A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes

Natalie R van Zuydam[1,2], Emma Ahlqvist[3], Niina Sandholm[4,5,6], Harshal Deshmukh[7], N William Rayner[1,2,8], Moustafa Abdalla[1,2,9], Claes Ladenvall[3], Daniel Ziemek[10], Eric Fauman[11], Neil R Robertson[1,2], Paul M McKeigue[12], Erkka Valo[4,5,6], Carol Forsblom[4,5,6], Valma Harjutsalo[4,5,6,13], FINNDIANE Study centres[14], Annalisa Perna[15], Erica Rurali[15], M Loredana Marcovecchio[16], Robert P Igo Jr[17], Rany M Salem[18], Norberto Perico[15], Maria Lajer[19], Annemari Käräjämäki[20,21], Minako Imamura[22,23,24], Michiaki Kubo[24], Atsushi Takahashi[25,26], Xueling Sim[27], Jianjun Liu[27,28,29], Rob M van Dam[27], Guozhi Jiang[30], Claudia H T Tam[30], Andrea O Y Luk[30,31,32], Heung Man Lee[30,31,32,33], Cadmon K P Lim[30], Cheuk Chun Szeto[30], Wing Yee So[30], Juliana C N Chan[30,31,32], Hong Kong Diabetes Registry TRS Project Group[34], Su Fen Ang[35], Rajkumar Dorajoo[28], Ling Wang[28], Tan Si Hua Clara[35], Amy-Jayne McKnight[36], Seamus Duffy[36], The Warren 3/UK GoKinD Study Group[37], Marcus G Pezzolesi[38], GENIE Consortium[39], Michel Marre[40], Beata Gyorgy[40], Samy Hadjadj[41,42,43], Linda T Hiraki[44], DCCT/EDIC group[45], Tarunveer S Ahluwalia[19,46], Peter Almgren[47], Christina-Alexandra Schulz[47], Marju Orho-Melander[47], Allan Linneberg[48,49,50], Cramer Christensen[51], Daniel R Witte[52,53], Niels Grarup[46], Ivan Brandslund[54,55], Olle Melander[56], Andrew D Paterson[44], David Tregouet[40], Alexander P Maxwell[36], Su Chi Lim[57,58,59], Ronald C W Ma[30,31,32,33], E Shyong Tai[27,58,60], Shiro Maeda[22,23,24], Valeriya Lyssenko[3,61], Tiinamaija Tuomi[6,62,63,64], Andrzej S Krolewski[65], Stephen S Rich[66], Joel N Hirschhorn[67,68,69], Jose C Florez[68,70,71], David Dunger[16,72], Oluf Pedersen[46], Torben Hansen[46,73], Peter Rossing[19,46], Giuseppe Remuzzi[15,74,75], SUMMIT Consortium[76], Mary Julia Brosnan[77], Colin N A Palmer[78], Per-Henrik Groop[4,5,6,79], Helen M Colhoun[80], Leif C Groop[3,63], Mark I McCarthy[1,2,81]

Affiliations

1. Wellcome Centre Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK.
2. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK.
3. Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, 20502, Sweden.
4. Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, FI-00290, Finland.
5. Abdominal Center: Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland, FI-00290, Finland.
6. Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland.
7. Newcastle University, Newcastle, U.K..
8. Department of Human Genetics, Wellcome Trust Sanger Institute, University of Cambridge, Hinxton, Cambridgeshire, CB10 1SA, UK.
9. Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK.
10. Inflammation and Immunology Research Unit, Pfizer, Berlin, 10785, Germany.
11. Computational Target Validation, Pfizer, Cambridge, MA, 02139, USA.
12. The Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH16 4UX, UK.
13. The Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, 00300, Finland.

14. Supplementary Table 12.

15. IRCCS- Istituto di Ricerche Farmacologiche "Mario Negri", Clinical Research Center for Rare Diseases "Aldo e Cele Daccò", Ranica, Bergamo, Italy.

16. Department of Paediatrics, University of Cambridge, Cambridge, CB2 0QQ, UK.

17. Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, 44106, USA.

18. Department of Family Medicine and Public Health, University of California, San Diego, U.S.A.

19. Steno Diabetes Center Copenhagen, Gentofte, 2820, Denmark.

20. Department of Primary Health Care, Vaasa Central Hospital, Vaasa, Finland.

21. Diabetes Center, Vaasa Health Care Center, Vaasa, Finland.

22. Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Japan.

23. Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Japan.

24. RIKEN Center for Integrative Medical Sciences, Japan.

25. Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Japan.

26. Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Japan.

27. Saw Swee Hock School of Public Health, National University of Singapore, Singapore.

28. Division of Human Genetics, Genome Institute of Singapore, A*STAR, 138672, Singapore.

29. Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

30. Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China.

31. Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China.

32. Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China.

33. Integrated Bioinformatics Laboratory for Cancer and Metabolic Diseases, The Chinese University of Hong Kong, China.

34. Supplementary Table 13.

35. Clinical Research Unit, Khoo Teck Puat Hospital, National Healthcare Group, 768828, Singapore.

36. Centre for Public Health, Queen's University Belfast, Belfast, Co. Down, BT9 7AB, UK.

37. Supplementary Table 14.

38. Division of Nephrology and Hypertension, Diabetes and Metabolism Center, University of Utah, Salt Lake City, Utah, USA.

39. Supplementary Table 15.

40. Sorbonnes Université, UPMC Univ. Paris 06, INSERM UMR_S 1166, ICAN Institute for Cardiometabolism and Nutrition, Department of Genomics and Pathophysiology of Cardiovascular Diseases, Paris, 75013, France.

41. Endocrinology, Diabetology, CHU Poitiers, Poitiers, France.

42. CIC 1402 & U1082, INSERM, Poitiers, FRANCE.

43. UFR médecine pharmacie, University of Poitiers, Poitiers, France.

44. Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada.

45. Supplementary Table 16.

46. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2100, Denmark.

47. Department of Clinical Sciences, Diabetes and Cardiovascular Disease, Genetic Epidemiology, Lund University, Malmö, 20502, Sweden.

48. Research Centre for Prevention and Health, Capital Region of Denmark, Glostrup, 2600, Denmark.

49. Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark.

50. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

51. Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, 7100, Denmark.

52. Department of Public Health, Aarhus University, Aarhus C, 8000, Denmark.

53. Danish Diabetes Academy, Odense, 5000, Denmark.

54. Institute of Regional Health Research, University of Southern Denmark, Odense, 5000, Denmark.

55. Department of Clinical Biochemistry, Vejle Hospital, Vejle, 7100, Denmark.

56. Department of Clinical Sciences, Hypertension and Cardiovascular Disease, Lund University, Malmö, 20502, Sweden.

57. Diabetes Centre; Clinical Research Unit; Department of Medicine, Khoo Teck Puat Hospital, National Health Group, Singapore.

58. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, 119228, Singapore.

59. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

60. Cardiovascular & Metabolic Disorders Program, Duke-NUS Medical School Singapore, 169857, Singapore.

61. Department of Clinical Science, K. G. Jebsen Center for Diabetes Research, University of Bergen, Norway, N-5020, NO.

62. Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki, 00029, Finland.

63. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.

64. Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Helsinki, Finland.

65. Joslin Diabetes Center, Harvard Medical School, Boston, MA, 02215, USA.

66. Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, 22908-0717, USA.

67. Center for Basic and Translational Obesity Research and Division of Endocrinology, Boston Children's Hospital, Boston, MA, 02115, USA.

68. Programs in Medical Genetics and Metabolism, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.

69. Department of Genetics, Harvard Medical School, Boston, Massachusetts, 02115, USA.

70. Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA.

71. Department of Medicine, Harvard Medical School, Boston, Massachusetts, 02115, USA.

72. Wellcome Trust-Medical Research Council Institute of Metabolic Science, University of Cambridge, Cambridge, UK.

73. Faculty of Health Sciences, University of Southern Denmark, Odense, 5000, Denmark.

74. Unit of Nephrology and Dialysis, Azienda Socio-Sanitaria Territoriale Papa Giovanni XXIII, Bergamo, Italy

75. Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy.

76. Supplementary Table 17.

77. CVMET Research Unit, Pfizer, Cambridge, MA, 02139, USA.

78. Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, DD1 9SY, UK.

79. Baker IDI Heart and Diabetes Institute, Melbourne.

80. Institute of Genetics and Molecular Medicine, University of Edinburgh,U.K .

81. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK.

Abstract

Identification of sequence variants robustly associated with predisposition to diabetic kidney disease (DKD) has the potential to provide insights into the pathophysiological mechanisms responsible. We conducted a genome-wide association study (GWAS) of DKD in type 2 diabetes (T2D) using eight complementary dichotomous and quantitative DKD phenotypes: the principal dichotomous analysis involved 5,717 T2D subjects, 3,345 with DKD. Promising association signals were evaluated in up to 26,827 subjects with T2D (12,710 with DKD). A combined (T1D+T2D) GWAS was performed using complementary data available for subjects with T1D, which, with replication samples, involved up to 40,340 diabetic subjects (and 18,582 DKD cases).

Analysis of specific DKD phenotypes identified a novel signal near *GABRR1* (rs9942471, $p=4.5\times10^{-8}$) associated with 'microalbuminuria' in European T2D cases. However, no replication of this signal was observed in Asian subjects with T2D, or in the equivalent T1D analysis. There was only limited support, in this substantially enlarged analysis, for association at previously-reported DKD signals, except for those at *UMOD* and *PRKAG2*, both associated with 'eGFR'.

We conclude that, despite challenges in addressing phenotypic heterogeneity, access to increased sample sizes will continue to provide more robust inference regarding risk-variant discovery for DKD.

Introduction

Progressive loss of renal function represents one of the most serious complications of diabetes, yet strategies for prevention and management are suboptimal. One of the principal obstacles to improved clinical interventions remains rudimentary understanding of the processes whereby sustained exposure to elevated levels of glucose (and/or other manifestations of the diabetic state) leads to progressive disturbance of renal morphology and function (1).

There is considerable variation in the progression and severity of renal complications of diabetes (collectively, diabetic kidney disease [DKD]). The prevalence of DKD in subjects with T2D is ~30-50%: some patients experience a relatively rapid decline in renal function, whilst others maintain normal renal function despite decades of suboptimal glycemic control (2). The factors influencing this variation in outcome have not been fully characterised, but substantial evidence supports a genetic contribution. As in type 1 diabetes (T1D), DKD in those with type 2 diabetes (T2D) aggregates in families (3, 4), and the prevalence of DKD in T2D differs considerably between ethnic groups (5-7).

These observations indicate that the identification of genetic variants influencing DKD predisposition should accelerate characterization of the biological basis of DKD. In contrast with most complex multifactorial traits, efforts to apply candidate gene and genome wide association studies (GWAS) approaches to DKD have met with limited success (8-11). Many genetic associations have been reported, but few robustly replicated loci have emerged. This likely reflects the comparatively small sample sizes of previous studies, such that power would have been limited to detection of common loci of unusually large effect. In the case of DKD in T2D, this is likely to have been compounded by the heterogeneity of the phenotype: autopsy studies indicate that only ~50% of chronic kidney disease (CKD) in T2D can be attributed to

classical diabetic nephropathy (12). The success of equivalent GWAS efforts for CKD (for which several replicated loci have been described) provides reassurance that it is possible to identify variants with broad impact on the progression of renal disease, irrespective of the dominant pathology (13).

Reduced kidney function, reflected by the estimated glomerular filtration rate (eGFR) and end-stage renal disease (ESRD), and dysfunction of the glomerular filtration barrier, reflected by albuminuria, can develop independently. This suggests that the two cardinal features of DKD involve distinct disease mechanisms and may be subject to different genetic effects. Albuminuria is known to be a poor predictor of diabetes-related ESRD, especially in the early stages, and regression to normoalbuminuria is common in patients with microalbuminuria (14).

These observations provide confidence that the combination of increased sample size and improved definition of DKD phenotypes should enable risk-variant detection and uncover mechanisms that contribute to renal dysfunction in diabetes. In particular, the separation of cases into phenotypic classes based on disease stage and/or phenotype manifestations, incorporating information on both albumin excretion and eGFR, can be expected to increase etiological homogeneity and augment power for locus identification (14-16).

The SUMMIT (SUrrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools) consortium adopted such a strategy to perform a GWAS for DKD in subjects with T1D (17). Here, we report on equivalent analyses conducted in the context of T2D, as well as those from a combined (T1D+T2D) analysis involving up to 40,340 subjects.

Methods

Diabetic kidney disease phenotype definitions

Not all patients with DKD will develop every form of the disease or progress to the most severe stage of ESRD. Dysfunction of the glomerular barrier, represented by albuminuria, and reduced kidney function, represented by eGFR, can develop independently. To explore the disease severity spectrum and the different disease processes represented by eGFR and albuminuria, we defined seven binary phenotypes using clinical measures of albumin creatinine ratio (ACR), AER and eGFR (Table 1 [T2D-only] and Supplementary Table 8 [T1D+T2D]). The phenotype definitions were aligned to other large-scale genetic studies of T1D-DKD in SUMMIT (17) and the Diabetic Nephropathy Collaborative Research Initiative (DNCRI) (18). The definition of chronic kidney disease was also aligned to that used by the CKDGen consortium (eGFR<60 mL/min/1.73m^2) although we restricted cases and controls to those with diabetes (13).

We used AER measured over-night (µg/min), during 24 hours (mg/24 h), or as a spot measurement of ACR (mg/mmol) or eGFR calculated using the Modification of Diet Renal Disease Study Group (MDRD) formula (eGFR=32788 x Serum Creatinine (µmol/L)$^{-1.154}$ x Age$^{-0.203}$ x [0.742 if female]) to classify disease stage and severity. We based the control definition on either AER or ACR as most studies had measured either. In the studies that had measured both, 2/3 measures for AER and ACR had to meet the control criteria (Table 1). We were unable to exclude albuminuric patients that presented as normoalbuminuric due to prescribed renin-angiotensin system blockers. Since reduced kidney function (reflected by eGFR) and dysfunction of the glomerular filtration barrier (reflected by albuminuria) can develop independently, we did not exclude individuals with albuminuria from the controls for the eGFR-defined phenotypes and vice versa. In subjects with T2D, ~46% of normoalbuminuric controls

had an eGFR<60 mL/min/1.73m^2 (1,098/2,372).

In all, we defined seven dichotomous phenotypes:

- the 'all DKD' phenotype, our primary phenotype, designed to capture the broadest set of DKD phenotypes;
- the 'microalbuminuria' phenotype (equivalent to 'early DKD' from Sandholm et al, 2017) (17) to identify variants that contribute to early dysfunction of the glomerular barrier;
- the 'late DKD' phenotype to identify variants that contribute to severe glomerular barrier dysfunction;
- two 'ESRD' related phenotypes focused on identification of variants associated with end stage renal failure, comparing those with ESRD either to control subjects without any DKD ('ESRD vs controls') and relative to control subjects without ESRD ('ESRD vs no ESRD');
- the chronic kidney disease ('CKD') phenotype to identify variants that contribute to reduced kidney function (eGFR);
- the 'CKD and DKD' phenotype to identify any variants that may contribute to the development of kidney disease irrespective of glomerular barrier dysfunction or reduced kidney function; and
- estimated glomerular filtration rate ('eGFR'), a continuous phenotype, to identify variants that play a role in kidney function that may not be detected by the analysis of the binary DKD phenotypes. The eGFR measures were not transformed as they

approximated a normal distribution (Supplementary Figure 1).

Study populations

We identified DKD cases and controls in subjects with T2D from the Scania diabetes registry (SDR) (19), the Genetics of Diabetes and Audit Research in Tayside Scotland (GoDARTS) study (20), the Steno Diabetes Centre (21) and the BENEDICT (Bergamo Nephrologic Diabetes Complications Trial, Italy) A and B studies (22). We identified independent replication studies in populations of European descent (deCODE, the Family investigation of nephropathy and diabetes [FIND] study, the DIabetes REgister in VAsa region [DIREVA] study, the Diagnostic Optimization and treatment of diabetes and its complications in the Chernihiv region [DOLCE] study, the Malmo Diet and Cancer Study [MDC], Inter99, Vejle Diabetes Biobank and the Anglo-Danish-Dutch study of Intensive Treatment In PeOple with screeN detected diabetes in primary care [ADDITION]), and Asian descent (RIKEN, the Singapore Diabetic Cohort Study [SDCS], the Hong Kong Diabetes Registry [HKDR] and the Singapore Study of Macro-angiopathy and Micro-vascular Reactivity in Type 2 Diabetes [SMART2D] study) (Supplementary Table 1).

We combined the subjects with T2D with non-overlapping samples from the study of DKD in subjects with T1D (17). Replication studies (of DKD in subjects with T1D) (17), were also used for replication in the combined analysis of T1D+T2D (Supplementary Table 1). None of these studies overlapped with samples included in the analysis of eGFR and CKD by the CKDGen consortium (13).

Genome-wide genotyping and imputation

The T2D discovery cohorts were genotyped on the Affymetrix SNP 6.0, the Illumina Omni express array and the Illumina 610Quad arrays (Supplementary Table 4). Individual study centres excluded SNPs for minor allele frequency (MAF) <1%. SNPs with a MAF 1-5% were

excluded if the Hardy-Weinberg equilibrium (HWE) test $p<1\times10^{-4}$ or the call rate <99%. SNPs with MAF ≥5% were excluded if the HWE $p<5.7\times10^{-7}$ or the call rate <95% (23). Samples were excluded if: their call rate was <95%; genotype heterozygosity was >3SD from the study sample mean; or they failed gender checks. Based on principal component analysis: population outliers were removed if they were not of European descent (compared to the 1000G populations) or fell >3SD away from the population means of the first two principal components for samples of European descent. Duplicates were removed but related individuals were retained for genotype imputation.

Genotypes were prephased using SHAPE-IT (v2) (24) and imputed using IMPUTEv2 (25) against the March 2013 1000G version 1 reference panel using standard protocols and recommended settings.

Replication genotyping

Direct typing of twelve SNPs (rs11622435, rs12917707, rs17421627, rs1989248, rs2194025, rs2206136, rs4977388, rs61277444, rs6865390, rs7222331, rs9939609 and rs9942471) was performed in DIREVA samples using TaqMan allelic discrimination assays according to the manufacturer's protocol (Applied Biosystems, Carlsbad, CA). Sequenom multiplex genotyping was performed for the same SNPs in DOLCE using the standard protocol (26).

Statistical analysis

Heritability of diabetic kidney disease phenotypes

Narrow sense heritability was estimated by GCTA (v1.26) (27) from 4.5 million directly typed and imputed markers (info>0.75) in GoDARTS (Supplementary Table 1) for 'all DKD', 'CKD' and

'eGFR'. The sample size for these phenotypes exceeded the recommended threshold for reliable heritability estimates (N=3,160 based on a standard error ≤ 0.1) (28).

Genome-wide association analysis

Genome-wide association analyses were performed by individual study centres using an additive model whilst correcting for age, gender and duration of diabetes. We estimated allelic effects using the score test from SNPTESTv2 in unrelated samples for dichotomous traits (29). Association *p* values were calculated using EMMAX from a larger sample of related individuals whilst correcting for a kinship matrix (30). For 'eGFR' we estimated allelic effects and association *p* values using EMMAX (30).

Power calculations

We performed power calculations for dichotomous traits based on a MAF of 8%, an allelic OR range 1.05-2.00, and $\alpha=5\times10^{-8}$ (genome-wide significance). The power calculations were performed, for the discovery meta-analysis of 'all DKD', separately for the T2D-only (3,345 DKD cases and 2,372 DKD controls) and the combined (T1D+T2D; 5,908 DKD cases and 4,965 DKD controls) meta-analyses

At $\alpha\leq5\times10^{-8}$, we had >80% power to detect an allelic OR>1.40 in the T2D-only discovery analysis (Supplementary Figure 2C) and an allelic OR>1.25 (Supplementary Figure 2B) in the combined discovery analysis. We also performed power calculations for the reported DKD loci, as above, but using $\alpha=9\times10^{-4}$ (this α accounts for the number of loci tested but not the number of phenotypes analysed). In the combined analysis (T1D+T2D, 'all DKD') we had >80% power to detect variants with an allelic OR>1.20 (Supplementary Figure 2A).

Discovery meta-analysis

Two discovery meta-analyses were performed: one that included summary statistics estimated from subjects with T2D-only and a second that combined T2D-only analyses with equivalent analyses in subjects with T1D (17). Individual study summary statistics were centrally filtered for a minor allele count in either cases or controls <10 and an info score <0.4 for imputed variants.

EMMAX *p* values were combined in a sample size weighted z-statistic meta-analysis using METAL (version 25/03/2011) (31). Effect estimates were combined in a fixed-effect-inverse-variance weighted meta-analysis using GWAMA (v2.1) (32). Meta-analysis results were restricted to allelic effects estimated in ≥ two studies. For binary traits, independent variants (>100Kb apart) were selected for replication from the T2D-only analysis based on association $p \leq 5\times10^{-6}$ and from the combined (T1D+T2D) analysis based on $p \leq 1\times10^{-6}$. For 'eGFR', SNPs were chosen for replication based on association $p \leq 5\times10^{-6}$ in subjects with T2D or $p < 1\times10^{-6}$ in the combined analysis. SNPs associated with 'eGFR' at $p \leq 5\times10^{-4}$ in either 'eGFR' analysis (T2D only or T1D+T2D) which had also been reported at $p \leq 5\times10^{-8}$ with eGFR by the CKDGen consortium were also included in the list of SNPs for replication (13).

Replication

We sought replication for 164 lead variants in thirteen studies of T2D-DKD for which it was possible to obtain *in silico* replication from available GWAS data or replication from where *de novo* genotyping (DIREVA and DOLCE) (Figure 1). Replication studies aligned their DKD phenotypes with those employed in the SUMMIT GWAS. Although association results for the lead variants were recovered for all compatible DKD phenotypes available in the replication

samples (Supplementary Table 1), joint meta-analysis results were reported for those phenotypes where the primary GWAS associations exceeded the thresholds above.

As with the discovery, meta-analysis effect estimates from replication studies were combined using GWAMA (v2.1) (32), and EMMAX *p* values using METAL (version 25/03/2011) (31).

Known DKD variants

We examined the literature for variants that have been associated with DKD from candidate gene ($p<0.05$) and GWA ($p\leq5\times10^{-8}$) studies. Sixty-one variants were identified and aligned to the reported risk allele for binary traits (or the trait-raising allele for quantitative traits). We assessed both direction of effect and strength of association in the present study for those phenotypes that most closely matched the original report (but irrespective of type of diabetes).

Genetic risk score analysis

We included variants ($p\leq5\times10^{-8}$) from GWAS to generate genetic risk scores (GRS) for: coronary artery disease (CAD) (33); body mass index (BMI) (34); waist-hip-ratio adjusted for BMI (WHR) (35); low-density lipoprotein cholesterol (LDL-C); triglycerides (TRIG); high-density lipoprotein cholesterol (HDL-C) (36); fasting insulin (FI); insulin resistance (IR) (37-39); fasting glucose (FG) (38); T1D (40); T2D (41); and systolic blood pressure (SBP) (42). The relationship between the GRS and the DKD phenotype was calculated using an inverse-variance weighted method described in Ehret et al., 2011(42).

Results

DKD definitions: We considered seven dichotomous phenotypes designed to capture the spectrum of DKD (see **Methods**), and 'eGFR'. We aimed to identify variants that influence multiple stages in DKD progression, as well as those that have more stage-specific effects. The principal definition ('all DKD') included 3,345 T2D subjects with any form of DKD (ranging from microalbuminuria through to ESRD) as cases, and 2,372 T2D subjects, normoalbuminuric despite >10 years duration of diabetes, as controls. The other six dichotomous phenotypic comparisons are described in Table 1 (see also **Methods**).

Contribution of Genetic Variants to DKD: The genetic variation, explained by the SNPs on the genotyping array and estimated using GCTA (v1.26) (30) in up to 6,335 subjects with T2D from the GoDARTS, was highest in 'CKD' (h^2=0.12) and similar for 'all DKD' (h^2=0.08) and 'eGFR' (h^2=0.07) (Supplementary Table 2). We restricted analyses to phenotypes with sample sizes deemed sufficient for accurate estimation of heritability (N≥3,160 to obtain an SE≤0.1) (28).

GWAS for DKD in T2D: The DKD discovery analysis combined GWAS data from four studies of European descent: GoDARTS (20), SDR (19), STENO (Denmark) (21) and the BENEDICT study (phases A and B, Italy) (22) (Table 1, Supplementary Table 3). For the principal ('all DKD') analysis, the sample size of the discovery T2D-only meta-analysis had >80% power to detect variants with MAF≥8% and allelic OR>1.40 (Supplementary Figure 2C). The number of variants meta-analysed for each DKD phenotype varied between 5,864,445 in the 'ESRD vs no ESRD' phenotype and 9,263,264 in the 'all DKD' phenotype (Supplementary Table 4). These differences reflect the minor allele count exclusion filter.

Manhattan and QQ plots of discovery *p*-values for each of the eight DKD phenotypes were well calibrated, and several showed a modest excess of significant associations (Supplementary Figure 3). In the discovery GWAS, only one locus reached genome-wide significance: *PLCB4 (*encoding 1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-4) on chromosome 20. The lead variant rs2206136 was associated with the 'CKD' phenotype (EAF 42%; OR 1.20 [1.08, 1.34]; $p=2.1\times10^{-8}$) (Table 2; Supplementary Figure 3A).

To extend power to detect associations of lesser effect, and to replicate the *PLCB4* association, we identified 139 loci with SNP associations exceeding $p\leq5\times10^{-6}$ in at least one of the seven dichotomous DKD analyses. We also identified 22 loci (25 lead variants) for replication from the 'eGFR' analysis (based on either $p<5\times10^{-6}$ in our 'eGFR' analyses alone, or $p<5\times10^{-4}$ in our analysis and a genome-wide association [$p<5\times10^{-8}$] reported by the CKDGen consortium) (Supplementary Figure 3Q) (13). We sought replication for 164 lead variants in thirteen studies of T2D-DKD (nine involving European subjects, and four involving Asian subjects) for which it was possible to obtain association analyses based on either *in silico* (from existing GWAS) or *de novo* genotyping (Figure 1). Replication studies recoded their DKD phenotypes to align definitions with those employed in the SUMMIT GWAS. Although association results for the lead variants were recovered for all compatible DKD phenotypes available in the replication samples (Supplementary Table 1), joint meta-analysis results are reported for only those phenotypes where the primary GWAS associations exceeded the thresholds above (Supplementary Table 5). The replication samples available for the 'all DKD' phenotype included up to 3,999 T2D subjects of European ancestry (1,270 cases) and 17,111 (8,095 cases) from Asia (Supplementary Table 1).

The 'CKD' association near *PLCB4* did not replicate in either European or Asian data (joint analysis ; $OR_{Asian+Euro}$ 1.12 [1.05, 1.19]; $p=2.1\times10^{-4}$) (Table 2). Joint analysis of dichotomous DKD phenotypes identified one novel SNP association that marginally exceeded genome-wide significance ($p=5\times10^{-8}$, without adjustment for the multiple GWAS we performed) (Table 2). This signal, on chromosome 6, is centred on rs9942471 and lies ~7kb upstream of *GABRR1* (encoding the rho1 subunit of the GABA type a receptor). The major allele was associated with increased risk of 'microalbuminuria' in subjects of European ancestry (Joint analysis; EAF 64%; OR_{Euro} 1.25 [1.16, 1.34]; $p=4.5\times10^{-8}$) (Figure 2, Table 2). Associations of rs9942471 with other DKD phenotypes are given in Supplementary Table 6.

Rs9942471 is in high LD ($r^2>0.8$) with the lead eQTL variant for *GABRR1* expression in artery, oesophagus and skin ($p\leq4\times10^{-8}$) and the major allele is associated with decreased expression (42). However, there was no evidence for replication of this SNP in T2D subjects of Asian ancestry only (EAF 90%; OR_{Asian} 0.99 [0.87, 1.13]; $p=0.91$) although the higher frequency of the effect allele in Asians (90%) compared to Europeans (64%) reduces the power to detect an effect in subjects of Asian descent. Ethnic differences in regional LD could have contributed to failed replication: rs9942471 may be a better marker of the shared causal variant in subjects of European descent. However, this seems unlikely given broad similarity of LD patterns across subjects of European and Asian descent (estimated separately from the 1000G population).

Replication samples for the 'eGFR' phenotype included 8,749 subjects of European and 9,071 subjects of Asian ancestry with T2D (Figure 1). Joint analysis of discovery and replication results captured the well-established association with variants near *UMOD* (uromodulin), centred on rs11864909 ($\beta_{Asian+Euro}$ 2.34 [1.68, 3.00] mL/min/1.73m^2; $p=4.4\times10^{-12}$) (Table 2). There was no difference in effect by diabetes type: the effect estimate in subjects with T1D (β_{T1D} 1.23 (-0.05,

2.51), p=0.06) overlapped the effect size in subjects with T2D (17). We also compared the effects of variants associated with DKD phenotypes in subjects with T2D from Table 1 with their effects in equivalent DKD phenotypes in subjects with T1D (17) (Supplementary Table 7).

Combined T1D and T2D analysis: To increase power to detect loci that contribute to processes involved in the development of DKD irrespective of diabetes subtype, we combined the results from the primary GWAS meta-analysis for T2D-DKD phenotypes with those for the corresponding T1D-DKD phenotypes (Supplementary Table 4, 9) (17). The combined discovery meta-analysis of 'all-DKD' included 10,873 diabetic subjects of European descent (5,908 cases) and provided >80% power ($\alpha=5\times10^{-8}$) to detect a SNP association with an allelic OR>1.25 for variants with MAF >8% (Supplementary Figure 2B). The number of variants meta-analysed ranged from 7,959,015 for 'ESRD vs no ESRD' to 9,364,702 for the 'all DKD' phenotype (Supplementary Table 4).

No significant associations were detected for dichotomous DKD phenotypes in the combined (T1D+T2D) meta-analysis (Supplementary Figure 4; Supplementary Table 9). The combined meta-analysis for 'eGFR' highlighted a novel genome-wide significant association involving a cluster of variants on chromosome 2 led by rs1974990 (EAF 8%; β 4.07 [2.61, 5.52] mL/min/1.73m^2; $p=4.8\times10^{-8}$) and mapping near *SSB* (encoding Sjogren syndrome antigen B) (Table 2).

As in the T2D-only analysis, we selected 47 loci for replication (30 with $p<1\times10^{-6}$ with at least one of the DKD phenotypes) from the combined (T1D+T2D) GWAS, and an additional 17 loci from the equivalent analysis of 'eGFR'. The combined association *p*-value for rs9942471 ('microalbuminuria'; OR 1.10 [1.02, 1.19]; p=0.001) did not reach the threshold for replication. Lead variants at these 47 loci were tested for all DKD phenotypes available in the relevant

replication samples in subjects with T1D or T2D (Supplementary Table 1): meta-analysis results were only reported for those phenotypes that contributed to discovery-stage associations. This joint, combined (T1D+T2D) analysis generated a substantially enlarged data set for the 'all-DKD' phenotype (40,640 subjects [18,582 cases]) (Figure 1). However, none of the variants selected for replication from the dichotomous phenotypes reached genome-wide significance ($p \leq 5\times10^{-8}$).

The joint, combined analysis for 'eGFR' in subjects European and Asian descent included 31,562 subjects, and replicated known associations near *UMOD* (rs11864909; $\beta_{Asian+Euro}$ 2.11 [1.52, 2.70]; $p=2.3\times10^{-12}$) and *PRKAG2* (rs10224002 $\beta_{Asian+Euro}$ 2.01 [1.30, 2.72], $p=2.7\times10^{-8}$) (Table 2; Supplementary Figures 5 and 6). The *PRKAG2* was non-significant ($p \leq 5\times10^{-8}$) in individual analyses of 'eGFR' in T2D-only (β_{Euro} 2.13 [1.28, 2.98]; $p=8.5\times10^{-7}$) or T1D-only (β_{Euro} 1.23 [-0.19, 2.65]; p=0.09) and effect sizes did not differ by type of diabetes.

The association at *SSB*, detected in the combined 'eGFR' analysis, did not replicate (rs1974990, β 0.04[-2.69, 2.76] mL/min/1.73m^2; p=0.98) and, in the joint, combined analysis was no longer genome-wide significant ($\beta_{Asian+Euro}$ 3.17 [1.88, 4.45] mL/min/1.73m^2; $p=1.4\times10^{-6}$) (Table 2).

Evaluating previous association claims: Of the 61 published loci, for which there are published claims of association with T1D-DKD or T2D-DKD (8), 55 of these associations were represented by variants contributing to our meta-analyses of DKD phenotypes in either subjects with T2D-only or T1D+T2D. Two of these, the 'eGFR' associations at *UMOD* and *PRKAG2,* replicate at genome-wide significance in our data (Table 2). We tested the association of the remaining 53 lead variants in the T2D-only and combined analyses (Supplementary Figure 6). Fourteen variants were associated with a DKD phenotype corresponding to the original report at nominal significance ($p<0.05$) but only 10 of these were directionally consistent with previous reports

(Supplementary Table 10). At a more stringent significance level ($p<9\times10^{-4}$) that accounts for the 55 variants tested (but not the multiple phenotypic categories), only two variants were associated with a DKD phenotype that corresponded to the original report, both of them in the combined (T1D+T2D) analysis, and both directionally consistent with previous reports. These two SNPs were rs2838302, near *SIK1,* associated with 'ESRD vs no ESRD' (EAF 8%; OR 1.39 [1.12, 1.74]; $p=3.9\times10^{-4}$) and rs7583877, near *AFF3*, associated with 'ESRD versus no ESRD' (OR 1.22 [1.13, 1.32]; $p=4.8\times10^{-4}$) (Supplementary Table 10). When we took account of the substantial participant overlap between the original reports and the samples in the present study, apparent replications failed to reach nominal (p<0.05) significance (though, for these, the sample sizes available for independent replication were often small). Thus, other than the 'eGFR' associations at *UMOD* and *PRKAG2,* we found limited evidence in this study to corroborate previously-reported DKD associations, despite, for most variants, sample sizes considerably larger than those included in the original report. Validation of previously-reported DKD associations could be complicated by differences in phenotype definitions and/or analytical methods between this study and published reports. We could not assess whether the *UMOD* or *PRKAG2* allelic effects were different in this study compared to those reported by CKDGen consortium as the allelic effects were not on the same scale (e.g. untransformed vs log transformed).

Genetic overlap with risk factors: Several exposures and diseases have been reported to increase DKD risk in epidemiological studies (1, 2, 43). To explore the extent to which these reflect shared genetic background, we constructed weighted genetic risk scores (GRS) for twenty traits related to diabetes (37, 39-41), insulin resistance (38), obesity (34, 35), hypertension (42), coronary artery disease (33), and lipids (36). These GRS, constructed from

signals identified ($p<5\times10^{-8}$) in previously-published GWAS, included between 10 and 96 SNPs per phenotype. We tested the association of these GRS with each of the DKD phenotypes from this study, in both T2D-only and combined (T1D+T2D) data sets (42).

After Bonferroni correction ($p\leq2.5\times10^{-3}$, which accounts for the number of trait GRS but not the number of DKD phenotypes): In subjects with T2D a GRS for increased waist-to-hip ratio (WHR) ($p=4.8\times10^{-4}$) was associated with increased risk of 'ESRD vs no ESRD'; and a GRS for increased BMI was associated with 'all DKD' ($p=1.8\times10^{-4}$) and 'late DKD' ($p=1.8\times10^{-3}$) phenotypes. A similar pattern of association for the BMI GRS was observed in the combined (T1D+T2D) 'all DKD' analysis ($p=2.4\times10^{-5}$) (Supplementary Table 11 and Figure 3). This last result survives additional correction ($\alpha=1.6\times10^{-4}$) for the 16 DKD phenotypic comparisons considered.

There is evidence implicating insulin resistance in the pathogenesis of DKD, and we wanted to understand whether the BMI GRS associations might reflect obesity-related insulin resistance (44, 45). We focused on the effects two alternative GRS for insulin resistance on DKD. The first, comprising lead variants (N=10) associated with increased fasting insulin (BMI-adjusted) (37), was associated with increased risk of ESRD in subjects with T2D ('ESRD vs no ESRD' $p=1.6\times10^{-3}$; 'ESRD vs controls' $p=1.7\times10^{-3}$) (Supplementary Table 11 and Figure 3). The second, comprising lead variants from 53 loci associated with high fasting insulin (BMI-adjusted), low HDL-C and high triglycerides (39), failed to show any association with DKD phenotypes. These findings provide some support for the causal contribution of insulin resistance and obesity to DKD pathogenesis. However, there is potential that some of these effects reflect collider bias (46) and additional larger studies will be required to substantiate this inference.

Discussion

This study represents the largest study of the genetic basis of DKD in subjects with T2D to date, extending previous reports with respect to sample size and range of DKD phenotypes. We aimed to overcome some of the limitations of earlier studies in this area, and to develop insights into the pathogenesis of DKD. Despite sample sizes that exceeded 40,000, the yield of novel discoveries was modest. There were no significant ($p<5x10^{-8}$) genetic associations with 'all DKD' that was best-powered definition on sample size. The relatively large sample size came with increased phenotypic (and likely genetic) heterogeneity: it was for this reason that we examined a range of DKD phenotypes that might offer better power to detect genetic associations with more restricted phenotypic impacts.

This approach successfully identified a novel locus, *GABRR1* (led by rs9942471), for 'microalbuminuria' in European subjects with T2D. The variants, near *GABRR1*, reached a level of significance ($p<5x10^{-8}$) that has typically been associated with robust, reproducible association in common disease GWAS. *GABRR1* expression is upregulated in renal biopsies form DKD subjects (compared to controls) and in other non-diabetic kidney diseases characterised by glomerular scarring and inflammation (47). The variants were associated with *GABRR1* expression in aorta, oesophageal mucosa and skin in GTEx. However, we found no replication of the *GABRR1* association in subjects of European ancestry with T1D-DKD, nor amongst subjects of Asian ancestry with T2D, though differences in risk-allele frequencies between these two ancestries and the modest size of the replication datasets at this locus reduce the power of the latter analysis. Our overall assessment is that this association should be considered provisional until it is possible to undertake further rounds of adequately-

powered replication that could establish the definitive status of this variant and this locus should also be assessed for effects on DKD progression in longitudinal studies.

Even in the absence of specific signals of association with DKD, it is possible to use the aggregate pattern of association across the genome to identify more subtle genetic effects. The GRS analyses described here provide genetic support for the causal contribution of obesity to the development of T2D-DKD. This echoes strong epidemiological data, and mirrors equivalent analyses in T1D-DKD (48, 49). However, we cannot exclude that these associations may partly reflect collider bias (46): subjects with high BMI are likely to have a longer duration of diabetes and thus a higher chance of developing complications. Analyses using genetic instruments (GRS) for variation in insulin sensitivity produced variable results with respect to T2D-DKD, but indicate that the BMI effects may be partially mediated via obesity-related insulin resistance (37). There is substantial epidemiological data to support this link between insulin resistance and DKD risk (44, 45).

The modest yield of association signals, and the limited replication of previous claims of DKD association, emphasises challenges associated with the identification of DKD-risk variants. For many complex traits, these have been overcome through a combination of increased sample size and phenotypic precision. Published genetic association studies of DKD have often used different definitions of DKD, which makes replication of previous findings difficult. In this study, we used phenotype definitions aligned to those used in the study of DKD in subjects with T1D (17). Standardising the phenotype definitions in this way allowed for seamless combination of the GWAS data across the two studies and may streamline subsequent efforts to study the genetics of DKD. The phenotype definitions applied to this study address some of the challenges associated with increasing sample size while maintaining phenotype precision,

and should, in due course, support the identification of robust associations with DKD. It is clear that these phenotype definitions are not without limitations, in the absence of strong genetic signals we have few clues to which particular diagnostic configurations will be most productive for genetic discovery. Targeting the phenotypes that show the greatest heritability may provide a guide (14).

Acknowledgments

Author Contributions

Central data analysis was performed by: N.R.v.Z., E.A., N.S., N.W.R., D.Z., E.F., S.C. and M.I.M. Data generation was performed by N.R.v.Z., E.A., H.D., C.L., F.S.c., M.K., J.L., G.J., A.O.Y.L., H.M.L., C.K.P.L., J.C.N.C., H.K.D.R.T.P.G., S.F.A., R.D., T.S.H.C., A.-J.M., T.W.3.G.S.G., G.C., S.H., D.g., T.S.A., M.O.-M., A.L., C.C., N.G., I.B., O.M., S.C.L., R.C.W.M., V.L., S.S.R., J.C.F., O.P., T.H., S.C., C.N.A.P., P.-H.G., L.C.G. and M.I.M. Individual study design was performed by E.A., N.S., H.D., P.M.M., C.F., N.P., R.M.v.D., G.J., J.C.N.C., M.M., M.O.-M., A.L., C.C., D.R.W., I.B., S.C.L., R.C.W.M., E.S.T., S.M., T.T., O.P., T.H., G.R., S.C., M.J.B., C.N.A.P., P.-H.G., H.M.C. and L.C.G. Local data analysis was peformed by N.R.v.Z., E.A., N.S., H.D., N.W.R., C.L., N.R.R., P.M.M., E.V., A.P., R.P.I.Jr, N.P., M.I., A.T., X.S., J.L., G.J., J.C.N.C., H.K.D.R.T.P.G., S.F.A., R.D., L.W., A.-J.M., S.D., M.G.P., G.C., M.M., S.H., L.T.H., T.S.A., P.A., C.-A.S., O.M., A.D.P., D.T., A.P.M., S.C.L., R.C.W.M., V.L., S.S.R., J.C.F., G.R., S.C., C.N.A.P. and H.M.C. The paper was prepared by N.R.v.Z., E.A., N.S., M.A., N.R.R., M.L., J.C.N.C., L.T.H., A.D.P., S.C.L., R.C.W.M., J.C.F., P.R., S.C., H.M.C., L.C.G. and M.I.M. Sample collection was conducted by C.F., V.H., F.S.c., E.R., M.L.M., N.P., M.L., R.M.v.D., A.O.Y.L., C.K.P.L., C.C.S., W.Y.S., J.C.N.C., H.K.D.R.T.P.G., S.F.A., T.S.H.C., A.-J.M., T.W.3.G.S.G., G.C., M.M., S.H., D.g., M.O.-M., A.L., C.C., D.R.W., I.B., O.M., A.P.M., S.C.L., R.C.W.M., E.S.T., V.L., T.T., A.S.K., S.S.R., J.C.F., D.D., O.P., T.H., P.R., G.R., S.C., C.N.A.P., P.-H.G., H.M.C., L.C.G., A.K., G.J., A.P.M., R.C.W.M., E.S.T. and L.C.G. N.R.v.Z and M.I.M are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Duality of Intereast

P. R. has given lectures for Astra Zeneca, BMS and Boehringer Ingelheim, has served as a consultant for AbbVie, Astra Zeneca, BMS, Eli Lilly, Boehringer Ingelheim, Astellas, Janssen, and

Novo Nordisk; all fees were given to the Steno Diabetes Center that has equity interest in Novo Nordisk. E.F. is an employee of and owns stock in Pfizer, Inc. W.Y.S. is co-founder of GemVCare, established under the Technology Start-up Support Scheme for Universities (TSSSU) from the Hong Kong Government Innovation and Technology Commission. J.C.N.C. is co-founder of GemVCare, established under the Technology Start-up Support Scheme for Universities (TSSSU) from the Hong Kong Government Innovation and Technology Commission. R.C.W.M. is co-founder of GemVCare, established under the Technology Start-up Support Scheme for Universities (TSSSU) from the Hong Kong Government Innovation and Technology Commission. J.C.F. has received a consulting honorarium from Merck. P.-H.G. has received lecture honoraria from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Genzyme, MSD, Novartis, Novo Nordisk, and Sanofi, and research grants from Eli Lilly and Roche. P-HG is also an advisory board member for AbbVie, Boehringer Ingelheim, Eli Lilly, Janssen, Medscape, MSD, Novartis and Sanofi. M.I.M. serves on advisory panels for Pfizer and Novo Nordisk; has received honoraria from Lilly, Pfizer, and Novo Nordisk; and M.I.M has received research support from Lilly, Pfizer, Novo Nordisk, Servier, Takeda, Roche, Merck, Janssen, Abbvie, Boehringer Ingelheim, Astra Zeneca and Sanofi Aventis.

Funding

The research was supported by the European Union's Seventh Framework Program (FP7/2007–2013) for the Innovative Medicine Initiative under grant agreement IMI/115006 (the SUMMIT consortium); Academy of Finland (grants 263401 and 267882); ADRV Paris; Agency for Science & Technology and Research (A*STAR) of Singapore; Albert Påhlsson Foundation and Diabetesfonden; Alexandra Health (Private Limited) (SIGII/08005; SIGII/11001; SIG/11029; SIG/12024; SIG II/15205); Chinese University of Hong Kong focused investment

scheme; Danish Diabetes Academy; DOLOrisk (European Union's Horizon 2020 research and innovation programme grant No 633491); ERC-Adv res grant 269045-GENE TARGET T2D; Ernhold Lundström; European Research Council (Consolidator grant nr 649021, Orho-Melander); Finska Läkaresällskapet; Folkhälsan Research Foundation; French Ministry of Health; Heart foundation of Jakobstad region; Helsinki University Central Hospital Research Funds (EVO); the Hong Kong Food and Health Bureau; (01120796) Hong Kong Foundation for Research and Development in Diabetes; Hong Kong Government Research Grant Committee and Innovation and Technology Grant Committee; Hong Kong Research Grants Council Theme-based Research Scheme (T12-402/13N); Japan Agency for Medical Research and Development; Juvenile Diabetes Research Foundation (17-2012-542, 17-2013-7, 2-SRA-2014-276-Q-R and 17-2013-9); Knut and Alice Wallenberg Foundation; Leading Project of Ministry of Education, Culture, Sports, Science and Technology, Japan; Liao Wun Yuk Memorial Fund; Linneus Foundation for the Lund University Diabetes Center; Liv och Hälsa Foundation; Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese Government; National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre; National Medical Research Council, Singapore: PPG/AH(KTPH)/2011; CIRG13nov045; Natural Sciences and Engineering Research Council of Canada; NHS Tayside; National Institute of Diabetes and Digestive and Kidney Diseases (R01-DK105154); National Institute for Health (R01-MH101814); Novo Nordisk Foundation (NNF14SA0003 and NNF15CC0018486); Påhlsson Foundation; Region Skåne; Rhodes Trust; Signe and Ane Gyllenberg Foundation; Sigrid Juselius Foundation; Skåne University Hospital; Société Francophone du Diabète; Swedish Diabetes Foundation; Swedish Heart and Lung Foundation; Swedish Research Council; Turku University Hospital Research Funds; The University of Dundee; Vasa Hospital district; Wellcome (072960/Z/03/Z,

084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z, 098381, 090532 and 106310); and Wilhelm and Else Stockmann Foundation.

References

1. Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev. 2013;93(1):137-88.
2. Krolewski AS, Skupien J, Rossing P, Warram JH. Fast renal decline to end-stage renal disease: an unrecognized feature of nephropathy in diabetes. Kidney Int. 2017;91(6):1300-11.
3. Seaquist ER, Goetz FC, Rich S, Barbosa J. Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. N Engl J Med. 1989;320(18):1161-5.
4. Quinn M, Angelico MC, Warram JH, Krolewski AS. Familial factors determine the development of diabetic nephropathy in patients with IDDM. Diabetologia. 1996;39(8):940-5.
5. Vijay V, Snehalatha C, Shina K, Lalitha S, Ramachandran A. Familial aggregation of diabetic kidney disease in Type 2 diabetes in south India. Diabetes Res Clin Pract. 1999;43(3):167-71.
6. Pettitt DJ, Saad MF, Bennett PH, Nelson RG, Knowler WC. Familial predisposition to renal disease in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia. 1990;33(7):438-43.
7. Freedman BI, Spray BJ, Tuttle AB, Buckalew VM, Jr. The familial risk of end-stage renal disease in African Americans. Am J Kidney Dis. 1993;21(4):387-93.
8. Mooyaart AL, Valk EJ, van Es LA, Bruijn JA, de Heer E, Freedman BI, et al. Genetic associations in diabetic nephropathy: a meta-analysis. Diabetologia. 2011;54(3):544-53.
9. Pezzolesi MG, Poznik GD, Mychaleckyj JC, Paterson AD, Barati MT, Klein JB, et al. Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. Diabetes. 2009;58(6):1403-10.
10. McKnight AJ, Currie D, Patterson CC, Maxwell AP, Fogarty DG, Warren UKGSG. Targeted genome-wide investigation identifies novel SNPs associated with diabetic nephropathy. The HUGO Journal. 2009;3(1-4):77-82.
11. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, et al. New susceptibility loci associated with kidney disease in type 1 diabetes. PLoS Genet. 2012;8(9):e1002921.
12. Pham TT, Sim JJ, Kujubu DA, Liu IL, Kumar VA. Prevalence of nondiabetic renal disease in diabetic patients. Am J Nephrol. 2007;27(3):322-8.
13. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun. 2016;7:10023.
14. Boger CA, Sedor JR. GWAS of diabetic nephropathy: is the GENIE out of the bottle? PLoS Genet. 2012;8(9):e1002989.
15. Placha G, Canani LH, Warram JH, Krolewski AS. Evidence for different susceptibility genes for proteinuria and ESRD in type 2 diabetes. Adv Chronic Kidney Dis. 2005;12(2):155-69.
16. Ellis JW, Chen MH, Foster MC, Liu CT, Larson MG, de Boer I, et al. Validated SNPs for eGFR and their associations with albuminuria. Hum Mol Gen. 2012;21(14):3293-8.

17. Sandholm N, Van Zuydam N, Ahlqvist E, Juliusdottir T, Deshmukh HA, Rayner NW, et al. The Genetic Landscape of Renal Complications in Type 1 Diabetes. JASN. 2017;28(2):557-74.
18. Todd JN, Salem R, Sandholm N, Valo EA, Hiraki LT, Di Liao C, et al. Novel Genetic Determinants of Diabetic Kidney Disease. Diabetes. 2016;65(S1):A100.
19. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh CD. Classifying diabetes according to the new WHO clinical stages. Eur J Epidemiol. 2001;17(11):983-9.
20. Morris AD, Boyle DI, MacAlpine R, Emslie-Smith A, Jung RT, Newton RW, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. BMJ. 1997;315(7107):524-8.
21. Rossing P, Hougaard P, Parving HH. Risk factors for development of incipient and overt diabetic nephropathy in type 1 diabetic patients: a 10-year prospective observational study. Diabetes Care. 2002;25(5):859-64.
22. Ruggenenti P, Remuzzi G. Nephropathy of type 1 and type 2 diabetes: diverse pathophysiology, same treatment? Nephrol Dial Transplant. 2000;15(12):1900-2.
23. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.
24. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013;10(1):5-6.
25. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature Genet. 2012;44(8):955-9.
26. Bradic M, Costa J, Chelo IM. Genotyping with Sequenom. Methods Mol Biol. 2011;772:193-210.
27. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82.
28. Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet. 2014;10(4):e1004269.
29. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature Genet. 2007;39(7):906-13.
30. Kang H, Sul J, Service SK, Zaitlen NA, Kong S-y, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nature Genet. 2010;42(4):348-54.
31. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
32. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics. 2010;11:288.
33. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nature Genet. 2013;45(1):25-33.
34. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genet. 2010;42(11):937-48.

35. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nature Genet. 2010;42(11):949-60.
36. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nature Genet. 2008;40(2):161-9.
37. Scott RA, Fall T, Pasko D, Barker A, Sharp SJ, Arriola L, et al. Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. Diabetes. 2014;63(12):4378-87.
38. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature Genet. 2012;44(9):991-1005.
39. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nature Genet. 2017;49(1):17-26.
40. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nature Genet. 2009;41(6):703-7.
41. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nature Genet. 2014;46(3):234-44.
42. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9.
43. Hill CJ, Cardwell CR, Maxwell AP, Young RJ, Matthews B, O'Donoghue DJ, et al. Obesity and kidney disease in type 1 and 2 diabetes: an analysis of the National Diabetes Audit. QJM. 2013;106(10):933-42.
44. Groop L, Ekstrand A, Forsblom C, Widen E, Groop PH, Teppo AM, et al. Insulin resistance, hypertension and microalbuminuria in patients with type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia. 1993;36(7):642-7.
45. Karalliedde J, Gnudi L. Diabetes mellitus, a complex and heterogeneous disease, and the role of insulin resistance as a determinant of diabetic kidney disease. Nephrol Dial Transplant. 2016;31(2):206-13.
46. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. PLoS Genet. 2017;13(10):e1006944.
47. Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, et al. Defining cell-type specificity at the transcriptional level in human disease. Genome Res. 2013;23(11):1862-73.
48. Maric-Bilkan C. Obesity and diabetic kidney disease. Med Clin North Am. 2013;97(1):59-74.
49. Todd JN, Dahlstrom EH, Salem RM, Sandholm N, Forsblom C, FinnDiane Study G, et al. Genetic Evidence for a Causal Role of Obesity in Diabetic Kidney Disease. Diabetes. 2015;64(12):4238-46.

Table 1: Genome-wide association study characteristics by diabetic kidney disease phenotypes in subjects with type 2 and type 1 diabetes

Analysis	Case definition	Control definition	Subjects with type 2 diabetes		Subjects with type 1 diabetes	
			#Cases	#Controls	#Cases	#Controls
All Diabetic kidney disease (DKD)	**All DKD**: Microalbuminuria OR Late DKD OR end-stage renal disease (ESRD)	Normoalbuminuria (Albumin excretion rate [AER] <20 µg/min OR AER <30 mg/24 h OR ACR <2.5/3.5 mg/mmol for men/women) AND duration of T2D >10 years‡	3,345	2,372	2,563	2,593
Microalbuminuria*	**Microalbuminuria**: At least 2 out of 3 consecutive measurements with albumin excretion rate (AER) ≥20 AND <200 µg/min OR AER ≥30 AND <300 mg/24 hr OR albumin to creatinine ratio (ACR) ≥2.5/3.5 AND <25/35 mg/mmol for men/women;	Normoalbuminuria AND duration of T2D >10 years‡	1,989	2,238†	806	2,593
Late DKD	**Late DKD**: At least one measurement with AER ≥200 µg/min OR AER ≥300 mg/24 h OR ACR ≥25/35 mg/mmol for men/women) or end-stage renal disease (ESRD, estimated glomerular filtration rate (eGFR) < 15 mL/min/1.73m^2 OR kidney transplantation OR dialysis)	Normoalbuminuria AND duration of T2D >10 years‡	1,339	2,372	1,757	2,593
ESRD vs. controls	**ESRD**: eGFR<15 mL/min/1.73m^2 or renal dialysis or kidney transplant	No DKD AND duration of T2D >10 years‡	371	2,076	813	2,398
ESRD vs. no ESRD	ESRD (see above)	No ESRD AND duration of T2D >10 years‡	371	4,471	813	3,995
Chronic Kidney Disease (CKD)	**CKD**: eGFR < 60 mL/min/1.73m^2	No CKD AND duration of T2D >10 years‡	3,094	2,906	2,460	774
CKD and DKD	**CKD and DKD**: eGFR < 45 mL/min/1.73m^2 AND all DKD	No CKD AND no ESRD AND normoalbuminuria AND duration of T2D >10 years‡	897	1,610	1,750	1,385
eGFR	32788 x Serum Creatinine(µmol/L)$^{-1.154}$ x Age$^{-0.203}$ x [0.742 if female] (mL/min/1.73m^2)		9,197		3,961	

*Equivalent to the 'early DKD' phenotype from Sandholm et al., 2017 (17); †Not all studies were able to define microalbuminuria (due to limited information on microalbuminuric status) thus the case and control number is smaller than 'all DKD' and 'late DKD'; ‡The duration of diabetes for subjects with T1D was >15 years

Table 2: Five loci were associated ($p \leq 5\times10^{-8}$) with chronic kidney disease ('CKD'), 'microalbuminuria' and estimated glomerular filtration rate ('eGFR') in subjects with type 2 diabetes (T2D) or the combined analysis of T2D and type 1 diabetes (T1D+T2D)

CHR:BP	Phenotype	SNP *Locus*	Discovery			Replication			Joint analysis		
			EA/NEA (Info) EAF	OR/Beta (95%CI)	P	Ancestry	OR/Beta (95%CI)	P	OR/Beta (95%CI)	P	N
20: 9351150	T2D 'CKD'	rs2206136 (*PLCB4*)	A/T (0.98) 0.42	1.20 (1.08-1.34)	2.1×10^{-8}	European	1.02 (0.91,1.15)	0.69	1.13 (1.05,1.21)	9.0×10^{-5}	11,900
						Asian and European	1.03 (0.94,1.13)	0.68	1.12 (1.05,1.19)	2.1×10^{-4}	13,813
6: 89948232	T2D 'microalbu-minuria'	rs9942471 (*GABRR1*)	A/C (0.99) 0.64	1.24 (1.15-1.34)	2.1×10^{-7}	European	1.32 (0.99,1.75)	0.06	1.25 (1.16,1.34)	4.5×10^{-8}	4,801
						Asian and European	1.11 (0.99,1.23)	0.12	1.15 (1.08,1.23)	1.2×10^{-5}	5,559
16: 20400839	T2D 'eGFR'	rs11864909 (*UMOD*)	T/C (1.00) 0.28	2.42 (1.28-3.56)	2.7×10^{-5}	European	2.22 (1.16,3.28)	4.1×10^{-5}	2.31 (1.54,3.09)	4.6×10^{-9}	12,343
						Asian and European	2.30 (1.48,3.12)	3.6×10^{-8}	2.34 (1.68,3.00)	4.4×10^{-12}	19,747
2: 170646916	T1D+T2D 'eGFR'	rs1974990* (*SSB*)	G/T (0.98) 0.08	4.07 (2.61,5.52)	4.8×10^{-8}	European	No replication available		4.07 (2.61,5.52)	4.8×10^{-8}	13,158
						Asian and European	0.04 (-2.69,2.76)	0.98	3.17 (1.88,4.45)	1.4×10^{-6}	14,828
7: 151415041	T1D+T2D 'eGFR'	rs10224002 (*PRKAG2*)	A/G (0.92) 0.74	1.75 (0.85-2.66)	1.5×10^{-4}	European	2.15 (0.93-3.37)	5.8×10^{-4}	1.89 (1.17,2.62)	3.4×10^{-7}	20,495
						Asian and European	2.42 (1.28,3.56)	3.2×10^{-5}	2.01 (1.30, 2.72)	2.7×10^{-8}	22,165
16: 20400839	T1D+T2D 'eGFR'	rs11864909 (*UMOD*)	T/C (0.99) 0.29	1.90 (1.05-2.74)	1.1×10^{-5}	European	2.22 (1.16,3.28)	4.1×10^{-5}	2.02 (1.36,2.69)	2.1×10^{-9}	16,304
						Asian and European	2.30 (1.48,3.12)	3.6×10^{-8}	2.11 (1.52,2.70)	2.3×10^{-12}	23,708

*rs1974490 was only available in the 1000G reference panel and was not imputed in the European studies used in the replications

Figures

Figure 1: Eight diabetic kidney disease (DKD) phenotypes were analysed in subjects with type 2 diabetes (T2D, blue boxes) and in a combined (green boxes) analysis of subjects with T2D or type 1 diabetes (T1D, yellow box). N indicates the total sample count for either the 'all DKD' (number of cases are given in brackets) or the 'eGFR' phenotypes and may vary by variant as well as by DKD phenotype. Replication was sought for 164 loci and 47 loci from each analysis respectively in subjects of European and Asian ancestry with either T1D or T2D.

Figure 2: A) Manhattan plot of p values from the meta-analysis of allelic effect on 'early diabetic kidney disease' in subjects with type 2 diabetes of European descent. The red line represents genome-wide significance ($p<5x10^{-8}$) and the blue line suggestive significance ($p<1x10^{-6}$). The peak represented by rs9942471 ($p=4.5x10^{-8}$), near *GABRR1* is highlighted in orange; B) A forest plot of allelic odds ratio (OR) and imputation information scores (RSQ) from individual studies (Study) that contributed to the discovery and replication (DIREVA) analyses of rs9942471 in 'microalbuminuria'; Rs9942471 genotypes were not available in Steno; and C) a Locuszoom plot of the signal near near *GABRR1* led by rs9942471 that was associated with early diabetic kidney disease in European subjects with T2D.

Figure 3: A heat map of genetic risk score associations with diabetic kidney disease (DKD) phenotypes in subjects with either type 1 diabetes or type 2 diabetes. A GRS for body mass index was significant after correction for multiple testing while other traits including systolic blood pressure were not associated with DKD phenotypes. Abbreviations used: chronic kidney disease ('CKD'), end stage renal disease ('ESRD') and estimated glomerular filtration rate ('eGFR').