A novel splice-affecting *HNF1A* variant with large population impact on diabetes in Greenland

Anne Cathrine Baun Thuesen,[a,b,q] Frederik Filip Stæger,[c,q] Alba Kaci,[d,e,f] Marie Holm Solheim,[d,g] Ingvild Aukrust,[d,h] Emil Jørsboe,[a,i,j] Cindy G. Santander,[c] Mette K. Andersen,[a] Zilong Li,[c] Arthur Gilly,[k] Sara Elizabeth Stinson,[a] Anette Prior Gjesing,[a] Peter Bjerregaard,[l] Michael Lynge Pedersen,[m,n] Christina Viskum Lytken Larsen,[l,n] Niels Grarup,[a] Marit E. Jørgensen,[b,l,m] Eleftheria Zeggini,[k,o] Lise Bjørkhaug,[p] Pål Rasmus Njølstad,[d] Anders Albrechtsen,[c,***] Ida Moltke,[c,**] and Torben Hansen[a,*]

[a]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[b]Clinical Research, Copenhagen University Hospital – Steno Diabetes Center Copenhagen, Herlev, Denmark
[c]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark
[d]Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway
[e]Department of Pediatrics and Adolescents, Haukeland University Hospital, Bergen, Norway
[f]Center for Laboratory Medicine, Østfold Hospital Trust, Kalnes, Norway
[g]Department of Neuronal Control of Metabolism, Max Planck Institute for Metabolism Research, Cologne, Germany
[h]Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway
[i]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, United Kingdom
[j]Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
[k]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[l]Centre for Public Health in Greenland, National Institute of Public Health, University of Southern Denmark, Denmark
[m]Steno Diabetes Center Greenland, Nuuk, Greenland
[n]Greenland Center for Health Research, Institute for Health and Nature, University of Greenland, Nuuk, Greenland
[o]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany
[p]Department of Safety, Chemistry and Biomedical Laboratory Science, Western Norway University of Applied Sciences, Norway

Summary

Background The genetic disease architecture of Inuit includes a large number of common high-impact variants. Identification of such variants contributes to our understanding of the genetic aetiology of diseases and improves global equity in genomic personalised medicine. We aimed to identify and characterise novel variants in genes associated with Maturity Onset Diabetes of the Young (MODY) in the Greenlandic population.

Methods Using combined data from Greenlandic population cohorts of 4497 individuals, including 448 whole genome sequenced individuals, we screened 14 known MODY genes for previously identified and novel variants. We functionally characterised an identified novel variant and assessed its association with diabetes prevalence and cardiometabolic traits and population impact.

Findings We identified a novel variant in the known MODY gene *HNF1A* with an allele frequency of 1.9% in the Greenlandic Inuit and absent elsewhere. Functional assays indicate that it prevents normal splicing of the gene. The variant caused lower 30-min insulin (β = −232 pmol/L, β_{SD} = −0.695, P = 4.43×10^{-4}) and higher 30-min glucose (β = 1.20 mmol/L, β_{SD} = 0.441, P = 0.0271) during an oral glucose tolerance test. Furthermore, the variant was associated with type 2 diabetes (OR 4.35, P = 7.24×10^{-6}) and HbA1c (β = 0.113 HbA1c%, β_{SD} = 0.205, P = 7.84×10^{-3}). The variant explained 2.5% of diabetes variance in Greenland.

Interpretation The reported variant has the largest population impact of any previously reported variant within a MODY gene. Together with the recessive *TBC1D4* variant, we show that close to 1 in 5 cases of diabetes (18%) in Greenland are associated with high-impact genetic variants compared to 1–3% in large populations.

*Corresponding author. Blegdamsvej 3B, 07-8-26, DK-2200 København N, Denmark.
**Corresponding author. Ole Maaløes Vej 5, DK-2200 København N, Denmark.
***Corresponding author. Ole Maaløes Vej 5, DK-2200 København N, Denmark.
E-mail addresses: aalbrechtsen@bio.ku.dk (A. Albrechtsen), ida@binf.ku.dk (I. Moltke), torben.hansen@sund.ku.dk (T. Hansen).
[q]These authors contributed equally.

Greenlandic Summary

Qulequtaq *HNF1A*-mi pinngoqqaammi allannguut suussusersineqarlaaq, pinngoqqaatinik allannguinermut sunniuteqartoq, Kalaallit Nunaanni diabetesimik nappaateqarnermut annertuumik sunniutilik.

Tunuliaqutaasoq Nagguweqatigiit Inuit akornanni nappaatit pinngoqqaatikkut katitigaanerat, pinngoqqaatikkut allannguutaagajuttunik amerlaqisunik nappaateqalersinnaanermut annertuumik sunniuteqartunik akoqarput. Allannguutinik taamaattunik suussusersineq, pinngoqqaatinik nappaatinut patsisaasunik paasinnititseqataavoq, nunarsuarmilu nakorsaanermut nappatinillu suussusersiniaanermut atatillugu pinngoqqaatinik kingornuttakkanik misissuinermi naliliinnerulersitsilluni. Misissuinermi matumani, pinngoqqaatini allannguutinik, diabetesimik pinngoqqaammi akornuteqarnermik patsiseqartumi Maturity Onset Diabetes of the Young (MODY), Kalaallit Nunaanni innuttaasuni nalinginnaanerpaamit ilisimaneqartunik suussusersinissaq naliliinissarlu siunertaavoq.

Periuseq Innuttaasut katillugit 4497-it taakkunannga 448-it, pinngoqqaataat kingornuttakkat, pinngoqqaateqarfik tamakkerlugu misissukkat, kalaallit peqqissusaannik misissuisitsinernit paasissutissanik ataqatigiissitanik atuinikkut, pinngoqqaatit MODY-it ilisimaneqartut 14-it, ilisimaneqartunik ilisimaneqanngitsunillu allannguuteqarnersut misissorpagut. Allannguutip suussusersineqarlaap qanoq ittuuneranik nassuiaavugut, diabetisillu atugaaneranut uummalluuteqarnermullu inooriaatsimik patsiseqartumut ilisarnaatit innuttaasunilu diabetesimut atassutaa misissoqqissaarlugit.

Paasisaq Nagguweqatigiinni kalaallini pinngoqqaammi allannguut ilisimaneqartumi MODY- mi *HNF1A* maannamut suussusersineqanngitsoq, pinngoqqaammi allannguut 1.9%-imik akulikissusilik, inuiannilu allani nassassaanngitsoq suussusersivarput. Pinngoqqaammi allannguutip pinngoqqaatip pissusissamisut allangornissaanik akornusiinera qanoq ittuuneranik misissuinerit paasinarsisippaat. Sukkumik arrortitsisinnaanermut misissuinermi pinngoqqaammi allannguutip 30-min insulinip appasinnerusarnera (β = −232 pmol/L, $β_{SD}$ = −0.695, P = 4.43 × 10⁻⁴) 30-min glucose-llu qaffasinnerunera (β = 1.20 mmol/L, $β_{SD}$ = 0.441, P = 0.0271) nassataraa. Allannguut type 2 diabetesimut (OR 4.35, P = 7.24 × 10⁻⁶) kiisalu HbA1c (β = 0.113 HbA1c%, $β_{SD}$ = 0.205, P = 7.84 × 10⁻³) atassuteqarpoq. Allannguutip Kalaallit Nunaanni diabetesimut allannguutip 2.5%-ia nassuiarpaa.

Nassuiaat Pinngoqqaammi allannguut nalunaarutigineqartoq pinngoqqaammi MODY-mi allannguummit siornatigut nassuiarneqartumit innuttaasunut annertunerusumik sunniuteqarpoq. Pinngoqqaammi allannguut kinguaanit kingornunneqarsinnaasoq *TBC1D4* ilanngullugu, Kalaallit Nunaanni diabetesimik nappaateqartut tallimaappata ataatsip pallingajattup (18%)- innuttaasunut amerlasuunut 1-3%-inut sanilliullugit-pinngoqqaammi allannguutinut nappaateqalernissamut annertuumik sunniutilinnut atassuteqarnera takutipparput.

Aningaasaliisut Novo Nordisk Fonden, Danmarks Frie Forskningsfond kiisalu Karen Elise Jensens Fond."

Introduction

Diabetes is a global health burden that disproportionately affects certain population groups.[1] Identifying population-specific genetic susceptibility is necessary to recognize diabetes subtypes that may have implications for prevention, diagnosis, and clinical care, thereby improving equity in genomic medicine.

Population-specific genetic variants have been identified, both in known and novel monogenic diabetes genes. In Mexican and US Latino individuals, the E508K variant of the known Maturity Onset Diabetes of the Young (MODY) gene *HNF1A* is associated with an increased risk (OR = 5.5) of type 2 diabetes (T2D).[2] Analysis of diverse populations revealed that this variant is very rare or absent in other populations and thus unique to individuals with Native American ancestry. In Oji-Cree individuals, the *HNF1A* G319S variant is associated with an increased risk of diabetes (OR 4.00 for homozygous carriers and OR 1.97 for heterozygous carriers)[3] and earlier age of onset.[4] Whole genome sequencing data from Pima Indians identified the R1420H variant of the known neonatal diabetes gene *ABCC8*, which was associated with a moderately increased risk (OR = 2.0) of diabetes, earlier age of diagnosis, and lower BMI.[5]

Genetic studies in diverse populations have advanced the discovery of novel disease-associated genes and loci. A novel haplotype in *SLC16A11* has been reported,

Research in context

Evidence before this study
The risk of developing diabetes differs between populations, and distinct genetic compositions contribute to the population-specific diabetes aetiology. Studies of individuals of European ancestry have primarily identified either common, low-impact variants that contribute moderately to diabetes risk, or extremely rare variants that cause monogenic diabetes. In contrast, studies of other diverse populations have identified variants within genes implicated in monogenic diabetes that are more frequent and have larger effect sizes that contribute to population-specific risk, such as the Mexican/US Latino E508K *HNF1A* variant, which is associated with a high diabetes risk. In particular, small and historically isolated populations such as the Greenlandic are more likely to harbour deleterious variants at high frequency. Studies of these populations are a powerful tool to address the genetic contribution to diabetes development as the group of deleterious variant carriers is larger and more homogenous, which simplifies functional inference.

Added value of this study
This study reports the finding of a novel *HNF1A* variant which is unique to the Greenlandic population. The variant is associated with an increased risk of diabetes and is found in 6% of diabetes patients. Variant-carriers showed decreased beta-cell function, but had no indications of altered insulin sensitivity. Due to the combination of a large effect and high frequency in the population the variant has the largest population impact of any previously identified variant in a known MODY gene.

Implications of all the available evidence
Combined with the previously described Greenland-specific variant in *TBC1D4*, close to 1/5 diabetes patients (18%) in Greenland have a disease aetiology associated with high-impact genetic variants, compared to 1–3% in large populations. This knowledge provides new avenues to subgroup patients, detect diabetes in family members, and pursue precision treatment trials. In a broader societal context, the finding also has implications for the organisation of healthcare in Greenland, where a focus on accessibility to a genetic diagnosis for broad disease groups is important in order to ensure that every patient is given the right diagnosis and best possible care. Beyond Greenland, the findings in this study underline the importance of performing genetic discovery studies in smaller populations covering all genetic ancestries, to ensure that genetic discovery is conducted equitably. Furthermore, this study challenges the notion that non-communicable diseases are primarily polygenic in aetiology, and emphasises the importance of considering ancestry in the aetiology of diabetes.

which moderately increases diabetes risk and is common among the Mexican and Latin American population, but rare in Europeans.[6] Conversely, population-specific genetic factors may also protect against disease. This is the case for two independent protein-truncating variants in *SLC30A8*, most common among Finnish and Icelandic individuals,[7] and a variant near *CCND2* also identified in Icelandic individuals.[8]

We have previously shown that unique genetic variants contribute substantially to diabetes development in Greenland. A loss-of-function variant in *TBC1D4* has a recessive odds ratio of ten for T2D, and leads to decreased protein expression of the GLUT4 glucose transporter and consequently postprandial muscle insulin resistance.[9] The variant is population-specific to Inuit with an allele frequency of 17% and accounts for more than 10% of all diabetes cases in Greenland. Additionally, a low frequency splice-site variant in *ADCY3* substantially increases the risk of T2D and overweight among Greenlanders,[10] and two common variants in *LARGE1* and *ITGA1* are associated with T2D with moderate effect sizes.[11] These examples show that the contribution of high-impact genetic effects related to diabetes is high in Greenland compared to larger outbred populations, where the common loci most strongly associated with diabetes, such as *TCF7L2*, only infer a moderately increased risk.[12,13]

Motivated by the unique genetic composition of the Greenlandic population, we used whole-genome sequencing data from 448 Greenlandic individuals to screen for population-specific variants in known MODY genes, in order to investigate to which degree variation within these genes contribute to the genetic burden of diabetes in the Greenlandic population. This approach allowed us to identify a novel variant that was not captured by previous genome-wide association studies in the Greenlandic population.[9] We here characterise this variant in terms of its associations to cardiometabolic traits, its functional consequences and its impact on diabetes risk in the Greenlandic population.

Methods
Ethics statement
The studies were approved by the Scientific Ethics Committee in Greenland (project 2011–13 (ref. no. 2011–056978), project 2013–13 (ref. no. 2013–090702), project 2015–22 (ref. no. 2015-16426), and project 2012–16/17 (ref. no. 2017-12997)), and conducted in accordance with the Declaration of Helsinki, second revision.

Study participants
Participants were pooled from two cross-sectional cohorts: B99 (n = 1788, 1998–2001) and Inuit Health in

Transition IHIT (n = 2709, 2005–2010), both collected as part of a general population health survey of the adult (18+ years of age) Greenlandic population based on random population samples.[14,15] For 263 individuals who participated in both studies, phenotypes from B99 were used and a separate cohort label was used. We further included phenotypic data from B2018; a follow-up study of a subset of the B99 and IHIT individuals (n = 1334, 2017–2019[16]).

Anthropometric and biochemical measurements
Anthropometric and biochemical data collection have been described previously.[14,15] HbA1c, serum total cholesterol HDL-cholesterol and triglyceride concentrations were measured at fasting.[14,15] LDL cholesterol was calculated using the Friedewald formula.[17] Plasma glucose, serum insulin, and C-peptide were measured at all timepoints.[14,15]

Participants without medication-treated diabetes >18 years of age (IHIT) or ≥35 years of age underwent an oral glucose tolerance test (OGTT) as previously described,[14,15] sampled at fasting and 2 h (2-h) after glucose ingestion. In the B2018 follow-up study, the OGTT was further sampled at 30 min.[16]

In the OGTT from the IHIT and B99 cohorts,[18] HOMA-IR,[18] insulin sensitivity index $ISI_{0,120}$[19] and HOMA-B[18] were calculated as crude measures of insulin resistance, insulin sensitivity, and beta-cell function. In the B2018 cohort, we calculated the insulinogenic index (IGI)[20] as an expression of beta cell function, and the Matsuda insulin sensitivity index (ISI)[21] as an expression of insulin sensitivity.

For case/control analyses, a diagnosis of T2D was defined as 1) having known T2D prior to the examination or 2) WHO 1999 criteria from the OGTT, whereas controls were normoglycemic (i.e. absence of diabetes, impaired fasting glucose, or impaired glucose tolerance).[22]

Whole genome sequencing
DNA for whole genome sequencing (WGS) and genotyping (see below) was extracted from buffy coat stored at −80°C.

A subset of 448 Greenlandic individuals from IHIT and B99 cohorts, selected based on sampling location independently of phenotype and disease status, underwent Illumina WGS with average sequencing depth of ~35×. Reads were cleared for adapters using biobambam tools[23] and mapped with BWA-MEM[24] to GRCh38 (*bwa mem -t 24 -p -Y -K 100000000*). After mapping, duplicated reads were marked. Genotype calling was done using GATK haplotype caller and variant quality score recalibration (VQSR) tools based on the GATK resource bundle. Only variants in the T98 tranche and above were used. The sites were parsed through Plink (v1.90b6), keeping the two most common alleles of multiallelic sites. The remaining variants were annotated using the Ensembl Variant Effect Predictor release 104.3.[25]

Genetic screening
Fourteen genes in which functionally disruptive variants are known to cause MODY[26] (OMIM entry #606391) were screened in the WGS data. Among these, we extracted variants predicted to have moderate or high impact. Variants that were already reported (found in gnomAD v3.0.0 or dbSNP build 155) and only observed in one individual were removed. The distribution of ClinVar (20210821) MODY variants in the data was analysed in order to assess the presence of known MODY variants. Given the population subset, there was >80% power to detect variants with a minor allele frequency (MAF) >0.2%.

Genotyping
A novel c.1108G>T variant was genotyped in all individuals using the KASP Genotyping Assay (LGC Genomics, Berlin, Germany). The genotyping call rate was 99.8% and there was a concordance rate of >99.9% between genotyping and imputation.

We used a previously described genome-wide genotype dataset[27] generated from the Multi-Ethnic Global Array (MEGA chip, Illumina). After quality control, it comprised 1.6 mil. variants. We used phased 1000 Genomes[28] and the phased Greenlandic WGS as reference for imputing common (MAF > 0.5%) variants. The imputed data was used to estimate a genetic similarity matrix (GSM) and admixture proportions. For local ancestry, we performed additional phasing and imputation which included inferred family relationships to avoid switch errors.

Estimation of allele frequency
To estimate the allele frequency of the c.1108G>T variant in the Greenlandic population, we first estimated the frequency as the fraction of variant copies among the individuals genotyped for the variant. We also estimated the frequency separately for the Inuit ancestry component and for the European component of the population using ADMIXTURE v1.3.0,[29] assuming two ancestral populations (Inuit and European).

To investigate the allele frequency of the c.1108G>T variant in other populations, we looked up frequencies in gnomAD,[30] Human Genetic Diversity Panel[31] (HPDP), Simons Genome Diversity Panel,[32] and 1000 Genome Project.[28] We used samtools,[33] BGT[34] and vcftools[35] to estimate allele frequencies in each population. Finally, we used samtools to estimate allele frequencies from whole genome data from several ancient and contemporary genomes from the Americas and the Arctic.[36–42]

Local ancestry

To visualise the Inuit-specific ancestry of the c.1108G>T variant, we estimated local ancestry for admixed individuals. Individuals with >99% inuit ancestry were used as Inuit reference, and unrelated individuals from the 1000 Genome populations TSI, CEU, GBR, and IBS were used as European reference. Only sites present in both references and admixed individuals with a MAF > 5% across all three groups were used as input to RFMix (v2.03-r0).[43]

Association analyses

To test for associations between the c.1108G>T variant and quantitative traits, we used a linear mixed model to take relatedness and admixture into account (GEMMA[44] v0.95alpha). The GSM was estimated from SNPs with <1% missingness and a MAF >5%. Association tests were performed using a score test. Effect sizes and standard errors were estimated using a restricted maximum likelihood approach. Traits were quantile transformed to a normal distribution independently for men and women using a rank-based inverse normal transformation; $y_{trans} = \phi^{-1}\left(rank(y) - 0.5/n\right)$. P-values are from analyses of transformed data and effect size estimates are reported in standard deviations as β_{SD}. We also estimated effect sizes from non-transformed data, reported as β. To test for association with T2D, we used a logistic mixed model (GMMAT[45]). P-value and effect size (odds ratio) was obtained using Wald test. The GSM was estimated as above. In all analyses, sex, age, and cohort were included as covariates. Given the allele frequency, the power was 88.7% to detect a relative risk of 4 (similar to the estimated OR of 5.5 found for the E508K variant) using a dominant model, a T2D disease prevalence of 10%, and a significance threshold of 0.001.[46]

Transcriptional activity assay and protein abundance

Transcriptional activity was evaluated as previously described.[47] Briefly, the c.1108G>T p.(V370F) variant was introduced into the human *HNF1A* WT cDNA isoform A (NM_000545.6), harbouring the variants c.51C>G, L17, and c.79A>C, I27L, in the pcDNA3.1/HisC vector using the QuikChange II XL Site Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA). Transcriptional activity was measured on cell lysates from HeLa cells transiently transfected with a rat albumin promoter-linked Firefly Luciferase reporter plasmid (pGL3-RA) along with either variants c.1108G>T p.(V370F), P112L,[48] P447L,[48] E508K,[2,49] or wildtype (WT) *HNF1A* plasmid. Renilla Luciferase reporter pRL-SV40 was used as an internal control. Luciferase activity was measured 24-h post-transfection using the Dual-Luciferase Assay System (Promega,

Madison, WI) on a Centro XS3 LB 960 luminometer (Berthold Technologies, Germany). The same cell lysates were used to measure HNF1A protein abundance by SDS-PAGE and immunoblotting (anti-HNF1A from Cell Signaling Technologies, Beverly, MA), normalised against alpha-tubulin-HRP (Abcam, Cambridge, MA). Transcriptional activity and protein expression experiments were performed in triplicate and duplicate respectively, and repeated on three individual days (n = 3).

Minigene assay

To assess the c.1108G>T variant on splicing, exon 6 of *HNF1A* and flanking intron sequences (86bp and 167bp) was amplified from genomic DNA of two carriers (WT and c.1108G>T) and cloned into the pCAS2 splicing minigene vector as previously described.[50] As carriers were heterozygous, variant splicing effect was evaluated by comparing the effect of the corresponding WT sequence of the same individual. Total RNA was isolated from HeLa cells transiently transfected with plasmids (WT or c.1108G>T). Samples were analysed by RT-PCR followed by agarose gel electrophoresis and Sanger sequencing.

Population effect of diabetes variants

To compare population effects, we collected GWAS summary statistics of associated variants from two large studies of 898,130 individuals of European ancestry[12] and 433,540 East Asian individuals.[51] Summary statistics from previously reported population-specific T2D variants were collected from Amerindian,[2,5,6] Icelandic,[8] and Finnish[7] populations. The effect of rare variant burden on T2D in the UK biobank (UKBB) was collected for *HNF1A* and *GCK*.[52] To have comparable estimates for the explained variance, we assumed a population diabetes prevalence of 10% for all populations. For the E508K variant, the allele frequency was calculated separately for individuals with and without diabetes and weighted by diabetes prevalence of 10%. For visualisation of variants with recessive effects, the frequency of homozygous individuals was used instead of the allele frequency. The R-package Mangrove (v.1.21) was used to calculate liability-scale variance explained (LVE) for each variant separately.

Role of the funding source

None of the funding agencies had any role in the study design or collection, analysis, or interpretation of the data.

Results

Screening for known and novel functionally disruptive variants in known MODY genes

We screened fourteen MODY genes (Fig. 1A) for known and novel genetic variants using WGS data from 448
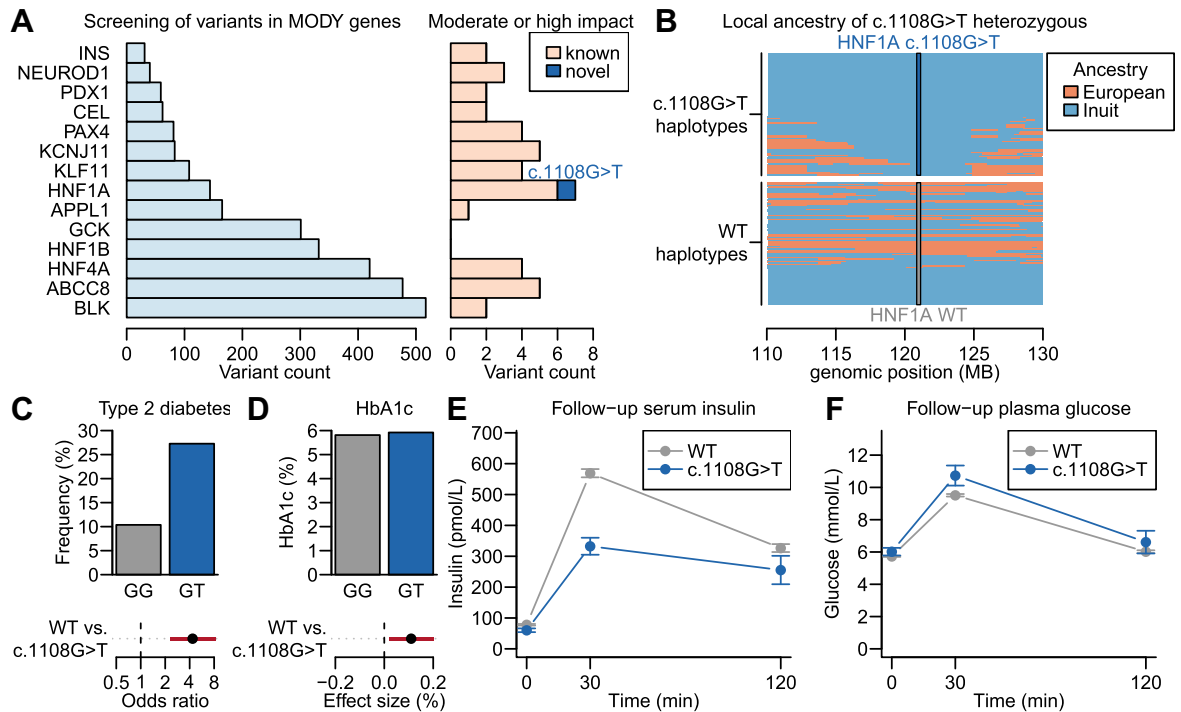
Fig. 1: (A) Screening of variants from whole genome sequencing in MODY genes. Either all variants or only showing moderate or high impact. Novel variants are variants without a dbSNP id or otherwise reported in literature. (B) Local ancestry haplotypes at the c.1108G>T variant region for all 87 admixed and c.1108G>T heterozygous individuals. Each individual is represented with two lines, one for the c.1108G>T haplotype and one for the WT haplotype. The vertical line represents the position of the c.1108G>T variant. (C) Frequency of T2D separated by genotype and effect size (OR) with 95% CI below. (D) Mean value of HbA1c separated by genotype and effect size with 95% CI below. (E) Serum insulin measurements at fasting (0 min), 30-min, and 2-h after glucose ingestion of the follow-up participants. (F) Plasma glucose measurements at fasting (0 min), 30-min, and 2-h after glucose ingestion of the follow-up participants. In E and F errors are mean ± SEM.

Greenlandic individuals (Supplementary Figure S1). First, we explored the presence and frequency of known, ClinVar-annotated variants. We found 91 benign variants at varying allele frequencies, nine variants with conflicting interpretations of pathogenicity at low allele frequencies (MAF<1%) and a single variant annotated as likely pathogenic (Supplementary Table S1). The likely pathogenic variant was present in only one allele in one individual who did not have diabetes. Secondly, we screened for novel functionally disruptive variants in the MODY genes. This screening led to the discovery of a variant, located in exon 6 of *HNF1A*, which was predicted to be a missense variant (hg19: chr 12 pos 121434344, c.1108G>T, p.Val370Phe; V370F, ENST00000257555/NM_000545.8) near a splice-site (Fig. 1A).

Ancestry component analysis and population screening

To investigate the frequency of the c.1108G>T variant, the variant was genotyped in 4497 Greenlanders from the population-based cohorts (Supplementary Figure S1). We identified 119 heterozygous, and no homozygous carriers, leading to an estimated allele frequency in the Greenlandic population of 1.3%. The Greenlandic population is admixed and we estimated the allele frequency to be 1.9% in the Inuit component and 0.0% in the European. We did not find the variant in sequencing data from more than 200,000 ancient and contemporary individuals from across the world (Supplementary Table S2). Local ancestry of all admixed and c.1108G>T-heterozygous Greenlandic individuals (n = 87) showed that the variant was exclusively found on a background of Inuit ancestry, whereas the reference allele (*HNF1A* WT) was found on both European and Inuit ancestry (Fig. 1B). Collectively, this suggests that the variant is common in the Greenlandic population and absent, or extremely rare, in non-Inuit populations.

Associations with T2D and related phenotypes

We tested the association of the variant with T2D and related phenotypes (Table 1). The variant showed strong association with T2D (OR = 4.35, P = 7.24×10^{-6}) and

Anthropometric and metabolic traits (IHIT + B99 cohorts)				
Trait	n	β_{SD} (SE)	β	P-value
BMI (kg/m^2)	4443	0.0850 (0.0957)	0.2926	0.3741
Waist-hip-ratio	4412	0.0715 (0.0874)	0.0051	0.4130
HbA1c (%)	4443	0.2047 (0.0770)	0.1127	7.84×10^{-3}
Fasting serum total cholesterol (mmol/L)	4337	−0.0443 (0.0908)	−0.0662	0.6253
Fasting serum HDL-cholesterol (mmol/L)	4468	−0.0002 (0.0916)	0.001	0.9979
Fasting serum LDL-cholesterol (mmol/L)	3818	−0.0470 (0.0987)	−0.0660	0.6337
Fasting serum triglyceride (mmol/L)	3979	0.0485 (0.1011)	0.0016	0.6314
Diastolic blood pressure (mmHg)	3981	0.0879 (0.0968)	1.4425	0.3638
Systolic blood pressure (mmHg)	3981	−0.0098 (0.0928)	0.5572	0.9162
	Cases/control	β (SE)	OR	P-value
T2D	311/2565	1.47 (0.33)	4.35	7.24×10^{-6}
T2D (no adj. for age & sex)	311/2565	1.15 (0.28)	3.18	5.61×10^{-5}
OGTT data and BCF/IS indices (B2018 (follow-up) cohort)				
Trait	n	β_{SD} (SE)	β	P-value
Fasting plasma glucose (mmol/L)	903	0.1366 (0.1848)	0.2334	0.4592
30-min plasma glucose (mmol/L)	842	0.4409 (0.1991)	1.2048	0.0271
2-h plasma glucose (mmol/L)	838	0.1084 (0.1925)	0.6996	0.5726
Fasting serum insulin (pmol/L)	903	−0.1783 (0.1944)	−15.9627	0.3587
30-min serum insulin (pmol/L)	843	−0.6953 (0.1962)	−232.1417	4.43×10^{-4}
2-h serum insulin (pmol/L)	839	−0.2978 (0.1926)	−40.5706	0.1221
Insulinogenic index	801	−0.8600 (0.1929)	−4.1923	1.15×10^{-5}
Insulin sensitivity index	798	0.2964 (0.1988)	0.9877	0.1359

SE = standard error, n = number of individuals available for the association analysis. The effect size is estimated without transformation (β) or after rank-based inverse normal transformation (β_{SD}).

Table 1: Associations between the c.1108G>T and measures of glucose metabolism, lipid metabolism, and body composition.

higher HbA1c levels (β = 0.11 HbA1c%, β_{SD} = 0.20, P = 7.84×10^{-3}) (Fig. 1C and D). Among the T2D cases, 6.8% were heterozygous for the c.1108G>T variant whereas 2.2% of glucose-tolerant individuals were heterozygous (phenotypes stratified by diabetes status are presented in Supplementary Table S3). Combined with the previously described recessive *TBC1D4* variant,[9] 18% of diabetes cases in the cohort were either heterozygous for the c.1108G>T variant or homozygous for the *TBC1D4* variant, compared to 4% of the glucose-tolerant individuals. The median age of NGT carriers was lower than that of the T2D carriers (NGT median age 38.5 versus T2D median age 55.5).

In the B2018 follow-up study, we further assessed indices of insulin sensitivity and beta cell function (Table 1). Interestingly, c.1108G>T carriers had a much lower 30-min insulin (β = −232 pmol/L, β_{SD} = −0.70, P = 4.43×10^{-4}) and a higher 30-min glucose (β = 1.20 mmol/L, β_{SD} = 0.44, P = 0.027), while the fasting and 2-h timepoints showed no significant differences (Fig. 1E and F). We found strong association with estimated beta cell function IGI (β = −4.19, β_{SD} = −0.860, P = 1.15×10^{-5}) but no association with estimated insulin sensitivity ISI (β = 0.99, β_{SD} = 0.296, P = 0.1359). The lack of association to fasting and 2-h measures was further supported in the OGTT data

from the B99 and IHIT cohorts (Supplementary Table S4).

Functional analyses

The c.1108G>T variant is located within the transactivation domain of *HNF1A* in the first nucleotide of exon 6 (Fig. 2A) and could have functional effects either through the amino acid substitution and/or by affecting splicing. We analysed the potential effects of the amino acid substitution on transcriptional activity and protein abundance, and the effect on splicing in a minigene splice assay.

The c.1108G>T variant showed near-normal levels of activity compared to WT *HNF1A* protein (mean normalized luciferase activity 105.5% ± 4.5) (Fig. 2B). We compared this to the Mexican/US Latino T2D-risk variant E508K, and two known MODY variants (P112L and P447L). The E508K variant exhibited an almost 50% reduction in activity (53.5% ± 7.9) whereas the P112L and P447L variants, showed a severe effect on *HNF1A* activity (16.3% ± 1.8 and 19.6% ± 2.8), similar to what has been described earlier.[49]

When assessing the effect on protein expression, levels were unchanged for the c.1108G>T/V370F variant compared to WT *HNF1A* levels (mean relative protein
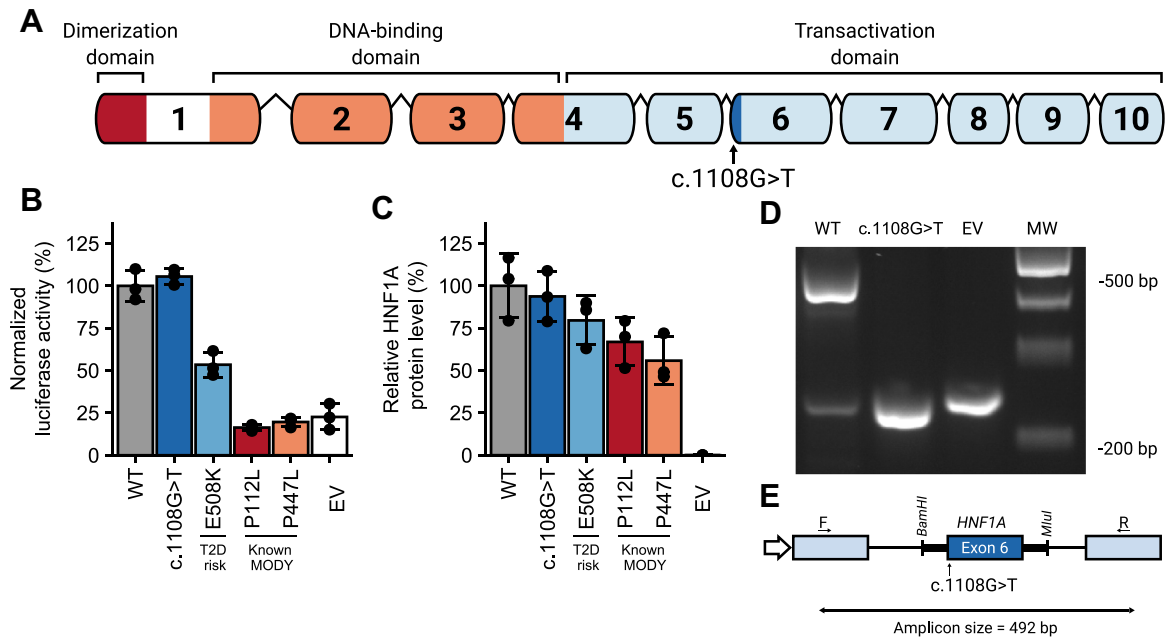
Fig. 2: (A) Structure of the *HNF1A* gene and placement of the c.1108G>T variant. (B) Transcriptional activity of the *HNF1A* c.1108G>T variant relative to wildtype (WT) and control variants E508K, P112L, P447L, and empty vector (EV) in transiently transfected HeLa cells, using a luciferase reporter assay. Each bar represents the mean level of transcriptional activity normalised to WT levels (set to 100%), ±SD (n = 3) (C) c.1108G>T effect on HNF1A protein expression level. Each bar represents the mean level of HNF1A protein expression normalised to WT levels (set to 100%), ±SD (n = 3). (D) RT-PCR fragments produced in the minigene assay, analysed by agarose gel electrophoresis (representative image shown), on the right molecular weight (MW) marker. (E) Schematic design of *HNF1A* exon 6 minigene assay.

expression 93.7% ± 14.9) (Fig. 2C). In comparison, the P447L control variant showed lower protein expression levels (55.8% ± 14.0) while a slight reduction was observed for the E508K and P112L variants.

To address the variant's effect on potential splice events, we performed a minigene splice assay (Fig. 2E). This assay revealed that the c.1108G>T variant allele yielded a transcript around 200bp shorter than the WT allele suggesting that the variant caused skipping of the entire 202bp length of exon 6 within the context of the minigene assay (Fig. 2D). This is suggestive of a splice-affecting variant although the assay cannot determine the exact nature of the full mutant *HNF1A* transcript.

Population impact

In order to contextualise the population impact of c.1108G>T compared to T2D variants identified in other large and small populations, we plotted the odds ratio as a function of the minor allele frequency (Fig. 3A). For the majority of variants with an allele frequency above 1% we observed moderate odds ratios in the range of 0.8 to 1.4. In the European and Asian populations, larger effects were only found for variants with allele frequencies below 1% or in the burden of rare variants in *HNF1A* and *GCK* from UKBB. We also included other population-specific variants (Mexican/ US Latino *HNF1A*, *SLC16A11* haplotype, Pima *ABCC8*, Finnish and Icelandic *SLC30A8*, Icelandic *CCND2*, and previously identified diabetes-associated Greenlandic variants). For these population-specific variants, the observed effects were higher compared to other variants with similar allele frequencies. However, the Greenlandic-specific novel *HNF1A* c.1108G>T and the recessive *TBC1D4* variant had strongly increased odds ratios relative to the allele frequencies. This difference is further accentuated when comparing the population T2D variance explained by each variant (Fig. 3B). Here, *TBC1D4* and *HNF1A* c.1108G>T variants explained 7.3% and 2.5% of the T2D variance in Greenlanders, respectively, whereas the variance of T2D explained by the European *TCF7L2* in Europeans was less than half that (1.1%) and all other variants explained less than 1% of the T2D variance. The G319S Oji-Cree variant[3] was identified in 55 unrelated diabetic and 148 unrelated controls. The study did not take population structure into account and we were unable to obtain corrected effect size estimates and allele frequencies. For this reason the variant was not included in Fig. 3. However, assuming the reported MAF of 9.85% and OR of 1.97 are not biassed by population structure, we estimate the liability-scale variance explained to 2.2%.
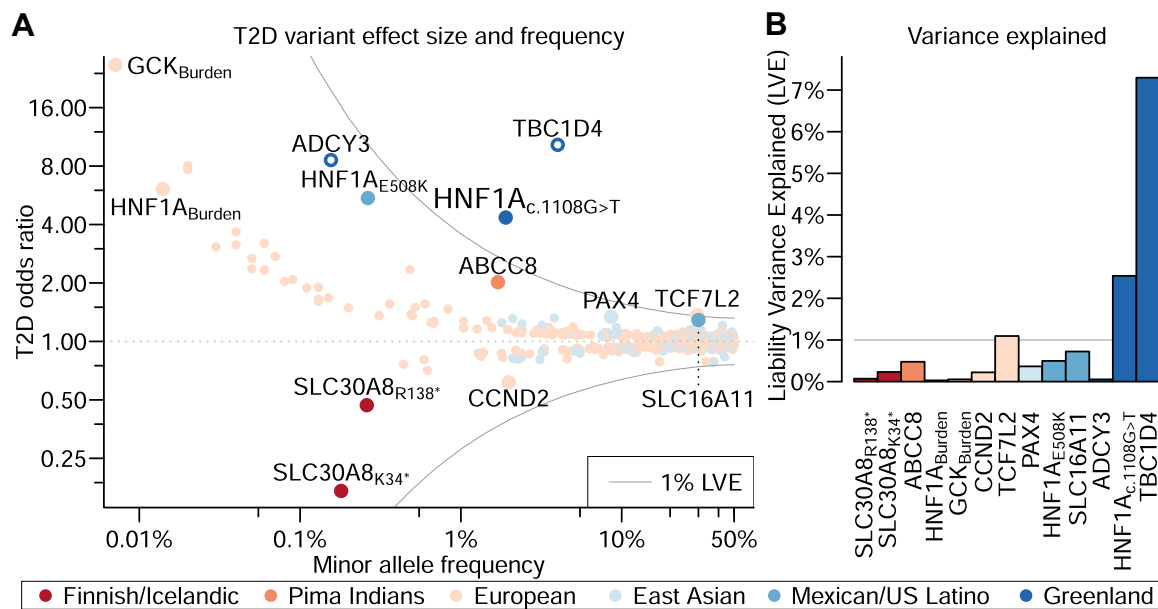
Fig. 3: (A) T2D odds ratio as a function of the minor allele frequency from 898130 European-descent individuals,[12] 433540 East Asian individuals.[51] Selected population variants are highlighted by gene name.[2,5–7,9,10,52] Minor allele frequency for recessive variants is the frequency of homozygous individuals. Gray line indicates the combination of minor allele frequency and odds ratio resulting in a 1% liability variance explained (LVE). Filled circles indicate additive variants while empty circles indicate recessive variants. (B) The liability variance explained for all highlighted variants.

Discussion

In this study, we analysed the genomes of Greenlandic individuals for the presence of variants in known MODY genes in order to address whether common variants in these genes, which have previously been shown to harbour population-specific variants that increase the risk of diabetes, contribute to the burden of T2D in Greenland. We identified a previously unknown variant, c.1108G>T in the known MODY gene HNF1A, which was associated with a highly increased risk (OR = 4.4) of developing T2D and decreased beta-cell function, and which was revealed in ancestry analysis to be unique to the Greenlandic population. We showed that the variant is likely to affect splicing and that it had the largest population-impact of any previously identified variant within a MODY gene and that almost 1 in 5 diabetes cases in this representative population cohort are associated with high-impact genetic variants, compared to 1–3% in European populations.[26]

In the present study, the c.1108G>T variant was associated with significantly lower 30-min insulin levels and higher 30-min glucose levels but not with fasting or 2-h measures of glucose, insulin or C-peptide. The 30-min results from an OGTT are closely associated with first-phase insulin secretion derived from an intravenous glucose load[53,54] unlike the 2-h measures which are affected by both the insulin secretory capacity and peripheral insulin resistance. Accordingly, we observed strong associations between

the variant and the beta cell function index IGI, indicative of a primary beta-cell defect, which is consistent with the known effects of HNF1A variants.[55,56] Patients with HNF1A-MODY have lower insulin levels compared to non-diabetic controls,[57] as do non-diabetic carriers of pathogenic MODY variants compared to WT individuals,[58] and carriers of less disruptive variants in HNF1A, such as the common A98V have lower C-peptide at 30-min in an OGTT,[59] similar to the lower 30-min insulin levels found for the identified c.1108G>T variant.

MODY is characterised by a lack of features associated with T2D, such as insulin resistance, higher body weight or abdominal fat deposition,[60] and further by an altered blood lipid composition.[61] In this study, we did not see any association between the c.1108G>T variant and any measure of lipid metabolism, insulin sensitivity, or body composition. However, screenings of MODY genes in an unselected diabetic population indicate that common clinical features are inconsistent in discriminating between clinical MODY and a T2D-like risk phenotype.[62–64] Similarly, the population-specific HNF1A E508K variant was not associated with any examined clinical characteristic between carriers and non-carriers, including age of diagnosis, BMI, or fasting glucose levels despite the substantial impact on diabetes risk.[2]

Our functional studies indicate that the molecular effects of the c.1108G>T variant are caused by defective splicing, rather than by the isolated effect of the amino

acid substitution. While splice variants are well-established in *HNF1A*-MODY, there are limited *in vitro* functional studies reported on these variants.[48] The most likely molecular mechanism of the identified c.1108G>T variant is haploinsufficiency, consistent with the known pathogenesis of *HNF1A* variants.[65] The variant effect indicated by the functional assays and the variant effect size may seem incongruent. However, the effect of a genetic variant may be modified by other genetic or environmental effects,[66–69] and, as mentioned above, it is challenging to compare effect sizes, penetrance, and defining clinical characteristics between case–control and population studies. The functional assays used in this study are standard in the functional characterization of *HNF1A*, which is only expressed in tissues that are difficult to extract. However, an ideal functional characterization would be based on beta cells cultured from Greenlandic carriers, in order to ensure that the variant is characterised in the relevant tissue and in the unique biological context of Greenlanders. Future studies are necessary to address the exact composition of the mutant transcript, and the mechanism of disease for example through gene-editing or studies of induced pluripotent stem cells.

Importantly, we show that the combination of the large effect size and high frequency of the *HNF1A* c.1108G>T variant constitutes the largest population impact of any variant within a known MODY gene reported to date. Collectively, the previously reported recessive *TBC1D4* variant and the c.1108G>T variant explain the largest amount of variance for single variants by a large margin compared to all other replicable previously reported T2D-associated variants. A possible exception is the G319S variant from the small Sandy Lake Oji-Cree community with an estimated LVE% only slightly lower than c.1108G>T.[3,4] However, a direct comparison is difficult because estimates of allele frequency and effect size for the G319S variant do not take population structure into account. The ancestral Inuit population to the present day Greenlanders is known to have been isolated and extremely small for thousands of years and was therefore strongly affected by genetic drift.[70,71] As a consequence, high impact disease-causing genetic variants segregate at higher frequencies in the Greenlandic population than in larger populations. These results underline that while the genetic component of T2D disease burden within Europe and East Asia is highly polygenic, the Greenlandic population harbours common high penetrance variants, and serves as a reminder that this may be the case for other small populations. In fact, close to 1 in 5 of diabetes cases in Greenland carry either the *HNF1A* c.1108G>T variant or are homozygous for the previously identified Inuit-specific *TBC1D4* stop-gain variant. Therefore the combination of these two common high impact variants alone contribute to a large fraction of the heritability of T2D in Greenland. It also serves as a reminder of the strengths of studies based in small populations, where the number of carriers of the same deleterious variants is higher and the genetic causes of disease are more homogenous.

While individuals with *HNF1A*-MODY generally respond well to treatment with sulphonylureas,[72–74] it is not always the case that variants in *HNF1A* result in an increased insulin secretory response to sulphonylurea. Notably, a study on the short-term response to the sulphonylurea glipizide showed that neither non-diabetic nor diabetic carriers of the Mexican/US Latino *HNF1A* E508K variant were sensitive to sulphonylurea stimulation.[75] Genome-wide meta-analysis also does not implicate any common *HNF1A* variants in the glycemic response to sulphonylurea treatment.[76] Whether carriers of the c.1108G>T variant could benefit from treatment with sulphonylurea should be pursued within the context of a randomised clinical trial establishing both short and long-term efficacy of sulphonylurea in these patients. Beyond potential pharmacogenetic benefits lie further personalised medicine potentials for diagnosis and management that take into account the individual's ancestral, genetic, and environmental background, and underlines the importance of a genetically-driven approach to diabetes management in Greenland. This includes tailored early detection approaches in family members, and a more precise substratification of diabetes within the population that provides an avenue for directed health care efforts in the future. These potentials are in line with the multifaceted areas defined within precision medicine, including precision diagnosis, prognostics, as well as treatment.[77]

In conclusion, the novel *HNF1A* c.1108G>T is unique to Greenlandic Inuit and has the largest population T2D impact of previously reported variants in known MODY genes to date. This finding emphasises that the Greenlandic population is uniquely impacted by common high-penetrance variants, as the previously identified *TBC1D4* variant and the newly identified c.1108G>T variant affect 18% of Greenlanders with diabetes, and underlines the importance of population-specific genetic discovery as a driver of genomic-based personalised medicine.

- Validation: ACBT, FFS, AK, MHS, IA, MKA, AMPG, NG, MEJ, LB, PRN, AA, IM, TH
- Visualization: ACBT, FFS, AK, MHS, IA, AA, IM
- Writing – original draft: ACBT, FFS, MHS, AA, IM, TH
- Writing – review & editing: ACBT, FFS, AK, MHS, IA, EJ, CGS, MKA, ZL, AG, SES, AMPG, PB, MLP, CVLL, NG, MEJ, EZ, LB, PRN, AA, IM, TH

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanepe.2022.100529.

References
1 Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract*. 2010;87:4–14.
2 The SIGMA Type 2 Diabetes Consortium, Estrada K, Aukrust I, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA*. 2014;311:2305–2314.
3 Hegele RA, Cao H, Harris SB, Hanley AJG, Zinman B. The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. *J Clin Endocrinol Metab*. 1999;84:6.
4 Triggs-Raine BL, Kirkpatrick RD, Kelly SL, et al. HNF-1α G319S, a transactivation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. *Proc Natl Acad Sci U S A*. 2002;99:4614–4619.
5 Baier LJ, Muller YL, Remedi MS, et al. ABCC8 R1420H loss-of-function variant in a Southwest American Indian community: association with increased birth weight and doubled risk of type 2 diabetes. *Diabetes*. 2015;64:4322–4332.
6 SIGMA Type 2 Diabetes Consortium, Williams AL, Jacobs SBR, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*. 2014;506:97–101.
7 Flannick J, Thorleifsson G, Beer NL, et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet*. 2014;46:357–363.
8 Steinthorsdottir V, Thorleifsson G, Sulem P, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet*. 2014;46: 294–298.
9 Moltke I, Grarup N, Jørgensen ME, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature*. 2014;512:190–193.
10 Grarup N, Moltke I, Andersen MK, et al. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet*. 2018;50:172–174.
11 Grarup N, Moltke I, Andersen MK, et al. Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population. *Diabetologia*. 2018;61:2005–2015.
12 Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50:1505–1513.
13 Vujkovic M, Keaton JM, Lynch JA, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52:680–691.
14 Bjerregaard P, Curtis T, Borch-Johnsen K, 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*. 2003;62:3–79.
15 Bjerregaard P. *Inuit health in transition Greenland survey 2005-2010*. 2nd ed. National Institute of Public Health; 2011.
16 Larsen CVL, Hansen CB, Ingemann C, et al. Befolkningsundersøgelsen i Grønland 2018 – Levevilkår, livsstil og helbred. National Institute of Public Health. https://www.sdu.dk/da/sif/rapporter/2019/befolkningsundersoegelsen_i_groenland; 2019.
17 Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18:499–502.
18 Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28:412–419.
19 Gutt M, Davis CL, Spitzer SB, et al. Validation of the insulin sensitivity index (ISI0,120): comparison with other measures. *Diabetes Res Clin Pract*. 2000;47:177–184.
20 Seltzer HS, Allen EW, Herron Jr AL, Brennan MT. Insulin secretion in response to glycemic stimulus: relation of delayed initial release to carbohydrate intolerance in mild diabetes mellitus. *J Clin Invest*. 1967;46:323–335.
21 Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care*. 1999;22:1462–1470.
22 WHO Consultation. Definition, diagnosis and classification of diabetes mellitus and its complications. 1999.
23 Tischler G, Leonard S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code Biol Med*. 2014;9:13.
24 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/abs/1303.3997; 2013.
25 McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. *Genome Biol*. 2016;17:122.
26 Naylor R, Knight Johnson A, del Gaudio D. Maturity-onset diabetes of the young overview. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews®*. Seattle (WA): University of Washington, Seattle; 2018.
27 Waples RK, Hauptmann AL, Seiding I, et al. The genetic history of Greenlandic-European contact. *Curr Biol*. 2021;31:2214–2219.e4.
28 Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
29 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–1664.
30 Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–443.
31 Bergström A, McCarthy SA, Hui R, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science*. 2020;367:eaay5012.
32 Mallick S, Li H, Lipson M, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. 2016;538:201–206.
33 Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
34 Li H. BGT: efficient and flexible genotype query across many samples. *Bioinformatics*. 2016;32:590–592.
35 Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–2158.
36 Posth C, Nakatsuka N, Lazaridis I, et al. Reconstructing the deep population history of Central and South America. *Cell*. 2018;175:1185–1197.e22.

37 Rasmussen M, Li Y, Lindgreen S, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. 2010;463:757–762.

38 Raghavan M, DeGiorgio M, Albrechtsen A, et al. The genetic pre-history of the new world Arctic. *Science*. 2014;345:1255832.

39 Wong EHM, Khrunin A, Nichols L, et al. Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res*. 2017;27:1–14.

40 de Barros Damgaard P, Martiniano R, Kamm J, et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*. 2018;360:eaar7711.

41 Sikora M, Pitulko VV, Sousa VC, et al. The population history of northeastern Siberia since the Pleistocene. *Nature*. 2019;570:182–188.

42 Wall JD, Stawiski EW, Ratan A, et al. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019;576:106–111.

43 Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet*. 2013;93:278–288.

44 Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;44:821–824.

45 Chen H, Wang C, Conomos MP, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet*. 2016;98:653–666.

46 Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006;38:209–213.

47 Kaci A, Keindl M, Solheim MH, Njølstad PR, Bjørkhaug L, Aukrust I. The E3 SUMO ligase PIASγ is a novel interaction partner regulating the activity of diabetes associated hepatocyte nuclear factor-1α. *Sci Rep*. 2018;8:12780.

48 Colclough K, Bellanne-Chantelot C, Saint-Martin C, Flanagan SE, Ellard S. Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia. *Hum Mutat*. 2013;34:669–685.

49 Najmi LA, Aukrust I, Flannick J, et al. Functional investigations of *HNF1A* identify rare variants as risk factors for type 2 diabetes in the general population. *Diabetes*. 2017;66:335–346.

50 Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. In: Webb M, ed. *Cancer susceptibility: methods and protocols*. Totowa, NJ: Humana Press; 2010:249–257.

51 Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*. 2020;582:240–245.

52 Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599:628–634.

53 Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes*. 2009;58:1212–1221.

54 Henriksen JE, Alford F, Handberg A, et al. Increased glucose effectiveness in normoglycemic but insulin-resistant relatives of patients with non-insulin-dependent diabetes mellitus. A novel compensatory mechanism. *J Clin Invest*. 1994;94:1196–1204.

55 Wang H, Antinozzi PA, Hagenfeldt KA, Maechler P, Wollheim CB. Molecular targets of a human HNF1 alpha mutation responsible for pancreatic beta-cell dysfunction. *EMBO J*. 2000;19:4257–4264.

56 Shih DQ, Screenan S, Munoz KN, et al. Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism. *Diabetes*. 2001;50:2472–2480.

57 Østoft SH, Bagger JI, Hansen T, et al. Incretin effect and glucagon responses to oral and intravenous glucose in patients with maturity-onset diabetes of the young–type 2 and type 3. *Diabetes*. 2014;63:2838–2844.

58 Stride A, Ellard S, Clark P, et al. Beta-cell dysfunction, insulin sensitivity, and glycosuria precede diabetes in hepatocyte nuclear factor-1alpha mutation carriers. *Diabetes Care*. 2005;28:1751–1756.

59 Urhammer SA, Fridberg M, Hansen T, Clausen JO, Pedersen O. A prevalent amino acid polymorphism at codon 98 in the hepatocyte nuclear factor-la gene is associated with reduced serum C-peptide and insulin responses to an oral glucose challenge. *Diabetes*. 1997;46:5.

60 Carroll RW, Murphy R. Monogenic diabetes: a diagnostic algorithm for clinicians. *Genes*. 2013;4:522–535.

61 Pearson ER, Badman MK, Lockwood CR, et al. Contrasting diabetes phenotypes associated with hepatocyte nuclear factor-1 and -1 mutations. *Diabetes Care*. 2004;27:1102–1107.

62 Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536:41–47.

63 Flannick J, Beer NL, Bick AG, et al. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant mendelian form of diabetes. *Nat Genet*. 2013;45:1380–1385.

64 Bonnefond A, Boissel M, Bolze A, et al. Pathogenic variants in actionable MODY genes are associated with type 2 diabetes. *Nature Metabolism*. 2020;2:1126–1134.

65 Ferrer J. A genetic switch in pancreatic beta-cells: implications for differentiation and haploinsufficiency. *Diabetes*. 2002;51:2355–2362.

66 Kettunen JLT, Rantala E, Dwivedi OP, et al. Correction to: a multigenerational study on phenotypic consequences of the most common causal variant of HNF1A-MODY. *Diabetologia*. 2022;65:912.

67 Stride A, Shepherd M, Frayling TM, Bulman MP, Ellard S, Hattersley AT. Intrauterine hyperglycemia is associated with an earlier diagnosis of diabetes in HNF-1a gene mutation carriers. *Diabetes Care*. 2002;25:2287–2291.

68 Kim S-H, Ma X, Klupa T, et al. Genetic modifiers of the age at diagnosis of diabetes (MODY3) in carriers of hepatocyte nuclear factor-1α mutations map to chromosomes 5p15, 9q22, and 14q24. *Diabetes*. 2003;52:2182–2186.

69 Klupa T, Warram JH, Antonellis A, et al. Determinants of the development of diabetes (maturity-onset diabetes of the young-3) in carriers of HNF-1α mutations: evidence for parent-of-origin effect. *Diabetes Care*. 2002;25:2292–2301.

70 Moltke I, Fumagalli M, Korneliussen TS, et al. Uncovering the genetic history of the present-day Greenlandic population. *Am J Hum Genet*. 2015;96:54–69.

71 Pedersen C-ET, Lohmueller KE, Grarup N, et al. The effect of an extreme and prolonged population bottleneck on patterns of deleterious variation: insights from the Greenlandic inuit. *Genetics*. 2017;205:787–801.

72 Bacon S, Kyithar MP, Rizvi SR, et al. Successful maintenance on sulphonylurea therapy and low diabetes complication rates in a HNF1A–MODY cohort. *Diabet Med*. 2016;33:976–984.

73 Shepherd M, Pearson ER, Houghton J, Salt G, Ellard S, Hattersley AT. No deterioration in glycemic control in HNF-1α maturity-onset diabetes of the young following transfer from long-term insulin to sulphonylureas. *Diabetes Care*. 2003;26:3191–3192.

74 Pearson ER, Liddell WG, Shepherd M, Corrall RJ, Hattersley AT. Sensitivity to sulphonylureas in patients with hepatocyte nuclear factor-1α gene mutations: evidence for pharmacogenetics in diabetes. *Diabet Med*. 2000;17:543–545.

75 Martagón AJ, Bello-Chavolla OY, Arellano-Campos O, et al. Mexican carriers of the *HNF1A* p.E508K variant do not experience an enhanced response to sulfonylureas. *Diabetes Care*. 2018;41:1726–1731.

76 Dawed AY, Yee SW, Zhou K, et al. Genome-wide meta-analysis identifies genetic variants associated with glycemic response to sulfonylureas. *Diabetes Care*. 2021;44:2673–2682.

77 Chung WK, Erion K, Florez JC, et al. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*. 2020;43:1617–1635.