SHORT REPORT

Type 2 diabetes susceptibility loci in the Ashkenazi Jewish population

**Michal Bronstein · Anne Pisanté ·
Benjamin Yakir · Ariel Darvasi**

Abstract   Until last year, type 2 diabetes (T2D) suscepti-bility loci have hardly been identified, despite great effort. Recently, however, several whole-genome association (WGA) studies jointly uncovered 10 robustly replicated loci. Here, we examine these loci in the Ashkenazi Jewish (AJ) population in a sample of 1,131 cases versus 1,147 con-trols. Genetic predisposition to T2D in the AJ population was found similar to that established in the previous studies. One SNP, rs7754840 in the *CDKAL1* gene, presented a sig-nificantly stronger effect in the AJ population as compared to the general Caucasian population. This may possibly be due to the increased homogeneity of the AJ population. The use of the SNPs considered in this study, to identify individuals at high (or low) risk to develop T2D, was found of limited value. Our study, however, strongly supports the robustness of WGA studies for the identification of genes affecting complex traits in general and T2D in particular.

Introduction

Type 2 diabetes [T2D (MIM 222100)] is a common disease and a major medical burden on society (Narayan et al. 2006). As a means to better understand the disease towards the development of novel therapies and prophylactic treat-ments, studying the genetic basis of T2D has for long been a top scientific priority. Until recently, in spite of the significant heritability of the disease, only a small fraction of the genes affecting susceptibility to the disease have been identified. Last year, however, a novel technology was applied at its currently full capacity to decipher the genetic makeup of T2D. Several whole-genome association (WGA) studies were conducted on thousands of samples (Salonen et al. 2007; Saxena et al. 2007; Scott et al. 2007; Steinthorsdottir et al. 2007; Zeggini et al. 2007; Sladek et al. 2007; The Wellcome Trust Case Control Consortium 2007). Subsequently, a combined analysis of several WGA studies resulted in the identification of 10 loci, all of them robustly replicated as T2D risk factors (Scott et al. 2007). T2D is the first extensively studied disease, using the WGA approach. Consequently, it is important to examine how these findings replicate in the other populations.

In the current study, we looked at these 10 loci and assessed their effect on T2D in the Ashkenazi Jewish (AJ) population. The AJ population is a unique population in terms of its demographic history and genetic architecture. Historical studies suggest that the AJ population descended from a small number of founders and then rapidly expanded 700–1,000 years ago. The AJ population founder bottleneck eliminated many rare haplotypes, causing a reduction in the population's genetic heterogeneity and a modest increase in linkage disequilibrium (LD) (Shifman and Darvasi 2001; Olshen et al. 2008). Therefore, a replication study in the AJ population appears both important and efficient.

M. Bronstein · A. Pisanté · A. Darvasi (✉)
Department of Genetics,
The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
e-mail: arield@cc.huji.ac.il

B. Yakir
Department of Statistics, The Hebrew University of Jerusalem,
91905 Jerusalem, Israel

Materials and methods

Subjects

The 1,131 T2D subjects included in the study were col-lected in Israel by the patients' physicians in specialized

clinics. The 1,147 healthy controls were collected from the Israeli blood bank. All subjects analyzed are of AJ ancestry, as self declared for all four grandparents. All samples are part of the Hebrew University Genetic Resource (HUGR), www.hugr.org.

SNP genotyping

Genotyping was carried out using the KASPar technology by KBioscience (http://www.kbioscience.co.uk).

Analysis

A standard chi-square allele test was performed to calculate one-tail P values using the R environment. The statistical power for testing association was estimated based on the allele frequencies in the AJ control sample and the odds ratios given by Scott et al. (2007), using formula (12.2) on page 285 of Siegmund and Yakir (2007). To assess the predictive value of the seven SNPs exhibiting a T2D susceptibility effect ($P < 0.1$, Table 1) we fitted, similar to the simulation analysis conducted by Scott et al. (2007), a logistic regression model to our data (assuming additive non-interacting effects), simulating nine times as much weight to AJ healthy individuals as compared to AJ T2D patients, in order to reflect the underlying assumption of a 0.1 population prevalence for the disease. Likewise, for the frequency estimation in the population of subjects in each risk interval, healthy subjects were counted nine times and cases only once. The standard deviation of the frequency estimation in each risk interval was calculated using 1,000 bootstrap samples.

Results

Average call rate was above 98% for all SNPs and for all cases and controls, except for rs5219 that had a 93% call rate in the controls. Five of the SNPs here genotyped were also independently genotyped for approximately 200 of the cases used in the current study, as part of our T2D WGA study using Sentrix HumanHap300 BeadChips and Infinium II genotyping assays by Illumina (Salonen et al. 2007). Concordant genotypes were observed for 99.8% of the genotypes called by both technologies. Except for rs5219, none of the other SNPs exhibited a significant deviation ($P < 0.05$) from Hardy–Weinberg equilibrium (in either cases or controls). As presented in Table 1, allele frequencies in the controls of the AJ population were generally only slightly different from those of the general Caucasian population. Table 1 presents the association results for the 10 SNPs analyzed based on 1,131 cases versus 1,147 controls. We tested for allele association of the previously established risk allele and therefore present the one-tail P value. The expected power to detect these SNPs at $P = 0.05$ (presented for each SNP in Table 1) was on average 0.65. Therefore, nominal statistical significance is not necessarily expected for every SNP in our study. Nevertheless, in seven of the 10 SNPs the P value was below 0.1; sevenfold more than what is expected by chance. The rs5219 SNP was the only SNP to exhibit an effect in the opposite direction to what has been reported in the previous studies. This can possibly be interpreted as a consequence of a biased allele calling rate between the cases and controls for this particular SNP. For example, if the difference in allele calling between the cases and the controls was caused by a higher proportion of genotyping fails for the CC

Table 1 The association results for the 10 SNPs analyzed based on 1,131 cases versus 1,147 controls

SNP	Gene	OR[a]	95% CI[b]	One-tail P value	Power[c]	Meta-analysis OR[d]
rs7903146	TCF7L2	1.27 (T:0.28)	1.12–1.43	0.0001	1.00	1.37 (T:0.18)
rs1111875	HHEX	1.03 (C:0.64)	0.92–1.16	0.30	0.64	1.13 (C:0.52)
rs13266634	SLC30A8	1.14 (C: 0.74)	1.00–1.30	0.02	0.51	1.12 (C:0.61)
rs9300039	–	1.03 (A:0.07)	0.81–1.32	0.58	0.53	1.25 (C:0.89)
rs1801282	PPARG	1.22 (C:0.93)	0.95–1.57	0.06	0.28	1.14 (C:0.82)
rs5219	KCNJ11	1.13 (C:0.61)	1.00–1.28	0.97	0.7	1.14 (T:0.46)
rs8050136	FTO	1.09 (A:0.50)	0.97–1.23	0.08	0.83	1.17 (A:0.38)
rs4402960	IGF2BP2	1.09 (T:0.39)	0.97–1.23	0.08	0.7	1.14 (T:0.30)
rs10811661	CDKN2A/B	1.15 (T:0.84)	0.98–1.35	0.05	0.67	1.20 (T:0.85)
rs7754840	*CDKAL1*	1.29 (C:0.33)	1.14–1.45	0.00003	0.57	1.12 (C:0.36)

[a] Odds ratio, the risk allele and its frequency in the controls is indicated in parentheses

[b] Confidence interval

[c] Expected power to detect the effects reported by Scott et al. (2007) with allele frequencies and sample sizes of the current study

[d] Meta-analysis odds ratio, the risk allele and its frequency in the controls is indicated in parentheses, from Scott et al. (2007)

genotype in the controls, the effect of this SNP could disappear. Therefore, we will not refer to this SNP hereafter. One of the polymorphisms tested here, a SNP (rs7754840) in the *CDKAL1* gene, exhibited a significantly increased OR as compared to the average OR calculated in the meta-analysis of the general Caucasian population (Scott et al. 2007). This may occur as a consequence of the homogeneity of the AJ population, differences in LD patterns or simply by chance. Figure 1 presents the LD patterns around this SNP in the AJ population and in the Caucasian population. Data for the LD patterns were obtained for the AJ population from a previous study on this population (Salonen et al. 2007) and for the Caucasian population from the Hap-Map database (www.hapmap.org). As shown in Fig. 1, LD patterns for the two populations are similar, thus unlikely to be the source of the difference in ORs. Interestingly, in a totally independent study of 513 T2D cases versus 475 controls, again all collected in the AJ population, the estimated OR was 1.3, with a 95% confidence interval of 1.08–1.56 (B. Glaser, personal communication). This result is strikingly similar to the OR here observed (1.29), strengthening the possibility that the higher OR observed in the AJ population is a true effect. All other SNPs in the current study (Table 1) did not differ significantly with regards to OR when compared to the meta-analysis values.

Figure 2 presents the distribution of the various risk levels, given genetic information for the seven SNPs that exhibit an effect ($P < 0.1$, Table 1). Each of the seven loci may increase or decrease the risk to develop T2D, and the particular combination of genotypes at those seven may jointly alter the individual's risk. Based on the observed OR and assuming a population prevalence of 10%, one can estimate the risk of any individual to develop T2D given the particular genotypes at all the risk loci. With allele and genotype frequencies in hand for all the SNPs, one can also estimate the frequency in the population of such an individual with a particular seven-SNP genotype combination. Grouping together individuals with similar risks and presenting their combined frequency in the population serves

Fig. 2 Distribution of the risk of developing T2D in the AJ population. The *y*-axis presents the percentage of the population carrying a given risk to develop T2D (*x*-axis) as estimated from the genetic information at seven loci. Each value on the *x*-axis refers to the central point of a 0.02-wide interval. Standard error *bars* are also shown

as an insight to the value of the genetic information for identifying individuals at high or low risk to develop T2D. Without any genetic information, the risk of an individual to develop T2D is 0.1 in this model. As shown in Fig. 2, even with genetic information available, most individuals will carry a risk close to that base line (i.e. between 0.08 and 0.12). Only a small percentage (about 2%) will carry a significantly lower risk (about 0.05) and another small percentage (about 1%) will carry a significantly higher risk (about 0.2).

Discussion

Recent advances provide the opportunity to uncover the genetic basis of common diseases. In the current study, we looked at well-established associations recently discovered, using WGA in T2D. The associations were tested in a large cohort of AJ cases and controls. Our cases and controls are unlikely to suffer from population stratification or other systematic difference between them (causing false positive

Fig. 1 LD patterns at the *CDKAL1* locus in the Caucasian population (*upper figure*) and the AJ population (*bottom figure*). The location of the SNP tested in this study is indicated by an *arrow*
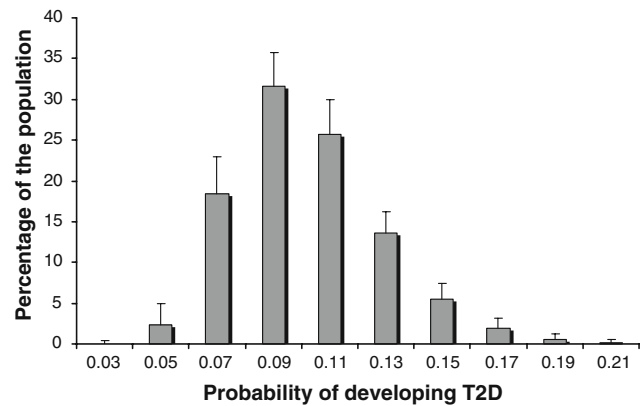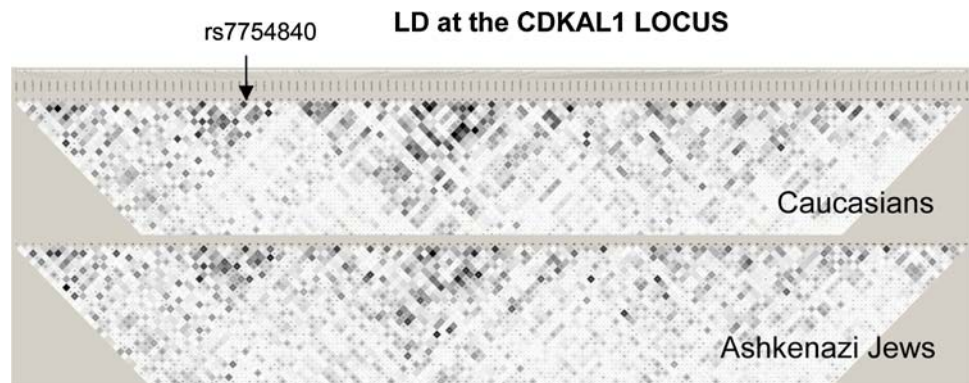
results), as all samples have been collected from the AJ population at the same small region in central Israel. Also, WGA studies that have been conducted on such sample collections (including a subset of the samples used in the current study) did not exhibit false-positive results (Salonen et al. 2007; Shifman et al. 2008). The genetic effects tested were typically replicated with similar OR. Furthermore, our data cannot confidently discard any of the previously reported risk loci as not affecting susceptibility for T2D in the AJ population. One locus, rs7754840 in the *CDKAL1* gene, exhibits a stronger effect in the AJ population compared to that of the general Caucasian population. This may be a result of the specific homogeneous genetic background of the AJ population (Shifman and Darvasi 2001). It is worth noting that OR can also be affected by covariates such as sex or body mass index (BMI). We did not include covariates in our analysis for comparative purposes but also since phenotypic records of the controls were limited and did not include parameters such as BMI. Until recently, the identification of genes affecting common diseases suffered from inconsistencies and lack of replication. It is very reassuring to see that well-designed studies, such as the meta-analysis reported by (Scott et al. 2007), are indeed reproducible.

Consistent with the previous findings showing that such set of SNPs may explain only a small fraction (2–3%) of the trait variation (Saxena et al. 2007), we have also found that the predictive value of these findings is relatively limited. For most individuals, the genotypic information will only modestly alter their baseline risk to develop T2D. Only for a small fraction of the population (about 3%), the genetic information may indicate a decrease or increase of the baseline risk by twofold (upwards or downwards). Consequently, at present, the use of genetic information by clinicians for the identification of patients at risk to develop T2D does not seem too relevant. Nevertheless, it is important to emphasize that the value of gene discovery lies primarily in the understanding of the disease as a means to develop efficient therapies.

References

Narayan KM, Boyle JP, Geiss LS, Saaddine JB, Thompson TJ (2006) Impact of recent increase in incidence on future diabetes burden: U.S., 2005–2050. Diabetes Care 29:2114–2116

Olshen AB, Gold B, Lohmueller KE, Struewing JP, Satagopan J, Stefanov SA, Eskin E, Kirchhoff T, Lautenberger JA, Klein RJ, Friedman E, Norton L, Ellis NA, Viale A, Lee CS, Borgen PI, Clark AG, Offit K, Boyd J (2008) Analysis of genetic variation in Ashkenazi Jews by high density SNP genotyping. BMC Genet 9:14

Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, Hypponen J, Korhonen VP, Asikainen J, Devine C, Tuomainen TP, Luedemann J, Nauck M, Kerner W, Stephens RH, New JP, Ollier WE, Gibson JM, Payton A, Horan MA, Pendleton N, Mahoney W, Meyre D, Delplanque J, Froguel P, Luzzatto O, Yakir B, Darvasi A (2007) Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. Am J Hum Genet 81:338–345

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345

Shifman S, Darvasi A (2001) The value of isolated populations. Nat Genet 28:309–310

Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A (2008) Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. PLoS Genet 4:e28

Siegmund D, Yakir B (2007) The statistics of gene mapping. Springer, NY

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885

Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 39:770–775

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341