ARIC Manuscript Proposal #4033

PC Reviewed: 4/12/22	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Type 2 Diabetes Phenotype Polygenic Risk Scores: An Examination Of Proteomics

b. Abbreviated Title (Length 41 characters): T2D phenotype risk scores and proteomics

2. Writing Group: Brian Steffen, Jim Pankow, Weihong Tang, Elizabeth Selvin. Nathan Pankratz, Alanna Morrison, David Couper, Faye Norby, Pam Lutsey, Kunihiro Matsushita, and Ryan Demmer. We welcome any further nominations.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __

First author: Brian Steffen
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. 2nd St., Suite 300
Minneapolis, MN 55454
Phone: 612-626-0371
E-mail: steff293@umn.edu

Senior author: Jim Pankow
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. 2nd St., Suite 300
Minneapolis, MN 55454

3. Timeline: Begin analysis in August, 2020
Draft of paper by November 2020

4. Rationale:

Type 2 diabetes (T2D) development has a complex pathophysiology characterized by disruptions in cell signaling and metabolic homeostasis across multiple organ systems. Peripheral and hepatic insulin resistance, pancreatic beta cell dysfunction, perturbations in macronutrient handling, aberrant distributions of adipose tissue, elevated BMI, and blood lipid abnormalities are all typical manifestations; however, affected individuals present with different combinations of these phenotypes (1-3). And while lifestyle factors such as physical inactivity and diet are well-known risk factors for disease development (4,5), genetic susceptibilities also influence risk (6-8) and, critically, may contribute to the phenotypic heterogeneity of T2D (9,10).

So far genome-wide association studies have identified over 450 gene variants significantly related to T2D (7-9, 11,12), though most variants have shown relatively modest magnitudes of association with disease, <10% increased risk per allele (12-14). In an effort to more thoroughly capture genetic risk, investigators have aggregated dozens to millions of common gene variants, most of which do not meet the genome-wide significance threshold ($p < 5 \times 10^{-8}$), in polygenic risk scores (PRSs) (6, 7, 9). PRSs for T2D have shown varying degrees of efficacy with early scores providing marginal improvement risk discrimination beyond that of typical clinical factors (15, 16), while others have been proven informative—e.g., a recent study showed that individuals with PRS values in the top 2.5% were at 3.4-fold and 9.4-fold greater risk compared to those at the median and bottom 2.5%, respectively (7).

Further refining this approach, PRSs comprised of clusters of gene variants have been developed to interrogate the distinct manifestations and disease heterogeneity of T2D. In a recent study by Goodarzi et al. (10), 106 gene variants were partitioned by their pathophysiological phenotypes, generating four PRSs with mutually exclusive variants: 1) beta cell dysfunction: 52 variants shown to be related to glucose-stimulated insulin secretion, proinsulin, beta-cell differentiation or proliferation, insulin processing or hyperglycemia; 2) insulin resistance: 30 variants identified by associations derived from oral glucose tolerance test, euglycemic-hyperinsulinemic clamp, insulin suppression test, or the frequently sampled intravenous glucose tolerance test; 3) lipodystrophy: 12 variants related lipoatrophy or adipose tissue storage shifting from peripheral and subcutaneous compartments to visceral and hepatic compartments; and finally 4) 12 variants related to BMI and blood lipids. Building upon this research, we propose to explore these phenotype-derived PRSs and their associations with proteomic signature outcomes.

T2D proteomics research has so far examined aspects including, but not limited to, associations with prevalent T2D, its complications, and responses to pharmacotherapies (17-19), yet no studies have interrogated the PRSs of T2D. We hypothesize that this approach may identify pathways through which gene variants contribute to pathophysiology, provide a more complete understanding of T2D heterogeneity, and reveal potential upstream signaling cascades for targeted pharmacotherapies. However, given the novelty in this area, it must be acknowledged that this proposal has a number of limitations. Perhaps most critically, our hypotheses are contingent on the variants within each PRS sharing pathways or converging on a common pathway. PRSs comprised of variants with heterogeneous mechanisms will show weaker, if any, associations with proteins.

5. Main Hypothesis/Study Questions:

Main Hypothesis: PRSs for beta cell dysfunction, insulin resistance, lipodystrophy and BMI+lipid aberrations will be associated with distinct proteomic signatures.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Design: cross-sectional. Proteomic data were measured at Visit 3 samples using SOMAscan v 4.0. This approach quantifies 4,931 human proteins through an aptamer-based platform. The substantial protein coverage of SOMAscan v 4.0 and the over 11,000 ARIC participants with available data provide the opportunity to characterize proteomic signatures associated with the genetic risk for distinct T2D manifestations. Visit 3 was selected for the primary analysis due to the larger sample size, fewer comorbidities, and less extensive medication use.

Data from Visit 5 will serve as an internal replication sample.

Inclusion/Exclusion: We will exclude participants who are not black or white, do not have SOMAscan data at visit 3 or the corresponding genetic variants, or failed QC for SNPs or proteins. To avoid finding medication effects, participants taking diabetes medications at the sample visit for SOMAscan will also be excluded.

Exposure: PRSs will be generated from the below gene variants for each of the four distinct T2D pathogenic phenotypes:

Beta cell dysfunction:

PROX1 rs340874	GCK rs878521	KCNQ1 rs2237897
PROX1 rs114526150	LEP rs791595	KCNQ1 rs445084
THADA rs80147536	ANK1 rs13262861	KCNJ11 rs5213
ADCY5 rs11708067	ANK1 rs4736819	TMEM258 rs102275
IGF2BP2 rs6780171	SLC30A8 rs3802177	CENTD2/ARAP1 rs77464186
IGF2BP2 rs150111048	GLIS3 rs10974438	MTNR1B rs10830963
IGF2BP2 rs11717959	CDKN2A/B rs10811660	MTNR1B rs57235767
WFS1 rs1801212	CDKN2A/B rs10757283	KLHDC5 rs10842994
WFS1 rs10937721	ABO rs505922	HNF1A rs56348580
ANKH rs146886108	GPSM1 rs28505901	SPRY2 rs1359790
ZBED3 rs4457053	CDC123/CAMK1D rs11257655	C2CD4A/B rs8037894
RREB1 rs9379084	HHEX/IDE rs10882101	HMG20A rs1005752
RREB1 rs9505097	HHEX/IDE rs1112718	PRC1 rs12910825

CDKAL1 rs7756992	TCF7L2 rs7903146	ZZEF1 rs1377807
CENPW rs11759026	TCF7L2 rs34855922	HNF1B rs10908278
DGKB rs10228066	KCNQ1 rs234853	TTLL6 rs35895680
JAZF1 rs1708302	KCNQ1 rs2237895	
DUSP9 rs5945326	ZHX3 rs17265513	

Insulin resistance:

MACF1 rs3768321	FAM13A rs1903002	MPHOSPH9 rs4148856
FAM63A rs145904381	PDGFC rs28819812	BCAR1 rs72802342
GCKR rs1260326	ARL15 rs702634	BCAR1 rs3115960
CEP68 rs2249105	ANKRD55 rs465002	CMIP rs2925979
CEP68 rs2052261	ANKRD55 rs9687832	BCL2A rs12454712
GRB14/COBLL1 rs10195252	VEGFA rs11967262	TM6SF2 rs8107974
IRS1 rs2972144	KLF14 rs1562396	PEPD rs10406327
PPARG rs11709077	PLEKHA1 rs2280141	GIPR rs10406431
KIF9 rs11926707	PLCB3 rs35169799	GIPR rs2238689
ADAMTS9 rs9860730	CCND1 rs61881115	PIM3 rs112915006

Lipodystrophy:

LYPLAL1 rs2820446	VEGFA rs6458354	ITPR2 rs718314
PIK3R1 rs4976033	SOGA3 rs2800733	ZNF664 rs7978610
YTHDC2 rs10077431	MIR3668 rs2982521	MAP2K7 rs4804833
EBF1 rs3934712	LPL rs10096633	EYA2 rs6063048

BMI and blood lipids:

POC5 rs2307111	NFAT5 rs862320	BPTF rs61676547
TFAP2B rs3798519	MC4R rs523288	TOMM40/APOE rs429358
NRXN3 rs17836088	ZMIZ1 rs703967,	HNF4A rs11696357
FTO rs1421085	MAP3K11 rs1783541	PNPLA3 rs738408

Outcome: Natural log2-transformed SomaScan protein levels will serve as the outcome variable, and ARIC analytic recommendations for the SomaScan data will be followed. The proteins were assessed using a Slow Off-rate Modified Aptamer (SOMAmer)-based capture array (SomaLogic, Inc, Boulder, Colorado). Non-human proteins and proteins with unacceptable QC will be removed (eg have large CV (e.g. > 20%), poor reproducibility between the blind duplicate pairs, or non-specific binding).

Data analysis:
Multiple linear regression analysis in which the PRS will serve as the independent variable and SOMAscan protein levels will serve as dependent variables. For each SNP, the number copy of the risk allele as reported by Goodarzi et al. (10) will be modeled. The PRS will be calculated as

the weighted sum of the number of risk alleles. Weighted sums will be derived from relative risk estimates for T2D as previously reported (10, 20, 21). Models will be adjusted for age at visit 3, sex, principal components for ancestry, estimated glomerular filtration rate, and field center. We will analyze the PRSs as both continuous and categorical variables (i.e., top quintile vs remaining). The categorical PRSs may capture additional proteins associated with high genetic risk for that metabolic phenotype. Statistical significance will be determined after Bonferroni correction for the number of proteins. Analyses in Black and white individuals will be run separately and associations compared between groups. As a replication step, we will repeat the analyses using data collected at visit 5. We will also seek replication of significant findings in external cohorts that have published GWAS for SOMAscan proteins.

To aid in the interpretation of the results, we will use Ingenuity Pathway Analysis to identify the most relevant signaling and metabolic pathways, molecular networks, and biological functions for associated proteins.

Limitations: It must be acknowledged that relying on PRSs categorized by T2D phenotypes may too broadly presume that common pathways influence each pathophysiology through the selected clusters of gene variants. For example, PRSs comprised of variants that share pathways or converge on a common pathway would be expected to more strongly associate with proteomic signatures; however, PRSs comprised of variants with heterogeneous pathways which converge on the overt phenotype will show weaker associations. Caution will have to be exercised in interpreting these latter results.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes __x__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __x__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __x__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status.

ARIC Investigators have access to the publications lists under the Study Members Area of the web site at:

___x___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

No investigators have so far proposed to examine proteomic signatures associated with these phenotypic PRSs.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _x___ Yes ____ No

11.b. If yes, is the proposal

_x__ A. primarily the result of an ancillary study (list number* _AS2017.27___)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at https://www2.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References

1) Faerch K, Hulmán A, Solomon TP. Heterogeneity of Pre-diabetes and Type 2 Diabetes: Implications for Prediction, Prevention and Treatment Responsiveness. Curr Diabetes Rev. 2016;12(1):30-41.
2) Ali S, Nafis S, Kalaiarasan P, Rai E, Sharma S, Bamezai RN. Understanding Genetic Heterogeneity in Type 2 Diabetes by Delineating Physiological Phenotypes: SIRT1 and its Gene Network in Impaired Insulin Secretion. Rev Diabet Stud. 2016;13(1):17-34.
3) Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, Vikman P, Prasad RB, Aly DM, Almgren P, Wessman Y, Shaat N, Spégel P, Mulder H, Lindholm E, Melander O, Hansson O, Malmqvist U, Lernmark Å, Lahti K, Forsén T, Tuomi T, Rosengren AH, Groop L. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. 2018;6(5):361-369.
4) Jannasch F, Kröger J, Schulze MB. Dietary Patterns and Type 2 Diabetes: A Systematic Literature Review and Meta-Analysis of Prospective Studies. J Nutr. 2017;147(6):1174-1182.
5) Bennett DA, Du H, Bragg F, Guo Y, Wright N, Yang L, Bian Z, Chen Y, Yu C, Wang S, Meng F, Lv J, Chen J, Li L, Clarke R, Chen Z; China Kadoorie Biobank Study Collaborative Group . Physical activity, sedentary leisure-time and risk of incident type 2 diabetes: a prospective study of 512 000 Chinese adults. BMJ Open Diabetes Res Care. 2019;7(1):e000835.
6) Läll K, Mägi R, Morris A, Metspalu A, Fischer K. Personalized risk prediction for type 2 diabetes: the potential of genetic risk scores. Genet Med. 2017;19(3):322-329.
7) Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet. 2018;50(4):559-571.
8) Hale PJ, López-Yunez AM, Chen JY. Genome-wide meta-analysis of genetic susceptible genes for Type 2 Diabetes. BMC Syst Biol. 2012;6 Suppl 3:S16.
9) Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. Endocr Rev. 2019;40(6):1500-1520.
10) Goodarzi MO, Palmer ND, Cui J, Guo X, Chen YI, Taylor KD, Raffel LJ, Wagenknecht LE, Buchanan TA, Hsueh WA, Rotter JI. Classification of Type 2 Diabetes Genetic Variants and a Novel Genetic Risk Score Association With Insulin Clearance. J Clin Endocrinol Metab. 2020;105(4). pii: dgz198.
11) Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, Pervjakova N, Pers TH, Johnson AD, Eicher JD, Jackson AU, Ferreira T, Lee Y, Ma C, Steinthorsdottir V, et al ; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 2017;66(11):2888-2902.
12) Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, Yengo L, Lloyd-Jones LR, Sidorenko J, Wu Y; eQTLGen Consortium, McRae AF, Visscher PM, Zeng J, Yang J. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018;9(1):2941.
13) Morris AP. Progress in defining the genetic contribution to type 2 diabetes susceptibility. Curr Opin Genet Dev. 2018;50:41-51.
14) Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. Genes (Basel). 2015 Mar 12;6(1):87-123.

15) Lango H; UK Type 2 Diabetes Genetics Consortium, Palmer CN, Morris AD, Zeggini E, Hattersley AT, McCarthy MI, Frayling TM, Weedon MN. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes. 2008;57(11):3129-35.
16) Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB Sr, Cupples LA. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med. 2008;359(21):2208-19.
17) Merchant ML, Klein JB. Proteomics and diabetic retinopathy. Clin Lab Med. 2009 Mar;29(1):139-49.
18) Sundsten T, Ortsäter H. Proteomics in diabetes research. Mol Cell Endocrinol. 2009;297(1-2):93-103.
19) Bergsten P. Islet protein profiling. Diabetes Obes Metab. 2009;11 Suppl 4:97-117.
20) Bonàs-Guarch S, Guindo-Martínez M, Miguel-Escalada I, et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. Nat Commun. 2018;9(1):321.
21) Tsai FJ, Yang CF, Chen CC, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. PloS Genet. 2010;6(2):e1000847.