Leveraging an Electronic Health Record-Linked Biorepository to Generate a Metformin Pharmacogenomics Hypothesis

Matthew K. Breitenstein, PhD[1,2], Liewei Wang, MD, PhD[1], Gyorgy Simon, PhD[2], Euijung Ryu, PhD[1], Sebastian M. Armasu, MS[1], Balmiki Ray, MBBS[1], Richard M. Weinshilboum, MD[1], Jyotishman Pathak, PhD[1]
[1]Mayo Clinic, Rochester, MN; [2]University of Minnesota, Minneapolis, MN

Abstract

Metformin is a first-line antihyperglycemic agent commonly prescribed in type 2 diabetes mellitus (T2DM), but whose pharmacogenomics are not clearly understood. Further, due to accumulating evidence highlighting the potential for metformin in cancer prevention and treatment efforts it is imperative to understand molecular mechanisms of metformin. In this electronic health record(EHR)-based study we explore the potential association of the flavin-containing monooxygenase(FMO)-5 gene, a biologically plausible biotransformer of metformin, and modifying glycemic response to metformin treatment. Using a cohort of 258 T2DM patients who had new metformin exposure, existing genetic data, and longitudinal electronic health records, we compared genetic variation within FMO5 to change in glycemic response. Gene-level and SNP-level analysis identified marginally significant associations for FMO5 variation, representing an EHR-driven pharmacogenetics hypothesis for a potential novel mechanism for metformin biotransformation. However, functional validation of this EHR-based hypothesis is necessary to ascertain its clinical and biological significance.

Introduction

Metformin is a first-line antihyperglycemic agent commonly prescribed for type 2 diabetes mellitus (T2DM) patients[1], whose pharmacogenomics are not clearly understood[2], but are thought to be absent of biotransformation[3]. Further, glycemic response to metformin is variable[3] and serious adverse reactions to metformin have been known to occur[4]. Due to increasing evidence highlighting the potential for metformin in cancer prevention and treatment, it is imperative to understand molecular mechanisms of metformin further.

Background

Metformin is primarily utilized to regain glycemic control in diabetic or pre-diabetic patients. Metformin is a relatively safe antidiabetic therapy[5]. However, serious adverse reactions can occur[4] and there is considerable variation in glycemic response to metformin, with ~30% of patients unable to achieve glycemic control with metformin[3]. While genetic factors may partially explain clinical glycemic response to metformin due to pharmacokinetic(PK) determinants[3], the transportation throughout the body variation, the identification and impact of metformin pharmacodynamic(PD) determinants, the physiological and biochemical impact of metformin in the body, remains uncertain[2]. Regarding PKs, Metformin is thought to not be metabolized[3], with absorption of metformin known to occur in the small and large intestines[5]. Uptake of metformin from the blood is known to occur in the kidneys and liver[2], but can be reasonably assumed to occur in any tissue with abundance of organic cation transporters (OCT). Eventually metformin is excreted unchanged in the urine[5]. Regarding PDs, metformin works primarily by inhibiting hepatic glucose production by reducing gluconeogenesis in the liver[6] and is also known to reduce intestinal glucose absorption[7]. Further, metformin appears to improve glucose uptake and utilization systemically[3].

Metformin is a nitrogen-rich biguanide. Flavin-containing monooxygenases(FMO)-5 has demonstrated narrow substrate specificity, but has been known to catalyze oxygenation of nitrogen-containing drugs[8]. FMO5 is expressed in the kidneys and liver[8]. The FMO5 gene exists near PRKAB2, a known PD regulator of metformin response, away from the single gene cluster for the remaining FMOs in chromosome 1q23-q25 region. Metformin is excreted unchanged in the urine[5], hinting that metformin does not undergo biotransformation. However, studies such as these do not produce 100% yield, hinting at room for deviation from this paradigm. While metformin is thought to be absent of biotransformation[3], it is biologically plausible that FMO5 might carry out N-oxygenation of metformin.

FMOs show overlapping substrate specificity among family members[8]; a signal corresponding to FMO5 might also correspond to an additional FMO gene. All FMOs contain eight coding exons that share 50 to 80% sequence identity, with mutant FMOs are known to react to alternative chemical sites[9]. FMOs are localized in the endoplasmic

reticulum of the cell whose expression is tissue-specific[8]. The extent of which reactions are catalyzed by FMOs in vivo cannot be determined by measuring end products excreted in bile or urine[10].

The primary purpose of this study was to add clarity to metformin pharmacogenomics by understanding the impact of common variants in the FMO5 gene on altered glycemic response in a clinical population derived from an EHR-linked biorepository. Due to some shared functional similarity among genes in the FMO gene family, we selected the remaining FMO genes (FMO1 – FMO4) as exploratory gene candidates as our secondary hypothesis.

Methods

In this EHR-linked genetic study, both the approaches for obtaining clinical phenotypes and genotypes had important considerations for both study design and study interpretation. Our primary hypothesis of interest holds that genetic variation within FMO5 has potential to modify glycemic response to metformin monotherapy. Secondary to the primary hypothesis is an exploratory hypothesis that posits similar potential associations for FMO1 – FMO4 due to functional similarity[8]. However, their function is not identical. Further, due to the close proximity of the FMO1 – FMO4 to each other and their relative distance from FMO5 on chromosome 1q21 our secondary hypothesis is considerably weaker than our primary hypothesis for FMO5. In this study, we utilized the longitudinal EHR at Mayo Clinic and genome-wide association study (GWAS) data from the subjects enrolled in the Mayo Genome Consortia[11].

Clinical Phenotypes

The application of EHR-based phenotypes dramatically impacts study design and interpretability of findings. In this study we had 4 key phenotype aspects to consider: 1) T2DM phenotype, 2) metformin exposure phenotype, and 3) change in A1c. First, attribution of a T2DM phenotype was performed using a modified methodology developed by eMERGE[12]. A key point of differentiation is that our T2DM phenotype relied on diagnosis codes and did not initially consider laboratory values or medication. However, our second and third considerations relied on lab values and medication exposure events that were more specific than the criteria for the eMERGE T2DM phenotype algorithm. Second, our metformin exposure period was designated as a new prescription of metformin that extended ≥6 months to ensure adequate primary care visits, multiple A1c measures, and maintenance dose achievement. Since our study aimed to understand genomic variation in relation to patients who respond or do not respond to metformin, maintenance dose was not a consideration. To accurately populate this metformin exposure phenotype our study design required longitudinal data access from primary care patients. Specifically, study inclusion criteria required ≥1 year of patient history and ≥2 primary care visits to ensure accurate capture of the first date of metformin exposure, which aimed to exclude patients that were false positives for a new recorded exposure to metformin due to medication reconciliation that occurred at transfer of primary care. Metformin exposure events were ascertained using a combination of validated structured and semi-structured EHR data collection methodologies that leveraged our prior work[13,14] where a total of 1 generic name (metformin) and 4 brand name medications (Fortamet®, Glucophage®, Glumetza®, and Riomet®) were queried. Patients with <6 months of metformin exposure or on combination drugs that included metformin or other prescribed antidiabetic drugs during the ≥6 month exposure period were excluded from the study. Third, to compare the association of genetic modification to glycemic response to metformin, measures of A1c were compared prior to metformin exposure and during the period of metformin exposure following a 6-month period of delay to allow for the achievement of maintenance dosage. A1c measures were required ≤6 months prior to metformin exposure and ≥6 months after metformin exposure. A1c measures were averaged across sections that occurred before and up to the date of metformin exposure. A1c measures were averaged across the period occurring ≥6 months after initial metformin exposure and until either metformin exposure ceased or anti-diabetic combination therapy

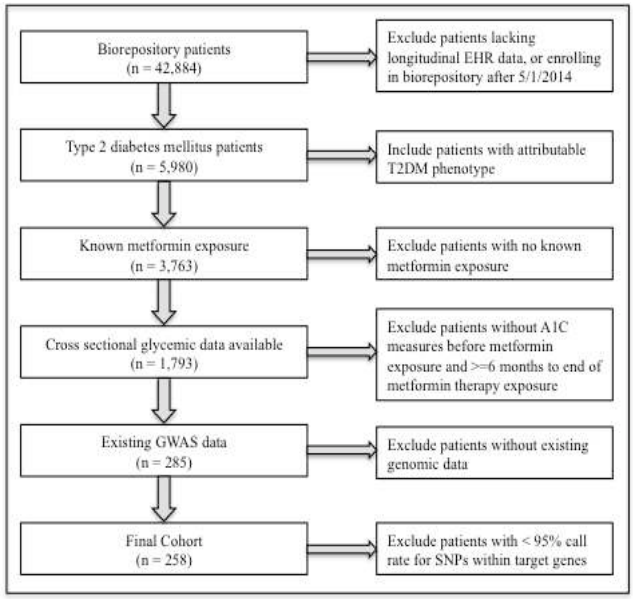

Figure 1: Study Cohort Development Process

was initiated. This approach minimizes the impact of any one A1c measure and biases change in A1c measures towards the null.

Genotyping and Quality Control

The MayoGC stores existing GWAS data generated from multiple studies. These data were harmonized to the forward strand mapped to become on the same strand as the 1000 genome cosmopolitan reference population. Genotypes for unmappable or ambiguous SNPs were excluded. We selected SNPs 20 kb upstream and downstream of each gene using 1000 genomes project variants and NCBI build 37 as the reference genome. By this mapping rule, a total of 1,381 SNPs were mapped to the 5 genes, but only 205 SNPs were available in the genotype data. Further, due to their proximity the FMO1, FMO2, FMO3, and FMO4 genes some SNPs belong to multiple genes. For the remaining SNPs, two main quality control filters were applied:(i) SNPs with unacceptable high rates of missing genotype calls (>10%); and (ii) monomorphic SNPs were excluded. The quality control of the genotype data was performed by PLINK v1.07[15]. A detailed diagram of cohort development is found in **Figure 1.**

Analysis

The SNP-level and gene-level analyses were performed on the final analysis cohort where 258 Caucasian subjects had metformin exposure, complete EHR data, and 90 SNPs after quality control. In the analysis, we adjusted for age, gender, and morbid obesity (BMI ≥35), a known modifier of T2DM state[1] as fixed covariates in our model. Age and BMI measures were calculated at first recorded exposure to metformin. The endpoint of change in A1c was transformed using Van der Waerden rank, otherwise known as rank based inverse Gaussian, to normalize and accommodate linear regression modeling. Batch adjustment did not change the results of GWAS data (data not shown) and was not adjusted in the displayed results. SNP-level and gene-level results were described, but not displayed, after application of Bonferroni correction.

Table 1: Demographics (n=258)

Variable	**n (%)**
Female, N(%)	89 (34.5)
Male, N (%)	169 (64.5)
BMI <30, N (%)	64 (24.8)
BMI(≥30 to <35 kg/m^2)	100 (38.8)
BMI ≥35 (kg/m^2)	93 (36.1)
Median A1c >7.0 (DCCT %), N (%)	101 (39.1)
Change in A1c (DCCT %), median (range)	0.07 (-6.45, 3.51)
Age (years), median (range)	64 (30, 84)

SNP-Level Analysis

SNP-level analyses were performed on each SNP in FMO genes pertaining to both our primary and secondary hypothesis to identify top SNPs and determine directionality of their associations. Using Van der Waerden rank transformation on change in A1c, linear regression models were applied adjusting for age, gender and morbid obesity. Coefficient estimates were calculated per minor allele, that is, with each minor allele, the A1c level changes by 'beta'. SNP-level results are displayed as unadjusted for multiple testing. Finally, conditional analysis was performed to identify potentially independent SNPs in each gene. Locus Zoom plots were also created for better visualization using the LD in the 1000 Genomes European reference population from March 2012 release.

Gene-Level Analysis

Gene-level tests were performed using principal component analysis (PCA)[16]. For each gene, principal components (PC) were created using linear combinations of ordinally scaled SNPs (i.e., 0, 1, 2 copies of minor allele) and the smallest set of resulting principal components that explained at least 90% of the SNP variance within the gene was included in linear regression models. Instead of including the entire set of SNPs for each gene, the PC approach reduces the degrees of freedom, avoids model fitting issues due to multi-collinearity of the SNPs from linkage disequilibrium (LD) and potentially improves the statistical power. Finally, we computed the likelihood ratio test (LRT) to assess overall significance of a gene by comparing the null model containing only the covariates with the full model containing covariates and the set of resulting principal components. The statistical package R 2.15.0 was utilized for the gene-level analysis. Plots of LD displaying r^2 for FMO5 gene was created using Haploview v 4.2.

Results

Our EHR-based phenotyping algorithm identified 1,793 T2DM subjects **(Figure 1)**. Among those, 258 subjects had 90 SNP data that passed quality control criteria. Cohort demographics can be found in **Table 1.** The estimates for male (Coefficient=0.0435, P-value=0.737), age (Coefficient=-0.0009, P-value=0.881), morbid obesity (Coefficient=0.2214, P-value=0.083) were not significantly associated with change in A1c at alpha=0.05 significance level in the univariate analysis. Further, none of the covariates were associated with change in A1c at alpha=0.05 significance level in a multivariate model.

SNP-Level Results

Of the 5 candidate genes, only FMO5 had SNPs that demonstrated a potentially significant association **(Table 2)**. After adjusting for multiple testing rs7541245, the top SNP in FMO5, was marginally significant, but since this signal is very close to passing correction (0.00188-observed vs. 0.00161-Bonferonni threshold) it was deemed appropriate for consideration. None of the SNPs in FMO1-FMO4 gene cluster were found to be significant. Among 31 genotyped SNPs within FMO5 gene **(Figure 2)**, 4 SNPs had p-values less than 0.05 for the association with a decrease in glycemic response during metformin exposure, with rs7541245 having the most significant signal. The FMO5 linkage disequilibrium (LD) plot (not shown due to space constraints) contained 4 LD blocks and appeared to show 9 independent SNPs. The conditional analysis that adjusted for the top most significant SNP in each gene and clinical covariates was performed. FMO5 rs7541245 was the main signal on FMO5 gene as no SNPs reached p-values less than 0.05 which pointed to the remaining SNPs within FMO5 being in high LD with rs7541245 and hence, not independent.

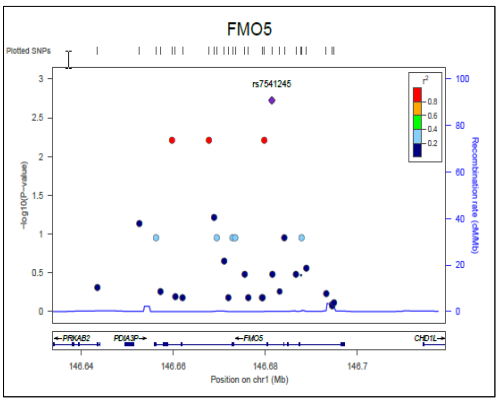

Figure 2: Locus Zoom Plot for FMO5 association with glycemic response to metformin

Table 2: SNP-Level Analysis[+]

Gene Name	Genotyped SNPs (n)	Top SNP	Minor Allele	Major Allele	MAF	BETA	95% CIs	P-value
FMO5	31	rs7541245	A	C	0.0311	-0.7885	(-1.28;-0.297)	**0.00188***
FMO4	15	rs2076322	G	A	0.1395	0.2061	(-0.055;0.467)	0.12270
FMO3	19	rs1920145	C	T	0.3346	-0.1076	(-0.296;0.081)	0.26530
FMO2	14	rs12752688	T	C	0.1434	0.2256	(-0.025;0.476)	0.07885
FMO1	12	rs13376631	G	A	0.1376	0.1997	(-0.062;0.461)	0.13560

MAF = minor allele frequency, * marginally significant after correction for multiple testing
[+]Only top SNPs displayed due to manuscript space constraints

Gene-Level Results

Our primary hypothesis for the FMO5 gene, represented by 5 PCs and 31 genotyped SNPs, was marginally significantly associated (p=0.0185) with glycemic response **(Table 3)** after controlling for age, gender and morbid obesity. No significant associations were identified for our secondary hypothesis tests of the remaining FMO genes.

Table 3: Principal Component Analysis

Gene	Genotyped SNPs (n)	nPCs	P-value
FMO5	31	5	**0.0185**
FMO4	12	4	0.5623
FMO3	14	5	0.5464
FMO2	19	4	0.3581
FMO1	15	6	0.5479

Discussion

In this study, we leverage EHR-linked biorepository data and EHR-based phenotyping methods to study common variants within FMO5, our gene of primary interest. While the FMO5 gene appeared to be of marginal significance in relation to glycemic response to metformin, our secondary hypothesis for the remaining FMO genes demonstrated no significance. Given the study design and execution of phenotypes, results of this study can be interpreted most accurately as pharmacogenetics hypothesis generating. However, this hypothesis could represent a novel mechanism for the biotransformation of metformin or other potential mechanism of metformin action that has been previously unidentified. Additional studies are needed; functional studies are potentially warranted.

In our study not all SNPs within candidate genes were available for analysis due to GWAS genotyping being originally performed for other studies. No effect difference was observed between cohort batches which hint that our findings were not biased due to original patient selection criteria or genotyping criteria, however potential for heterogeneity remains. Having all patients with T2DM and metformin allowed for us to identify genetic variation as the consideration of interest. However, the limited sample size paired with a relatively weak clinical outcome had potential to bias associations towards the null. While utilizing a clinical endpoint enabled us to engage in exploratory research, our signal strength was limited by modest cohort size (n=258) and the study criteria design. Specifically, by removing patients with <6 months of metformin exposure during metformin exposure we potentially removed patients who were complete non-responders to metformin or who experienced an adverse

reactions to metformin. By study design these would not have been able to attain glycemic control with metformin, biasing our outcome phenotype towards positive glycemic response (i.e. decreased A1c) to metformin.

Alterations in FMO genes are known to induce differential biotransformation of nitrogen-rich compounds, such as metformin[10]. In this study, it appeared that the utility of metformin (i.e. glycemic response) is impaired by alterations in the FMO5 gene, hinting that potential biotransformation of metformin might be occurring in the normal FMO5 gene product. Our finding hints that metformin conjugates resulting from metformin biotransformation via FMO5 might be responsible for the anti-diabetic effects of metformin. Should these findings be confirmed by functional studies, this hypothesis could represent a novel mechanism for the biotransformation of metformin and mechanism of metformin action that has been previously unidentified.

Conclusion

FMO5 appears to be marginally significantly associated with decreases in glycemic response after exposure to metformin, representing an EHR-driven pharmacogenetics hypothesis that could represent a novel mechanism for the biotransformation of metformin that has been previously unidentified. Functional validation of this hypothesis is warranted to ascertain its clinical and biological significance.

Acknowledgments

This work has been supported by NIH grant U19-GM61388-13, PGRN Network Resource, Pharmacogenomics of Phase II Drug Metabolizing Enzymes.

References

1. Inzucchi SE, Bergenstal RM, Buse JB, et al. Management of Hyperglycemia in Type 2 Diabetes: A Patient-Centered Approach. Diabetes Care. 2012;35:1364-1379.
2. Todd JN, Florez JC. An update on the pharmacogenomics of metformin: progress, problems and potential. Pharmacogenomics. 2014;15(4):529-539.
3. Gong L, Goswami S, Giacomini KM, Altman RB, Klein TE. Metformin pathways: pharmacokinetics and pharmacodynamics. Pharmacogenetics and genomics. Nov 2012;22(11):820-827.
4. Bailey CJ, Path MRC, Turner RC. Metformin. The New England Journal of Medicine. 1996;334(9):574-579.
5. Graham G.C., Punt J, Arora M, et al. Clinical Pharmacokinetics of Metformin. Clinical Pharmacokinetics. 2011;50(2):81-98.
6. Hundal RS, Krssak M, Dufour S, et al. Mechanism by Which Metformin Reduces Glucose Production in Type 2 Diabetes. Diabetes. 2000;49.
7. Sakar Y, Meddah B, Faouzi MYA, Cherrah Y, Bado A, Ducroc R. Metformin-Induced Regulation of Intestinal D-Glucose Transporters. Journal of Physiology and Pharmacology. 2010;61(3):301-307.
8. Lattard V, Zhang J, Cashman JR. Alternative Processing Events in Human FMO Genes. Molecular Pharmacology. 2004;65(6):1517-1525.
9. Joosten V, van Berkel WJH. Flavoenzymes. Current Opinion in Chemical Biology. 2007;11:195-202.
10. Ziegler DM. Flavin-Containing Monooxygenases: Catalytic Mechanism and Substrate Specificities. Drug Metabolism Reviews. 1988;19(1):1-33.
11. Bielinski SJ, Chai HS, Pathak J, et al. Mayo Genome Consortia: A Genotype-Phenotype Resource for Genome-Wide Association Studies With an Application to the Analysis of Circulating Bilirubin Levels. Mayo Clinic proceedings. 2011;86(7):606-614.
12. Kho AN, Hayes MG, Rasmussen-Torvik L, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. Journal of the American Medical Informatics Association : JAMIA. Mar-Apr 2012;19(2):212-218.
13. Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. JAMIA. Mar-Apr 2010;17(2):131-135.
14. Pathak J, Murphy SP, Willaert BN, et al. Using RxNorm and NDF-RT to Classify Medication Data Extracted from Electronic Health Records: Experiences from the Rochester Epidemiology Project. American Medical Informatics Association Annual Symposium. 2011:1089-1098.
15. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. Sep 2007;81(3):559-575.
16. Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. Genetic epidemiology. Jul 2007;31(5):383-395.