Large-Scale Association Analysis Provides Insights Into the Genetic Architecture and Pathophysiology of Type 2 Diabetes Mellitus

Nehal N. Mehta, MD, MSCE


Study Hypothesis
We are amidst an epidemic of cardiometabolic diseases. Despite aggressive lifestyle modifications, the incidence of type 2 diabetes mellitus (T2DM), a potent risk factor for cardiovascular diseases, continues to rise. Simple family histories in patients demonstrate that there is a clustering of the disease in families, and prior studies using genome-wide association (GWA) approaches have identified 56 genetic loci that are associated with T2DM. In a recently published herculean effort, the authors stated that prior GWA study results captured 10% of familial aggregation of the disease. Therefore, to enhance fine mapping of established loci, the authors asked whether characterization of increasing numbers of risk loci may provide evidence, at the functional level, that susceptibility to T2DM involves a limited set of molecular processes.

How Was the Hypothesis Tested?
The authors conducted a 2-stage meta-analysis; stage-1 meta-analysis consisted of 12 171 T2DM cases and 56 862 controls across 12 GWA studies of individuals of European descent, and the stage-2 meta-analysis consisted of 21 491 T2DM cases and 55 647 controls across 25 studies of individuals of European descent, and 1178 T2DM cases and 2472 controls from 1 study of individuals of Pakistani descent (PROMIS). These participants all underwent genotyping on the Metabochip, a custom array of 196 725 variants designed to facilitate cost-effective follow-up of nominal associations for T2DM, and other metabolic and cardiovascular traits and to enhance fine mapping of established loci. The authors asked whether characterization of increasing numbers of risk loci may provide evidence, at the functional level, that susceptibility to T2DM involves a limited set of molecular processes.

Principal Findings
T2DM-susceptibility loci reaching genome-wide significance: Combining stage-1 and stage-2 meta-analyses, the authors identified 8 new T2DM-susceptibility loci at genome-wide significance (P<5×10−8). The strongest signals mapped to ZMIZ1 (P=1.0×10−10), ANK1 (P=2.5×10−10), KLHDC5 (P=6.1×10−10), HMG20A (P=4.6×10−9), and GRB14 (P=1.0×10−8). The lead SNPs from both meta-analyses were in strong linkage disequilibrium (HMG20A r2=0.89 and GRB14 r2=0.77 in European [CEU]) and likely represented the same association signals.

Several of these signals mapped to loci previously implicated in T2DM-related metabolic traits. For example, the lead SNP at MC4R was in strong linkage disequilibrium with...
variants associated with body mass index (CEU \( r^2 = 0.80 \)) and triglyceride concentration (CEU \( r^2 = 0.84 \)) and has been associated with waist circumference and insulin resistance. The lead SNP at \( GRB14 \) was highly correlated with variants associated with waist-to-hip ratio and high-density lipoprotein cholesterol (CEU \( r^2 = 0.93 \)). At \( CILP2 \), the lead SNP for T2DM was also associated with triglyceride, low-density lipoprotein, and total cholesterol concentrations. In contrast, the previously reported association signals for hemoglobin A1C concentrations near \( ANK1 \) were both independent (CEU \( r^2 < 0.01 \)) of the lead T2DM-associated SNP from their meta-analysis suggesting that variation at this locus also has direct effects on glucose homeostasis. Finally, the authors demonstrated that by including their 8 new loci with assumed population prevalence for T2DM of 8%, the total 63 loci together account for 5.7% of variance in disease susceptibility. The authors performed additional models to determine the extent to which additional common variant associations contributed to the overall variance explained.

**Sex differentiated analyses:** Next, the authors performed sex-differentiated meta-analysis to test for association of each SNP with T2DM, allowing for heterogeneity in allelic effects between men (20219 cases, 54640 controls) and women (14621 cases, 60377 controls). They identified 2 additional loci achieving genome-wide significance mapping near \( CCND2 \) in men (men \( P = 1.1 \times 10^{-9} \), women \( P = 0.036 \); heterogeneity \( P = 0.013 \)) and upstream of \( GIPR \) in women (women \( P = 2.2 \times 10^{-7} \), men \( P = 0.0037 \); heterogeneity \( P = 0.057 \)).

**Understanding the biology of T2DM-susceptibility loci:** Next, the authors applied a variety of approaches to the newly discovered and established T2DM-susceptibility loci to identify mechanisms involved in disease pathogenesis. Using physiological analyses, they studied the effect of SNPs on glycomic traits using fasting glucose concentration in 133010 non-T2DM individuals. In addition to the 9 loci previously reported (\( MTNR1B, DGKB, ADCYS, PROXI, GCK, GCKR, TCF7L2, SLCO3A8, \) and \( C2CD4A \)), 4 more T2DM-association signals reached genome-wide significance for fasting glucose: \( CDKNA2A-CDKN2B \) (\( P = 5.7 \times 10^{-13} \)), \( ARAP1 \) (\( P = 2.0 \times 10^{-9} \)), \( IGF2BP2 \) (\( P = 1.8 \times 10^{-8} \)), and \( CDKAL1 \) (\( P = 2.0 \times 10^{-8} \)). The \( ZBED3 \) locus also attained genome-wide significance with fasting glucose concentration, after adjustment for body mass index (\( P = 1.2 \times 10^{-8} \)). In contrast, lead T2DM-associated SNPs at 27 of the newly discovered and established loci showed no evidence of association with fasting glucose (\( P > 0.05 \)). Finally, lead T2DM-associated SNPs at the remaining 24 loci were nominally associated with fasting glucose concentrations (\( P < 0.05 \)), suggesting that the genetic landscapes of pathological and physiological variation in glyemia are only partially overlapping. They then examined how homeostatic assessment insulin resistance (HOMA-IR), an estimation of insulin resistance, and homeostatic assessment beta-cell function (HOMA-B), an estimate of pancreatic \( \beta \)-cell function, were related to SNPs. \( ANK1 \) was nominally significant in reduction of HOMA-B, indicating a primary effect on \( \beta \)-cell function, whereas those at \( GRB14 \) and \( AKNRD55 \) increased HOMA-IR. The only lead SNP to show convincing evidence of association (\( P < 1 \times 10^{-5} \)) with adiposity was at \( MC4R \).

**Mapping potential causal transcripts and variants:** As the most comprehensive effort to date, the authors embarked on various approaches to understand functional variants. To identify promising regional transcripts, they examined expression quantitative trait locus data from a variety of tissues. At 6 of the newly discovered loci, the lead T2DM-associated SNP showed strong cis-expression quantitative trait locus associations and was highly correlated (CEU \( r^2 > 0.8 \)) with the lead cis-expression quantitative trait locus SNP that implicated \( GRB14 \) (omental fat), \( ANK1 \) (omental and subcutaneous fat, liver, and prefrontal cortex), \( KLFDC5 \) (blood, T-cells, and CD4+ lymphocytes), \( BCAR1 \) (blood), \( ATP1A3A1 \) (at the \( CILP2 \) locus; blood and monocytes), \( HMGA2A \) (liver), and \( LINGO1 \) (also at the \( HMGA2A \) locus; adipose tissue).

Finally, to extend previous efforts to define pathways and networks involved in T2DM pathogenesis, they combined meta-analysis data with protein–protein interactions, semantic relationships within the published literature, and annotated pathways. All direct interactions and common interactors between direct connections were extracted from the larger network of 314 proteins defined in an established network analysis of 77 selected transcripts mapping nearest to lead SNPs at T2DM-susceptibility loci or implicated in monogenic diabetes. They detected an excess of physical interactions in the network, both direct and indirect. The transcriptional coactivator protein CREBBP, implicated in the coupling of chromatin remodeling to transcription factor recognition, did not map to any T2DM-susceptibility locus. However, it was the most connected gene for protein-level interactions (\( P < 0.005 \)) in the protein–protein interaction network, interacting with 9 primary transcripts, 8 implicated in monogenic diabetes, or mapping to established T2DM-susceptibility loci (\( HNF1A, HNF1B, HNF4A, PLAG1, TCF7L2, PPARG, PROXI, \) and \( NOTCH2 \)), suggesting that modulation of CREBBP-binding transcription factors has an important role in T2DM susceptibility. When the authors used this set of 77 genes as a seed to query a list of 77 secondary transcripts, they found significant connections between the primary associated transcripts and 4 other genes: \( LEPR \) (leptin obesity pathways), \( MYC \) (cell cycle pathway), \( GATA6 \) (pancreas development pathway), and \( DLL4 \) (Notch signaling target).

They tested for enrichment of GWA-associated transcripts in pathway data in 16 biological hypotheses chosen for assumed relevance to T2DM pathogenesis, and 2 showed reproducible enrichment of T2DM associations. The strongest enrichment was observed for a broader set of primary and secondary transcripts mapping to T2DM-associated loci in the adipocytokine signaling that includes the adiponectin, leptin, and tumor necrosis factor-\( \alpha \) signaling. This analysis highlighted 8 genes in this pathway most likely to be the cause for T2DM susceptibility: \( IRS1, LEPR, RELA, RRXG, ACSL1, NFkB1, CAMKK1, \) and a monogenic diabetes gene, \( AKT2 \). Lastly, they found modest but robust enrichment observed for genes influencing cell cycle, in particular, regulators of the G1 phase, such as cyclin-dependent kinase inhibitors (\( CDKN2A, CDKN2B, CDKN1C, \) and \( CDKN2C \)) and cyclins that activate cyclin-dependent kinases (\( CCNE2, CCND2, \) and \( CCNA2 \)). Many of these regulate \( CDK4 \) or \( CDK6 \), which are known to have a role in pancreatic \( \beta \)-cell proliferation. They did not observe evidence of enrichment for other processes implicated in T2DM pathogenesis, including amyloid formation, endoplasmic reticulum stress, and insulin signaling.
Implications

These findings further our understanding of the potential genetic basis of T2DM in >150000 individuals by adding another 10 loci to the list of confirmed common variant signals. These data support the notion of a large number of causal variants of modest effect impacting susceptibility to T2DM moreso than the contribution of rare and low-frequency risk variants. This study elegantly demonstrates how using clinical research, several large datasets along with simultaneous application of in silico, biomarker and genomic technologies can effectively advance steps to underpinning biology. Utilization of outcomes related to cardiometabolic diseases, such as insulin resistance, lipids, and anthropometrics, support observations within the epidemiology of T2DM, metabolic syndrome, and obesity. Variants biologically expected to be involved in the pathogenesis of T2DM, such as those involved in cell cycle regulation and adipocytokine signaling, not only confirm the established reports in these areas, but also demonstrate the critical need to perform targeted, tissue-specific experiments to further delineate mechanisms of these associations. However, notably missing were variants related to inflammation. This finding may reject the hypothesis that inflammation is causal in these disease states, despite recent evidence demonstrating evoked inflammation resembling a diabetes mellitus-like state and inflammatory diseases such as psoriasis being associated with T2DM. Indeed, ongoing resequencing studies will address the contribution of rare and low-frequency variants, and it will be important to determine whether further discovered loci, in fact, coalesce around a limited set of core pathways and networks.

Acknowledgments

The author is a member of the Early Career Committee of the American Heart Association Functional Genomics and Translational Biology Council.

Disclosures

Dr Mehta is an employee of the National Heart, Lung and Blood Institute.

Key Words: diabetes mellitus ▪ genomics
Large-Scale Association Analysis Provides Insights Into the Genetic Architecture and Pathophysiology of Type 2 Diabetes Mellitus
Nehal N. Mehta

_Circ Cardiovasc Genet._ 2012;5:708-710
doi: 10.1161/CIRCGENETICS.112.965350

_Circulation: Cardiovascular Genetics_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/5/6/708

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation: Cardiovascular Genetics_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation: Cardiovascular Genetics_ is online at:
http://circgenetics.ahajournals.org//subscriptions/