Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes

Niels Grarup1, Ida Moltke2, Mette K. Andersen1, Maria Dalby2, Kristoffer Vitting-Seerup2,3, Timo Kern1, Yuvaraj Mahendran1, Emil Jørsboe1,2, Christina V. L. Larsen4,5, Inger K. Dahl-Petersen4, Arthur Gilly6, Daniel Suveges6, George Dedoussis7, Eleftheria Zeggini6, Oluf Pedersen1, Robin Andersson2, Peter Bjerregaard4,5, Marit E. Jørgensen4,5,8*, Anders Albrechtsen2, Mette K. Andersen2, Christina V. L. Larsen4,5, Inger K. Dahl-Petersen4,5

We have identified a variant in ADCY3 (encoding adenylate cyclase 3) associated with markedly increased risk of obesity and type 2 diabetes in the Greenlandic population. The variant disrupts a splice acceptor site, and carriers have decreased ADCY3 RNA expression. Additionally, we observe an enrichment of rare ADCY3 loss-of-function variants among individuals with type 2 diabetes in trans-ancestry cohorts. These findings provide new information on disease etiology relevant for future treatment strategies.

Identification of homozygous loss-of-function mutations in humans may readily provide information about the biological impact of specific genes and point to novel drug targets. We previously identified a loss-of-function variant in TBC1D4 segregating at high frequency in the Greenlandic population and displaying a large impact on risk of type 2 diabetes1, confirming the advantage of studying the Greenlandic population due to its extreme demographic history2. Motivated by this, we screened for new loss-of-function variants in exome sequencing data from the 27 individuals in nine trios which were used to identify the causal TBC1D4 loss-of-function variant1. We identified 46 such variants (Supplementary Table 1) and intersected the location of these variants with loci previously known to associate with obesity or body mass index (BMI) (Supplementary Fig. 1). One of the variants, which was present in one copy in one of the parents in one trio, was situated in a locus where a common noncoding variant has been shown to be associated with BMI in adults and children in genome-wide association studies (GWAS)3,4. The variant (hg19: chromosome 2: 25050478(C>T), c.2433–1G>A) is predicted to abolish a splice acceptor site in exon 14 (NM_004036.4) of ADCY3 (Fig. 1a). For this reason, we investigated the specific variant further by genotyping this site in two Greenlandic cohorts. This analysis showed that the variant has an overall minor allele frequency (MAF) of 2.3% in the Greenlandic study population (Supplementary Table 2). Notably, the seven homozygous carriers had BMI that was 7.3 kg/m² higher (P = 0.00994) than the BMI of the remaining study population (Table 1). We also observed that three of the seven homozygous carriers had type 2 diabetes (P = 7.8 × 10⁻⁵; Table 1), while one had impaired fasting glucose and one had impaired glucose tolerance. Notably, the association with type 2 diabetes remained significant after adjustment for BMI (P = 6.5 × 10⁻⁴), suggesting that it is not simply mediated by increased BMI. The effects on BMI and type 2 diabetes were also observed, although with smaller sizes, when data were analyzed according to an additive genetic model (Table 1). However, when we compared the recessive and additive models with the full genotype model, we rejected the additive model (BMI, P = 0.002; type 2 diabetes, P = 0.004) but not the recessive model (BMI, P = 0.17; type 2 diabetes, P = 0.095). This suggests that the recessive model is appropriate for explaining the effect of the ADCY3 c.2433–1G>A variant.

To further characterize homozygous carriers of ADCY3 c.2433–1G>A in the Greenlandic cohorts, we analyzed a number of additional traits related to BMI and type 2 diabetes. The homozygous carriers had a body fat percentage that was 8.1 percentage points higher (P = 0.0024) and a waist circumference that was 17 cm larger (P = 0.0017). In addition, the homozygous carriers had nominally higher levels of fasting and 2-h plasma glucose in an oral glucose tolerance test (P = 0.022 and P = 0.035, respectively; Table 1). Finally, we also observed nominally significant effects of the variant on dyslipidemia and insulin resistance (Supplementary Table 2).

The c.2433–1G>A ADCY3 variant was not observed in sequencing data of up to 138,000 individuals from non-Greenlandic populations collected by the Genome Aggregation Database Consortium (gnomAD). To generalize our findings to other populations, we therefore investigated the effect of loss-of-function variants in ADCY3 more generally by analyzing 18,176 samples with exome sequencing data of up to 138,000 individuals from non-Greenlandic populations collected by the Genome Aggregation Database Consortium (gnomAD). To generalize our findings to other populations, we therefore investigated the effect of loss-of-function variants in ADCY3 more generally by analyzing 18,176 samples with exome sequencing data generated by the GoT2D, T2DGenes, SIGMA and LuCAMP consortia5,6, which are available at the Accelerating Medicines Partnership Type 2 Diabetes Knowledge Portal (AMP-T2D). No homozygous carriers of ADCY3 loss-of-function variants were observed in this dataset, but the analysis included seven predicted ADCY3 loss-of-function variants (Fig. 1a) observed in the heterozygous form in eight individuals and we observed an enrichment of carriers among type 2 diabetes cases as compared with

1Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 2Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark. 3Biotech Research and Innovation Centre, Hinxton, UK. 4Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece. 5Greenland Centre for Health Research, University of Greenland, Nuuk, Greenland. 6Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, UK. 7Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece. 8Steno Diabetes Center Copenhagen, Gentofte, Denmark. 9Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark. Niels Grarup and Ida Moltke contributed equally to this work.

*e-mail: marit.eika.joergensen@regionh.dk; albrecht@binf.ku.dk; torben.hansen@sund.ku.dk
Brief Communication

NATuRE GeNeTics

Correspondence to the first and third quartiles of data, respectively, while the group in and c.2433–1G>A variant genotype. The number of individuals in each quantified isoforms—the canonical isoform, the novel exon-skipping transcript isoform fractions for the three ADCY3, c.2433–1G>A transcript ADCY3-202 (NM_001320613). Transcripts individuals from two isolated Greek populations that were part of 17 Greenlandic individuals (7 GG carriers, 6 GA carriers and 4 AA carriers). The RNA sequencing data confirmed that ADCY3 was expressed in homozygous carriers of the wild-type allele (GG carriers) and that exon 14 (NM_0040306.4) of ADCY3 was expressed and spliced in its canonical form (Supplementary Fig. 2a). Inclusion of exon 14 in the mature mRNA was further confirmed by RNA sequence data from the adipose tissue of a healthy European-ancestry female donor. Notably, we found that the overall RNA expression level of ADCY3 was severely decreased in homozygous carriers of the variant (AA), while heterozygous carriers (GA) showed an intermediate expression level (Fig. 1b). The RNA sequencing data further confirmed that the predicted disruption of the exon 14 splice acceptor site by the c.2433–1G>A variant has an impact on molecular phenotype. Specifically, the data predict that two novel ADCY3 isoforms are expressed in variant carriers: one transcript where exon 14 is skipped and an alternative splice acceptor site at exon 15 is used and one transcript in which the intron between exon 13 and exon 14 is retained (Supplementary Fig. 2b). We quantified these alternative splicing events by comparing expression of the three predicted ADCY3 isoforms using isofrom fraction (IF) (Fig. 1c) and the percentage spliced in (PSI) at the relevant splice sites (Supplementary Fig. 2b). Both analyses demonstrated that homozygous AA carriers had a severely affected splicing pattern and mainly used intron retention (median IF = 0.38, median PSI = 75%) or exon skipping (IF = 0.53, PSI = 24%), while wild-type GG carriers had the canonical splicing pattern (IF = 0.88, PSI = 87%). In all analyses, heterozygous carriers showed an intermediate level of alternative splicing (Fig. 1c and Supplementary Fig. 2b). Notably, we predict the isoform with intron retention to be sensitive to nonsense-mediated decay owing to the introduction of a premature stop codon (Fig. 1a). This predicted degradation naturally would lead to further reduction of ADCY3 protein levels.

ADCY3 encodes an adenylate cyclase with a wide tissue distribution showing high levels in subcutaneous and visceral adipose tissue, intermediate levels in brain, and rather low levels in pancreas and skeletal muscle in Genotype-Tissue Expression (GTEX) project data. ADCY3 catalyzes the synthesis of cyclic AMP (cAMP) from ATP. cAMP is an essential second messenger in intracellular signaling downstream of key metabolic mediators such as glucagon-like peptide 1, ghrelin and α-melanocyte-stimulating hormone, and cAMP signaling has been linked to control of adipose tissue development and function, as well as insulin secretion in beta cells. In addition, mouse models have indicated that ADCY3 has an important role in the regulation of adiposity and glucose homeostasis. Hence, in mice, Adcy3 haploinsufficiency causes impaired insulin sensitivity and dyslipidaemia, Adcy3 gain of function protects against diet-induced obesity and Adcy3-knockout mice show increased fat mass, hyperphagia, depression-like phenotypes and leptin resistance. Possibly, leptin resistance occurs through disrupted CAMP signaling in primary cilia in hypothalamus, affecting the downstream signaling and morphology of neurons. Interestingly, previously described syndromic forms of obesity, including Bardet–Biedl and Alström syndromes, have been found to be caused by altered function of primary cilia and are, in addition to obesity, characterized by diabetes. Furthermore, common variation in ADCY5, a gene in the same family as ADCY3, is known to be associated with fasting plasma glucose levels and risk of type 2 diabetes.

In humans, common variants in the ADCY3 locus have been associated with higher BMI and total as well as truncal fat mass. Thus, the phenotype observed in Greenlandian individuals homozygous for the loss-of-function variant, characterized by truncal adiposity, insulin resistance, dyslipidaemia and type 2 diabetes, is in accordance with and elaborates on the phenotype observed for GWAS-identified variants. Our findings for carriers of loss-of-function variants implicate ADCY3 as the causal transcript in the reported GWAS-identified locus, and coherent evidence from genetic and biological studies suggest that pharmacological modulation

Fig. 1 AADCY3 isoforms, observed loss-of-function variants and functional consequences based on RNA sequencing of leucocytes from 17 Greenlandic individuals. a. Schematic illustration of ADCY3 displaying the three relevant transcript isoforms with their predicted functional consequences annotated to the left (coding or nonsense-mediated decay (NMD) sensitive). The exons that correspond to the guanylate cyclase protein domain are shown as gray-filled boxes, while the rest of the exons are shown as black-filled boxes. The red box encompasses the exons affected by the Greenlandic ADCY3 c.2433–1G>A variant. The locations of the identified loss-of-function variants in ADCY3 in the Greenlandic and trans-ancestry cohorts are shown with red and black arrows, respectively. Variants were annotated to canonical transcript ADCY3-201 (NM_0040306) except for c.1072–1G>A, which is annotated to alternative transcript ADCY3-202 (NM_001320613). b. ADCY3 transcripts per million (TPM) normalized gene expression, stratified according to ADCY3 c.2433–1G>A variant genotype (WT, wild type; HE, heterozygous; HO, homozygous). c. ADCY3 transcript isoform fractions for the three quantified isoforms—the canonical isoform, the novel exon-skipping isoform and the novel intron-retention isoform—stratified according to ADCY3 c.2433–1G>A variant genotype. The number of individuals in each group in b and c is as follows: WT, 7 GG carriers; HE, 6 GA carriers; HO, 4 AA carriers. The lower and upper hinges of the boxes in b and c correspond to the first and third quartiles of data, respectively, while the middle line is the median. The whiskers extend to the largest and smallest data points no further than 1.5 times the interquartile range.
of this target may possibly be a valid future therapy for obesity and type 2 diabetes.

In conclusion, we identified an ADCY3 loss-of-function variant in a Greenlandic Inuit population that increases adiposity and risk of type 2 diabetes in homozygous carriers and, to a lesser extent, in heterozygous carriers. Concomitantly, we detected decreased ADCY3 RNA expression levels in homozygous carriers and, again to a lesser extent, in heterozygous carriers. Furthermore, we show that the variant disrupts a splice acceptor site, triggering exon skipping or intron retention, where the latter is predicted to confer nonsense-mediated decay. The association with type 2 diabetes was substantiated in heterozygous carriers of rare ADCY3 variants across ancestry groups, underlining the possible role of ADCY3 as a future target for the prevention and treatment of obesity and type 2 diabetes.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0022-7.

Acknowledgements

We gratefully acknowledge the participants in the Greenlandic health surveys. We thank J. Hannick (Broad Institute) for help with obtaining the AMP-T2D exome sequencing data. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (http://www.metabol.ku.dk). This project was also funded by the Danish Council for Independent Research (Sapere Aude DFF-Ung Eliteforsker grant 11-120969 from FSS to I.M.; DFF-4094-0044 from ENU to I.M.; Sapere Aude DFF-Forskningsleder 6108-00038B from ENU to R.A.; DFF-4181-0038 to A.A.), the Steno Diabetes Center Copenhagen (http://www.steno.dk), Simon Fouger Hartmanns Familiefond to N.G., the Lundbeck Foundation (R215-2015-4174) to P.A., the Novo Nordisk Foundation (NNF15OC0017918 to N.G.; NNF16OC0006986 to N.G.; NNFCC00184663 to T.H.) and the European Research Council (ERC) (grant agreement 638173) to R.A. The Greenlandic health surveys (HIT and B99) were supported by Karen Elise Jensen’s Foundation, the Department of Health in Greenland, NunaFonden, the Medical Research Council of Denmark, the Medical Research Council of Greenland and the Commission for Scientific Research in Greenland.

Author contributions

T.H. and A.A. conceived and headed the project. I.M. and A.A. designed the statistical setup for the association testing, while T.H., N.G., M.K.A. and O.P. designed the experimental setup for the DNA extraction, genotyping and sequencing. M.E.I. and P.B. were principal investigators of the population studies in Greenland and, together with C.V.L.L. and I.K.D.-P., provided the Greenlandic samples, collected and defined the phenotypes, and provided context for these samples. I.M., N.G., E.J. and A.A. performed the association analyses. T.K. and Y.M. designed the experimental setup for RNA extraction and sequencing. A.G., D.S., G.D. and E.Z. performed the loss-of-function analysis in the Greek cohorts. K.V.-S., M.D. and R.A. performed the RNA sequence analysis. N.G., I.M., M.K.A., A.A. and T.H. wrote the majority of the manuscript with input from all authors. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-017-0022-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.E.I. or A.A. or T.H.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Study samples. The Greenlandic samples are from two different cohorts: B99 (N = 1,401) and Inuit Health in Transition (IHT) (N = 3,115), which were collected in Greenland as part of a general population health surveys conducted in 1999–2000 and 2001–2002, respectively. The two cohorts overlapped between the two cohorts and were assigned to the B99 dataset. Clinical characteristics of the participants are shown in Supplementary Table 6.

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

Phenotypic data. Height and weight were measured while participants were wearing light indoor clothes. Waist circumference was measured midway between the rib cage and the iliac crest, and hip circumference was measured at its maximum. Weight was measured on a standard electronic clinical scale. Fat percentage was calculated from biometric measurements (Tanita, TBF-300MA) in IHT participants. Intra-abdominal and subcutaneous adipose tissue was further assessed by ultrasonography, which is considered to be a valid and reproducible method as compared with computerized axial tomography (CT) scans and magnetic resonance imaging (MRI)10, using a 3.5-MHz transducer with the participant in supine position and at the end of a normal expiration. Tests for intra- and inter-observer variation were performed in IHT, and results were in the range of 1.9–5.6%10. In IHT, all participants underwent an oral glucose tolerance test (OGTT). In B99, participants above 24 years of age had fasting blood samples taken and participants above 35 years of age underwent an OGTT. At the baseline health examination, venous blood samples were drawn after an overnight fast of at least 8h. After this, participants received a standard 75-g OGTT, with blood samples taken 2h after glucose intake. Only fasting venous plasma glucose was measured in participants with known diabetes. Fasting and 2-h plasma glucose values were analyzed with the Hitachi 912 system (Roche Diagnostics). Fasting and 2-h serum insulin levels were analyzed by an immunoassay method excluding des-31,32 split products and intact proinsulin (AutoDELFIA; Perkin Elmer). Indices of insulin sensitivity and insulin secretion were derived from excluding des-31,32 split products and intact proinsulin (AutoDELFIA; Perkin Elmer). Indices of insulin sensitivity and insulin secretion were derived from excluding des-31,32 split products and intact proinsulin (AutoDELFIA; Perkin Elmer).

Identification and selection of loss-of-function variants to study. To identify new loss-of-function variants in the Greenlandic population, we used the exome data from 27 Greenlandic individuals in nine trios presented previously3,4. In this dataset, we identified 46 loss-of-function variants which were not present in 1000 Genomes Project (version 138 and below) and that were not present in the NHLBI Exome Sequencing Project (ESP) (Supplementary Table 1). We then used the GWAS catalog (downloaded January 2013) to assign known associations to the loci with the loss-of-function variants using associations with P-values less than 10E-5. Finally, we selected the variants with the strongest association with BMI or BMI-related traits. This selection procedure left us with one predicted loss-of-function variant, ADCY3 c.2433–1G>A (Supplementary Fig. 1). This variant is located in ADCY3 and was present in a single parent in one of the nine trios in heterozygous form.

Genotype data. Subsequently, we genotyped ADCY3 c.2433–1G>A in the two Greenlandic cohorts, IHT and B99, using the KASP Genotyping Assay (LGC Genomics). The genotyping reaction was performed in duplicate for the association testing, we used quality-filtered Metabochip genotype data on which to base our genetic similarity matrix estimates. The genotyping and quality filtering of the Metabochip data have been described in detail earlier11. The filtered Metabochip dataset contains data from 2,791 individuals from IHT and 1,336 individuals from B99, and we have ADCY3 c.2433–1G>A genotype data for 4,038 of these individuals, 2,779 from IHT and 1,259 from B99.

Association testing and replication. Greenlandic cohorts. To test for association between the ADCY3 variant and the different phenotypes of interest, we used a linearized mixed model in the software GEMMA to account for relatedness and admixture12. For each phenotype, GEMMA was applied to data from the subset of individuals across the two cohorts with information available about that specific phenotype. The genetic similarity matrix required as input to GEMMA was estimated from Metabochip genotype data12. For each phenotype, we performed two association tests: a test where we assumed a recessive effect model and a test where we assumed an additive effect model. All tests were performed with sex, age and cohort included as covariates. An additional analysis also included BMI as a covariate. Before all tests for association with quantitative traits, we quantile transformed the phenotype data to a normal distribution separately for each sex, which gave the results reported as βsex, SESD and P. However, to get effect sizes in traits in which the sex effect was also present, we performed analyses without transforming the phenotype data, which resulted in the results reported as β. For the binary trait type 2 diabetes, we did not perform any transformation and the test results are reported as β and P.

To assess the appropriateness of the recessive effect model and the additive effect model for BMI and type 2 diabetes, we compared each of these effect models against a full genotype model, where each of the three genotypes have an independent effect. To do this, we used the same linear model framework as we did for the association tests. These additional tests were only performed in the IHT cohort because there were no homozygous carriers in B99, which in practice means the tests cannot be performed if B99 is included.

Accelerating Medicines Partnership Type 2 Diabetes Knowledge Portal (AMP-T2D). Data from AMP-T2D generated as part of the GoT2D, T2DGenes, SIGMA and LaCAMP consortia13 were queried for association between a burden of loss-offunction variants in ADCY3 and type 2 diabetes. The data contained 18,869 samples with exome sequence data, of which 18,176 samples were informative for type 2 diabetes. Sample quality was assessed from five ancestors (6,356 European, 5,722 Hispanic, 2,199 South Asian, 2,158 East Asian and 1,741 African American). Variants annotated as stop-gain or frameshift variants or variants in a splice adpator/donor site were considered loss of function. Burden association analyses were performed with each individual coded as carrying a loss-of-function variant or not. We used logistic regression and adjusted for principal components 1–4 as well as age and sex by including these as covariates. We also analyzed all other genes with at least five loss-of-function variants with at least one mutation carrier for each variant. Only loss-of-function variants with MAF <5% were used. This analysis showed that there is no general inflation in the test statistic (Supplementary Fig. 3).

Greek isolated populations. As part of the HELIC study14, 1,642 samples from the Pomak villages in northern Greece and 1,482 samples from the Mylopotamos villages in Crete were sequenced at an average depth of 18.6x and 22.5x, respectively, using the HiSeqX platform. Adapter-ligated libraries were amplified by six cycles of PCR and subjected to DNA sequencing according to the manufacturer's instructions. Genomic (G)VCFs were created for 200 equally sized chunks using GATK HaploTypeCaller v3.5, combined into batches of 150 samples using GATK CombineGVCFs and called using GATK GenotypeGVCFs. Variant-level quality control was performed using GATK VQSR v3.5.

Estimation of admixture proportions, ancestral allele frequencies and relatedness. For the Greenlanders, all admixture proportions were obtained from ref. 1. To estimate the ancestral Inuit allele frequencies for the ADCY3 variant, we performed maximum-likelihood estimation using the likelihood model from admIXTURE15 with the admixture proportions fixed and including only genotype data from the variant of interest. To achieve the maximum-likelihood estimates, we applied the original algorithm15, as well as the Admixture-based5 on the estimated admixture proportions. None of the seven homozygous carriers are closely related to each other (second-degree relatives or closer); however, one pair may be first cousins with an estimated kinship coefficient of 0.08.

RNA sequencing analysis. RNA sequencing was performed on leukocytes from 17 Greenlandic individuals (7 GG carriers, 6 GA carriers and 4 AA carriers). Total RNA was extracted from 2.5 ml of peripheral blood with the PAXGene miRNA kit (Qagen) according to the manufacturer's protocol and subjected to on-column DNase I treatment with RNase-free DNase (Qagen). RNA quality and purity were checked using an Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit) and NanoDrop, respectively.

The RNA sequencing library preparation was performed following the instructions of the TruSeq RNA Sample Prep Kit v2 (Illumina). For mRNA isolation and fragmentation, 200 ng of total RNA was purified on oligo(dT) beads before fragmentation with Enzyme, Prime, Fragment Mix. First-strand cDNA was generated using First-Strand Mix and SuperScript II (Invitrogen) reverse transcription master mix for 50 min at 42 °C and 40 min at 70 °C. The second strand was synthesized by adding Second-Strand Master Mix (16 °C for 15 min). The fragmented cDNA was end-repaired and purified with AMPure XP Beads (Agencourt). A-Tailing Mix was added, and reactions were incubated at 37 °C for 30 min. For adapter ligation, Adenylate 3’ Ends DNA, RNA Index Adaptor and Ligation Mix were mixed and reactions were incubated at 30 °C for 10 min. Each sample was then amplified in the so-called PCR program. Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix were performed to enrich the cDNA fragments. The PCR products were purified.
with AMPure XP Beads (Agencourt). The average molecule length was measured using the Agilent 2100 Bioanalyzer instrument (Agilent DNA 1000 Reagents) and by real-time qPCR (TaqMan Probe). The qualified libraries were amplified on a cBot to generate the cluster on the flow cell (TruSeq PE Cluster Kit V3–cBot–HS, Illumina). The amplified flow cell was sequenced using paired-end reads on the HiSeq 4000 system (TruSeq SBS Kit–HS V3, Illumina).

The obtained RNA-seq libraries, as well as one RNA-seq library from adipose tissue of a healthy European-ancestry female11, were processed by Trimmomatic v 0.3233, trimming the reads using HEADCROP:11 LEADING:22 SLIDINGWINDOW:4:22 MINLEN:25. The quality of all libraries was checked with FastQC, before and after trimming. Trimmed reads were mapped to the human reference genome (hg19) using Hisat2 v2.0.1-beta14 with splice-site information from GENCODE v1929 and set to annotate properly paired reads as those with a minimum insert size from 0 to 1,000 nt. Sashimi plots29 were generated from the IGV genome browser17. PSI was calculated as \( r_{\text{exon:intron}} / (r_{\text{exon:intron}} + r_{\text{exon:exon}}) \), where \( r \) is the number of reads indicating the different splice patterns. \( r_{\text{exon:intron}} \) was obtained by extracting relevant GENCODE v1935 exon:intron boundaries \( \pm 3 \) nt. Reads mapping to these regions were quantified using featureCounts from the R package Rsubread38 specifying countChimericFragments = FALSE and minOverlap = 6, ensuring that reads mapped across the exon boundary. \( r_{\text{exon:exon}} \) was calculated by running featureCounts with the juncCounts = TRUE parameter. Plots were generated using ggplot2. For transcript quantification, we constructed a gtf file containing three versions of GENCODE v19 transcript ENST00000260600.5, the longest transcript for ADCY3. The corresponding genomic sequences were extracted using the Cufflinks gffread tool39, and a Salmon index was built using Salmon. The corresponding genomic sequences were extracted transcript for ADCY3 three versions of GENCODE v19 transcript ENST00000260600.5, the longest featureCounts with the juncCounts = TRUE parameter. Plots were generated using ggplot2. For transcript quantification, we constructed a gtf file containing three versions of GENCODE v19 transcript ENST00000260600.5, the longest transcript for ADCY3. The corresponding genomic sequences were extracted using the Cufflinks gffread tool39, and a Salmon index was built using Salmon. The corresponding genomic sequences were extracted transcript for ADCY3 three versions of GENCODE v19 transcript ENST00000260600.5, the longest featureCounts with the juncCounts = TRUE parameter. Plots were generated using ggplot2. For transcript quantification, we constructed a gtf file containing three versions of GENCODE v19 transcript ENST00000260600.5, the longest transcript for ADCY3. The corresponding genomic sequences were extracted using the Cufflinks gffread tool39, and a Salmon index was built using Salmon. The corresponding genomic sequences were extracted transcript for ADCY3 three versions of GENCODE v19 transcript ENST00000260600.5, the longest featureCounts with the juncCounts = TRUE parameter. Plots were generated using ggplot2. For transcript quantification, we constructed a gtf file containing three versions of GENCODE v19 transcript ENST00000260600.5, the longest transcript for ADCY3. The corresponding genomic sequences were extracted using the Cufflinks gffread tool39, and a Salmon index was built using Salmon.

References
22. Bjerrregaard, P. et al. Int. J. Circumpolar Health 62 (Suppl. 1), 3–79 (2003).
23. Bjerrregaard, P. Inuit Health in Transition—Greenland Survey 2005–2010 2nd edn (National Institute of Public Health, Copenhagen, 2011).
24. Philipsen, A. et al. PLoS One 10, e0123062 (2015).
25. Jørgensen, M. E., Borch-Johnsen, K., Stolk, R. & Bjerrregaard, P. Diabetes Care 36, 2988–2994 (2013).
26. Matthews, D. R. et al. Diabetologia 28, 412–419 (1985).
27. Gutt, M. et al. Diabetes Res. Clin. Pract. 47, 177–184 (2000).
28. World Health Organization Study Group. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications (World Health Organization, Geneva, 1999).
29. Andersen, M. K. et al. PLoS Genet. 12, e1006119 (2016).
30. Zhou, X. & Stephens, M. Nat. Genet. 44, 821–824 (2012).
31. Alexander, D. H., Novembre, J. & Lange, K. Genome Res. 19, 1655–1664 (2009).
32. Möltke, I. & Albrechtsen, A. Bioinformatics 30, 1027–1028 (2014).
33. Bolger, A. M., Lohse, M. & Usadel, B. Bioinformatics 30, 2114–2120 (2014).
34. Kim, D., Langmead, B. & Salzberg, S. L. Nat. Methods 12, 357–360 (2015).
35. Harrow, J. et al. Genome Res. 22, 1760–1774 (2012).
36. Katz, Y. et al. Bioinformatics 31, 2400–2402 (2015).
37. Robinson, J. T. et al. Nat. Biotechnol. 29, 24–26 (2011).
38. Liao, Y., Smyth, G. K. & Shi, W. Bioinformatics 30, 923–930 (2014).
39. Trapnell, C. et al. Nat. Biotechnol. 28, 511–515 (2010).
40. Pato, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Nat. Methods 14, 417–419 (2017).
41. Punta, M. et al. Nucleic Acids Res. 40, D290–D301 (2012).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

### Experimental design

1. **Sample size**
   
   Describe how sample size was determined.

   Under an additive effect model the sample size of ~3,000 is large enough to have the power (>80%) to detect associations with a single genetic variant with MAF=2% with modest to high effect sizes (0.4 s.d.) comparable to the known genetic variants associated with lipid levels in Europeans.

2. **Data exclusions**
   
   Describe any data exclusions.

   In the association analyses of the Greenlandic cohorts, we excluded individuals for whom we did not have quality filtered Metabochip data, because such data was needed to be able to correct for admixture and relatedness.

3. **Replication**
   
   Describe whether the experimental findings were reliably reproduced.

   The variant was found through exome sequencing and validated by genotyping and with RNA sequencing. Since the identified variant is extremely rare, if at all present, in other populations we were not able to perform replication of the association signal for this exact variant in other populations. Instead, we performed a burden test for other loss-of-function variants in the same gene in other populations. This test supported our initial finding and conclusion.

4. **Randomization**
   
   Describe how samples/organisms/participants were allocated into experimental groups.

   NA

5. **Blinding**
   
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

| Parameter | Confirmed |
|-----------|-----------|
| n/a       |           |

☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)

☐ ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.

☐ ☒ A statement indicating how many times each experiment was replicated

☐ ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)

☐ ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons

☐ ☒ The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted

☐ ☒ A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)

☒ ☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software
Policy information about availability of computer code

Describe the software used to analyze the data in this study. As described in the online methods, we used the publicly available software package GEMMA to perform all the association tests in the Greenlandic cohorts. To perform the burden test, we used the standard implementation of logistic regression in the statistical software R. Software used for RNA seq analysis included: Trimmomatic, FastQC, Salmon v0.8.2 and R.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

8. Materials and reagents
Policy information about availability of materials

8. Materials availability
Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No restrictions

9. Antibodies
Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

NA

10. Eukaryotic cell lines
a. State the source of each eukaryotic cell line used.

NA

b. Describe the method of cell line authentication used.

NA
c. Report whether the cell lines were tested for mycoplasma contamination.

NA
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

NA
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.
   NA

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.
   The Greenlandic samples are from two different cohorts: B99 (N=1,401) and Inuit Health in Transition (IHIT) (N=3,115), which were collected in Greenland as a part of general population health surveys conducted in 1999-2001 and 2005-2010, respectively. Two hundred and ninety-five individuals overlapped between the two cohorts and were assigned to the B99 dataset. Clinical characteristics of the participants with genotype data are shown in Supplementary Table 6.