Title

Genome-Wide Association Study Identifies Pharmacogenomic Variants Associated With Metformin Glycemic Response in African American Patients With Type 2 Diabetes.

Journal

Diabetes Care, 47(2)

Authors

Wu, Baojun
Yee, Sook Wah
Xiao, Shujie
et al.

Publication Date

2024-02-01

DOI

10.2337/dc22-2494

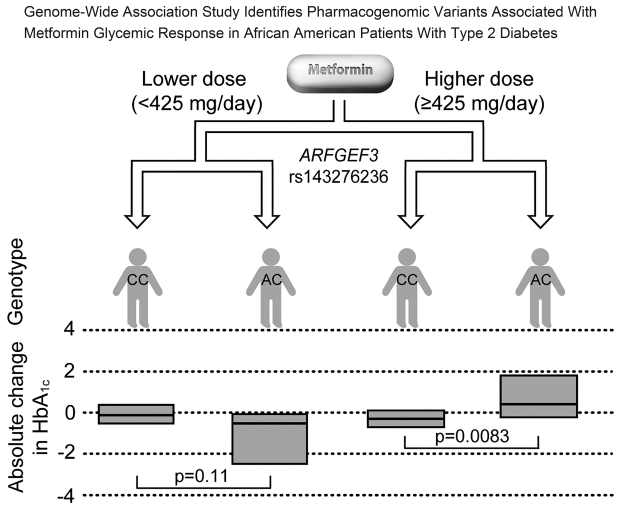
Peer reviewed

Diabetes Care
American Diabetes Association

Genome-Wide Association Study Identifies Pharmacogenomic Variants Associated With Metformin Glycemic Response in African American Patients With Type 2 Diabetes

Baojun Wu, Sook Wah Yee, Shujie Xiao, Fei Xu, Sneha B. Sridhar, Mao Yang, Samantha Hochstadt, Whitney Cabral, David E. Lanfear, Monique M. Hedderson, Kathleen M. Giacomini, and L. Keoki Williams

Diabetes Care 2024;47(2):208–215 | https://doi.org/10.2337/dc22-2494



ARTICLE HIGHLIGHTS

- There are few existing pharmacogenomic studies in populations of color.
- This study is one of the few pharmacogenomic studies of metformin for the treatment of type 2 diabetes and the first in which African American individuals alone comprised the discovery population.
- Variants in the gene *ARFGEF3* appeared to impact metformin treatment response as measured by the change in glycated hemoglobin levels.
- These findings may help to direct type 2 diabetes treatment in the future and may lead to new therapeutic targets.

Genome-Wide Association Study Identifies Pharmacogenomic Variants Associated With Metformin Glycemic Response in African American Patients With Type 2 Diabetes

Baojun Wu,[1] Sook Wah Yee,[2] Shujie Xiao,[1] Fei Xu,[3] Sneha B. Sridhar,[3] Mao Yang,[1] Samantha Hochstadt,[1] Whitney Cabral,[1] David E. Lanfear,[1] Monique M. Hedderson,[3] Kathleen M. Giacomini,[2] and L. Keoki Williams[1]

Diabetes Care 2024;47:208–215 | https://doi.org/10.2337/dc22-2494

ORIGINAL ARTICLE

OBJECTIVE

Metformin is the most common treatment for type 2 diabetes (T2D). However, there have been no pharmacogenomic studies for T2D in which a population of color was used in the discovery analysis. This study sought to identify genomic variants associated with metformin response in African American patients with diabetes.

RESEARCH DESIGN AND METHODS

Patients in the discovery set were adult, African American participants from the Diabetes Multi-omic Investigation of Drug Response (DIAMOND), a cohort study of patients with T2D from a health system serving southeast Michigan. DIAMOND participants had genome-wide genotype data and longitudinal electronic records of laboratory results and medication fills. The genome-wide discovery analysis identified polymorphisms correlated to changes in glycated hemoglobin (HbA_{1c}) levels among individuals on metformin monotherapy. Lead associations were assessed for replication in an independent cohort of African American participants from Kaiser Permanente Northern California (KPNC) and in European American participants from DIAMOND.

RESULTS

The discovery set consisted of 447 African American participants, whereas the replication sets included 353 African American KPNC participants and 466 European American DIAMOND participants. The primary analysis identified a variant, rs143276236, in the gene *ARFGEF3*, which met the threshold for genome-wide significance, replicated in KPNC African Americans, and was still significant in the meta-analysis ($P = 1.17 \times 10^{-9}$). None of the significant discovery variants replicated in European Americans DIAMOND participants.

CONCLUSIONS

We identified a novel and biologically plausible genetic variant associated with a change in HbA_{1c} levels among African American patients on metformin monotherapy. These results highlight the importance of diversity in pharmacogenomic studies.

In the U.S., ∼34 million people (10.5% of the total population) have diabetes, and the majority of these individuals (90–95%) have had type 2 diabetes (T2D) (1). For

[1]Center for Individualized and Genomic Medicine Research (CIGMA), Department of Internal Medicine, Henry Ford Health System, Detroit, MI
[2]Department of Bioengineering and Therapeutic Sciences and Institute for Human Genetics, School of Pharmacy, University of California San Francisco, San Francisco, CA
[3]Division of Research, Kaiser Permanente Northern California, Oakland, CA

Corresponding authors: Baojun Wu, bwu3@hfhs.org, and L. Keoki Williams, kwillia5@hfhs.org

Received 22 December 2022 and accepted 3 August 2023

This article contains supplementary material online at https://doi.org/10.2337/figshare.23925912.

B.W., S.W.Y., and S.X. contributed equally as first authors.

This article is featured in a podcast available at diabetesjournals.org/care/pages/diabetes_care_on_air.

See accompanying article, p. 193.

more than a decade, the American Diabetes Association and the European Association for the Study of Diabetes have recommended metformin as the initial drug of choice for the treatment of T2D (2,3). According to the 2018 Medical Expenditure Panel Survey, metformin is estimated to be the sixth most commonly used medication in the U.S. in terms of the number of patients filling a prescription and the fourth most common medication in terms of total prescriptions filled (4). Nevertheless, the mechanism by which metformin exerts its effect on blood glucose levels and gluconeogenesis is not completely known and may involve both AMPK-dependent and AMPK-independent pathways (5).

Heritability for the change in glycated hemoglobin (HbA_{1c}) levels while on metformin has been estimated to be between 20% and 34%, suggesting that a considerable proportion of treatment response is genetically determined (6). To date, there have been four published genome-wide association studies (GWAS) to identify pharmacogenomic variants associated with metformin treatment response (7–11). Of note, using the additive genetic model, the variants identified in these studies did not meet genome-wide statistical significance in the discovery analysis alone but, rather, reached statistical significance only after meta-analyzing results with other cohorts.

In other pharmacogenomic analyses, we and others have demonstrated that risk variants identified in one population group often do not replicate in other groups (12). We have also demonstrated the advantages of genetic association studies in diverse population groups, particularly African Americans, where lower linkage disequilibrium (LD) can facilitate finer mapping resolution (13).

In this study, we performed an ancestry-stratified genome-wide analysis for pharmacogenomic variants associated with glycemic response among African American individuals treated with metformin. Study participants were from the Diabetes Multiomic Investigation of Drug Response (DIAMOND), a diverse cohort of individuals with T2D from southeast Michigan and the Detroit metropolitan area.

RESEARCH DESIGN AND METHODS

Study Populations and Settings

The cohorts included in this study were approved by the institutional review boards of the Henry Ford Health System; the University of California, San Francisco; and Kaiser Permanente. All participants provided written informed consent at the time of enrollment prior to the collection of study-related data and samples. Eligible DIAMOND participants were aged ≥18 years, were members of the health system and possessed affiliated health insurance with pharmaceutical coverage, and resided in southeast Michigan. Among this group, individuals with two or more recorded diagnoses of T2D and at least one HbA_{1c} value ≥6.5% were invited to participate. Individuals with a diagnosis of type 1 diabetes, gestational diabetes mellitus alone, or drug-associated diabetes (e.g., because of corticosteroid use) were excluded. For the discovery analysis, we restricted the study sample to individuals who self-identified as African American and who had been treated with metformin alone (i.e., metformin monotherapy) for diabetes.

The replication cohort consisted of African American patients who received care at Kaiser Permanente Northern California (KPNC), and White individuals of European descent (European Americans) from the DIAMOND cohort. The KPNC cohort has been described in detail elsewhere (9,14). Electronic medical records were used to identify patients with T2D who had received at least 6 months of care in the health system and who had not used other diabetes medications prior to their first metformin prescription. Patients had to have an HbA_{1c} measurement within 90 days and ≥90 days following metformin initiation. Patients who met these criteria were contacted via letter and invited to participate.

DNA Collection, Genotyping, and Quality Control

At the time of enrollment, participants underwent a detailed evaluation that included completing a research survey, a clinical examination, and specimen collection. DNA was primarily obtained from blood, but occasionally from saliva (Oragene DNA Kit; DNA Genotek Inc.). Genomic DNA was isolated and assessed for its quality and quantity using a Nanodrop spectrophotometer and PicoGreen dsDNA quantification assay (Thermo Fisher Scientific). Genome-wide genotyping was performed with the Axiom Precision Medicine Research Array (Thermo Fisher Scientific) at the University of North Carolina's Functional Genomics Core. The genotyping quality control (QC) threshold for including an individual's results was a dish QC >0.82 and an overall call rate >97%. A total of 2,158 DIAMOND participants were genotyped; 13 (0.6%) samples did not pass genotyping QC.

Variant QC included assessment of overall call rate and Hardy-Weinberg equilibrium ($P < 10^{-6}$). Of 855,849 single nucleotide polymorphisms (SNPs) genotyped, 797,368 (93%) passed our QC criteria. The internal QC filter of the Michigan Imputation Server (reference panel TOPMed-r2 version 1.0.0 [hg38]) excluded an additional 70,265 variants, and the remaining set of variants were used to impute missing genotypes (15). Variants with an imputation score ≥0.8 were used in the analysis. We also restricted our association analyses to biallelic SNPs with a minor allele frequency (MAF) ≥1%.

Population structure was characterized using principal components (PCs) (16,17). Individuals ≥6 SDs from their self-reported race group in the first 10 PCs were removed from the analysis (no African Americans and eight European American participants were excluded based on this criterion). An additional eight samples were removed (three African American and five European American participants) based on an inbreeding coefficient ≥6 SDs from the rest of the group. Relatedness among participants was estimated using kinship coefficients computed in PLINK 2.0 (18). In situations where participants were related (i.e., pairwise kinship coefficient >0.1), one individual's sample was randomly excluded from the analysis set. Twenty-six individuals (20 African American and 6 European American participants) were removed based on relatedness.

Individuals in the KPNC cohort and their samples underwent a similar QC process. These individuals were genotyped on the Illumina Infinium OmniExpress-24 version 1.2 array. After QC, 725,468 SNPs were used to impute missing genotypes on the Michigan Imputation Server.

Local and global genetic ancestries were assessed in both the discovery and replication groups using the programs RFMix version 1.5.4 (19) and ADMIXTURE 1.3.0 (20). The reference populations for ancestry estimation were the YRI (Yoruba in Ibadan, Nigeria) and CEU (Northern Europeans from Utah) groups from the 1000 Genomes Project (21).

Clinical Data and Drug Exposure Measures

The HbA_{1c} laboratory results used for assessing metformin response had to be at least 120 days apart; metformin use had to span this interval such that a 120-day window of drug exposure could be calculated and related to the change in HbA_{1c}. The first set of HbA_{1c} measurements fulfilling these criteria (i.e., the interval closest to or including treatment initiation) was used in calculating the primary outcome.

Metformin exposure was estimated using available pharmacy claims data in the manner similar to what we have done previously (22); a schematic is shown in Supplementary Fig. 1. We used the claims information on each metformin fill (i.e., preparation strength, total amount dispensed) to calculate the total amount of medication received. The time interval between refills was then used to equally distribute the total amount dispensed over that interval, thereby assigning an estimated daily amount consumed to each day between fills. Anchoring on the date of the second HbA_{1c} test result (for the observation window used), we averaged the estimated daily amount of metformin consumed for the preceding 120-day period (i.e., the period of glycemic control reflected in the HbA_{1c} result). Baseline HbA_{1c} was the first HbA_{1c} measurement in the selected observation window. Serum creatinine measures taken within the 6 months of baseline HbA_{1c} were used for the estimated glomerular filtration rate (eGFR) (23).

Statistical Analysis

The primary outcome was the change in HbA_{1c} levels while on metformin monotherapy. In our initial analysis, we assessed the relationship between genotype and the primary outcome, restricted to individuals whose estimated daily metformin use was $\geq$425 mg. This dose was selected to represent the equivalent of 80% adherence to the lowest possible daily dose of metformin (i.e., 500 mg/day) or 50% adherence of the next lower pill size (850 mg) to define a consistent level of exposure; this is the approximate range of adherence observed in studies of metformin use by us and others (24). To validate this threshold, we also evaluated the relationship between metformin use and change in HbA_{1c} levels (Supplementary Fig. 2) and the distribution of metformin exposure (Supplementary Fig. 3). These analyses suggested that 425 mg/day of metformin constituted an inflection point in terms of treatment response (i.e., the point above which increasing metformin use lowered HbA_{1c} levels). Change in HbA_{1c} was normalized using rank-based inverse normal transformation (25). The transformed change in HbA_{1c} levels was regressed on genotype, using an additive genetic model and restricted to biallelic SNPs. Additional covariates included patient age, sex, eGFR, and baseline HbA_{1c} levels (i.e., the initial HbA_{1c} level in the interval assessed). Association testing was performed using the software package GENESIS (26) and was restricted to regions identified as homozygous for African ancestry based on results from RFMix. Genome-wide statistical significance was defined as $P < 5 \times 10^{-8}$.

Variants that met the threshold for statistical significance were reassessed in the KPNC cohort and in European American DIAMOND participants. A P value of 0.05 was used as the significance threshold for the replication. The regression models used in the replication analysis were the same as those used for the discovery analysis. Results from African American participants in the discovery and replication groups were meta-analyzed using the program METAL (27). Based on a discovery group sample size of 447, $P < 5 \times 10^{-8}$ significance threshold, and an MAF of either $\geq$0.01 or $\geq$0.03, we estimated 80% power to detect effect sizes of $\geq$2.07 and $\geq$1.21, respectively. For the replication group using a sample size of 353, P = 0.05 significance threshold, and an MAF of either $\geq$0.01 or $\geq$0.03, we estimated 80% power to detect effect sizes of $\geq$1.05 and $\geq$0.61, respectively. We also performed a post hoc sensitivity analysis to account for the periods used in our evaluation with respect to metformin initiation. These analyses included separately stratifying the group by the time of assessment, as well as by adjusting for time since initiation in the regression model.

Our second approach for identifying pharmacogenomic variants associated with metformin treatment response involved evaluating for gene (genotype) × drug interactions associated with change in HbA_{1c}. Rank-based inverse normal transformed change in HbA_{1c} levels was regressed on genotype, metformin exposure (i.e., <425 vs. $\geq$425 mg/day), and an interaction term between genotype and metformin use. The models adjusted for patient age, sex, eGFR, and baseline HbA_{1c} levels. Statistical significance was assessed by comparing the full model (i.e., the model with age, sex, eGFR, baseline HbA_{1c} level, metformin exposure, genotype, and a genotype × metformin interaction) with a parsimonious model (i.e., the model without the genotype and genotype × metformin interaction terms). Therefore, the joint association test assessed the combined effect of adding genotype and the interaction term. The genome-wide statistical significance threshold for the joint association was also set at $P < 5 \times 10^{-8}$. See the Supplementary Methods and Discussion for additional commentary on the analyses performed.

RESULTS

Study Populations

The discovery set comprised 447 African American DIAMOND participants with genome-wide genotype data and consistent metformin use (i.e., an estimated $\geq$425 mg/day during the observation period). Similarly, there were 353 African American participants from the KPNC cohort and 466 European American participants from the DIAMOND cohort with similar data for replication. The characteristics of these participants are shown in Table 1 and Supplementary Table 1. The average age among African American DIAMOND participants, African American KPNC participants, and European American DIAMOND participants was 56.4, 55.2, and 59.5 years, respectively. The sex distribution in these groups was 69.1%, 66.3%, and 41.8% female, respectively. Baseline HbA_{1c} levels in these groups were 7.6%, 7.3%, and 7.3%, and the average absolute change in HbA_{1c} among individuals whose average metformin use was $\geq$425 mg/day during the study observation period was -0.57, -0.66, and -0.44, respectively.

Genome-Wide Association Identifies Relationship Between Variant rs143276236 and HbA_{1c} Change Among Individuals With T2D on Metformin Therapy

Results from the genome-wide association analysis are shown in Table 2 and Fig. 1. The quantile-quantile plot is shown in Supplementary Fig. 4. In the discovery analysis, 11 SNPs had an association P value less than the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ in regions homozygous for African ancestry.

Table 1—Characteristics of discovery and replication groups used in the primary analysis*

	Discovery group	Replication groups	
	African American DIAMOND participants (n = 447)	African American KPNC participants (n = 353)	European American DIAMOND participants (n = 466)
Age (years)†	56.4 ± 10.3	55.2 ± 10.0	59.5 ± 10.5
Female sex, n (%)	309 (69.1)	234 (66.3)	195 (41.8)
BMI (kg/m^2)‡	35.0 ± 7.1 (n = 328)	35.3 ± 7.2 (n = 287)	34.2 ± 6.0 (n = 306)
Proportion of African ancestry§	81.7 ± 10.3	79.4 ± 14.3	0.06 ± 2.0
Baseline HbA_{1c} level (%)‖	7.6 ± 1.8	7.3 ± 1.8	7.3 ± 1.4
Change in HbA_{1c} on metformin treatment**	−0.57 ± 1.76	−0.66 ± 1.78	−0.44 ± 1.40
Metformin use (mg/day)††	937.5 ± 419.0	963.5 ± 471.7	1,088.5 ± 502.9
eGFR (mL/min/1.73 m^2)‡‡	93.7 ± 20.4 (n = 447)	94.4 ± 19.1 (n = 353)	87.0 ± 18.5 (n = 466)

Data are mean ± SD unless otherwise indicated. *This is among the study sample with an average metformin use ≥425 mg/day. †Age at the time of metformin response was assessed. ‡BMI was available for 344 (72%), 298 (81%), and 319 (64%) of the African American DIAMOND participants, the African American KPNC participants, and the European American DIAMOND participants, respectively. §Using the approach described in the *Research Design and Methods*, the proportion of African ancestry refers to the proportion of each individual's genome determined to be of African continental origin. This was then averaged over all participants to determine the average proportion across the subset of participants. ‖The result of the first HbA_{1c} level in the interval used to assess metformin response. **Absolute change in the HbA_{1c} percent in the observation period. ††Average estimated metformin daily use in mg/day during the 120-day period preceding the second HbA_{1c} measurement in the observation interval. ‡‡Derived using creatinine levels drawn within the 6 months preceding the exposure observation period used for each individual.

Variants rs143276236 and rs116251012 in the gene *ARFGEF3* on chromosome 6 had a consistent direction of effect in both DIAMOND and KPNC cohorts, and the variants had a replication $P < 0.05$ in KPNC cohort. A meta-analysis of the signal in regions homozygous for African ancestry in DIAMOND and KPNC participants was statistically significant for rs143276236 ($P = 1.17 \times 10^{-9}$) and for rs116251012 ($P = 1.59 \times 10^{-8}$). A locus zoom plot of the region (Supplementary Fig. 5) demonstrates that the lead variant rs143276236 was in high LD with variant rs116251012 ($r^2 = 0.78$). Only two of the variants identified in African ancestry regions were sufficiently frequent to assess in European American DIAMOND participants (Supplementary Table 2). Neither of these variants reached statistical significance among European American participants or when meta-analyzed in all three study groups.

Unique African and European Haplotypes in Chromosome 6 Region Associated With Metformin Glycemic Response

Supplementary Fig. 6 demonstrates the haplotype structure in the region surrounding the leading variant rs143276236 on chromosome 6. As can be seen, the extent of LD was lower and the size of the haplotype blocks smaller in a background homozygous for African ancestry (Supplementary Fig. 6*A*) compared with the same region homozygous for European ancestry (Supplementary Fig. 6*B*).

An Ancestry-Informed Gene × Metformin Interaction Association Analysis for Glycemic Response Identifies Similar Top Genetic Variant Associations

In DIAMOND, we had 203 additional African American participants who were on metformin monotherapy but had an average estimated exposure <425 mg/day in the observation period (Supplementary Table 3). Average estimated metformin use in this group was 224.2 mg/day. As would be expected, the absolute improvement in HbA_{1c} levels was lower in individuals whose average exposure was <425 mg/day compared with those using ≥425 mg/day (−0.22 vs. −0.57, respectively; $P = 0.024$). Both the high and low metformin exposure groups (i.e., ≥425 mg/day and <425 mg/day) were used to assess for gene × metformin interactions (i.e., variants whose effect on HbA_{1c} appeared to differ based on metformin exposure). The association analysis was again restricted to intra-individual genomic regions homozygous for African ancestry. The quantile-quantile plot is shown in Supplementary Fig. 7. A joint test was used to assess the combined effect of genotype and genotype × drug interaction (Supplementary Table 4). The effect of metformin on the change in HbA_{1c} was evinced by the consistent inverse relationship between level of metformin exposure and the outcome variable. Many of the same variants identified in the previous analysis (Table 2) were again observed in the interaction analysis. *ARFGEF3* was also identified in the gene × metformin interaction analysis; significant variants in this gene included rs143276236 (noted previously) and the two SNPs rs141012141 and rs80340144 located within 137 base pairs of each other. The joint test P values for these three *ARFGEF3* variants were 1.08×10^{-8}, 3.48×10^{-8}, 3.48×10^{-8}, respectively; however, only rs143276236 suggested replication in the KPNC cohort (joint test $P = 0.06$).

Association Between Variants in *ARFGEF3* and HbA_{1c} Change by Degree of Metformin Use

Differences in the apparent relationship among degree of metformin use, genotype, and HbA_{1c} improvement are shown in Fig. 2. Among individuals with ≥425 mg/day average daily exposure to metformin, individuals with the *ARFGEF3* rs143276236 CC genotype had an average absolute HbA_{1c} reduction of 0.59%,

Table 2—Genome-wide significant associations for change in HbA_{1c} levels among individuals on metformin monotherapy*

					DIAMOND cohort (discovery set)				KPNC cohort (replication set)					
Variant	Chr	Position (Hg38)	Alleles	Nearest gene	Functional consequence	*n*	MAF	Parameter estimate†	*P*	*n*	MAF	Parameter estimate†	*P*	Direction of effect	Meta-analysis *P*
rs150979496	18	8501049	T/C	*THEMIS3P*	None known	300	0.012	1.90	5.04×10^{-9}	—	NA	NA	NA	NA	NA
rs143276236	6	138176343	A/C	*ARFGEF3*	Intron variant	292	0.027	1.20	6.39×10^{-9}	227	0.024	0.60	0.0072	+/+	1.17×10^{-9}
rs59699971	12	40649666	G/A	*MUC19*	None known	288	0.031	1.16	9.57×10^{-9}	219	0.041	−0.32	0.0977	+/−	0.006
rs115567566	12	40671268	G/A	*MUC19*	None known	287	0.031	1.16	1.13×10^{-8}	218	0.039	−0.39	0.0482	+/−	0.011
rs142017898	16	83500006	C/T	*CDH13*	Intron variant	287	0.017	1.52	1.15×10^{-8}	219	0.025	−0.09	0.6884	+/−	0.001
rs138642807	9	109970865	T/G	*PALM2-AKAP2*	Intron variant	285	0.012	1.88	1.25×10^{-8}	221	0.018	0.20	0.4432	+/+	3.17×10^{-5}
rs146090661	9	109975903	G/A	*PALM2-AKAP2*	Intron variant	285	0.012	1.88	1.25×10^{-8}	221	0.018	0.20	0.4432	+/+	3.17×10^{-5}
rs12297682	12	40920062	G/A	*CNTN1*	Intron variant	287	0.052	0.85	4.09×10^{-8}	219	0.055	−0.24	0.1506	+/−	0.003
rs79162841	12	40926234	G/A	*CNTN1*	Intron variant	287	0.052	0.85	4.09×10^{-8}	219	0.055	−0.24	0.1506	+/−	0.003
rs116251012	6	138191883	T/C	*ARFGEF3*	Intron variant	292	0.031	1.08	4.80×10^{-8}	227	0.035	0.49	0.0092	+/+	1.59×10^{-8}
rs149911554	10	106606746	A/G	*SORCS1*	Intron variant	274	0.013	1.54	4.99×10^{-8}	210	0.024	0.44	0.0775	+/+	8.45×10^{-7}
rs114019854	10	106606829	T/C	*SORCS1*	Intron variant	274	0.013	1.54	4.99×10^{-8}	210	0.024	0.44	0.0775	+/+	8.45×10^{-7}

Chr, chromosome; NA, not available. *The dependent variable was the rank-based inverse normal transformation of the change in HbA_{1c}. *P* values are the association results for a given genotype adjusted for patient age, sex, eGFR, and baseline HbA_{1c} levels. Association and meta-analysis results for African American participants were from individuals who were homozygous for African ancestry at that location. Analysis was restricted to individuals with a level of metformin exposure ≥425 mg/day. The genome-wide significance threshold was $P < 5.0 \times 10^{-8}$. †Parameter estimates represent the approximate effect size of the genotype (coded 0, 1, and 2 for the number of copies of effect allele). A negative parameter estimate (−) suggests that the effect allele is associated with a reduction in HbA_{1c} over the observation period, and a positive parameter estimate (+) is associated with an increase in HbA_{1c}.

whereas individuals with the AC genotype had an average absolute HbA_{1c} reduction of 0.28%; these differences were statistically significant ($P = 0.0083$) (Fig. 2*A*). The genotype relationship was not statistically significant among individuals with average metformin exposure <425 mg/day ($P = 0.11$). A similar pattern was also observed using the HbA_{1c} change adjusted for baseline HbA_{1c}. The difference between the two genotypes were statistically significant ($P = 0.012$) (Fig. 2*B*) with average metformin exposure ≥425 mg/day, but the difference was not statistically significant among individuals with average metformin exposure <425 mg/day ($P = 0.91$).

Post Hoc Analyses Accounting for Time of Treatment

Since the exposure window used in our assessment may not have coincided with metformin initiation, we performed a post hoc analysis to assess the potential effect of timing on our results. Most assessments were performed at the time of metformin initiation (Supplementary Fig. 8). The durability of metformin response appeared to last to day 403 following treatment initiation, but it was most evident in the first 146 days (Supplementary Fig. 9). The lead variant rs143276236 in the association analysis had a consistent magnitude and direction of effect between 0 and 146 days and between 147 and 403 days postinitiation (Supplementary Table 5), and it demonstrated genome-wide significance within the 403 days in which the metformin treatment response was evident ($P = 9.72 \times 10^{-9}$) (Supplementary Table 6). Variant rs143276236 was also associated with change in HbA_{1c} after adjusting for time since treatment initiation in both the discovery ($P = 9.09 \times 10^{-9}$) and replication sets ($P = 0.035$) (Supplementary Table 7).

CONCLUSIONS

Despite rapid progress in genomics, U.S. minority populations are considerably underrepresented in existing studies (28). For example, individuals of African descent constitute ∼2% of all GWAS participants (29,30), and these figures are much smaller when one considers the number of studies where people of color comprise the discovery population or are sufficiently represented for adequately powered subgroup analyses. Without such studies, these population groups would not share in precision

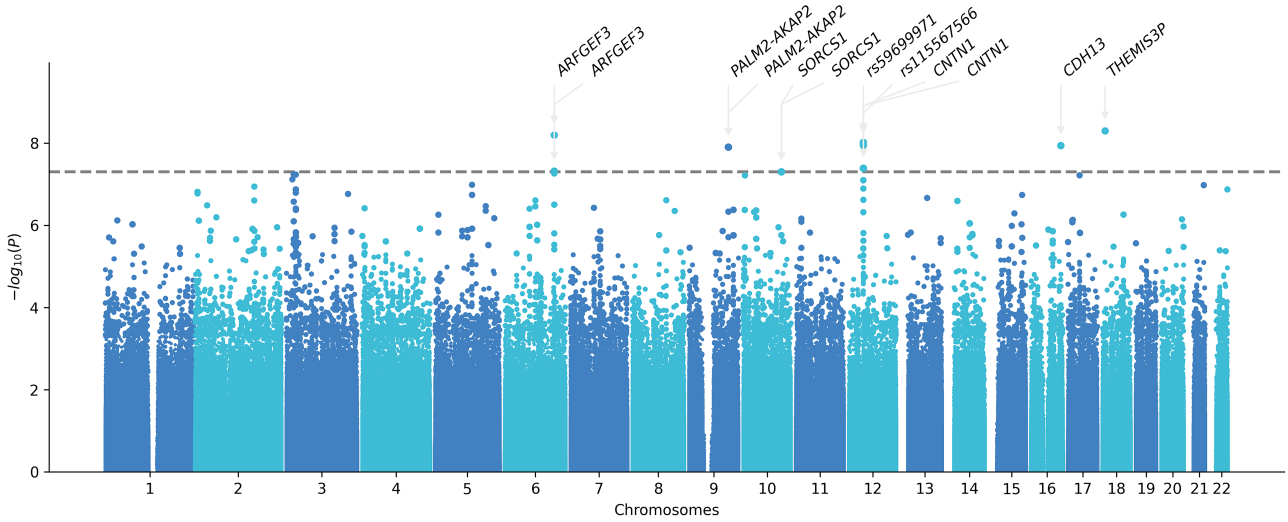


Figure 1—Manhattan plot of the discovery analysis relating genome-wide polymorphisms to changes in HbA_{1c} among African American patients with T2D on metformin monotherapy from the DIAMOND cohort. Discovery association results were restricted to regions where individuals were homozygous for African ancestry. The dashed line denotes the threshold for genome-wide statistical significance ($P < 5.0 \times 10^{-8}$).

medication advancements that are predicated on this foundational research (31).

To leverage the strength of our study's diversity, we used a technique that we have used successfully on other projects involving fine mapping of variant associations in regions homozygous for African ancestry (13). This approach has multiple potential advantages. First, it can be used to assess for associations in an ancestral background not previously evaluated, thereby identifying novel or population-specific risk variants. Second, it may permit better mapping resolution of causal variants, especially if ancestral group differences in allele frequency, LD, or genetic variation permit better signal isolation (32). We also used methods developed by us to assess metformin exposure in large patient populations to ensure that medication use was taken into account (22,24).

Using the above approach, we evaluated for variants associated with a change in HbA_{1c} levels while on metformin monotherapy. The most significant allele replicated was rs143276236, an intronic SNP located in the gene *ARFGEF3* that is also known as brefeldin A-inhibited guanine nucleotide-exchange protein 3 (*BIG3*). Experiments in mice have found *BIG3* to be highly expressed in islet α- and β-cells, and *BIG3* expression is inversely associated with both glucagon and insulin granule biogenesis in both types of cells, respectively (33,34). *BIG3* knockout mice also had a 35% increase in insulin granule content in β-cells and enhanced insulin secretion; however, they were also more likely to develop postprandial hyperglycemia and hyperinsulinemia, consistent with peripheral insulin resistance attributable to chronic oversecretion (33). In short, *BIG3* appears to function as a negative regulator of insulin secretion through its role in regulating early granule biogenesis in the Golgi apparatus. While the mechanistic relationship between *ARFGEF3*/*BIG3* and metformin use remain to be determined, the gene itself is a plausible mediator of blood glucose levels.

To investigate the possible effects that our lead variant rs143276236 may have on gene expression, we performed a post hoc analysis using HaploReg version 4.1 (35) to assess potential effects on transcription factor binding. The algorithm estimated that the variant may alter activating transcription factor 3 (*ATF3*) and *CACD2* binding motifs with reduced binding affinity. Studies have suggested that *ATF3* may influence both T2D and metformin's effect on hepatic gluconeogenesis (36,37). Furthermore, a study showed that *ATF3* is the top gene that is regulated by metformin in human liver hepatocytes (38).

A previous GWAS identified variant rs8192675 in the glucose transporter gene *SLC2A2* to be associated with metformin glycemic response among White European individuals with T2D (9). We did not find any significant variants within the *SLC2A2* among participants of African ancestry, and the association of rs8192675 with change in HbA_{1c} obtained a *P* value of 0.81 in our discovery set. The discrepancy highlights the potential importance of diversity in pharmacogenomic studies. In addition, Mahajan et al. (39) performed a meta-analysis of 122 separate studies comprising 180,834 individuals with T2D and 1,159,055 control subjects; the composition of the group was 51.1% White European, 28.4% East Asian, 8.3% South Asian, 6.6% African (including African Americans), and 5.6% Latino (of any ancestry group). The investigators identified 277 separate genetic loci associated with T2D at a significance level of $P < 5 \times 10^{-8}$. Again, we did not find any significant variants overlapping with these 277 loci, suggesting that the variant that we identified may be only relevant in the presence of metformin, as was observed in our study. Recently, Li et al. (11) conducted a GWAS of individuals with prediabetes from the Diabetes Prevention Program (DPP). The investigators assessed both for variants associated with various glycemic traits, including HbA_{1c}, in the treatment arm and for gene × drug interactions (using both treatment and placebo arms). Variants in *ENOSF1* and *LOC101928519* were associated with a 1-year change in HbA_{1c} levels in the metformin-only group, and variants in *LINC01093*, *MAN2B2*, and *SOX5* were associated with change in HbA_{1c} in the interaction model. Only *LOC101928519* was found to be statistically significant among African American

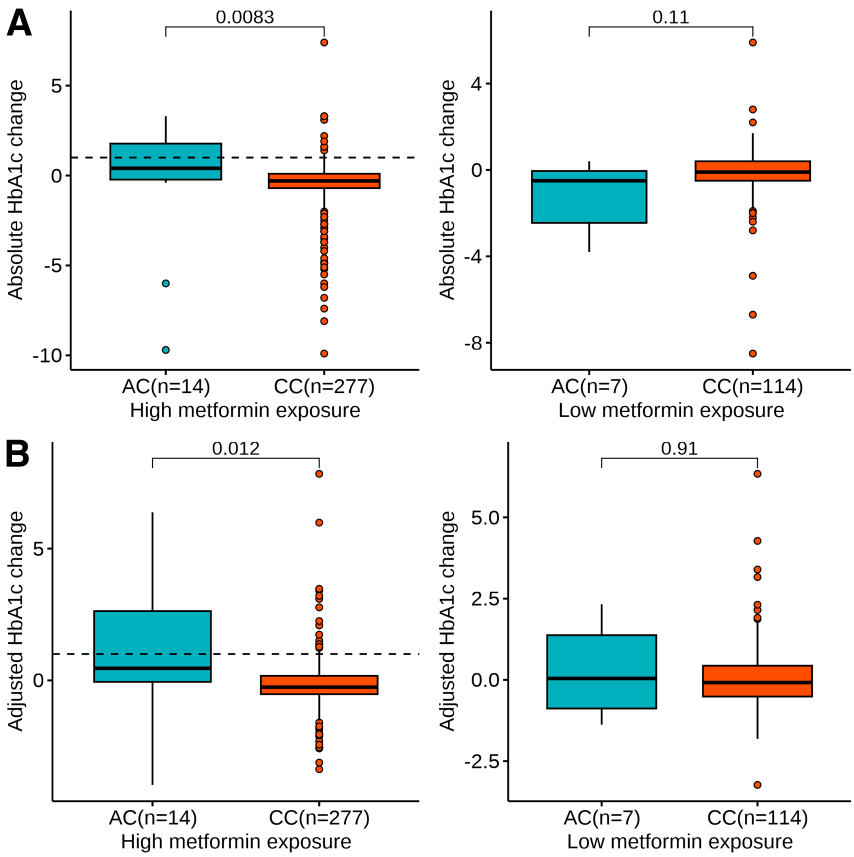


Figure 2—Absolute changes (*A*) and adjusted changes (*B*) in HbA_{1c} levels with metformin treatment among African American DIAMOND participants stratified by the *ARFGEF3* variant genotype rs143276236. Results are also stratified by average levels of daily metformin use as assessed using pharmacy records of patient medication fills. Among individuals on metformin monotherapy for T2D, high exposure was considered ≥425 mg/day (left), and low exposure was considered <425 mg/day (right). Only one individual was AA homozygous for the *ARFGEF3* rs143276236 variant (noted by dashed line).

DPP participants, and none of these variants were identified in our current study.

Although the mechanism of metformin's effect on blood glucose levels is not known, it is widely speculated that the principal effect involves hepatic gluconeogenesis (40). Curiously, our top associations in *ARFGEF/BIG3* suggest that metformin may also affect glucose and insulin levels through islet cell granule formation. However, our observational study design did not allow us to investigate whether metformin has a direct or indirect effect on these processes. Nevertheless, the consistency of our findings in terms of site of action, robustness to different analytic methods, and replication in an independent sample suggests that the pharmacogenomic variants identified here deserve additional functional validation.

The use of observational data to assess medication response has its limitations. For example, we had to use strict criteria to define drug exposure using retrospective electronic medical record data. We have previously demonstrated that these exposure measures are strongly related to clinical outcomes (22); hence, these data provide an opportunity to evaluate population groups that are understudied in clinical trials. We also acknowledge that 425 mg/day does not correspond to an actual metformin pill size; however, this level of exposure did appear to constitute an inflection point for metformin treatment response (Supplementary Fig. 2). It is important to note that since patients often do not take their full dose as prescribed (24), average levels of exposure may not perfectly coincide with pill strengths. To best leverage the data available to us, we also used exposure times that did not always include metformin initiation. Nevertheless, in post hoc analyses, we demonstrated that our lead variant was robust to various methods to account for time since treatment initiation. Most importantly, we demonstrated that rs143276236 was a leading pharmacogenomic variant using two very different analytic approaches: an exposure-only analysis and a genotype × drug interaction analysis.

In summary, we have identified at least one variant that appears to influence metformin's glycemic effect in African American individuals with T2D. Performing pharmacogenomic studies in diverse population groups is necessary for a number of reasons. First, pharmacogenomic information for U.S. minority patients is lacking, so without these studies, these population groups will not equally benefit from downstream advances in precision medicine. For example, genetic predictive risk scores developed using data from European ancestry groups perform poorly in other population groups (31); therefore, more diverse pharmacogenomic studies are needed as a basis for more broadly applicable risk scores (22). Second, LD differences between population groups may actually improve fine-mapping resolution and causal variant identification (13). While further studies are needed to replicate our findings, our approach provides a useful roadmap for others to follow in terms of leveraging diversity to improve pharmacogenomic discovery.

Acknowledgments. The authors acknowledge Dr. Michiaki Kubo, Dr. Atsushi Takahashi, Dr. Shiro Maeda, Dr. Yukihide Momozawa, and Dr. Yoichiro Kamatani at the RIKEN Institute for performing genotyping of the KPNC cohort on the Illumina Infinium OmniExpress-24, version 1.2 array.

Funding. This project was supported by the Fund for Henry Ford Hospital and National Institute of Diabetes and Digestive and Kidney Diseases grants R01DK064695 and R01DK113003 (to L.K.W.); National Heart, Lung, and Blood Institute grants R01HL079055, R01HL118267, and R01HL141845 (to L.K.W.), National Institute of Allergy and Infectious Diseases grants R01AI079139 and R56AI165903 (to L.K.W.), and National Institute of General Medical Sciences grants U01GM061390 and R01GM117163 (to S.W.Y., K.M.G., and M.M.H.).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. B.W., S.W.Y., S.X., F.X., S.B.S., M.Y., S.H., W.C., D.E.L., M.M.H., K.M.G., and L.K.W. contributed to the data collection. B.W., S.W.Y., S.X., and L.K.W. contributed to the conception and design of the study and to the data analysis. B.W., S.X., and L.K.W. contributed to the initial manuscript writing. All authors revised the manuscript for important content and approved of the final version. B.W. and L.K.W. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the analysis.

References

1. Centers for Disease Control and Prevention, Ed. *National Diabetes Statistics Report, 2020.* Atlanta, GA, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2020
2. Nathan DM, Buse JB, Davidson MB, et al.; American Diabetes Association; European Association for Study of Diabetes. Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetes Care 2009;32:193–203
3. Buse JB, Wexler DJ, Tsapas A, et al. 2019 Update to: Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2020;43:487–493
4. Agency for Healthcare Research and Quality (Ed.). Total purchases in thousands by prescribed drug, United States, 1996-2018. In: *Medical Expenditure Panel Survey*. Rockville, MD, Agency for Healthcare Research and Quality, Rockville, Maryland, 2021. Accessed 1 August 2023. Available from https://datatools.ahrq.gov/meps-hc/?tab=prescribed-drugs&dash=18
5. Rena G, Hardie DG, Pearson ER. The mechanisms of action of metformin. Diabetologia 2017;60:1577–1585
6. Zhou K, Donnelly L, Yang J, et al.; Wellcome Trust Case Control Consortium 2. Heritability of variation in glycaemic response to metformin: a genome-wide complex trait analysis. Lancet Diabetes Endocrinol 2014;2:481–487
7. GoDarts and UKPDS Diabetes Pharmacogenetics Study Group; Wellcome Trust Case Control Consortium 2. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nat Genet 2011;43:117–120
8. van Leeuwen N, Nijpels G, Becker ML, et al. A gene variant near ATM is significantly associated with metformin treatment response in type 2 diabetes: a replication and meta-analysis of five cohorts. Diabetologia 2012;55:1971–1977
9. Zhou K, Yee SW, Seiser EL, et al.; MetGen Investigators; DPP Investigators; ACCORD Investigators. Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. Nat Genet 2016;48:1055–1059
10. Rotroff DM, Yee SW, Zhou K, et al.; MetGen Investigators; ACCORD/ACCORDion Investigators. Genetic variants in *CPA6* and *PRPF31* are associated with variation in response to metformin in individuals with type 2 diabetes. Diabetes 2018;67:1428–1440
11. Li JH, Perry JA, Jablonski KA, et al.; Diabetes Prevention Program Research Group. Identification of genetic variation influencing metformin response in a multiancestry genome-wide association study in the Diabetes Prevention Program (DPP). Diabetes 2023;72:1161–1172
12. Levin AM, Gui H, Hernandez-Pacheco N, et al. Integrative approach identifies corticosteroid response variant in diverse populations with asthma. J Allergy Clin Immunol 2019;143:1791–1802
13. Gui H, Levin AM, Hu D, et al. Mapping the 17q12-21.1 locus for variants associated with early-onset asthma in African Americans. Am J Respir Crit Care Med 2021;203:424–436
14. Goswami S, Yee SW, Xu F, et al. A longitudinal HbA1c model elucidates genes linked to disease progression on metformin. Clin Pharmacol Ther 2016;100:537–547
15. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–1287
16. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190
17. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet 2010;11:459–463
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7
19. 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;93:278–288
20. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–1664
21. Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65
22. Williams LK, Padhukasahasram B, Ahmedani BK, et al. Differing effects of metformin on glycemic control by race-ethnicity. J Clin Endocrinol Metab 2014;99:3160–3168
23. Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. Am J Kidney Dis 2022;79:268–288.e1
24. Pladevall M, Williams LK, Potts LA, Divine G, Xi H, Lafata JE. Clinical outcomes and adherence to medications measured by claims data in patients with diabetes. Diabetes Care 2004;27:2800–2805
25. McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. Biometrics 2020;76:1262–1272
26. Gogarten SM, Sofer T, Chen H, et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics 2019;35:5346–5348
27. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–2191
28. Morales J, Welter D, Bowler EH, et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. Genome Biol 2018;19:21
29. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. Cell 2019;177:26–31
30. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. Nat Genet 2020;52:242–243
31. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51:584–591
32. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. Nat Rev Genet 2019;20:520–535
33. Li H, Wei S, Cheng K, et al. BIG3 inhibits insulin granule biogenesis and insulin secretion. EMBO Rep 2014;15:714–722
34. Li H, Liu T, Lim J, Gounko NV, Hong W, Han W. Increased biogenesis of glucagon-containing secretory granules and glucagon secretion in BIG3-knockout mice. Mol Metab 2015;4:246–252
35. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016;44:D877–D881
36. Kim JY, Park KJ, Hwang JY, et al. Activating transcription factor 3 is a target molecule linking hepatic steatosis to impaired glucose homeostasis. J Hepatol 2017;67:349–359
37. Allen-Jennings AE, Hartman MG, Kociba GJ, Hai T. The roles of ATF3 in liver dysfunction and the regulation of phosphoenolpyruvate carboxykinase gene expression. J Biol Chem 2002;277:20020–20025
38. Luizon MR, Eckalbar WL, Wang Y, et al. Genomic characterization of metformin hepatic response. PLoS Genet 2016;12:e1006449
39. Mahajan A, Spracklen CN, Zhang W, et al.; FinnGen; eMERGE Consortium. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 2022;54:560–572
40. LaMoia TE, Shulman GI. Cellular and molecular mechanisms of metformin action. Endocr Rev 2021;42:77–96