Genome-wide association analysis identifies ancestry-specific genetic variation associated with acute response to metformin and glipizide in SUGAR-MGH

Li, J.H.; Brenner, L.N.; Kaur, V.; Figueroa, K.; Schroeder, P.; Huerta-Chagoya, A.; ... ; Diab Prevent Program DPP Res Grp

Diabetologia (2023) 66:1260–1272
https://doi.org/10.1007/s00125-023-05922-7

ARTICLE

Genome-wide association analysis identifies ancestry-specific genetic variation associated with acute response to metformin and glipizide in SUGAR-MGH

Josephine H. Li[1,2,3,4] · Laura N. Brenner[1,3,4,5] · Varinderpal Kaur[1,2,3] · Katherine Figueroa[1,3] · Philip Schroeder[1,2,3] · Alicia Huerta-Chagoya[1,2,3] · MAGIC Investigators · Diabetes Prevention Program (DPP) Research Group · Miriam S. Udler[1,2,3,4] · Aaron Leong[1,2,3,4,6] · Josep M. Mercader[1,2,3,4] · Jose C. Florez[1,2,3,4]

Received: 14 December 2022 / Accepted: 6 March 2023 / Published online: 26 May 2023

Abstract

Aims/hypothesis Characterisation of genetic variation that influences the response to glucose-lowering medications is instrumental to precision medicine for treatment of type 2 diabetes. The Study to Understand the Genetics of the Acute Response to Metformin and Glipizide in Humans (SUGAR-MGH) examined the acute response to metformin and glipizide in order to identify new pharmacogenetic associations for the response to common glucose-lowering medications in individuals at risk of type 2 diabetes.

Methods One thousand participants at risk for type 2 diabetes from diverse ancestries underwent sequential glipizide and metformin challenges. A genome-wide association study was performed using the Illumina Multi-Ethnic Genotyping Array. Imputation was performed with the TOPMed reference panel. Multiple linear regression using an additive model tested for association between genetic variants and primary endpoints of drug response. In a more focused analysis, we evaluated the influence of 804 unique type 2 diabetes- and glycaemic trait-associated variants on SUGAR-MGH outcomes and performed colocalisation analyses to identify shared genetic signals.

Results Five genome-wide significant variants were associated with metformin or glipizide response. The strongest association was between an African ancestry-specific variant (minor allele frequency [MAF_{Afr}]=0.0283) at rs149403252 and lower fasting glucose at Visit 2 following metformin ($p=1.9\times10^{-9}$); carriers were found to have a 0.94 mmol/l larger decrease in fasting glucose. rs111770298, another African ancestry-specific variant (MAF_{Afr}=0.0536), was associated with a reduced response to metformin ($p=2.4\times10^{-8}$), where carriers had a 0.29 mmol/l increase in fasting glucose compared with non-carriers, who experienced a 0.15 mmol/l decrease. This finding was validated in the Diabetes Prevention Program, where rs111770298 was associated with a worse glycaemic response to metformin: heterozygous carriers had an increase in HbA_{1c} of 0.08% and non-carriers had an HbA_{1c} increase of 0.01% after 1 year of treatment ($p=3.3\times10^{-3}$). We also identified associations between type 2 diabetes-associated variants and glycaemic response, including the type 2 diabetes-protective C allele of rs703972 near *ZMIZ1* and increased levels of active glucagon-like peptide 1 (GLP-1) ($p=1.6\times10^{-5}$), supporting the role of alterations in incretin levels in type 2 diabetes pathophysiology.

Conclusions/interpretation We present a well-phenotyped, densely genotyped, multi-ancestry resource to study gene–drug interactions, uncover novel variation associated with response to common glucose-lowering medications and provide insight into mechanisms of action of type 2 diabetes-related variation.

Data availability The complete summary statistics from this study are available at the Common Metabolic Diseases Knowledge Portal (https://hugeamp.org) and the GWAS Catalog (www.ebi.ac.uk/gwas/, accession IDs: GCST90269867 to GCST90269899).

Josephine H. Li, Laura N. Brenner and Varinderpal Kaur contributed equally to this work.

Josep M. Mercader and Jose C. Florez jointly directed this work.

Members of the MAGIC Consortium and the DPP Research Group are included as collaborators and listed in the electronic supplementary material (ESM) text.

Extended author information available on the last page of the article

Keywords Genetics · Genome-wide association study · Glipizide · Incretin · Metformin · Multi-ancestry · Pathophysiology · Pharmacogenetics · Sulfonylurea · Type 2 diabetes

Research in context

What is already known about this subject?

- Treatment of type 2 diabetes is currently algorithmic and does not consider an individual's underlying genetics or the disease pathophysiology that might benefit from a tailored intervention
- Genome-wide association studies (GWAS) have uncovered genetic loci influencing metformin and sulfonylurea response but were largely performed in European populations with established disease

What is the key question?

- Can a genome-wide approach identify new pharmacogenetic associations and generate insight into the functional relevance of known genetic risk factors for type 2 diabetes in a multi-ethnic acute drug perturbation study of individuals at increased risk of type 2 diabetes?

What are the new findings?

- We identified novel genomic regions associated with acute metformin and glipizide response at genome-wide significance
- Several top findings were more common in participants of African ancestry, underscoring the importance of studying non-European populations
- Established type 2 diabetes and glycaemic trait loci were associated with differences in incretin levels in SUGAR-MGH with evidence of colocalisation; these findings provide further insight into incretin physiology as a potential mechanism by which these variants influence type 2 diabetes risk, with implications for use of incretin-based medication

How might this impact on clinical practice in the foreseeable future?

- Our study provides initial proof of concept for considering ancestry-specific genetic variation in the choice of pharmacotherapy for type 2 diabetes and advances precision medicine

Abbreviations

AOC	Area over the curve
DPP	Diabetes Prevention Program
EAF	Effect allele frequency
gePS	Global extended polygenic score
GLP-1	Glucagon-like peptide 1
GWAS	Genome-wide association study
LD	Linkage disequilibrium
MAF	Minor allele frequency
PC	Principal component
PP	Posterior probability
pPS	Process-specific polygenic score
SUGAR-MGH	Study to Understand the Genetics of the Acute Response to Metformin and Glipizide in Humans
V1	Visit 1
V2	Visit 2

Introduction

Treatment of type 2 diabetes currently follows a standard algorithm that begins with metformin [1], but involves the trial and error of additional drug regimens as the disease progresses. The choice of agent is based on several considerations, including an individual's comorbidities, the drug's side effect profile and costs of the therapy, but does not include information about the molecular target of the agent or genetic factors that might predict response or development of adverse effects [2]. The understanding of who responds best to each medicine is instrumental to furthering and optimising care of patients with diabetes.

Large-scale genome-wide association studies (GWAS) have identified over 700 genetic variants influencing type 2 diabetes risk and glycaemic traits. Data on how genetic variation influences response to glucose-lowering medications are starting to emerge. In individuals with established type 2 diabetes, GWAS have revealed novel loci for glycaemic response to metformin [3, 4]. With respect to sulfonylureas, candidate gene studies have uncovered genetic predictors of glycaemic response [5, 6] as well as sulfonylurea-induced hypoglycaemia [7, 8]. Recently, a GWAS of sulfonylurea response identified two independent loci associated with HbA_{1c} reduction [9]. Since the majority of pharmacogenetic studies have been conducted in those

with established type 2 diabetes, a genome-wide approach evaluating the response to metformin and sulfonylureas in a population at risk for developing type 2 diabetes has not previously been carried out.

Moreover, the functional relevance of many type 2 diabetes and glycaemic loci is not fully understood. The mechanisms leading to the development of type 2 diabetes are complex, both intrinsic and extrinsic to the beta cell [10]. For instance, an intronic variant in *TCF7L2* is the strongest common genetic risk factor for type 2 diabetes [11], yet multiple mechanisms have been proposed, including reduced beta cell mass, diminished insulin secretion and alterations in the incretin response [12]. In the Study to Understand the Genetics of the Acute Response to Metformin and Glipizide in Humans (SUGAR-MGH), we previously observed that an impaired incretin effect may contribute to the increased risk of type 2 diabetes in carriers of the high-risk allele at *TCF7L2* [5, 13].

In this study, we applied a genome-wide approach to comprehensively identify novel genetic predictors of acute metformin and glipizide response in individuals at risk of type 2 diabetes but naive to these medications. We examined the effects of known genetic variants associated with type 2 diabetes and glycaemic traits across all outcomes in SUGAR-MGH to gain further insights into the mechanisms by which they confer increased risk of type 2 diabetes or glycaemic dysregulation. Overall, we present and make available a resource for studying how genetic variation influences the biochemical response to two common glucose-lowering agents.

Methods

Study design and participants SUGAR-MGH is a pharmacogenetic study in which 1000 individuals who were naive to type 2 diabetes medications received a single-dose glipizide challenge and a short course of metformin [5, 14]. Participants at risk for diabetes, defined as participants with the metabolic syndrome, obesity, a history of gestational diabetes, a history of polycystic ovarian syndrome or a family history of type 2 diabetes, were preferentially enrolled. The rationale for selecting an at-risk population was twofold: (1) a recruitment strategy to increase participation; and (2) individuals with relatively intact beta cell function may have a more robust response to sulfonylureas and metformin. Figure 1 summarises the study design, which is described in detail in the electronic supplementary material (ESM) Methods. The study has been registered on ClinicalTrials.gov (NCT01762046) and is approved by the Mass General Brigham Human Research Committee Institutional Review Board (IRB).

Genotyping and imputation One thousand samples underwent genome-wide genotyping on the Multi-Ethnic Genotyping Array (Illumina, San Diego, CA, USA), which covers over 1.7 million genetic markers. A three-step quality control protocol was applied using PLINK 1.9 [15]. This included two stages of variant removal and an intermediate stage of sample exclusion. Variants were filtered for minor allele frequency (MAF) <0.01, low call rate <95% and failure to meet Hardy–Weinberg equilibrium within each self-described ancestry group ($p<5\times10^{-7}$). Samples were excluded for sex discrepancies, close relatedness (pairs with $\pi\hat{}$ [pi-hat]≥0.125, from which we removed the individual with the highest proportion of missingness) and call rate <98%. Phasing was performed using SHAPEIT2 [16]. Imputation was performed with the Michigan Imputation Server using the TOPMed reference panel [17]. After post-imputation quality control, excluding variants with imputation $R^2<0.8$ and MAF<0.005, ~12 million variants were available for analyses in 890 individuals. Genome annotations were generated using the GRCh38 assembly.

Endpoints of metformin and glipizide response As previously described [14], the primary endpoint of metformin response was defined as the fasting glucose at Visit 2 (V2), adjusted for fasting glucose at Visit 1 (V1). For the primary outcome of glipizide response, we selected the following closely related endpoints: insulin peak adjusted for baseline insulin, glucose trough adjusted for baseline glucose and time to glucose trough. We identified secondary outcomes of metformin and glipizide response based on measurements taken during the glipizide challenge and the 75 g OGTT following metformin (ESM Table 1), including insulin, incretin and homeostasis model assessments.

Genome-wide association analysis We performed genome-wide association analyses to assess the role of genetic variation in the acute response to metformin and glipizide. Multiple linear regression using an additive model tested for association between genetic variants and the primary endpoints, implemented using SNPTEST v2.5.4. Analyses were adjusted for age, sex, BMI and the first ten ancestry principal components (PCs) to account for population stratification. Quantitative traits were rank-inverse normalised to avoid spurious associations driven by outliers or skewed distributions and β estimates reflect rank-inverse normalisation. When relevant, we adjusted for the baseline trait at V1. Genome-wide significance was set at $p<5\times10^{-8}$ and an experiment-wide threshold was set at $p<2.5\times10^{-8}$, accounting for two drugs. Manhattan and quantile–quantile plots were produced with R (version 4.0) [18], and regional association plots were generated in LocusZoom [19] using the linkage disequilibrium (LD) reference panel for the ancestry that had the highest allele frequency for each variant.

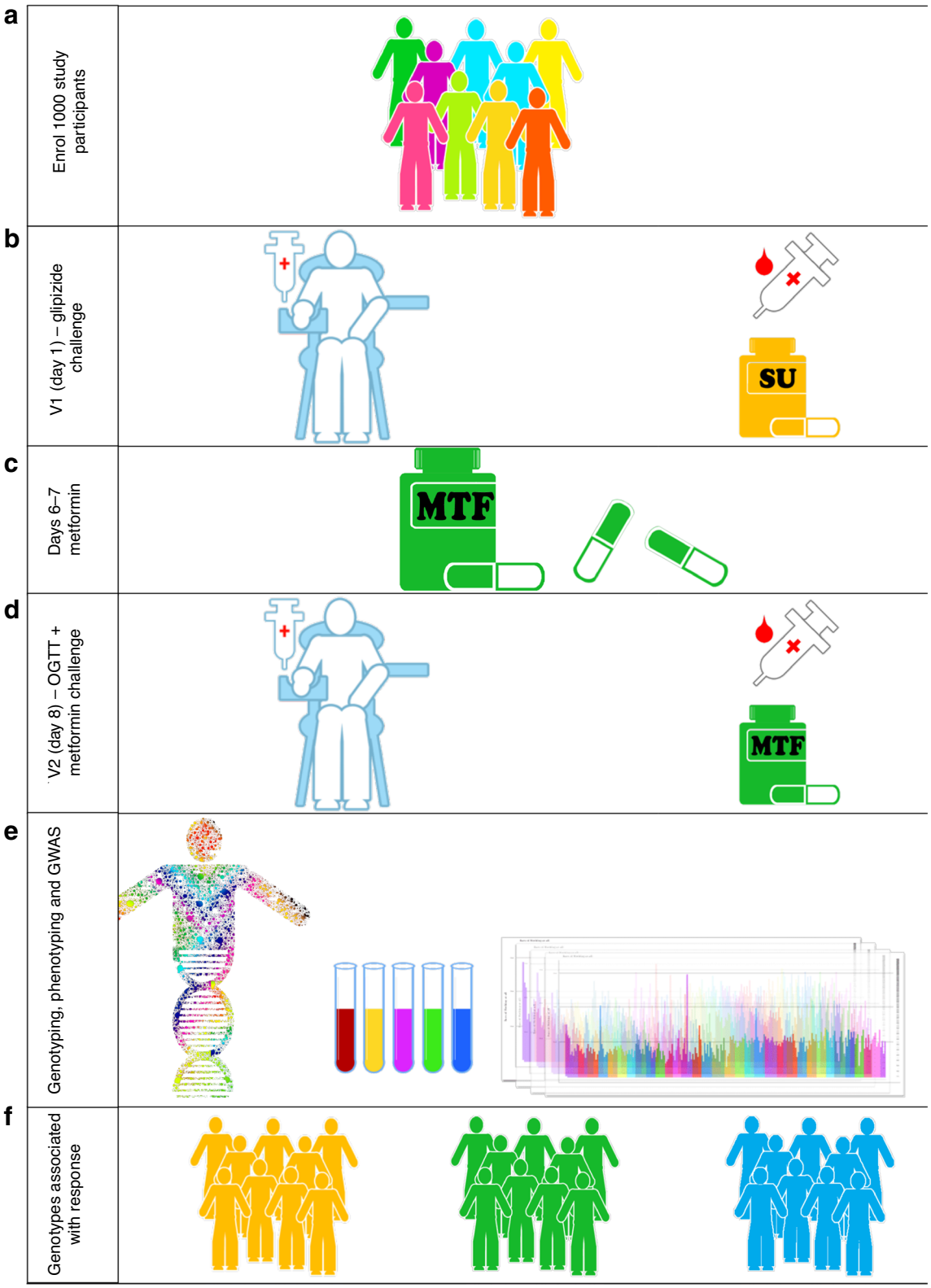

Fig. 1 Study schema. (**a**) We recruited 1000 individuals at risk of developing type 2 diabetes. (**b**) At V1, participants had their vital signs monitored, provided whole blood for DNA and underwent fasting measurements. Individuals with a fasting blood sugar >4.4 mmol/l received a dose of 5 mg of glipizide orally, followed by additional measurements. (**c**) After a 5 day wash-out period, participants received three doses of metformin of 500 mg. (**d**) At V2, participants returned for the final (fourth) dose of metformin and a 75 mg OGTT. (**e**) We performed genome-wide genotyping, constructed phenotypes of drug response and performed a GWAS, in order to (**f**) identify genotypes associated with outcomes of drug response. MTF, metformin; SU, sulfonylurea

In an exploratory analysis, we tested for the association of variants with the pre-defined secondary drug outcomes and assessed findings that reached both genome-wide significance for at least one trait, and suggestive significance threshold of $p<1\times10^{-6}$ for another trait. For top variants of interest, we examined their association with glucose and insulin curves during the glipizide challenge and the OGTT following metformin. Multiple linear regression assessed for differences in outcomes by genotype groups, adjusted for similar covariates.

We also assessed the association between previously reported genome-wide significant loci for type 2 diabetes and quantitative glycaemic phenotypes and all available traits in SUGAR-MGH. We evaluated 429 genetic variants associated with type 2 diabetes [20, 21] and 375 genetic variants associated with glycaemic traits [22]. We used an r^2 threshold of 0.5 to prune variants based on LD, using the full 1000 Genomes as a reference panel and LDlink [23], resulting in 563 independent effective markers. Based on their higher prior probability for glycaemic associations, we did not demand genome-wide significance in these analyses. While we corrected for the number of variants, we did not correct for the number of traits in SUGAR-MGH because they are highly correlated. The resultant threshold ($p<8.9\times10^{-5}$ [0.05/563]) was used to prioritise associations for which we proceeded with colocalisation analyses of the SUGAR-MGH trait and the relevant type 2 diabetes/glycaemic trait in order to confirm the presence of shared genetic risk factors (ESM Methods) [24].

We generated weighted global extended polygenic scores (gePSs) for type 2 diabetes, fasting glucose, fasting insulin and HbA_{1c}, based on summary statistics from published GWAS of type 2 diabetes and glycaemic traits [20–22]. To construct the gePS, we used PRS-CS using auto as a global shrinkage parameter [25]. We constructed five process-specific polygenic scores (pPSs) derived from physiologically driven clusters [26]. We tested these scores against the primary endpoints of metformin and glipizide response and set an experiment-wide significance threshold of $p<0.003$ to account for multiple comparisons (two drugs × nine polygenic scores). We adjusted for the same covariates as in the primary GWAS.

Replication of metformin response variants in the Diabetes Prevention Program We attempted to replicate the genome-wide significant variants associated with metformin response in the Diabetes Prevention Program (DPP), a multicentre randomised controlled trial that evaluated the impact of intensive lifestyle modification and pharmacologic intervention on development of type 2 diabetes in high-risk individuals [27, 28]. A GWAS of metformin response, defined as diabetes incidence and change in quantitative traits (fasting glucose, 2 h glucose on OGTT, HbA_{1c}, fasting insulin, insulin sensitivity index and weight), has been completed in the DPP [29]. The full study details of the GWAS completed in the DPP are described in the ESM Methods. For the replication of SUGAR-MGH findings, we tested the association of our top metformin findings with the 1 year change (follow-up minus baseline) in fasting glucose and HbA_{1c} in the metformin treatment arm only. Multiple linear regression using an additive model was performed, adjusting for baseline trait, age, sex and ten ancestry PCs. In the published GWAS completed in the DPP, a sensitivity analysis showed that GWAS findings did not change with additional adjustment for BMI, so it was not included in the model [29]. To account for multiple testing, we set a replication significance threshold based on two outcomes and the number of variants tested in replication.

Results

Participant characteristics Baseline demographics of the 890 participants with complete GWAS data are summarised (ESM Table 2). Approximately 53% of participants were female, the mean age was 47 years and 37% of participants self-reported as non-white. The mean BMI was 30.2 kg/m^2 and mean fasting glucose was 5.14 mmol/l, consistent with a population at risk for requiring future treatment of type 2 diabetes. The HOMA-B score at baseline was 91.3, comparable to that observed in healthy individuals without type 2 diabetes and normal fasting glucose [30, 31]. Of the 890 participants with genetic data, 20 participants did not receive the glipizide challenge due to a low baseline fasting glucose and 298 terminated the challenge early for hypoglycaemia, in accordance with study protocol.

Association of genetic variation with primary outcomes of drug response We identified five genome-wide significant variants associated with primary endpoints of acute metformin and glipizide response, four of which met experiment-wide significance of $p<2.5\times10^{-8}$ (Table 1). Three variants (rs149403252, rs111770298 and rs117207651) were associated with metformin response, as measured by fasting glucose at V2, adjusted for fasting glucose at V1; two variants (rs9954585 and rs150628520) were associated with glipizide response, as measured by the time to glucose trough. For each of the five variants, the allele counts by self-reported race/ethnicity are listed in ESM Table 3.

Among the variants associated with metformin response at genome-wide significance, rs149403252 (MAF_{Afr}=0.0283, $\beta=-1.3$, $p=1.9\times10^{-9}$) is an African ancestry-specific variant located in chromosome 3 near *ERC2* (ESM Fig. 1a). Carriers of the T effect allele had a lower fasting glucose at V2, adjusted for baseline glucose, indicating that they had an

Table 1 Genome-wide significant variants ($p<5\times10^{-8}$) associated with primary endpoints of acute metformin or glipizide response in SUGAR-MGH

rsID	Chr	Position[a]	Nearest gene	NEA	EA	EAF	*N*	Genotype counts[b]	AFR[c]	AMR[c]	EAS[c]	EUR[c]	SAS[c]	Trait	β[d]	*p* value
Metformin response																
rs149403252	3	55883717	*ERC2*	G	T	0.006	805	796/9/0	0.0283	0.0018	0	<0.0001	0.0002	Fasting glucose at V2, adjusted for V1	−1.3	1.9×10^{-9}
rs111770298	2	28307503	*BABAM2/FOSL2*	A	G	0.013	805	784/21/0	0.0536	0.0048	0	0.0001	0.0002	Fasting glucose at V2, adjusted for V1	0.8	2.4×10^{-8}
rs117207651	16	82250950	*MPHOSPH6*	T	C	0.009	805	794/11/0	0.0030	0.0082	0	0.0158	0.0008	Fasting glucose at V2, adjusted for V1	−1.0	4.5×10^{-8}
Glipizide response																
rs9954585	18	56245092	*TXNL1*	C	T	0.013	550	536/14/0	0.0602	0.0052	0.0002	0.0010	0.0008	Time to reach glucose trough at V1	−1.5	7.0×10^{-9}
rs150628520	4	187296094	*FAT1*	A	G	0.009	550	543/7/0	0.0020	0.0068	0.0002	0.0114	0.0025	Time to reach glucose trough at V1	1.7	9.8×10^{-9}

[a]GRCh38 assembly

[b]Reported as the number of individuals carrying 0, 1 or 2 copies of the effect allele. For all imputed variants, fractional alleles were converted to hard calls for calculation of genotype counts

[c]Ancestry-specific allele frequencies as reported in gnomAD version 3.1.2

[d]β estimates are rank-inverse normalised. A negative β when evaluating metformin response implies that the effect allele is associated with an enhanced response to metformin, i.e. lower fasting glucose following metformin exposure. A negative β when evaluating glipizide response implies that the effect allele is associated with an enhanced response to glipizide, i.e. shorter time to reach glucose trough following glipizide exposure

AFR, African; AMR, admixed American; Chr, chromosome; EA, effect allele; EAF, effect allele frequency; EAS, East Asian; EUR, European; NEA, non-effect allele; rsID, reference SNP cluster identifier; SAS, South Asian

enhanced metformin response. This was particularly apparent when examining the change in fasting glucose (ESM Fig. 1b), in which heterozygous individuals had a decrease of 1.1 mmol/l after four doses of metformin compared with a decrease of 0.12 mmol/l in non-carriers (β of difference=−0.94 mmol/l [p=1.1×10^{-6}]). During the OGTT following metformin, heterozygous individuals had lower insulin AUC (p=0.005) despite statistically similar glucose AUC. Another African ancestry-specific genetic variant influencing metformin response was rs111770298 (MAF$_{Afr}$=0.0536, β=0.8 [p=2.4×10^{-8}]), located in an intron of *BABAM2* in chromosome 2 (Fig. 2). Carriers of the G allele had a reduced metformin response, as evidenced by a higher fasting glucose at V2, adjusted for baseline glucose at V1. We calculated that whereas individuals homozygous for the A (common) allele experienced a 0.15 mmol/l decrease in fasting glucose after metformin, heterozygous individuals had a 0.29 mmol/l increase (β of difference=0.43 mmol/l [p=9.4×10^{-7}]). Finally, rs117207651 near *MPHOSPH6* was associated with a better response to metformin (ESM Fig. 2): whereas TT individuals experienced a 0.13 mmol/l decrease in fasting glucose after metformin, TC individuals had a greater decrease of 0.50 mmol/l (β of difference=−0.50 mmol/l [p=1.8×10^{-4}]).

We attempted validation of our top three variants associated with metformin response in the DPP, a randomised controlled trial of lifestyle intervention or pharmacologic therapy (metformin) conducted in individuals with impaired glucose tolerance at high risk for developing type 2 diabetes. The outcomes examined were changes in fasting glucose and HbA$_{1c}$ after 1 year of follow-up. We set a replication significance threshold of p<0.008 (two outcomes × three variants). Results are summarised in ESM Table 4. rs111770298 was significantly associated with worse metformin response, where heterozygous carriers experienced a 0.08% increase in the 1 year change in HbA$_{1c}$ after 1 year of metformin treatment, compared with an increase of 0.01% in non-carriers

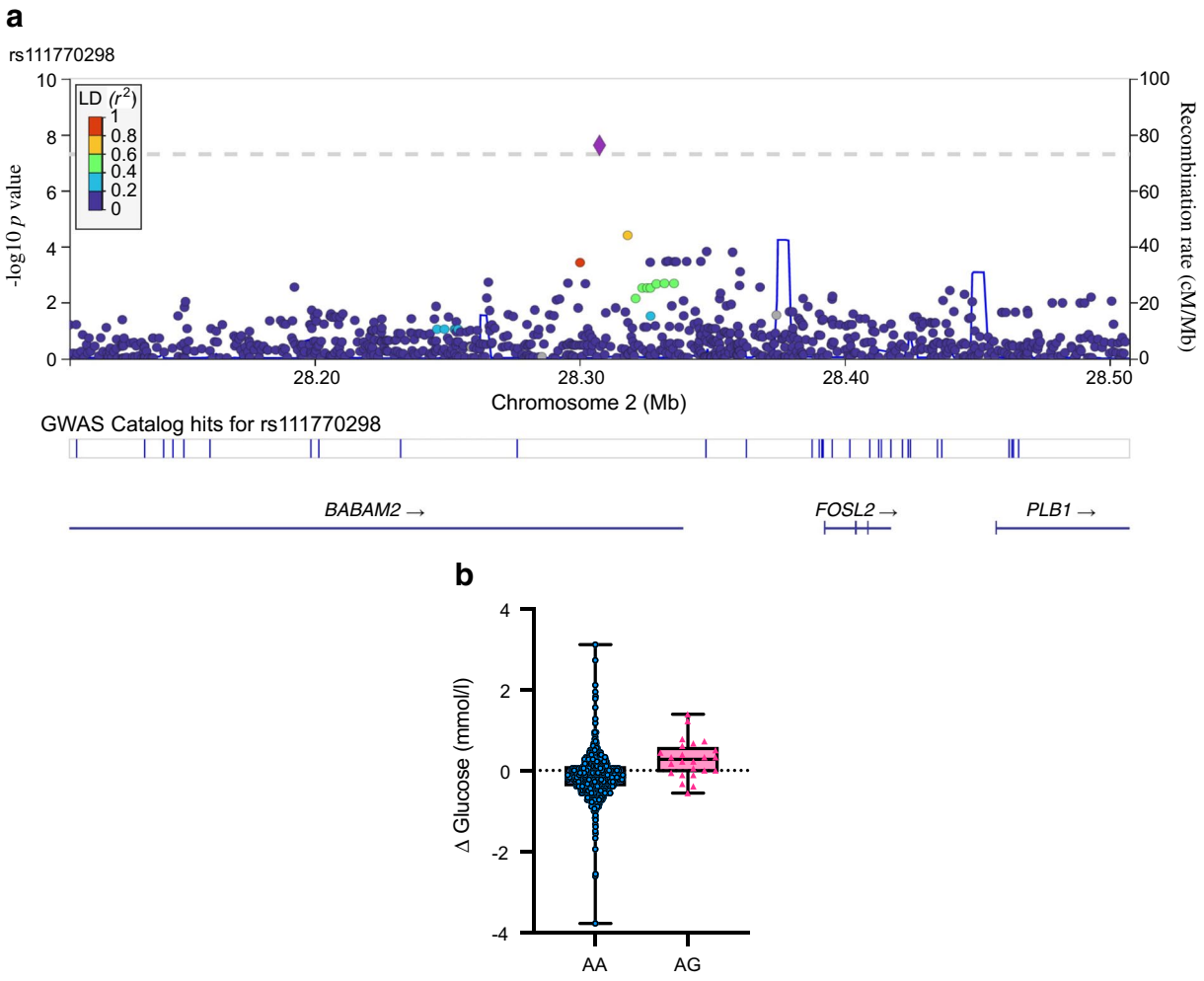

Fig. 2 (**a**) Regional association plot of rs111770298. (**b**) Box plot illustrating mean change in fasting glucose (V2 minus V1) by rs111770298 genotype

(p=3.3×10^{-3}), thus confirming our findings in SUGAR-MGH. For rs149403252 and rs117207651, we did not replicate our findings in the DPP.

For the glipizide challenge, the strongest glipizide-associated variant was rs9954585, near *TXNL1* in chromosome 18 (ESM Fig. 3). Being a carrier of the T allele was associated with a shorter time to glucose trough (EAF=0.013, β=−1.5, p=7.0×10^{-9}). In addition, rs150628520, a low-frequency variant near *FAT1* in chromosome 4 (ESM Fig. 4a), was associated with increased time to glucose trough, consistent with a diminished glipizide response (EAF=0.009, β=1.7, p=9.8×10^{-9}). In agreement with this, carriers of the G allele had a significantly decreased cumulative drop in glucose, measured by glucose area over the curve (AOC) (ESM Fig. 4b, p=0.004), as well as a decreased insulin AUC (ESM Fig. 4c, p=0.006).

Association of genetic variation with secondary outcomes of drug response Beyond the primary outcomes, we examined associations reaching genome-wide significance (p<5×10^{-8}) for either the primary or secondary outcomes (ESM Table 1). We curated a list of variants that were additionally associated with at least one other secondary outcome at a suggestive p<1×10^{-6} and in consistent direction of effect (e.g. both associations pointing toward enhanced metformin response). ESM Table 5 describes the resultant set of markers that met these criteria and were therefore considered to be likely true associations with metformin or glipizide response. Notably, two of the primary GWAS findings (rs150628520 and rs111770298) were also associated with secondary outcomes (ESM Table 5).

Evaluation of known type 2 diabetes and glycaemic variation and polygenic scores in SUGAR-MGH We next focused on assessing the pharmacological response to variants previously associated with type 2 diabetes and quantitative glycaemic traits. Within the associations meeting our screening threshold (p<8.9×10^{-5}) for colocalisation analysis, we were able to confirm through colocalisation analyses that ten of them share the same genetic signal between a SUGAR-MGH outcome and type 2 diabetes/glycaemic traits with a posterior probability (PP) of ≥75% (ESM Table 6). As an example, we found that the type 2 diabetes-protective C allele of rs703972 near *ZMIZ1* was associated with increased levels of active glucagon-like peptide 1 (GLP-1) (p=1.6×10^{-5}), with high evidence of colocalisation (PP=90.3%, Fig. 3a–c).

When we evaluated the influence of polygenic scores on drug response, we confirmed our previously reported nominal association between a higher type 2 diabetes polygenic score and greater glucose AOC, representing a greater cumulative drop in glucose following glipizide (p=0.02) [32]. In addition, we observed an association between the fasting glucose gePS and the primary outcome for metformin response meeting experiment-wide significance (p=0.002): after adjusting for baseline glucose, individuals with higher fasting glucose gePS had a 0.03 mmol/l lower drop in fasting glucose after metformin per SD increase in polygenic score, consistent with a worse metformin response. In a subgroup analysis, this was found to be driven by individuals who began the study with a fasting glucose over 5 mmol/l, who experienced a mean drop of 0.07 mmol/l (p=0.04). No associations between any of the pPSs and metformin or glipizide response met experiment-wide significance (ESM Table 7).

Discussion

SUGAR-MGH is a pharmacogenetic resource for characterising genetic influences on pharmacological perturbations relevant to type 2 diabetes. In prior work, SUGAR-MGH has contributed to the understanding of the influence of *TCF7L2* and *CYP2C9*, as well as a restricted-to-significant (i.e. using only genome-wide significant variants) type 2 diabetes polygenic risk score, on drug response [5, 8, 32]. With the completion of genome-wide genotyping, we report new genetic associations with acute metformin and sulfonylurea response in an ancestrally diverse population at risk for type 2 diabetes and naive to commonly prescribed glucose-lowering medications.

We identified three variants that were significantly associated with acute response to metformin, of which two were African ancestry-specific variants. The strongest association was between rs149403252, an intronic variant located in *ERC2*, and lower fasting glucose following metformin, but unfortunately this finding did not replicate. *ERC2* encodes a protein in the CAZ-associated structural protein (CAST) family, which has been implicated in the calcium-dependent exocytosis of neurotransmitters [33]; one family member is present in pancreatic beta cells and may be involved in the regulation of insulin secretion [34]. More robustly, rs111770298 was associated with both a reduced response to metformin in SUGAR-MGH, as measured by a higher fasting glucose after metformin, and a rise in HbA_{1c} in independent replication in the DPP. rs111770298 is an intronic variant located near *BABAM2* and *FOSL2*, the latter of which has been shown to promote leptin gene expression in mouse adipocytes [35]. Rare coding and common variants in or near *FOSL2* are associated with lower triglyceride levels [36]. In the Type 2 Diabetes Knowledge Portal [36], this variant has a nominal association with diastolic blood pressure. Fine-mapping analyses, phenome-wide association analyses and functional experiments will be needed to confirm the implication of these loci in metformin response.

We uncovered several promising variants of interest for glipizide response. T allele carriers at rs9954585 have a shorter time to glucose trough, indicating a more robust

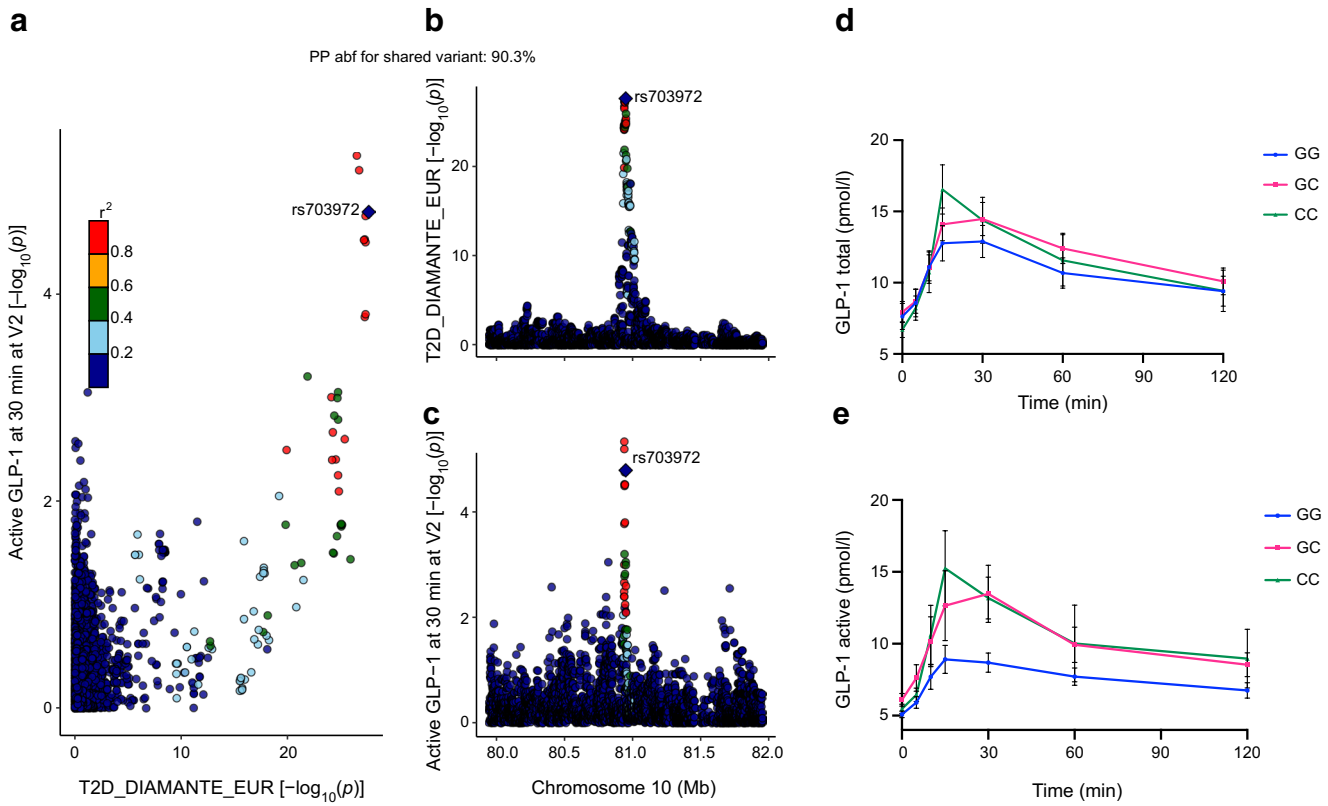

Fig. 3 (**a**) Colocalisation plot showing that rs703972 near *ZMIZ1* colocalises with active GLP-1 and type 2 diabetes risk. (**b**, **c**) LocusZoom plots of association of rs703972 with (**b**) type 2 diabetes and (**c**) active GLP-1. (**d**) Change in total and (**e**) active GLP-1 by rs703972 at V2 during OGTT. abf, approximate Bayes factor; T2D_DIAMANTE_EUR, GWAS results in European population from Diabetes, Meta-Analysis Trans-Ethnic Consortium

response to glipizide. G allele carriers at rs150628520 near *FAT1* appear to have an attenuated response to glipizide, having not only an increased time to glucose trough but also a more gradual slope to glucose trough. We also note that the presence of concordant associations across multiple primary and secondary outcomes for glipizide response provides support for our genetic findings. However, it is unclear whether the observed differences by genotype are due to a decrease in glipizide action or impairment in glipizide absorption. To further elucidate the mechanisms responsible for these effects, future directions include quantifying glipizide drug levels, and comparing carriers and non-carriers at these loci.

We tried to characterise the biological mechanisms of known type 2 diabetes and glycaemic loci by leveraging the phenotypic outcomes constructed in this physiological study. We identified established genome-wide significant type 2 diabetes and glycaemic variation that met our screening criteria for association with highly correlated traits in SUGAR-MGH and proceeded with colocalisation analysis to confirm the presence of shared genetic risk factors. We demonstrated that the protective C allele of rs703972 near *ZMIZ1* was associated with increased levels of active GLP-1. Interestingly, *ZMIZ1* has been previously reported to play a role in regulation of beta cell function, with expression of *ZMIZ1* reducing insulin secretion [37]. Thus, an augmented incretin response may explain how C allele carriers are able to mitigate their type 2 diabetes risk. Incretins have been implicated in the pathophysiology of type 2 diabetes; however, it is unknown whether the incretin effect is impaired due to a reduction in functional beta cell mass or due to a defect in incretin action leading to resistance [38]. Our findings provide support for altered incretin physiology in the pathogenesis of type 2 diabetes and shed additional insight on a potential mechanism underlying the effect of the *ZMIZ1* variant. This example demonstrates that our research can be used to determine endophenotypes of already known genetic associations and can serve as a useful resource for characterisation of future associations.

Previously, we reported that a higher type 2 diabetes polygenic score of 65 variants was associated with several measures of glipizide response at nominal significance [32], but we did not identify any associations with phenotypes of metformin response. With the availability of genome-wide genotyping and access to full summary statistics from larger meta-analyses for type 2 diabetes and glycaemic traits, we expanded our analysis to incorporate large numbers of sub-significant variants across the genome. With a type 2 diabetes gePS, we confirmed the

previous association between a higher genetic burden for type 2 diabetes and greater glucose AOC, indicating an enhanced response to glipizide at this early stage of dysglycaemia. This is consistent with the findings of Dennis et al in the A Diabetes Outcome Progression Trial (ADOPT), in which participants with a phenotypically defined severe insulin-deficient diabetes demonstrated a robust early response to sulfonylureas, which was attenuated as their beta cell function deteriorated over time [39]. Moreover, we found that individuals with a greater burden of risk variants for higher fasting glucose, possibly representing a genetic susceptibility for lower beta cell function, had a diminished response to metformin. Our ability to detect this pharmacogenetic association was likely bolstered by the vast increase in the number of variants included in the polygenic score and may have clinical implications for the effectiveness of metformin as a first-line therapy in those genetically predisposed to fasting dysglycaemia. While we hypothesised that physiologically derived clusters related to the drug's mode of action may have an influence on the acute drug response (i.e. association between beta cell function clusters and glipizide response), we did not detect associations with primary outcomes of metformin or glipizide response, possibly due to lower statistical power of the pPSs, which comprise a smaller number of variants. Given the increasing availability of genotype information, future studies are needed to validate the utility and predictive value of polygenic scores for drug response.

Our study is the first GWAS of acute metformin and glipizide response including participants at risk of type 2 diabetes from multiple ancestries. In contrast to existing type 2 diabetes pharmacogenetic GWAS performed in European populations [3, 4, 9], over a third of SUGAR-MGH participants were of non-European descent. The value of analysing cohorts that span multiple ancestries is exemplified by the identification of novel associations in genetic variants that are more prevalent in non-European populations. Several of our genome-wide significant findings (rs149403252 near *ERC2*, rs9954585 near *TXNL1* and rs111770298 near *BABAM2/FOSL2*) had minor allele frequencies that were common to low frequency in African populations and rare in European populations. Associations near these genes have not previously been identified as related to type 2 diabetes risk or response to glucose-lowering medications, which may be due to the dearth of studies in non-European populations. Understanding the impact of such ancestry-specific variants may guide treatment decisions for type 2 diabetes in these population subgroups in the future, but also provide drug targets suitable for all ancestries. One major barrier to translating ancestry-specific variants to their function is the lack of ancestry-specific genetic and genomic data. For instance, the Genotype-Tissue Expression (GTEx) project largely contains individuals with European ancestry [40], limiting our ability to characterise the effects of genome-wide significant variants not present or at low frequency in Europeans on the transcriptome across human tissues. Similarly, the lack of phenome-wide association data on diverse ancestries hinders follow-up of identified variants. Expansion of existing datasets to include non-European populations will be valuable for linking pharmacogenetic associations to functional mechanisms.

Given the global dearth of pharmacogenomic GWAS, especially those conducted in non-European populations, one major challenge of this work was identifying a suitable replication venue. Due to the unique characteristics of this study examining acute drug response, no comparable replication venue was readily available. However, we sought replication in the DPP, a study of individuals with prediabetes who received longitudinal metformin exposure. We illustrated that the influence of rs111770298 on 1 year change in HbA_{1c} validated our findings for fasting glucose in SUGAR-MGH, with a consistent direction of effect on metformin response. However, we did not observe a differential impact of this same variant on 1 year change in fasting glucose in the DPP. One explanation is that the variant's effect on fasting glucose might be more pronounced and better detected initially in the acute setting; perhaps in the DPP, the long-term effect is better captured by average blood glucose levels as measured by change in HbA_{1c}.

We also recognise that our study examined those at risk of type 2 diabetes, and it is unclear whether our findings would have the same magnitude of effect in people with overt or long-standing type 2 diabetes, as disease stage may affect the metabolic state of a person who carries the same genetic profile. For example, variants in drug transporter genes that influence response to metformin in healthy individuals [41] were not found to affect HbA_{1c} in people with type 2 diabetes [42]. Another limitation is that the study design did not incorporate a baseline OGTT, which limited our ability to assess the impact of metformin on a dynamic glucose challenge. This was due to the financial and time constraints of enrolling participants for an additional OGTT. A final limitation is that our sample size was small for measurements of incretin levels, which restricted our ability to detect additional findings relevant to incretin physiology.

In summary, we identified novel genetic variation in a multi-ethnic human drug perturbation study which requires validation in ancestry-specific cohorts but has the potential to influence the selection of glucose-lowering medications in specific populations. We demonstrated the utility of our pharmacogenetic resource for understanding the underlying mechanisms of known genetic variation for type 2 diabetes and glycaemic traits. Beyond the primary drug endpoints, we created a public resource to permit the organisation and sharing of genetic association results across a wide variety of traits in SUGAR-MGH, which can be used as a validation cohort for future pharmacogenetic discoveries by others as well as for functional characterisation of newly identified genes implicated in the pathogenesis of type 2 diabetes.

Supplementary Information The online version contains peer-reviewed but unedited supplementary material available at https://doi.org/10.1007/s00125-023-05922-7.

Acknowledgements We would like to thank J. Del Rio (freelance graphic designer, Cambridge, MA, USA) for his support with Fig. 1 generation. Portions of this study were previously presented as an oral presentation at the 81st Virtual Scientific Sessions of the American Diabetes Association, 25–29 June 2021.

Data availability The complete summary statistics from this study will be deposited and made available at the Common Metabolic Diseases Knowledge Portal (https://hugeamp.org) and the GWAS Catalog (www.ebi.ac.uk/gwas/, accession IDs: GCST90269867 to GCST90269899) following article publication. Related study documents, including the original study protocol and informed consent forms, are available [14]. Additional data requests should be sent by email to the corresponding author.

Funding This work was conducted with support from National Institutes of Health/NIDDK awards R01 GM117163, R01 DK088214, R03 DK077675 and P30 DK036836; from the Joslin Clinical Research Center from its philanthropic donors; and from the Harvard Catalyst: the Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH Awards M01-RR-01066, 1 UL1 RR025758-04 and 8UL1TR000170-05, and financial contributions from Harvard University and its affiliated academic healthcare centres). JHL received individual support from NIH T32DK007028 and NIDDK K23DK131345. LNB is supported by NIDDK K23DK125839. MSU is supported by NIDDK K23DK114551. AL is supported by grant 2020096 from the Doris Duke Charitable Foundation and the American Diabetes Association grant 7-22-ICTSPM-23. JMM is supported by American Diabetes Association Innovative and Clinical Translational Award 1-19-ICTS-068, American Diabetes Association grant #11-22-ICTSPM-16, and by NHGRI U01HG011723. JCF is supported by NHLBI K24HL157960.

Authors' relationships and activities The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement All authors took part in designing the experiments presented in this manuscript. VK, LNB, MSU, AL and JCF recruited participants in SUGAR-MGH. VK supervised participant recruitment, data collection, and IRB review and approval, and performed DNA extractions and managed GWAS genotyping. Quality control, imputation of the genetic data and GWAS analyses were performed by JMM. JHL, LNB, VK, KF, PS, AH-C and JMM performed follow-up of GWAS data analysis. JHL, LNB, VK, JMM and JCF contributed to the interpretation of the results. JHL, LNB, VK and JMM wrote and prepared the manuscript. All authors revised and approved the final manuscript. JMM and JCF jointly supervised this study. JCF is the guarantor of this work.

References

1. Davies MJ, D'Alessio DA, Fradkin J et al (2018) 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 41(12):2669–2701. https://doi.org/10.2337/dci18-0033
2. Chung WK, Erion K, Florez JC et al (2020) Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 43(7):1617–1635. https://doi.org/10.2337/dci20-0022
3. Zhou K, Yee SW, Seiser EL et al (2016) Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. Nat Genet 48(9):1055–1059. https://doi.org/10.1038/ng.3632
4. Zhou K, Bellenguez C, Spencer CC et al (2011) Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nat Genet 43(2):117–120. https://doi.org/10.1038/ng.735
5. Srinivasan S, Kaur V, Chamarthi B et al (2018) TCF7L2 Genetic variation augments incretin resistance and influences response to a sulfonylurea and metformin: the study to understand the genetics of the acute response to Metformin and Glipizide in Humans (SUGAR-MGH). Diabetes Care 41(3):554–561. https://doi.org/10.2337/dc17-1386
6. Zhou K, Donnelly L, Burch L et al (2010) Loss-of-function CYP2C9 variants improve therapeutic response to sulfonylureas in type 2 diabetes: a Go-DARTS study. Clin Pharmacol Ther 87(1):52–56. https://doi.org/10.1038/clpt.2009.176
7. Dujic T, Zhou K, Donnelly LA, Leese G, Palmer CNA, Pearson ER (2018) Interaction between variants in the CYP2C9 and POR genes and the risk of sulfonylurea-induced hypoglycaemia: A GoDARTS Study. Diabetes Obes Metab 20(1):211–214. https://doi.org/10.1111/dom.13046
8. Chen L, Li JH, Kaur V et al (2020) The presence of two reduced function variants in CYP2C9 influences the acute response to glipizide. Diabet Med 37(12):2124–2130. https://doi.org/10.1111/dme.14176
9. Dawed AY, Yee SW, Zhou K et al (2021) Genome-wide meta-analysis identifies genetic variants associated with glycemic response to sulfonylureas. Diabetes Care 44(12):2673–2682. https://doi.org/10.2337/dc21-1152
10. Xue A, Wu Y, Zhu Z et al (2018) Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun 9(1):2941. https://doi.org/10.1038/s41467-018-04951-w
11. Florez JC, Jablonski KA, Bayley N et al (2006) TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. N Engl J Med 355(3):241–250. https://doi.org/10.1056/NEJMoa062418
12. Lyssenko V, Lupi R, Marchetti P et al (2007) Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest 117(8):2155–2163. https://doi.org/10.1172/JCI30706
13. Villareal DT, Robertson H, Bell GI et al (2010) TCF7L2 variant rs7903146 affects the risk of type 2 diabetes by modulating incretin action. Diabetes 59(2):479–485. https://doi.org/10.2337/db09-1169
14. Walford GA, Colomo N, Todd JN et al (2015) The study to understand the genetics of the acute response to metformin and glipizide in humans (SUGAR-MGH): design of a pharmacogenetic resource for type 2 diabetes. PloS One 10(3):e0121553. https://doi.org/10.1371/journal.pone.0121553
15. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8
16. Delaneau O, Zagury J-F, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 10(1):5–6. https://doi.org/10.1038/nmeth.2307
17. Taliun D, Harris DN, Kessler MD et al (2021) Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590(7845):290–299. https://doi.org/10.1038/s41586-021-03205-y
18. R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria
19. Pruim RJ, Welch RP, Sanna S et al (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England) 26(18):2336–2337. https://doi.org/10.1093/bioinformatics/btq419

20. Vujkovic M, Keaton JM, Lynch JA et al (2020) Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. Nat Genet 52(7):680–691. https://doi.org/10.1038/s41588-020-0637-y
21. Mahajan A, Spracklen CN, Zhang W et al (2022) Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 54(5):560–572. https://doi.org/10.1038/s41588-022-01058-3
22. Chen J, Spracklen CN, Marenne G et al (2021) The trans-ancestral genomic architecture of glycemic traits. Nat Genet 53(6):840–860. https://doi.org/10.1038/s41588-021-00852-9
23. Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics (Oxford, England) 31(21):3555–3557. https://doi.org/10.1093/bioinformatics/btv402
24. Giambartolomei C, Zhenli Liu J, Zhang W et al (2018) A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics (Oxford, England) 34(15):2538–2545. https://doi.org/10.1093/bioinformatics/bty147
25. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW (2019) Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 10(1):1776. https://doi.org/10.1038/s41467-019-09718-5
26. Udler MS, Kim J, von Grotthuss M et al (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med 15(9):e1002654. https://doi.org/10.1371/journal.pmed.1002654
27. The Diabetes Prevention Program Research Group (2000) The Diabetes Prevention Program: baseline characteristics of the randomized cohort. Diabetes Care 23(11):1619–1629. https://doi.org/10.2337/diacare.23.11.1619
28. The Diabetes Prevention Program (1999) Design and methods for a clinical trial in the prevention of type 2 diabetes. Diabetes Care 22(4):623–634. https://doi.org/10.2337/diacare.22.4.623
29. Li JH, Perry JA, Jablonski KA et al (2022) Identification of genetic variation influencing metformin response in a multi-ancestry genome-wide association study in the Diabetes Prevention Program (DPP). Diabetes. https://doi.org/10.2337/db22-0702
30. Sung K-C, Reaven GM, Kim SH (2009) Utility of homeostasis model assessment of β-cell function in predicting diabetes in 12,924 healthy Koreans. Diabetes Care 33(1):200–202. https://doi.org/10.2337/dc09-1070
31. Song Y, Manson JE, Tinker L et al (2007) Insulin sensitivity and insulin secretion determined by homeostasis model assessment and risk of diabetes in a multiethnic cohort of women: the Women's Health Initiative Observational Study. Diabetes Care 30(7):1747–1752. https://doi.org/10.2337/dc07-0358
32. Li JH, Szczerbinski L, Dawed AY et al (2021) A polygenic score for type 2 diabetes risk is associated with both the acute and sustained response to sulfonylureas. Diabetes 70(1):293–300. https://doi.org/10.2337/db20-0530
33. Ohara-Imaizumi M, Aoyagi K, Ohtsuka T (2019) Role of the active zone protein, ELKS, in insulin secretion from pancreatic β-cells. Mol Metab 27(Suppl):S81–S91. https://doi.org/10.1016/j.molmet.2019.06.017
34. Omar-Hmeadi M, Idevall-Hagren O (2021) Insulin granule biogenesis and exocytosis. Cell Mol Life Sci 78(5):1957–1970. https://doi.org/10.1007/s00018-020-03688-4
35. Wrann CD, Eguchi J, Bozec A et al (2012) FOSL2 promotes leptin gene expression in human and mouse adipocytes. J Clin Invest 122(3):1010–1021. https://doi.org/10.1172/jci58431
36. Accelerating Medicines Partnership. T2D Knowledge Portal. Available from https://t2d.hugeamp.org/region.html?chr=2&end=28690184&phenotype=TG&start=28565315. Accessed 21 Sep 2022
37. Thomsen SK, Ceroni A, van de Bunt M et al (2016) Systematic functional characterization of candidate causal genes for type 2 diabetes risk variants. Diabetes 65(12):3805–3811. https://doi.org/10.2337/db16-0361
38. Ruetten H, Gebauer M, Raymond RH et al (2018) Mixed meal and intravenous L-arginine tests both stimulate incretin release across glucose tolerance in man: lack of correlation with β cell function. Metab Syndr Relat Disord 16(8):406–415. https://doi.org/10.1089/met.2018.0022
39. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT (2019) Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. Lancet Diabetes Endocrinol 7(6):442–451. https://doi.org/10.1016/s2213-8587(19)30087-7
40. The GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (New York, NY) 369(6509):1318–1330. https://doi.org/10.1126/science.aaz1776
41. Shu Y, Brown C, Castro RA et al (2008) Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics. Clin Pharmacol Ther 83(2):273–280. https://doi.org/10.1038/sj.clpt.6100275
42. Zhou K, Donnelly LA, Kimber CH et al (2009) Reduced-function SLC22A1 polymorphisms encoding organic cation transporter 1 and glycemic response to metformin: a GoDARTS study. Diabetes 58(6):1434–1439. https://doi.org/10.2337/db08-0896

Authors and Affiliations

Josephine H. Li[1,2,3,4] · Laura N. Brenner[1,3,4,5] · Varinderpal Kaur[1,2,3] · Katherine Figueroa[1,3] · Philip Schroeder[1,2,3] · Alicia Huerta-Chagoya[1,2,3] · MAGIC Investigators · Diabetes Prevention Program (DPP) Research Group · Miriam S. Udler[1,2,3,4] · Aaron Leong[1,2,3,4,6] · Josep M. Mercader[1,2,3,4] · Jose C. Florez[1,2,3,4]

✉ Jose C. Florez
jcflorez@mgh.harvard.edu

[1] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[2] Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA

[3] Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[4] Harvard Medical School, Boston, MA, USA

[5] Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston, MA, USA

[6] Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, USA