|  |  |
| --- | --- |
| **eMERGE Network: Manuscript Concept Sheet** |  **m** |
| **Reference Number** *(to be assigned by CC)* | NT413.2 |  |
| **Submission Date** | 01/13/2022 |  |
| **Project Title** | Genomic Variance of Type 2 Diabetes Mellitus in African Americans  |  |
| **Tentative Lead Investigator** *(first author)* | Nicole D. Armstrong |  |
| **Tentative Lead Investigator Email Address** | nmda@uab.edu |  |
| **Tentative Senior Author** *(last author)* | Marguerite (Ryan) Irvin |  |
| **All Other Authors**  | UAB: Marguerite (Ryan) Irvin; Hemant Tiwari; Nita A. Limdi; Amit PatkiMGH: Elizabeth W. Karlson; Tian Ge; James B. MeigsOther eMERGE investigators interested in the project |  |
| **Sites Participating** | UAB, Mass General Brigham, and any other eMERGE sites that want to participate |  |
| **Background / Significance** | Diabetes refers to a heterogeneous group of metabolic and health conditions characterized by glucose dysregulation and defects in insulin secretion and/or insulin action(1). Chronic hyperglycemia has been associated with long-term damage, dysfunction, and failure of several organs, including the kidneys, heart, and blood vessels(2), and is a major risk factor for cardiovascular disease, particularly coronary heart disease and stroke(3). In the United States, type 2 diabetes (T2D) is the most common form, which constitutes 90-95% of cases (1). T2D is more often associated with increased age (4); however, the T2D epidemic can largely be attributed to a worldwide increase in obesity(5).It is well established that T2D clusters in families and the risk of developing T2D depends on both genetic and environmental factors. In the Framingham Heart Study, having a first degree relative conferred 3.4 times increased risk of T2D compared to the general population, which increased to a 6.1 times increase if both parents were affected (6). The genetic nature of the disease is supported by stronger heritability in monozygotic twins compared to dizygotic twins in previous studies (7-10). A 2005 study demonstrated that family history, together with a BMI ≥30 kg/m2 and fasting plasma glucose ≥5.5 mmol/L, is associated with a significantly increased risk of T2D compared to those with one or two risk factors (hazard ratio (HR) (95% CI) 3.7 (2.3-6.1), p<0.0001)(11). Furthermore, African Americans (AAs) tend to have higher mortality rates and a higher risk of T2D complications compared to individuals of European descent, where AAs are at least twice as likely to die due to T2D (12, 13). Narrow-sense heritability is the proportion of phenotypic variance due to additive genetic variation. Traditionally, heritability has been estimated from family-based studies, which have suggested for many complex traits that much of the phenotypic variance is due to additive genetic variance(14, 15). However, these family-based heritability estimates may be biased by factors shared by close relatives, such as non-additive genetic and common environmental effects(15). Previous heritability estimates of T2D and related clinical traits (e.g. fasting glucose, fasting insulin) have varied between 25-80%, depending on the study and follow-up periods, but have been calculated predominately in individuals of European ancestry (2, 16-19). Genetic correlations are genome-wide aggregate effects of causal variants affecting related traits. In the past, genetic correlations between complex traits were estimated from pedigree studies, but estimates from family designs can also be confounded by shared environmental factors. Statistical methods that used GWAS based on linear mixed models to obtain unbiased estimates of the genetic correlation between pairs of quantitative or binary traits, are now the standard to give insight into genetic correlation (20). Previous studies have used these approaches to estimate the genetic correlation between chronic cardiometabolic diseases and traits in large samples of EAs. For instance, the joint risk of T2D and hypertension was evaluated in the Wellcome Trust Case Control Consortium (positive correlation ~0.31; SE 0.14; *p*=0.024)(21). Similar estimates have not been reported among AAs. In the present study, we seek to capture the T2D variance associated with genetic variation from >28,000 AA participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study, the Genetics of Hypertension Associated Treatments (GenHAT) study, the Warfarin Pharmacogenomics Cohort (WPC) and the electronic medical records and genomics network phase III (eMERGE III), making the proposed study the largest study to estimate T2D heritability in AAs.  |  |
| **Outline of Project** | UAB will lead narrow-sense heritability analyses for 28,784 individuals of African ancestry across the REGARDS study (2,683 T2D cases and 6,066 controls), the GenHAT study (2,776 T2D cases and 4,132 controls), the WPC (300 T2D cases and 355 controls), and eMERGE III (2,688 T2D cases and 9,784 controls). We will compare these estimates to those we calculate in the subset of individuals of European descent from the REGARDS study (310 T2D cases, 1,237 controls) and the WPC (325 cases, 659 controls). We will also evaluate the bivariate genetic correlation between T2D and a cluster of related cardiometabolic traits (CKD, HTN, Obesity). First, we will harmonize the T2D definition across all AA datasets, using T2D ICD codes, a single measurement of glucose (fasting glucose ≥ 126 mg/dL or non-fasting glucose ≥ 200 mg/dL) or use of any glucose-lowering medications.Second, we will utilize two heritability estimating toolkits: the genome-wide complex trait analysis (GCTA) (22) and LDAK (23), to estimate the genetic component of phenotypic variance that is explained by the additive genetic effects using the available TOPMed imputed genetic variants. We will compare our heritability estimates to previously published estimates in European populations, as well as in our REGARDS/WPC European dataset.Lastly, we will explore the GCTA toolkit to conduct bivariate genome-based restricted maximum likelihood (GREML) to estimate the genetic correlation between T2D and 1.) BMI, 2.) hypertension, and 3.) chronic kidney disease using eMERGE data where cohort sample size allows.  |  |
| **Desired Data - Common Variables\*** *(Available from the CC)* | [x] Demographics [x] ICD9/10 codes[ ] CPT codes[x] Phecodes[x] BMI | [ ] Common Variable Labs[ ] Common Variable Meds[x] Other: Case/Control status on Phase I and Phase II phenotypes |  |
| **Other Desired Data *(Available from participating sites)*** | *Please specifically list out any data elements that participating sites would collect or extract from clinical or other sources for this project (i.e. not common variables above)*  |  |
| **Desired Genetic Data** | [x] eMERGE I-III Merged set (HRC imputed, GWAS) [ ] eMERGE PGx/PGRNseq data set [ ] eMERGEseq data set (Phase III)[ ] eMERGE Whole Genome sequencing data set[ ] eMERGE Exome chip data set[ ] eMERGE Whole Exome sequencing data set[ ] Other (not listed above): |  |
| **Does project pertain to an existing eMERGE Phenotype?** | [x] Yes, if so please list: Type 2 Diabetes Mellitus [ ] No |  |
| **Planned Statistical Analyses** | Using GCTA Toolkit to estimate heritability using imputed genetic variants:1. Construct a genomic relationship matrix (GRM) within the GCTA toolkit, as previously described (15, 22).
2. Detect cryptic relatedness among participants. Remove individuals at relatedness values > 0.025.
3. Generate top 10 principal components using EIGENSTRAT (24) to account for genetic ancestry in mixed models (for use as covariates).
4. Estimate narrow-sense heritability using mixed models, where T2D phenotypic covariance between any two individuals in a study will be modeled as a function of the genomic relatedness of the individual. We will include age, sex and genetic ancestry covariates as fixed effects.

Generate heritability estimates using LDAK using imputed genetic variants:1. Calculate kinship matrix for each study
2. Estimate the heritability contributed by each kinship matrix using the restricted maximum likelihood (REML) or phenotype-correlation, genotype-correlation (PCGC) regression (while REML is generally preferred because it produces the most precise estimates of heritability, i.e. the smallest standard deviation, it is more computationally intensive and makes more assumptions, i.e. the effect sizes are Gaussian, when compared to PCGC. Therefore, PCGC produces less-biased estimates for binary traits)(23, 25). We will adjust for age, sex, and genetic ancestry.

Investigate the genetic correlation between T2D and comorbid traits:1. Using bivariate-GREML in GCTA, we will calculate the genetic correlation of T2D and BMI, hypertension, and chronic kidney disease in the eMERGE AA dataset.
 |  |
| **Ethical Considerations** | N/A |  |
| **Target Journal** | PLoS Genetics |  |
| **Milestones***(This section should include the key dates for completion of project, including approval, project duration, draft completion, and submission.)* | * 01/2022: Perform narrow-sense heritability estimates for T2D on eMERGE III
* 01/2022: Perform a combined, narrow-sense heritability estimate for T2D on participants from eMERGE III, GenHAT, WPC, and REGARDS.
* 02/2022: Perform genetic correlation analyses
* 02/2022: Manuscript draft completion
* 03/2022: Manuscript draft submission
 |  |

**\*Common Variables available across all datasets:**

* Demographics: sex, year of birth, a decade of birth, race, ethnicity
* Codes: (repeated values & age at the event): ICD9/10, CPT, Phecodes
* BMI: (repeated value & age at the event) height, weight, BMI, for potential sensitivity analyses
* HTN: (repeated value & age at event): SBP, DBP, HTN case-control status
* CKD: (repeated value & age at event): eGFR, albuminuria
* Medications: (medication name, repeated, & age at event) Blood pressure medications for adjustment to baseline blood pressure values
* Other: Case/Control status on Phase I and Phase II phenotype: only on GWAS dataset participants

**References**

1. Virani SS, Alonso A, Aparicio HJ, et al. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. *Circulation*. 2021;143(8):e254-e743. doi:10.1161/CIR.0000000000000950

2. Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes (Basel)*. 2015;6(1):87-123. doi:10.3390/genes6010087

3. Emerging Risk Factors C, Sarwar N, Gao P, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet*. 2010;375(9733):2215-22. doi:10.1016/S0140-6736(10)60484-9

4. Stankov K, Benc D, Draskovic D. Genetic and epigenetic factors in etiology of diabetes mellitus type 1. *Pediatrics*. 2013;132(6):1112-22. doi:10.1542/peds.2013-1652

5. Lyssenko V, Jonsson A, Almgren P, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med*. 2008;359(21):2220-32. doi:10.1056/NEJMoa0801869

6. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes*. 2000;49(12):2201-7. doi:10.2337/diabetes.49.12.2201

7. Kaprio J, Tuomilehto J, Koskenvuo M, et al. Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland. *Diabetologia*. 1992;35(11):1060-7. doi:10.1007/BF02221682

8. Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia*. 1999;42(2):139-45. doi:10.1007/s001250051131

9. Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD. Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia*. 1987;30(10):763-8. doi:10.1007/BF00275741

10. Medici F, Hawa M, Ianari A, Pyke DA, Leslie RD. Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia*. 1999;42(2):146-50. doi:10.1007/s001250051132

11. Lyssenko V, Almgren P, Anevski D, et al. Predictors of and longitudinal changes in insulin sensitivity and secretion preceding onset of type 2 diabetes. *Diabetes*. 2005;54(1):166-74. doi:10.2337/diabetes.54.1.166

12. Mansour O, Golden SH, Yeh HC. Disparities in mortality among adults with and without diabetes by sex and race. *J Diabetes Complications*. 2020;34(3):107496. doi:10.1016/j.jdiacomp.2019.107496

13. Lanting LC, Joung IM, Mackenbach JP, Lamberts SW, Bootsma AH. Ethnic differences in mortality, end-stage complications, and quality of care among diabetic patients: a review. *Diabetes Care*. 2005;28(9):2280-8. doi:10.2337/diacare.28.9.2280

14. Polderman TJ, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015;47(7):702-9. doi:10.1038/ng.3285

15. Evans LM, Tahmasbi R, Jones M, et al. Narrow-sense heritability estimation of complex traits using identity-by-descent information. *Heredity (Edinb)*. 2018;121(6):616-630. doi:10.1038/s41437-018-0067-0

16. Miljkovic-Gacic I, Wang X, Kammerer CM, et al. Genetic determination of adiponectin and its relationship with body fat topography in multigenerational families of African heritage. *Metabolism*. 2007;56(2):234-8. doi:10.1016/j.metabol.2006.09.019

17. Freedman BI, Rich SS, Sale MM, et al. Genome-wide scans for heritability of fasting serum insulin and glucose concentrations in hypertensive families. *Diabetologia*. 2005;48(4):661-8. doi:10.1007/s00125-005-1679-5

18. Poveda A, Chen Y, Brandstrom A, et al. The heritable basis of gene-environment interactions in cardiometabolic traits. *Diabetologia*. 2017;60(3):442-452. doi:10.1007/s00125-016-4184-0

19. Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet*. 2012;8(3):e1002637. doi:10.1371/journal.pgen.1002637

20. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. *Nat Rev Genet*. 2019;20(10):567-581. doi:10.1038/s41576-019-0137-z

21. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012;28(19):2540-2. doi:10.1093/bioinformatics/bts474

22. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

23. Speed D, Holmes J, Balding DJ. Evaluating and improving heritability models using summary statistics. *Nat Genet*. 2020;52(4):458-462. doi:10.1038/s41588-020-0600-y

24. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-9. doi:10.1038/ng1847

25. Speed D. <https://dougspeed.com/ldak/>.