All we need is GWAS: Genome-Wide Association Studies in Type 2 Diabetes Mellitus presented on the 2008 EASD Meeting in Rome

Tomasz Klupa and Maciej T. Malecki

Department of Metabolic Diseases, Jagiellonian University, Medical College, 15 Kopernika Street, 31-501 Krakow, Poland.

Address correspondence to: Tomasz Klupa, e-mail: tomasz_klupa@yahoo.com

Manuscript submitted October 31, 2008; resubmitted November 21, 2008; accepted November 26, 2008

■ Abstract

Several lines of evidence suggest that genetic susceptibility plays a major role in the pathogenesis of type 2 diabetes mellitus (T2DM). Limited success of candidate gene approach and linkage analysis in identifying the genetic background of T2DM has caused many research groups to apply the genome-wide association studies (GWAS) approach. Recently, several GWAS have identified and validated novel gene variants highly associated with T2DM. Unfortunately, most of the genetic variance in risk of T2DM still remains undiscovered. The main topic for discussion concerning genetics of T2DM during the 2008 EASD sessions was how to get more data from existing GWAS, and how to improve

GWAS for the future. The presentations and subsequent discussions highlighted some clear weaknesses of GWAS that need to be overcome before further progress can be made. One recognized minefield is the inability to pick up the signals from true diabetes susceptibility variants, against the background statistical noise produced by a large number of other analyzed markers. Also, there are the problems of low sensitivity in identifying the signals from relatively rare variants, and the low effectiveness in identifying epistatic gene-gene interactions.

Keywords: type 2 diabetes • gene • genome-wide association study • SNP • IRS1 • TCF7L2 • FTO • HMGA2 • CDKAL1 • SLC30A8 • insulin resistance

Introduction

everal lines of evidence suggest that genetic susceptibility plays a major role in the pathogenesis of T2DM. Limited success of candidate gene approach and linkage analysis in identifying the genetic background of common T2DM forms has caused many research groups to apply the genome-wide association studies (GWAS) approach. Recently, several GWAS have identified and validated novel gene variants highly associated with T2DM. The gene variants are single nucleotide polymorphisms (SNP).

The results of those GWAS were presented and discussed at past ADA and EASD meetings. Unfortunately, most of the genetic variance in risk for T2DM still awaits discovery. The main questions for the 2008 EASD meeting in Rome, concerning GWAS, were: “What research strategy should we apply to utilize available GWAS data; and, how can we improve GWAS in future?” Other strategies in the search for T2DM susceptibility genes were also intended to be up for discussion. The results, and future GWAS projects, were discussed during oral and poster sessions on the 2008 EASD meeting.

Late breaking GWAS—can we handle the data?

Philippe Froguel led the session entitled, “Late breaking GWAS—can we handle the data?” Dr. Froguel presented the results of newly published GWAS led by his group. The study was successful in identifying novel type 2 diabetes susceptibility genes [1]. The key to success was careful selection of cases and truly normoglycemic controls that increase power and decreases heterogeneity. Professor Froguel concluded that the strong association found in a single large population may be real, although it did not fully meet the criteria for “genome-wide significance”. Applying proper study design resulted in finding IRS1 locus, the first real insulin resistance susceptibility locus found in GWAS. Professor Froguel discussed the problem of under representation of genes influencing insulin resistance among type 2 diabetes genes. This phenomenon may be due to the following reasons:

- Improper study design regarding matching of cases and controls for BMI; in consequence, the genes influencing BMI could not be found.
- Lower heritability of insulin resistance and stronger influence of environmental factors.
- Lower number of variants that affect insulin resistance.
- Possible lower population frequency of insulin resistance variants.
- It may be more difficult to distinguish signals from insulin resistance variants as they do not stand out against the surrounding statistical noise.
- Definition of insulin resistance is not clear, and may lead to biased ascertainment criteria.
- The measures of insulin resistance may not correlate well with insulin resistance at tissue or molecular level.

Genome-wide scans and type 2 diabetes-related traits

Issues concerning GWAS were presented during session entitled “Genome-wide scans in type 2 diabetes”. Most presentations were focused on GWAS based analysis of diabetes-related traits.

Lindgren *et al.* presented a meta-analysis of 13 GWAS (~29,000 individuals) focused on identification of common genetic variants, and pathways influencing individual patterns of fat distribution and central obesity [2]. The phenotype analysis (including waist cir-

cumference and waist-to-hip ratio) has identified 18 independent loci associated with fat distribution. Results of particular interest suggest that genetic variants with a primary effect on central obesity differ from those with an effect on overall obesity (weight, BMI).

Randall *et al.* presented the meta-analysis of GWAS data for weight, showing 18 independent loci associated with adult weight at levels of genome-wide significance [3]. Top hits included the previously known fat mass- and obesity-associated (*FTO*) gene and *HMG42*. Several novel loci are implicated in biologically interesting pathways such as lipid metabolism, glycogen storage and neural signaling.

Bell *et al.* presented results of a first-generation GWAS that allowed for epistasis priorities of multiple interacting candidate loci in T2DM [4]. The analysis was based on 1,924 cases and 2,938 controls. The authors took novel T2DM loci implicated by GWAS and assessed pairwise interaction within and between them, against previously known T2DM loci. Then, the authors performed a genome-wide, gene-gene effect scan, using a joint two-locus test of association. There was no evidence for interaction between novel and “old” T2DM susceptibility loci. The authors performed comparisons involving pairs of inter-chromosomal SNPs, and found 79 pairs showing evidence for two-locus association. All 79 loci showed evidence for joint effect in cooperation with previously known T2DM susceptibility gene: *TCF7L2*.

Fernandez-Cadenas *et al.* analyzed T2DM-oriented GWAS data to test the hypothesis that age of diagnosis (AOD) of T2DM may be genetically influenced [5]. The authors used an age of diagnosis cut off at 45 years. Subjects diagnosed <45 years were compared with controls, and those diagnosed >45 years with controls. Then, the difference in effect between cases diagnosed <45years and those diagnosed >45years was compared. The T2DM-associated genes *FTO* and *TCF7L2* were associated with the disease regardless of AOD. However, for AOD < 45yr, the authors found 3 novel T2DM susceptibility loci, whilst for AOD > 45yr the strongest signals came from established T2DM gene regions. The authors concluded that stratified analysis of GWAS revealed age-related susceptibility loci that may impact on age of diagnosis of diabetes.

Complex diseases: new mechanisms, new interactions

During the session entitled “Genetics: type 2 diabetes” Langeberg *et al.* showed a population-based study

on more than two thousand children who were genotyped for variants in novel T2DM susceptibility loci. The authors found evidence that two established type 2 diabetes susceptibility genes, *CDKAL1* and *SLC30A8*, may influence susceptibility to T2DM in an early age, as well as in late adulthood [6].

There were interesting presentations during the session entitled: "Complex diseases: how to proceed after a genome-wide association study?" David Evans presented a review lecture concerning the role of gene-gene (epistatic) interactions in GWAS [7]. The presence of epistatic interaction may be one of the important factors responsible for GWAS failure in defining the majority of the genetic background of T2DM. The main message of the presentation was that we should model epistasis in GWAS because single locus tests will not always detect interacting loci. Explicitly modeling epistasis may improve the power to detect gene-gene interaction. However, we are faced with multiple testing and computational problems when including epistatic interactions into GWAS analysis. Whilst these are often seen as major problems, their severity is often overestimated.

In the same session Prof. Stylianos Antonarakis debated the influence of human gene polymorphisms on gene expression [8]. He raised the important question of whether our knowledge concerning the structure of human genes is complete. Dr. Antonarakis presented the results of the ENCODE study, which was targeted on a small part of the human genome, and included generation and analysis of functional data from multiple, diverse experiments. The data have been integrated and augmented by a number of evolutionary and computational analyses. The ENCODE study resulted in some major conclusions: genes have numerous additional exons, genes extend hundreds of kilo bases upstream, genome is pervasively transcribed, which means that majority of its bases can be found in primary transcripts, including non-protein-coding transcripts. The results of ENCODE may therefore have significant impact on future genetic tests and may indicate a need to re-evaluate some previous studies.

Factors influencing type 2 diabetes risk and diabetes-related traits

Other, interesting T2DM genetics-orientated sessions included a poster session concerning single and multiple gene influence on diabetes risk. Morris *et al.* presented a poster concerning analysis of the T2DM-related region on chromosome 9p21 [9]. The aim of the study was to explain the nature of association of

the disease with the region, and to fine map casual variants. So far, the stronger signal of association with T2DM mapped to ~10kb interval flanked by recombination hotspots. The authors sequenced the region in a panel of 96 samples and found 30 SNPs, which were included in the following T2DM association study. The study revealed evidence for T2DM associations, with the results pointing to the presence of at least two independent causative mutations.

Lund *et al.* analyzed the role of *TCF7L2* variants in the response to anti-hyperglycemic oral treatment [10]. They performed a randomized, double masked, crossover study of metformin 1g twice daily, versus repaglinide 2 mg thrice daily in 96 insulin naïve T2DM patients. Common *TCF7L2* variants were genotyped. The authors concluded that clinical response to both metformin and repaglinide in non-obese T2DM patients may be influenced by *TCF7L2* gene variants.

Hattersley *et al.* addressed the issue of potential molecular mechanisms that might link low birth weight and increased risk of developing T2DM in future [11]. The authors analyzed 5 mechanisms recently identified with T2DM susceptibility and found risk haplotypes in 2 of them (*CDKAL1* and *HHEX-IDE*) that have a strong influence on birth weight. The authors provided strong evidence that birth weight may be significantly reduced through predominant effect of fetal genotype. Along with previously published data concerning the predominantly maternal effect of *TCF7L2*, the study results suggest that T2DM risk loci have an heterogeneous effect on fetal insulin secretion and growth.

Zeggini *et al.* carried out a meta-analysis of 3 T2DM GWA scans across 10,128 individuals and ~2.2 million SNPs, which detected 6 novel T2DM susceptibility loci with robust evidence for association [12]. The authors calculated individual lambda-s (sibling recurrence risk ratio) values for these loci, finding them to vary between 1.001 to 1.002. Together with the previously established 11 loci, their cumulative lambda-s is 1.07, assuming a log-additive model. Taking into consideration that the overall lambda-s for T2DM has been estimated to be approximately 3, the authors state that the known susceptibility variants account for only a small proportion of observed familial aggregation.

Some genes causal for monogenic diabetes have known roles in cancer pathogenesis. Elliot *et al.* showed evidence that GWA studies support the genetic overlap between type 2 diabetes and prostate cancer [13]. At least 5 of T2DM loci may be associated with T2DM and prostate cancer. Alternate alleles of the same variant in the *HNF1b* gene support a shared genetic etiology between two diseases. To approach

the issue of their common background, the authors compared the results of T2DM and prostate cancer. GWA studies looked for evidence of association signals shared between two. They found 42 signals, of which 22 showed discordant, and 20 showed concordant, direction of associations between the two diseases. The authors conclude that non-random co-localization of T2DM and prostate cancer signals at the genome-wide level indicates overlaps in etiological mechanisms.

Bowden *et al.* tested the role of T2DM susceptibility loci established for European populations, in African Americans [14]. SNPs in 12 loci associated with T2DM in European-derived populations were genotyped in 993 T2DM African American cases and 1054 controls. The authors concluded that except for *TCF7L2*, "European" T2DM variants do not contribute significantly to inter-individual susceptibility to T2DM in African Americans. Unfortunately, due to the relatively small sample size, the study was perhaps "underpowered".

Summary

The 2008 EASD sessions confirmed that concerning genetics of type 2 diabetes, GWAS are still the major tool for identifying T2DM susceptibility loci and T2DM-related traits. However, some weaknesses of GWAS need to be overcome before further progress can be made. In particular, there is the problem of sen-

sitivity to, and picking up, the signal from true diabetes susceptibility variants against the statistical noise produced by the many other analyzed markers. Low sensitivity in identifying signals may result from relatively rare variants or low effectiveness in identifying epistatic gene-gene interactions. Another GWAS-related problem is the difficulty in choosing a significance threshold and avoiding false-positive results. It is now generally recognized, that in GWA studies no finding should be considered "positive" if the p-value is above 5×10^{-8} . The need for proper and thorough study design features strongly.

An important question was raised, as to whether our knowledge of human gene structure is really complete. The ENCODE study has resulted in some very significant conclusions, which may impact on future genetic tests, and may also imply a need to re-evaluate some previous studies.

Other sessions reported on factors influencing type 2 diabetes risk and diabetes-related traits. One of the many papers presented, examined the potential molecular mechanisms that might link low birth-weight and increased risk of developing of T2DM. Another paper concluded that non-random co-localization of T2DM and prostate cancer signals at the genome-wide level indicates overlaps in etiological mechanisms.

Overall the 2008 EASD sessions proved to be an extremely valuable forum, encouraging a wide exchange and sharing of present knowledge.

References

1. **Froguel P.** Late breaking GWAS - can we handle the data? The French/Canadian angle. Oral presentation, 44th EASD Annual Meeting, Rome, September 8-11, 2008.
2. **Lindgren CM, Heid IM, Randall JC, Lamina C, Speliotes EK.** Analysis of genome-wide scans identifies potential novel common susceptibility loci for central obesity. *Diabetologia* 2008. 51(Suppl 1):S9.
3. **Randall JC, Lindgren CM, Speliotes EK, Li S.** Meta-analysis of genome-wide association data for weight from over 32000 individuals. *Diabetologia* 2008. 51(Suppl 1):S9.
4. **Bell JT, Timpson N, Rayner NW, Zeggini E, Morris AP, McCarthy MI.** First-generation scan of genome-wide association data allowing for epistasis prioritises multiple interacting candidate loci in T2D. *Diabetologia* 2008. 51(Suppl 1):S10.
5. **Fernandez-Cadenas I, Prokopenko I, Timpson NJ, Boraska V, Rayner NW, Hattersley AT, Frayling TM, Zeggini E, Lindgren CM, McCarthy MI.** Stratified analysis of the Wellcome Trust Case Control Consortium scan for type 2 diabetes reveals susceptibility loci that may affect age of diagnosis. *Diabetologia* 2008. 51(Suppl 1):S10.
6. **Langenberg C, Loos RJ, Brage S, Sardinha LB, Andersen LB, Ekelund U, Wareham NJ.** Novel type 2 diabetes susceptibility loci and measures of glycemic control in a population based study of healthy children. *Diabetologia* 2008. 51(Suppl 1):S31.
7. **Evans D.** Detecting gene-gene interactions in GWAS. Oral presentation, 44th EASD Annual Meeting, Rome, September 8-11, 2008.
8. **Antonarakis S.** Influence of human gene polymorphisms on gene expression. Oral presentation, 44th EASD Annual Meeting, Rome, September 8-11, 2008.
9. **Morris A, Prokopenko I, Akan P, Lango H, Zeggini E, Frayling TM, Deloukas P, Hattersley AT, McCarthy MI.** Fine-mapping type 2 diabetes causal variants on chromosome 9p21 in 2000 UK cases and 3000 unselected controls. *Diabetologia* 2008. 51(Suppl 1):S125.
10. **Lund SS, Tarnow L, Stehouwer CDA, Schalkwijk CG, Frandsen M, Szmidt UM, Hansen T, Pedersen O, Parkin HH, Vaag AA.** Impact of transcription-factor 7-like 2 (*TCF7L2*) gene-variants on the response to anti-hyperglycaemic treatment in non-obese patients with type 2 diabetes. *Diabetologia* 2008. 51(Suppl 1):S126.
11. **Hattersley AT, Freathy RM, Bennett AJ, Ring SM, Timpson NJ, Pouta A, Ruukonen A, Hypponen E, Power C, Elliott P et al.** Direct evidence to support the fetal insulin hypothesis as the type 2 diabetes risk alleles at the *CDKAL1* and *HHX-IDE* gene loci reduce birth weight. *Diabetologia* 2008. 51(Suppl 1):S127.
12. **Zeggini E, Lawrence R, Hattersley AT, Frayling TM, Elliott KS, McCarthy MI.** Extended analysis of genome wide

- scans provide clues about novel common and rare susceptibility loci for type 2 diabetes. *Diabetologia* 2008. 51(Suppl 1):S128.
13. **Elliott KS, Zeggini E, Rayner N, Groves C, Frayling T, Hattersley AT, McCarthy MI.** Studies of genome-wide association data support a genetic overlap between type 2 diabetes and prostate cancer. *Diabetologia* 2008. 51(Suppl 1):S128.
14. **Bowden DW, Lewis JP, Alfred ND, Langefeld CD, Divers J, Friedman BI.** Genetic analysis of "European" type 2 diabetes SNPs from whole genome association studies in African Americans. *Diabetologia* 2008. 51(Suppl 1):S129/304.