A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes

Xin Jin

Nature

LETTER

doi:10.1038/nature13425

A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes

Ida Moltke[1,2]*, Niels Grarup[3]*, Marit E. Jørgensen[4], Peter Bjerregaard[5], Jonas T. Treebak[6], Matteo Fumagalli[7], Thorfinn S. Korneliussen[8], Marianne A. Andersen[6], Thomas S. Nielsen[6], Nikolaj T. Krarup[3], Anette P. Gjesing[3], Juleen R. Zierath[6,9], Allan Linneberg[10], Xueli Wu[11], Guangqing Sun[11], Xin Jin[11], Jumana Al-Aama[11,12], Jun Wang[3,11,12,13,14], Knut Borch-Johnsen[15], Oluf Pedersen[3], Rasmus Nielsen[7,16], Anders Albrechtsen[1] & Torben Hansen[3,17]

The Greenlandic population, a small and historically isolated founder population comprising about 57,000 inhabitants, has experienced a dramatic increase in type 2 diabetes (T2D) prevalence during the past 25 years[1]. Motivated by this, we performed association mapping of T2D-related quantitative traits in up to 2,575 Greenlandic individuals without known diabetes. Using array-based genotyping and exome sequencing, we discovered a nonsense p.Arg684Ter variant (in which arginine is replaced by a termination codon) in the gene *TBC1D4* with an allele frequency of 17%. Here we show that homozygous carriers of this variant have markedly higher concentrations of plasma glucose ($\beta = 3.8$ mmol l^{-1}, $P = 2.5 \times 10^{-35}$) and serum insulin ($\beta = 165$ pmol l^{-1}, $P = 1.5 \times 10^{-20}$) 2 hours after an oral glucose load compared with individuals with other genotypes (both non-carriers and heterozygous carriers). Furthermore, homozygous carriers have marginally lower concentrations of fasting plasma glucose ($\beta = -0.18$ mmol l^{-1}, $P = 1.1 \times 10^{-6}$) and fasting serum insulin ($\beta = -8.3$ pmol l^{-1}, $P = 0.0014$), and their T2D risk is markedly increased (odds ratio (OR) = 10.3, $P = 1.6 \times 10^{-24}$). Heterozygous carriers have a moderately higher plasma glucose concentration 2 hours after an oral glucose load than non-carriers ($\beta = 0.43$ mmol l^{-1}, $P = 5.3 \times 10^{-5}$). Analyses of skeletal muscle biopsies showed lower messenger RNA and protein levels of the long isoform of TBC1D4, and lower muscle protein levels of the glucose transporter GLUT4, with increasing number of p.Arg684Ter alleles. These findings are concomitant with a severely decreased insulin-stimulated glucose uptake in muscle, leading to postprandial hyperglycaemia, impaired glucose tolerance and T2D. The observed effect sizes are several times larger than any previous findings in large-scale genome-wide association studies of these traits[2–4] and constitute further proof of the value of conducting genetic association studies outside the traditional setting of large homogeneous populations.

Genetic association studies have traditionally been performed in large homogeneous populations. However, several studies have shown that it can be valuable to use founder populations[5], and there are similar advantages to using small and historically isolated populations. These advantages include increased statistical power to detect associations, owing to extended linkage disequilibrium and to an increased probability that deleterious variants overcome their selective disadvantage and reach high allele frequencies as a result of substantial genetic drift over many generations. Therefore, we aimed at identifying genetic variants associated with glucose homeostasis in the Greenlandic population, which is a small and historically isolated founder population, by association mapping of four T2D-related traits: plasma glucose and serum insulin levels at fasting and 2 h after an oral glucose load.

We successfully genotyped 2,733 participants in the Inuit Health in Transition (IHIT) cohort[6], sampled from 12 regions in Greenland (Fig. 1a), with the Illumina Cardio-Metabochip[7] (Metabochip) (Extended Data Table 1). We observed a high degree of linkage disequilibrium compared with Europeans (Extended Data Fig. 1a) and a high degree of European admixture (Fig. 1b). Additionally, as a natural consequence of the fact that the cohort constitutes almost 5% of the population, we identified more than 1,000 close relationships (siblings or parent–offspring) (Extended Data Fig. 1b). Population structure can lead to both decreased statistical power and increased type II error rates[8]. To avoid the latter, we therefore performed association analyses using a linear mixed model, which takes both admixture and relatedness into account. Genomic control inflation factors showed no inflation (range, 0.937–0.995; Fig. 2a and Extended Data Fig. 2).

Discovery analyses were performed for up to 2,575 IHIT participants without previously known T2D using an additive model. We found that the minor allele of rs7330796 was strongly associated with a higher 2-h plasma glucose level ($P = 4.2 \times 10^{-17}$) and a higher 2-h serum insulin level ($P = 6.4 \times 10^{-10}$) (Fig. 2a, b and Extended Data Fig. 2). These associations were replicated in the B99 Greenlandic cohort[9] ($P = 4.2 \times 10^{-6}$ for 2-h plasma glucose and $P = 2.7 \times 10^{-5}$ for 2-h serum insulin).

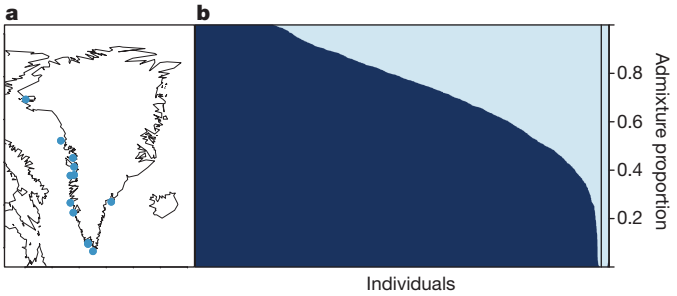


Figure 1 | Greenlandic study population. **a**, Sampling locations in Greenland. **b**, Estimated admixture proportions of Inuit and European ancestry. The admixture proportions were estimated assuming two source populations ($K = 2$). The estimates are both for the 2,733 individuals in the Greenlandic sample (IHIT), depicted to the left of the vertical line, and for 50 Danes, to the right of the vertical line.

[1]The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark. [2]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [3]The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark. [4]Steno Diabetes Center, 2820 Gentofte, Denmark. [5]National Institute of Public Health, University of Southern Denmark, 1353 Copenhagen, Denmark. [6]The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Integrative Physiology, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. [7]Department of Integrative Biology, University of California, Berkeley, California 94720, USA. [8]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark. [9]Department of Molecular Medicine and Surgery, Karolinska Institute, 171 77 Stockholm, Sweden. [10]Research Centre for Prevention and Health, Glostrup University Hospital, 2600 Glostrup, Denmark. [11]BGI-Shenzhen, Shenzhen 518083, China. [12]The Department of Genetic Medicine, Faculty of Medicine and Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia. [13]Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark. [14]Macau University of Science and Technology, Macau 999078, China. [15]Holbaek Hospital, 4300 Holbaek, Denmark. [16]Department of Statistics, University of California, Berkeley, California 94720, USA. [17]Faculty of Health Sciences, University of Southern Denmark, 5000 Odense, Denmark.
*These authors contributed equally to this work.

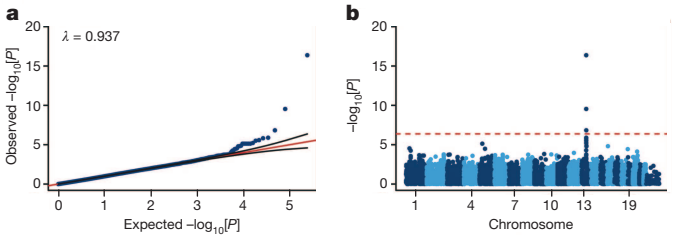


Figure 2 | Associations between 2-h plasma glucose levels and genotypes, as determined by Metabochip assay. Tests were performed using an additive linear mixed model in 2,540 individuals from the IHIT study who were not known to have T2D and for whom valid 2-h plasma glucose data were available. **a**, A quantile–quantile (QQ) plot of the observed $-\log_{10}[P]$ values (y axis) versus the $-\log_{10}[P]$ values expected under the null hypothesis of no association (x axis). The red line shows $x = y$, and the black lines demarcate the 95% confidence interval. The λ value is the genomic control inflation factor. **b**, A Manhattan plot of the observed $-\log_{10}[P]$ values. The dashed horizontal line indicates a 0.05 significance threshold after Bonferroni correction for multiple testing. The lowest P value is for rs7330796 on chromosome 13.

The rs7330796 variant was selected for inclusion on the Metabochip because it was in the top 5,000 candidate single nucleotide polymorphisms (SNPs) for association with waist-to-hip ratio[7] and it has not previously been reported to be associated with any of the four examined T2D-related traits or with T2D. The variant is located in intron 11 of *TBC1D4* and is neither in high linkage disequilibrium with neighbouring variants on the Metabochip nor situated inside a long range linkage disequilibrium block (Extended Data Fig. 3a, b). To locate the causal variation in the region, we performed exome sequencing of nine trios, and we identified four coding SNPs in high linkage disequilibrium ($r^2 > 0.8$) with rs7330796 (Extended Data Table 2). We genotyped these SNPs and found that p.Arg684Ter, a nonsense polymorphism in *TBC1D4* (c.2050C>T, rs61736969), was strongly associated with 2-h plasma glucose levels ($P = 3.6 \times 10^{-25}$) in the IHIT cohort (Table 1 and Fig. 3a). Conditional analyses demonstrated that p.Arg684Ter was significantly associated with 2-h plasma glucose and 2-h serum insulin levels when conditioning on rs7330796 ($P = 1.3 \times 10^{-9}$ and $P = 8.9 \times 10^{-9}$, respectively), whereas rs7330796 was not associated with the two traits when conditioning on p.Arg684Ter ($P = 0.47$ and $P = 0.09$) (Fig. 3a). Additionally, the mean 2-h plasma glucose levels for individuals with two copies of the minor rs7330796 allele increased with increasing Inuit admixture proportion (Extended Data Fig. 4a), which is expected if a variant is not causative and if the linkage disequilibrium patterns differ between Inuit and Europeans. By contrast, the same was not true for p.Arg684Ter (Extended Data Fig. 4b). These findings suggest that p.Arg684Ter is the causative variant.

The mean 2-h plasma glucose levels stratified by p.Arg684Ter genotype suggested that the variant mainly has an effect in homozygous carriers, indicating a recessive inheritance (Fig. 3b). We therefore also performed analyses using a recessive model: that is, we compared homozygous

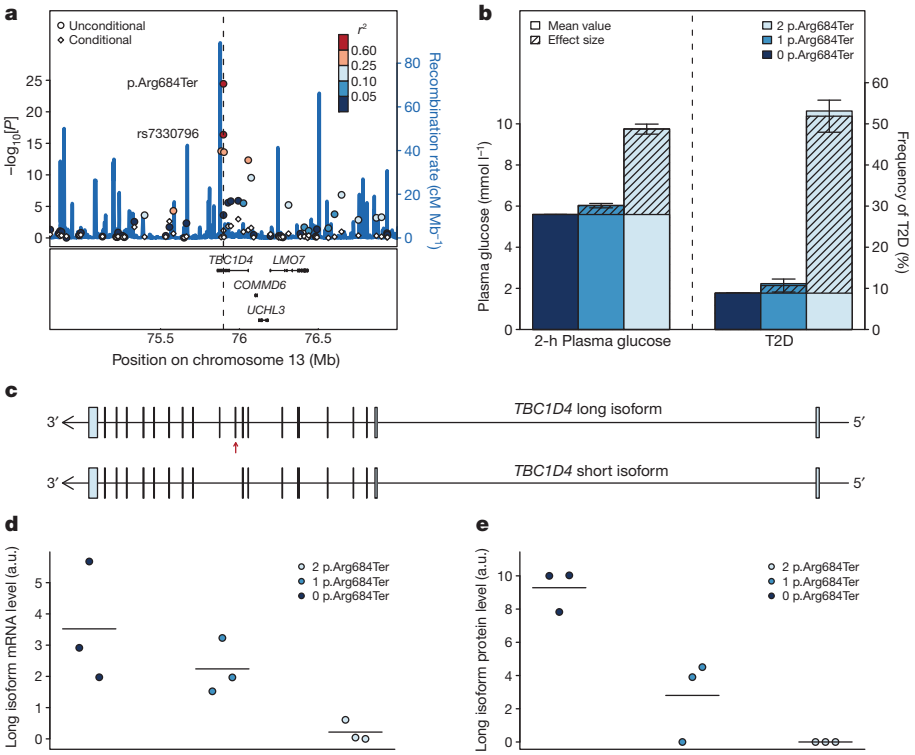


Figure 3 | Effect of the p.Arg684Ter nonsense polymorphism in *TBC1D4*. **a**, Association test results for all tested Metabochip SNPs in a 2-megabase (Mb) region around the p.Arg684Ter polymorphism (shown as a dashed vertical line). Each SNP is represented by a coloured circle. The position of the circle on the x axis shows the genomic position of the SNP. The position of the circle on the left y axis shows the $-\log_{10}[P]$ value of the SNP when testing for association with 2-h plasma glucose levels, as determined using an additive model. The colour of the circle indicates the extent of correlation (r^2) between the SNP and p.Arg684Ter. The circles representing p.Arg684Ter and rs7330796 are labelled (to the left of the circles). For every SNP, except for p.Arg684Ter, there is also a white diamond, which illustrates the P value obtained by testing for association conditional on p.Arg684Ter. The solid blue line illustrates the recombination rate from the Chinese HapMap (CHB) panel (in centimorgan (cM) per Mb, right y axis). The protein-coding genes in the genetic region are shown below the plot. **b**, The mean 2-h plasma glucose and the frequency of T2D for three genotypes (zero, one or two p.Arg684Ter alleles). Superimposed are the estimated effect sizes from the mixed model ± s.e.m. **c**, The two predominant isoforms of the *TBC1D4* gene illustrating which exons are transcribed: the long isoform, which has two additional exons (top), and the short isoform (bottom). Exons are depicted as boxes, and the location of the p.Arg684Ter polymorphism is indicated by a red arrow. **d**, The mRNA expression level of the long *TBC1D4* isoform in skeletal muscle from nine Greenlandic individuals (measured in arbitrary units (a.u.)). The mean value for each genotype group is shown as a horizontal line. **e**, Abundance of the long TBC1D4 protein isoform in skeletal muscle from nine individuals as quantified from western blot (measured in a.u.). The mean value for each genotype group is shown as a horizontal line.

Table 1 | Association of *TBC1D4* p.Arg684Ter with metabolic traits in the IHIT cohort

Trait	*n*	Additive model			Recessive model		
		$\beta_{s.d.}$ (95% CI)	β	*P*	$\beta_{s.d.}$ (95% CI)	β	*P*
Fasting plasma glucose (mmol l^{-1})	2,546	−0.13 (−0.2 to −0.064)	−0.048	0.00011	−0.45 (−0.63 to −0.27)	−0.18	1.1×10^{-6}
2-h Plasma glucose (mmol l^{-1})	2,511	0.35 (0.29 to 0.42)	1.1	**3.6×10^{-25}**	1.2 (0.99 to 1.4)	3.8	**2.5×10^{-35}**
Fasting serum insulin (pmol l^{-1})	2,546	−0.14 (−0.21 to −0.07)	−2.3	0.00012	−0.33 (−0.53 to −0.13)	−8.3	0.0014
2-h Serum insulin (pmol l^{-1})	2,511	0.29 (0.22 to 0.36)	57	**6.7×10^{-17}**	0.90 (0.71 to 1.1)	160	**1.5×10^{-20}**
Fasting serum C-peptide (pmol l^{-1})	2,546	−0.12 (−0.19 to −0.049)	−28	0.00092	−0.32 (−0.52 to −0.13)	−85	0.0012
2-h Serum C-peptide (pmol l^{-1})	2,511	0.30 (0.24 to 0.36)	360	**4.4×10^{-20}**	0.82 (0.65 to 1)	1,000	**8.5×10^{-20}**
HbA_{1C} (%)	2,692	0.015 (−0.047 to 0.078)	0.015	0.63	0.2 (0.036 to 0.37)	0.10	0.017
HOMA-IR (mmol l^{-1} × pmol l^{-1})	2,546	−0.15 (−0.22 to −0.076)	−0.078	6.4×10^{-5}	−0.36 (−0.56 to −0.16)	−0.37	0.00047
$ISI_{0,120}$	2,487	−0.32 (−0.39 to −0.26)	−0.47	**1.4×10^{-20}**	−1.0 (−1.2 to −0.86)	−1.4	**1.6×10^{-27}**
HOMA-B (%)	2,545	−0.085 (−0.15 to −0.018)	−2.3	0.013	−0.12 (−0.31 to 0.062)	−4.2	0.19
T2D (cases/controls)	220/1,810	0.083 (0.059 to 0.11)	0.083	**2.1×10^{-11}**	0.37 (0.3 to 0.44)	0.37	**1.6×10^{-24}**
Fasting serum HDL-cholesterol (mmol l^{-1})	2,702	0.032 (−0.035 to 0.099)	0.019	0.34	0.098 (−0.082 to 0.28)	0.064	0.29
Fasting serum total cholesterol (mmol l^{-1})	2,566	−0.013 (−0.081 to 0.056)	−0.018	0.71	0.25 (0.064 to 0.44)	0.30	0.0086
Fasting serum triglyceride (mmol l^{-1})	2,702	−0.040 (−0.11 to 0.032)	−0.020	0.27	0.022 (−0.17 to 0.22)	0.038	0.82
BMI (kg m^{-2})	2,673	−0.036 (−0.11 to 0.036)	−0.19	0.32	0.047 (−0.15 to 0.24)	0.25	0.63

Results are shown for an additive and a recessive genetic model. For each trait, *n* is the number of individuals with genotype data for the specific variant and phenotype data for the specific trait. $\beta_{s.d.}$ is the effect size estimated using quantile-transformed values of the trait (except for the binary trait T2D), and β is the effect size estimated using untransformed values. The *P* values were obtained from the quantile-transformed-value-based analyses. All *P* values $<1 \times 10^{-6}$ are marked in bold and were all successfully replicated in the B99 cohort (Extended Data Table 3). Individuals with known T2D were removed from the analysis of quantitative traits. Highly significant associations were also observed when removing both individuals with known and screen-detected T2D from the analyses (data not shown). BMI, body mass index; HDL, high-density lipoprotein; HOMA-IR, homeostasis model assessment-estimated insulin resistance.

carriers with carriers of other genotypes. These analyses demonstrated that homozygous carriers of p.Arg684Ter in the IHIT had a 3.8 mmol l^{-1} higher 2-h plasma glucose level ($P_{\text{recessive model (rec)}} = 2.5 \times 10^{-35}$) (Table 1). Although the main effect was seen when comparing homozygous p.Arg684Ter carriers with all other individuals, even heterozygous carriers displayed a 0.43 mmol l^{-1} higher 2-h plasma glucose level than non-carriers ($P = 5.3 \times 10^{-5}$) (Fig. 3b). To further investigate the metabolic implications of p.Arg684Ter, we analysed additional metabolic traits (Table 1). Analyses of 220 individuals with T2D and 1,810 non-diabetic control individuals showed a strong association of p.Arg684Ter with increased risk of T2D ($P_{\text{additive model (add)}} = 2.1 \times 10^{-11}$). Similar to the 2-h plasma glucose levels, the data suggested a recessive inheritance for T2D (OR_{rec}, 10.3; $P_{rec} = 1.6 \times 10^{-24}$) (Fig. 3b). Interestingly, when using an alternative definition of T2D that is based on recent HbA_{1C} criteria and does not include plasma glucose data, the association was modest ($P = 0.0084$). This finding is in line with the modest association of p.Arg684Ter with HbA_{1C} as a quantitative trait (Table 1). We also found that p.Arg684Ter was associated with decreased peripheral insulin sensitivity, as estimated by the Gutt insulin sensitivity index (ISI)[10] ($ISI_{0,120}$: $\beta_{add} = -0.32$ s.d., $P_{add} = 1.4 \times 10^{-20}$; $\beta_{rec} = -1.0$ s.d., $P_{rec} = 1.6 \times 10^{-27}$) (Table 1). We replicated these findings in the B99 cohort and found consistent results (Extended Data Table 3). Finally, we found associations of p.Arg684Ter with lower fasting plasma glucose and serum insulin levels in the IHIT cohort but with substantially lower effect sizes than the glucose-stimulated effects ($\beta_{add} = -0.048$ mmol l^{-1}, $P_{add} = 0.00011$; $\beta_{rec} = -0.18$ mmol l^{-1}, $P_{rec} = 1.1 \times 10^{-6}$; and $\beta_{add} = -2.3$ pmol l^{-1}, $P_{add} = 0.00012$; $\beta_{rec} = -8.3$ pmol l^{-1}, $P_{rec} = 0.0014$, respectively) (Table 1). Thus, our findings indicate that the p.Arg684Ter *TBC1D4* variant confers increased risk of a subset of diabetes that features deterioration of postprandial glucose homeostasis. In this context, it is of interest that 2-h glucose levels appear to be a better predictor of cardiovascular disease than do fasting plasma glucose levels[11]. The p.Arg684Ter variant showed no significant association with HOMA-B, an estimate of basal β-cell function. Similarly, no convincing associations were detected with measures of adiposity, fasting lipid levels or other components of metabolic syndrome (Table 1).

The impact of p.Arg684Ter in its recessive form on 2-h plasma glucose (3.8 mmol l^{-1}) and T2D risk (OR, 10.3) are several times larger than any effects that have been reported in large-scale genome-wide association studies for these traits[2–4]. Furthermore, the p.Arg684Ter polymorphism has a high population impact, as 3.8% of Greenlanders are homozygous carriers of the risk allele. In the IHIT cohort, 15.5% of the patients with T2D are homozygous carriers of the risk allele, in contrast to 1.6% among glucose-tolerant individuals, indicating that p.Arg684Ter accounts for more than 10% of all cases of T2D in Greenland. Between 40 and 60 years of age, more than 60% of the homozygous carriers have T2D, and this increases to more than 80% above the age of 60. The effect of p.Arg864Ter thus mirrors a Mendelian-disease-like pattern of inheritance.

The p.Arg684Ter variant has a minor allele frequency (MAF) of 17% in the IHIT cohort, and we estimated it to have a MAF of 23% and 0% in the unobserved Inuit and European populations that are ancestral populations to the Greenlanders. In comparison, this variant was found in only 1 Japanese individual (NA18989) out of the 1,092 individuals sequenced in the 1000 Genomes Project[12], and it was not present in exome sequencing data from 2,000 Danish individuals[13], 448 Han Chinese individuals or ~6,500 European and African American individuals[14]. Thus, the variant is not unique to the Greenlandic population but is probably common only among Greenlanders and other related populations. This finding raises the question of whether the variant has been favoured by natural selection or whether it has increased in frequency as a result of genetic drift. A test for selection showed weak evidence for positive selection (Extended Data Fig. 5 and Supplementary Notes 1).

TBC1D4, also known as AS160, acts as a mediator of insulin-stimulated Akt-induced glucose uptake through Rab-mediated regulation of GLUT4 mobilization[15]. *Tbc1d4*-knockout mice have decreased basal plasma glucose levels and are resistant to insulin-stimulated glucose uptake in muscle and adipose tissue[16]. Furthermore, the overall GLUT4 levels in these mice are markedly lower than those of *Tbc1d4*-sufficient mice[16]. Two isoforms of the *TBC1D4* gene have been reported: one encodes a full-length protein, and the other encodes a short form lacking exons 11 and 12 (Fig. 3c). It is predicted that the p.Arg684Ter variant results in termination of *TBC1D4* transcription in exon 11; thus, this variant is expected to affect only the long isoform. In line with previous findings[17], expression analyses in humans showed that while the short isoform of *TBC1D4* is widely expressed, the long isoform is primarily expressed in skeletal muscle and not in other major tissues associated with glucose metabolism, such as adipose tissue, the liver or the pancreatic islets (Extended Data Fig. 6a, b). Thus, it is unlikely that p.Arg684Ter affects the latter tissues. We measured the expression levels in skeletal muscle tissue in groups of three individuals carrying zero, one or two copies of p.Arg684Ter. As predicted, the levels of long *TBC1D4* isoform mRNA and protein decreased with increasing number of p.Arg684Ter alleles (Fig. 3d, e). The short isoform was observed at very low levels in skeletal muscle regardless of the genotype (Extended Data Fig. 6c), indicating that this isoform is unlikely to contribute to the observed phenotype. Further analyses showed that GLUT4 protein levels in the skeletal muscle decreased with increasing number of p.Arg684Ter alleles (Extended Data Fig. 6d). Thus, the phenotype of global *Tbc1d4*-knockout mice—lower fasting glucose levels and markedly lower insulin-stimulated glucose uptake than *Tbc1d4*-sufficient mice[16]—is comparable to the phenotype observed in homozygous carriers of the p.Arg684Ter variant. Our data indicate that disruption of the full-length TBC1D4 protein in skeletal muscle results in severely decreased insulin-stimulated

glucose uptake, leading to postprandial hyperglycaemia, impaired glucose tolerance and T2D.

The effect of *TBC1D4* p.Arg684Ter is in line with a reported familial case of postprandial hyperinsulinaemia caused by a *TBC1D4* p.Arg363Ter variant[18]. However, the reported p.Arg363Ter variant affects both TBC1D4 isoforms and consequently many tissues. Moreover, this variant has a large effect on insulin-stimulated glucose uptake in heterozygous carriers but not on fasting glucose levels. By contrast, the p.Arg684Ter variant discovered here has a large effect only in homozygous carriers and is restricted to the long isoform of TBC1D4, thereby affecting TBC1D4 signalling in skeletal muscle but not in β-cells, the liver or adipose tissue. Furthermore, the p.Arg684Ter variant affects fasting glucose levels. Finally, the high frequency of the nonsense variant in Greenlanders has allowed us to assess the physiological impact of this variant with high statistical confidence.

In summary, our study demonstrates the strength of conducting genetic association mapping outside the traditional setting of large homogeneous populations. We report a novel association of a common *TBC1D4* nonsense variant with T2D and elevated circulating glucose and insulin levels after an oral glucose load. The effect sizes of the variant markedly exceed previously reported associations of common genetic variants with metabolic traits. The variant leads to a prematurely terminated transcript of the long isoform of *TBC1D4*, which in homozygous carriers causes insulin resistance in skeletal muscle and confers a high risk of a subtype of T2D that is characterized by a deterioration of postprandial glucose homeostasis.

METHODS SUMMARY

Discovery association analyses were performed on the IHIT cohort data[6] and replicated with the B99 cohort data[9] (Extended Data Table 1). Participants underwent an oral glucose tolerance test, with plasma glucose and serum insulin levels measured at fasting and after 2 h. Diabetes was classified according to the World Health Organization. Samples were genotyped using the Cardio-Metabochip (Illumina)[7] with standard protocols. Samples with mis-specified gender, high rates of missing data and duplicates were removed using the software toolset PLINK. For association testing, we used a linear mixed model, implemented in the software GEMMA, to control for admixture and relatedness. Only participants without known diabetes were analysed (IHIT/B99: fasting plasma glucose, $n = 2{,}575/n = 1{,}064$; fasting serum insulin, $n = 2{,}575/n = 1{,}062$; 2-h plasma glucose, $n = 2{,}540/n = 845$; and 2-h serum insulin, $n = 2{,}540/n = 845$). All quantitative traits were quantile-transformed to a standard normal distribution; age and sex were included as covariates, and tests were performed using a likelihood ratio test. Effect sizes and their standard errors were estimated using a restricted maximum likelihood approach. Conditional analyses were performed by including the SNP that we conditioned on as an additional covariate. A meta-analysis was performed using the inverse-variance-based method. To estimate the OR for T2D under the recessive model, we used a linear mixed model without covariates, estimated α (the intercept) and β (the genotype effect) and set $\mathrm{OR} = \{(\alpha+\beta)/[1-(\alpha+\beta)]\}/[\alpha/(1-\alpha)]$. After the discovery studies, we selected nine trios that we inferred to have no European ancestry, enriching for carriers of the rs7330796 minor allele. We exome-sequenced the trios by using SureSelect capture (Agilent) followed by HiSeq2000 sequencing (Illumina), aligned the data using bwa software and called genotypes using SAMtools software. Variants that were in high linkage disequilibrium with rs7330796 were identified, genotyped in all individuals and tested for association. Admixture proportions were estimated using the software ADMIXTURE, assuming that there are two ancestral populations. Relatedness was estimated using the software RelateAdmix, which takes admixture into account. Methods for the biological studies are described in Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 January; accepted 30 April 2014.
Published online 18 June 2014.

1. Jørgensen, M. E., Bjerregaard, P. & Borch-Johnsen, K. Diabetes and impaired glucose tolerance among the Inuit population of Greenland. *Diabetes Care* **25,** 1766–1771 (2002).
2. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genet.* **44,** 991–1005 (2012).
3. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genet.* **44,** 981–990 (2012).
4. Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genet.* **44,** 659–669 (2012).
5. Rafnar, T. *et al.* Mutations in *BRIP1* confer high risk of ovarian cancer. *Nature Genet.* **43,** 1104–1107 (2011).
6. Jørgensen, M. E., Borch-Johnsen, K., Stolk, R. & Bjerregaard, P. Fat distribution and glucose intolerance among Greenland Inuit. *Diabetes Care* **36,** 2988–2994 (2013).
7. Voight, B. F. *et al.* The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8,** e1002793 (2012).
8. Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nature Genet.* **36,** 512–517 (2004).
9. Bjerregaard, P. *et al.* Inuit health in Greenland: a population survey of life style and disease in Greenland and among Inuit living in Denmark. *Int. J. Circumpolar Health* **62** (suppl. 1), 3–79 (2003).
10. Gutt, M. *et al.* Validation of the insulin sensitivity index ($ISI_{0,120}$): comparison with other measures. *Diabetes Res. Clin. Pract.* **47,** 177–184 (2000).
11. Cederberg, H. *et al.* Postchallenge glucose, A1C, and fasting glucose as predictors of type 2 diabetes and cardiovascular disease: a 10-year prospective cohort study. *Diabetes Care* **33,** 2077–2083 (2010).
12. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
13. Lohmueller, K. E. *et al.* Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *Am. J. Hum. Genet.* **93,** 1072–1086 (2013).
14. NHLBI. *NHLBI GO Exome Sequencing Project (Exome Variant Server)* http://evs.gs.washington.edu/EVS/.
15. Sano, H. *et al.* Insulin-stimulated phosphorylation of a Rab GTPase-activating protein regulates GLUT4 translocation. *J. Biol. Chem.* **278,** 14599–14602 (2003).
16. Wang, H. Y. *et al.* AS160 deficiency causes whole-body insulin resistance via composite effects in multiple tissues. *Biochem. J.* **449,** 479–489 (2013).
17. Baus, D. *et al.* Identification of a novel AS160 splice variant that regulates GLUT4 translocation and glucose-uptake in rat muscle cells. *Cell. Signal.* **20,** 2237–2246 (2008).
18. Dash, S. *et al.* A truncation mutation in *TBC1D4* in a family with acanthosis nigricans and postprandial hyperinsulinemia. *Proc. Natl Acad. Sci. USA* **106,** 9350–9355 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank I. Kleist, as well as all colleagues at the Physician's Clinic, Nuuk, Greenland. We also thank J. F. Wojtaszewski for comments and X. Zhou for discussions about GEMMA. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent research centre at the University of Copenhagen and is partly funded by an unrestricted donation from the Novo Nordisk Foundation. This project was also funded by the Danish Council for Independent Research (Medical Sciences), the Steno Diabetes Center and the Villum Foundation. The IHIT study was supported by Karen Elise Jensen's Foundation, NunaFonden, the Medical Research Council of Denmark, the Medical Research Council of Greenland and the Commission for Scientific Research in Greenland. None of the funding agencies had any role in the study design or in the collection or interpretation of the data.

Author Contributions T.H. and A.A. conceived and headed the project. I.M., R.N. and A.A. designed the statistical set-up, and T.H., N.G., A.P.G. and O.P. designed the experimental set-up for the DNA extraction, genotyping and sequencing. A.L. provided the Danish samples. M.E.J., P.B. and K.B.-J. provided the Greenlandic samples, collected and defined the phenotypes and provided context for these samples. I.M. and A.A. performed the admixture, relatedness and linkage disequilibrium analyses. I.M. carried out the statistical part of the association analysis and N.G. carried out the medical part, with input from A.A., T.H., O.P. and R.N. The Chinese samples were analysed by X.W., G.S., X.J., J.A.-A. and J.W. N.G. analysed the Danish samples. X.W., G.S., X.J., J.A.-A. and J.W. performed the library constructions and sequencing. T.S.K. performed the mapping and genotyping for the sequencing data. M.F. and R.N. performed the selection analysis. N.T.K. collected muscle biopsies, and J.T.T., T.S.N., M.A.A. and J.R.Z. experimentally analysed and interpreted the *TBC1D4* and *GLUT4* expression data. N.G., I.M., A.A. and T.H. wrote most of the manuscript, with input from R.N., O.P., M.E.J. and P.B. All authors approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H. (torben.hansen@sund.ku.dk) or A.A. (albrecht@binf.ku.dk).

METHODS

Study samples. The primary data set, the IHIT cohort ($n = 3{,}115$), was collected in 2005–2010 as part of a general population health survey of the Greenlandic population[6]. A random sample of adults in 12 regions, consisting of 21 selected towns and villages, was invited to participate (Fig. 1a). The data set used for replication ($n = 1{,}401$) was collected in 1999–2001 as part of a general population health survey (B99)[9]. Two hundred and ninety-five individuals overlapped between the two sample sets and were assigned to the B99 sample set. The clinical characteristics of the participants with genotype data are shown in Extended Data Table 1. Note that the sample size of ~3,000 is large enough to have the power (>80%) to detect common (MAF > 5%) variants with modest to high effect sizes (0.4 s.d.), comparable to the known genetic variations associated with lipid levels in Europeans. When estimating admixture proportions, in addition to the Greenlandic samples, we included a set of 50 Danish samples from the Inter99 study[19] to represent a European reference population.

Ethical considerations. The study has received ethics approval from the Commission for Scientific Research in Greenland (project 2011-13, ref. no. 2011-056978; and project 2013-13, ref.no. 2013-090702) and was conducted in accordance with the ethical standards of the Declaration of Helsinki, second revision. Participants gave written consent after being informed about the study orally and in writing.

Phenotypic data and biochemical variables. In IHIT, all participants underwent an oral glucose tolerance test (OGTT). In B99, participants above 24 years of age had blood samples taken when fasting, and participants above 35 years underwent an OGTT. At the baseline health examination, venous blood samples were drawn after an overnight fast of at least 8 h. After this, participants received a standard 75-g OGTT, with blood samples drawn 2 h after the glucose intake. Only fasting venous plasma glucose was measured in participants with known diabetes. Plasma glucose values (0 and 2 h) were analysed with the Hitachi 912 system (Roche Diagnostics). Serum insulin levels (0 and 2 h) were analysed by an immunoassay method excluding des-31,32 split products and intact proinsulin (AutoDELFIA, PerkinElmer).

Indices of insulin sensitivity and insulin secretion were derived from plasma glucose and serum insulin measures from the OGTT. Insulin sensitivity was estimated by the homeostasis model assessment (HOMA)[20] and by an insulin sensitivity index[10]: HOMA-IR = [(fasting glucose level × fasting insulin level)/6.945]/22.5, where insulin levels are expressed in pmol l^{-1} and glucose levels are expressed in mmol l^{-1}; insulin sensitivity index ($ISI_{0,120}$) = {[75,000 + (fasting glucose level × 18 – 2-h glucose level × 18) × 0.19 × weight]/120}/[0.5(fasting glucose level × 18 + 2-h glucose level × 18)]/log[(fasting insulin level/6.945 + 2-h insulin/6.945)/2], where insulin levels are expressed in pmol l^{-1}, glucose levels are expressed in mmol l^{-1} and weight is expressed in kg. Basal β-cell function was estimated by the HOMA-B index[20] calculated as HOMA-B = [20(fasting insulin level/6.945)]/(fasting glucose level − 3.5), where insulin levels are expressed in pmol l^{-1} and glucose levels are expressed in mmol l^{-1}.

Analyses of quantitative glycaemic traits were performed excluding individuals with previously diagnosed diabetes, leaving up to 2,575 IHIT participants for discovery analysis (fasting plasma glucose, $n = 2{,}575$; fasting serum insulin, $n = 2{,}575$; 2-h plasma glucose, $n = 2{,}540$; and 2-h serum insulin, $n = 2{,}540$) and up to 1,064 B99 participants for replication analysis (fasting plasma glucose, $n = 1{,}064$; fasting serum insulin, $n = 1{,}062$; 2-h plasma glucose, $n = 845$; and 2-h serum insulin $n = 845$). Individuals taking lipid-lowering medication were removed for the analysis of fasting serum lipids. Dichotomous glycaemia variables were constructed to test for association with T2D, classified according to the World Health Organization criteria[21]. This analysis included 223 individuals with T2D and 1,832 control individuals with normal glucose tolerance for the discovery IHIT cohort, as well as 84 individuals with T2D and 543 control individuals for the B99 replication sample (Extended Data Table 1).

Genotyping. DNA was purified from blood leukocytes. Both the Greenlandic and the Danish individuals were genotyped on the Cardio-Metabochip (Metabochip; Illumina), which is a custom iSelect genotyping array of 196,725 SNPs for genetic studies of metabolic, cardiovascular and anthropometric traits[7]. Genotyping was performed using the HiScan system (Illumina), and genotypes were called using the GenCall module of the GenomeStudio software (Illumina) using default cluster data. In total, 2,812 individuals from the IHIT cohort and 1,376 from the B99 cohort were genotyped. Four additional SNPs identified in exome sequencing were genotyped in both the IHIT and B99 cohorts, by applying the KASP genotyping assay (LGC Genomics).

Data filtering. Before performing any analyses, we used PLINK software tools to identify and remove samples with mis-specified gender and high rates of missing data (>2% of all SNPs with a MAF > 0.01). This led to the removal of 79 samples from the IHIT data set and 45 samples from the B99 data set, leaving us with 2,733 and 1,331 samples, respectively, to analyse.

Before performing association testing, we filtered out all SNPs with MAFs <1% or with >5% missing genotypes. This filtering was performed separately for each test: that is, the MAF and 'missingness' estimates on which the filtering was based for a given test were calculated from the individuals included in that specific test. Additionally, all SNPs with >10% difference in allele frequency between the IHIT and B99 data sets were removed from both data sets.

Association testing. For association testing, we used a standard linear mixed model, implemented in GEMMA[22]: that is, we tested for association between each SNP and each trait by modelling the effects of genotypes and the additional covariates, age and sex, as fixed effects, and admixture and relatedness as random effects. These random effects were assumed to be distributed as a multivariate normal distribution with variance proportional to an $n \times n$ relatedness matrix estimated from the standardized genotypes of all SNPs with a MAF >5% and <1% missingness. We tested for association using a likelihood ratio test of the alternative hypothesis H_1 ($\beta \neq 0$) against the null hypothesis H_0 ($\beta = 0$), where β is the effect size of the genotype. The effect sizes and standard errors of the effect sizes were estimated using a restricted maximum likelihood (REML) approach.

Before performing the analyses, all of the quantitative traits were quantile-transformed to a standard normal distribution. Analyses were performed by applying an additive model (that is, with genotypes coded as 0, 1 and 2 denoting the number of minor alleles) and a recessive model (that is, with homozygotes for the minor allele coded as 1, and the two other genotypes coded as 0). Conditional analyses were performed by including the genotypes of the SNP, which we conditioned on, as an additional covariate.

We analysed the two data sets (IHIT and B99) separately, IHIT with the purpose of discovery and B99 with the purpose of replication. We also performed a meta-analysis using the inverse-variance-based method.

Estimation of OR. To test for association with T2D, we used a linear mixed model to properly control for the substantial amount of admixture and relatedness present in the sample. Even though a linear model does not properly model binary traits, such as T2D, it can still give valid results. To estimate the OR of p.Arg684Ter for T2D under the recessive model, we used a linear mixed model without the covariates age and sex and estimated the intercept, α, and the genotype effect, β. Using these estimates, we calculated the OR as $OR = \{(\alpha + \beta)/[1 - (\alpha + \beta)]\}/[\alpha/(1 - \alpha)]$. Note that ORs obtained from such a model are not always comparable to those obtained from the standard logistic regression model.

Exome sequencing. We selected nine trios for exome sequencing. One of the trios had a minor estimated amount of European ancestry (average of 5%), while the others were inferred to have no European admixture. We enriched for carriers of the rs7330796 minor allele such that seven individuals carried two copies of the minor allele and nine individuals carried one copy, giving a MAF of close to 50% in these individuals. The nine trios were exome-sequenced using SureSelect capture (Agilent) followed by HiSeq2000 sequencing (Illumina). The sequence data were aligned to the human genome assembly hg19 using the software bwa[23] (version 0.7.5a-r405) with default parameters. Duplicates were removed, and the mate-pair information was properly defined using picard tools (http://picard.sourceforge.net) (version 1.95). Subsequently, local realignment surrounding known indel regions was performed, and the quality score for each read was recalibrated with gatk[24] (version 2.4-7-g5e89f01) using the same resource files as in the 1000 Genomes Project[12]. Finally genotypes were called using SAMtools[25] (version 0.1.19-44428cd) and bcftools with standard parameters (-A -c -e -g -v). We observed no discordance between the lead SNP from the sequencing data and the SNP chip data; we also did not find any discordance between the four SNPs found to be in linkage disequilibrium with the lead SNP, which were also genotyped.

The mean individual sequencing depth was 90.1× for all target regions for the 27 samples. This is comparable to that of the exomes for *TBC1D4*, which had a mean individual sequencing depth of 86.8×. The mean phred scale quality scores were 33.2 and 32.7 for *TBC1D4* and all target regions, respectively. These scores correspond to a base error probability of approximately 0.05%.

Trio phasing. Phasing of the exome-sequenced individuals was done by first calling genotypes, as described above. Then, the inferred genotypes were filtered and, finally, phasing was done using the trio information based on the Beagle software[26]. We filtered the sites based on the following criteria: average sequenced depth above 10× and below 500×, SNP quality score above 30, SNPs with a mappability of 1 (100mer), pv4 (strand and mapQ, end of read) with a P value higher than 10^{-8} and less than nine missing genotypes. Individual genotypes were set to missing if the individual sequencing depth was less than 5×, if there were Mendelian errors, if a site was triallelic and if the genotype quality score was less than 20.

***TBC1D4* and *GLUT4* expression studies.** Three individuals in each p.Arg684Ter genotype group (carriers of zero, one or two copies of p.Arg684Ter) were selected from the IHIT cohort for examination of *TBC1D4* and *GLUT4* skeletal muscle expression after an overnight fast. Under local anaesthesia, samples of the vastus lateralis muscle were acquired from each of these participants, snap-frozen and stored at −80 °C.

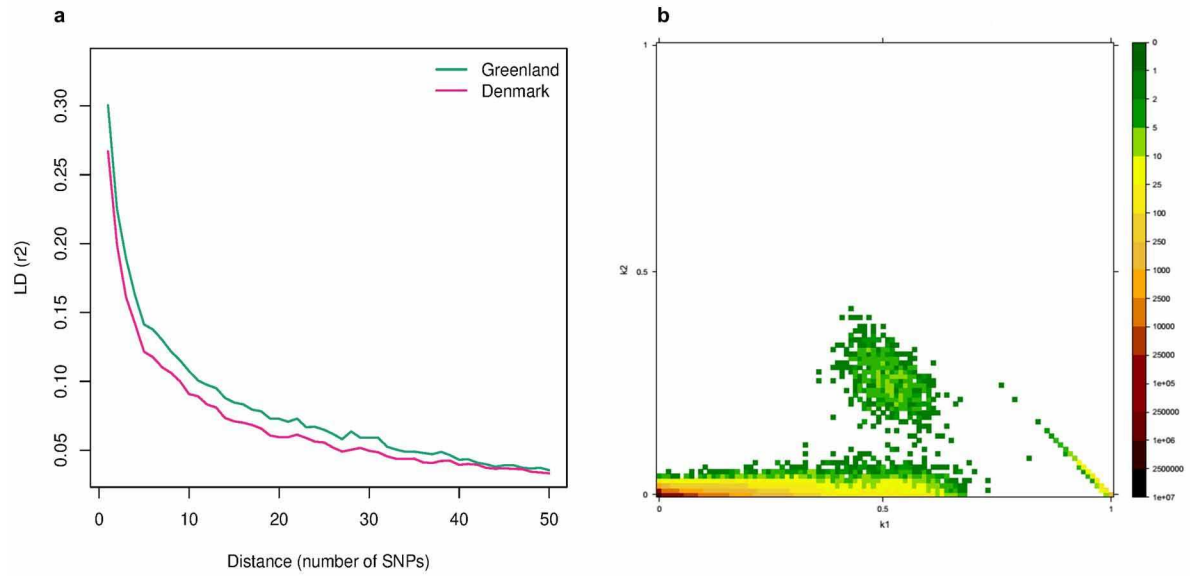
mRNA analyses. Total RNA was extracted from the skeletal muscle samples and real-time PCR was used to quantify mRNA expression levels of the two *TBC1D4* isoforms using primer sets that specifically amplify either the short or the long isoform of *TBC1D4*, as previously described[27]. In addition, the gene expression levels of the *TBC1D4* isoforms were determined from commercially available human tissue RNA (Clontech Laboratories): hypothalamus (catalogue no., 636144), subthalamic nucleus (catalogue no., 636167), thalamus (catalogue no., 636135), hippocampus (catalogue no., 636134), cerebellum (catalogue no., 6361), cerebral cortex (catalogue no., 636164), postcentral gyrus (catalogue no., 636573), pons (catalogue no., 636166), temporal lobe (catalogue no., 63616), parietal lobe (catalogue no. 636571) and tissues as previously described[28]. Finally, the *TBC1D4* isoform expression patterns in human islets were determined from RNA extracted from five donors. The human islets were provided through the Juvenile Diabetes Research Foundation (JDRF) Islet Distribution Program (JDRF award 6-2005-1178) by Islet Cell Resource Centres in Milan (Italy) and Lille (France). Total RNA was extracted from intact human islets using the TRIzol method (Invitrogen). cDNA synthesis was performed using the Taqman Gold RT–PCR Kit (PerkinElmer).

Protein analyses. Muscle biopsies were homogenized (TissueLyser II, QIAGEN) in lysis buffer (50 mM HEPES, pH 7.5, 10% glycerol, 1% IGEPAL CA-630, 150 mM NaCl, 20 mM $Na_4P_2O_7$, 20 mM β-glycerophosphate, 10 mM NaF, 5 mM nicotinamide, 2 mM Na_3VO_4, 1 mM EDTA, 1 mM EGTA, 1 mM sodium butyrate, 4 μM thiamet G and 1× SIGMAFAST protease inhibitor cocktail (Sigma Aldrich, catalogue no. S8820)) and incubated end-over-end at 4 °C for 45 min. In a separate round of homogenizations, membrane-bound proteins were extracted by adding SDS to homogenates (final concentration 2.5%) followed by incubations at 1,200 r.p.m. (Thermomixer, Eppendorf) at 30 °C for 45 min. All homogenates were cleared by centrifugation (16,000*g*, 20 min, 4 °C), and the lysates were recovered and stored at −80 °C. Protein concentrations were determined (BCA Protein Assay, Thermo Scientific, catalogue no. 23225), and western blot analyses were performed using equal amounts of protein to determine TBC1D4 content (5% SDS–PAGE, 40 μg lysate per lane) and GLUT4 content (10% SDS–PAGE, 10 μg lysate per lane). The abundance of full-length TBC1D4 was quantified using an antibody targeting the carboxy terminus of TBC1D4 (Abcam, catalogue no. ab24469); and GLUT4 proteins, using an antibody targeting the amino terminus of GLUT4 (Thermo Scientific, catalogue no. PA1-1065). Band intensities were evaluated using Image Lab software (Bio-Rad).

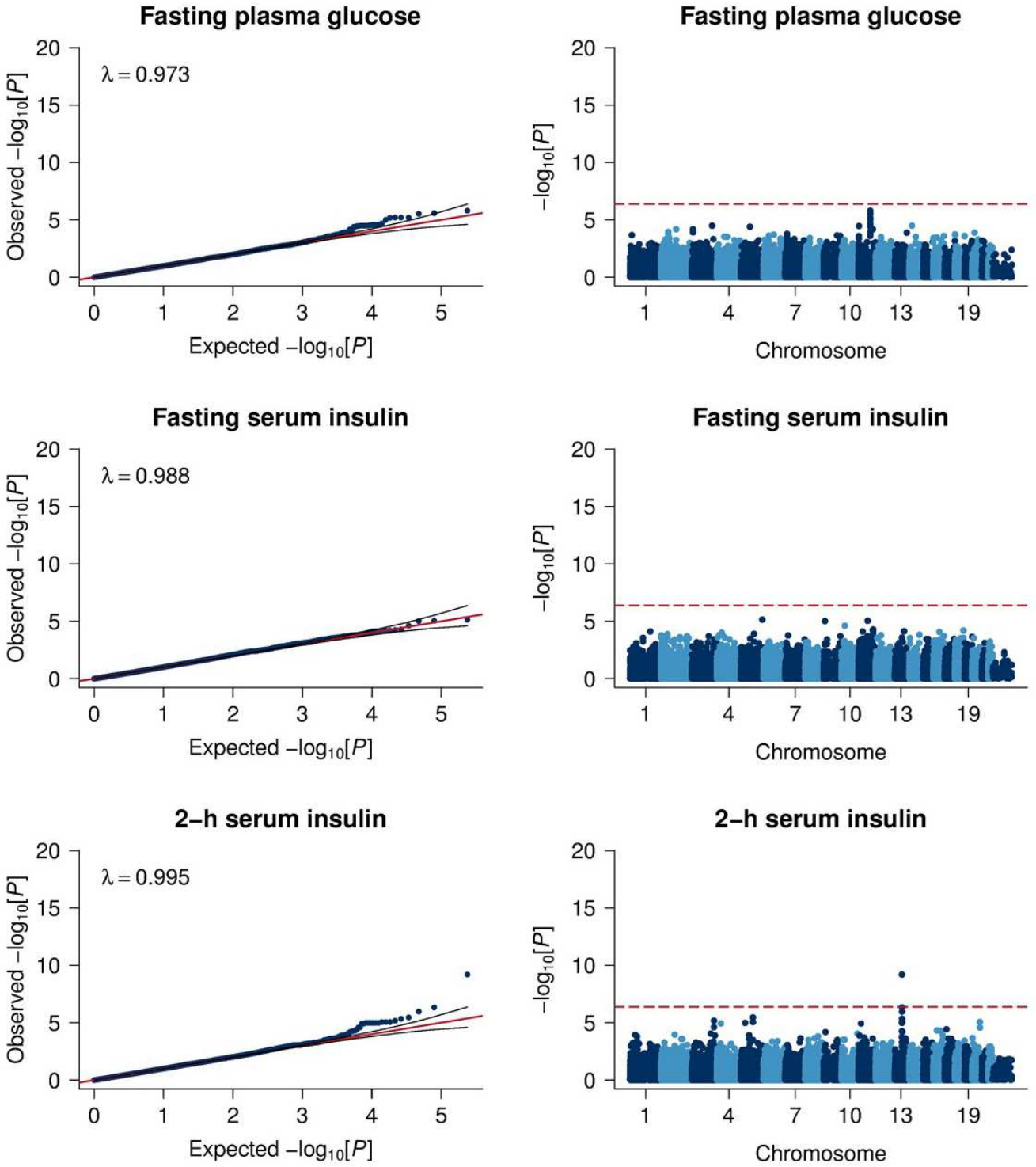
Admixture analysis and allele frequency estimation. We estimated admixture proportions for the IHIT cohort from the Metabochip data using the software ADMIXTURE[29] with $K = 2$: that is, assuming that all individuals have ancestry from two populations, the European population and the ancestral Inuit population. We also included 50 Danish individuals in the analyses as a population reference. We ran the ADMIXTURE program 50 times with different starting points to ensure that convergence was reached. This program also estimates the frequencies of the ancestral alleles. Thus, it is possible to obtain the allele frequencies in the Inuit population before the admixture with Europeans. For the *TBC1D4* p.Arg684Ter variant, the MAF in the IHIT cohort was 17%, but the MAF was estimated as 0% in the ancestral European population and 22.6% in the ancestral Inuit population. The European estimate was confirmed by the absence of this variant in 2,000 Danish individuals, as determined by exome sequencing[13]. In the IHIT cohort, 681 individuals showed no signs of European admixture, and the MAF in those individuals was 25.0% (95% CI, 22.9–27.6).

Relatedness estimation. Relatedness was estimated from the Metabochip data for all pairs of individuals using the program RelateAdmix[30], which takes admixture into account. The admixture proportions and population-specific allele frequencies that RelateAdmix needs as input were estimated using the ADMIXTURE program assuming two ancestral populations (see Admixture analysis above).

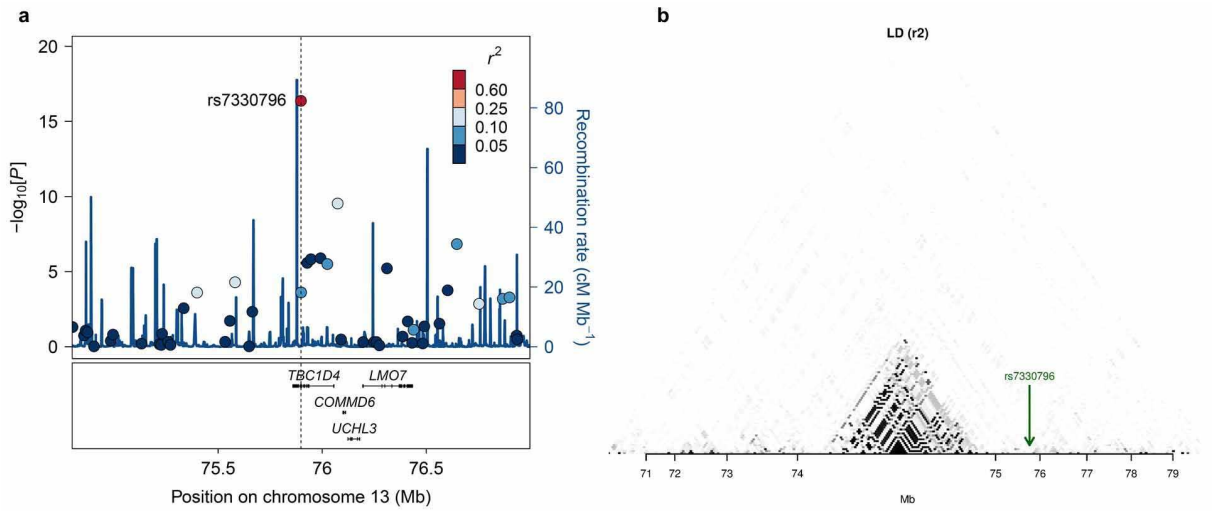
19. Jørgensen, T. *et al.* A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. *Eur. J. Cardiovasc. Prev. Rehabil.* **10,** 377–386 (2003).
20. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28,** 412–419 (1985).
21. World Health Organization Study Group. *Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus* Technical Report Series WHO/NCD/NCS/99.2 (World Health Organization, 1999).
22. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genet.* **44,** 821–824 (2012).
23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
24. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
25. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).
26. Browning, B. L. & Browning, S. R. A unified approach to genotype imputation and haplotype-phase Inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84,** 210–223 (2009).
27. Treebak, J. T. *et al.* Potential role of TBC1D4 in enhanced post-exercise insulin action in human skeletal muscle. *Diabetologia* **52,** 891–900 (2009).
28. Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56,** 298–310 (2013).
29. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
30. Moltke, I. & Albrechtsen, A. RelateAdmix: a software tool for estimating relatedness between admixed individuals. *Bioinformatics* **30,** 1027–1028 (2014).



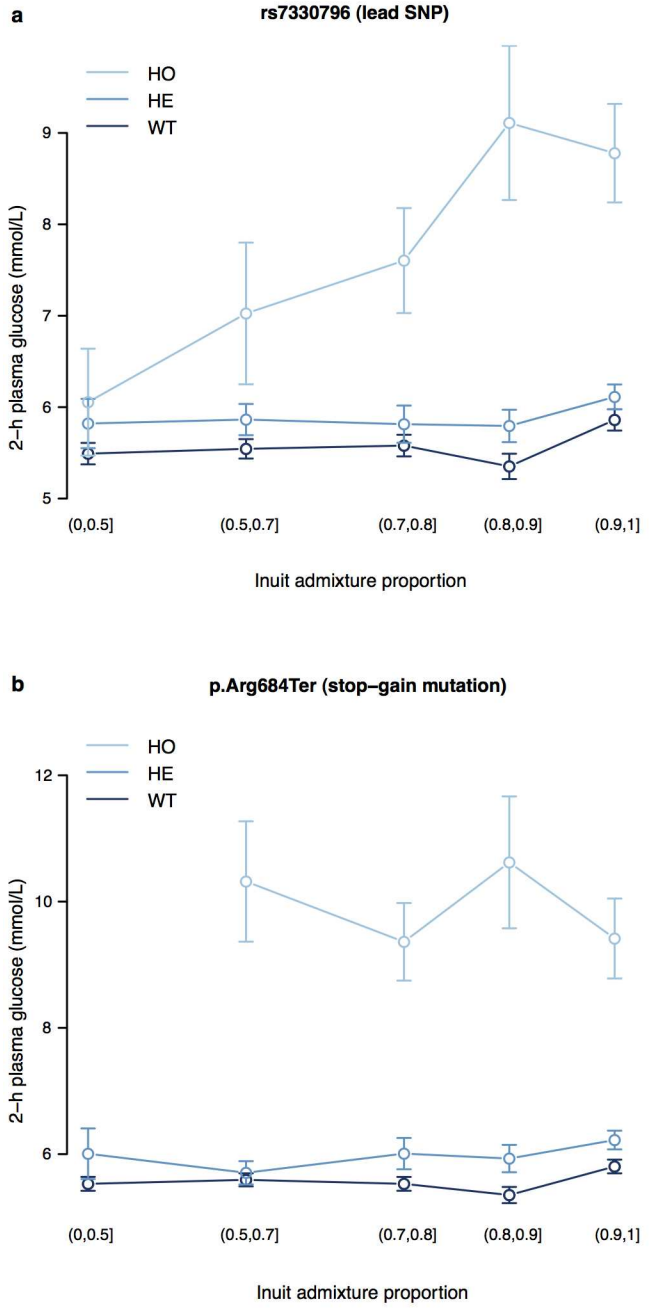
Extended Data Figure 1 | Linkage disequilibrium (LD) and relatedness in the IHIT cohort. **a**, The mean LD at chromosome 13 as a function of distance estimated for a Danish population sample (Denmark) and for the IHIT cohort (Greenland). LD is measured in r^2 values, and distance is measured as number of SNPs. **b**, A heat map of pairwise relatedness estimates based on the Metabochip SNP data from the individuals from the IHIT cohort and their estimated admixture proportions. *k1* is an estimate of the fraction of sites where two individuals share one allele identity by descent (IBD), and *k2* is an estimate of the fraction of sites where they share two alleles IBD. The estimates were achieved using the software tool RelateAdmix, which provides maximum likelihood estimates that take admixture into account by allowing alleles to be IBD only if the alleles are from the same population. The vast majority of the approximately 4 million pairwise estimates have a relatedness coefficient *r* of less than 1%, where $r = 0.5k1 + k2$.



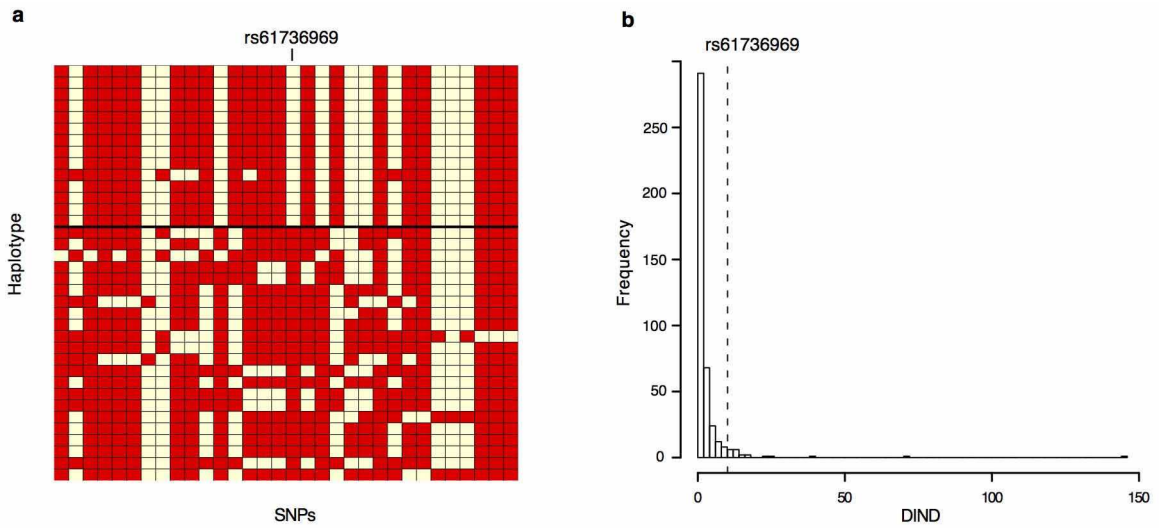
Extended Data Figure 2 | Association between fasting plasma glucose level, fasting serum insulin level and 2-h serum insulin level and the SNPs on the Metabochip for the IHIT cohort. Analyses were done using an additive model and quantile-transformed phenotypes. Left, QQ plots with 95% confidence interval. The λ value is the genomic control inflation factor. Right, Manhattan plots showing the $-\log_{10}[P]$ values. The red dashed horizontal line indicates the 0.05 significance threshold after Bonferroni correction for multiple testing.



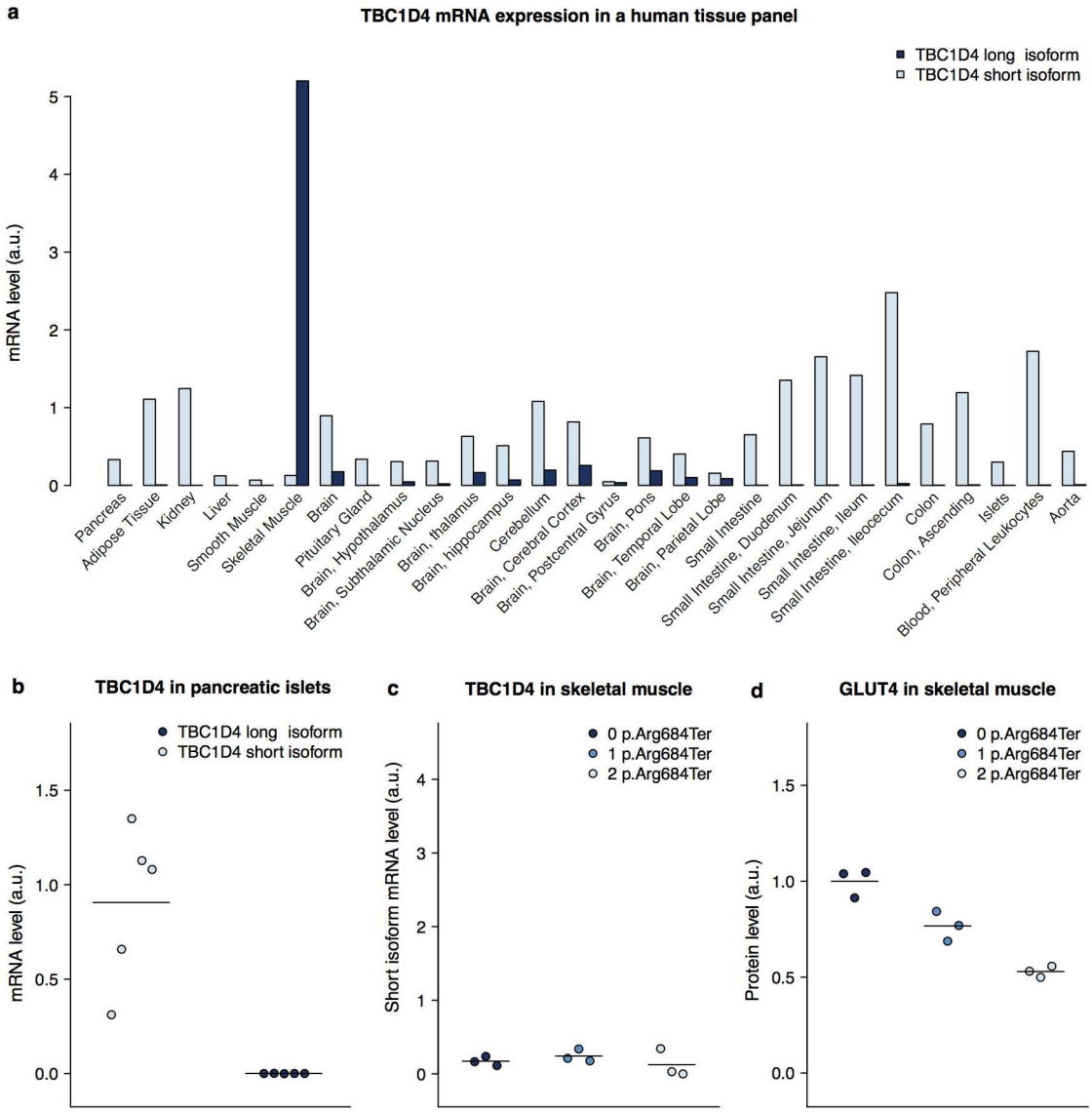
Extended Data Figure 3 | The region surrounding rs7330796. a, Association results for all tested Metabochip SNPs in a 2-Mb region surrounding rs7330796 (shown as a dashed vertical line). Each SNP is represented by a coloured circle. The position of the circle along the *x* axis shows the genomic position of the SNP. The position of the circle along the left *y* axis shows the $-\log_{10}[P]$ value of the SNP when testing for association with 2-h plasma glucose levels using an additive model. The colour of the circle indicates the r^2 value between the SNP and rs7330796. The circle representing rs7330796 has a label to the left of it so that it can be identified. The solid blue curve (right *y* axis) illustrates the recombination rates from the Chinese HapMap (CHB) panel. Bottom, the protein-coding genes in the region are shown. **b**, Pairwise LD in the Greenlandic population for all SNPs with a MAF >0.05 in a 10-Mb region surrounding rs7330796. LD is measured in r^2 and was estimated from the IHIT study sample.



Extended Data Figure 4 | Mean 2-h plasma glucose levels in the IHIT cohort stratified by Inuit admixture proportion. **a**, The three possible genotypes of rs7330796 (zero copies of the minor allele (WT), one copy (HE) and two copies (HO)). **b**, The three possible genotypes of p.Arg684Ter (zero copies of the p.Arg684Ter variant (WT), one copy (HE) and two copies (HO)). Error bars are s.e.m. and were estimated for each admixture proportion using a standard linear regression model.



Extended Data Figure 5 | Selection analysis. **a**, Haplotype structure of a 500-kilobase pair (kbp) region around the coding SNP rs61736969. For each SNP, the ancestral state is coloured red, while the derived allele is coloured yellow. Haplotypes are ordered according to the coding SNP rs61736969 (p.Arg684Ter), whose position is highlighted. Haplotypes above the black line carry the derived allele for rs61736969. **b**, The empirical distribution of the derived intra-allelic nucleotide diversity (DIND) values, computed on a 500-kbp region, for all SNPs with the same derived allele frequency as the coding SNP rs61736969. DIND is a measure of the ratio of the nucleotide diversity of haplotypes carrying the derived allele to the nucleotide diversity of haplotypes carrying the ancestral allele. The DIND value for the coding SNP rs61736969 is highlighted by a dashed line and lies in the top 5% of the distribution.



Extended Data Figure 6 | Expression of *TBC1D4* and *GLUT4* in humans. **a**, The mRNA expression levels of the long and short *TBC1D4* isoforms in a range of different human tissues (measured in arbitrary units, a.u.). **b**, The mRNA expression levels of both *TBC1D4* isoforms in human pancreatic islet cell preparations from five individuals (in a.u.). The five individuals are not Greenlandic and do not carry the p.Arg684Ter variant. The mean value for each of the isoforms is shown as a horizontal line. **c**, The mRNA expression levels of the short *TBC1D4* isoform in skeletal muscle from nine Greenlanders: three with zero copies of the p.Arg684Ter variant (WT), three with one copy of the p.Arg684Ter variant (HE) and three with two copies of the p.Arg684Ter variant (HO). The arbitrary units (a.u.) are on the same scale as in Fig. 3d. The mean value for each genotype group is shown as a horizontal line. **d**, The protein abundance of GLUT4 in the same nine Greenlanders as in **c** (measured in a.u.). The mean value for each genotype group is shown as a horizontal line.

Extended Data Table 1 | Clinical characteristics of genotyped individuals from the IHIT and B99 Greenland population studies

	IHIT	B99
n	2,733	1,331
n (men) / *n* (women)	1,227 / 1,506	575 / 756
Age (years)	44 (33-54)	42 (33-54)
BMI (kg m^{-2})	25.4 (22.4-29.4)	25.5 (22.7-29)
Fasting plasma glucose (mmol l^{-1})	5.6 (5.2-6)	5.6 (5.3-6.1)
2-h plasma glucose (mmol l^{-1})	5.4 (4.3-6.7)	5.4 (4.3-6.8)
Fasting serum insulin (pmol l^{-1})	37 (25-56)	38 (26-57)
2-h serum insulin (pmol l^{-1})	111 (48-213)	100 (43-201)
Fasting serum C-peptide (pmol l^{-1})	535 (390-738)	407 (301-565)
2-h serum C-peptide (pmol l^{-1})	1,800 (1,200-2,580)	1,320 (868-2,030)
HbA_{1C} (%)	5.6 (5.4-5.9)	6 (5.7-6.4)
HOMA-IR (mmol l^{-1} × pmol l^{-1})	1.33 (0.869-2.04)	1.4 (0.91-2.14)
$ISI_{0,120}$	2.63 (1.92-3.88)	2.72 (1.91-4.15)
HOMA-B (%)	53.2 (35.3-78)	53.5 (36.3-77.2)
Fasting serum HDL-cholesterol (mmol l^{-1})	1.6 (1.3-1.97)	1.5 (1.3-1.9)
Fasting serum total cholesterol (mmol l^{-1})	5.7 (5.0-6.6)	5.9 (5.2-6.7)
Fasting serum triglyceride (mmol l^{-1})	1.02 (0.76-1.42)	0.985 (0.74-1.36)
Type 2 diabetes: *n* (cases) / *n* (ctrls)	223 / 1,832	84 / 543

HOMA-B = [20(fasting insulin level/6.945)]/(fasting glucose level − 3.5), where insulin levels are expressed in pmol l^{-1} and glucose levels are expressed in mmol l^{-1}. HOMA-IR = [(fasting glucose level × fasting insulin level)/6.945]/22.5, where insulin levels are expressed in pmol l^{-1} and glucose levels are expressed in mmol l^{-1}. Insulin sensitivity index ($ISI_{0,120}$) = {[75,000 + (fasting glucose level × 18 − 2-h glucose level × 18) × 0.19 × weight]/120}/[0.5(fasting glucose level × 18 + 2-h glucose level × 18)]/log[(fasting insulin level/6.945 + 2-h insulin/6.945)/2], where insulin levels are expressed in pmol l^{-1}, glucose levels are expressed in mmol l^{-1} and weight is expressed in kg.

Extended Data Table 2 | LD and allele frequency estimates from the 1000 Genomes Project data for SNPs in *TBC1D4*

Position	ID	REF	ALT	MAF	Function	r^2	CEU	CHB	CHS	FIN	JPT	YRI
75813348	rs475484	G	A	0.315	Intergenic	0.23	0.506	0.716	0.69	0.554	0.59	0.335
75880384	rs33992075	AAA	.,A,AA	0.278	Intron	0.29						
75880386	rs33992075	AA	AACA	0.167	Intron	0.35						
75880667	rs2297209	G	A	0.426	Intron	0.55	0.541	0.655	0.57	0.522	0.556	0.324
75884216	rs1062087	C	T	0.426	Missense	1	0.894	0.825	0.79	0.812	0.758	0.341
75884290	rs2297203	G	A	0.426	Intron	0.55	0.394	0.165	0.195	0.317	0.202	0.04
75898521	rs61736969	G	C	0	missense		0	0	0	0	0	0.034
75898521	rs61736969	G	A	0.389	stop-gained	0.86	0	0	0	0	0.0048	0
75901749	rs2297206	G	T	0.426	Intron	1	0.9	0.82	0.775	0.839	0.77	0.347
75910909	rs693134	C	A	0.426	Intron	0.55	0.441	0.665	0.67	0.478	0.697	0.244
75910950	rs9573525	T	G	0.3	Intron	0.34	0.429	0.665	0.67	0.446	0.697	0.074
75915179	rs140300462	AG	A	0.333	Intron	0.36	0.024	0.113	0.08	0.091	0.096	0.102
75915250	rs67719506	TA	TAA	0.389	Intron	0.47	0.465	0.686	0.695	0.462	0.702	0.068
75915261	rs9565152	A	C	0.352	near-splice	0.40	0.024	0.113	0.08	0.091	0.096	0.102
75915859	rs9888477	C	A	0.413	Intron	0.49	0.465	0.67	0.675	0.462	0.697	0.068
76055602	rs77685055	G	A	0.352	Missense	0.40	0.071	0.299	0.235	0.081	0.191	0

The MAF column contains the MAF in the 27 (18 independent) sequenced individuals, and the six last columns contain the MAFs from the various populations from the 1000 Genomes Project. The r^2 column contains the LD between the SNPs and rs7330796. To estimate LD, all variations have to be diallelic. This was achieved by first defining the most common allele as the base allele. The genotypes were then defined as the number of non-base alleles. The populations shown are the following: Utah residents with Northern and Western European ancestry (CEU); Han Chinese in Beijing, China (CHB); Southern Han Chinese, China (CHS); Finnish in Finland (FIN); Japanese in Tokyo, Japan (JPT); and Yoruban in Ibadan, Nigeria (YRI).

Extended Data Table 3 | Association of *TBC1D4* p.Arg684Ter with metabolic traits in the B99 cohort and in both the IHIT and B99 combined by meta-analysis

	B99							B99 + IHIT (meta-analysis)	
		Additive model			Recessive model				
Trait	*n*	$\beta_{s.d.}$ (95% CI)	β	*P*	$\beta_{s.d.}$ (95% CI)	β	*P*	P_{add}	P_{rec}
Fasting plasma glucose (mmol l^{-1})	1,009	-0.061 (-0.17 to 0.049)	-0.0054	0.28	0.00098 (-0.34 to 0.34)	0.13	0.99	1.0×10^{-4}	1.5×10^{-5}
2-h plasma glucose (mmol l^{-1})	809	0.3 (0.18 to 0.42)	0.98	1.1×10^{-6}	1.7 (1.2 to 2.1)	5.5	**2.4×10^{-13}**	**8.0×10^{-31}**	**1.6×10^{-47}**
Fasting serum insulin (pmol l^{-1})	1,007	-0.025 (-0.14 to 0.091)	-0.93	0.67	0.14 (-0.22 to 0.5)	3.4	0.45	0.00046	0.015
2-h serum insulin (pmol l^{-1})	809	0.24 (0.12 to 0.36)	32	9.6×10^{-5}	1.3 (0.86 to 1.7)	200	**6.7×10^{-9}**	**2.3×10^{-20}**	**9.7×10^{-28}**
Fasting serum C-peptide (pmol l^{-1})	1007	-0.09 (-0.21 to 0.027)	-16	0.13	-0.13 (-0.49 to 0.22)	-18	0.47	3.0×10^{-4}	0.0014
2-h serum C-peptide (pmol l^{-1})	809	0.17 (0.055 to 0.29)	150	0.0041	0.93 (0.5 to 1.4)	890	2.9×10^{-5}	**1.9×10^{-21}**	**6.9×10^{-24}**
HbA_{1C} (%)	1,255	0.019 (-0.07 to 0.11)	0.016	0.67	0.044 (-0.22 to 0.31)	0.043	0.74	0.52	0.028
HOMA-IR (mmol l^{-1} × pmol l^{-1})	1,002	-0.031 (-0.15 to 0.086)	-0.048	0.61	0.14 (-0.22 to 0.5)	0.19	0.46	0.00025	0.0072
$ISI_{0,120}$	800	-0.25 (-0.37 to -0.13)	-0.42	4.9×10^{-5}	-1.5 (-2 to -1.1)	-2.6	**1.0×10^{-11}**	**2.5×10^{-24}**	**1.4×10^{-37}**
HOMA-B (%)	994	0.006 (-0.11 to 0.12)	2.2	0.92	0.099 (-0.25 to 0.44)	3.8	0.57	0.037	0.38
Type 2 diabetes (cases/controls)	79/516	0.056 (0.0056 to 0.11)	0.056	0.029	0.67 (0.48 to 0.86)	0.67	**4.6×10^{-12}**	**2.3×10^{-12}**	**4.2×10^{-34}**
Fasting serum HDL-cholesterol (mmol l^{-1})	1,256	0.072 (-0.029 to 0.17)	0.029	0.17	0.021 (-0.28 to 0.32)	0.0057	0.89	0.12	0.33
Fasting serum total cholesterol (mmol l^{-1})	1,256	0.12 (0.024 to 0.22)	0.15	0.015	0.27 (-0.022 to 0.56)	0.32	0.069	0.28	0.0014
Fasting serum triglyceride (mmol l^{-1})	1,256	0.053 (-0.054 to 0.16)	0.035	0.33	0.2 (-0.12 to 0.51)	0.083	0.23	0.71	0.41
BMI (kg m^{-2})	1,251	0.042 (-0.063 to 0.15)	0.17	0.43	0.019 (-0.3 to 0.34)	0.18	0.9	0.71	0.64

Results are shown for an additive and a recessive genetic model. For each trait, *n* is the number of individuals with genotype data for the specific variant and phenotype data for the specific trait. $\beta_{s.d.}$ is the effect size estimated using quantile-transformed values of the trait (except for the binary trait T2D), and β is the effect size estimated using untransformed values. The *P* values were obtained from the quantile-transformed-value-based analyses. The meta-analysis was performed using the inverse-variance-based method. All *P* values below 1×10^{-6} are marked in bold.