


RESEARCH

Open Access



# Ancestry-driven metabolite variation provides insights into disease states in admixed populations

Kaylia M. Reynolds<sup>1,2†</sup>, Andrea R. V. R. Horimoto<sup>3†</sup>, Bridget M. Lin<sup>1</sup>, Ying Zhang<sup>4</sup>, Nuzulul Kurniansyah<sup>4</sup>, Bing Yu<sup>5</sup>, Eric Boerwinkle<sup>5</sup>, Qibin Qi<sup>6</sup>, Robert Kaplan<sup>6,7</sup>, Martha Daviglus<sup>8</sup>, Lifang Hou<sup>9</sup>, Laura Y. Zhou<sup>1</sup>, Jianwen Cai<sup>1</sup>, Saame Raza Shaikh<sup>10</sup>, Tamar Sofer<sup>4,11</sup>, Sharon R. Browning<sup>3</sup> and Nora Franceschini<sup>2\*</sup> 

## Abstract

**Background** Metabolic pathways are related to physiological functions and disease states and are influenced by genetic variation and environmental factors. Hispanics/Latino individuals have ancestry-derived genomic regions (local ancestry) from their recent admixture that have been less characterized for associations with metabolite abundance and disease risk.

**Methods** We performed admixture mapping of 640 circulating metabolites in 3887 Hispanic/Latino individuals from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Metabolites were quantified in fasting serum through non-targeted mass spectrometry (MS) analysis using ultra-performance liquid chromatography-MS/MS. Replication was performed in 1856 nonoverlapping HCHS/SOL participants with metabolomic data.

**Results** By leveraging local ancestry, this study identified significant ancestry-enriched associations for 78 circulating metabolites at 484 independent regions, including 116 novel metabolite-genomic region associations that replicated in an independent sample. Among the main findings, we identified Native American enriched genomic regions at chromosomes 11 and 15, mapping to *FADS1/FADS2* and *LIPC*, respectively, associated with reduced long-chain polyunsaturated fatty acid metabolites implicated in metabolic and inflammatory pathways. An African-derived genomic region at chromosome 2 was associated with N-acetylated amino acid metabolites. This region, mapped to *ALMS1*, is associated with chronic kidney disease, a disease that disproportionately burdens individuals of African descent.

**Conclusions** Our findings provide important insights into differences in metabolite quantities related to ancestry in admixed populations including metabolites related to regulation of lipid polyunsaturated fatty acids and N-acetylated amino acids, which may have implications for common diseases in populations.

**Keywords** Admixture mapping, Local ancestry, Metabolites, Hispanics/Latino populations

<sup>†</sup>Kaylia M. Reynolds and Andrea R. V. R. Horimoto shared co-authors.

\*Correspondence:  
Nora Franceschini  
noraf@unc.edu

Full list of author information is available at the end of the article



## Background

Circulating metabolites can provide insights into biological processes in health and disease states [1]. Metabolites are end-products of metabolic cellular pathways regulated by genetic variation and environmental factors such as diet, microbiome, and exposure to exogenous compounds [2]. Studies have identified metabolite profiles for complex diseases including diabetes, obesity, and chronic kidney disease [3, 4]. The integration of genotypes and metabolites has provided insights into disease mechanisms [5]. This research has been facilitated by the generation of high-throughput metabolomics in large datasets with genome-wide genotypes. Using genome-wide association approaches, several large-scale studies have identified rare and common genetic variants regulating blood metabolite levels in populations [6–8], and genes related to disease processes or drug targets [9].

However, the role of genetic ancestry in metabolite abundance and regulation is not well known. This question is relevant to populations with recent admixture, such as Hispanic/Latino populations, who have a high burden of metabolic diseases. Hispanic/Latino populations were shaped by a long history of colonization and migration across America, which introduced high diversity in both cultural aspects and genetic ancestry that can be observed in the Hispanic/Latino individuals currently in the USA [10, 11]. They carry genetic variants from three continental ancestries (West African, European, and Native American), and their genome is composed of varying segments of these different ancestral origins that can be mapped to local ancestries, i.e., each chromosomal segment can be attributed to a source, local (segment-specific) ancestry. Genomic regions associated with a specific ancestry are expected to be enriched for ancestry-derived variants. There are several examples of ancestry-derived genetic variations that confer disease risk in populations, some of them driven by adaptation to environmental factors such as dietary restrictions or exposure to pathogens [12, 13]. For example, in Hispanic/Latino individuals, Native American single nucleotide variants (SNVs) at the *SLC16A11* gene are associated with risk of diabetes, and African ancestry SNVs in the *APOL1* gene (related to plasmodium pathogen resistance) are associated with chronic kidney disease [14, 15]. Therefore, research focusing on genetic ancestry could point to differences in metabolic regulation across populations and provide insight into differences in disease risk.

We leveraged local ancestries for a comprehensive study of ancestry-derived genomic regions associated with circulating metabolites in admixed Hispanic/Latino individuals. Recent admixture creates long-range blocks of linkage disequilibrium (admixture-LD) between

genetic variants with differences in allele frequencies in the parental populations. The extent of admixture depends on the population admixture dynamics and time since admixture, and recombination rates. Causal variants should occur more frequently on chromosomal segments inherited from ancestral populations with higher disease burden. Therefore, associations of phenotypes with local ancestry regions are likely to improve the discovery of causal variants enriched in their population of origin for both rare and common SNVs. We studied the genetic ancestry influences on 640 circulating metabolites in 3887 Hispanic/Latino individuals from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) using admixture mapping. This approach has been successfully applied in HCHS/SOL, with the identification of novel loci associated with blood pressure [16], chronic kidney disease traits [17, 18], and pulmonary traits [19, 20]. We used summary statistics from a published genome-wide association study (GWAS) of metabolites in the HCHS/SOL for fine-mapping loci and replicated novel findings in a non-overlapping HCHS/SOL sample.

## Methods

### HCHS/SOL study design, population, and covariates

The HCHS/SOL is a multi-center community-based prospective cohort study of 16,415 self-identified Hispanic/Latino individuals aged 18–74 years who were recruited from households in predefined census-block groups in four US field centers (Chicago, Miami, the Bronx, and San Diego) between 2008 and 2011 [21]. The study used a two-stage area probability sample of households selected with stratification and oversampling incorporated at each stage to provide a broadly diverse sample of Hispanics/Latinos and to ensure that the target age distribution was obtained [22]. Participants self-reported their country of origin as Central America ( $n=1730$ ), Cuba ( $n=2348$ ), Dominican Republic ( $n=1460$ ), Mexico ( $n=6471$ ), Puerto Rico ( $n=2728$ ), or South America ( $n=1068$ ). A baseline clinical examination included clinical, behavioral, and sociodemographic assessments, and collection of fasting blood and spot urine samples. Estimated glomerular filtration rate (eGFR) was calculated using the race-free Chronic Kidney Disease Epidemiology Collaboration serum creatinine-based equation and used as a covariate in analyses.

### Metabolomics data and processing

We used two non-overlapping datasets sampled from the overall HCHS/SOL study for discovery and replication. The discovery included a sample of 3972 participants which was randomly selected from HCHS/SOL for serum metabolomic profiling. Serum metabolites

( $n=1136$ ; 782 with known and 354 unknown biochemical identities) were quantified in fasting serum through non-targeted mass spectrometry (MS) analysis using ultra-performance liquid chromatography (UPLC)-MS/MS (DiscoveryHD4™ platform, Metabolon Inc, NC) [23]. Identification and classification of metabolites used a comparison of the ion features in the experimental samples to a reference library of chemical standard entries (e.g., molecular weight (m/z), preferred adducts, in-source fragments, and associated MS spectra) and known chemical entities. Peaks were quantified using area-under-the-curve. Raw area counts for each metabolite in each sample were normalized to correct for variation resulting from instrument inter-day tuning differences. To avoid batch effects, samples were randomly allocated across the platform. Replicas were used to determine endogenous variability, with representative relative standard deviation of 10% across all biochemicals. Metabolites were inverse normally transformed to approximate a normal distribution. We excluded unknown metabolites and metabolites with 25% or more missing data (amino acids,  $n=12$ ; carbohydrates,  $n=2$ ; cofactor a,  $n=5$ ; lipid,  $n=23$ ; nucleotide,  $n=2$ ; peptide,  $n=3$ ; xenobiotics,  $n=95$ ). For metabolites with less than 25% missing data, missing values were imputed with the observed minimum value of the metabolite in the sample. A total of 640 metabolites were used in the analyses. We provided the Human Metabolome Database (HMDB) annotation when available and used LIPID MAPS to provide a RefMet-driven lipid standardized name [24].

The replication dataset included 2,330 HCHS/SOL participants who were profiled using the DiscoveryHD4™ platform (Metabolon Inc, NC) using the same protocols described for the discovery dataset. After removing individuals overlapping with the discovery dataset, duplicates, and individuals without genotypes, 1856 non-overlapping participants with the discovery sample were used for replication. We excluded metabolites with >25% missing values across individuals and imputed missing values for the remaining metabolites using the same methods as described in the discovery dataset.

### Genotyping and local ancestry

Participants were genotyped at over 2.5 million SNVs using a custom-built Illumina array. Details on genotyping, quality control proceedings, and imputation have been published [25–27]. Local ancestry references were estimated from 195 West African, 527 European, and 63 Native American samples from the Human Genome Diversity Project [28] and 1000 Genomes Project [29]. BEAGLE (v.4) was employed for phasing and imputation of sporadic missing genotypes in the HCHS/SOL and reference-panel datasets [30]. Local ancestry calls at

each locus were estimated using RFMix 1.5.4 [31], with the PopPhased option and a minimum node size of 5, as recommended in the documentation, and previously described [27]. The estimation of kinship coefficients, principal components (PCs) of ancestry, and genetic analysis groups are published [25].

### Statistical analyses

Descriptive statistics for continuous data were presented as the mean  $\pm$  standard deviation (SD) or median with interquartile range, and categorical variables as number and percentage. The HCHS/SOL study was developed under a complex sampling design [22]. To account for the correlation structure of the data, all regression models included three random effects: the pairwise kinship coefficients, household, and census block group (the geographic cluster of the households) to represent the correlation between participants to genetic relatedness and shared environmental effects.

### Admixture mapping

Admixture mapping analyses for each metabolite were performed using a linear mixed model framework implemented in the GENESIS R package [32] in which African, European, and Native American ancestries were tested simultaneously in a joint admixture mapping analysis [18]. Briefly, we first fit the models under the null hypothesis of no genetic ancestry effect while including multiple random (pairwise kinship coefficients, household, and census block group) and fixed (age, sex, eGFR, recruitment center, genetic analysis group, and the first five PCs) effects. We then used the null models to run the joint test for associations between ancestries at each locus and metabolites using a Wald test. Here, local ancestry is defined as the locus-specific ancestry allelic dosages (0, 1, or 2 copies of African, European, or Native American alleles, estimated from the genotype data) at each genomic interval. After identifying associations using the joint test, follow-up admixture mapping analyses then tested each ancestry against the others to determine the ancestry driving the signal at each associated local ancestry region. We tested a total of 15,500 local ancestry regions. We applied a significance threshold of  $5.04 \times 10^{-9}$ , which controls for a family-wise error rate at a level of 0.05 for multiple testing for the 640 metabolites and 15,500 local ancestry regions. The effect sizes of the associated loci were estimated using the allelic dosage of the ancestry driving the signal.

In a sensitivity analysis, we ran the admixture mapping models adjusting for global ancestry, defined as ancestry proportions estimated by averaging the local ancestry calls across the genome, instead of using the first five PCs, to ensure that our significant local ancestry

associations were not spurious. We obtained similar results adjusting by either PC or global ancestry proportions, and although 18 local ancestry regions were no longer significant, *p*-values had small changes. The sensitivity analysis results indicated that the PCs accurately adjusted for population structure and increased our confidence that observed associations were not spurious.

#### Conditional analysis of admixture mapping regions

RFMix infers local ancestry in a series of intervals, referred to here as local ancestry regions. Inferred local ancestry is constant within each region. However, local ancestry is not independent across neighboring regions because local ancestry tracts have average lengths of more than six centiMorgans in Hispanic individuals due to the onset of admixture occurring within the past 15 generations. To determine whether significant results from local ancestry regions located close in proximity were independent, we performed a conditional admixture mapping analysis on local ancestry regions located within ten centiMorgans of the most significant region associated with a metabolite. The allelic dosage of the ancestry driving the association signal was included as a covariate. We conservatively chose a significance threshold of  $5 \times 10^{-5}$  for this analysis and kept only ancestry-derived regions that remained significantly associated with a metabolite after conditioning.

#### Overlap of local ancestry regions with genetic variants identified in GWAS

We used the recently published GWAS of metabolites to identify the overlap of significant local ancestry regions with genetic variants significant in GWAS [33]. Local ancestry regions without any significant GWAS variants were considered novel associations discovered through admixture mapping.

#### Conditional analyses using GWAS summary statistics of metabolites in HCHS/SOL for fine-mapping

We used existing GWAS summary statistics from the HCHS/SOL cohort to fine-map potential variants within identified ancestry-specific regions associated with metabolites. The methods for the GWAS were described previously [8]. We extracted GWAS results for all SNVs located within significant local ancestry regions. We then used the conditional and joint association analysis (COJO) application from the software package Genome-wide Complex Trait Analysis (GCTA v1.93.2) to determine which SNVs (separately for each local ancestry region) were independent. The number of copies of the reference allele for each independent SNV was included as a covariate in a conditional admixture mapping analysis for its corresponding region [34]. The

reference population for calculating linkage disequilibrium was a sample of 5879 unrelated individuals from the HCHS/SOL study. Only GWAS SNVs from COJO modeling with *p*-values lower than  $5 \times 10^{-8}$  were included as covariates in admixture mapping conditional analyses. We considered a SNV to explain the admixture mapping association with a metabolite if the admixture mapping joint *p*-value increased to greater than  $5 \times 10^{-5}$  when conditioning on the SNV. For local ancestry regions with multiple COJO SNVs, if none of the COJO variants individually changed the admixture mapping joint *p*-value to more than  $5 \times 10^{-5}$ , all of the COJO SNVs for that region were tested together in a single admixture mapping conditional model.

#### Genotype annotation

We used several tools to annotate SNVs identified in conditional analysis for their functional impact including ANNOVAR [35] and the Ensembl (<http://uswest.ensembl.org/index.html>) and refGene annotation databases (<https://varianttools.sourceforge.net/Annotation/RefGene>). For non-coding variants, we used resources to assess evidence for enrichment in regulatory elements, including enhancers, transcription factor binding sites, and histone modification in tissues using a range of approaches implemented in FORGE2 [36].

#### Replication of admixture mapping findings

We performed admixture mapping in an independent sample from HCHS/SOL using the same statistical methods and covariates used in the discovery analysis. The threshold for significance was based on Bonferroni correction for 404 association tests of 64 metabolites with associations spread across 169 genomic regions. In addition, we compared the direction of effects (beta coefficients) and the ancestry driving the association between the discovery and replication samples.

## Results

### Admixture mapping of metabolites

We performed admixture mapping of metabolites in 3887 HCHS/SOL Hispanic/Latino individuals. The mean age was 45.9 years old, 57% were women, and the mean eGFR was 96.4 ml/min/1.73 m<sup>2</sup> (Additional File 1: Table S1). Participants lived in the USA and reported originated from the Mainland (Central and South America, and Mexico) and the Caribbean (Cuba, Dominican Republic, and Puerto Rico), and they had varying African, European, and Native American global ancestry proportions, as shown in Additional File 2: Fig. S1 and previously described [25]. Note that individuals from the Mainland had a higher proportion of Native American ancestry,



while those from the Caribbean had a higher proportion of African ancestry.

The overall study design and approach for analyses are shown in Fig. 1. Associations were performed among 640 metabolites with known biochemical identities and 15,500 local ancestry regions using a joint admixture mapping linear mixed model. A total of 651 local ancestry regions within twelve chromosomes were significantly associated with at least one metabolite, and 78 of the 640 tested metabolites were associated with at least one significant local ancestry region (Additional File 1: Table S2). Summarizing associations while allowing for multiple metabolite associations per region, our study identified a total of 2127 significant metabolite-local ancestry pair results (Additional File 1: Table S3).

#### Local ancestry independent associations

Given that several metabolites were frequently associated with multiple neighboring local ancestry regions, we tested associations conditioning on local ancestry regions of the most significant region, which reduced the number of significant local ancestry regions from 2127 to 484 independent regions (Additional File 3: Extended Data Table S1). The remaining local ancestry regions were removed from future analyses.

#### Ancestry-driven regions for metabolites

We next tested the ancestry driving the association in each independent local ancestry region. The most significant ancestry associations were attributed to Native American ancestry (Table 1). There was notable clustering of association regions by ancestry where multiple nearby independent regions across metabolites were attributed to the same ancestry (Fig. 2). On several chromosomes, there was a concordant association between the driving ancestry and the direction of the association, i.e., one ancestry was associated with increased metabolite abundance while the other was associated with reduced metabolite abundance (Additional File 2: Fig. S2). A large number of metabolites were associated with two local ancestry regions located on chromosomes 2 and 11. Of the metabolites, 97.4% of those associated with regions on chromosome 11 were lipids, and 84.6% of those associated with chromosome 2 were amino acids (Table 1). On chromosome 2, 91% of the associations were attributed to African local ancestry, with 89% of those associations being related to increased metabolite

abundance. Similarly, on chromosome 11, all but one of the 264 local ancestry region associations were driven by Native American ancestry, with 86% of these associations being related to low metabolite quantities (Table 1).

#### Fine-mapping results using GWAS variants on genomic regions overlapping with GWAS findings

Among 484 metabolite-genomic associations, 252 (51%) of the local ancestry regions overlapped with a significant variant reported in GWAS, and 232 without overlap were considered novel. We used GWAS summary statistics for metabolites in HCHS/SOL from a recent publication [8] to test if genome-wide significant SNVs explained our admixture mapping findings. Using a stepwise selection model implemented in GCTA-COJO, we first identified all independent associated SNVs within the boundaries of each local ancestry region. These analyses identified 361 SNVs in the 252 local ancestry regions, which were then used in admixture mapping conditional analyses.

Among these SNVs, 46 explained the admixture association in 42 local ancestry regions (Table 2), and two examples are shown in Fig. 3. The significance of a locus was explained in a further two local ancestry regions when conditioning on all COJO-selected SNVs in a region (Additional File 1: Table S4).

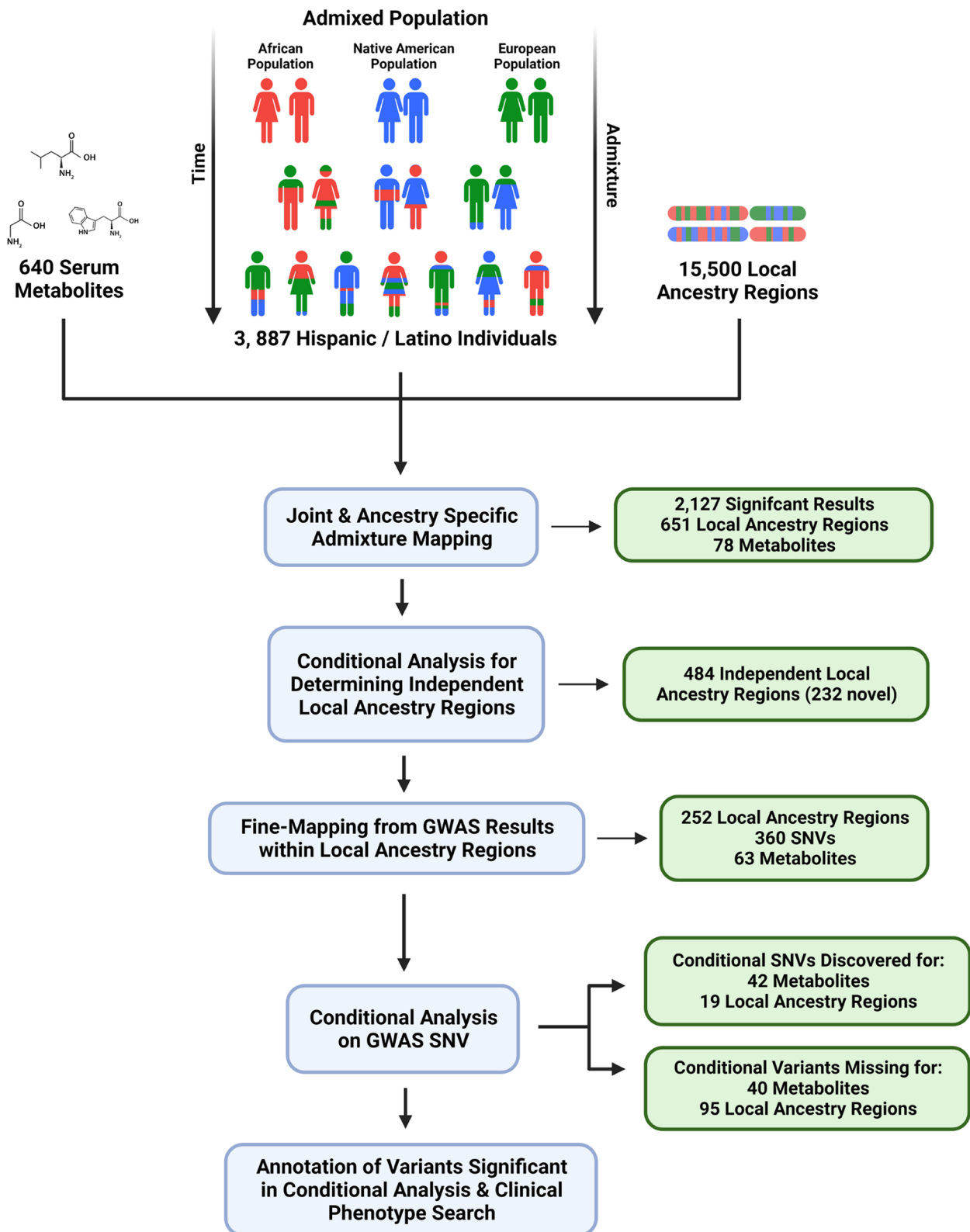
Interestingly, 31 significant local ancestry associations with metabolites were driven by only four local ancestry regions on chromosomes 2, 11, 15, and 16. Often, the same SNV explained the association with local ancestry for multiple metabolites (Additional File 1: Tables S5 to S8). With only one exception (*alliin* on chromosome 2), local ancestry associations explained by the same SNV in multiple metabolites had the same driving ancestry (Additional File 1: Table S6). For 208 metabolite and local ancestry region associations, the GWAS SNVs did not explain the admixture mapping association (Additional File 3: Extended Data Table S2), suggesting the presence of additional SNVs, likely rare and ancestry-enriched, contributing to the association in these regions.

#### Replication of admixture mapping findings

Among 64 metabolites available in the replication sample associated at 169 local ancestry regions, we identified significant associations for 116 of 190 novel metabolite-genomic region associations and 187 of 211 known metabolite-region associations based on a Bonferroni adjusted  $p$ -value  $< 1 \times 10^{-4}$  (Additional File 3: Extended

(See figure on next page.)

**Fig. 1** Study design and results. Hispanic/Latino populations are descendants from African, European, and Native American ancestral populations. Joint (all-ancestry) and ancestry-specific admixture mapping was performed using genetic, metabolic, and clinical data from 3887 Hispanic/Latino individuals. Conditional admixture mapping analysis determined individual local ancestry regions. These local ancestry regions were fine-mapped using GWAS results from a previous study. Annotation and clinical phenotypes were obtained for significant variants from conditional admixture mapping analysis



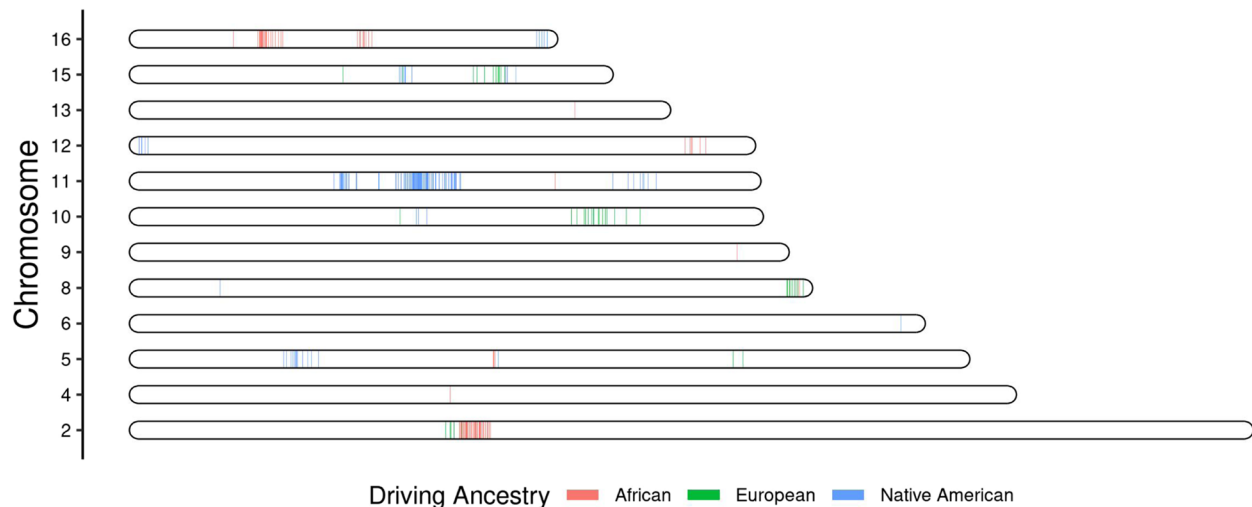
**Fig. 1** (See legend on previous page.)

**Table 1** All admixture mapping results by ancestry and super pathway in independent local ancestry regions

Chr	Metabolites <i>n</i>	Significant LA regions		Driving ancestry from ancestry-specific analysis			Super pathway		
		Novel <i>n</i> (%)	Total <i>n</i>	NAM <i>n</i> (%)	AFR <i>n</i> (%)	EUR <i>n</i> (%)	Lipids <i>n</i> (%)	Amino acids <i>n</i> (%)	Other <i>n</i> (%)
2	13	41 (48)	85	-	78 (92)	7 (8)	1 (8)	11 (85)	1 (7)
4	1	-	1	-	1 (100)	-	1 (100)	-	-
5	3	14 (74)	19	14 (74)	3 (16)	2 (11)	-	2 (67)	1 (33)
6	1	-	1	1 (100)	-	-	-	-	1 (100)
8	4	8 (53)	15	1 (7)	5 (33)	9 (60)	-	4 (100)	-
9	1	-	1	-	1 (100)	-	-	-	1 (100)
10	2	13 (65)	20	3 (15)	-	17 (85)	1 (50)	-	1 (50)
11	38	114 (70)	164	263 (99.6)	1 (0.4)	-	37 (97)	1 (3)	-
12	4	7 (58)	12	6 (50)	6 (50)	-	2 (50)	2 (50)	-
13	1	-	1	-	1 (100)	-	-	-	1 (100)
15	8	15 (58)	26	14 (54)	-	12 (46)	6 (75)	2 (25)	-
16	4	20 (51)	39	6 (15)	33 (85)	-	-	3 (75)	1 (25)

Novel regions are local ancestry regions containing no significant SNVs from GWAS. The total number of significant local ancestry regions allows for duplicate significant regions across metabolites for the corresponding chromosome. Ancestry-specific analysis was performed on all metabolites significant in joint (all-ancestries) admixture mapping. The driving ancestry was the ancestry group with the smallest *p*-value in ancestry-specific testing. Metabolic super pathways represent overall metabolite function in the body

Chr chromosome, LA local ancestry, NAM Native American driving ancestry, AFR African driving ancestry, EUR European driving ancestry



**Fig. 2** Location of associated independent local ancestry regions. Each vertical line represents an independent local ancestry region, colored by driving ancestry (the ancestry with the lowest *p*-value in ancestry-specific analyses). Regions located in relatively close proximity tend to have the same driving ancestry

Data Table S3). All metabolite-local ancestry associations had the same direction of effects for the ancestry-driving associations.

#### Disease relevance

We selected three local ancestry regions for more in-depth analysis based on the large number of metabolites associated and their relevance to disease traits (chromosomes 2, 11, and 15). Two genomic regions

(chromosomes 11 and 15) were associated with lipids metabolites driven by Native American ancestry. The chromosome 11 locus included twelve lipid metabolites and seven fine-mapped SNVs that were located at intronic regions of *TMEM258*, *FADS1*, and *FADS2*, and an intergenic SNV that was located near *FTH1* (Additional File 1: Table S5). The chromosome 15 had four associated lipids with a fine-mapped SNV located upstream to *LIPC*. Two lipid metabolites,

**Table 2** COJO genetic variants that explain association between local ancestry regions and metabolites

Metabolite	Chr	LA ID	COJO rsID	Gene	Gene function	Driving ancestry (MAF)	AM joint P-value	Conditional joint P-value
Alliin	2	1579	rs10189885	<i>ALMS1P1</i>	Intronic	EUR (—)	$2.77 \times 10^{-10}$	$9.00 \times 10^{-2}$
N-acetylkynurenine (2)	2	1579	rs10189885 <sup>a</sup>	<i>ALMS1P1</i>	Intronic	AFR (—)	$4.13 \times 10^{-12}$	$2.87 \times 10^{-1}$
N2-acetylysine	2	1579	rs12620091	<i>ALMS1P1</i>	Intronic <sup>b</sup>	AFR (0.17)	$2.87 \times 10^{-17}$	$1.19 \times 10^{-3}$
N-acetylglutamine	2	1579	rs13431529	<i>ALMS1P1</i>	Intronic <sup>b</sup>	AFR (0.6)	$6.05 \times 10^{-12}$	$6.29 \times 10^{-1}$
N2-acetylysine	2	1579	rs1881244	<i>ALMS1</i>	Intronic	AFR (0.85)	$2.87 \times 10^{-17}$	$3.47 \times 10^{-3}$
N-acetylarginine	2	1579	rs1881244	<i>ALMS1</i>	Intronic	AFR (0.85)	$4.85 \times 10^{-25}$	$3.75 \times 10^{-3}$
N-acetylasparagine	2	1579	rs1881244	<i>ALMS1</i>	Intronic	AFR (0.85)	$1.74 \times 10^{-22}$	$7.78 \times 10^{-4}$
N-acetylglutamine	2	1579	rs1881244	<i>ALMS1</i>	Intronic	AFR (0.85)	$6.05 \times 10^{-12}$	$1.42 \times 10^{-2}$
N-acetylleucine	2	1579	rs1881244	<i>ALMS1</i>	Intronic	AFR (0.85)	$2.07 \times 10^{-19}$	$1.52 \times 10^{-4}$
N-acetyltyrosine	2	1579	rs1881244	<i>ALMS1</i>	Intronic	AFR (0.85)	$1.73 \times 10^{-14}$	$1.04 \times 10^{-1}$
N-delta-acetylornithine	2	1579	rs28525015	<i>ALMS1P1</i>	Intronic <sup>b</sup>	AFR (0.61)	$6.15 \times 10^{-24}$	$3.17 \times 10^{-4}$
N-acetylleucine	2	1579	rs28879089	<i>ALMS1</i>	Intronic	AFR (—)	$2.07 \times 10^{-19}$	$6.41 \times 10^{-3}$
N-acetyltyrosine	2	1579	rs6546854	<i>ALMS1</i>	Intronic	AFR (0.62)	$1.73 \times 10^{-14}$	$1.86 \times 10^{-1}$
2-Aminooctanoic acid	2	1581	rs72903325	<i>ACTG2</i>	Intronic	AFR (0.35)	$6.26 \times 10^{-11}$	$1.73 \times 10^{-4}$
Androstenediol (3alpha, 17alpha) monosulfate (3)	4	3592	rs138040976	<i>GNRHR</i>	Intronic	AFR (—)	$4.07 \times 10^{-11}$	$3.10 \times 10^{-4}$
3-aminoisobutyrate	5	4386	rs191679549	<i>AGXT2</i>	Intronic	NAM (—)	$7.11 \times 10^{-29}$	$6.87 \times 10^{-3}$
Betaine	5	4546	rs7736027	<i>DMGDH</i>	Intronic	AFR (0.66)	$2.16 \times 10^{-11}$	$2.12 \times 10^{-2}$
N-acetylglucosaminylasparagine	6	5848	rs2346120 <sup>a</sup>	<i>UNC93A</i>	Intronic	NAM (—)	$3.20 \times 10^{-9}$	$1.67 \times 10^{-2}$
S-1-pyrroline-5-carboxylate	8	7373	rs2242090	<i>PYCR3</i>	Exonic <sup>c</sup>	EUR (0.52)	$1.43 \times 10^{-24}$	$1.01 \times 10^{-1}$
6-oxopiperidine-2-carboxylate	8	7376	rs11993782	<i>PLEC</i>	Intronic	AFR (0.83)	$2.24 \times 10^{-10}$	$9.18 \times 10^{-2}$
2'-O-methyluridine	9	7977	rs56023505	<i>PHYHD1</i>	UTR5	AFR (0.44)	$3.35 \times 10^{-10}$	$4.56 \times 10^{-3}$
Carnitine	10	8412	rs12415764	<i>ARID5B</i>	Intergenic	NAM (0.24)	$2.32 \times 10^{-11}$	$5.37 \times 10^{-5}$
N-methylpipercolate	10	8587	rs2147896	<i>PYROXD2</i>	Exonic <sup>c</sup>	EUR (0.35)	$2.02 \times 10^{-46}$	$1.39 \times 10^{-3}$
MG 20:4	11	9120	rs102274	<i>TMEM258</i>	Intronic	NAM (0.59)	$2.15 \times 10^{-37}$	$5.72 \times 10^{-3}$
PC 16:0/18:2	11	9120	rs11320420	<i>MYRF</i>	Intronic	NAM (—)	$1.53 \times 10^{-15}$	$6.34 \times 10^{-2}$
Stearidonic acid	11	9120	rs174556	<i>FADS1</i>	Intronic	NAM (0.57)	$1.30 \times 10^{-19}$	$3.60 \times 10^{-4}$
1-arachidonoyl-GPC (20:4n6) <sup>b</sup>	11	9120	rs174562	<i>FADS2</i>	Intronic	NAM (0.59)	$7.02 \times 10^{-76}$	$1.02 \times 10^{-4}$
PC 16:0/20:4	11	9120	rs174562	<i>FADS2</i>	Intronic	NAM (0.59)	$1.04 \times 10^{-69}$	$1.72 \times 10^{-4}$
Arachidonoylcholine	11	9120	rs174567	<i>FADS2</i>	Intronic	NAM (0.61)	$8.82 \times 10^{-28}$	$9.13 \times 10^{-3}$
PC 18:2/18:2	11	9120	rs3834458	<i>FADS2</i>	Intronic	NAM (0.58)	$5.97 \times 10^{-26}$	$4.87 \times 10^{-3}$
LPC 18:2/0:0	11	9120	rs3834458	<i>FADS2</i>	Intronic	NAM (0.58)	$8.16 \times 10^{-11}$	$1.23 \times 10^{-1}$
Sphingomyelin (d18:1/20:2, d18:2/20:1, d16:1/22:2) <sup>b</sup>	11	9120	rs3834458	<i>FADS2</i>	Intronic	NAM (0.58)	$2.24 \times 10^{-10}$	$1.61 \times 10^{-2}$
PE 16:0/18:2	11	9120	rs5792235	<i>FADS2</i>	Intronic	NAM (0.59)	$5.40 \times 10^{-29}$	$4.02 \times 10^{-2}$
PC 18:0/18:2	11	9120	rs5792235	<i>FADS2</i>	Intronic	NAM (0.59)	$3.80 \times 10^{-11}$	$1.43 \times 10^{-1}$
PE 18:0/18:2	11	9120	rs5792235	<i>FADS2</i>	Intronic	NAM (0.59)	$2.32 \times 10^{-22}$	$7.67 \times 10^{-2}$
LPC P-16:0 or LPC O-16:1	11	9121	rs11230873	<i>FTH1</i>	Intergenic	NAM (0.26)	$1.60 \times 10^{-10}$	$9.49 \times 10^{-5}$
MG 20:4	11	9149	rs142021492	—	—	NAM (—)	$2.93 \times 10^{-9}$	$2.21 \times 10^{-3}$
3beta-hydroxy-5-cholestenoate	11	9317	chr11:110244360	<i>RDX</i>	—	NAM (—)	$6.11 \times 10^{-26}$	$4.75 \times 10^{-2}$
Ethylmalonate	12	10132	chr12:121176083	<i>ACADS</i>	Exonic <sup>c</sup>	AFR (—)	$5.72 \times 10^{-13}$	$1.41 \times 10^{-1}$
Ethylmalonate	12	10132	rs73415734	—	—	AFR (—)	$5.72 \times 10^{-13}$	$1.82 \times 10^{-2}$
LPE 16:0/0:0	15	11487	rs1077834	<i>LIPC</i>	Intronic <sup>c</sup>	NAM (0.43)	$2.32 \times 10^{-14}$	$2.90 \times 10^{-3}$
LPE 18:0/0:0	15	11487	rs1077834	<i>LIPC</i>	Intronic <sup>c</sup>	NAM (0.43)	$1.85 \times 10^{-11}$	$4.26 \times 10^{-2}$
PE 16:0/20:4	15	11487	rs2070895	<i>LIPC</i>	Intronic	NAM (—)	$5.53 \times 10^{-18}$	$3.20 \times 10^{-2}$
PE 18:0/20:4	15	11487	rs2070895	<i>LIPC</i>	Intronic	NAM (—)	$5.10 \times 10^{-13}$	$1.98 \times 10^{-1}$
Cys-gly, oxidized	16	12348	chr16:89909429	<i>SPIRE</i>	Intronic	NAM (—)	$1.46 \times 10^{-29}$	$4.68 \times 10^{-2}$
Cysteinylglycine	16	12348	rs62068366	<i>SPATA2L</i>	Intronic	NAM (0.28)	$8.45 \times 10^{-10}$	$3.14 \times 10^{-2}$

COJO conditional and joint association analysis, Chr chromosome, LA ID local ancestry region ID, MAF minor allele frequency, AFR African, NAM Native American, EUR European, AM admixture mapping

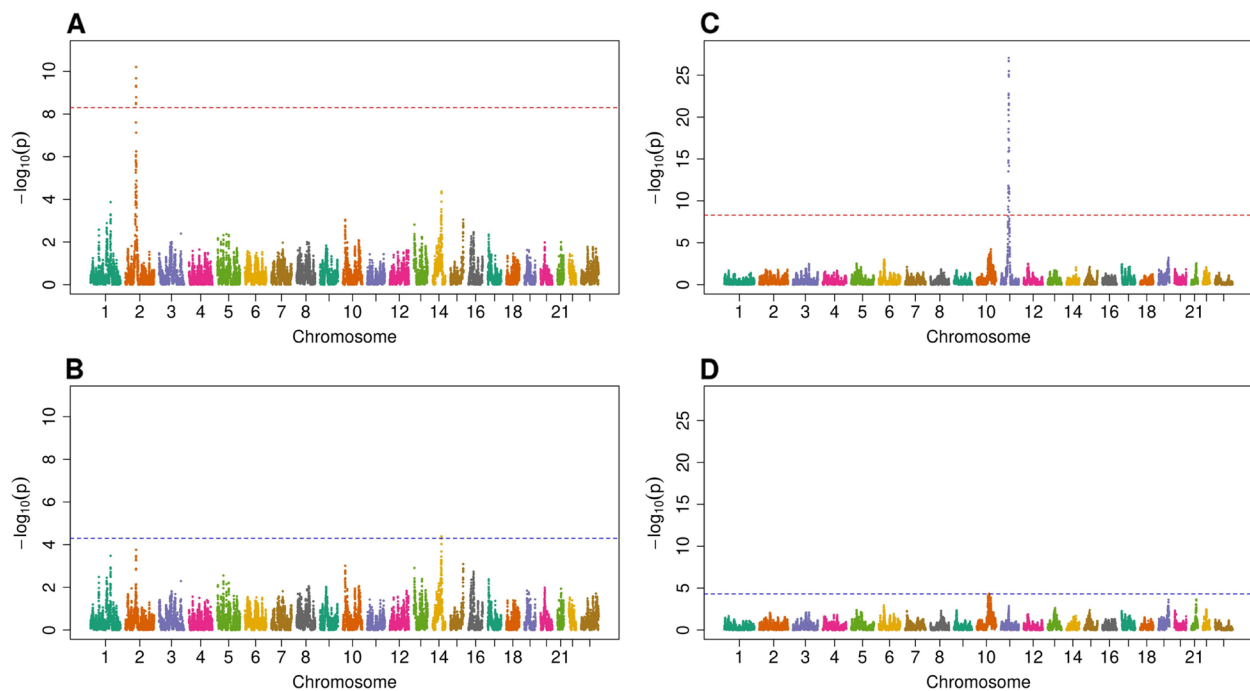
<sup>a</sup> Genetic variant not significant in GWAS

<sup>b</sup> ncRNA

<sup>c</sup> Missense

<sup>c</sup> Upstream





**Fig. 3** COJO SNVs account for admixture mapping signals. The horizontal red line in **A** and **C** represents the significance level of  $5.04 \times 10^{-9}$  for all-ancestry admixture mapping. The horizontal blue line in **B** and **D** represents the significance level of  $5 \times 10^{-5}$  for conditional admixture mapping. **A** Admixture mapping results for 2-aminooctanoic acid. **B** The SNV rs72903325 accounts for the admixture mapping signal in local ancestry region 1581 on chromosome 2 for 2-aminooctanoic acid. **C** Admixture mapping results for arachidonoylcholine. **D** The SNV rs174567 accounts for the admixture mapping signal in local ancestry region 9120 on chromosome 11 for arachidonoylcholine

1-palmitoyl-2-linoleoyl-GPE (PE) 16:0/18:2 and phosphatidylethanolamine (PE) 18:0/18:2, were associated with local ancestry regions on both chromosomes 11 and 15. *FADS1/FADS2* are involved in the desaturation of fatty acids to generate long-chain polyunsaturated fatty acids (LC-PUFAs). *LIPC* encodes the hepatic triglyceride lipase. Both *FADS1* and *LIPC* code for major enzymes of the LC-PUFA metabolism and have been previously associated with circulating lipids and metabolites [5, 37]. Prior studies have shown evidence for Native American haplotypes at the *FADS1* locus related to reduced PUFA metabolism [38–40], and we additionally identified novel Native American ancestry metabolite associations for SNVs at the *FADS2* and *LIPC*. The ancestral haplogroup of the *FADS* genes, which is associated with a deficient biosynthesis of the biologically active form of PUFA (LC-PUFA), is nearly fixed in Native American populations and has replicated signals of positive selection possibly related to dietary conditions. Within the *FADS* cluster, rs102274, a SNV intronic to *TMEM258* fine-mapped in our analysis (Additional File 1: Table S5), has been considered a causal SNV for LC-PUFA biosynthesis [41]. Signatures of selection in Native American populations for other genes related to diet have also been found in modern populations of Latin America [42].

At the chromosome 2 region, eight N-acetylated amino acids were associated with African local ancestry. Conditional analysis identified seven GWAS SNVs, intronic to *ALMS1/ALMS1P*, that explained the admixture mapping results in the region (Additional File 1: Table S6). Prior studies using blood transcriptome data support associations of N-acetylated amino acid metabolites with *ALMS1* [43]. This gene has been associated to chronic kidney disease [44], a disease with a high burden in individuals of West African descent in the USA.

Additional novel associations of disease relevance are homoarginine at chromosome 15 and carnitine at chromosome 10, for which European ancestry was associated with low circulation levels of these metabolites. Low homoarginine levels are related to endothelial cell dysfunction and cardiovascular disease [45]. The identified association is outside the *GATM* locus at chromosome 15, previously associated with this metabolite in GWAS [46, 47]. Carnitine levels are related to metabolic diseases and mitochondrial function [48].

## Discussion

This study supports the presence of metabolic differences across populations based on ancestry admixture, which are likely driven by genetic ancestral diversity and

are potential adaptations to environmental stressors. We identified 2127 significant metabolite-local ancestry associations for 78 metabolites and 651 local ancestry regions within twelve chromosomes. Of the 484 independent local ancestry regions, 232 were novel associations, and several of them were replicated in an independent sample of Hispanics/Latinos. Several associated regions were driven by Native American ancestry, for which genetic variation is less known. Native American is a population less often included in genetic studies. Our study supports the approach of leveraging genetic ancestry to map genes and putative causal variants, as well as to better understand differences in metabolic processes driven by ancestry in admixed populations. The metabolic pathways identified in our study are related to a variety of physiological functions that may be altered in complex diseases or involved in response to environmental stressors, such as those related to diet restrictions and exposure to pathogens. Therefore, our findings have relevance to health and disease.

We identified a wide range of lipid-based metabolites either negatively or positively associated with ancestry-specific genomic regions, predominately on chromosomes 11 and 15, including some newly identified to be associated to these regions. A notable feature of the findings was that many of the metabolites contain LC-PUFAs. For instance, the LC-PUFA arachidonic acid (20:4) of the n-6 biosynthetic pathway (esterified to the glycerol-based backbone of phosphatidylcholines, phosphatidylethanolamines, diacylglycerols, and lysophospholipids) displayed a reduced abundance in individuals with Native American ancestry. A reduction in arachidonic acid in this population could have several biological implications. Arachidonic acid is central in the initiation of inflammatory pathways as it undergoes liberation from phospholipids and serves as a substrate for enzymes, such as cyclooxygenases and lipoxygenases, that generate prostaglandins, leukotrienes, and the newer class of pro-resolution lipoxins [49]. A reduction in these metabolites could suggest an increased production of pro-inflammatory molecules. Some literature suggests differences in inflammatory status in adults and children may be driven by genetic and environmental factors [50]. Future mechanistic studies that tease apart each of these metabolites in primary culture models may shed light on their role in controlling the inflammatory response in the context of specific populations.

Prior literature supports that genetic Native American ancestry haplotypes contribute to a fatty acid desaturase SNV that is associated with low levels of LC-PUFAs of the n-3 biosynthetic pathway [51]. Interestingly, our findings also identified long-chain n-3 PUFAs eicosapentaenoic acid (EPA, 20:5), docosapentaenoic acid,

and docosahexaenoic acid (DHA, 22:6) as having a low abundance association with Native American ancestry on chromosome 11. A reduction in EPA and DHA levels is of biological significance as these fatty acids control downstream metabolites (such as resolvins, protectins, and maresins) that drive the resolution of inflammation [52]. Thus, a reduction in these fatty acids may contribute to impaired resolution of inflammation. Finally, it is important to point out that our data revealed changes in PUFA metabolism, notably arachidonic acid, across a wide range of lipid pools, which are likely impacting biological processes outside of inflammation.

In another region, the *ALMS1* locus was associated with an increased abundance of N-acetylated amino acids in African ancestry-derived regions, including some newly associated metabolites. Prior studies have shown evidence of population-specific signals of adaptation in Niger-Congo West Africans at this locus [53]. The mechanisms relating the identified metabolic changes and gene to disease are not fully understood. *ALMS1* has been consistently associated with chronic kidney disease in population studies [44, 54] and in a monogenic disorder (Alstrom syndrome, OMIM #203,800) [33, 55], and obesity and insulin resistance in experimental models [56]. Chronic kidney disease is more common in individuals of African descent in the U.S. Prior studies support African-derived SNVs at another gene (*APOLI*), related to resistance to infectious diseases in Africa, conferring risk to chronic kidney disease in Hispanics/Latinos with West African admixture. A study has shown that *ALMS1* protein is involved in the regulation of kidney sodium transport and blood pressure, through interaction with the  $\text{Na}^+/\text{K}^+/\text{2Cl}^-$ -cotransporter (NKCC2) in the nephron loop of Henle [57]. NKCC2 sodium reabsorption is increased in Blacks [58]. These findings provide some mechanistic pathways for the relation among African ancestry at the *ALMS1* locus and chronic kidney disease, but the association with N-acetylated amino acids will require further studies. A nearby gene in this locus, *NAT8*, related to N-acetyltransferase activity, has been associated with N-acetylated amino acids abundance in blood and urine, but gene expression studies showed stronger associations of metabolite-associated SNVs with the *ALMS1* gene [43]. In addition, a study of admixed Brazilians also identified *ALMS1* but not *NAT8* as most significantly associated gene for these metabolites at the region [59].

Several other ancestry-driven genomic regions associated with metabolites were not previously reported, including novel associations for homoarginine and carnitine that have implications for cardiometabolic diseases. These findings support loci discovery and complementary information to GWAS obtained when testing

ancestry-driven genomic regions for loci discovery. For some regions previously reported, the GWAS SNV did not explain the admixture mapping associations, supporting the presence of additional causal SNV(s) in these regions. Other approaches that leverage local ancestry for loci discovery and fine-mapping including BMIX [60], which tests association at single markers while stratifying by local ancestry patterns at each interval, could be extended to test three-way ancestry in studies of Hispanics/Latinos and for mixed models. New strategies for fine mapping these regions for potential causal genetic variants are needed that may include integration of sequencing for rare and low-frequency variants and gene expression data generated in admixed populations to better query causal variants that are ancestry enriched. This effort is particularly important to capture Native American ancestry-enriched genetic variants, given the large Native American ancestry proportions in Hispanics/Latinos.

## Conclusions

We identified several ancestry-enriched genomic regions associated with metabolites including Native American-driven regions at chromosomes 11 and 15 related to PUFAs that may contribute to metabolic and inflammatory disease in individuals with Native American ancestry components, and an African-driven genomic region related to N-acetylated amino acid compounds previously identified in chronic kidney disease. These findings support ancestry differences in metabolite regulation of lipid PUFAs and N-acetylated amino acids, which may have implications for common diseases in populations.

## Abbreviations

eGFR	Estimated glomerular filtration rate
HCHS/SOL	Hispanic Community Health Study/Study of Latinos
HMDB	Human Metabolome Database
GWAS	Genome-wide association study
LC-PUFA	Long-chain polyunsaturated fatty acid
SNV	Single nucleotide variant

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-023-01209-z>.

**Additional file 1: Table S1.** Descriptive statistics of 3,887 HCHS / SOL participants at visit 1. **Table S2.** Significant metabolites from admixture mapping. **Table S3.** All-ancestries and ancestry-specific admixture mapping results. **Table S4.** Regions whose significance was explained by adding all COJO SNVs to model.

**Additional file 2: Fig. S1.** Global ancestry proportions of African, European, and Native American ancestries for participants based on their country of origin: Mainland (Mexico, Central and South America) or Caribbean (Cuba, Dominican Republic, and Puerto Rico). Note that participants from Mainland had a higher proportion of Native American ancestry, while those from Caribbean had a higher proportion of African ancestry. **Fig. S2.**

Volcano plots showing relationship between the direction of association and driving ancestry in the three chromosomes with the largest numbers of associated metabolites. The driving ancestry was the ancestry with the smallest p-value in ancestry-specific testing. In chromosome 2, most of the associations with African ancestry were positive. In chromosome 11, most of the associations with Native American ancestry were negative.

**Additional file 3: Extended Data Table S1.** Significant independent local ancestry regions from admixture mapping. **Extended Data Table S2.** Genetic variants that do not explain association between local ancestry regions and metabolites. **Extended Data Table S3.** Replication of significant independent local ancestry regions from admixture mapping for 64 metabolites available in the replication dataset.

## Authors' contributions

NF conceived and designed the study and obtained the funding. KMR, ARVRH, BML and NK applied statistical methods or produced the figures. TS, YZ, JC provided support for statistical analysis and code for HCHS/SOL complex design. EB, BY generated the metabolite data. NF, KMR, ARVRH, TS, SRB, SRS wrote the paper. All authors read and approved the final manuscript.

## Funding

HCHS/SOL is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN2682013000011 / N01-HC-65233), University of Miami (HHSN2682013000041 / N01-HC-65234), Albert Einstein College of Medicine (HHSN2682013000021 / N01-HC-65235), University of Illinois at Chicago (HHSN2682013000031 / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN2682013000051 / N01-HC-65237). The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements. This project was supported by NIH R01 DK117445 and R01 MD012765 (to NF). Support for metabolomics data was graciously provided by the JLH Foundation (Houston, Texas).

## Availability of data and materials

This research was conducted using genotype and phenotype data from HCHS/SOL, which is publicly available through the dbGap (access phs000810.v1.p1). Metabolomics data is available upon request from Dr. Eric Boerwinkle (Eric.Boerwinkle@uth.tmc.edu). The summary results for discovery and replication are included in Additional File 3: Extended Data.

## Declarations

### Ethics approval and consent to participate

The HCHS/SOL study was approved by the institutional review boards at each field center and the coordinating center, and all subjects included in this study provided written informed consent for the use of phenotype and genetic data for research. The research conforms to the principles of the Declaration of Helsinki.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. <sup>2</sup>Department of Epidemiology, University of North Carolina, 123 W Franklin St, Suite 401, NC NC 27516 Chapel Hill, USA. <sup>3</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA. <sup>4</sup>Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. <sup>5</sup>Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>6</sup>Department of Epidemiology and Population Health, Albert Einstein

College of Medicine, Bronx, NY, USA. <sup>7</sup>Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. <sup>8</sup>Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, USA. <sup>9</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>10</sup>Department of Nutrition, University of North Carolina, Chapel Hill, NC, USA. <sup>11</sup>Departments of Medicine and Biostatistics, Harvard University, Boston, MA, USA.

Received: 22 November 2022 Accepted: 10 July 2023

Published online: 17 July 2023

## References

- Wishart DS. Metabolomics for investigating physiological and pathophysiological processes. *Physiol Rev.* 2019;99(4):1819–75.
- Suhre K, Gieger C. Genetic variation in metabolic phenotypes: study designs and applications. *Nat Rev Genet.* 2012;13(11):759–69.
- Newgard CB. Metabolomics and metabolic diseases: where do we stand? *Cell Metab.* 2017;25(1):43–56.
- Pietzner M, Stewart ID, Raffler J, Khaw KT, Michelotti GA, Kastenmuller G, et al. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat Med.* 2021;27(3):471–9.
- Cadby G, Giles C, Melton PE, Huynh K, Mellett NA, Duong T, et al. Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. *Nat Commun.* 2022;13(1):3124.
- Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet.* 2014;46(6):543–50.
- Long T, Hicks M, Yu HC, Biggs WH, Kirkness EF, Menni C, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet.* 2017;49(4):568–78.
- Feofanova EV, Chen H, Dai Y, Jia P, Grove ML, Morrison AC, et al. A genome-wide association study discovers 46 loci of the human metabolome in the hispanic community health study/study of Latinos. *Am J Hum Genet.* 2020;107(5):849–63.
- Bomba L, Walter K, Guo Q, Surendran P, Kundu K, Nongmaithem S, et al. Whole-exome sequencing identifies rare genetic variants associated with human plasma metabolites. *Am J Hum Genet.* 2022;109(6):1038–54.
- Spear ML, Diaz-Papkovich A, Ziv E, Yracheta JM, Gravel S, Torgerson DG, et al. Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. *Elife.* 2020;9:e56029.
- Dai CL, Vazifeh MM, Yeang CH, Tachet R, Wells RS, Vilar MG, et al. Population Histories of the United States Revealed through Fine-Scale Migration and Haplotype Analysis. *Am J Hum Genet.* 2020;106(3):371–88.
- Rees JS, Castellano S, Andres AM. The genomics of human local adaptation. *Trends Genet.* 2020;36(6):415–28.
- Fan S, Hansen ME, Lo Y, Tishkoff SA. Going global by adapting local: a review of recent human adaptation. *Science.* 2016;354(6308):54–9.
- Consortium STD, Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature.* 2014;506(7486):97–101.
- Kramer HJ, Stilp AM, Laurie CC, Reiner AP, Lash J, Daviglus ML, et al. African ancestry-specific alleles and kidney disease risk in Hispanics/Latinos. *J Am Soc Nephrol.* 2017;28(3):915–22.
- Sofer T, Baier LJ, Browning SR, Thornton TA, Talavera GA, Wassertheil-Smoller S, et al. Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits. *PLoS ONE.* 2017;12(11):e0188400.
- Brown LA, Sofer T, Stilp AM, Baier LJ, Kramer HJ, Masindova I, et al. Admixture mapping identifies an Amerindian ancestry locus associated with albuminuria in Hispanics in the United States. *J Am Soc Nephrol.* 2017;28(7):2211–20.
- Horimoto AVR, Xue D, Cai J, Lash JP, Daviglus ML, Franceschini N, et al. Genome-wide admixture mapping of estimated glomerular filtration rate and chronic kidney disease identifies European and African ancestry-of-origin loci in Hispanic and Latino individuals in the United States. *J Am Soc Nephrol.* 2022;33(1):77–87.
- Burkart KM, Sofer T, London SJ, Manichaikul A, Hartwig FP, Yan Q, et al. A genome-wide association study in Hispanics/Latinos identifies novel signals for lung function. *The Hispanic Community Health Study/Study of Latinos. Am J Respir Crit Care Med.* 2018;198(2):208–19.
- Wang H, Cade BE, Sofer T, Sands SA, Chen H, Browning SR, et al. Admixture mapping identifies novel loci for obstructive sleep apnea in Hispanic/Latino Americans. *Hum Mol Genet.* 2019;28(4):675–87.
- Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8):629–41.
- Lavage LM, Kalsbeek WD, Sorlie PD, Aviles-Santa LM, Kaplan RC, Barnhart J, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.* 2010;20(8):642–9.
- Evans AM, Bridgewater BR, Liu Q, Mitchell MW, Robinson RJ, Dai H, et al. High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling. *Metabolomics.* 2014;4:132.
- Liebisch G, Fahy E, Aoki J, Dennis EA, Durand T, Ejsing CS, et al. Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *J Lipid Res.* 2020;61(12):1539–55.
- Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet.* 2016;98(1):165–84.
- Laurie CC, Doherty KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol.* 2010;34(6):591–602.
- Browning SR, Grinde K, Plantinga A, Gogarten SM, Stilp AM, Kaplan RC, et al. Local ancestry inference in a large US-based Hispanic/Latino study: Hispanic community health study/study of Latinos (HCHS/SOL). G3 (Bethesda). 2016;6(6):1525–34.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 2008;319(5866):1100–4.
- Consortium GP, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56–65.
- Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 2007;81(5):1084–97.
- Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet.* 2013;93(2):278–88.
- Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics.* 2019;35(24):5346–8.
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Mangano C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005–12.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
- Breeze CE, Haugen E, Reynolds A, Teschendorff A, van Dongen J, Lan Q, et al. Integrative analysis of 3604 GWAS reveals multiple novel cell type-specific regulatory associations. *Genome Biol.* 2022;23(1):13.
- Gieger C, Geistlinger L, Altmaier E, Hrabec de Angelis M, Kronenberg F, Meisinger T, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* 2008;4(11):e1000282.
- Chilton FH, Manichaikul A, Yang C, O'Connor TD, Johnstone LM, Blomquist S, et al. Interpreting Clinical Trials With Omega-3 Supplements in the Context of Ancestry and FADS Genetic Variation. *Front Nutr.* 2021;8:808054.
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature.* 2015;528(7583):499–503.
- Amorim CE, Nunes K, Meyer D, Comas D, Bortolini MC, Salzano FM, et al. Genetic signature of natural selection in first Americans. *Proc Natl Acad Sci U S A.* 2017;114(9):2195–9.
- Harris DN, Ruczinski I, Yanek LR, Becker LC, Becker DM, Guio H, et al. Evolution of Hominin Polyunsaturated Fatty Acid Metabolism: From Africa to the New World. *Genome Biol Evol.* 2019;11(5):1417–30.

42. Mendoza-Revilla J, Chacon-Duque JC, Fuentes-Guajardo M, Ormond L, Wang K, Hurtado M, et al. Disentangling Signatures of Selection Before and After European Colonization in Latin Americans. *Mol Biol Evol.* 2022;39(4):msac076.
43. Sonmez Flitman R, Khalili B, Kotalik Z, Rueedi R, Brummer A, Bergmann S. Untargeted Metabolome- and Transcriptome-Wide Association Study Suggests Causal Genes Modulating Metabolite Concentrations in Urine. *J Proteome Res.* 2021;20(11):5103–14.
44. Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun.* 2019;10(1):29.
45. Karetnikova ES, Jarzebska N, Markov AG, Weiss N, Lentz SR, Rodionov RN. Is homoarginine a protective cardiovascular risk factor? *Arterioscler Thromb Vasc Biol.* 2019;39(5):869–75.
46. Tahir UA, Katz DH, Avila-Pachecho J, Bick AG, Pampana A, Robbins JM, et al. Whole genome association study of the plasma metabolome identifies metabolites linked to cardiometabolic disease in black individuals. *Nat Commun.* 2022;13(1):4923.
47. Kleber ME, Seppala I, Pilz S, Hoffmann MM, Tomaschitz A, Oksala N, et al. Genome-wide association study identifies 3 genomic loci significantly associated with serum levels of homoarginine: the AtheroRemo Consortium. *Circ Cardiovasc Genet.* 2013;6(5):505–13.
48. McCann MR, George De la Rosa MV, Rosania GR, Stringer KA. L-carnitine and acylcarnitines: mitochondrial biomarkers for precision medicine. *Metabolites.* 2021;11(1):51.
49. Serhan CN, Levy BD. Resolvins in inflammation: emergence of the pro-resolving superfamily of mediators. *J Clin Invest.* 2018;128(7):2657–69.
50. Alderete TL, Toledo-Corral CM, Goran MI. Metabolic basis of ethnic differences in diabetes risk in overweight and obese youth. *Curr Diab Rep.* 2014;14(2):455.
51. Yang C, Hallmark B, Chai JC, O'Connor TD, Reynolds LM, Wood AC, et al. Impact of Amerind ancestry and FADS genetic variation on omega-3 deficiency and cardiometabolic traits in Hispanic populations. *Commun Biol.* 2021;4(1):918.
52. Dyall SC, Balas L, Bazan NG, Brenna JT, Chiang N, da Costa SF, et al. Polyunsaturated fatty acids and fatty acid-derived lipid mediators: Recent advances in the understanding of their biosynthesis, structures, and functions. *Prog Lipid Res.* 2022;86:101165.
53. Scheinfeldt LB, Soi S, Lambert C, Ko WY, Coulibaly A, Ranciaro A, et al. Genomic evidence for shared common ancestry of East African hunting-gathering populations and insights into local adaptation. *Proc Natl Acad Sci U S A.* 2019;116(10):4166–75.
54. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2017;45(D1):D896–901.
55. Choudhury AR, Munonye I, Sanu KP, Islam N, Gadaga C. A review of Alstrom syndrome: a rare monogenic ciliopathy. *Intractable Rare Dis Res.* 2021;10(4):257–62.
56. Nesmith JE, Hostalley TL, Leitch CC, Matern MS, Sethna S, McFarland R, et al. Genomic knockout of *alms1* in zebrafish recapitulates Alstrom syndrome and provides insight into metabolic phenotypes. *Hum Mol Genet.* 2019;28(13):2212–23.
57. Jaykumar AB, Caceres PS, King-Medina KN, Liao TD, Datta I, Maskey D, et al. Role of Alstrom syndrome 1 in the regulation of blood pressure and renal function. *JCI Insight.* 2018;3(21):e95076.
58. Tu W, Pratt JH. A consideration of genetic mechanisms behind the development of hypertension in blacks. *Curr Hypertens Rep.* 2013;15(2):108–13.
59. Montoliu I, Genick U, Ledda M, Collino S, Martin FP, le Coutre J, et al. Current status on genome-metabolome-wide associations: an opportunity in nutrition research. *Genes Nutr.* 2013;8(1):19–27.
60. Shriner D, Adeyemo A, Rotimi CN. Joint ancestry and association testing in admixed individuals. *PLoS Comput Biol.* 2011;7(12):e1002325.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

