



**HAL**  
open science

## **A novel rare CUBN variant and three additional genes identified in Europeans with and without diabetes: results from an exome-wide association study of albuminuria**

Tarunveer S Ahluwalia, Christina-Alexandra Schulz, Johannes Waage, Tea Skaaby, Niina Sandholm, Natalie van Zuydam, Romain Charmet, Jette Bork-Jensen, Peter Almgren, Betina H Thuesen, et al.

### **► To cite this version:**

Tarunveer S Ahluwalia, Christina-Alexandra Schulz, Johannes Waage, Tea Skaaby, Niina Sandholm, et al.. A novel rare CUBN variant and three additional genes identified in Europeans with and without diabetes: results from an exome-wide association study of albuminuria. *Diabetologia*, 2019, 62 (2), pp.292-305. 10.1007/s00125-018-4783-z . hal-01976819

**HAL Id: hal-01976819**

**<https://hal.sorbonne-universite.fr/hal-01976819>**

Submitted on 10 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# A novel rare *CUBN* variant and three additional genes identified in Europeans with and without diabetes: results from an exome-wide association study of albuminuria

Tarunveer S. Ahluwalia<sup>1,2,3</sup> · Christina-Alexandra Schulz<sup>4</sup> · Johannes Waage<sup>3</sup> · Tea Skaaby<sup>5</sup> · Niina Sandholm<sup>6,7,8</sup> · Natalie van Zuydam<sup>9,10</sup> · Romain Charmet<sup>11</sup> · Jette Bork-Jensen<sup>2</sup> · Peter Almgren<sup>4</sup> · Betina H. Thuesen<sup>5</sup> · Mathilda Bedin<sup>12</sup> · Ivan Brandslund<sup>13</sup> · Cramer K. Christensen<sup>14</sup> · Allan Linneberg<sup>5</sup> · Emma Ahlqvist<sup>4</sup> · Per-Henrik Groop<sup>6,7,8,15</sup> · Samy Hadjadj<sup>16</sup> · David-Alexandre Tregouet<sup>11</sup> · Marit E. Jørgensen<sup>1,17</sup> · Niels Grarup<sup>2</sup> · Oluf Pedersen<sup>2</sup> · Matias Simons<sup>12</sup> · Leif Groop<sup>4,18</sup> · Marju Orho-Melander<sup>4</sup> · Mark I. McCarthy<sup>9,10</sup> · Olle Melander<sup>4</sup> · Peter Rossing<sup>1,19</sup> · Tuomas O. Kilpeläinen<sup>2</sup> · Torben Hansen<sup>2,20</sup>

Received: 7 August 2018 / Accepted: 22 October 2018 / Published online: 13 December 2018  
© The Author(s) 2018

## Abstract

**Aims/hypothesis** Identifying rare coding variants associated with albuminuria may open new avenues for preventing chronic kidney disease and end-stage renal disease, which are highly prevalent in individuals with diabetes. Efforts to identify genetic susceptibility variants for albuminuria have so far been limited, with the majority of studies focusing on common variants.

**Methods** We performed an exome-wide association study to identify coding variants in a two-stage (discovery and replication) approach. Data from 33,985 individuals of European ancestry (15,872 with and 18,113 without diabetes) and 2605 Greenlanders were included.

**Results** We identified a rare (minor allele frequency [MAF]: 0.8%) missense (A1690V) variant in *CUBN* (rs141640975,  $\beta = 0.27$ ,  $p = 1.3 \times 10^{-11}$ ) associated with albuminuria as a continuous measure in the combined European meta-analysis. The presence of each rare allele of the variant was associated with a 6.4% increase in albuminuria. The rare *CUBN* variant had an effect that was three times stronger in individuals with type 2 diabetes compared with those without ( $p_{\text{interaction}} = 7.0 \times 10^{-4}$ ,  $\beta$  with diabetes = 0.69,  $\beta$  without diabetes = 0.20) in the discovery meta-analysis. Gene-aggregate tests based on rare and common variants identified three additional genes associated with albuminuria (*HES1*, *CDC73* and *GRM5*) after multiple testing correction ( $p_{\text{Bonferroni}} < 2.7 \times 10^{-6}$ ).

**Conclusions/interpretation** The current study identifies a rare coding variant in the *CUBN* locus and other potential genes associated with albuminuria in individuals with and without diabetes. These genes have been implicated in renal and cardiovascular dysfunction. The findings provide new insights into the genetic architecture of albuminuria and highlight target genes and pathways for the prevention of diabetes-related kidney disease.

**Keywords** Albuminuria · Diabetes · DKD · Exome chip · Genetics · Genome-wide association study · Kidney disease · GWAS · Rare variant · SKAT · Type 2 diabetes

Tuomas O. Kilpeläinen and Torben Hansen contributed equally to this study.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00125-018-4783-z>) contains peer-reviewed but unedited supplementary material, which is available to authorised users.

✉ Tarunveer S. Ahluwalia  
tarun.veer.singh.ahluwalia@regionh.dk

Extended author information available on the last page of the article

## Abbreviations

ACR	Urinary albumin/creatinine ratio
Addition	Anglo–Danish–Dutch Study of Intensive Treatment In People with Screen Detected Diabetes in Primary Care
AER	Urinary albumin excretion rate
CADD	Combined Annotation Dependent Depletion
CKD	Chronic kidney disease
DanFunD	Danish study of functional disorders
DIAGRAM	DIAbetes Genetics Replication And Meta-analysis

## Research in context

### What is already known about this subject?

- Increased albuminuria is a key indicator of chronic kidney disease, which is prevalent in diabetic individuals and is thus also known as diabetic kidney disease (DKD)
- A few common genetic variants associated with albuminuria have been identified, it being a heritable trait. However, much of the heritability is still unexplained

### What is the key question?

- Are there other common or rare coding gene variants that are associated with albuminuria and do their effects vary between individuals with and without diabetes?

### What are the new findings?

- The current study describes the discovery and validation of a new rare mutation (~1%) in the *CUBN* gene that associates with increased albuminuria
- The effect of this variant is three times higher in people with diabetes compared with non-diabetic individuals
- The study further uncovers three additional genes modulating albuminuria levels in humans

### How might this impact on clinical practice in the foreseeable future?

- The study findings provide new insights into the genetic architecture of albuminuria and highlight novel genes/pathways for prevention of DKD

DKD	Diabetic kidney disease
EAF	Effect allele frequency
Eur	All individuals of European ancestry (in combined meta-analysis)
eQTL	Expression quantitative trait loci
Eur–GL	Eur pooled with Greenlandic data (in combined meta-analysis)
ExWAS	Exome-wide association study
FinnDiane	Finnish Diabetic Nephropathy Study
GWAS	Genome-wide association study
IMI-SUMMIT	Innovative Medicines Initiative – Surrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools
LD	Linkage disequilibrium
MAF	Minor allele frequency
MDCS	Malmö Diet and Cancer Study
PCs	Principal components
$p_{\text{het}}$	$p$ for heterogeneity
SKAT	Set (sequence) kernel association test
SNP	Single nucleotide polymorphism

## Introduction

Albuminuria is a manifestation of chronic kidney disease (CKD), a major health burden worldwide with a current prevalence of 14.8% in the USA [1]. In individuals with CKD, changes in albuminuria are strongly associated with the risk

of end-stage renal disease and death [2]. Diabetic individuals have an increased risk of developing CKD (referred to as diabetic kidney disease [DKD]); in the USA, the prevalence of CKD is ~41% among individuals with diabetes in comparison with ~10% in individuals without diabetes [3]. DKD proceeds in stages: (1) an increase in albuminuria (or microalbuminuria, 30 to 300 mg/g urinary albumin); (2) progressing to macroalbuminuria or proteinuria (>300 mg/g); (3) loss of kidney function (GFR < 30 ml/min); and finally (4) requiring renal replacement. Recent evidence suggests a new facet of nephron function, with the proximal tubule playing a part in DKD pathophysiology [4] in addition to having a role as a filtration barrier in glomerular haemodynamics.

DKD development may be primarily determined by proximal tubule injury, which is connected to glomerulus hyperfiltration and glomerular barrier damage via mechanisms modulating albumin excretion and re-uptake [4]. Glomerular hyperfiltration, which occurs early in the course of DKD, is augmented by the hyperglycaemic state in diabetes via increased filtering of glucose. This stimulates the proximal tubule to reabsorb glucose which, coupled with sodium reabsorption, results in vasorelaxation of the afferent artery and increased renal blood flow [5]. Albuminuria is a pivotal biomarker among diabetic individuals who develop DKD, reflecting glomerular and tubular dysfunction [6]. It may also reflect a generalised endothelial dysfunction and is associated with an increased risk of cardiovascular events in diabetic individuals [7, 8].

Family studies suggest that genetic factors explain 16–49% of albuminuria [9]. While several genome-wide association

studies (GWASs) of albuminuria have been performed to date, most have focused on identifying common genetic variants (minor allele frequency [MAF]  $\geq 5\%$ ) for albuminuria [10–12]. Recently, we identified rare coding variants for kidney function (estimated GFR [eGFR]) and development in an exome-wide association study (ExWAS) [13]. Here, we used a similar approach to identify rare (MAF  $< 1\%$ ) or low-frequency (MAF 1–5%) coding variants for albuminuria in 33,985 individuals of European ancestry with ( $n = 15,872$ ) and without ( $n = 18,113$ ) diabetes.

## Methods

### Study populations

The present study comprises a two-stage design: discovery and replication. The discovery set includes five cohorts from Denmark (Inter99, Health2006, Health2008, Vejle Biobank and the Anglo–Danish–Dutch Study of Intensive Treatment In People with Screen Detected Diabetes in Primary Care (Addition)-DK, [Table 1]), with a total of 13,226 participants (3896 with and 9330 without type 2 diabetes), as described previously [14, 15] (electronic supplementary material [ESM] Methods 1.1 Discovery Phase). The first three are population based while the last two are type 2 diabetes case cohorts.

The replication set includes multiple studies of European descent ( $n = 20,759$ ) involving 11,976 individuals with and 8783 without diabetes (Table 1). These comprise the Danish study of functional disorders (DanFunD) [16], Malmö Diet and Cancer Study (MDCS) [17], Genesis/Genediab [18], Innovative Medicines Initiative – Surrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools (IMI-SUMMIT) Consortia (Europe/UK-based consortia on diabetes studies) [19] and Greenlandic Inuit populations ( $n = 2605$ ) [20]. The DanFunD, MDCS and Greenlandic studies were population based, whereas the Genesis/Genediab studies involved people with type 1 diabetes and the IMI-SUMMIT comprised four type 1 and five type 2 diabetes studies. All studies are described in ESM Methods 1.2 Replication Phase.

None of the studies overlapped with the previous albuminuria GWASs except two replication cohorts within the IMI-SUMMIT Consortia (Finnish Diabetic Nephropathy Study [FinnDiane] and Scania Diabetes Register) that participated in the type 1 diabetes albuminuria GWAS [12].

The present study was conducted in accordance with the Helsinki Declaration and all the participating studies were approved by their respective data protection boards and by the regional scientific ethics committees. Informed consent from all participants was obtained.

### Albuminuria measurements

Albuminuria was diagnosed from a 24 h urine collection (mg/24 h), also called the urinary albumin excretion rate (AER) or from spot urine samples measuring urinary albumin and creatinine concentrations and calculating the urinary albumin/creatinine ratio (ACR in mg/mmol). The summary measures for AER and ACR in the participating cohorts have been described in Table 1 and methods described in ESM Table 1.

### Genotyping and SNP quality control

Genotyping of the discovery stage studies was performed on the Illumina HumanExome BeadChip 12V.1.0 containing 263,894 single nucleotide polymorphisms (SNPs) and including an additional 16,340 custom-typed SNPs from the Danish Exome Sequencing Project as described previously [14, 21] and briefly in ESM Methods 1.3 Danish Exome Sequencing based SNP Selection. Most SNPs were exome based (non-synonymous/coding) gene variants ( $\sim 90\%$ ); thus, we refer to the present association study as an ExWAS. Genotype calling on the discovery set cohorts was performed using the Illumina GenCall plus the zCall algorithm to improve rare variant calling [22]. We excluded SNPs based on: (1) cluster separation score  $< 0.4$ ; (2) Hardy–Weinberg equilibrium  $p < 10^{-6}$ ; and (3) call rate  $< 98\%$ . We also excluded individuals with: (1) sex mismatches; (2) genetic duplicates; (3) call rate  $< 95\%$ ; and (4) no clustering with the European-ancestry-specific SNPs through a principal component analyses (PCA) approach seeded with ancestry informative markers (AIMs), as described previously [23]. After quality control, a total of 142,397 SNPs remained for a total of 13,226 individuals with complete phenotype and genotype data in the discovery set. Details of the replication cohorts are provided in ESM Table 2. All SNP positions are based on the Genome Reference Consortium Human Build 37 (GRCh37) of dbSNP ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.13/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/)).

One replication (MDCS) cohort used Illumina Exome array-based genotyped data, the Greenlandic cohort used Illumina Metabochip-based genotype data [20] while other cohorts had SNP data available from exome-/genome-wide array imputation.

### Statistical analyses

**Discovery stage** The discovery stage exome-wide association analysis was first performed in each of the five participating studies individually using additive linear regression model and adjusting for sex, age and population sub-structure (principal components; see Project analyses plan in ESM Methods 1.4). Albuminuria measures were natural-log transformed to correct for non-normalised data. The study-specific

**Table 1** Clinical characteristics of the individual cohorts: pooled and stratified on diabetes status: discovery and replication stages

Study name	Sample set type	<i>n</i>	Women, %	Age, years	ACR <sup>a</sup> (mg/mmol) or AER <sup>b</sup> (mg/24 h)	Diabetes, %	Diabetes type	Diabetes duration, years
Stage 1 (discovery)								
Addition-DK <sup>a</sup>	All	2013	47.3	59.6 (7.0)	0.6 (0.2–1.59) <sup>a</sup>	81.6	Type 2	–
	DM	1643	43.9	60.1 (6.8)	0.6 (0.2–1.7) <sup>a</sup>			–
	Non-DM	370	62.4	57.6 (7.4)	0.5 (0.2–1.39) <sup>a</sup>			–
Health2006 <sup>a</sup>	All	2658	51.3	49.6 (12.8)	0.57 (0.45–0.9) <sup>a</sup>	0	–	–
Health2010 <sup>a</sup>	All	642	55.6	46.6 (8.2)	0.45 (0.34–0.57) <sup>a</sup>	0	–	–
Inter99 <sup>a</sup>	All	5971	50.9	46.1 (7.9)	0.34 (0.23–0.57) <sup>a</sup>	5.2	Type 2	–
	DM	311	37.6	51.0 (7.1)	0.45 (0.34–1.13) <sup>a</sup>			–
	Non-DM	5660	51.6	45.9 (7.8)	0.34 (0.23–0.57) <sup>a</sup>			–
Vejle Biobank <sup>c</sup>	DM	1942	38.1	63.4 (8.7)	0.96 (0.54–1.83) <sup>a</sup>	100	Type 2	–
Stage 2 (replication 1): Europeans								
DanFund <sup>a</sup>	All	7364	47.3	52.1 (13.1)	0.57 (0.57–0.57) <sup>a</sup>	6.3	Type 2	–
	DM	449	43.9	60.4 (8.8)	0.57 (0.57–1.47) <sup>a</sup>			–
	Non-DM	6915	62.4	51.4 (13.3)	0.57 (0.57–0.57) <sup>a</sup>			–
Genesis/Genediab <sup>b</sup>	DM	1249	52.2	42.2 (11.9)	0.11 (0.51–1.36) <sup>a</sup>	100	Type 1	25.6 (10.2)
MDCS <sup>a</sup>	All	2641	56.5	73.0 (5.6)	0.6 (0.4–1.2) <sup>a</sup>	21	Type 2	–
	DM	547	48.3	73.6 (5.4)	0.8 (0.4–2.2) <sup>a</sup>			–
	Non-DM	2094	58.6	72.9 (5.6)	0.6 (0.4–1.1) <sup>a</sup>			–
SUMMIT Consortium (diabetes cohorts)								
Benedict (Phase A and B) <sup>c</sup>	DM	324	32.7	68.1 (7.7)	17.3 (4.6–55.6) <sup>b</sup>	100	Type 2	17.2 (7.1)
Cambridge <sup>b</sup>	DM	245	46.9	23.1 (9.9)	0.78 (0.52–1.71) <sup>a</sup>	100	Type 1	15.7 (6.5)
Eurodiab <sup>b</sup>	DM	680	50.0	42.5 (9.9)	23.9 (22.6–25.3) <sup>b</sup>	100	Type 1	24.7 (8.4)
FinnDiane <sup>b</sup>	DM	2840	50.6	43.9 (11.7)	11.0 (5.0–89.6) <sup>b</sup>	100	Type 1	29.1 (10.0)
GoDarts 1 <sup>c</sup>	DM	3530	46.0	67.0 (0.7)	46.1 (38.7–52.6) <sup>a</sup>	100	Type 2	7.5 (6.0)
GoDarts 2 <sup>c</sup>	DM	2805	43.0	66.8 (11.8)	50.5 (40.7–59.3) <sup>a</sup>	100	Type 2	7.3 (6.2)
SDR (Type1) <sup>b</sup>	DM	598	43.0	48.6 (13.7)	8.6 (4.3–50.4) <sup>b</sup>	100	Type 1	31.9 (13.3)
SDR (Type2) <sup>c</sup>	DM	1426	41.3	65.6 (10.7)	18.7 (7.2–100.8) <sup>b</sup>	100	Type 2	14.3 (7.6)
Steno Type 2 Diabetes <sup>c</sup>	DM	295	39.3	61.5 (8.1)	50.5 (13.9–893.0) <sup>b</sup>	100	Type 2	15.1 (6.8)
Stage 2 (replication 2): Greenlanders								
Greenlanders <sup>a</sup> (IHIT + B99)	All	2605 (IHIT = 2519; B99 = 86)	53.7	44.1 (14.5)	0.9 (0.68–1.58) <sup>a</sup>	8.1	–	–

Data for age and diabetes duration are shown as mean ± standard deviation; ACR (mg/mmol)/AER (mg/24 h) are represented as median (interquartile range). For some studies, AER was converted from µg/min to mg/24 h with a multiplication factor of 1.44 (µg/min × 1.44 = mg/24 h). ACR was converted from mg/g to mg/mmol by a multiplication factor of 0.113 (mg/g × 0.113 = mg/mmol)

In sample set type, ‘All’ is all the individuals in the cohort (with and without diabetes). These sets are further stratified based on presence or absence of diabetes. The number of individuals in the phenotyping summary may not match the association summary numbers in actual analyses because genetic information is missing for some individuals

<sup>a</sup> Population-based studies

<sup>b</sup> Type 1 diabetes study

<sup>c</sup> Type 2 diabetes study

DM, diabetes mellitus (with); Go-DARTS, Genetics of Diabetes Audit and Research in Tayside Scotland; IHIT, Inuit Health in Transition study; non-DM, without diabetes mellitus; SDR, Scania Diabetes Registry

results were meta-analysed using inverse variance-weighted fixed-effects meta-analysis with weights proportional to the squared standard errors of the effect estimates. The genomic inflation factor ( $\lambda$ ) was at acceptable levels both in the

individual association analysis ( $\lambda_{\text{Inter99}} = 1.01$ ,  $\lambda_{\text{Health2006}} = 1.0$ ,  $\lambda_{\text{Health2008}} = 1.0$ ,  $\lambda_{\text{Vejle}} = 0.99$ ,  $\lambda_{\text{Addition-DK}} = 1.01$ ) and in the combined discovery meta-exome-wide association analysis ( $\lambda_{\text{discovery}} = 1.0$ ). A  $\chi^2$  test for heterogeneity was



implemented to estimate the heterogeneity in effect size across the participating studies. The proportion of phenotypic variance ( $r^2$  or the coefficient of determination) associated with the top SNP was estimated through a linear regression model (trait  $\sim$  SNP + covariates) with the covariates age + sex + principal components (PCs). METAL software [24] was used for the meta-analysis and the R meta package [25] for constructing meta-forest plots. Ancestry-specific linkage disequilibrium (LD) between variants was extracted using the National Institutes of Health (NIH)-based LDlink database [26].

**Replication stage** SNPs with  $p < 5.0 \times 10^{-5}$  in the discovery meta-analysis were tested for replication in European ( $n = 20,759$ ) and Greenlandic ( $n = 2605$ ) study populations using a similar approach to that used in the discovery analysis. The covariates used for each analysis are given in ESM Table 2. SNPs with imputation quality  $r^2 < 0.3$  were not used for replication (Genesis/Genediab *CUBN* SNP). Following this, a fixed-effects meta-analysis using either inverse variance weighting (wherever possible) or a weighted sum of  $z$  scores was performed. For the IMI-SUMMIT Consortium, no effect sizes or standard errors were available in the summary results. Hence, we performed the replication meta-analysis for the *KCNK5* rs10947789 and *LMX1B* rs140177498 SNPs, for which the IMI-SUMMIT Consortium contributed data, by using the weighted sum of  $z$  scores. Replication meta-analysis with  $p_{\text{Bonferroni}} = 0.017$  (three SNPs) was considered significant.

**Combined meta-analysis** The combined meta-analysis was performed with all individuals of European ancestry (Eur) followed by pooling with Greenlandic data (Eur–GL).

Any SNP with  $p_{\text{replication}} < 0.017$  and  $p_{\text{meta\_Eur/Eur-GL}} < 5.0 \times 10^{-8}$  was considered overall significant.

**Diabetes-stratified analysis** Diabetes-stratified SNP–albuminuria association (for index SNPs) was assessed in the discovery set (with diabetes,  $n = 3896$ ; non-diabetes,  $n = 9330$ ) after pooling individual genotype data on all participants and verified through an interaction regression model (trait  $\sim$  SNP + diabetes\_status + [SNP  $\times$  diabetes\_status] + age + sex + cohort + PC<sub>1</sub>–PC<sub>4</sub>).

**Conditional analyses** Conditional analyses for novel SNPs identified in known loci (and/or in low LD,  $r^2 < 0.01$ ) were performed to determine if the signal was independent. The following linear model was used: trait  $\sim$  top identified SNP + secondary known SNP + sex + age + PCs. If the top SNP retained the association estimates and  $p$  value it was considered an independent signal.

**Gene-aggregate tests** Gene-based multi-marker association testing for rare and common exonic and intragenic variants

(after removing monomorphic variants) was performed using the Meta Analysis for SNP–set (sequence) kernel association test (MetaSKAT) R package [27]. SNPs were filtered based on their annotation status in the Genome Variation Server (GVS: SeattleSeq Annotation 138/hg19), where SNPs belonging to the following categories were taken forward to the gene-based analyses: missense, missense-near-splice, splice-3, splice-5, coding-synonymous, stop-gained, stop-gained-near-splice, stop-lost, stop-lost-near-splice, untranslated region (UTR)-3, UTR-5 and intronic. Intergenic variants or singletons that did not fall under the annotated gene sets were removed, with 18,026 valid gene sets remaining.

At the study-specific level, the gene-based analyses were performed against the null model (using SKAT-O method) [27], accounting for sex and ten PCs, generating SKAT objects individually for each cohort with available genotyped data (discovery + MDCS study, six studies) which were then meta-analysed in a single stage to incorporate maximum power for testing the rare variants cumulatively.

The meta-analysis of the summary-level score statistics was run using the Hom-Meta-SKAT-O ‘optimal’ method, which assumes that different studies share the same causal variant, weighting them equally. A Bonferroni threshold ( $p_{\text{adjusted}} < 2.7 \times 10^{-6}$ ) based on 18,026 annotated gene sets was used.

**Additional SNP–trait associations** We examined the associations of the exome-wide significant index SNPs with type 2 diabetes risk and kidney function (eGFR) in the discovery set. The Pheno Scanner database ([www.phenoscanter.medschl.cam.ac.uk/phenoscanter](http://www.phenoscanter.medschl.cam.ac.uk/phenoscanter)), comprising publicly available results for GWASs and expression quantitative trait loci (eQTL) studies [28], as well as the GWAS summary results for type 2 diabetes from the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium, were also accessed to mine known SNP–trait associations for the index SNPs.

**SNP functionality prediction** We used the Combined Annotation Dependent Depletion (CADD) database to predict variants as per their functional categories (deleterious, disease causal, pathogenicity) through integrating multiple annotations into one metric [29]. A scaled CADD or C score/PHRED [ $-10 \times \log_{10}(\text{rank}/\text{total})$ ] is a ranking for a variant relative to all possible substitutions of the human genome ( $8.6 \times 10^9$ ).

## Results

### Albuminuria ExWAS

**Discovery stage** In the discovery stage meta-analysis, three independent SNPs, including two rare variants (MAF  $< 1\%$ ) in *CUBN* and *LMX1B* and a common variant in *KCNK5*, attained  $p_{\text{discovery}} < 5.0 \times 10^{-5}$  (Table 2). The Manhattan, QQ

**Table 2** Associations for the top SNPs from EXWAS discovery, replication and combined meta-analyses, including Europeans and Europeans + Greenlanders and Greenlanders, for albuminuria

SNP characteristics																
SNP (rsID) <sup>a</sup>	Gene	Chr (BP)	Anno	Discovery <i>n</i> = 13,226			Replication <i>n</i> = 20,759		Combined (Eur) <i>n</i> = 33,985		Combined (Eur+GL) <i>n</i> = 36,590		<i>P</i> <sub>het</sub>			
				EA/OA (%)	EAF (%)	$\beta$	SE	<i>P</i> <sub>discovery</sub>	EAF (%)	<i>n</i>	$\beta$ /DOE	<i>P</i> <sub>replication</sub>		<i>n</i>	<i>P</i> <sub>Eur-GL</sub>	
rs141640975	<i>CUBN</i>	10 (16992011)	missense (A1690V)	A/G	0.8	0.266	0.061	$1.2 \times 10^{-5}$	0.8	9742	0.279	$2.9 \times 10^{-7}$	22,866	$1.3 \times 10^{-11}$	–	–
rs10947789	<i>KCNK5</i>	6 (39174922)	Intronic	C/T	23	0.054	0.012	$1.6 \times 10^{-5}$	24	20,759	+	0.03	33,985	$1.5 \times 10^{-5}$	36,590	$9.1 \times 10^{-6}$
rs140177498	<i>LMXB</i>	9 (129372974)	Upstream	T/C	0.9	0.260	0.058	$8.7 \times 10^{-6}$	1.1	13,233	+	0.43	–	–	–	–

SNPs with  $P_{Eur+GL} < 5.0 \times 10^{-8}$  are novel

Discovery set is based on up to 13,226 individuals (3896 with and 9330 without diabetes)

EAF (Greenlanders, rs10947789) = 45%

EAF values are in %

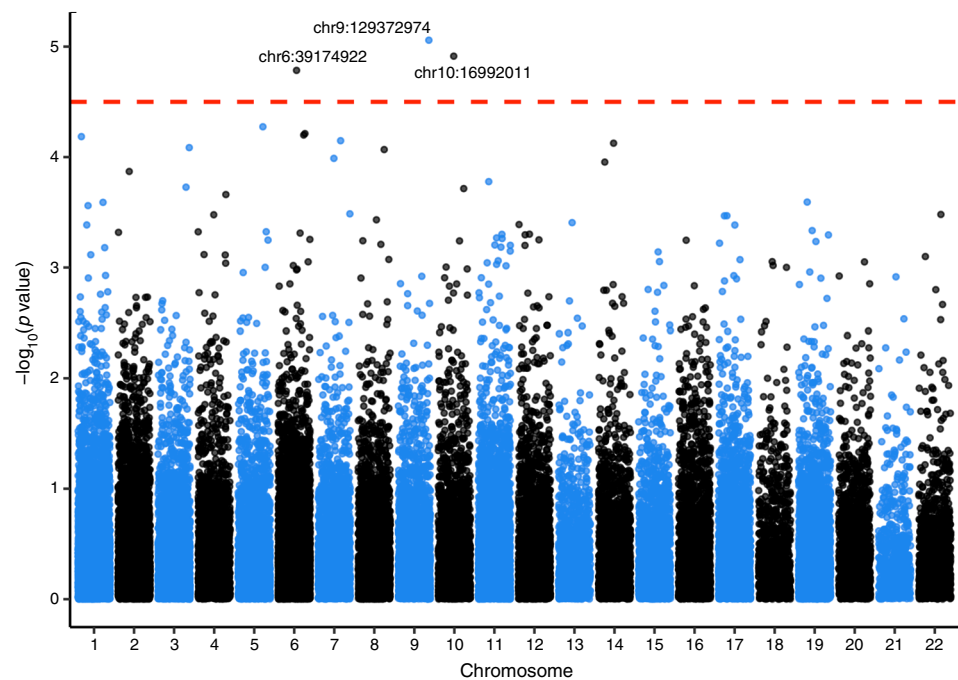
$\beta$  values correspond to natural log transformed albuminuria levels (ACR in mg/mmol or AER in mg/24 h)

*P* values are for association

<sup>a</sup> SNPs were selected for replication based on  $P_{discovery} < 5.0 \times 10^{-5}$

Anno, SNP annotation; BP, SNP position in base pairs from the Genome Reference Consortium Human Build 37 (GRCh37) in dbSNP; Chr, chromosome; DOE, direction of effect corresponding to the association summary meta-analysis based on *z* scores for the respective SNP; EA/OA, effect allele/other allele; ID, identity

**Fig. 1** Manhattan Plot for the discovery stage meta-ExWAS at  $p_{\text{discovery}} < 5.0 \times 10^{-5}$ . The x-axis shows the chromosome number and the y-axis shows  $-\log_{10}(p \text{ values})$  for the SNP–albuminuria association. Index SNPs are named as chromosome number:position (GRCh37 of dbSNP): chr10:16992011, *CUBN* rs141640975; chr6:39174922, *KCNK5* rs10947789; chr9:129372974, *LMX1B* rs140177498. Chr, chromosome



and LocusZoom (<http://locuszoom.sph.umich.edu/>) plots for the discovery meta-analysis are shown in Fig. 1, ESM Fig. 1 and Figs 2 and 3 and genotype-stratified study characteristics are given in ESM Table 3.

The *CUBN* rs141640975 is a rare (MAF 0.83%) missense (A1690V) SNP with the A allele associated with increased albuminuria ( $\beta = 0.25$ ;  $p_{\text{discovery}} = 1.2 \times 10^{-5}$ ). Genotype vs mean albuminuria levels for the *CUBN* rs141640975 are shown in ESM Fig. 2 (levels for *KCNK5* and *LMX1B* SNPs are shown in ESM Figs 3, 4, respectively), whereas a forest plot depicting the effect estimates for each study in the discovery meta-analysis is shown in ESM Fig. 5. The *KCNK5* (which encodes potassium two-pore domain channel subfamily K member 5) common (MAF 23%) intronic SNP rs10947789 C allele and the *LMX1B* (encoding LIM homeobox transcription factor 1  $\beta$ ) rare (MAF 0.89%) intronic SNP rs140177498 T allele associated with increased albuminuria (rs10947789,  $\beta = 0.05$ ,  $p_{\text{discovery}} = 1.6 \times 10^{-5}$ ; ESM Fig. 6; rs140177498,  $\beta = 0.26$ ,  $p_{\text{discovery}} = 8.7 \times 10^{-6}$ ; ESM Fig. 7). Systolic BP-adjusted associations are shown in ESM Table 4.

**Replication stage** Replication of the three SNPs was sought in up to 20,759 individuals (with diabetes,  $n = 11,976$ ; without diabetes,  $n = 8783$ ).

The *CUBN* rs141640975 replicated strongly ( $n_{\text{replication}_{CUBN}} = 9742$ ;  $p_{\text{replication}} = 2.8 \times 10^{-7}$ ), while the *KCNK5* rs10947789 was close ( $n_{\text{replication}_{KCNK5}} = 20,757$ ;  $p_{\text{replication}} = 0.03$ ). The *LMX1B* rs140177498 did not replicate ( $n_{\text{replication}_{LMX1B}} = 13,233$ ;  $p_{\text{replication}} = 0.43$ ) (Table 2) and was not analysed further, whereas we continued to look into *KCNK5*.

**Combined meta-analysis** The combined meta-analysis (discovery + replication) comprised 33,985 European individuals while the combined European–Greenland group comprised 36,590 individuals (Table 2). Only the *KCNK5* SNP was available in the Greenlandic Illumina Metabochip data, and here the *KCNK5* rs10947789 C allele had a frequency of 45%.

The *CUBN* rs141640975 remained significant overall after the Eur meta-analysis ( $p_{\text{meta}_{Eur}} = 1.3 \times 10^{-11}$ ) while the *KCNK5* SNP was non-significant with  $p_{\text{meta}_{Eur-GL}} = 9.1 \times 10^{-6}$  in the Eur–GL meta-analysis (Table 2). The overall study design for single SNP testing is shown in Fig. 3.

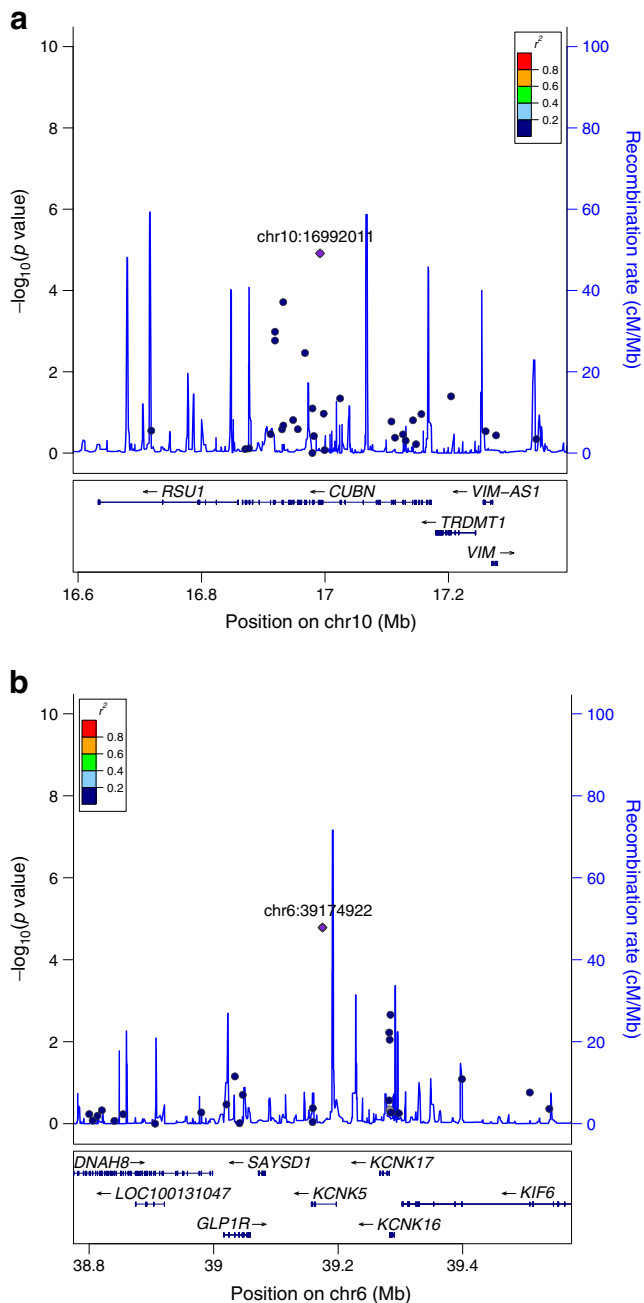
### Diabetes-stratified association

The effect estimates of *CUBN* rs141640975 were more than threefold higher in the type 2 diabetes group ( $\beta = 0.69$ ;  $p = 2.0 \times 10^{-5}$ ; ESM Fig. 8) in comparison with the non-diabetes group ( $\beta = 0.20$ ;  $p = 0.002$ ; ESM Fig. 9) with a significant interaction based on diabetes status ( $p_{\text{interaction}} = 7.0 \times 10^{-4}$ ) (Fig. 4). No significant interaction was observed for the *KCNK5* SNP ( $p_{\text{interaction}} = 0.077$ ). Replication studies were not included in this analysis because individual-level genotype data were not available for all cohorts.

### Conditional analyses

Conditional analysis for the identified *CUBN* rare variant rs141640975 (GRCh37.p13 position: 16,992,011, MAF 0.83%) was carried out for the known *CUBN* common SNP rs1801239 (GRCh37.p13 position 16,919,052, MAF 10%), which occurs at a distance of  $\sim 73$  kbp (LD  $r^2 = 0.0002$ ,  $D' =$





**Fig. 2** (a) Regional plot for *CUBN* rs141640975 on chromosome 10 (for space reasons, the *ST8SIA6* gene is omitted from the right of the key below). (b) Regional plot for *KCNK5* rs10947789 on chromosome 6. Index SNPs are named as chromosome number:position (GRCh37 of dbSNP): chr10:16992011, *CUBN* rs141640975; chr6:39174922, *KCNK5* rs10947789. Chr, chromosome. The diamond (shown in purple) is the most significant SNP in the region (index SNP), based on  $p$  value; all other SNPs are shown as circles. The colours represent the degree of LD with the index SNP (see  $r^2$  values in the key); grey represents unavailability of LD data for a SNP. For additional documentation please see <http://locuszoom.sph.umich.edu/>

1.0). The effect estimates for the novel variant rs141640975 did not change before ( $p_{rs141640975} = 8.8 \times 10^{-7}$ ,  $\beta = 0.33$ ) or after ( $p_{rs141640975\_condition} = 8.5 \times 10^{-7}$ ,  $\beta = 0.33$ ) conditioning with

the known *CUBN* SNP rs1801239 ( $p_{rs1801239} = 0.0002$ ,  $\beta = 0.05$ ), which is also in LD with another reported SNP rs17343073 (LD  $r^2 = 0.92$ ,  $D' = 1.0$ ). Similar conditioning for other known *CUBN* common SNPs (rs6602163/rs10795433) was performed in the same or proxy SNP (LD,  $r^2 \geq 0.6$ , ESM Table 5).

### Power calculations

We performed power calculations for the single SNP analyses using the Genetic Association Study (GAS) power calculator for GWAS [30]. For the rare variant (rs141640975), assuming an additive disease model,  $\alpha = 5 \times 10^{-8}$ ,  $n \sim 33,000$ , MAF  $\sim 1\%$  and relative risk  $\sim 1.5$ , with disease prevalence = 0.15, the expected power was 81%. For the common variant (rs10947789)  $\alpha = 5 \times 10^{-8}$ ,  $n \sim 33,000$ , MAF  $\sim 23\%$  and relative risk  $\sim 1.1$ , the expected power was 62%.

### Gene-aggregate tests

Applying the Hom-O-SKAT Meta (weighted) optimal test from six studies comprising a total of 15,867 individuals, we identified three genes *HES1* ( $p = 3.7 \times 10^{-9}$ ) in chromosome 3, *CDC73* ( $p = 6.4 \times 10^{-9}$ ) in chromosome 1 and *GRM5* ( $p = 1.6 \times 10^{-6}$ ) in chromosome 11 (Table 3) surviving Bonferroni correction ( $p_{\text{Bonferroni}} < 2.7 \times 10^{-6}$ , 18,026 gene sets). Association results for individual SNPs included in the gene-aggregate tests for these three genes are shown in ESM Table 6, whereas the overall analysis design is shown in ESM Fig. 10.

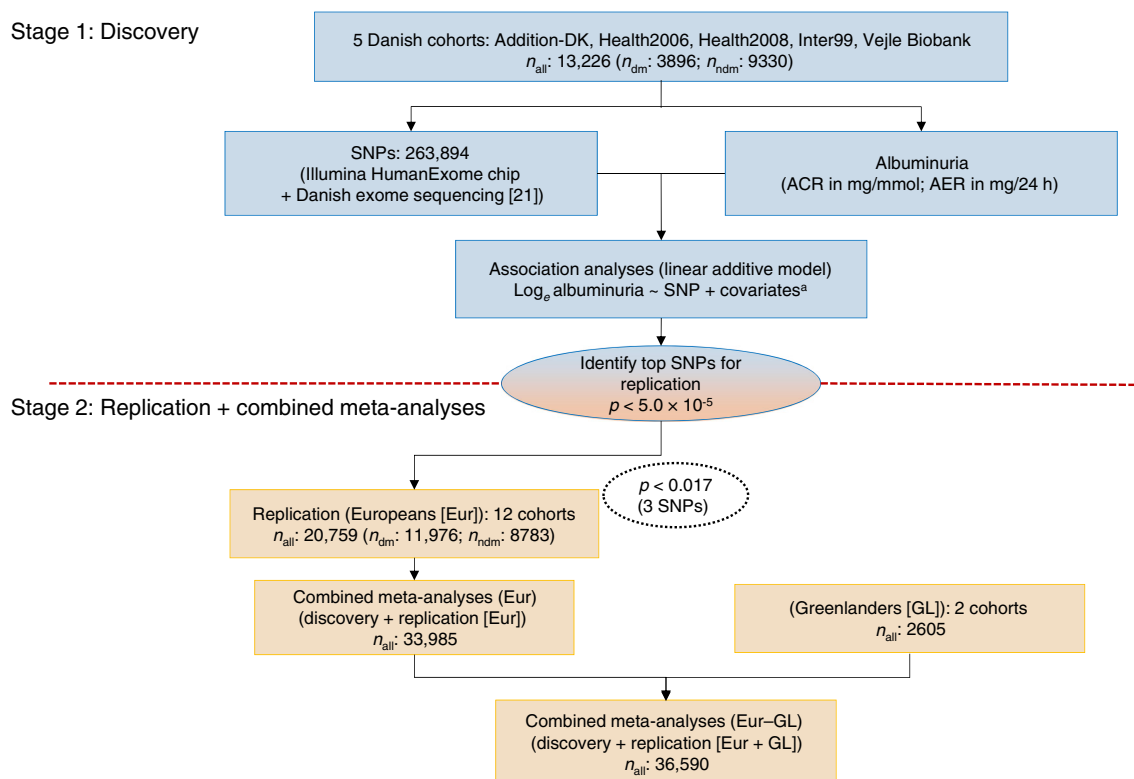
### Additional SNP–trait associations

The *CUBN* rs141640975 A allele was not associated with type 2 diabetes risk in the discovery set ( $p > 0.05$ ) or in the DIAGRAM Consortium summary results ( $p > 0.05$  in models unadjusted and adjusted for BMI; ESM Table 7). However, the A allele showed a nominally significant association with increased eGFR ( $p = 0.04$ ,  $\beta = 0.026$ ) within the discovery set (ESM Table 7).

The publicly available report suggests that the *CUBN* rare SNP A allele is associated with reduced serum creatinine levels in a blood-based metabolomics study ( $p = 0.014$ ,  $\beta = -0.28$ ).

The *KCNK5* rs10947789 C allele associated with increased urinary ACR ( $\beta = 0.025$ ,  $p = 5.6 \times 10^{-4}$ ) in CKD Genetics Consortia-based research. Other *KCNK5*-associated traits include myocardial infarction (CARDIoGRAMplusC4D Consortia,  $\beta = -0.059$ ,  $p = 1.4 \times 10^{-6}$ ), visual refractive error ( $\beta = 0.11$ ,  $p = 0.003$ ) and birthweight (Early Growth Genetics [EGG] Consortium,  $\beta = 0.025$ ,  $p < 0.007$ ) in addition to *KCNK5* being a GWAS locus for coronary artery disease.

eQTL data suggest rs10947789 C allele-specific gene expression associations within the adrenal gland ( $\beta = 0.51$ ,  $p =$



**Fig. 3** Study design overview. dm, with diabetes; ndm, without diabetes. <sup>a</sup>covariates: age, sex, PCs

$2.5 \times 10^{-5}$ ), subcutaneous adipose tissue ( $\beta = 0.34$ ,  $p = 7.2 \times 10^{-5}$ ), lymphoblastoid cell line ( $\beta = 0.03$ ,  $p = 0.0019$ ) and tibial nerve ( $\beta = 0.15$ ,  $p = 0.002$ ).

Detailed SNP–trait associations with references are documented in ESM Tables 8–10.

### SNP functionality prediction

The *CUBN* missense SNP rs141640975 was described as functionally ‘deleterious’ with SIFT, ‘probably damaging’ with PolyPhen and had a CADD (scaled C score) of 24.5.

### Discussion

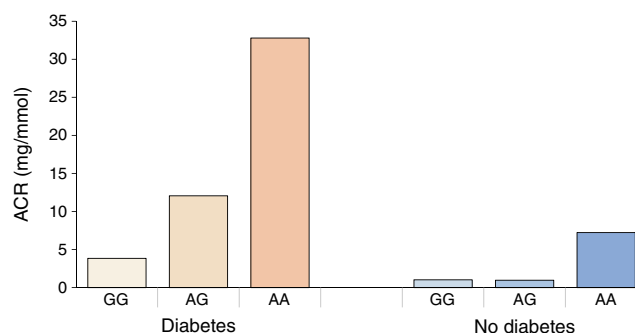
In the combined meta-ExWAS of 33,985 Europeans (five discovery and 12 replication cohorts), we identified one novel *CUBN* variant associated with albuminuria levels and exerting >3.5-fold increased effects among individuals with type 2 diabetes compared with non-diabetic individuals.

Although *CUBN* is a known locus for albuminuria, the identified rare missense variant shows independent effects (with respect to known SNPs in *CUBN*) that are stronger within the diabetes vs the non-diabetes group ( $p_{\text{interaction}} = 7.0 \times 10^{-4}$ ). This rare variant explains up to 6.4% of variance per rare allele (in a model adjusted for age and sex) in albuminuria levels (natural log transformed). Also, the gene-based

tests identify three additional genes (*HES1*, *CDC73* and *GRM5*) that associate with albuminuria ( $p_{\text{Bonferroni}} < 2.7 \times 10^{-6}$ ) in a meta-analysis comprising six Scandinavian cohorts.

There have been a few albuminuria GWASs in the past decade [10–12, 31], all exploring the common genetic variants (MAF > 5%), but in the current study we examined low-frequency and rare variants, particularly from the coding region (exome) of the genome.

While common variants in *CUBN* have been previously reported to associate with albuminuria in individuals of European, African and Hispanic ancestry [10, 32], the rare missense (A1690V) SNP rs141640975 in *CUBN* that we



**Fig. 4** ACR levels stratified by *CUBN* rs141640975 genotype (GG, AG and AA) and type 2 diabetes status in the discovery cohorts. Diabetes, with GG  $n = 3806$ , with GA  $n = 66$ , with AA  $n = 1$ ; no diabetes, with GG  $n = 9270$ , with GA  $n = 152$ , with AA  $n = 1$

**Table 3** Genes associated with albuminuria through gene-aggregate tests

Gene	Chr	<i>p</i> value	No. SNPs/set	Full gene name	Alternative gene symbols	Gene ID (NCBI) 21 Dec 2017
<i>HES1</i>	3q29	$3.7 \times 10^{-9}$	2	Hes family bHLH transcription factor 1	<i>HHL; HRY; HES-1; bHLHb39</i>	3280
<i>CDC73</i>	1q312	$6.4 \times 10^{-9}$	2	Cell division cycle 73	<i>HYY; FIHP; HPTJT; HRPT1;</i> <i>HRPT2; C1orf28</i>	79,577
<i>GRM5</i>	11q142–q143	$1.6 \times 10^{-6}$	7	Glutamate metabotropic receptor 5	<i>mGlu5; GPRC1E; MGLUR5;</i> <i>PPP1R86</i>	2915

All results significant at  $p < 2.7 \times 10^{-6}$ , after correction for (Bonferroni's) multiple testing and are based on six study cohorts (discovery set + MDCS) Inclusions for gene-aggregate tests are based on annotated exonic or intragenic SNPs, totalling  $n = 18,026$  gene sets examined

Detailed results for the identified genes are available in ESM Table 6

The SKAT-cohort function was used to run each cohort, adjusting for age, sex, PCs + study-specific covariates. Following this, the Meta SKAT function and Hom-O-SKAT Meta (weighted) model was used to meta-analyse individual cohort data. The rest of the replication studies did not have data/resources available to run these analyses and therefore could not be incorporated. A brief overview of the gene-aggregate testing is presented in ESM Fig. 10

Chr, chromosome; ID, identity; no., number

identified is not in LD with the recently reported *CUBN* SNPs rs1801239 ( $r^2_{LD} = 0.0002$ ,  $D'_{LD} = 1.0$ ) [10] and rs6602163 ( $r^2_{LD} = 0.0008$ ,  $D'_{LD} = 1.0$ ) [11] for European ancestry. This is confirmed in the current study through conditional analyses (rs1801239/rs17343073,  $p_{\text{conditional}} = 8.5 \times 10^{-7}$  and rs6602163/rs10795433,  $p_{\text{conditional}} = 4.9 \times 10^{-7}$ ), with the minor allele associated with increased albuminuria.

A strong interaction between diabetes status and *CUBN* missense rs141640975 observed with respect to albuminuria in the current study suggests potential clinical implications. Cubilin, encoded by the *CUBN* gene, is expressed in the apical brush border of proximal renal tubule cells and forms a complex with megalin protein to promote albumin re-uptake [10, 33]. An important mechanism underlying DKD is tubulointerstitial damage involving the proximal tubule [4]. A diabetic or hyperglycaemic state alters the tubular function by augmenting glomerular hyperfiltration [4, 5], while the proximal tubule may determine the level of glomerular hyperfiltration through glucose reabsorption [34]. Excess albumin in the urine may be a consequence of defective tubular reabsorption in diabetic individuals, which could be protective against DKD by reducing the reabsorption of glucose. A recent study on diabetic mice identified a lower expression of cubilin (part of the endocytic machinery) in the renal cortex and proximal tubule and a correlation between tubular endocytosis dysfunction and higher urinary excretion of albumin, transferrin and total protein [35], highlighting the critical role of cubilin and the proximal tubule in the diabetic state.

As cubilin protein is a co-receptor not only for tubular resorption but also for the intestinal vitamin B<sub>12</sub>-intrinsic factor complex, *CUBN* mutations lead to a hereditary form of megaloblastic anaemia (or Imerslund–Gräsbeck syndrome) characterised by tubular proteinuria and vitamin B<sub>12</sub> malabsorption [36, 37]. Moreover, a recent exome-sequencing study revealed a homozygous frameshift mutation in *CUBN* associated with the only

cause of proteinuria in affected family members [37]. Despite *CUBN* being a disease gene, recent exome-sequencing studies and related reference databases (ExAC [38]) have shown that damaging variants are rather frequent in 'non-diseased' populations and are thus well tolerated by humans [38, 39]. On this basis, it was recently hypothesised that the tubular proteinuria caused by cubilin deficiency could actually be protective against tubular overload, seen, for example, in nephrotic syndrome or even DKD [39]. As *CUBN* rs141640975 has been associated with lower serum creatinine ( $p = 0.014$ ) in a recent meta-GWAS of circulating metabolites [40] and causes albuminuria also in the general population group, our study supports the idea that functional variations in *CUBN* might not be damaging but instead protective. The CADD/SIFT database testing of functionality of the non-synonymous SNPs that result in amino acid changes suggests *CUBN* rs141640975, with a high C score and deleterious nature, to be of functional importance. Indeed, we found that the albuminuria-increasing rs141640975 A allele was associated with more efficient kidney function in the discovery set ( $p = 0.04$ ), suggesting it may be protective against DKD. However, further functional and validation studies are required to shed light on the potential protective effect of this *CUBN* variant.

Although no other SNP was replicated in the single SNP analyses, the *KCNK5* gene was close to the replication threshold ( $p = 0.03$ ), though it did not reach the GWAS threshold in the combined Eur–GL meta-analysis ( $p = 9.1 \times 10^{-6}$ ). *KCNK5* is a known coronary artery disease GWAS locus [41], encoding the potassium two-pore domain channel subfamily K member 5 protein, which is mainly expressed in the cortical distal tubules and collecting ducts of the kidney [42]. This protein is highly sensitive to pH, and functional inactivation may lead to renal acidosis [43]. Data mining revealed the rs10947789 minor allele (C) to be associated with increased albuminuria in the CKDGen Consortia [11], supporting our findings. eQTL-based look-ups indicate rs10947789 associated strongly with *KCNK5* expression

in the adrenal gland and subcutaneous adipose tissue ( $p < 8.0 \times 10^{-5}$ ), suggesting a functional role in the kidneys. Albeit *KCNK5* rs10947789 has promising roles with respect to the cardio–renal axis, the power calculation in the current study (62%) suggests that the variant has smaller effects; a larger sample size may be required to validate its association with albuminuria.

The *HES1* gene identified through the gene-aggregate tests is a transcription factor ubiquitously expressed in most organs, including the kidneys; it has been documented to be involved in Notch signalling pathways that play a role in renal fibrosis [44], glomerulosclerosis [45] and other forms of kidney disease [44, 46]. The *CDC73* gene is a tumour suppressor gene, mutations in which have been associated with hyperparathyroidism–jaw tumour syndrome and familial hyperparathyroidism [47]. Albuminuria is associated with hyperparathyroidism, which is a complication of CKD [48], and the present findings thus suggest a plausible link between the two.

*GRM5* encodes glutamate metabotropic receptor 5, which is a G protein-coupled receptor involved in second messenger signalling. Variants in the metabotropic glutamate receptor group I pathway, including *GRM1* and *GRM5*, were enriched in the pathway analysis of a recent albuminuria GWAS among people with type 1 diabetes from the FinnDiane study [12] (no individuals overlapped with the current gene-aggregate meta-analysis). *GRM5* is also expressed in podocytes and is associated with podocyte apoptosis in animals [49] and pharmacological effects in humans [50].

Albeit the strength of the current study is the identification and validation of a rare variant with optimal power, more statistical power could have improved *KCNK5* validation. Also, testing type 2 diabetes individuals in the discovery stage and including type 1 diabetes cohorts in the replication stage may have introduced some heterogeneity. However, our main finding, the association of the *CUBN* rare variant with albuminuria, remains unaffected as the meta-analysis for this variant only included type 2 diabetes cohorts because of the lack of genotyping or poor imputation quality in type 1 diabetes replication cohorts. Furthermore, the *KCNK5* SNP showed the same direction of effect in both type 1 and type 2 diabetes cohorts, and there was no significant difference in the association of *KCNK5* with albuminuria between these cohorts ( $p$  for heterogeneity [ $p_{het}$ ]  $> 0.05$ ). For the gene-aggregate testing we used six cohorts (having genotype data) in a single-stage meta-analysis to ensure quality and maximise statistical power for rare variant analysis.

In summary, we identified a rare coding *CUBN* variant implicated in elevated albuminuria levels, especially in individuals with type 2 diabetes. Further, we identified additional novel genes associated with albuminuria through an alternative gene-aggregate approach among Europeans. Our findings provide fresh insights into the genetic architecture of albuminuria and highlight new targets, genes and pathways for the prevention and treatment of DKD.

**Acknowledgements** Study-specific acknowledgements are available in the ESM Table 11 and SUMMIT Consortia members are listed in ESM Table 12. TSA acknowledges P. G. Ahluwalia (Project Manager [Quality], Novo Nordisk, Bagsvaerd, Denmark) for helpful feedback on the study from a layman's point of view.

Some of the data were presented as an abstract at the American Society of Nephrology (ASN) annual meeting in 2018. The results presented in this paper have not been published elsewhere in whole or part, except in abstract form.

**Data availability** All data generated or analysed during this study are included in this published article (and the electronic supplementary material). The summary data for the associations of the index SNP with other traits and with type 2 diabetes that support the findings of this study are available from the PhenoScanner and DIAGRAM Consortium websites ([www.phenoscan.com](http://www.phenoscan.com) and [www.diagram-consortium.org/downloads.html](http://www.diagram-consortium.org/downloads.html)). Additional data are available on request from the authors (T. S. Ahluwalia and/or T. Hansen).

**Funding** TSA was supported by Lundbeck Foundation travel grant (reference 2013-14471). This study was partially funded by the Immunometabolism grant from the Novo Nordisk Foundation (NNF15CC0018486). Study-specific funding details are shown in ESM Table 11.

**Duality of interest** P-HG has received investigator-initiated research grants from Lilly and Roche, is an advisory board member for AbbVie, AstraZeneca, Boehringer Ingelheim, Cebix, Lilly, Janssen, Medscape, Merck Sharp & Dohme, Novartis, Novo Nordisk and Sanofi and has received lecture fees from AstraZeneca, Boehringer Ingelheim, Lilly, Elo Water, Genzyme, Merck Sharp & Dohme, Medscape, Novo Nordisk and Sanofi. PR received consultancy and/or speaking fees (to institution) from AbbVie, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol-Myers Squibb, Eli Lilly, MSD, Novo Nordisk and Sanofi Aventis and has received research grants from AstraZeneca and Novo Nordisk. PR also has shares in Novo Nordisk. All other authors declare that there is no duality of interest associated with this manuscript.

**Contribution statement** TSA and TH were responsible for the conception and design of the study. TH, OP, NG, AL, TS, BHT, MEJ, IB and CKC were responsible for recruitment, cohort management and genotyping of the discovery set cohorts along with the DanFunD and Greenlanders study. TSA drafted and revised the paper. JB-J performed data quality control for the discovery cohorts. TSA performed data analyses for the discovery set and the DanFunD cohort in the replication set and interpreted data. C-AS and PA performed replication data analyses for the MDCS cohort. JW and TSA analysed the gene-aggregate tests and interpreted results. RC and D-AT performed replication for Genesis/Genediab cohort. NS, NvZ and EA performed replication for the SUMMIT Consortia cohorts. NG performed replication data analyses for Greenlanders. TSA, TOK, NG, NS, P-HG, MB, MS, SH, LG, MIM, OM, MO-M, PR and TH were responsible for the data interpretation and useful comments/suggestions. All authors reviewed and approved the final version of the manuscript. TSA is the guarantor of this work.

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

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.




## References

- Saran R, Robinson B, Abbott KC et al (2018) US renal data system 2017 annual data report: epidemiology of kidney disease in the United States. *Am J Kidney Dis* 71(3):A7. <https://doi.org/10.1053/j.ajkd.2018.01.002>
- Carrero JJ, Grams ME, Sang Y et al (2017) Albuminuria changes are associated with subsequent risk of end-stage renal disease and mortality. *Kidney Int* 91(1):244–251. <https://doi.org/10.1016/j.kint.2016.09.037>
- Plantinga LC, Crews DC, Coresh J et al (2010) Prevalence of chronic kidney disease in US adults with undiagnosed diabetes or prediabetes. *Clin J Am Soc Nephrol* 5(4):673–682. <https://doi.org/10.2215/CJN.07891109>
- Zeni L, Norden AGW, Cancarini G, Unwin RJ (2017) A more tubulocentric view of diabetic kidney disease. *J Nephrol* 30(6):701–717. <https://doi.org/10.1007/s40620-017-0423-9>
- Anders HJ, Davis JM, Thurau K (2016) Nephron protection in diabetic kidney disease. *N Engl J Med* 375(21):2096–2098. <https://doi.org/10.1056/NEJMcibr1608564>
- Gluhovschi C, Gluhovschi G, Petrica L et al (2016) Urinary biomarkers in the assessment of early diabetic nephropathy. *J Diabetes Res* 2016:4626125
- Nichols GA, Deruaz-Luyet A, Hauske SJ, Brodovicz KG (2018) The association between estimated glomerular filtration rate, albuminuria, and risk of cardiovascular hospitalizations and all-cause mortality among patients with type 2 diabetes. *J Diabetes Complicat* 32(3):291–297. <https://doi.org/10.1016/j.jdiacomp.2017.12.003>
- de Zeeuw D, Remuzzi G, Parving HH et al (2004) Albuminuria, a therapeutic target for cardiovascular protection in type 2 diabetic patients with nephropathy. *Circulation* 110(8):921–927. <https://doi.org/10.1161/01.CIR.0000139860.33974.28>
- Mottl AK, Vupputuri S, Cole SA et al (2009) Linkage analysis of albuminuria. *J Am Soc Nephrol* 20(7):1597–1606. <https://doi.org/10.1681/ASN.2008080895>
- Boger CA, Chen MH, Tin A et al (2011) CUBN is a gene locus for albuminuria. *J Am Soc Nephrol* 22(3):555–570. <https://doi.org/10.1681/ASN.2010060598>
- Teumer A, Tin A, Sorice R et al (2016) Genome-wide association studies identify genetic loci associated with albuminuria in diabetes. *Diabetes* 65(3):803–817. <https://doi.org/10.2337/db15-1313>
- Sandholm N, Forsblom C, Makinen VP et al (2014) Genome-wide association study of urinary albumin excretion rate in patients with type 1 diabetes. *Diabetologia* 57(6):1143–1153. <https://doi.org/10.1007/s00125-014-3202-3>
- Li M, Li Y, Weeks O et al (2017) SOS2 and ACP1 loci identified through large-scale exome Chip analysis regulate kidney development and function. *J Am Soc Nephrol* 28(3):981–994. <https://doi.org/10.1681/ASN.2016020131>
- Ahluwalia TS, Allin KH, Sandholt CH et al (2015) Discovery of coding genetic variants influencing diabetes-related serum biomarkers and their impact on risk of type 2 diabetes. *J Clin Endocrinol Metab* 100(4):E664–E671. <https://doi.org/10.1210/jc.2014-3677>
- Ahluwalia TS, Troelsen JT, Balslev-Harder M et al (2017) Carriers of a VEGFA enhancer polymorphism selectively binding CHOP/DDIT3 are predisposed to increased circulating levels of thyroid-stimulating hormone. *J Med Genet* 54(3):166–175. <https://doi.org/10.1136/jmedgenet-2016-104084>
- Dantoft TM, Ebstrup JF, Linneberg A et al (2017) Cohort description: the Danish study of functional disorders. *Clin Epidemiol* 9:127–139. <https://doi.org/10.2147/CLEP.S129335>
- Rosvall M, Persson M, Ostling G et al (2015) Risk factors for the progression of carotid intima-media thickness over a 16-year follow-up period: the Malmo Diet and Cancer Study. *Atherosclerosis* 239(2):615–621. <https://doi.org/10.1016/j.atherosclerosis.2015.01.030>
- Charnet R, Duffy S, Keshavarzi S et al (2018) Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes. *Cardiovasc Diabetol* 17(1):61. <https://doi.org/10.1186/s12933-018-0705-0>
- van Zuydam NR, Ahlqvist E, Sandholm N et al (2018) A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes* 67(7):1414–1427. <https://doi.org/10.2337/db17-0914>
- Grarup N, Moltke I, Andersen MK et al (2018) Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet* 50(2):172–174. <https://doi.org/10.1038/s41588-017-0022-7>
- Albrechtsen A, Grarup N, Li Y et al (2013) Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* 56(2):298–310. <https://doi.org/10.1007/s00125-012-2756-1>
- Goldstein JI, Crenshaw A, Carey J et al (2012) zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 28(19):2543–2545. <https://doi.org/10.1093/bioinformatics/bts479>
- Tian C, Gregersen PK, Seldin MF (2008) Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 17(R2):R143–R150. <https://doi.org/10.1093/hmg/ddn268>
- Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26(17):2190–2191. <https://doi.org/10.1093/bioinformatics/btq340>
- Schwarzer G (2007) Meta: an R package for meta-analysis. *R News* 7:40–45
- Machiela MJ, Chanock SJ (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31(21):3555–3557. <https://doi.org/10.1093/bioinformatics/btv402>
- Lee S, Teslovich TM, Boehnke M, Lin X (2013) General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* 93(1):42–53. <https://doi.org/10.1016/j.ajhg.2013.05.010>
- Staley JR, Blackshaw J, Kamat MA et al (2016) PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* 32(20):3207–3209. <https://doi.org/10.1093/bioinformatics/btw373>
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46(3):310–315. <https://doi.org/10.1038/ng.2892>
- Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38(2):209–213. <https://doi.org/10.1038/ng1706>
- Pattaro C (2018) Genome-wide association studies of albuminuria: towards genetic stratification in diabetes? *J Nephrol* 31(4):475–487. <https://doi.org/10.1007/s40620-017-0437-3>
- Kramer HJ, Stilp AM, Laurie CC et al (2017) African ancestry-specific alleles and kidney disease risk in Hispanics/Latinos. *J Am Soc Nephrol* 28(3):915–922. <https://doi.org/10.1681/ASN.2016030357>
- Amsellem S, Gburek J, Hamard G et al (2010) Cubilin is essential for albumin reabsorption in the renal proximal tubule. *J Am Soc Nephrol* 21(11):1859–1867. <https://doi.org/10.1681/ASN.2010050492>
- Vallon V, Gerasimova M, Rose MA et al (2014) SGLT2 inhibitor empagliflozin reduces renal growth and albuminuria in proportion to hyperglycemia and prevents glomerular hyperfiltration in diabetic Akita mice. *Am J Physiol Ren Physiol* 306(2):F194–F204. <https://doi.org/10.1152/ajprenal.00520.2013>
- Figueira MF, Castiglione RC, de Lemos Barbosa CM et al (2017) Diabetic rats present higher urinary loss of proteins and lower renal expression of megalin, cubilin, CIC-5, and CFTR. *Physiol Rep* 5(13):e13335. <https://doi.org/10.14814/phy2.13335>



36. Grasbeck R (2006) Imlerslund-Grasbeck syndrome (selective vitamin B(12) malabsorption with proteinuria). *Orphanet J Rare Dis* 1(1):17. <https://doi.org/10.1186/1750-1172-1-17>
37. Ovunc B, Otto EA, Vega-Warner V et al (2011) Exome sequencing reveals cubilin mutation as a single-gene cause of proteinuria. *J Am Soc Nephrol* 22(10):1815–1820. <https://doi.org/10.1681/ASN.2011040337>
38. Lek M, Karczewski KJ, Minikel EV et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536(7616):285–291. <https://doi.org/10.1038/nature19057>
39. Simons M (2018) The benefits of tubular proteinuria: an evolutionary perspective. *J Am Soc Nephrol* 29:710–712
40. Kettunen J, Demirkan A, Wurtz P et al (2016) Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 7:11122. <https://doi.org/10.1038/ncomms11122>
41. Consortium CAD, Deloukas P, Kanoni S et al (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 45:25–33
42. Bittner S, Bobak N, Herrmann AM et al (2010) Upregulation of K2P5.1 potassium channels in multiple sclerosis. *Ann Neurol* 68(1):58–69. <https://doi.org/10.1002/ana.22010>
43. Warth R, Barriere H, Meneton P et al (2004) Proximal renal tubular acidosis in TASK2 K+ channel-deficient mice reveals a mechanism for stabilizing bicarbonate transport. *Proc Natl Acad Sci U S A* 101(21):8215–8220. <https://doi.org/10.1073/pnas.0400081101>
44. Huang R, Zhou Q, Veeraragoo P, Yu H, Xiao Z (2011) Notch2/Hes-1 pathway plays an important role in renal ischemia and reperfusion injury-associated inflammation and apoptosis and the gamma-secretase inhibitor DAPT has a nephroprotective effect. *Ren Fail* 33(2):207–216. <https://doi.org/10.3109/0886022X.2011.553979>
45. Ueno T, Kobayashi N, Nakayama M et al (2013) Aberrant Notch1-dependent effects on glomerular parietal epithelial cells promotes collapsing focal segmental glomerulosclerosis with progressive podocyte loss. *Kidney Int* 83(6):1065–1075. <https://doi.org/10.1038/ki.2013.48>
46. Kobayashi T, Terada Y, Kuwana H et al (2008) Expression and function of the Delta-1/Notch-2/Hes-1 pathway during experimental acute kidney injury. *Kidney Int* 73(11):1240–1250. <https://doi.org/10.1038/ki.2008.74>
47. van der Tuin K, Tops CMJ, Adank MA et al (2017) CDC73-related disorders: clinical manifestations and case detection in primary hyperparathyroidism. *J Clin Endocrinol Metab* 102(12):4534–4540. <https://doi.org/10.1210/jc.2017-01249>
48. Inker LA, Coresh J, Levey AS, Tonelli M, Muntner P (2011) Estimated GFR, albuminuria, and complications of chronic kidney disease. *J Am Soc Nephrol* 22(12):2322–2331. <https://doi.org/10.1681/ASN.2010111181>
49. Gu L, Liang X, Wang L et al (2012) Functional metabotropic glutamate receptors 1 and 5 are expressed in murine podocytes. *Kidney Int* 81(5):458–468. <https://doi.org/10.1038/ki.2011.406>
50. Collett VJ, Collingridge GL (2004) Interactions between NMDA receptors and mGlu5 receptors expressed in HEK293 cells. *Br J Pharmacol* 142(6):991–1001. <https://doi.org/10.1038/sj.bjp.0705861>

## Affiliations

Tarunveer S. Ahluwalia<sup>1,2,3</sup>  · Christina-Alexandra Schulz<sup>4</sup> · Johannes Waage<sup>3</sup> · Tea Skaaby<sup>5</sup> · Niina Sandholm<sup>6,7,8</sup> · Natalie van Zuydam<sup>9,10</sup> · Romain Charmet<sup>11</sup> · Jette Bork-Jensen<sup>2</sup> · Peter Almgren<sup>4</sup> · Betina H. Thuesen<sup>5</sup> · Mathilda Bedin<sup>12</sup> · Ivan Brandslund<sup>13</sup> · Cramer K. Christensen<sup>14</sup> · Allan Linneberg<sup>5</sup> · Emma Ahlqvist<sup>4</sup> · Per-Henrik Groop<sup>6,7,8,15</sup> · Samy Hadjadj<sup>16</sup> · David-Alexandre Tregouet<sup>11</sup> · Marit E. Jørgensen<sup>1,17</sup> · Niels Garup<sup>2</sup> · Oluf Pedersen<sup>2</sup> · Matias Simons<sup>12</sup> · Leif Groop<sup>4,18</sup> · Marju Orho-Melander<sup>4</sup> · Mark I. McCarthy<sup>9,10</sup> · Olle Melander<sup>4</sup> · Peter Rossing<sup>1,19</sup> · Tuomas O. Kilpeläinen<sup>2</sup> · Torben Hansen<sup>2,20</sup>

<sup>1</sup> Steno Diabetes Center Copenhagen, Niels Steensens Vej 2, 2820 Gentofte, Denmark

<sup>2</sup> Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup> Copenhagen Prospective Studies on Asthma in Childhood, Gentofte and Herlev Hospital, University of Copenhagen, Copenhagen, Denmark

<sup>4</sup> Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden

<sup>5</sup> Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Capital Region, Copenhagen, Denmark

<sup>6</sup> Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland

<sup>7</sup> Abdominal Center, Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

<sup>8</sup> Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland

<sup>9</sup> Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>10</sup> Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

<sup>11</sup> Inserm UMR-S 1166, Sorbonne Universités, UPMC Université Paris, Paris, France

<sup>12</sup> Paris Descartes University–Sorbonne Paris Cité, Imagine Institute, Paris, France

- <sup>13</sup> Department of Clinical Immunology and Biochemistry, Lillebaelt Hospital, Vejle, Denmark
- <sup>14</sup> Department of Internal Medicine and Endocrinology, Lillebaelt Hospital, Vejle, Denmark
- <sup>15</sup> Department of Diabetes, Central Clinical School, Monash University, Melbourne, VIC, Australia
- <sup>16</sup> L'institut du thorax, Department of Endocrinology, CIC 1413 INSERM, CHU Nantes, Nantes, France
- <sup>17</sup> National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark
- <sup>18</sup> Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
- <sup>19</sup> University of Copenhagen, Copenhagen, Denmark
- <sup>20</sup> Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark