university of groningen

umcg:

University of Groningen

A variant in CDKAL1 influences insulin response and risk of type 2 diabetes

Steinthorsdottir, V.; Thorleifsson, G.; Reynisdottir, I.; Benediktsson, R.; Jonsdottir, T.; Walters, G.B.; Styrkarsdottir, U.; Gretarsdottir, S.; Emilsson, V.; Ghosh, S.

Published in:
Nature Genetics

DOI:
10.1038/ng2043

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S., Baker, A., Snorradottir, S., Bjarnason, H., Ng, M. C., Hansen, T., Bagger, Y., Wilensky, R. L., Reilly, M. P., Adeyemo, A., ... Stefansson, K. (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics*, *39*(6), 770-775. https://doi.org/10.1038/ng2043

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 10-11-2026

nature genetics

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes

Valgerdur Steinthorsdottir[1,15], Gudmar Thorleifsson[1,15], Inga Reynisdottir[1], Rafn Benediktsson[2,3], Thorbjorg Jonsdottir[1], G Bragi Walters[1], Unnur Styrkarsdottir[1], Solveig Gretarsdottir[1], Valur Emilsson[1], Shyamali Ghosh[1], Adam Baker[1], Steinunn Snorradottir[1], Hjordis Bjarnason[1], Maggie C Y Ng[4], Torben Hansen[5], Yu Bagger[6], Robert L Wilensky[7], Muredach P Reilly[7], Adebowale Adeyemo[8], Yuanxiu Chen[8], Jie Zhou[8], Vilmundur Gudnason[3], Guanjie Chen[8], Hanxia Huang[8], Kerrie Lashley[8], Ayo Doumatey[8], Wing-Yee So[4], Ronald C Y Ma[4], Gitte Andersen[5], Knut Borch-Johnsen[5,9,10], Torben Jorgensen[10], Jana V van Vliet-Ostaptchouk[11], Marten H Hofker[11,12], Cisca Wijmenga[13,14], Claus Christiansen[6], Daniel J Rader[7], Charles Rotimi[8], Mark Gurney[1], Juliana C N Chan[4], Oluf Pedersen[5,9], Gunnar Sigurdsson[2,3], Jeffrey R Gulcher[1], Unnur Thorsteinsdottir[1], Augustine Kong[1] & Kari Stefansson[1]

We conducted a genome-wide association study for type 2 diabetes (T2D) in Icelandic cases and controls, and we found that a previously described variant in the transcription factor 7-like 2 gene (*TCF7L2*) gene conferred the most significant risk. In addition to confirming two recently identified risk variants[1], we identified a variant in the *CDKAL1* gene that was associated with T2D in individuals of European ancestry (allele-specific odds ratio (OR) = 1.20 (95% confidence interval, 1.13–1.27), $P = 7.7 \times 10^{-9}$) and individuals from Hong Kong of Han Chinese ancestry (OR = 1.25 (1.11–1.40), $P = 0.00018$). The genotype OR of this variant suggested that the effect was substantially stronger in homozygous carriers than in heterozygous carriers. The ORs for homozygotes were 1.50 (1.31–1.72) and 1.55 (1.23–1.95) in the European and Hong Kong groups, respectively. The insulin response for homozygotes was approximately 20% lower than for heterozygotes or noncarriers, suggesting that this variant confers risk of T2D through reduced insulin secretion.

We recently described a variant in *TCF7L2* associated with T2D[2,3]. To look for additional genetic variants that increase the risk of developing T2D, we performed a genome-wide association study on Icelandic individuals with T2D using the Illumina HumanHap300 chip. We tested 313,179 SNPs individually for association with T2D in a sample of 1,399 individuals with T2D and 5,275 controls. We tested an additional 339,846 two-marker haplotypes identified as efficient surrogates ($r^2 > 0.8$) for a set of SNPs that were not included on the Hap300 chip but that were typed in the HapMap project[4]. In addition to analyzing the entire group of individuals with T2D, separately we tested 700 non-obese individuals with T2D and 531 obese individuals with T2D for association. Overall, we performed a total of 1,959,075 (653,025 variants × 3 phenotypes) tests. The results were adjusted for relatedness between individuals and potential population stratification by genomic control[5] (see Methods). A previously identified SNP, rs7903146, in *TCF7L2* gave the most significant results, with $OR = 1.38$ and $P = 1.82 \times 10^{-10}$ in all individuals with T2D. Although no other SNP or haplotype was significant after adjustment for the number of tests performed, we observed more borderline-significant signals than expected by chance alone (**Supplementary Fig. 1** online). A comprehensive follow-up strategy would require genotyping a large number of SNPs[6], so we decided to pursue the top signals quickly in a fast-tracking effort.

For each phenotype tested, we selected all single SNPs and two-marker haplotypes with $P < 0.00005$ for replication in a case-control sample from Denmark (Denmark B). After eliminating redundant markers, we selected a total of 46 SNPs for replication (**Supplementary Table 1** online). In addition, we included the five most significant nonsynonymous SNPs present on the Illumina Hap300 chip. Of these 51 SNPs, we successfully genotyped 47 in 1,110 Danish T2D cases and 2,272 controls. In the Danish group of all individuals with T2D, SNPs rs7756992 and rs13266634 stood out and were significantly replicated ($P = 0.00013$ and

[1]deCODE genetics, Sturlugata 8, 101 Reykjavik, Iceland. [2]Landspitali-University Hospital, 101 Reykjavik, Iceland. [3]Icelandic Heart Association, 201 Kopavogur, Iceland. [4]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong. [5]Steno Diabetes Center, DK-2820 Copenhagen, Denmark. [6]Center for Clinical and Basic Research A/S, DK-2750 Ballerup, Denmark. [7]University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [8]National Human Genome Center, Howard University, Department of Community and Family Medicine, Washington, DC 20060, USA. [9]Faculty of Health Science, University of Aarhus, DK-8000 Aarhus, Denmark. [10]Research Centre for Prevention and Health, Glostrup University Hospital, DK-2600 Glostrup, Denmark. [11]Department of Molecular Genetics, Maastricht University, 6200 MD Maastricht, The Netherlands. [12]Department of Pathology and Laboratory Medicine and [13]Department of Genetics, University Medical Center Groningen (UMCG), 9700 RB Groningen, The Netherlands. [14]Complex Genetics Section, Department of Biomedical Genetics, University Medical Centre Utrecht, 3508 AB Utrecht, The Netherlands. [15]These authors contributed equally to this work. Correspondence should be addressed to K.S. (kstefans@decode.is) or V.S. (vstein@decode.is).

Received 14 March; accepted 17 April; published online 26 April 2007; doi:10.1038/ng2043

Table 1 Association results for rs7756992 and rs13266634 in five T2D case-control groups of European ancestry and in case-control groups from Hong Kong and West Africa

	Controls		Affected individuals			
	Frq	AA/Aa/aa[b]	Frq	AA/Aa/aa[b]	OR (95% c.i.)	*P* value
Iceland (1,399/ 5,275)						
rs7756992 (G)	0.232	3,107/1,887/277	0.270	751/539/108	1.23 (1.10–1.37)	0.00021
rs13266634 (C)	0.646	700/2,339/2,236	0.685	143/596/660	1.19 (1.08–1.31)	0.0006
Denmark A (263/597)						
rs7756992 (G)	0.297	292/255/50	0.331	111/99/30	1.17 (0.93–1.47)	0.18
rs13266634 (C)	0.686	62/242/279	0.672	35/99/124	0.94 (0.75–1.17)	0.58
Denmark B (1,359/4,825)						
rs7756992 (G)	0.279	2,503/1,884/394	0.320	624/564/144	1.21 (1.10–1.33)	0.000054
rs13266634 (C)	0.673	555/1,997/2,204	0.692	128/566/639	1.09 (0.99–1.19)	0.073
Philadelphia (447/950)						
rs7756992 (G)	0.262	492/331/68	0.295	216/174/40	1.18 (0.98–1.42)	0.073
rs13266634 (C)	0.678	85/377/387	0.760	29/145/249	1.51 (1.25–1.81)	1.5×10^{-5}
The Netherlands (368/915)						
rs7756992 (G)	0.270	475/359/63	0.280	186/138/30	1.05 (0.86–1.29)	0.64
rs13266634 (C)	0.717	80/349/469	0.736	28/136/199	1.10 (0.91–1.33)	0.33
European ancestry combined[a] (3,836/12,562)						
rs7756992 (G)	0.258		0.295		1.20 (1.13–1.27)	7.7×10^{-9}
rs13266634 (C)	0.666		0.700		1.15 (1.08–1.22)	3.3×10^{-6}
Hong Kong (1,457/986)						
rs7756992 (G)	0.462	293/446/220	0.517	351/681/400	1.25 (1.11–1.40)	0.00018
rs13266634 (C)	0.523	214/497/259	0.566	276/686/464	1.19 (1.06–1.33)	0.0035
West Africa[a] (865/1,106)						
rs7756992 (G)	0.612	160/499/397	0.625	137/349/344	1.02 (0.92–1.14)	0.72
rs13266634 (C)	0.962	4/74/1004	0.971	2/45/804	1.26 (0.88–1.81)	0.21

Numbers in parentheses next to population names represent the number of individuals with T2D and controls, respectively. Also shown are the allelic frequency (Frq) and genotype counts in the affected and control individuals, the allelic OR with 95% confidence intervals (c.i.) and two-sided *P* values based on the multiplicative model.
[a]For the combined European ancestry groups and the five West African groups, ORs and *P* values were combined using a Mantel-Haenszel model, and frequency in affected individuals and controls was estimated as a weighted average over the different study groups. [b]For rs7756992, the genotype counts are for AA/AG/GG individuals; for rs13266634, the counts are for TT/TC/CC individuals.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

OR = 1.24 and *P* = 0.0012 and OR = 1.20, respectively; **Supplementary Table 2** online), compared with *P* = 0.00021 and OR = 1.23 and *P* = 0.000061 and OR = 1.19, respectively, in the initial Icelandic study. All of the other SNPs genotyped had *P* > 0.01 in the Danish group, and we chose not to pursue them further. The first SNP, rs7756992, is located in intron 5 of the CDK5 regulatory subunit associated protein 1-like 1 (*CDKAL1*) gene on 6p22.3. It resides in a large LD block of 201.7 kb that includes exons 1–5 of the *CDKAL1* gene and the minimal promoter region but no other known genes (**Fig. 1**). The second SNP, rs13266634, is a nonsynonymous SNP causing an arginine to tryptophan change at position 325 in the last exon of the solute carrier family 30 (zinc transporter), member 8 (*SLC30A8*) gene on 8q24. SLC30A8 is specific to the pancreas and is expressed in beta cells, where it facilitates accumulation of zinc from the cytoplasm into intracellular vesicles[7]. The risk allele of rs13266634 on 8q24 has recently been found to confer risk of T2D in a genome-wide association study of French individuals with T2D and controls[1]. Of other significantly associated SNPs in that study, we also replicated, in the initial Icelandic samples, association with two SNPs close to the *HHEX* gene (**Supplementary Table 3** online). However, in our samples, we did not replicate with significance the reported associations to markers in the *LOC387761* and *EXT2* genes also described in that study.

We typed rs7756992 and rs13266634 in three other T2D case-control groups of European ancestry from Denmark (Denmark A), the Netherlands and Philadelphia as well as in case-control groups from Hong Kong and West Africa. Furthermore, we expanded the size of the Denmark B study group mostly by increasing the number of genotyped controls. The association of the G allele of rs7756992 was replicated with significance in the Hong Kong case-control group (OR = 1.25; *P* = 0.00018; **Table 1**). Association in other study groups was not individually significant, but all were in the same direction. Because some of the replication groups are not very large individually, the study should be considered as a whole in order to meaningfully interpret the results. Specifically, the observed association from combining all five case-control groups of European ancestry gave an OR of 1.20 with a corresponding *P* value of 7.7×10^{-9} (**Table 1**). Given that approximately 2 million tests were performed in the initial genome scan, this association remained significant with Bonferroni adjustment[6]. Moreover, the Chinese data provided further support for the association. Even when combined with the West African data, which did not show a significant effect, it yielded a *P* value of 0.0050. Attempts at refining the association observed with rs7756992 by genotyping additional markers that correlate with the original signal in the HapMap CEPH (CEU) data set did not yield more significant results (**Supplementary Table 4** online). As we expected, the observed linkage disequilibrium was considerably lower for the West African population than for the Icelandic and Hong Kong groups (**Supplementary Table 4**). Further work is needed to determine if an associated variant with a higher OR than observed for rs7756992 can be identified in the West African group. In total, we genotyped 61 SNPs (of which 35 were on the Hap300 chip) in the LD block containing rs7756992 in the Icelandic case-control group (**Supplementary Table 5** online). After we adjusted for the observed association of rs7756992,

Table 2 Genotype-specific OR for rs7756992 and rs13266634

	Allelic OR	Genotype OR[a]				
	(95% c.i.)	00	0X (95% c.i.)	XX (95% c.i.)	P^b	PAR
European ancestry						
rs7756992 (G)	1.20 (1.13–1.27)	1	1.15 (1.06–1.24)	1.50 (1.31–1.72)	0.089	0.061
rs13266634 (C)	1.15 (1.08–1.22)	1	1.05 (0.93–1.19)	1.26 (1.10–1.43)	0.10	0.157
Hong Kong						
rs7756992 (G)	1.25 (1.11–1.40)	1	1.13 (0.97–1.31)	1.55 (1.23–1.95)	0.071	0.154
rs13266634 (C)	1.19 (1.06–1.33)	1	1.13 (0.96–1.34)	1.40 (1.11–1.76)	0.43	0.148

PAR, population attributable risk.

[a]Genotype OR for heterozygous (0X) and homozygous carriers (XX) compared with noncarriers (00). [b]Test of the multiplicative model (the null hypotheses) versus the full model (that is, the model that puts no constraints on the genotype-specific risks). This test has one degree of freedom.

none of the other SNPs were significantly associated with T2D. As was the case for rs7756992, the association with T2D for allele C of the non-synonymous SNP rs13266634 was replicated with significance in two of the five additional groups (from Philadelphia and Hong Kong) (**Table 1**). Even though the OR for Denmark B decreased with the larger sample size, and the estimated effect was in the opposite direction (only slightly, and nonsignificantly) for Denmark A, the combined results from all study groups of European ancestry yielded a P value of 3.3×10^{-6} and an OR of 1.15 (**Table 1**).

In the Icelandic study, the observed association to rs7756992 was greater in non-obese individuals with T2D (OR = 1.37 (1.20–1.57); $P = 9.0 \times 10^{-6}$) than in the group of all individuals with T2D (OR = 1.23 (1.10–1.37); P = 0.00021) (**Supplementary Table 1** and **Table 1**). We also observed a higher OR in non-obese individuals than in obese individuals with T2D for this variant in the other populations studied. For the combined populations of European origin, the OR was 1.24 (1.15–1.33) with $P = 3.0 \times 10^{-8}$ for the non-obese individuals with T2D compared with OR = 1.14 (1.04–1.25) and P = 0.004 for the obese group. We saw an even stronger effect in the Hong Kong non-obese T2D group (OR = 1.36 (1.19–1.56); $P = 7.48 \times 10^{-6}$) compared with the obese group (OR = 1.13 (0.98–1.30); P = 0.094). For the Hong Kong group, obesity was defined as a body mass index (BMI) ≥25. Furthermore, examination of the controls showed a very weak, but significant, negative correlation of the variant with BMI, a result that needs further confirmation. Most notably, the combined results indicate that this variant does not confer increased risk of T2D by increasing BMI.

We estimated genotype ORs for each of the two loci (**Table 2**). For the combined study of populations of European descent, the OR for heterozygotes for rs7756992 was 1.15 (1.06–1.24), which is smaller than that predicted by the multiplicative model, compared with an OR of 1.50 (1.31–1.72) for homozygotes, which is larger than that predicted by the multiplicative model. We observed similar results for the Hong Kong samples (**Table 2**). Combining the European and Hong Kong data, we were able to reject the multiplicative model (P = 0.011). A multiplicative model for the genotype relative risk provided an adequate fit for rs13266634.

The function of the gene product of *CDKAL1* is unknown. However, the protein product is similar to another protein, CDK5 regulatory subunit–associated protein 1 (encoded by *CDK5RAP1*). *CDK5RAP1* is expressed in neuronal tissues, where it inhibits cyclin-dependent kinase 5 (CDK5) activity by binding to the CDK5 regulatory subunit p35 (ref. 8). In pancreatic beta cells, CDK5 has been shown to have a role in the loss of beta cell function under glucotoxic conditions[9]. Furthermore, inhibition of the CDK5/p35 complex prevents a decrease of insulin gene expression that results from glucotoxicity[10]. It is tempting to speculate that CDKAL1 may have a role in the inhibition of the CDK5/p35 complex in pancreatic beta cells similar to that of CDK5RAP1 in neuronal tissue. Reduced expression of *CDKAL1* or reduced inhibitory function thus could lead to an impaired response to glucotoxicity.

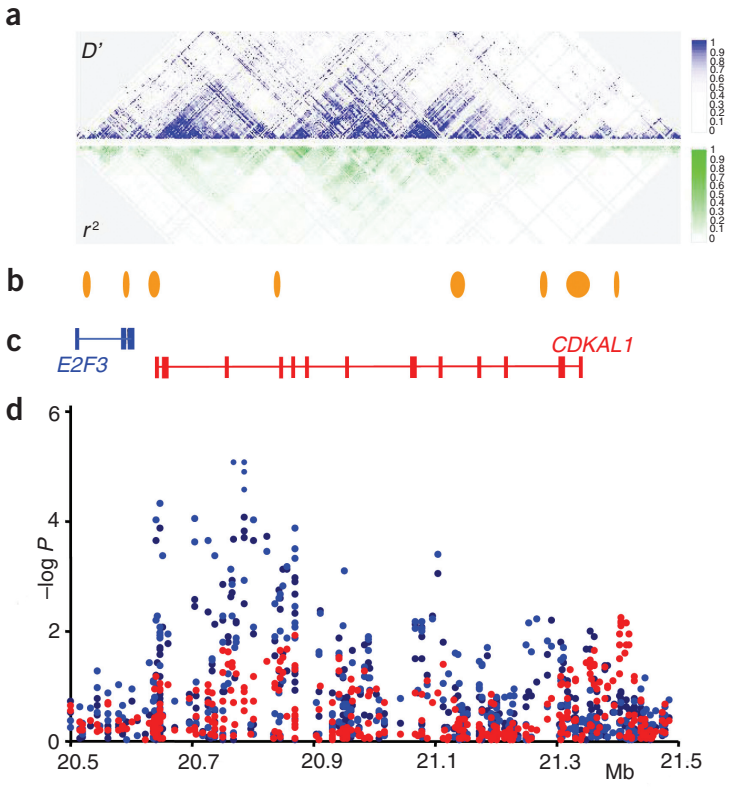

Figure 1 Schematic view of the association of T2D with variants in the 6p22.3 region. (**a**) Pairwise correlation structure in a 1-Mb interval (20.5–21.5 Mb, NCBI build34) on chromosome 6. The upper plot includes pairwise D′ for 1,047 common SNPs (with MAF >5%) from the HapMap release 19 for the CEU population; the lower plot includes pairwise r^2 values for the same set of SNPs. (**b**) Location of recombination hotspots in this interval based on the HapMap data set[26]. (**c**) Location of exons (vertical bars) of the two genes, *E2F3* (blue) and *CDKAL1* (red), that map to the interval. (**d**) Schematic view of the genome-wide association results in the interval for all T2D cases (black dots), non-obese T2D cases (blue dots) and obese T2D cases (red dots), respectively. –log P is plotted (where P is the adjusted P value) against the chromosomal location of the markers. All four panels use the same horizontal Mb scale indicated at the bottom of **d**.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

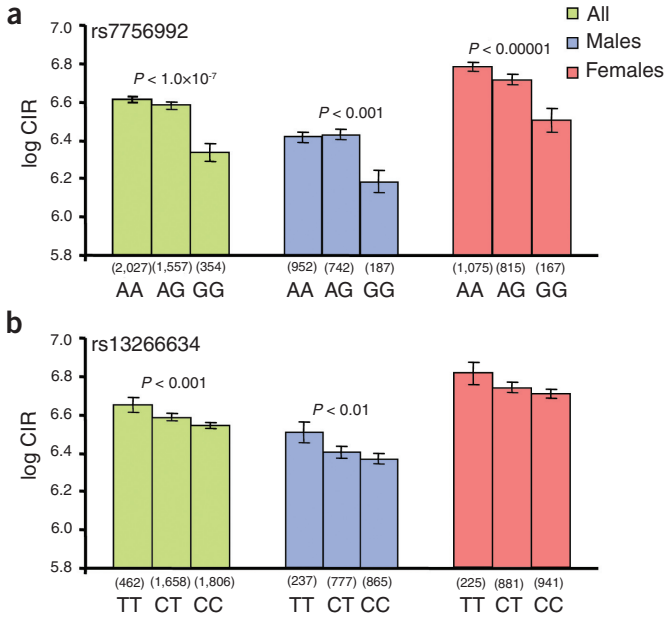

Figure 2 Association of rs7756992 and rs13266634 with insulin secretion. Mean log-transformed insulin secretion levels, estimated by corrected insulin response (CIR; see Methods), for the three different genotypes of the two SNPs, rs7756992 and rs13266634. Results are shown for 3,982 individuals (231 T2D cases and 3,751 controls) from the Danish Inter99 study that had an oral glucose tolerance test. Results are shown for all individuals and for males and females separately. The number of individuals analyzed for each genotype is shown in parentheses under each column, and the s.e.m. is indicated by vertical bars. *P* values are from a two–degree of freedom *F* test of the null model of no difference among the three genotype states against the full model.

In this study, we found that *CDKAL1* is expressed in the rat pancreatic beta cell line INS-1 (data not shown). Further studies are needed to determine if the effect of *CDKAL1* on risk of T2D is exerted through this pathway.

Based on the predicted function of *CDKAL1* and the known function of *SLC30A8*, we would expect both rs7756992 and rs13266634 to affect insulin secretion. To evaluate the effects of the two SNPs on insulin secretion, we analyzed the effect of genotype status on corrected insulin response (CIR) in a set of individuals from the Inter99 study (part of Denmark B) that had undergone an oral glucose tolerance test. For rs7756992, we found that homozygous carriers of the risk allele had an estimated 22% lower CIR than the noncarriers ($P = 3.5 \times 10^{-9}$). By contrast, heterozygous carriers showed a very small (2%) and nonsignificant ($P = 0.23$) reduction of CIR compared with noncarriers (**Fig. 2**). Testing the null hypothesis of no difference among all three genotypic states against the full model gave a *P* value of 2.5×10^{-8}. Hence, the effect of the variant on CIR is highly significant and is close to recessive. This observation is consistent with the observed effect of the variant in disease risk (that is, the increased risk for the heterozygous carriers is very modest). Furthermore, the effect on CIR was present in both males and females (**Fig. 2**) and in individuals with T2D as well as controls, and adjusting for BMI status did not affect the results (**Supplementary Table 6** online). The effect of rs13266634 on insulin response was smaller but significant, and for this risk variant, the reduction in CIR was consistent with an additive effect. We did not observe any effect on insulin sensitivity for either variant (**Supplementary Table 6**). For both variants, we obtained similar results by measuring insulin response in the form of the insulinogenic index (see Methods).

Based on our data from all five groups of European ancestry for *TCF7L2*, *CDKAL1* and *SLC30A8*, we found that the *TCF7L2* risk variant and the *CDKAL1* risk variant were positively correlated within the populations with T2D ($P = 0.0057$). Given that both have apparently stronger effects for non-obese cases, this suggests that they might work through a similar pathway. However, further investigation is necessary to confirm and understand this apparent correlation. By contrast, the risk conferred by the *SLC30A8* variant is consistent with its effects being multiplicative with the joint effects of *TCF7L2* and *CDKAL1*.

Considering that our fast-tracking strategy is not expected to be comprehensive and that the susceptibility variants identified so far for T2D, including the variant in *TCF7L2*, explain only a small fraction of the familial clustering of the disease, it is expected that there are many more variants with effects similar to those in *CDKAL1* and *SLC30A8* that have yet to be identified. Still, the identification of *CDKAL1* as a susceptibility gene for T2D adds a new piece to the puzzle of how genetic factors predispose to T2D. Although the function of this gene remains to be elucidated, we have shown that a variant within the gene is correlated with insulin secretion. The similarity to CDK5RAP1 further indicates that CDKAL1 may facilitate insulin production under glucotoxic conditions through interaction with CDK5. In conclusion, we have identified a variant in *CDKAL1* that predisposes to T2D and that blunts the insulin response in a nearly recessive manner.

METHODS

Icelandic study population. The Icelandic T2D group has been described previously[11]. A total of 1,500 individuals with T2D were recruited for this genome-wide association study, which used the Infinium II assay method and the Sentrix HumanHap300 BeadChip (Illumina). Of these, 1,399 were successfully genotyped according to our quality control criteria and were used in the present case control-analysis; 531 of the genotyped cases were obese (BMI ≥ 30), 700 were non-obese (BMI < 30) and information on BMI was missing for 168 cases. The controls used in this study consisted of 599 controls randomly selected from the Icelandic genealogical database and 4,676 individuals from other ongoing genome-wide association studies at deCODE. Specifically, approximately 1,400 of the controls came from studies on prostate cancer, and about 1,100 came from studies on breast cancer; studies on anxiety, addiction, schizophrenia and infectious diseases provided approximately 500 controls each. The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Written informed consent was obtained from all affected individuals and controls.

Other study populations. The Danish female study group of 282 cases and 629 controls, herein termed Denmark A, was selected from the Prospective Epidemiological Risk Factor (PERF) study in Denmark[12]. This is a group of postmenopausal women who took part in various placebo-controlled clinical trials and epidemiological studies at the Center for Clinical and Basic Research. In follow-up examinations of 5,847 women in 2000–2001, we collected medical histories (including type I or type II diabetes), family histories and information on current or previous long-term use of drugs through personal interviews using a preformed questionnaire. If a subject was diagnosed with diabetes of either type I or type II, the date of diagnosis or treatment was also noted. The study was approved by the Ethical Committee of Copenhagen County and was in accordance with the principles of the Helsinki Declaration.

The second Danish study population (Denmark B) of 1,359 individuals with T2D and 4,858 controls with normal glucose tolerance was from the Steno Diabetes Center in Copenhagen (1,016 cases and 374 controls) and from the Inter99 population-based sample of 30- to 60-year-old individuals living in the greater Copenhagen area, sampled at Research Centre for Prevention and Health (343 affected individuals and 4,484 controls)[13]. Diabetes and pre-diabetes were diagnosed according to the 1999 World Health Organization (WHO) criteria. An oral glucose tolerance test was performed on participants in the Inter99 study as described[13]. Informed written consent was obtained from all subjects before participation. The study was approved by the Ethical Committee of Copenhagen County and was in accordance with the principles of the Helsinki Declaration.

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

The Philadelphia study population consisted of 468 individuals with T2D and 1,024 control individuals. The study population was selected from the PENN CATH study, a cross-sectional study of the association of biochemical and genetic factors with coronary atherosclerosis in a study population of consecutive individuals undergoing cardiac catheterization at the University of Pennsylvania Medical Center. T2D was defined as a history of fasting blood glucose ≥ 126 mg dl^{-1}, 2 h postprandial glucose ≥ 200 mg dl^{-1}, use of oral hypoglycemic agents or use of insulin and oral hypoglycemic agents in a subject older than 40 years. The University of Pennsylvania Institutional Review Board approved the study protocol, and all subjects gave written informed consent. All affected individuals and controls were of European ancestry. Ethnicity was determined through self-report and has been validated by genotyping of ethnicity markers[14].

The Dutch Breda study population consisted of 370 T2D affected individuals and 916 control individuals. The affected individuals were recruited in 1998–1999 in collaboration with the Diabetes Service Breda and 80 general practitioners from the region around Breda. All patients were diagnosed according to WHO criteria (plasma glucose levels >11.1 mmol l^{-1} or fasting plasma glucose levels ≥7.0 mmol l^{-1}) and underwent clinical and laboratory evaluations for their diabetes at regular 3-month intervals. The Medical Ethics Committee of the University Medical Centre in Utrecht approved the study protocol. All probands gave written informed consent and filled out a questionnaire on clinical data, including any diabetes-related medication as well as height and weight at present and at the age of 20. The controls were healthy Dutch blood bank donors of European origin.

All subjects in the Hong Kong study population were of southern Han Chinese ancestry and resided in Hong Kong. The cases consisted of 1,500 individuals with T2D selected from the Prince of Wales Hospital Diabetes Registry[15]. Of these, 682 had young-onset diabetes (age at diagnosis ≤ 40 years) with a positive family history. An additional 818 cases were randomly selected from the same registry. The controls consisted of 1,000 subjects with normal glucose tolerance (fasting plasma glucose < 6.1 mmol l^{-1}). Of these, 617 were recruited from members of the general population participating in a community-based cardiovascular risk screening program as well as from hospital staff. In addition, 383 subjects were recruited from a cardiovascular risk screening program for adolescents. Informed consent was obtained for each participating subject. This study was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong.

The African study population comes from the Africa America Diabetes Mellitus study, which was originally designed as an affected sibling pair study with enrollment of available spouses as controls. It has since been expanded to include other family members of the affected pairs and population controls. Recruitment strategies and eligibility criteria for the families enrolled in this report have been described previously[16]. This West African case-control series consisted of individuals from the Yoruba (233 affected individuals, 432 controls) and Igbo (237 affected individuals, 276 controls) groups from Nigeria, and from the Akan (257 affected individuals, 248 controls), Ewe (22 affected individuals, 30 controls) and Gaa-Adangbe (123 affected individuals, 141 controls) groups from Ghana. Further characteristics of the seven case-control groups used in this study are shown in **Supplementary Table 7** online.

The DNA used for genotyping in all replication study populations was the product of whole-genome amplification (GenomiPhi Amplification kit, Amersham) of DNA isolated from the peripheral blood.

Illumina genome-wide genotyping. All Icelandic case and control samples were assayed with the Infinium HumanHap300 SNP chips (Illumina), containing 317,503 tagging SNPs derived from phase I of the International HapMap project. Of the SNPs assayed on the chip, 4,324 SNPs were excluded because they showed either (i) a call rate lower than 95% in cases or controls; (ii) a minor allele frequency <1% in the population or (iii) significant distortion from Hardy-Weinberg equilibrium in the controls ($P < 0.001$). Any samples with yield <98% were excluded from the analysis. Thus, the final analyses presented in the text use 313,179 SNPs.

Single-SNP genotyping. All single-SNP genotyping was carried out at deCODE Genetics on the Centaurus (Nanogen) platform[17]. The quality of each Centaurus SNP assay was evaluated by genotyping each assay in the CEU and/or YRI HapMap samples and comparing the results with the HapMap data. Assays with a mismatch rate >1.5% were not used, and a linkage disequilibrium (LD) test was used for markers known to be in LD.

Association analysis. For association analysis, we used standard likelihood ratio statistics, implemented in NEMO software[18], to calculate two-sided *P* values and allele-specific ORs for each individual allele, assuming a multiplicative model (that is, that the two alleles are independent, or in Hardy-Weinberg-Equilibrium, within the population of affected individuals). This corresponds to a setting where the ratio of the risks for homozygous carriers (AA) and heterozygous carriers (Aa) is the same as the ratio of the risks for heterozygous carriers and noncarriers, or (risk(AA) / risk(Aa)) = (risk(Aa) / risk(aa)). Allelic frequencies, rather than carrier frequencies, are presented for the markers, and *P* values are given after adjustment for the relatedness of the subjects. When estimating genotype-specific OR (**Table 2**), we estimated genotype frequencies in the control population assuming HWE after checking that the data were not inconsistent with this assumption.

In general, allele and haplotype frequencies were estimated by maximum likelihood, and tests of differences between cases and controls were performed using a generalized likelihood ratio test[19]. This method is particularly useful in situations where there are some missing genotypes for the marker of interest, and genotypes of another marker that is in strong LD with the marker of interest are used to provide some partial information. This was used in the association tests presented in **Supplementary Table 4** to ensure that the comparison of the highly correlated markers was done using the same number of individuals. To handle uncertainties with phase and missing genotypes, maximum likelihood estimates, likelihood ratios and *P* values are computed directly for the observed data, and hence the loss of information owing to uncertainty in phase and missing genotypes is automatically captured by the likelihood ratios.

Results from multiple case-control groups were combined using a Mantel-Haenszel model[20] in which the groups were allowed to have different population frequencies for alleles and for genotypes but were assumed to have common relative risks.

For both the *CDKAL1* and *SLC30A8* variants (rs7756992 and rs13266634), we did not detect any significant differences in frequencies among the disease groups (see description of the Icelandic study population) that make up the Icelandic genome-wide control sets ($P = 0.13$ and 0.19, respectively).

Correction for relatedness of the subjects and genomic control. Some of the individuals in both the affected and control Icelandic groups are related to each other, causing the χ^2 test statistic to have a mean >1 and median >0.675[2]. We estimated the inflation factor by calculating the average of the 653,025 χ^2 statistics, which was a method of genomic control[5] to adjust for both relatedness and potential population stratification. The inflation factors were estimated as 1.287 for all affected individuals, 1.204 for non-obese affected individuals and 1.184 for obese affected individuals. In addition to estimating the correction factors for the test statistics using the method of genomic control, we also applied a simulation method where genotypes are simulated through the genealogy of 708,683 Icelanders[21]. Specifically, the simulation procedure estimates the correction factor that is required owing to the known relatedness among the study participants. Based on 100,000 simulated data sets, the estimates were 1.211 for all affected individuals, 1.157 for non-obese affected individuals and 1.136 for obese affected individuals. We were not surprised that the correction factors estimated based on genomic control were somewhat larger than those based on simulations through genealogy, but it is comforting to us that the differences were not very substantial. This comparison shows that most of the adjustment was due to the relatedness of the participants; however, the higher estimates from the genomic controls indicate that there is some additional correction (possibly owing to genotyping quality, missing data or population stratification that, although small, is detectable in the Icelandic population)[22]. We do feel that the estimates based on genomic control, when available, are the most appropriate and conservative, and hence we have used that in the presentation in the paper.

Quantitative analysis. Data from oral glucose tolerance tests on individuals from the Danish Inter99 study were used to calculate insulin secretion as corrected insulin response (CIR) using the following equation: (100 × insulin at 30 min) / (glucose at 30 min × (glucose at 30 min – 3.89 mmol))[23,24]. Insulin sensitivity was estimated as the reciprocal of the insulin resistance according to the homeostasis model assessment (HOMA): 22.5 / (fasting insulin × fasting glucose)[25]. The association between CIR (HOMA) and genotype status was tested using multiple regression, where the log-transformed CIR (HOMA) was taken as the

response variable. For the full model, the two explanatory variables were the indicator variables for the heterozygous carriers and the homozygous carriers so that the fitted coefficients corresponded to the estimated effects for each of the two genotypic states relative to the noncarriers. Apart from estimated effects, standard errors and *P* values calculated for each of the two explanatory variables separately, a two–degree of freedom *P* value based on an *F*-test was calculated to test the null model (no difference among all three genotypic states) against the full model (**Supplementary Table 6**). We adjusted for sex, age and affection status by including the appropriate terms as explanatory variables. For comparison, insulin secretion was also calculated in the form of the insulinogenic index[24] (insulin at 30 min – insulin at 0 min) / (glucose at 30 min – glucose at 0 min), yielding comparable results.

Requests for materials: kstefans@decode.is or vstein@decode.is

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the individuals with T2D and other study participants whose contribution made this work possible. We also thank the nurses at Noatun (deCODE's sample recruitment center) and personnel at the deCODE core facilities for their hard work and enthusiasm. The Denmark B studies were supported by a grant from Eugene 2. Support for the Africa America Diabetes Mellitus (AADM) study is provided by the US National Institutes of Health, including the National Center on Minority Health and Health Disparities (3T37TW00041-03S2), the National Institute of Diabetes and Digestive and Kidney Diseases (DK072128), the National Human Genome Research Institute and the National Center for Research Resources (RR03048). The Hong Kong Diabetes case-control study was supported by the Hong Kong Research Grants Committee Central Allocation Scheme CUHK 1/04C.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

© 2007 Nature Publishing Group http://www.nature.com/naturegenetics

1. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
2. Grant, S.F. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**, 320–323 (2006).
3. Helgason, A. *et al.* Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet.* **39**, 218–225 (2007).
4. Pe'er, I. *et al.* Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat. Genet.* **38**, 663–667 (2006).
5. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
6. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
7. Chimienti, F., Devergnas, S., Favier, A. & Seve, M. Identification and cloning of a beta-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. *Diabetes* **53**, 2330–2337 (2004).
8. Ching, Y.P., Pang, A.S., Lam, W.H., Qi, R.Z. & Wang, J.H. Identification of a neuronal Cdk5 activator-binding protein as Cdk5 inhibitor. *J. Biol. Chem.* **277**, 15237–15240 (2002).
9. Wei, F.Y. *et al.* Cdk5-dependent regulation of glucose-stimulated insulin secretion. *Nat. Med.* **11**, 1104–1108 (2005).
10. Ubeda, M., Rukstalis, J.M. & Habener, J.F. Inhibition of cyclin-dependent kinase 5 activity protects pancreatic beta cells from glucotoxicity. *J. Biol. Chem.* **281**, 28858–28864 (2006).
11. Reynisdottir, I. *et al.* Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am. J. Hum. Genet.* **73**, 323–335 (2003).
12. Tanko, L.B., Bagger, Y.Z., Nielsen, S.B. & Christiansen, C. Does serum cholesterol contribute to vertebral bone loss in postmenopausal women? *Bone* **32**, 8–14 (2003).
13. Jorgensen, T. *et al.* A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. *Eur. J. Cardiovasc. Prev. Rehabil.* **10**, 377–386 (2003).
14. Helgadottir, A. *et al.* A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. *Nat. Genet.* **38**, 68–74 (2006).
15. Yang, X. *et al.* Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care* **30**, 65–70 (2007).
16. Rotimi, C.N. *et al.* In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann. Epidemiol.* **11**, 51–58 (2001).
17. Kutyavin, I.V. *et al.* A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.* **34**, e128 (2006).
18. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
19. Rice, J.A. *Mathematical Statistics and Data Analysis* (Wadsworth, Belmont, California, 1995).
20. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
21. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).
22. Helgason, A., Yngvadottir, B., Hrafnkelsson, B., Gulcher, J. & Stefansson, K. An Icelandic example of the impact of population structure on association studies. *Nat. Genet.* **37**, 90–95 (2005).
23. Sluiter, W.J., Erkelens, D.W., Reitsma, W.D. & Doorenbos, H. Glucose tolerance and insulin release, a mathematical approach I. Assay of the beta-cell response after oral glucose loading. *Diabetes* **25**, 241–244 (1976).
24. Hanson, R.L. *et al.* Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *Am. J. Epidemiol.* **151**, 190–198 (2000).
25. Matthews, D.R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
26. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).