecular functions (MF) for **(A)** 432 genes annotated to differentially hypermethylated regions, **(B)** 602 genes annotated to differentially hypomethylated regions and **(C)** 131 genes with gene expression associated to WAA (FDR < 0.10). *P*-values are BH-adjusted *p*-values obtained from gProfiler.

Table 1. Demographics and clinical characteristics of hepatocyte cohorts

Variable	RNA-Seq Cohort	Methylation Cohort	GTEx Liver Cohort
	AA (n = 60)	AA (n = 44)	AA (n = 12)
Age, years (mean ± SD)	39.44 ± 20.50	45.89 ± 12.22	52 ± 11.56
Sex (Female %)	48.3	52.3	33.3
West African ancestry (%) (mean ± SD)	78.17 ± 12.86	79.83 ± 11.86	72.09 ± 25.11

AA: African American

Table2. PharmGKB VIP gene expression levels associated with West African Ancestry

	AA Hepatocyte Cohort			GTEx AA Liver Cohort		
Gene	Coefficient (%)	P	FDR	Coefficient (%)	P	FDR
ALDH1A1	1.027	0.0320	0.3605	-2.648	0.0456	0.1769
*CYP2C19**	-1.362	0.0322	0.3605	-2.484	0.0040	0.0829
P2RY1	1.608	0.0022	0.0600	2.4899	0.2254	0.4157
PTGIS	-0.574	0.0021	0.0600	1.534	0.3670	0.4811
*VDR**	0.350	0.0032	0.0600	2.227	0.0194	0.1208

* Genes identified in the AA hepatocyte cohort replicated in GTEx AA liver cohort.

Table 3. Genome-wide gene expression associated with West African Ancestry

	AA Hepatocyte Cohort			GTEx AA Liver Cohort
Gene	Coefficient (%)	P	FDR	FDR
AKAP8	-0.812	1.07E-05	0.0307	0.7283
ALDH1A1	2.077	6.56E-05	0.0471	0.5803
APOL1	-2.848	3.83E-05	0.0380	0.6114
BBOX1	3.049	8.23E-05	0.0485	0.6217
CRYM	1.573	3.69E-05	0.0380	0.9219
EFNA1	-3.540	7.23E-05	0.0471	0.9191
ENSG00000205622	-1.547	2.04E-05	0.0380	0.7318
ENSG00000250764	-0.540	3.53E-05	0.0380	0.9961
ENSG00000259357	1.869	8.06E-05	0.0485	-
ENSG00000271239	5.823	4.16E-05	0.0392	0.8626
FUT11	-1.749	7.09E-05	0.0471	0.9783
GGT1	2.155	8.14E-05	0.0485	0.9287
GPI	-1.453	4.58E-05	0.0411	0.9236
HIC1	-1.567	9.64E-06	0.0307	0.9153
HILPDA	-2.583	9.83E-06	0.0307	0.7829
IL18	3.152	7.25E-05	0.0471	0.6804
IL33	-1.069	6.76E-05	0.0471	0.9067
LOX	-4.029	2.81E-05	0.0380	0.9586
MRO	1.131	2.57E-05	0.0380	0.9830
NPR2	1.235	6.11E-05	0.0471	0.7045
OSMR	-2.193	9.69E-06	0.0307	0.7933
P4HA1	-2.038	2.29E-05	0.0380	0.8679
PLAT	-1.883	3.49E-05	0.0380	0.8603
PLCL2	0.760	6.32E-05	0.0471	0.6250
PTGIS	-0.855	4.24E-06	0.0307	0.7748
RFLNB	-3.175	1.14E-05	0.0307	0.8333
RNF135	1.801	2.67E-05	0.0380	0.8876
SIVA1	-2.137	5.42E-06	0.0307	0.8836
SLC2A3	-1.961	2.99E-05	0.0380	0.8554
SLC39A11	0.491	3.02E-05	0.0380	0.7213
SPIN1	0.675	1.52E-05	0.0358	0.6449
VEGFA	-2.246	4.91E-05	0.0421	0.9167

References

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-169. doi:10.1093/bioinformatics/btu638

Andersen, M. R., Farin, F. M., & Omiecinski, C. J. (1998). Quantification of multiple human cytochrome P450 mRNA molecules using competitive reverse transcriptase-PCR. *DNA Cell Biol, 17*(3), 231-238. doi:10.1089/dna.1998.17.231

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., & Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nat Protoc, 5*(9), 1564-1573. doi:10.1038/nprot.2010.116

Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D., & Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics, 30*(10), 1363-1369. doi:10.1093/bioinformatics/btu049

Austin, H., Hooper, W. C., Lally, C., Dilley, A., Ellingsen, D., Wideman, C., . . . Evatt, B. (2000). Venous thrombosis in relation to fibrinogen and factor VII genes among African-Americans. *J Clin Epidemiol, 53*(10), 997-1001.

Baharian, S., Barakatt, M., Gignoux, C. R., Shringarpure, S., Errington, J., Blot, W. J., . . . Gravel, S. (2016). The Great Migration and African-American Genomic Diversity. *PLoS Genet, 12*(5), e1006059. doi:10.1371/journal.pgen.1006059

Barfield, R. T., Almli, L. M., Kilaru, V., Smith, A. K., Mercer, K. B., Duncan, R., . . . Conneely, K. N. (2014). Accounting for population stratification in DNA methylation studies. *Genet Epidemiol, 38*(3), 231-241. doi:10.1002/gepi.21789

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nat Genet, 40*(3), 340-345. doi:10.1038/ng.78

Beydoun, M. A., Hossain, S., Tajuddin, S. M., Canas, J. A., Kuczmarski, M., Beydoun, H. A., . . . Zonderman, A. B. (2018). Vitamin D Metabolism-Related Gene Haplotypes and Their Association with Metabolic Disturbances Among African-American Urban Adults. *Sci Rep, 8*(1), 8035. doi:10.1038/s41598-018-26230-w

Beydoun, M. A., Tajuddin, S. M., Dore, G. A., Canas, J. A., Beydoun, H. A., Evans, M. K., & Zonderman, A. B. (2017). Vitamin D Receptor and Megalin Gene Polymorphisms Are Associated with Longitudinal Cognitive Change among African-American Urban Adults. *J Nutr, 147*(6), 1048-1062. doi:10.3945/jn.116.244962

Boger, C. A., & Heid, I. M. (2011). Chronic kidney disease: novel insights from genome-wide association studies. *Kidney Blood Press Res, 34*(4), 225-234. doi:10.1159/000326901

Boulter, A. C., Quinlan, J., Miro-Herrans, A. T., Pearson, L. N., Todd, N. L., Committee, H. S., . . . Mulligan, C. J. (2015). Interaction of Alu Polymorphisms and Novel Measures of Discrimination in Association with Blood Pressure in African Americans Living in Tallahassee, Florida. *Hum Biol, 87*(4), 295-305.

Caliskan, M., Manduchi, E., Rao, H. S., Segert, J. A., Beltrame, M. H., Trizzino, M., . . . Brown, C. D. (2018). Genetic And Epigenetic Fine Mapping Of Complex Trait Associated Loci In The Human Liver. *bioRxiv doi: https://doi.org/10.1101/432823*.

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet, 9*, 403-433. doi:10.1146/annurev.genom.9.081307.164258

Cancer Genome Atlas Research, N., Linehan, W. M., Spellman, P. T., Ricketts, C. J., Creighton, C. J., Fei, S. S., . . . Zuna, R. (2016). Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med, 374*(2), 135-145. doi:10.1056/NEJMoa1505917

Chalubinski, M., Luczak, E., Wojdan, K., Gorzelak-Pabis, P., & Broncel, M. (2016). Innate lymphoid cells type 2 - emerging immune regulators of obesity and atherosclerosis. *Immunol Lett, 179*, 43-46. doi:10.1016/j.imlet.2016.09.007

Consortium, G. T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet, 45*(6), 580-585. doi:10.1038/ng.2653

Consortium, G. T., Laboratory, D. A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G. g., Fund, N. I. H. C., . . . Montgomery, S. B. (2017). Genetic effects on gene expression across human tissues. *Nature, 550*(7675), 204-213. doi:10.1038/nature24277

Cresci, S., Depta, J. P., Lenzini, P. A., Li, A. Y., Lanfear, D. E., Province, M. A., . . . Bach, R. G. (2014). Cytochrome p450 gene variants, race, and mortality among clopidogrel-treated patients after acute myocardial infarction. *Circ Cardiovasc Genet, 7*(3), 277-286. doi:10.1161/CIRCGENETICS.113.000303

Daw, J. (2017). Contribution of Four Comorbid Conditions to Racial/Ethnic Disparities in Mortality Risk. *Am J Prev Med, 52*(1S1), S95-S102. doi:10.1016/j.amepre.2016.07.036

De, T., Alarcon, C., Hernandez, W., Liko, I., Cavallari, L. H., Duarte, J. D., & Perera, M. A. (2018). Association of Genetic Variants With Warfarin-Associated Bleeding Among Patients of African Descent. *JAMA, 320*(16), 1670-1677. doi:10.1001/jama.2018.14955

De, T., Park, C. S., & Perera, M. A. (2018). Cardiovascular Pharmacogenomics: Does It Matter If You're Black or White? *Annu Rev Pharmacol Toxicol*. doi:10.1146/annurev-pharmtox-010818-021154

Dean, L. (2012). Prasugrel Therapy and CYP Genotype. In V. Pratt, H. McLeod, W. Rubinstein, L. Dean, B. Kattman, & A. Malheiro (Eds.), *Medical Genetics Summaries*. Bethesda (MD).

Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., . . . Feinberg, A. P. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet, 41*(12), 1350-1353. doi:10.1038/ng.471

Galanter, J. M., Gignoux, C. R., Oh, S. S., Torgerson, D., Pino-Yanes, M., Thakur, N., . . . Zaitlen, N. (2017). Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife, 6*. doi:10.7554/eLife.20532

Geeleher, P., Hartnett, L., Egan, L. J., Golden, A., Raja Ali, R. A., & Seoighe, C. (2013). Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics, 29*(15), 1851-1857. doi:10.1093/bioinformatics/btt311

Grossman, S. R., Andersen, K. G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., . . . Sabeti, P. C. (2013). Identifying recent adaptations in large-scale genomic data. *Cell, 152*(4), 703-713. doi:10.1016/j.cell.2013.01.035

Hernandez, W., Gamazon, E. R., Smithberger, E., O'Brien, T. J., Harralson, A. F., Tuck, M., . . . Perera, M. A. (2016). Novel genetic predictors of venous thromboembolism risk in African Americans. *Blood, 127*(15), 1923-1929. doi:10.1182/blood-2015-09-668525

Hooper, W. C., Lally, C., Austin, H., Renshaw, M., Dilley, A., Wenger, N. K., . . . Evatt, B. L. (2000). The role of the t-PA I/D and PAI-1 4G/5G polymorphisms in African-American adults with a diagnosis of myocardial infarction or venous thromboembolism. *Thromb Res, 99*(3), 223-230.

Huang, J., Huffman, J. E., Yamakuchi, M., Trompet, S., Asselbergs, F. W., Sabater-Lleal, M., . . . Group, C. C. H. F. W. (2014). Genome-wide association study for circulating tissue plasminogen activator levels and functional follow-up implicates endothelial STXBP5 and STX2. *Arterioscler Thromb Vasc Biol, 34*(5), 1093-1101. doi:10.1161/ATVBAHA.113.302088

Janicki, P. K., Eyileten, C., Ruiz-Velasco, V., Sedeek, K. A., Pordzik, J., Czlonkowska, A., . . . Postula, M. (2017). Population-Specific Associations of Deleterious Rare Variants in Coding Region of P2RY1-P2RY12 Purinergic Receptor Genes in Large-Vessel Ischemic Stroke Patients. *Int J Mol Sci, 18*(12). doi:10.3390/ijms18122678

Jarrar, M., Behl, S., Manyam, G., Ganah, H., Nazir, M., Nasab, R., & Moustafa, K. (2016). Cytochrome allelic variants and clopidogrel metabolism in cardiovascular diseases therapy. *Mol Biol Rep, 43*(6), 473-484. doi:10.1007/s11033-016-3983-1

Jiao, Y., Widschwendter, M., & Teschendorff, A. E. (2014). A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinformatics, 30*(16), 2360-2366. doi:10.1093/bioinformatics/btu316

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics, 8*(1), 118-127. doi:10.1093/biostatistics/kxj037

Kathiresan, S., Gabriel, S. B., Yang, Q., Lochner, A. L., Larson, M. G., Levy, D., . . . O'Donnell, C. J. (2005). Comprehensive survey of common genetic variation at the plasminogen activator inhibitor-1 locus and relations to circulating plasminogen activator inhibitor-1 levels. *Circulation, 112*(12), 1728-1735. doi:10.1161/CIRCULATIONAHA.105.547836

Keemink, J., Oorts, M., & Annaert, P. (2015). Primary Hepatocytes in Sandwich Culture. *Methods Mol Biol, 1250*, 175-188. doi:10.1007/978-1-4939-2074-7_12

Kruzel-Davila, E., Wasser, W. G., & Skorecki, K. (2017). APOL1 Nephropathy: A Population Genetics and Evolutionary Medicine Detective Story. *Semin Nephrol, 37*(6), 490-507. doi:10.1016/j.semnephrol.2017.07.002

Lane, S., Al-Zubiedi, S., Hatch, E., Matthews, I., Jorgensen, A. L., Deloukas, P., . . . Pirmohamed, M. (2012). The population pharmacokinetics of R- and S-warfarin: effect of genetic and clinical factors. *Br J Clin Pharmacol, 73*(1), 66-76. doi:10.1111/j.1365-2125.2011.04051.x

Lee, S. M., Schelcher, C., Demmel, M., Hauner, M., & Thasler, W. E. (2013). Isolation of human hepatocytes by a two-step collagenase perfusion procedure. *J Vis Exp*(79). doi:10.3791/50615

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics, 28*(6), 882-883. doi:10.1093/bioinformatics/bts034

Li, J. L., Fu, Y., Qin, S. B., Liang, G. K., Liu, J., Nie, X. Y., . . . Lu, Y. (2018). Association between P2RY12 gene polymorphisms and adverse clinical events in coronary artery disease patients treated with clopidogrel: A systematic review and meta-analysis. *Gene, 657*, 69-80. doi:10.1016/j.gene.2018.03.007

Li, Y. C. (2013). Vitamin D receptor signaling in renal and cardiovascular protection. *Semin Nephrol, 33*(5), 433-447. doi:10.1016/j.semnephrol.2013.07.005

Liew, F. Y., Girard, J. P., & Turnquist, H. R. (2016). Interleukin-33 in health and disease. *Nat Rev Immunol, 16*(11), 676-689. doi:10.1038/nri.2016.95

Liu, C. T., Garnaas, M. K., Tin, A., Kottgen, A., Franceschini, N., Peralta, C. A., . . . Fox, C. S. (2011). Genetic association for renal traits among participants of African ancestry reveals new loci for renal function. *PLoS Genet, 7*(9), e1002264. doi:10.1371/journal.pgen.1002264

Liu, Y., Wang, F., Liu, Y., Yao, Y., Lv, X., Dong, B., . . . Xu, Y. (2016). RNF135, RING finger protein, promotes the proliferation of human glioblastoma cells in vivo and in vitro via the ERK pathway. *Sci Rep, 6*, 20642. doi:10.1038/srep20642

Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., . . . consortium, F. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol, 16*, 22. doi:10.1186/s13059-014-0560-6

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Majumder, S., & Advani, A. (2017). VEGF and the diabetic kidney: More than too much of a good thing. *J Diabetes Complications, 31*(1), 273-279. doi:10.1016/j.jdiacomp.2016.10.020

Mannucci, P. M., Mari, D., Merati, G., Peyvandi, F., Tagliabue, L., Sacchi, E., . . . Franceschi, C. (1997). Gene polymorphisms predicting high plasma levels of coagulation and fibrinolysis proteins. A study in centenarians. *Arterioscler Thromb Vasc Biol, 17*(4), 755-759.

Mishra, D. K., Wu, Y., Sarkissyan, M., Sarkissyan, S., Chen, Z., Shang, X., . . . Vadgama, J. V. (2013). Vitamin D receptor gene polymorphisms and prognosis of breast cancer among African-American and Hispanic women. *PLoS One, 8*(3), e57967. doi:10.1371/journal.pone.0057967

Mondul, A. M., Weinstein, S. J., Layne, T. M., & Albanes, D. (2017). Vitamin D and Cancer Risk and Mortality: State of the Science, Gaps, and Challenges. *Epidemiol Rev, 39*(1), 28-48. doi:10.1093/epirev/mxx005

Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K., & Beck, S. (2014). ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics, 30*(3), 428-430. doi:10.1093/bioinformatics/btt684

Nagareddy, P., & Smyth, S. S. (2013). Inflammation and thrombosis in cardiovascular disease. *Curr Opin Hematol, 20*(5), 457-463. doi:10.1097/MOH.0b013e328364219d

Ndaw, V. S., Abebayehu, D., Spence, A. J., Paez, P. A., Kolawole, E. M., Taruselli, M. T., . . . Ryan, J. J. (2017). TGF-beta 1 Suppresses IL-33-Induced Mast Cell Function. *Journal of Immunology, 199*(3), 866-873. doi:10.4049/jimmunol.1601983

Nedelec, Y., Sanz, J., Baharian, G., Szpiech, Z. A., Pacis, A., Dumaine, A., . . . Barreiro, L. B. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell, 167*(3), 657-669 e621. doi:10.1016/j.cell.2016.09.025

Nguyen, T. A., Diodati, J. G., & Pharand, C. (2005). Resistance to clopidogrel: a review of the evidence. *J Am Coll Cardiol, 45*(8), 1157-1164. doi:10.1016/j.jacc.2005.01.034

Nordlund, J., Backlin, C. L., Wahlberg, P., Busche, S., Berglund, E. C., Eloranta, M. L., . . . Syvanen, A. C. (2013). Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol, 14*(9), r105. doi:10.1186/gb-2013-14-9-r105

Notarangelo, M. F., Bontardelli, F., & Merlini, P. A. (2013). Genetic and nongenetic factors influencing the response to clopidogrel. *J Cardiovasc Med (Hagerstown), 14 Suppl 1*, S1-7. doi:10.2459/JCM.0b013e328364bb04

O'Connor, M. Y., Thoreson, C. K., Ramsey, N. L., Ricks, M., & Sumner, A. E. (2013). The uncertain significance of low vitamin D levels in African descent populations: a review of the bone and cardiometabolic literature. *Prog Cardiovasc Dis, 56*(3), 261-269. doi:10.1016/j.pcad.2013.10.015

Ongen, H., Brown, A. A., Delaneau, O., Panousis, N. I., Nica, A. C., Consortium, G. T., & Dermitzakis, E. T. (2017). Estimating the causal tissues for complex traits and diseases. *Nat Genet, 49*(12), 1676-1683. doi:10.1038/ng.3981

Park, D. S., Eskin, I., Kang, E. Y., Gamazon, E. R., Eng, C., Gignoux, C. R., . . . Zaitlen, N. (2018). An ancestry-based approach for detecting interactions. *Genet Epidemiol, 42*(1), 49-63. doi:10.1002/gepi.22087

Park, L. (2016). Evidence of Recent Intricate Adaptation in Human Populations. *PLoS One, 11*(12), e0165870. doi:10.1371/journal.pone.0165870

Pendyala, L. K., Torguson, R., Loh, J. P., Devaney, J. M., Chen, F., Kitabata, H., . . . Waksman, R. (2013). Racial disparity with on-treatment platelet reactivity in patients undergoing percutaneous coronary intervention. *Am Heart J, 166*(2), 266-272. doi:10.1016/j.ahj.2013.04.008

Perera, M. A., Cavallari, L. H., Limdi, N. A., Gamazon, E. R., Konkashbaev, A., Daneshjou, R., . . . Johnson, J. A. (2013). Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. *Lancet, 382*(9894), 790-796. doi:10.1016/S0140-6736(13)60681-9

Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., R, V. L., . . . Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin, 8*, 6. doi:10.1186/1756-8935-8-6

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics, 197*(2), 573-589. doi:10.1534/genetics.114.164350

Rathi, A., Virmani, A. K., Harada, K., Timmons, C. F., Miyajima, K., Hay, R. J., . . . Gazdar, A. F. (2003). Aberrant methylation of the HIC1 promoter is a frequent event in specific pediatric neoplasms. *Clin Cancer Res, 9*(10 Pt 1), 3674-3678.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., & Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res, 44*(W1), W83-89. doi:10.1093/nar/gkw199

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol, 11*(3), R25. doi:10.1186/gb-2010-11-3-r25

Rodriguez-Antona, C., Jover, R., Gomez-Lechon, M. J., & Castell, J. V. (2000). Quantitative RT-PCR measurement of human cytochrome P-450s: application to drug induction studies. *Arch Biochem Biophys, 376*(1), 109-116. doi:10.1006/abbi.2000.1697

Sarkissyan, M., Wu, Y., Chen, Z., Mishra, D. K., Sarkissyan, S., Giannikopoulos, I., & Vadgama, J. V. (2014). Vitamin D receptor FokI gene polymorphisms may be associated with colorectal cancer among African American and Hispanic participants. *Cancer, 120*(9), 1387-1393. doi:10.1002/cncr.28565

Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y., . . . Ulrich, R. (2008). Mapping the genetic architecture of gene expression in human liver. *PLoS Biol, 6*(5), e107. doi:10.1371/journal.pbio.0060107

Scott, D. M., & Taylor, R. E. (2007). Health-related effects of genetic variations of alcohol-metabolizing enzymes in African Americans. *Alcohol Res Health, 30*(1), 18-21.

Shuldiner, A. R., O'Connell, J. R., Bliden, K. P., Gandhi, A., Ryan, K., Horenstein, R. B., . . . Gurbel, P. A. (2009). Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA, 302*(8), 849-857. doi:10.1001/jama.2009.1232

Siggens, L., & Ekwall, K. (2014). Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. *J Intern Med, 276*(3), 201-214. doi:10.1111/joim.12231

Simon, T., Verstuyft, C., Mary-Krause, M., Quteineh, L., Drouet, E., Meneveau, N., . . . Non, S. T. E. M. I. I. (2009). Genetic determinants of response to clopidogrel and cardiovascular events. *N Engl J Med, 360*(4), 363-375. doi:10.1056/NEJMoa0808227

Sivaskandarajah, G. A., Jeansson, M., Maezawa, Y., Eremina, V., Baelde, H. J., & Quaggin, S. E. (2012). Vegfa protects the glomerular microvasculature in diabetes. *Diabetes, 61*(11), 2958-2966. doi:10.2337/db11-1655

Slieker, R. C., Relton, C. L., Gaunt, T. R., Slagboom, P. E., & Heijmans, B. T. (2018). Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. *Epigenetics Chromatin, 11*(1), 25. doi:10.1186/s13072-018-0191-3

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol, 3*, Article3. doi:10.2202/1544-6115.1027

Stegle, O., Parts, L., Piipari, M., Winn, J., & Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc, 7*(3), 500-507. doi:10.1038/nprot.2011.457

Stephen, J. K., Chen, K. M., Shah, V., Havard, S., Kapke, A., Lu, M., . . . Worsham, M. J. (2010). DNA hypermethylation markers of poor outcome in laryngeal cancer. *Clin Epigenetics, 1*(1-2), 61-69. doi:10.1007/s13148-010-0005-3

Tastet, J., Decalonne, L., Marouillat, S., Malvy, J., Thepault, R. A., Toutain, A., . . . Vourc'h, P. (2015). Mutation screening of the ubiquitin ligase gene RNF135 in French patients with autism. *Psychiatr Genet, 25*(6), 263-267. doi:10.1097/YPG.0000000000000100

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics, 29*(2), 189-196. doi:10.1093/bioinformatics/bts680

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Gayther, S. A., Apostolidou, S., . . . Widschwendter, M. (2009). An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One, 4*(12), e8274. doi:10.1371/journal.pone.0008274

Timur, A. A., Murugesan, G., Zhang, L., Aung, P. P., Barnard, J., Wang, Q. K., . . . Kottke-Marchant, K. (2012). P2RY1 and P2RY12 polymorphisms and on-aspirin platelet reactivity in patients with coronary artery disease. *Int J Lab Hematol, 34*(5), 473-483. doi:10.1111/j.1751-553X.2012.01420.x

Vasiliou, V., & Pappa, A. (2000). Polymorphisms of human aldehyde dehydrogenases. Consequences for drug metabolism and disease. *Pharmacology, 61*(3), 192-198. doi:10.1159/000028400

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci, 131*(4), 281-285. doi:10.1007/s12064-012-0162-3

Wang, F. S., Fan, J. G., Zhang, Z., Gao, B., & Wang, H. Y. (2014). The global burden of liver disease: the major impact of China. *Hepatology, 60*(6), 2099-2108. doi:10.1002/hep.27406

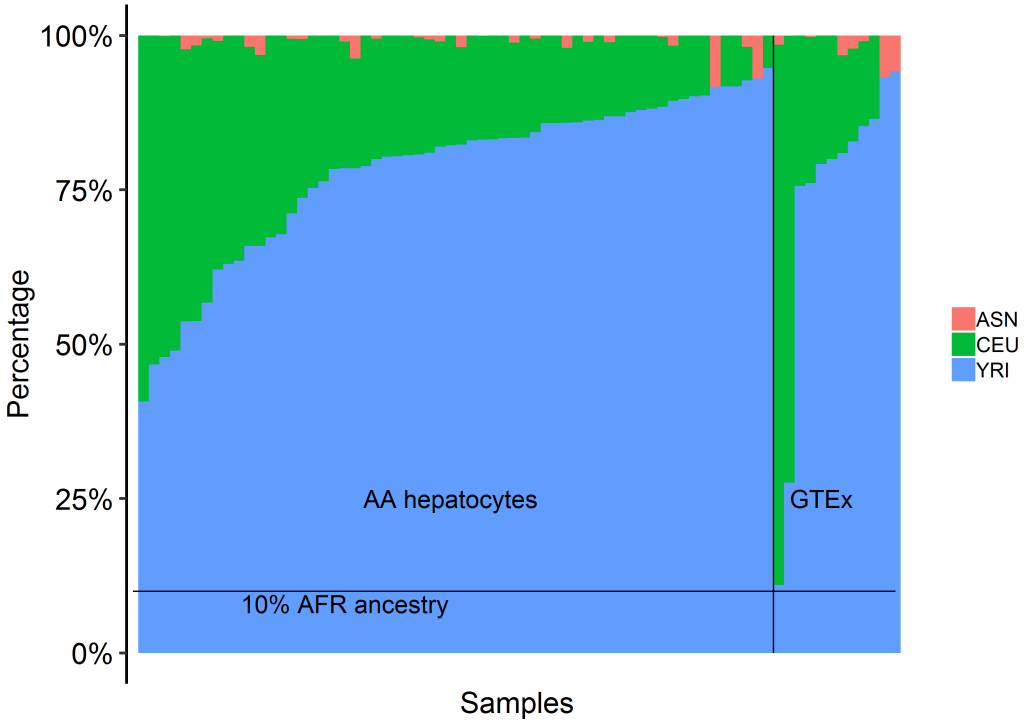
Wang, Y., Freedman, J. A., Liu, H., Moorman, P. G., Hyslop, T., George, D. J., . . . Wei, Q. (2017). Associations between RNA splicing regulatory variants of stemness-related genes and racial disparities in susceptibility to prostate cancer. *Int J Cancer, 141*(4), 731-743. doi:10.1002/ijc.30787

Whirl-Carrillo, M., McDonagh, E. M., Hebert, J. M., Gong, L., Sangkuhl, K., Thorn, C. F., . . . Klein, T. E. (2012). Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther, 92*(4), 414-417. doi:10.1038/clpt.2012.96

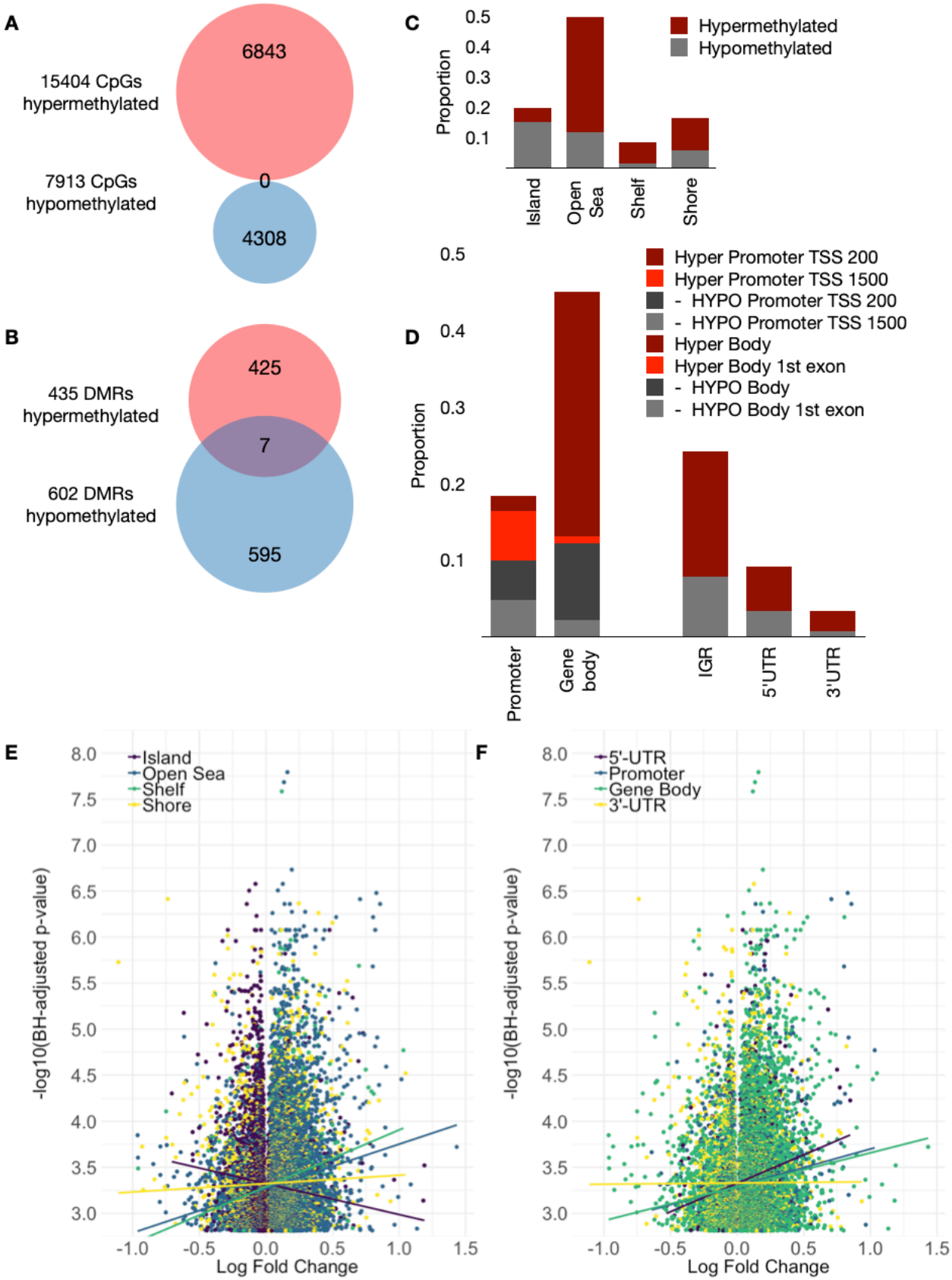
Yang, X., Han, H., De Carvalho, D. D., Lay, F. D., Jones, P. A., & Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell, 26*(4), 577-590. doi:10.1016/j.ccr.2014.07.028

Zhang, H., Ma, K., Liu, W., Yang, F., Liu, J., & Zhou, H. (2016). Impact of CYP2C19 gene polymorphism on warfarin maintenance doses in patients with non-valvular atrial fibrillation. *Gene, 591*(1), 80-84. doi:10.1016/j.gene.2016.06.046

Supplementary Files



Supplementary Figure 2. West African ancestry estimation. An admixture plot showing the ancestry components of the 60 African-American hepatocyte and 12 GTEx cohorts was calculated using fastStructure v1.3. A model with three parental ancestral components (K = 3) was used to estimate from three populations from HapMap phase 3 reference data: Han Chinese in Beijing, China (CHB) and Japanese in Tokyo, Japan (JPT) combined as Asian (ASN, pink); Utah residents with Northern and Western European ancestry (CEU, green); Yoruba in Ibadan, Nigeria (YRI) as African (AFR, blue). Percent WAA is plotted on the y-axis. Each sample is represented individually on the x-axis and partitioned into segments proportional to the contributions of the ancestral components to the individual's genome.


A
15404 CpGs hypermethylated
7913 CpGs hypomethylated
6843
0
4308
B
435 DMRs hypermethylated
602 DMRs hypomethylated
425
7
595
C
Hypermethylated
Hypomethylated
Proportion
Island
Open Sea
Shelf
Shore
D
Hyper Promoter TSS 200
Hyper Promoter TSS 1500
- HYPO Promoter TSS 200
- HYPO Promoter TSS 1500
Hyper Body
Hyper Body 1st exon
- HYPO Body
- HYPO Body 1st exon
Proportion
Promoter
Gene body
IGR
5'UTR
3'UTR
E
Island
Open Sea
Shelf
Shore
-log10(BH-adjusted p-value)
Log Fold Change
F
5'-UTR
Promoter
Gene Body
3'-UTR
-log10(BH-adjusted p-value)
Log Fold Change

Supplementary Figure 2. DNA methylation by genomic features and regions associated with West African ancestry. Venn diagrams of **(A)** unique genes comprised of 15,404 hyper- (pink) and 7913 hypomethylated (blue) CpG sites associated with WAA, and **(B)** unique genes comprised of 435 hyper- (pink) and 602 hypomethylated (blue) differentially methylated regions (DMRs) associated with WAA, **(C)** Bar graphs showing the proportion of hyper- and hypomethylated CpG sites by genomic features(islands, open sea, shelf and shore), and **(D)** proportion of hyper- and hypomethylated CpG sites by transcriptionally-regulated regions; promoter, gene body (with 200 or 1500 bp of the gene transcriptional start site [TSS]), intergenic region (IGR) 5' and 3' untranslated regions (UTR). Volcano plots showing the correlation of log fold change in DNA methylation of the 23,317 significant DM CpG sites (BH-corrected $p < 0.05$) and p-value for the association of each CpG site to WAA by **(E)** genomic feature and by **(F)** transcriptionally-regulated regions.