Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population

Grarup, Niels; Moltke, Ida; Andersen, Mette K.; Bjerregaard, Peter; Larsen, Christina Viskum Lytken; Dahl-Petersen, Inger K.; Jørsboe, Emil; Tiwari, Hemant K.; Hopkins, Scarlett E.; Wiener, Howard W.; Boyer, Bert B.; Linneberg, Allan; Pedersen, Oluf; Jørgensen, Marit Eika; Albrechtsen, Anders; Hansen, Torben

Published in:
Diabetologia

DOI:
10.1007/s00125-018-4659-2

Publication date:
2018

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

Document license
CC BY

Citation for published version (APA):
Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population

Niels Grarup1 · Ida Moltke2 · Mette K. Andersen1 · Peter Bjerregaard3,4 · Christina V. L. Larsen3,4 · Inger K. Dahl-Petersen3 · Emil Jørsboe2 · Hemant K. Tiwari5 · Scarlett E. Hopkins6 · Howard W. Wiener7 · Bert B. Boyer6 · Allan Linneberg8,9 · Oluf Pedersen1 · Marit E. Jørgensen3,4,10 · Anders Albrechtsen2 · Torben Hansen1,11

Received: 30 January 2018 / Accepted: 3 May 2018 © The Author(s) 2018

Abstract

Aims/hypothesis In a recent study using a standard additive genetic model, we identified a TBC1D4 loss-of-function variant with a large recessive impact on risk of type 2 diabetes in Greenlanders. The aim of the current study was to identify additional genetic variation underlying type 2 diabetes using a recessive genetic model, thereby increasing the power to detect variants with recessive effects.

Methods We investigated three cohorts of Greenlanders (B99, n = 1401; IHIT, n = 3115; and BBH, n = 547), which were genotyped using Illumina MetaboChip. Of the 4674 genotyped individuals passing quality control, 4648 had phenotype data available, and type 2 diabetes association analyses were performed for 317 individuals with type 2 diabetes and 2631 participants with normal glucose tolerance. Statistical association analyses were performed using a linear mixed model.

Results Using a recessive genetic model, we identified two novel loci associated with type 2 diabetes in Greenlanders, namely rs870992 in ITGA1 on chromosome 5 (OR 2.79, p = 1.8 × 10−8), and rs16993330 upstream of LARGE1 on chromosome 22 (OR 3.52, p = 1.3 × 10−7). The LARGE1 variant did not reach the conventional threshold for genome-wide significance (p < 5 × 10−8) but did withstand a study-wide Bonferroni-corrected significance threshold. Both variants were common in Greenlanders, with minor allele frequencies of 23% and 16%, respectively, and were estimated to have large recessive effects on risk of type 2 diabetes in Greenlanders, compared with additively inherited variants previously observed in European populations.

Niels Grarup, Ida Moltke and Mette K. Andersen contributed equally to this study.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00125-018-4659-2) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

1 Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark
2 The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark
3 National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark
4 Greenland Centre for Health Research, University of Greenland, Nuuk, Greenland
5 Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA
6 Center for Alaska Native Health Research, University of Alaska Fairbanks, Fairbanks, AK, USA
7 Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA
8 Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark
9 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
10 Steno Diabetes Center Copenhagen, Gentofte, Denmark
11 Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark

Published online: 20 June 2018
Conclusions/interpretation We demonstrate the value of using a recessive genetic model in a historically small and isolated population to identify genetic risk variants. Our findings give new insights into the genetic architecture of type 2 diabetes, and further support the existence of high-effect genetic risk factors of potential clinical relevance, particularly in isolated populations. 

Data availability The Greenlandic MetaboChip-genotype data are available at European Genome-Phenome Archive (EGA; https://ega-archive.org/) under the accession EGAS00001002641.

Keywords Genetic association · Genome-wide association study · Greenlanders · Inuit · ITGA1 · LARGE1 · Recessive genetic model · Type 2 diabetes

Abbreviations CANHR Center for Alaska Native Health Research EAF Effect allele frequency eQTL Expression quantitative trait loci IHIT Inuit Health in Transition GWAS Genome-wide association study/studies ISI Insulin sensitivity index MAF Minor allele frequency

Introduction Numerous genome-wide association studies (GWAS) have been performed to identify genetic factors predisposing to type 2 diabetes. These studies, mainly performed in European and Asian populations, have identified around 120 genetic variants associated with risk of type 2 diabetes [1–8].

Until now, GWAS of type 2 diabetes and glycaemic traits have almost exclusively been performed using an additive genetic model [1–5]. This model, however, has limited statistical power to detect associations with variants displaying recessive effects [9, 10], unless the effect is very large. Recently, a GWAS of type 2 diabetes in up to 4040 individuals with type 2 diabetes and 116,246 participants without diabetes from the UK Biobank applied a dominance deviation model to identify non-additive association signals [11]. Although no novel signals were identified, the paper reported a recessive effect on risk of type 2 diabetes at the previously identified CDKAL1 locus [11], which is in concordance with the findings from the original discovery [12]. Importantly, given the effective sample size of around 15,300, this study lacked statistical power to exclude the possibility that additional alleles with recessive effects predispose to type 2 diabetes in European populations.

We previously demonstrated the existence of a high-impact type 2 diabetes risk allele with predominately recessive effect in the Greenlandic population, despite using an additive genetic model in the discovery analysis [13]. Notably, however, this TBC1D4 loss-of-function variant was discovered via analyses of a type 2 diabetes-related trait, 2 h plasma glucose, on which it had an extremely large effect, and not via analyses of type 2
diabetes. Indeed, the study in which that variant was discovered had very limited statistical power to detect variants with recessive effects on type 2 diabetes unless they were very frequent and had very large effects. Here, we aimed to identify additional variants with recessive effects on risk of type 2 diabetes in the Greenlandic population by applying a recessive genetic model to an increased number of Greenlandic samples with type 2 diabetes information.

Methods

Study populations The Greenlandic study sample comprised individuals from two cohorts selected as part of a general population health survey of the Greenlandic population during the periods 1999–2001 (B99, n = 1401) and 2005–2010 (Inuit Health in Transition (IHit), n = 3115) and from a cohort of Greenlanders living in Denmark on whom information was collected during 1998–1999 (BBH, n = 547) [14, 15]. There was an overlap of individuals (n = 295) between the IHIT and B99 cohorts, and these were assigned to B99. Of the 4674 genotyped individuals who passed quality control, 4648 had phenotype data available. The characteristics of the three cohorts are shown in electronic supplementary material (ESM) Table 1. All participants gave informed consent, and the study was approved by the Commission for Scientific Research in Greenland (project 2011-13, ref. no. 2011-056978; and project 2013-13, ref. no. 2013-090702), and conducted in accordance with the ethical standards of the Declaration of Helsinki, second revision.

Yup’ik samples for quantitative trait replication came from the Center for Alaska Native Health Research (CANHR), which performs studies related to genetic, behavioural, and nutritional risk factors for obesity and cardiometabolic diseases among Yup’ik people in a community-based setting [16]. Recruitment of Yup’ik families was initiated in 2003 and continues in 11 Southwest Alaska communities, where all residents are invited to participate, resulting in a convenience sample. The present study sample comprised 1059 non-pregnant Yup’ik individuals aged 14 years or above at the time of enrolment. All participants signed informed consent documents, and the study protocols were approved by the Institutional Review Boards of the University of Alaska and the National and Alaska Area Indian Health Service Institutional Review Boards, as well as the Yukon Kuskokwim Health Corporation Human Studies Committee.

Additional replication analyses were performed on up to 23,776 Danish samples from the Inter99 study (n = 4947) (CT00289237, ClinicalTrials.gov) [17], from Health2006–Health2010 studies (n = 4776) [18], from the Danish study of Functional Disorders (DanFunD; n = 6004) [19], from the Vejle Diabetes Biobank (n = 6266) [20], and from the Steno Diabetes Center Copenhagen (n = 1783). Characteristics of the cohorts have previously been published [6]. The studies were approved by the appropriate Regional Ethical Committees and were performed in accordance with the scientific principles of the Helsinki Declaration, second revision.

Measurements and assays The Greenlandic participants underwent an OGTT in which blood samples were drawn after an overnight fast and after 2 h during a 75 g OGTT. Plasma glucose levels were analysed using a Hitachi 912 system (Roche Diagnostics, Indianapolis, IN, USA), serum insulin with an immunoassay excluding des-31,32 split products and intact proinsulin (AutoDELFIA; PerkinElmer, Waltham, MA, USA), and HbA1c by ion-exchange HPLC (G7 [Tosoh Bioscience, Tokyo, Japan] for IHIT samples; VARIANT [Bio-Rad, Hercules, CA, USA] for B99 samples). The thickness of visceral and subcutaneous adipose tissue was measured according to a validated protocol using a portable ultrasound scanner (Pie Medical, Maastricht, the Netherlands) with a 3.5 MHz transducer [21]. Serum cholesterol, HDL-cholesterol and triacylglycerol concentrations were measured using enzymatic calorimetric techniques (Hitachi 917, Roche Molecular Biochemicals, Indianapolis, IN, USA). LDL-cholesterol concentration was calculated according to Friedewald’s formula. Insulin resistance was estimated by either: (1) HOMA-IR [22], calculated as ([fasting glucose level (mmol/l) × fasting insulin (pmol/l)/6.945])/22.5; or (2) the insulin sensitivity index, ISI(0,120) [23], calculated as [(75,000 + [fasting glucose (mmol/l)] × (18 − 2 h glucose [mmol/l]) × 18) × 0.19 × weight [kg]/120] / ([fasting glucose (mmol/l) + 2 h glucose (mmol/l)]/2) / log([fasting insulin (pmol/l)/6.945] + [2 h insulin (pmol/l)/6.945])/2, where log is the natural logarithm.

In the CANHR Yup’ik study population, blood samples were collected after an overnight fast. Insulin was measured with a radioimmunoassay applying an 125I-iodinated insulin tracer, anti-human insulin-specific antibody and human insulin standards from Linco Research (Winchester, VA, USA). The Poly-Chem System Chemistry Analyzer (Polymedco, Courtlandt Manor, NY, USA) was applied to measure HDL-cholesterol, total cholesterol and triacylglycerols. Fasting blood glucose was measured on a Cholestech LDX analyser (Cholestech, Hayward, CA, USA), and glycosylated haemoglobin was measured on a Bayer HbA1c DCA 2000+ analyser (Bayer, Leverkusen, Germany).

In the Danish samples, plasma glucose was measured by the glucose oxidase method (Granutex; Merck, Darmstadt, Germany).

In the Greenlandic and Danish cohorts, type 2 diabetes was defined based on self-reported type 2 diabetes, fasting plasma glucose level >7 mmol/l, or 2 h plasma glucose during an OGTT >11.1 mmol/l. Control individuals were normal glucose tolerant with a fasting plasma glucose <6.1 mmol/l and 2 h plasma glucose during an OGTT <7.8 mmol/l.
**Genotyping** The Greenlandic samples were genotyped using Illumina MetaboChip (Illumina, San Diego, CA, USA), which contains 196,725 SNPs potentially related to metabolic, cardiovascular or anthropometric traits [24]. Details about the genotyping procedure and quality control have previously been described [25]. In total, 4674 individuals (IHIT, 2791; B99, 1336; BBH, 547) and 115,182 SNPs passed the quality control.

Detailed descriptions of the genotyping procedures and data cleaning of CANHR Yup’ik samples have also previously been published [26]. Briefly, we used the Illumina Linkage IV panel to genotype 6090 SNPs spanning the entire genome, with an average genetic distance of 0.58 cM. A total of 5632 autosomal SNPs from this Linkage IV panel passed the quality control. These SNPs were used to obtain ancestry information for the statistical analysis. Additionally, genotyping of SNPs for replication was performed by the KASPar Genotyping assay (LGC Genomics, Hoddesdon, UK).

The Danish samples were genotyped by the Illumina Infinium OmniExpress-24 v1.1 array. Genotypes were called by the Illumina GenCall algorithm, and variants with a call rate <98% and Hardy–Weinberg equilibrium $p < 1 \times 10^{-5}$ were removed. Samples were excluded if they were ethnic outliers, had a mismatch between genetic and phenotypic sex or had a call rate <95%. Genotype data were imputed using the Haplotype Reference Consortium (HRC) reference panel v1.1 [27] at the Michigan imputation server using Minimac3 v2.0.1 (https://genome.sph.umich.edu/wiki/Minimac3) after phasing with Eagle2 v2.4 (https://data.broadinstitute.org/alkesgroup/Eagle/) [28]. Post-imputation filtering of SNPs excluded variants with a minor allele frequency (MAF) <0.01 and info score <0.70.

**Statistical power simulations** To assess the statistical power of an additive and a recessive model-based test to detect association to a variant with recessive effect, we performed a range of simulations of data from a locus with a type 2 diabetes effect allele. For each of 540 combinations of effect allele frequencies (EAFs; EAF values 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3) and effect sizes (ORs ranging from 1.1 to 10 with a step size of 0.1), we simulated first genotypes for 2948 individuals to reflect the number of individuals with type 2 diabetes information available in this study. The Yup’ik cohort was also analysed using GEMMA software [29]. Here, however, the genetic similarity matrix required for the association analysis was not estimated using GEMMA, as we did in our analyses of the Greenlanders, but was instead calculated using the genotype data from the linkage panel merged with the additional genotypes of the SNPs genotyped for this study.

In the Danish data, association analyses of type 2 diabetes were carried out by logistic regression adjusting for age, sex and the first ten principal components. These analyses were performed on imputed genotype data, taking genotype uncertainty into account, applying the expected count test in SNPTEST v2.5.2 (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html).

**Allele frequency estimation** For the SNPs of interest, we estimated allele frequencies for each of the Inuit and European ancestry components of the Greenlandic population using a
two-step approach. In step 1, we estimated ancestry proportions for the Greenlandic study individuals as well as 50 Danish individuals. To do this, we applied ADMIXTURE v1.3.0 (https://www.genetics.ucla.edu/software/admixture/) [32] to all the SNP data from these individuals, assuming two ancestral populations—Inuit and Europeans. In step 2, we estimated ancestral allele frequencies with CIs for each SNP separately using bootstrap with replacement. We used 1000 bootstrap samples of individuals and, based on each, we performed maximum likelihood estimation of the allele frequencies, using the likelihood from ADMIXTURE with the ancestry proportions fixed to the estimates obtained in step 1. The CIs were based on the quantiles of these bootstrap estimates.

In CANHR Yup’ik samples, allele frequencies for the SNPs of interest were estimated using the Mendel program v16 (https://www.genetics.ucla.edu/software/mendel) [33, 34].

**Querying results from previous large GWAS of type 2 diabetes and metabolic traits** Additive genetic model GWAS results were queried online: type 2 diabetes results from the DIAGRAM Consortium (DIAbetes Genetics Replication And Meta-analysis; http://diagram-consortium.org) [8], glycaemic trait results from MAGIC (the Meta-Analyses of Glucose and Insulin-related traits; www.magicinvestigators.org) [35], anthropometric trait results from the GIANT Consortium (Genetic Investigation of Anthropometric Traits; http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium) [36–38], and lipid results from GLGC (Global Lipids Genetics Consortium; http://lipidgenetics.org) [39].

**Assessment of functional effects and expression quantitative trait loci** To investigate whether the associated variants were causal, RegulomeDB (http://www.regulomedb.org/, accessed 10 January 2018) and HaploReg v4.1 (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php) were applied to assess possible co-localisation of the genetic variants with regulatory elements, such as transcription factor binding sites, promoter regions and regions of DNAase hypersensitivity. We also used data from the GTEx Consortium available online to investigate associations between genetic variants and RNA expression (www.gtexportal.org; accessed 27 January 2018). All 48 tissues with more than 70 samples were queried. Sample sizes ranged from 80 to 388. Results were only available for an additive genetic model.

**Results**

We analysed association with type 2 diabetes for 317 individuals with type 2 diabetes and 2631 participants with normal glucose tolerance. Using a recessive genetic model on these Greenlandic data, we identified three loci associated with type 2 diabetes below the study-wide Bonferroni-corrected significance threshold ($p<4.3 \times 10^{-7}$) (ESM Fig. 1). Besides the previously described TBC1D4 locus [13], these comprised two novel loci, namely the common intron variant rs870992 in ITGA1 on chromosome 5 (MAF 23%, OR 2.79, $p=5.18 \times 10^{-8}$) and the common intergenic variant rs16993330 approximately 25 kb upstream of LARGE1 on chromosome 22 (MAF 16%, OR 3.52, $p=1.3 \times 10^{-7}$) (Fig. 1, Table 1). However, the LARGE1 variant did not reach the conventional threshold for genome-wide significance ($p<5 \times 10^{-8}$). Similar associations with type 2 diabetes were observed when adjusting for BMI (data not shown).

The Greenlandic population is admixed with ancestry from both Inuit and Europeans [30]. Interestingly, the estimated EAF were considerably higher in the Inuit ancestry component for ITGA1 rs870992 (31%) and for LARGE1 rs16993330 (19%), than in the European ancestry component among Greenlanders (3% and 7%, respectively) and in European populations (Table 1, ESM Table 2). The allele frequencies found in Europeans were so low that homozygous carriers were almost absent. The novel variants in ITGA1 and LARGE1 had a high impact on risk of type 2 diabetes in Greenlanders, evident by the estimated effect sizes (Table 1) and by the higher frequency of type 2 diabetes among homozygous carriers compared with non-carriers (24.9% and 28.9% vs 10%; Fig. 2). To gain further insights into the effect of the identified loci, we also performed type 2 diabetes association analyses of these assuming an additive model (Table 1). This led to low $p$ values as well, although none of them was low enough to pass the Bonferroni-corrected significance threshold ($p<4.3 \times 10^{-7}$). Based on genetic data and functional annotation analyses facilitated by in silico tools, we were unable to determine whether the identified variants were causal.

Next, to elucidate the diabetes-related mechanisms underlying the association of the novel variants, we analysed diabetes-related traits in individuals without diabetes from the Greenlandic cohorts with a recessive genetic model. Homozygous carriers of ITGA1 rs870992 had nominally increased HbA1c levels ($\beta=0.13$ SD [95% CI 0.027, 0.23], $p=0.012$) (Table 2). For LARGE1 rs16993330, the underlying physiological mechanism seems to be linked to increased fat accumulation, indicated by higher levels of visceral adipose tissue (0.29 SD [95% CI 0.076, 0.50], $p=0.0078$), larger hip circumference (0.22 SD [95% CI 0.048, 0.40], $p=0.013$) and waist circumference (0.21 SD [95% CI 0.041, 0.38], $p=0.015$) and higher BMI (0.18 SD [95% CI 0.0046, 0.35], $p=0.045$) (Table 2). Furthermore, homozygous LARGE1 rs16993330 carriers had increased insulin resistance, indicated by higher levels of fasting serum insulin (0.22 SD [95% CI 0.027, 0.41], $p=0.026$) and HOMA-IR (0.24 SD [95% CI 0.048, 0.44], $p=0.014$), and lower levels of the ISI (−0.21 SD [95% CI −0.016, −0.41], $p=0.034$) (Table 2). We then
performed the analyses of the same quantitative diabetes-related traits with an additive model and obtained similar results (ESM Table 3).

We subsequently attempted to validate our findings in another Arctic indigenous population by genotyping the ITGA1 and LARGE1 variants in 1059 Alaska Native Yup'ik. However, none of the potential associations with diabetes-related quantitative traits were replicated in this smaller sample (ESM Table 4), and analysis of type 2 diabetes was precluded due to the low number of participants with diabetes. A second line of replication was attempted in a Danish sample of 5220 individuals with type 2 diabetes and 18,556 control participants. However, we did not observe an association with type 2 diabetes using a recessive model (ESM Table 2). In addition, the results of a dominance deviance model GWAS of European UK Biobank samples [11] showed no significant rejection of the additive model for either ITGA1 rs870992 or LARGE1 rs16993330 (p = 0.54 and p = 0.36, respectively) based on 117,775 individuals. We also queried the two variants from the available summary data from a recent additive genetic model GWAS of type 2 diabetes [8] and found that the ITGA1 rs870992 G allele was nominally associated with increased risk of type 2 diabetes, whereas no effect was observed for LARGE1 rs16993330 (ESM Table 5). Additionally, queries of results from GWAS of other traits revealed associations of the ITGA1 rs870992 G allele with increased total cholesterol and LDL-cholesterol levels, and nominal associations of the LARGE1 rs16993330 A allele with increased total cholesterol and triacylglycerol levels (ESM Table 5). Owing to the

![Figure 1](https://example.com/fig1.png)

**Fig. 1** Association results with type 2 diabetes for those SNPs on the MetaboChip that are located in a 2 Mb region around the lead SNPs: (a) rs870992 in ITGA1, and (b) rs16993330 near LARGE1. Each SNP in the region is represented by a circle whose colour indicates the extent of correlation ($r^2$) between the SNP and the lead SNP, which is shown in red. 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 type 2 diabetes as determined using a recessive model. The solid blue line illustrates the recombination rate from the Chinese HapMap Phase III panel (www.sanger.ac.uk/resources/downloads/human/hapmap3.html). The protein-coding genes in the genetic region are shown below the plot.
Table 1  Variants on MetaboChip associated with type 2 diabetes in Greenlanders under a recessive genetic model

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chr</th>
<th>EA</th>
<th>All (EAF) (95% CI)</th>
<th>Inuit (EAF) (95% CI)</th>
<th>EU (EAF) (95% CI)</th>
<th>Type 2 diabetes</th>
<th>Additive model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>β (95% CI)</td>
<td>p value</td>
<td>OR</td>
<td>OR</td>
<td>β (95% CI)</td>
</tr>
<tr>
<td>TBC1D4 rs7330796</td>
<td>13</td>
<td>T</td>
<td>0.22 (0.21, 0.23)</td>
<td>0.27 (0.25, 0.29)</td>
<td>0.10 (0.08, 0.13)</td>
<td>0.18 (0.13, 0.23)</td>
<td>1.0 × 10^{-14}</td>
</tr>
<tr>
<td>ITGAI rs870992</td>
<td>5</td>
<td>G</td>
<td>0.23 (0.22, 0.24)</td>
<td>0.31 (0.30, 0.33)</td>
<td>0.03 (0.01, 0.06)</td>
<td>0.12 (0.079, 0.17)</td>
<td>1.8 × 10^{-8}</td>
</tr>
<tr>
<td>LARGE1 rs16993330</td>
<td>22</td>
<td>A</td>
<td>0.16 (0.15, 0.17)</td>
<td>0.19 (0.18, 0.21)</td>
<td>0.07 (0.05, 0.09)</td>
<td>0.17 (0.11, 0.23)</td>
<td>1.3 × 10^{-7}</td>
</tr>
</tbody>
</table>

Analyses of type 2 diabetes comprised 317 individuals with type 2 diabetes and 2631 control participants with normal glucose tolerance. ORs were estimated using logistic regression and reported for the entire Greenlandic study sample (OR_{All}) and the Inuit ancestry component (OR_{Inuit}). EAFs are given for the entire Greenlandic study sample (All), as well as the Inuit ancestry component (Inuit), and the European ancestry component (EU) of the Greenlandic study population.

rs7330796 is the discovery variant for TBC1D4 on the MetaboChip, not the causal variant [13].
expression quantitative trait loci (eQTL) associations were found for LARGE1 expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancreas and smooth muscle tissue in up to 361 samples, albeit with modest expression in nerve, pancrea
Table 2  Association of ITGA1 rs870992 and LARGE1 rs16993330 with quantitative metabolic traits in Greenlanders under a recessive genetic model

<table>
<thead>
<tr>
<th>Trait (measured unit)</th>
<th>ITGA1 rs870992</th>
<th>LARGE1 rs16993330</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n  βSD  95% CI</td>
<td>βSD  95% CI  β  p value</td>
</tr>
<tr>
<td>Fasting plasma glucose (mmol/l)</td>
<td>3693 0.12 -0.0034, 0.25 0.16</td>
<td>0.15 -0.028, 0.32 0.19 0.10</td>
</tr>
<tr>
<td>2 h plasma glucose (mmol/l)</td>
<td>3437 0.070 -0.063, 0.20 0.36</td>
<td>0.12 -0.069, 0.31 0.32 0.21</td>
</tr>
<tr>
<td>Fasting serum insulin (pmol/l)</td>
<td>3691 0.016 -0.12, 0.16</td>
<td>1.9 0.82</td>
</tr>
<tr>
<td>2 h serum insulin (pmol/l)</td>
<td>3437 0.0090 -0.13, 0.15</td>
<td>-5.0 0.91</td>
</tr>
<tr>
<td>HbA1c (mmol/mol)</td>
<td>4626 0.13 0.027, 0.23</td>
<td>NA 0.012</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>4624 0.13 0.027, 0.23</td>
<td>0.063, 0.20 0.36 0.31</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>3684 0.053 -0.086, 0.19</td>
<td>0.15 0.45</td>
</tr>
<tr>
<td>IS(0,120)</td>
<td>3404 -0.051 -0.19, 0.088</td>
<td>0.12 0.47</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>4631 -0.045 -0.17, 0.077</td>
<td>-0.59 0.47</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>4626 -0.078 -0.20, 0.047</td>
<td>-0.28 0.22</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>4594 -0.054 -0.18, 0.068</td>
<td>-0.67 0.39</td>
</tr>
<tr>
<td>Hip circumference (cm)</td>
<td>4592 -0.036 -0.16, 0.087</td>
<td>-0.37 0.57</td>
</tr>
<tr>
<td>WHR</td>
<td>4591 -0.041 -0.15, 0.073</td>
<td>-0.0040 0.48</td>
</tr>
<tr>
<td>Visceral adipose tissue (cm)</td>
<td>2693 0.0020 -0.15, 0.15</td>
<td>0.031 0.98</td>
</tr>
<tr>
<td>Subcutaneous adipose tissue (cm)</td>
<td>2683 -0.090 -0.24, 0.065</td>
<td>-0.16 0.26</td>
</tr>
<tr>
<td>Fasting serum total cholesterol (mmol/l)</td>
<td>4517 -0.012 -0.13, 0.11</td>
<td>-0.0080 0.84</td>
</tr>
<tr>
<td>Fasting serum HDL-cholesterol (mmol/l)</td>
<td>4652 0.011 -0.11, 0.13</td>
<td>0.028 0.86</td>
</tr>
<tr>
<td>Fasting serum LDL-cholesterol (mmol/l)</td>
<td>3957 -0.014 -0.14, 0.11</td>
<td>-0.018 0.83</td>
</tr>
<tr>
<td>Fasting serum triacylglycerol (mmol/l)</td>
<td>4124 -0.034 -0.16, 0.095</td>
<td>0.0060 0.61</td>
</tr>
</tbody>
</table>

Analyses were performed using a recessive genetic model. βSD is the effect size estimated from quantile-transformed values of the trait, and β is the effect size estimated from untransformed values. Values of p were obtained from the quantile-transformation based analyses. NA, not available.

studies of phenotypes reflecting liver function in Greenlanders would be of great interest to elucidate the biological mechanisms underlying the association between variation in ITGA1 and type 2 diabetes. In eQTL analyses, no effect of the type 2 diabetes-associated rs870992 variant on mRNA expression in liver tissue was observed. Instead, the variant was associated with increased ITGA1 expression in three other tissues; however, this was in the opposite direction to what was expected from the Itgal knockout mice. This difference might rely on the fact that rs870992 is unlikely to be the causal variant and therefore we cannot conclude that the causal variant acts through RNA regulation even if our top SNP is significantly associated with ITGA1 expression.

No previous studies have indicated a link between LARGE1 and type 2 diabetes in additive model GWAS [8], nor did we replicate our findings in recessive analysis in a Danish sample of 5220 individuals with diabetes and 18,556 control participants. This lack of association in European samples may rely on low statistical power for recessive effects, due to the low frequency of risk alleles (EAF<10%) in Europeans (ESM Fig. 2) or due to population-specific differences in linkage disequilibrium between the identified variant and the causal variant. It is also possible that the causal variant behind the observed type 2 diabetes association is Inuit-specific, similar to the risk variant in TBC1D4 [13]. However, we were unable to pinpoint such a causal variant. Moreover, the lack of replication of the diabetes-related quantitative traits associations in a small sample of Alaska Native Yup’ik illustrates one of the main challenges with studies of isolated populations, namely the difficulty of finding appropriate replication cohorts in terms of both genetic composition and sample size. It is, however, also possible that the LARGE1 association represents a false-positive discovery.

LARGE1 encodes the LARGE xylosyl- and glucuronyl-transferase 1 protein. This protein interacts with α-dystroglycan and functions as a glycosyltransferase stimulating glycosylation of α-dystroglycan, which in skeletal muscles connects the cytoplasm with the extracellular matrix as part of the dystrophin-glycoprotein complex [49, 50]. How this relates to the observed association with visceral fat accumulation and insulin resistance is unclear; therefore studies including a more comprehensive coverage of the genome and a larger sample size are warranted to verify the suggested associations, and possibly identify the causal variant in the locus.

In this study, we observed three loci, TBC1D4, ITGA1 and LARGE1, harbouring variants with large recessive effects on the risk of type 2 diabetes in Greenlanders, the latter, however,
having weaker statistical support. In addition to these loci, based on a loss-of-function screening of the same population, we recently identified a variant in ADCY3 recessively associated with obesity and type 2 diabetes [51]. Variants in all four loci have effect sizes that are much greater than what has been reported for TCF7L2, which is the greatest genetic risk factor for type 2 diabetes among Europeans [1]. Even though the estimated effects are likely to be inflated by the ‘winner’s curse’, these variants have the potential to be of clinical relevance by facilitating the prediction of diabetes development in Greenlandic population. More research is needed to identify the causal variants, to elucidate the biological mechanisms underlying the association with type 2 diabetes, and to clarify whether these novel variants distinguish type 2 diabetes subtypes with specific aetiology, as demonstrated for TBC1D4 [13]. This additional knowledge could have the potential to guide choice of treatment for subtypes of type 2 diabetes.

In conclusion, we demonstrate that common alleles with recessive effects play a role in the genetic architecture of type 2 diabetes in the small and historically isolated Greenlandic population. These findings reiterate the importance of considering non-additive genetic models when performing GWAS.

Acknowledgements We gratefully acknowledge the participants in the Greenlandic health surveys. From Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark, we wish to thank A. Forman, T. H. Lorentzen and G. J. Klavsen for laboratory assistance, P. Sandbeck for data management, G. Lademann for secretarial support, and T. F. Toldsted for grant management.

Data availability The Greenlandic MetaboChip-genotype data are available at European Genome-Phenome Archive (EGA; https://ega-archive.org/) under the accession EGAS00001026241.

Funding 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 (www.metabol.ku.dk). This project was also funded by the Danish Council for Independent Research (DFF-4090-00244, Sapere Aude grant DFF-11-120909 and DFF-4181-00383), the Steno Diabetes Center Copenhagen (www.steno.dk), the Lundbeck Foundation (R215–2015–4174) and the Novo Nordisk Foundation (NNF15OC0017918, NNF16OC019986 and NNF15CC0018486). The Greenlandic health surveys (IHIT and B99) were supported by Karen Elise Jensen’s Foundation, the Department of Health in Greenland, NunaFonden, Medical Research Council of Denmark, Medical Research Council of Greenland and Commission for Scientific Research in Greenland. The CANHR studies involving Alaska Native Yup’ik people were funded by the following National Institutes of Health grants: P30 GM103325, R01 DK104347 and R01 DK074842.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript.

Contribution statement TH and AA conceived and headed the project. IM and AA designed the statistical set-up, while TH, NG, EJ, and AA designed the experiment-set-up for the extraction, genotyping and sequencing of the DNA. MEJ and PB were principal investigators of the population studies in Greenland, and together with CVLL and IKD-P, they provided the Greenlandic samples, collected and defined the phenotypes and provided context of these samples. AL planned, recruited and acquired the Danish samples for replication. BBB and SEH led the initial study design for the Yup’ik studies and data collection. KHT and HWW led the statistical analyses for the Yup’ik studies. BB was the principal investigator of the population studies in Alaska and, together with SEH, provided the Yup’ik samples, collected and defined the phenotypes and provided the context of these samples. IM, NG, EJ, and AA performed the association analyses. NG, IM, MKA, AA and TH wrote the majority of the manuscript with input from all authors. All authors approved the final version of the manuscript. NG and TH are the guarantors of this work.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

11. Wood AR, Tyrrell J, Beaumont R et al (2016) Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. Diabetologia 59:1214–1221