



Genetic determinants of glycated hemoglobin levels in the Greenlandic Inuit population

Emil V. R. Appel¹ · Ida Moltke² · Marit E. Jørgensen³ · Peter Bjerregaard^{4,5} · Allan Linneberg^{6,7,8} · Oluf Pedersen¹ · Anders Albrechtsen² · Torben Hansen¹ · Niels Grarup¹

Received: 17 March 2016 / Revised: 12 January 2018 / Accepted: 23 January 2018 / Published online: 26 February 2018
© European Society of Human Genetics 2018

Abstract

We previously showed that a common genetic variant leads to a remarkably increased risk of type 2 diabetes (T2D) in the small and historically isolated Greenlandic population. Motivated by this, we aimed at discovering novel genetic determinants for glycated hemoglobin (HbA_{1C}) and at estimating the effect of known HbA_{1C}-associated loci in the Greenlandic population. We analyzed genotype data from 4049 Greenlanders generated using the Illumina Cardio-MetaboChip. We performed the discovery association analysis by an additive linear mixed model. To estimate the effect of known HbA_{1C}-associated loci, we modeled the effect in the European and Inuit ancestry proportions of the Greenlandic genome (EAPGG and IAPGG, respectively). After correcting for multiple testing, we found no novel significant associations. When we investigated loci known to associate with HbA_{1C} levels, we found that the lead variant in the *GCK* locus associated significantly with HbA_{1C} levels in the IAPGG ($P_{IAPGG} = 4.8 \times 10^{-6}$, $\beta_{IAPGG} = 0.13$ SD). Furthermore, for 10 of 15 known HbA_{1C} loci, the effects in IAPGG were similar to the previously reported effects. Interestingly, the *ANKK1* locus showed a statistically significant ancestral population differential effect, with opposing directions of effect in the two ancestral populations. In conclusion, we found only 1 of the 15 known HbA_{1C} loci to be significantly associated with HbA_{1C} levels in the IAPGG and that two-thirds of the loci showed similar effects in Inuit as previously found in European and East Asian populations. Our results shed light on the genetic effects across ethnicities.

Introduction

The Greenlandic population comprises about ~57,000 individuals of whom the majority (>80%) has a degree of European ancestry [1]. On average, a Greenlander has 25%

European ancestry [1], thus, the Greenlandic genome has both an Inuit ancestry proportion (Inuit Ancestry Proportion of the Greenlandic Genome (IAPGG)) and a European ancestry proportion (European Ancestry Proportion of the Greenlandic Genome (EAPGG)). The Greenlandic population is closely related, both genetically and culturally, to the Inuit population (Iñupiat) in Canada and Alaska, which in turn is culturally related to the Yup'ik people of Alaska and Siberia [2]. In contrast, the Greenlandic population is genetically far from any large population, has been small for many thousand years,

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0109-3>) contains supplementary material, which is available to authorized users.

✉ Emil V. R. Appel
vincent@sund.ku.dk

- ¹ The Novo Nordisk Foundation Center for Basic Metabolic Research, Section for Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
- ² The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark
- ³ Steno Diabetes Center, 2820 Gentofte, Denmark
- ⁴ National Institute of Public Health, University of Southern

Denmark, 1353 Copenhagen, Denmark

- ⁵ Greenland Centre for Health Research, University of Greenland, Nuuk, Greenland
- ⁶ Research Centre for Prevention and Health, Capital Region of Denmark, Copenhagen, Denmark
- ⁷ Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark
- ⁸ Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

and has been isolated in an inhospitable climate, making this population genetically interesting.

In the past 25 years, the Greenlandic population has experienced a dramatic increase in type 2 diabetes (T2D) prevalence [3–5]. We previously described a common loss-of-function variant in *TBC1D4* with a high isolated impact on postprandial circulating glucose and insulin levels and risk of T2D in the Greenlandic population [6]. In a recessive model, the variant increased the 2-h (h) plasma glucose levels during an oral glucose tolerance test (OGTT) by 4 mmol/l and increased the risk of T2D with an odds ratio of 10.3. The effect sizes shown for the *TBC1D4* variant are several fold higher than previously observed in genome-wide association studies (GWAS). Interestingly, the loss-of-function *TBC1D4* variant decreased fasting plasma glucose level by 0.18 mmol/l⁶.

T2D can be diagnosed from measurement of plasma glucose at fasting and at 2 h during an OGTT or by measurements of glycosylated hemoglobin (HbA_{1C}) levels [7]. The HbA_{1C} level reflects the average level of plasma glucose for the individual during the past 6–12 weeks [8]. From 2011, the World Health Organization (WHO) recommended the use of standardized measurement of HbA_{1C} levels to diagnose T2D with a cut-off point of 6.5% (48 mmol/mol) [7]. HbA_{1C} has also been shown to be a risk factor for cardiovascular diseases, even at non-diabetic levels [9].

GWAS of HbA_{1C} levels in the European and Asian populations with up to 31,000 individuals have identified a total of 18 genomic loci significantly associated with HbA_{1C} in either or both populations [10, 11] (Supplementary Table 1). The majority of these loci do not associate with plasma glucose levels or T2D [10, 11], conversely six loci (*ANKK1*, *GCK*, *G6PC2*, *MTNR1B*, *CDKALI*, and *FADS1-FADS2*) have been shown to associate with fasting glucose levels or T2D [12–14].

Motivated by the highly increased risk of T2D imposed by the common variant in the *TBC1D4* in the Greenlandic population [6], we aimed to discover novel genetic associations with HbA_{1C} in this population. Furthermore, we aimed to investigate genomic loci already known to associate with HbA_{1C} levels in European and East Asian populations by estimating the effect of the reported lead single-nucleotide polymorphisms (SNPs) in the Inuit ancestry proportion of the genomes of the Greenlandic population, IAPGG.

Materials and methods

Study samples

Data originated from two separate cohorts, “Inuit Health in Transition” (IHIT) [3] and B99 [15]. The IHIT cohort ($N = 3115$) was collected as a part of a health survey of the Greenlandic population during 2005–2010. Participants

were invited from 12 regions, consisting of 21 selected towns and villages. The B99 cohort ($N = 1401$) was recruited during 1999–2001 as a part of a general population health survey. There was an overlap of 295 individuals between the two cohorts, who were assigned to the B99 cohort. We removed participants previously diagnosed with T2D ($N = 67$), as well as individuals whose genetic data did not pass quality control (see details below) making the final number of individuals in the meta-analysis 4049 (IHIT: $N = 2730$; B99: $N = 1319$).

In both Greenlandic cohorts, the HbA_{1C} levels were measured using ion exchange high-performance liquid chromatography, in a Na-heparinized (5 μ l) capillary in Ebendorf tube with 1 ml EDTA and potassium cyanide solution (0.25 mmol/l). In the IHIT cohort, this was measured by Tosoh G7 [16], whereas in the B99 cohort the measurement was done using BioRad VARIANTTM [17]. HbA_{1C} was defined as the ratio of HbA_{1C} to the total hemoglobin and measured in accordance to the The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) reference method [18]. Basic information for each cohort is given in Supplementary Table 2.

The studies have received ethical approval from the Commission for Scientific Research in Greenland (project 2011–13, ref. no. 2011–056978 and project 2013–13, ref. no. 2013–090702) and were conducted in accordance with the ethical standards of the Helsinki II Declaration. Participants gave their written consent after being informed about the study both orally and in writing.

We also included the Danish cohort Inter99 in our analysis of ancestral-specific effect to be representative of the European population. The Inter99 cohort ($N = 6127$) is a population-based randomized controlled trial, investigating the effect on cardiovascular diseases, and diabetes of a non-pharmacological intervention [19]. After removing participants with known diabetes ($N = 121$), the number of individuals added to the ancestral-specific effect analysis was 6006, bringing the final number in the ancestral-specific effect analyses to 10,055 individuals.

Genotyping

We genotyped all participants with the Illumina Cardio-Metabochip [20] (MetaboChip) using the HiScan system (Illumina). Illumina GenomeStudio software was used for the genotype calling. The MetaboChip consists of 196,725 SNPs selected for genetic studies on cardiovascular, metabolic, and anthropometric traits.

We removed all SNPs with a call rate below 0.99 before analysis, but kept all other SNPs regardless of minor allele frequency (MAF). Furthermore, we removed SNPs that failed a Hardy–Weinberg equilibrium test with a P -value below 10^{-5} in either of the two cohorts. We removed

duplicate samples and excluded samples with mis-specified sex. Finally, we removed samples that had a call rate below 98% for a subset of SNPs with a MAF above 1%. A total of 426 samples and 84,448 SNPs were removed through these steps leaving 112,277 SNPs in 4049 individuals for analysis.

Loci known to associate with HbA_{1c} level

From the literature, we found 21 SNPs across 18 loci, shown to associate genome-wide significantly with HbA_{1c} levels in European or East Asian populations (Supplementary Table 1) [10, 11]. For three of these loci, two different lead SNPs were reported and for these we chose the SNP with the lowest reported *P*-value. We defined a locus as the ± 240 kilo-base-pairs (kbp) region surrounding the reported lead SNP. For 7 of 18 genomic loci, the reported lead SNP was available in the quality filtered MetaboChip data, whereas we for 8 of the 18 loci analyzed a proxy SNP in high ($R^2 > 0.9$) linkage disequilibrium (LD) in the population in which they were identified (Supplementary Table 3). For the remaining three loci (*TMEM79*, *MYO9B*, and *TMPRSS6*), the lead SNP was unavailable in the quality filtered MetaboChip data and we were unable to find a suitable proxy. Therefore, these three loci were omitted from our analyses. For the sake of clarity, we name these 15 loci as “known HbA_{1c} loci” as a shorthand version for loci known to associate with HbA_{1c} levels in European and/or East Asian populations.

Statistical analysis

The population structure of the Greenlandic population has been shown to be composed of an ancestral Inuit population with recent gene flow from the European population [1, 6] and as the genotyped individuals constitute almost 10% of the adult population, >1000 close relationships (parent–offspring or siblings) were identified [6]. To control the false-positive rate in the discovery analyses, we therefore performed association testing between each SNP and HbA_{1c} levels using an additive linear mixed model as implemented in GEMMA [21]. This model takes the relatedness between individuals, as well as admixture into account by using a general similarity matrix (GSM). We further adjusted the analysis for age and sex by adding these as covariates in the model. As the distribution of HbA_{1c} prior to transformation was not normal in either cohort (Skewness_{IHIT} = 3.70, Kurtosis_{IHIT} = 62.71, Skewness_{B99} = 4.44, Kurtosis_{B99} = 53.93) (see supplementary figure 2 and 3), we transformed the HbA_{1c} measurements to a standard normal distribution using a rank-based inverse normal transformation on the unadjusted variable. We used GEMMA’s maximum likelihood estimator method to

calculate *P*-values. Data from the two cohorts were analyzed separately and the results were subsequently combined in a fixed-effect meta-analysis. We reported the *P*-value for the two-tailed test on the combined effect. The inflation of *P*-values was calculated as $\lambda = \frac{\text{median}(\chi_1^2(1-P))}{\text{median}(\chi_1^2(0.5))}$, where χ_1^2 is the chi-squared distribution with one degree of freedom and *P* is the observed *P*-values in the discovery analysis.

We also looked for novel associations in the known loci by using the same linear mixed model but with the genotype of the reported lead SNP for the locus as an additional covariate (conditional model) in addition to the unconditional model, and reported the results from both models alongside each other.

When comparing the effect of the lead SNPs found in European and East Asian populations with the Greenlandic cohort, the effect of interest for the current study is in the IAPGG. We therefore used asaMap [22], a novel statistical method that can perform association testing on the ancestral parts of an admixed population. First, we combined both Greenlandic cohorts, IHIT and B99, with the Danish Inter99 cohort [19], to strengthen the statistical power of asaMap and ADMIXTURE. Second, we used a rank-based inverse normal transformation separately in each cohort before combining the data. Third, to determine the proportions of Inuit ancestry for each individual in each cohort needed for the asaMap analyses, we used the software program, ADMIXTURE [23] with $k = 2$, implying that we assume two ancestral populations (Inuit and European) in admixed Greenlandic genome. Finally, using asaMap we ran an additive linear regression with age, sex, cohort and the first 10 principal components as covariates to adjust for cohort bias, population, and relatedness structure in lieu of a GSM. We tested three hypotheses: (1) The existence of an ancestral-specific effect (asaMap M1vM4) of the variant on HbA_{1c} levels in the admixed Greenlandic population, (2) the variant’s effect only in IAPGG (asaMap M1vM2: $\beta_{\text{EAPGG}} = 0$), and (3) the variant’s effect only in the Inter99 cohort combined with the European ancestry proportion of the Greenlandic genome (Inter99 + EAPGG) (asaMap M1vM3: $\beta_{\text{IAPGG}} = 0$). The applied models have previously been described in more detail [22]. We assumed that only the minor allele could have ancestry-specific effects, whereas the most frequent allele has the same effect regardless of ancestry. Therefore, the minor allele was always used as the effect allele in all of the asaMap analyses.

The inflation of *P*-values for the ancestral-specific effect model was calculated for all SNPs with a call rate of at least 99% and a MAF above 5% across all three cohorts. The resulting lambda ($\lambda = 0.89$) was calculated as

before $\lambda = \frac{\text{median}(\chi_1^2(1-P))}{\text{median}(\chi_1^2(0.5))}$ (Supplementary Figure 1). We attribute the low lambda to the fact that many SNPs are not polymorphic in both the Inuit and the European ancestral populations.

To be able to compare the effect sizes from our analyses to the effect sizes reported in the literature, we translated the effects of known loci from percentage units to standard deviation (SD) units, as used in our study. For this, we used the 5783 non-diabetic participants of the Danish Inter99 cohort [19] where the SD of HbA_{1C} was 0.40%.

We calculated the statistical significance thresholds using the conservative Bonferroni correction, which was calculated as $P = \frac{0.05}{N_{\text{SNP}}}$, where N_{SNP} is the number of SNPs. We applied different significance thresholds depending on the hypothesis tested, to reflect the different number of test in each model. In the discovery analysis, we used the standard GWAS significance threshold of $P < 5 \times 10^{-8}$. For the analyses of novel additional association signals in known loci were corrected for the total number of SNPs (2612) across all known loci ($P = \frac{0.05}{2,612} = 1.9 \times 10^{-5}$). The threshold chosen for the GEMMA and asaMAP analyses of the lead SNPs were corrected for the number of lead SNPs (15) tested ($P = \frac{0.05}{15} = 3.3 \times 10^{-3}$).

We calculated the statistical power to detect association signals at different significance thresholds given a specific MAF and effect size, by running a linear additive model on a simulated data with a normal distribution with the same variation as the combined HbA_{1C} levels in the two Greenlandic cohorts. We then observed how often the model crossed a given significance level. We tested each of the three thresholds (genome-wide, loci-wide and lead SNP-wide) and for each effect size and MAF, we ran the simulation 100 times. To calculate statistical power for the ancestry-specific analyses, we used the asaMap model

instead of a linear model, and added the real admixture proportions from the combined cohorts of Inter99, B99, and IHIT.

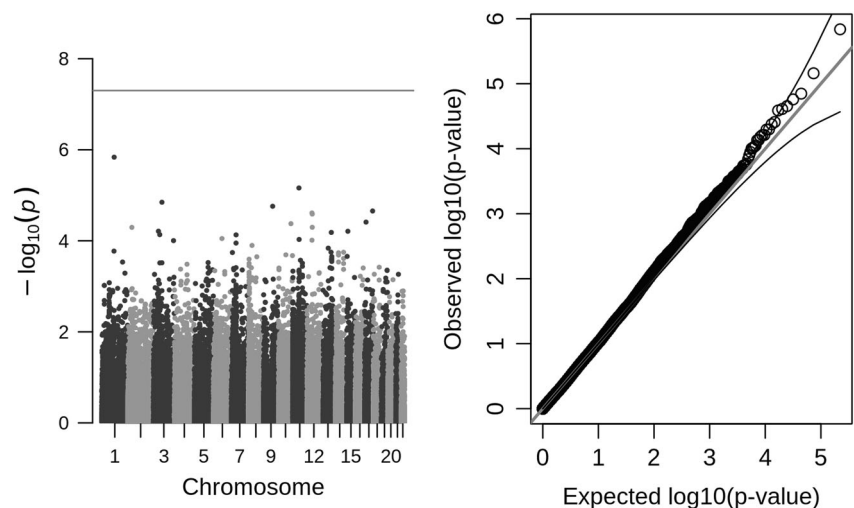
Results

We first performed a discovery analysis, in which we sought to detect novel associations with HbA_{1C} in the Greenlandic population when applying an additive linear mixed model. We tested 112,277 SNPs in 4049 individuals and observed no inflation of the test statistic ($\lambda=0.97$). We found no SNPs reaching statistical genome-wide significance ($P < 5 \times 10^{-8}$) (Fig. 1). The strongest signal was for rs1146347_G ($\beta = -0.145$ SD, $P = 1.5 \times 10^{-6}$), which is located in an intergenic region with ~1 Mbp (mega base pairs) to the nearest genes (*SPAG17* and *TBX15*; up- and downstream, respectively), none of which are known to associate with HbA_{1C} levels or other glycemic or hematological traits. None of the most associated variants was known from the literature to associate with HbA_{1C} levels (Supplementary Table 4) or other glycemic traits. None of these variants was significantly associated with HbA_{1C} levels in the Danish cohort Inter99 consisting of European individuals (data not shown).

Next, we looked for novel association signals in loci known to be associated with HbA_{1C}; however, we did not find any novel statistically loci-wide significant signals in neither in tests conditioned on the lead SNP in the locus nor in unconditional tests (Supplementary Table 5).

Finally, we investigated lead SNPs in known HbA_{1C} loci. When using a linear mixed model, we found that 2 of the 15 lead SNPs associated statistically significantly with HbA_{1C} levels in the Greenlandic cohorts (*GCK* rs730497_A: $\beta = 0.079$ SD, $P = 0.0013$, *MTNR1B* rs1387153_T: $\beta = 0.081$ SD, $P = 0.00036$) (Supplementary Table 6).

Fig. 1 Association analyses of 112,277 markers on the MetaboChip with HbA_{1C} levels in 4,049 Greenlanders. Left: Manhattan plot showing the $-\log_{10}(P)$ for each marker against their position, with each chromosome colored alternatively blue and green. The red line represents the genome-wide significant level ($P < 5 \times 10^{-8}$). Right: QQ plot showing the $-\log_{10}(P)$ from the discovery analysis against the expected distribution. (Color figure online)



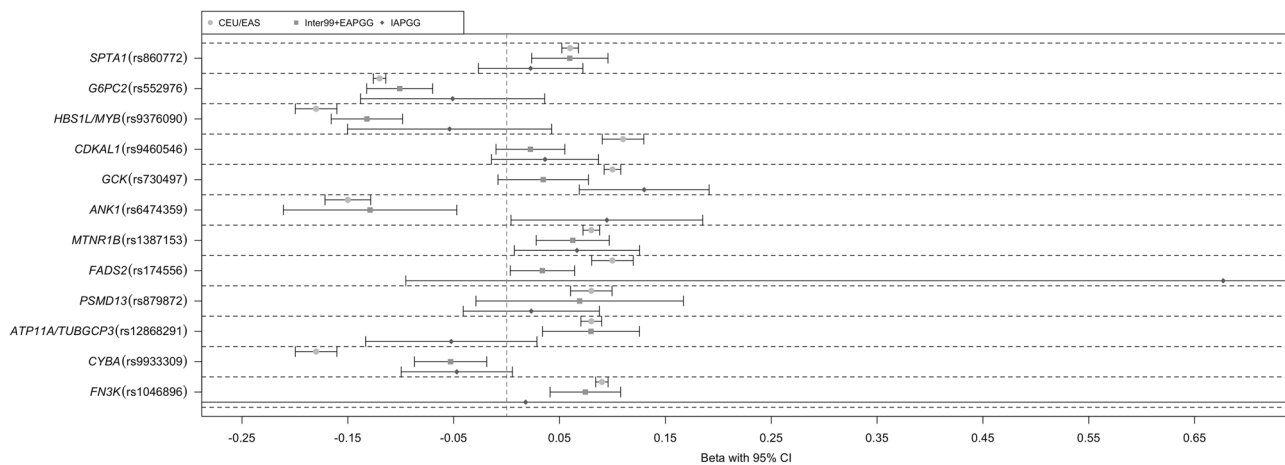


Fig. 2 Comparison of the reported effects of HbA_{1C}-associated lead SNPs with the effects found Inuit proportion of the Greenlandic genome (IAPGG) and the combination of Danish Inter99 cohort and the European proportion of the Greenlandic genome (Inter99 + EAPGG). The plot shows the effect size, with 95% confidence

intervals (CI), found in European (CEU) or East Asian (EAS), IAPGG and Inter99 + EAPGG for each known HbA_{1C} locus. The effect size, beta, was measured in standard deviations (SD). Loci with a SE above one were omitted in this plot for easier viewing. The SNP for each locus is either the reported lead SNP or a proxy on the chip (Table 1)

When further testing for ancestry-specific effects using the *asaMap* software, we found one locus with a statistically significant ancestral-specific effect (*ANK1*: $P_{AnSE} = 0.00051$). Additionally, we found one locus that associated significantly with HbA_{1C} levels in IAPGG (*GCK* rs730497_A: $\beta = 0.13$ SD, $P_{IAPGG} = 4.8 \times 10^{-6}$). In comparison, 9 of the 15 loci associated significantly ($P < 3.3 \times 10^{-3}$) in the European part of the *asaMap* analysis (Inter99 + EAPGG). Five of the six loci that did not associate in Europeans where originally reported in East Asian populations.

When comparing the 95% confidence interval (CI) of the effects reported in European (CEU) or East Asian (EAS) populations with the CI of association in the IAPGG, we observed that for 10 of the 15 known HbA_{1C} loci, the 95% CI of association with HbA_{1C} in IAPGG overlapped the 95% CI reported in CEU or EAS (Fig. 2). Of the five loci with non-overlapping CIs, two loci (*ANK1*, *ATP11A/TUBGCP3*) showed effect in the opposite direction in IAPGG compared with the effect found in CEU, however, only the *ANK1* locus was nominally significant in IAPGG ($P_{IAPGG} = 0.023$) (Fig. 2, Table 1).

We also investigated the lead SNP (rs7330796_T) in *TBC1D4* and its association with HbA_{1C} levels, but it showed no ancestral-specific effect ($P = 0.97$) nor was it significantly associated with HbA_{1C} levels in either IAPGG or EAPGG ($P = 0.7$ and $P = 0.75$, respectively) under an additive model.

Discussion

We investigated the genetic determinants of HbA_{1C} in the small and historically isolated Greenlandic population and

found no novel association signals in the discovery analysis. Two known HbA_{1C}-associated loci, *GCK* and *MTNR1B*, were significantly associated with HbA_{1C} in the Greenlandic population. Furthermore, in analysis of the Inuit ancestral proportion of the Greenlandic genome (IAPGG), we found the *GCK* locus to associate significantly with HbA_{1C} levels. In addition, we found that 10 of 15 known loci showed evidence of comparable effects in IAPGG when compared with the effects found in CEU or EAS populations. In contrast, two loci (*ANK1* and *ATP11A/TUBGCP3*) showed indications of a different effect in IAPGG than reported in the literature (Fig. 2). The most likely explanation for these findings is that the tested variants in these two loci are not the causal variants and LD patterns differ between the two ancestral populations. This is especially an issue when proxy SNPs are used for the analysis. However, it is also possible the multiple association signals are present at these loci (*ANK1* and *ATP11A/TUBGCP3*) although we find no evidence for this in the current data. Lastly, three loci (*HFE*, *HKL*, and *9q31.2*) had very low estimated effect allele frequencies (<0.01%) and are likely monomorphic in IAPGG.

In the discovery analysis, we did not find any statistically significant associations. This analysis was obviously limited by low statistical power (see Supplementary Table 7) for variants with moderate effects and by the non-genome-wide coverage of the MetaboChip. None of the genes in the region of the strongest signal had previously been shown to associate with HbA_{1C} levels or metabolic traits or had any known biological relations to glucose metabolism. The top association signal is located on chromosome 1 between *SPAG17* and *TBX15*. We recently detected a signal for selection in the region of *TBX15* [24], however, the two

Table 1 Lead SNPs in known loci associated with HbA_{1c} and their effect on HbA_{1c} in Inuit and European ancestries

Locus	Greenlanders										IAPGG					Inter99 + EPAGG				
	Lead SNP	EAF	Beta	Pop.	Proxy	N _{Inter99}	N _{Greenland}	Effect allele/Other allele	P(AnSE)	EAF	Beta	SE	P-value	EAF	Beta	SE	P-value			
<i>SPTA1</i>	rs2779116	0.27	0.060	CEU	rs860772	6127	4049	C/G	0.24	0.71	0.023	0.025	0.33	0.24	0.060	0.018	0.0019			
<i>G6PC2</i>	rs552976	0.36	-0.12	CEU	-	6126	4049	A/G	0.26	0.14	-0.051	0.044	0.19	0.36	-0.10	0.016	5.2 × 10 ⁻¹⁰			
<i>HFE</i>	rs1800562	0.060	-0.16	CEU	rs1408272	6123	4049	G/T	0.82	0.00010	0.45	16	0.85	0.057	-0.12	0.032	0.00022			
<i>HBS1L/MYB</i>	rs9399137	0.31	-0.18	EAS	rs9376090	6121	4049	C/T	0.12	0.12	-0.054	0.049	0.21	0.26	-0.13	0.017	3.6 × 10 ⁻¹³			
<i>CDKALI</i>	rs7772603	0.42	0.11	EAS	rs9460546	6125	4049	G/T	0.66	0.50	0.036	0.026	0.13	0.32	0.023	0.017	0.20			
<i>GCK</i>	rs730497	0.19	0.080	CEU	rs730497	6126	4049	A/G	0.013	0.25	0.13	0.031	4.8 × 10 ⁻⁶	0.16	0.035	0.022	0.13			
<i>ANK1</i>	rs6474359	0.030	-0.15	CEU	-	6126	4049	C/T	5.1 × 10 ⁻⁴	0.10	0.10	0.046	0.023	0.036	-0.13	0.042	0.0037			
<i>9q31.2</i>	rs1467311	0.23	0.10	EAS	-	6126	4049	G/A	0.70	0.00010	-1.3	8.6	0.70	0.35	0.034	0.015	0.028			
<i>HK1</i>	rs16926246	0.10	-0.22	CEU	rs10159477	6127	4049	A/G	0.44	0.00010	-2.8	5.6	0.44	0.15	-0.20	0.020	2.1 × 10 ⁻²⁰			
<i>MTNRI1B</i>	rs1387153	0.28	0.080	CEU	-	6125	4049	T/C	0.92	0.27	0.067	0.030	0.017	0.28	0.063	0.018	5.4 × 10 ⁻⁴			
<i>FADS2</i>	rs174570	0.55	0.10	EAS	rs174556	6127	4049	C/T	0.081	0.0047	0.68	0.39	0.065	0.71	0.034	0.016	0.030			
<i>PSMD13</i>	rs540078	0.44	0.080	EAS	rs879872	6127	4049	T/C	0.45	0.18	0.023	0.033	0.44	0.030	0.069	0.050	0.17			
<i>ATP11A/TUBGCP3</i>	rs7998202	0.14	0.080	CEU	rs12868291	6119	4049	C/T	0.0043	0.13	-0.052	0.041	0.15	0.14	0.080	0.023	6.5 × 10 ⁻⁴			
<i>CYBA</i>	rs9933309	0.37	-0.18	EAS	-	6125	4049	T/C	0.86	0.82	-0.047	0.027	0.062	0.30	-0.053	0.017	0.0034			
<i>FN3K</i>	rs1046896	0.31	0.090	CEU	-	6126	4049	T/C	0.83	0.0052	0.018	0.50	0.95	0.28	0.074	0.017	1.1 × 10 ⁻⁵			

Each locus from the literature that has been found to associate with HbA_{1c} levels in Europeans (CEU) or East Asian (EAS) populations was analyzed using asaMap, which allowed us to analyze the Inuit proportions in Greenlanders (IAPGG) and the combination of the Danish cohort and the European proportion in Greenlanders (Inter99 + EPAGG) separately. When the lead SNP could not be found on the MetaboChip, a proxy was used instead (see Supplementary Table 3). The effect allele frequencies (EAF) for IAPGG and Inter99 + EPAGG were estimated with the ADMIXTURE software. All of the beta values are in standard deviations. The P-value, P(AnSE), is a test of a difference ancestral-specific effect, whereas the P-values for IAPGG and Inter99 + EPAGG are tests of the specific ancestral effects, that is, is there an effect in the specific ancestral population. All genomic positions and loci are based on the GRCh37 build

SNPs (rs1146347 and rs4659153) are separated by ~800 kb and are not in LD in Greenlanders ($R^2=0.0005$, $D'=0.031$).

In the analyses of the ancestral-specific effect of Inuit and European ancestry in the Greenlandic genome, we found one locus, *GCK*, which associated significantly in IAPGG. *GCK* encodes glucokinase (a member of the hexokinase family) and is primarily expressed in the liver and pancreas. Glucokinase is an enzyme that facilitates phosphorylation of glucose to glucose-6-phosphate, the first step in the glucose metabolism pathway [25]. The association with HbA_{1C} is thus not surprising and with a larger effect allele frequency in IAPGG this locus may have a larger overall impact on the population, although the effect size per allele is similar to what was previously found in European populations. This finding is supported by previous reports showing a significantly higher level of HbA_{1C} for a given level of fasting plasma glucose and 2-h plasma glucose among Greenlanders compared with Danes [26]. In contrast, the lack of a strong association in the European subset of the current study (Inter99 + EAPGG) is surprising. When investigating the *GCK* locus in details in Inter99, we find that the strongest association with HbA_{1C} is carried by the rarer allele rs117316370 ($MAF = 0.02$, $P = 0.0058$, $\beta = -0.21$ SD), which is not in strong LD ($R^2=0.093$) with our chosen proxy SNP (rs730497) that does not associate with HbA_{1C} for Inter99 ($P = 0.29$, $\beta = 0.029$ SD). We speculate that this might be the cause of the surprising lack of association in the subset Inter99 + EAPGG.

We found that that the *ANK1* locus had the strongest difference in ancestral-specific effects among the examined known HbA_{1C} loci. *ANK1* was identified as an HbA_{1C}-associated locus in Europeans [11], and in IAPGG we find a nominally significant association of *ANK1* rs6474359_C in the opposite direction as reported in Europeans. This may indicate that the identified lead variant is not causal and that LD between the lead and causal variants differs among Inuit and European populations. Of interest, an opposite direction of effect than in Europeans of rs6474359_C was also observed in a study of the East Asian population [10]. The metaboChip does not cover the *ANK1* locus densely as only four additional SNPs within 240 kb of lead SNP rs6474359 were available for analysis. None of these had a significant ancestral-specific effect nor did any associate more strongly than rs6474359 in either ancestral population (see Supplementary Table 8). A recent study found that the T2D risk allele at rs516946 leads to an increased expression of short *ANK1* isoforms in skeletal muscle [27]. The T2D-associated rs516946 variant was not in strong LD ($R^2 < 0.2$) with any amino-altering variants in the study nor with the HbA_{1C}-associated rs6474359 in Europeans ($R^2 = 0.0076$, using LD-link [28]). The authors speculate that the association with T2D may be driven by expression of the short isoforms

of *ANK1*. Protein products of *ANK1* have been shown to interact with insulin receptor substrate 1, through pathways enriched in skeletal muscle [29]. However, more functional studies are needed to understand how the short isoforms influence this interaction. Together this points to that the lead SNP is not the causal SNP in the *ANK1* locus.

In the GEMMA meta-analyses of known HbA_{1C} loci, the *MTNR1B* locus was significantly associated with HbA_{1C}, whereas in our ancestry-specific analyses the association was nominally significant in IAPGG and statistically significant in the European proportion of the data (Inter99 + EAPGG). It has previously been suggested that the *MTNR1B* locus has a European-specific association with T2D [30]. Similarly, a GWAS of East Asians did not show association between SNPs in *MTNR1B* and HbA_{1C} [10]. Of interest, we found almost identical effect sizes of *MTNR1B* rs1387153_T on HbA_{1C} in IAPGG and Inter99 + EAPGG and showed no indication of ancestral-specific effects, which suggests that the association is not unique to Europeans. In addition, data on fasting plasma glucose in Inuit have shown an association between variants in *MTNR1B* and increased fasting plasma glucose level [31], which is in line with the current findings.

In conclusion, motivated by the previous finding of a high-effect T2D-associated variant in the Greenlandic population, we investigated the genetic determinants of HbA_{1C} levels in this population. We did not detect any novel loci associated with HbA_{1C} levels in the Greenlandic population when using data from the MetaboChip. However, we found that one known HbA_{1C}-associated locus, *GCK*, associated with HbA_{1C} in the Inuit ancestral proportion of the Greenlandic population. Secondly, we also found that the *ANK1* locus shows a significant ancestral-specific effect indicating that the lead variant may not be the causal variant in the locus. Finally, for 10 of 15 known HbA_{1C}-associated loci the effect was comparable between IAPGG and the reported effects in CEU or EAS. Our results shed light on genetic effects across populations and ethnicities and can contribute alongside other populations in future trans-ethnic meta-analyses.

Data availability

The data are available through European Genome-phenome Archive (EGA) at the url: <https://ega-archive.org/studies/EGAS00001002641>.

Acknowledgements 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). The B99 and IHIT study was supported by Karen Elise Jensen's Foundation, NunaFonden, the Medical Research Council of Denmark, the Medical Research Council of Greenland, and the Commission for Scientific Research in

Greenland. The study was partially funded by the Novo Nordisk Foundation (grants NNF15OC0017918 and NNF16OC0019986). None of the funding agencies had any role in the study design or in the collection or interpretation of the data.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- 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.
- Sturtevant WC, Damas D. Handbook of North American Indians vol 5. The Arctic. In: Sturtevant WC, Damas D, editors. Handbook of North American Indians vol 5. The Arctic. Washington, D.C.: Smithsonian Institution; 1984. p. 845.
- Jørgensen ME, Borch-Johnsen K, Stolk R, Bjerregaard P. Fat distribution and glucose intolerance among Greenland Inuit. *Diabetes Care.* 2013;36:2988–94.
- Jørgensen ME, Bjerregaard P, Borch-Johnsen K, et al. Diabetes and impaired glucose tolerance among the Inuit population of Greenland. *Diabetes Care.* 2002;25:1766–71.
- Jørgensen ME, Borch-Johnsen K, Witte DR, Bjerregaard P. Diabetes in Greenland and its relationship with urbanization. *Diabet Med.* 2012;29:755–60.
- Moltke I, Grarup N, Jørgensen ME, et al. A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature.* 2014;512:190–3.
- Report A, Consultation WHO. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. *Diabetes Res Clin Pract.* 2011;93:299–309.
- Nathan DM, Turgeon H, Regan S. Relationship between glycated haemoglobin levels and mean glucose levels over time. *Diabetologia.* 2007;50:2239–44.
- Selvin E, Steffes MW, Zhu H, et al. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N Engl J Med.* 2010;362:800–11.
- Chen P, Takeuchi F, Lee J-Y, et al. Multiple nonglycemic genomic loci are newly associated with blood level of glycated hemoglobin in East Asians. *Diabetes.* 2014;63:2551–62.
- Soranzo N, Sanna S, Wheeler E, et al. Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. *Diabetes.* 2010;59:3229–39.
- Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* 2010;42:105–16.
- Scott Ra, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet.* 2012;44:991–1005.
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet.* 2007;39:770–5.
- 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.
- Tosoh method sheet for Hemoglobin A1c.
- BioRad variant hemoglobin A1c program. Instruction manual.
- Jeppsson JO, Kobold U, Barr J, et al. Approved IFCC reference method for the measurement of HbA1c in human blood. *Clin Chem Lab Med.* 2002;40:78–89.
- Jørgensen T, Borch-Johnsen K, Thomsen TF, Ibsen H, Glumer C, Pisinger C. A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99 (1). *Eur J Cardiovasc Prev Rehabil.* 2003;10:377–86.
- Voight BF, Kang HM, Ding J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* 2012;8:e1002793.
- Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;44:821–4.
- Skotte L, Korneliussen TSS, Moltke I, Albrechtsen A. Ancestry specific association mapping in admixed populations. 2015. <https://doi.org/10.1101/014001>.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
- Fumagalli M, Moltke I, Grarup N, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science (80-).* 2015;349:1343–7.
- Hussain S, Richardson E, Ma Y, et al. Glucokinase activity in the arcuate nucleus regulates glucose intake. *J Clin Invest.* 2015;125:337–49.
- Jørgensen ME, Bjerregaard P, Borch-Johnsen K, Witte D. New diagnostic criteria for diabetes: is the change from glucose to HbA1c possible in all populations? *J Clin Endocrinol Metab.* 2010;95:E333–6.
- Scott LJ, Erdos MR, Huyghe JR, et al. The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nat Commun.* 2016;7:11764.
- Machiela MJ, Chanock SJ. LD link: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* 2014;31:3555–7.
- Caruso M, Ma D, Msallat Z, et al. Increased interaction with insulin receptor substrate 1, a novel abnormality in insulin resistance and type 2 diabetes. *Diabetes.* 2014;63:1933–47.
- Wang H, Liu L, Zhao J, et al. Large scale meta-analyses of fasting plasma glucose raising variants in GCK, GCKR, MTNR1B and G6PC2 and their impacts on type 2 diabetes mellitus risk. *PLoS ONE.* 2013;8:e67665.
- Grarup N, Moltke I, Albrechtsen A, Hansen T. Diabetes in population isolates: lessons from Greenland. *Rev Diabet Stud.* 2015;12:320–9.