HELDA
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Association of Coding Variants in Hydroxysteroid 17-beta Dehydrogenase 14 (HSD17B14) with Reduced Progression to End Stage Kidney Disease in Type 1 Diabetes

Mychaleckyj, Josyf C.

2021-10

Mychaleckyj , J C , Valo , E , Ichimura , T , Ahluwalia , T S , Dina , C , Miller , R G , Shabalin , I G , Gyorgy , B , Cao , J , Onengut-Gumuscu , S , Satake , E , Smiles , A M , Haukka , J K , Tregouet , D-A , Costacou , T , O'Neil , K , Paterson , A D , Forsblom , C , Keenan , H A , Pezzolesi , M G , Pragnell , M , Galecki , A , Rich , S S , Sandholm , N , Klein , R , Klein , B E , Susztak , K , Orchard , T J , Korstanje , R , King , G L , Hadjadj , S , Rossing , P , Bonventre , J V , Groop , P-H , Warram , J H & Krolewski , A S 2021 , ' Association of Coding Variants in Hydroxysteroid 17-beta Dehydrogenase 14 (HSD17B14) with Reduced Progression to End Stage Kidney Disease in Type 1 Diabetes ' , Journal of the American Society of Nephrology , vol. 32 , no. 10 , pp. 2634-2651 . https://doi.org/10.1681/ASN.2020101457

http://hdl.handle.net/10138/353497
https://doi.org/10.1681/ASN.2020101457

unspecified
acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Association of coding variants in hydroxysteroid 17-beta dehydrogenase 14 (*HSD17B14*) with reduced progression to end stage kidney disease in type 1 diabetes

Josyf C. Mychaleckyj[1], Erkka Valo[2,3,4], Takaharu Ichimura[5], Tarunveer S. Ahluwalia[6], Christian Dina[7], Rachel G. Miller[8], Ivan G. Shabalin[9], Beata Gyorgy[10], JingJing Cao[11], Suna Onengut-Gumuscu[1], Eiichiro Satake[12,13], Adam M. Smiles[12], Jani K. Haukka[2,3,4], David-Alexandre Tregouet[10,14], Tina Costacou[8], Kristina O'Neil[12], Andrew D. Paterson[11], Carol Forsblom[2,3,4], Hillary A. Keenan[12,13], Marcus G. Pezzolesi[12,13,15], Marlon Pragnell[16], Andrzej Galecki[17], Stephen S. Rich[1], Niina Sandholm[2,3,4], Ronald Klein[18], Barbara E. Klein[18], Katalin Susztak[19], Trevor J. Orchard[8], Ron Korstanje[20], George L. King[12,13], Samy Hadjadj[21,22], Peter Rossing[6,23], Joseph V. Bonventre[5], Per-Henrik Groop[2,3,4,24], James H. Warram[12], Andrzej S. Krolewski[12,13]

Author Affiliations:

[1] Center for Public Health Genomics, University of Virginia, Charlottesville, VA

[2] Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland

[3] Department of Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[4] Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Finland

[5] Renal Division, Brigham and Women's Hospital, Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

[6] Steno Diabetes Center Copenhagen, Copenhagen, Denmark

[7] Université de Nantes, CNRS INSERM, l'institut du thorax, F-44000, Nantes, France

[8] Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA

[9] Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA

[10] INSERM UMRS1166, ICAN - Institute of CardioMetabolism and Nutrition, Sorbonne Université, Paris, France

[11] Genetics & Genome Biology Research Institute, SickKids Hospital, Toronto, Canada

[12] Research Division, Joslin Diabetes Center, Boston, MA

[13] Department of Medicine, Harvard Medical School, Boston, MA

[14] Université de Bordeaux, INSERM, BPH, U1219, F-33000 Bordeaux, France

[15] Division of Nephrology and Hypertension, University of Utah, Salt Lake City, UT

[16] JDRF International, New York, NY

[17] Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI

[18] Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI

[19] Department of Medicine and Genetics, University of Pennsylvania, Philadelphia, PA
[20] The Jackson Laboratory, Bar Harbor, ME
[21] INSERM CIC 1402 and U 1082, Poitiers, France
[22] Department of Endocrinology, l'institut du thorax, INSERM, CNRS, UNIV Nantes, CHU Nantes, Nantes, France
[23] University of Copenhagen, Copenhagen, Denmark
[24] Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria, Australia

Running Title: *HSD17B14* gene variants and T1D ESKD

Abstract Word Count: 268
Text Word Count: 4229

Corresponding Authors:
Josyf C. Mychaleckyj MA, DPhil
Center for Public Health Genomics,
University of Virginia, Charlottesville, VA
Tel: 434-982-1107
Email: jcm6t@virginia.edu

or

Andrzej S. Krolewski MD, PhD
Section on Genetics & Epidemiology
Research Division, Joslin Diabetes Center
Boston, MA 02215
Tel: 617-309-2668
Email: andrzej.krolewski@joslin.harvard.edu

Keywords:
type 1 diabetes; hydroxysteroid 17-beta dehydrogenase 14; end stage kidney disease; gene-based tests; rare variants; diabetic nephropathy; diabetic kidney disease

ABSTRACT

Background: Rare genetic variants in gene coding regions likely have a greater impact on disease-related phenotypes than common variants through disruption of their encoded protein. We searched for rare variants associated with onset of end stage kidney disease (ESKD) in individuals with type 1 diabetes at advanced kidney disease stage.

Methods: We performed gene-based exome array analysis of 15,449 genes in 5 large incidence cohorts of individuals with type 1 diabetes and proteinuria analyzed for survival time-to-ESKD, testing the top gene in a 6th cohort (N=2,372/1,115 events all cohorts) and replicating in two retrospective case-control studies (N=1,072 cases, 752 controls). We followed up by deep resequencing of the top associated gene in 5 cohorts. We performed immunohistochemistry and gene expression experiments in human control and diseased cells, and in mouse ischemia reperfusion and aristolochic acid nephropathy models.

Results: Protein coding variants in hydroxysteroid 17-beta dehydrogenase 14 gene (*HSD17B14),* predicted to affect protein structure, had a net protective effect against development of ESKD at exome-wide significance (N=4,196; p-value=3.3×10^{-7}). The *HSD17B14* gene and encoded enzyme were robustly expressed in healthy human kidney, and highest in proximal tubular cells. Paradoxically, gene and protein expression were attenuated in human diabetic proximal tubules and in mouse kidney injury models. Expressed *HSD17B14* gene and protein levels remained low without recovery after 21 days in a murine ischemic reperfusion injury model. Decreased gene expression was found in other chronic kidney disease-associated renal pathologies.

Conclusions: We have discovered that the *HSD17B14* gene is mechanistically involved in diabetic kidney disease. The encoded sex steroid enzyme is a druggable target, potentially opening a new avenue for therapeutic development.

INTRODUCTION

Diabetic kidney disease (DKD) is one of the most prevalent, costly, and devastating for quality of life, of all the complications of type 1 diabetes (T1D)[1,2]. It affects about a third of individuals with T1D[3,4] with many of them developing end stage kidney disease (ESKD). Despite improvements in glycemic control and almost universal implementation of reno-protective therapies, the incidence of new cases of ESKD in T1D has not changed over the last 20 years[2]. Clearly, further knowledge is needed about the disease process underlying the development of DKD and the mechanisms of progression to ESKD, in order to develop more effective interventions to reduce the risk of ultimate kidney failure.

Recently it has been recognized that in addition to elevated urinary albumin excretion, the most significant clinical feature of DKD that predicts onset of ESKD is progressive decline in kidney function[5,6]. Although renal decline is progressive, there is profound heterogeneity in the rate at which different patients lose renal function, as measured by GFR slope and time to develop ESKD[6,7]. The disease process underlying this heterogeneity seems to be multifactorial and includes among other factors, variation in exposures such hyperglycemia, hypertension, and genetics[8].

To date, genome-wide discovery studies in DKD have almost universally focused on testing individual common SNPs to identify associations[9–13]. Progress using this approach has been slow, and published results from individual studies and consortium efforts have not been consistently replicated[9,10]. Possible reasons include modest sample sizes, pragmatic but imprecise case and control definitions, and phenotypes that do not adequately interrogate the heterogeneous nature of T1D kidney function decline[14–17]. Furthermore, most common variants are in non-

protein coding segments of the genome[18] that are difficult to functionally characterize. They may influence the expression of remote genes that are 100-1000s of kilobases distant from the landmark SNP[19].

More recently, a new approach has been developed to test the aggregate association of multiple rare variants within gene coding and splicing regions[20,21]. The basis of this method is the expectation that rare variants are likely to have larger individual effects on phenotypes through direct disruption of an encoded protein, while aggregate testing of all of the variants in a gene improves the statistical power. A genome scan using this method tests each individual gene in the genome (approx. 20,000) for net association of the variants with the phenotype. Two types of models are tested: 1) assume all the variants in a gene have the same direction of effect on the phenotype, either risk or protective; or 2) variants in a gene can act in opposite directions, some increasing risk and some protecting. Statistical power is also improved as a result of the reduced multiple testing correction (20,000 gene tests versus 1 million or more SNPs). Finally, since the variants are at least partially likely to act through direct disruption of the gene they are located within, the target protein and its biological actions are a logical starting point for functional interpretation, and for more detailed tissue and cellular experimentation.

Under the auspices of the JDRF Diabetic Nephropathy Collaborative Research Initiative (JDRF DNCRI), specifically the sub-project entitled 'Genes determining time of onset of ESRD in type 1 diabetes individuals with proteinuria', we assembled one of the largest longitudinal multi-cohort collections to investigate genetic influences contributing to fast progression to ESKD in cohorts with advanced DKD. By analogy with other chronic diseases[22–24] we reasoned that focusing on rare variants of larger effect size could more quickly lead to direct functional insight

compared to associated common variants and that a more precise phenotype that captured the heterogeneity in the rate of decline to ESKD would best exploit the power available in the examined cohorts with long-term follow-up.

METHODS

Study Design and Participants

We assembled 6 T1D cohorts, with a baseline of prevalent proteinuria as a biomarker of advanced DKD, to study their rates of progression to ESKD. The recruitment, follow-up, and renal function measurements in four of the six international clinical cohorts (Joslin Kidney Study: USA [Joslin]; Finnish Diabetic Nephropathy Study: Finland [FinnDiane]; T1D patients from Steno Diabetes Center: Denmark [Steno]; INSERM: France) have been described in detail previously. Briefly, individuals with T1D were eligible for the genetic screening cohorts if they had persistent proteinuria at baseline, generally defined as two urinary measurements out of three consecutive, or two within three years, having >300mg/g, but with slight differences between the cohorts, as described[8]. Undiagnosed ESKD events were defined as the first occurrence of eGFR < 10 mL/min/1.73m^2. Two additional US-based cohorts (Pittsburgh Epidemiology of Diabetes Complications Study: USA [EDC][25]; and Wisconsin Epidemiologic Study of Diabetic Retinopathy: USA [WESDR][26]) included individuals with less stringent inclusion criteria, to maximize sample size, of incident proteinuria (first occurrence >300mg/g) with an accompanying eGFR measurement and at least one other later eGFR measurement where the individual's diagnosed and undiagnosed ESKD status was known. Data from EDC was combined with Joslin, FinnDiane, Steno, and INSERM in the Discovery Stage, whilst WESDR, with limited genetic data, provided data for extension and cohort meta-analysis.

Two T1D case-control replication studies were assembled from T1D cases with prevalent ESKD at study recruitment, and controls without clinical evidence of DKD but with very long duration of T1D. In the first study, through a collaboration of the Joslin and the USA Fresenius Dialysis

Centers, East Coast patients with T1D and new-onset ESKD of non-Hispanic white ancestry were randomly selected as cases. They were 20-54 years of age at the initiation of dialysis, had been taking insulin, and had diagnosis of T1D. The controls were selected from among participants in the Joslin T1D Medalist Study. These individuals had a diabetes duration of at least 50 years without clinical evidence of DKD[27].

The second T1D case-control replication group was drawn from the Genetics of Kidneys in Diabetes US Study (GoKinD) recruited through George Washington University (GWU-GoKinD)[28]. Joslin-recruited GoKinD participants were omitted to avoid overlap with the Joslin Kidney Study cohort. Lacking longitudinal time to ESKD for this group and to enrich for more rapid ESKD in cases versus longer duration in controls, cases were selected as having prevalent ESKD at study entry (either dialysis or transplant) with diabetes duration less than 33 years (median of all GWU-GoKinD ESKD), while controls had prevalent normoalbuminuria and diabetes duration greater than 23 years (median of all GWU-GoKinD normoalbuminuria participants). Protocols for recruitment and data collection for participants in the above studies were approved by the relevant Institutional Review Boards or Ethics Committees.

Combined Genome and Exome Array Genotyping and Quality Control

All study groups were genotyped on a combined GWAS plus Exome array (Illumina Human CoreExome, San Diego, CA, USA) at the Center for Public Health Genomics (CPHG) Laboratory, University of Virginia. The genotyping and quality control methods have been described extensively in a previous publication from this JDRF DNCRI Consortium[11]. Briefly, the array contained 250K genome-wide SNPs and other variants, and more than 200K

gene/exome-centered variants. The samples were genotyped in batches using Illumina Gentrain2 algorithm and software, and then re-called with zCall, an algorithm specially designed for calling array-genotyped rare variants[29]. The sample batches were filtered for low quality samples (call rate<98%, gender misclassification, extreme heterozygosity) and variants for quality (call rate < 95%, extreme deviation from Hardy-Weinberg equilibrium). Duplicates and cryptic relatedness was identified using KING[30] and one of each pair was removed. Principal component analysis was performed in each cohort and study group to exclude outliers with evidence of non-European ancestry.

Gene Resequencing and Bioinformatics Processing

DNA samples were selected from five cohorts (Joslin, FinnDiane, Steno, INSERM, and WESDR) that had passed sample quality control (QC) in the prior genotyping array assays. EDC samples were not available for inclusion at the time of Resequencing Stage design. Cohort samples were resequenced at the same CPHG Laboratory using Illumina TruSeq custom amplicon assays. Primers for the amplification of targeted regions were designed using the DesignStudio sequencing assay design tool (Illumina, San Diego, CA, USA). Targeted amplicon libraries, consisting of 151 targets, were constructed using Illumina TruSeq custom amplicon assay kit (version 1.5) according to the manufacturer's protocol (Supplemental Methods and Figure S1). The amplicon sequencing libraries were sequenced with 150bp paired-end reads on an Illumina MiSeq Sequencer. The within-amplicon overlapping paired-end FASTQ sequence reads were checked and pairwise assembled using PEAR[31]. The assembled contigs were then aligned to genome build hg38 using BWA-MEM[32], and variants called using the Genome Analysis Toolkit (GATK) v4.1 best practices, modified for the amplicon sequencing protocol[33].

More details about the methods used for the pipeline and quality control are available in Supplemental Methods.

Variant Functional Annotation

Variants were annotated using the Ensembl Variant Effect Predictor (release 93) with LOFTEE plugin for loss-of-function estimation[34]. Only nonsense or splice site variants that were predicted with High Confidence by LOFTEE were included. The pre-selected primary genome scan variant risk set 'Missense' included all predicted protein coding missense and nonsense variants, and any intronic variant located in the canonical two base pair splice donor or acceptor site at the 5' and 3' ends of an intron, in any aligned transcript of a gene. Two other secondary risk sets were tested for comparison of the relative magnitudes of variant associations using data from the resequencing experiment. The 'Deleterious' variant risk set included any nonsense or splice variant, plus any missense variant that was predicted to be 'probably_damaging' (PolyPhen) and 'deleterious' (SIFT). The most restrictive 'LOF/GOF/Splice' set was defined to include only nonsense and splice site variants.

Outcomes

The primary outcome for the cohort genetic analysis was the time to ESKD event from the baseline proteinuria cohort entry time (either incident and persistent, or first incident). The ESKD event was either clinically diagnosed or was inferred as an undiagnosed event at the first occurrence of eGFR below 10mL/min/1.73m^2. Absent an ESKD event for a participant, follow-up time was censored at the last eGFR measure. eGFR was estimated using the Chronic Kidney Disease (CKD) Epidemiology Collaboration (CKD-EPI) serum creatinine formula[35].

Membership in the retrospective study groups was used for the dichotomous case-control replication analyses, with inclusion/exclusion criteria as described above.

Whole Exome Gene-Aggregated Analysis

Gene-aggregated tests of variants were performed locally within each cohort using a minimal proportional hazards model including adjustments for eGFR at study baseline and principal components appropriate for the cohort ancestral composition. For the retrospective case-control study groups, a logistic regression model was used with similar adjustments. More details are available in Supplementary Methods. For each gene, two standard tests were performed: a Burden test, more powerful for genes with multiple variants that wholly or predominantly confer either risk or protection (variant effects in the same direction); and SKAT, which is more powerful for genes that contain multiple variants that confer risk and protection (variant effects in different directions in the same gene)[21]. The inheritance model for each variant was assumed to be additive such that two allele copies of the rarer allele were modeled with double the effect size of one, and the variants in each risk set in each gene were combined using a standard beta(1,25) weighting scheme based on their minor allele frequency (MAF) such that very rare variants (MAF close to 0) were given a weight of almost 25, while common variant weights dropped to nearly 0 at MAF=0.5. No MAF filter was applied to the variants, hence both rare and common variants were included, but the common had these much-reduced weights in the variant-aggregated statistic. The burden test (but not the bidirectional SKAT test) also permitted an estimate of the variant weighted average single direction effect size per minor allele. The gene score and covariance matrix summary statistics were transmitted to the genetic analysis center at the University of Virginia and combined into a meta-analysis of all genes using published

methods. The statistical genetic models and meta-analyses were implemented in R 3.2 or later (R Core Team), using the seqMeta package (v1.6.5 and v1.6.7; Voorman, Brody, Chen, Lumley, and Davis; https://cran.r-project.org/web/packages/seqMeta/vignettes/seqMeta.pdf). Empirical genomic control was applied to the exome-wide gene-based test results by normalizing the null model residual standard error by the square root of the parameter ($\lambda^{1/2}$) thereby adjusting the median p-value to null expectation and all p-values towards the null[36]. More details are included in Supplemental Methods. For the whole exome gene-aggregated scan, the study-wide significance threshold was set at $P<1.6x10^{-6}$ calculated with Bonferroni correction (0.05/15449 non-monomorphic genes/2 tests, burden and SKAT). The overall meta-analysis of cohort survival and case-control statistics was performed using a Liptak-Stouffer method of standardized normal deviates weighted by inverse standard error[37]. The overall meta-analysis test was 2-sided for burden and 1-sided for SKAT.

Immunofluorescence Staining for HSD17B14 in Human Kidney

The preparation of human kidney samples and RT-PCR were previously described in detail[38]. Human kidney biopsy paraffin sections were deparaffinized with xylene and ethanol, and pressure cooker treated for antigen retrieval. The sections were blocked with 3% BSA-PBS and were incubated with primary antibodies (anti-HSD17B14 rabbit polyclonal antibody: a generous gift from Dr. A. Jansson, Linköping University, Linköping, Sweden; anti-human KIM-1 mouse monoclonal antibody, Clone AKG7) for 1 hour at room temperature. After washing with PBS, sections were incubated with secondary antibodies for 30 minutes. The sections were then incubated with Vector ABC Elite Kit followed by color development with Vector DAB kit.

Mouse Ischemia-Reperfusion Injury and Aristolochic Acid-Induced Nephrotoxicity Studies

The murine ischemia-reperfusion injury (IRI) and acute aristolochic acid-induced nephropathy (AAN) models have previously been described[39]. Details of the preparation of mouse kidney samples are also described[38]. Mouse kidney frozen sections were thawed and treated with 1% SDS/PBS, then washed in PBS. After blocking with 3% BSA, sections were incubated with primary antibodies for 1 hour at room temperature or overnight at 4 °C. The frozen sections were then incubated with secondary antibodies for 30 minutes, then washed. Vectashield (Vector Laboratories, Burlingame, CA) containing DAPI (12.5 µg/ml) was applied and slide cover-slips applied.

Protein Structure Modeling

Available crystal structures of HSD17B14 in the Protein Data Bank (PDB, http://www.rcsb.org) showed very similar conformation and good superposition. PDB entry 6EMM was chosen for structure visualization of HSD17B14 since it contained a fully-modelled C-terminal chain, possibly due to the preservation of the inter-subunit Cys255-Cys255 disulfide bond (Bertoletti N et al, PDB deposited 10/02/2018, publication in process). C-terminal residues Gly271 and Ser272 were excluded from the analysis and presumed to be a cloning product since no predicted translation product of *HSD17B14* transcripts (Gencode v32), nor the canonical UniProt sequence Q9BPX1 (270 aa), contain these. CCP4mg was used for structural visualizations[40].

Expression Quantitative Trait (eQTL) Analysis

To test the association of specific common variants with gene expression in kidney compartments, existing published datasets were interrogated by variant[41]. The methods for data

generation are described therein, but briefly, human kidney samples were obtained from surgical nephrectomies, stored at -80°C in RNAlater (Ambion), then microdissected into glomerular and tubular compartments. RNA-seq data was generated using Illumina TruSeq protocols and GWAS data from Affymetrix Axiom Biobank arrays. eQTL analyses were run on European ancestry samples with absence of significant kidney structural changes (tubular fibrosis<10%, glomerular sclerosis<10%). After QC and filtering, 121 and 119 samples were used for tubule and glomerular eQTL analysis respectively.

Single Nucleus RNA-seq Analysis

Single nucleus RNA-seq (snRNA-seq) data were downloaded from NCBI GEO (Series GSE131882) containing transcript counts from experiments on renal cortex from nephrectomies of 3 human non-diabetic controls and 3 patients with early diabetic nephropathy[42]. The diabetic patients had elevated A1c, evidence of mesangial sclerosis, and glomerular basement membrane thickening. The patient ages ranged from 52 to 74 yr. and eGFR from 56 to 85 mL/min/1.73m^2 and did not differ between groups. Two diabetic patients had proteinuria with an increased proportion of global glomerulosclerosis and interstitial fibrosis and tubular atrophy. The six samples were analyzed as a single group using Seurat v3.1 with default QC and log normalization. The top 30 principal components were used as the input dimensions into using the Uniform Manifold Approximation and Projection (UMAP) algorithm for dimensional reduction and visualization of the single nucleus cell types[43].

Publicly Available Data Sets

Bulk RNA-seq expression results were retrieved from GTEx (https://www.gtexportal.org) Oct 21, 2019; human Affymetrix U133 array data from Nephroseq (http://www.nephroseq.org), Oct 21, 2019, from published experiments[44]. Selected gene expression values were extracted together with ascertainment pathology and other covariates. Additional post-processed snRNA-seq results were retrieved from http://humphreyslab.com, Jan 2, 2020, and used previously published experimental data[42,45]. Previous observations of alleles and frequencies were downloaded from gnomAD v2.1 (http://gnomad.broadinstitute.org)[46]. CKDGen summary results were retrieved from https://ckdgen.imbi.uni-freiburg.de.

RESULTS

Study Design and Participant Characteristics

We used clinical, phenotype, and the exome variant subset of Illumina Human CoreExome genotyping data, for six international proteinuria cohorts with longitudinal eGFR data, and two case-control sets, all of European ancestry. The study design and participation in the genetic analysis stages is shown in Figure 1A, and the clinical characteristics of the study groups in Table 1. The five discovery cohorts (Joslin, FinnDiane, Steno, INSERM, EDC) contained 2,212 total participants that experienced 1,095 ESKD events. The extension cohort (WESDR) contributed 160 participants and 20 events giving 1115 events in 2,372 participants for the overall meta-analysis of the leading gene. The two case-control replication studies respectively contained 946 new-onset ESKD cases (Joslin-Fresenius) and 610 controls (Joslin Medalists); and 126 cases (lower 50th percentile of diabetes duration) versus 142 controls (upper 50th percentile of diabetes duration) from GWU-GoKinD.

Genes Associated with Survival Against ESKD in Participants with T1D and Advanced DKD

Using gene-aggregated tests of association of all coding and splice site-affecting variants on the Illumina HumanCoreExome Array, we identified *HSD17B14* (hydroxysteroid 17-beta dehydrogenase 14) as the most significantly associated gene in our initial whole exome scan of 15,449 annotated, non-monomorphic genes using five T1DKD Discovery cohorts (Figure 1B and Table 2). After genomic control correction within each cohort the burden test p-value was 8.6×10^{-6}, just above genome-wide significance accounting for all multiple testing (1.6×10^{-6}). The top 10 genes by significance from this screen are included in Supplemental Table S1. Variant-

weighted meta-analysis of the lead gene including the WESDR cohort yielded an overall model burden test beta for the log(HR) of -0.046 (SE=0.010), equivalent to a HR=0.955 [95% CI: 0.94-0.97]. The meta-analysis of the two case-control study sets yielded a nearly identical burden test log(OR) of -0.045 (SE=0.019), equivalent to OR=0.955 [95%CI: 0.93-0.99]. Overall meta-analysis of the standardized effect sizes of the cohort and case-control results gave an experiment-wide significant p-value of 3.3×10^{-7} for *HSD17B14*, Figure 1 Panel B, and remained the most significant gene by either test. The burden test showed a superior p-value to the SKAT test in the same gene by more than two orders of magnitude (Supplemental Results Table S2) suggesting that the rare variants were predominantly acting in one direction of effect. The detailed study group results in Table 2 demonstrate the consistency of effect direction, with 7/8 study groups having a negative beta for log(HR) or log(OR) indicating that the rare/minor alleles for the gene variants exerted an overall protective effect against progression to ESKD from an advanced state of DKD; in the cohorts, this was the ascertained state of proteinuria. The top genes from the combined cohort and case-control meta-analyses are included in Supplemental Results Table S3. Since the burden test was empirically more significant for *HSD17B14*, follow-up analyses focused on this model.

Association of *HSD17B14* Variants Discovered by Resequencing

Since the exome array HSD17B14 content was limited to only 6 variants, we resequenced the HSD17B14 gene exons in five of our T1D cohorts (Figure 1A) to generate a deeper catalog of coding variation. The samples submitted for resequencing and that passed QC were similar to those with array genotyping but were augmented with additional samples that originally did not meet the cohort inclusion criteria and the EDC cohort samples were not available, hence the

sample sizes, variants seen, and distribution of genotype missing data differs from, and therefore is not directly comparable with the results from the whole-exome array-based scan. In the study time interval, these additional samples had subsequent DKD progression and now met those criteria. Post-sequencing QC, the cohort sample numbers were Joslin n=620 (n=614 previous array); FinnDiane 820 (n=783); Steno 416 (n=414); INSERM 254 (n=257); WESDR 312 (N/A) for 2,422 samples which resulted in 2,239 samples with phenotypic data that were included in the sequence-based association meta-analyses. We increased the catalog of variation from 6 array genotyped variants (2 SNPs with MAF >0.05 and 4 rare SNVs with MAF<0.002) to 20 coding SNVs and 2 non-coding intronic SNVs within the 2 bp splice recognition sites directly flanking intron-exon boundaries (Figure 2, Table 3). Three of the variants were classified as loss or gain of function or within a splice site (LOF/GOF/splice site risk set) with high quality (technical sequencing quality statistics shown in Supplemental Results Table S4); an additional 6 were classified as nonsynonymous but deleterious (Deleterious risk set); and 13 were non-deleterious missense (Missense risk set). A frameshift variant (p.A249CfsTer55) was found in a Joslin cohort patient, and the clinical record revealed that this patient had the longest duration since diabetes diagnosis without progression to ESKD in the entire Joslin cohort of 620 patients (65.8 years, Figure 3A and Supplemental Results Table S5), suggesting this could be a protective mutation. The predicted functional consequence of the inserted base allele was to extend the protein length in the most highly expressed protein isoform from wild type 270aa to 302aa, with the C-terminal end 22aa mutated compared to the wild type (Figure 3B). This frameshift had been seen 14 times previously in 235,000 alleles (gnomAD v2.1, Supplemental Results Table S6) but no homozygotes had been seen. Using visualization of existing crystal structures for HSD17B14 we found the wild type C-terminal residues 249-270 in each subunit of the homo-

tetramer contacted with the opposite subunit for most of their length (Figure 3C), and residues Val263, Pro266, and Pro269 of the C-terminal tail formed complementary hydrophobic interactions with the core of the subunit (Figure 3D); moreover, Asp267 formed a salt bridge with Arg203 of the core. The most important interaction was possibly the disulfide bond between residues Cys255 of the adjacent subunits (Figure 3E). These models suggested that the p.A249CfsTer55 mutant could have a destabilized tetrameric complex structure and that the elongated mutated C-tail had the potential to hinder the entrance of the substrate into the active site (Figure 3E). Further commentary is available in Supplemental Results and Figure S3.

Two splice site substitution mutations were found, one in the FinnDiane cohort (g.48835804C>G; exon 2, +1 donor position) and one in INSERM (g.48835844C>T; exon 2, -1 acceptor position). Neither of the two splice site mutations resulted in as extreme a phenotype as in the Joslin case (Supplemental Figure S3 and Results Table S5), nor had they been described previously in gnomAD (Table S6). Of all variants, R130W (rs35299026) and D62Y (rs139987974) were individually nominally significant (P<0.05), both were consistently protective against ESKD progression in the cohorts and predicted to be deleterious, and the former was also a common SNP (MAF=0.052). In the Joslin cohort, under a Weibull model of increasing hazard with age and all other factors held constant, each minor allele of the R130 variant was estimated to increase the median time to ESKD by 20%. There were no differences between genders in the association effect size, Table 4. Stratification of the variants by MAF and equal weighting in the Burden test showed that the rare variants (MAF < 0.01) were associated with protection overall (beta=-0.65, p=0.0033) while the only protective common variant was

rs35299026 (R130W). Further comparison of the risk set results is available in Supplemental Results and Table S7.

Association of *HSD17B14* Common Variant R130W in CKDGen GWAS Results

Since rs35299026 (R130W) was a common SNP, predicted to be deleterious, with the most significant single variant association in our data (log(HR) beta(SE) -0.3(0.1), P=0.0077, Table 3) we reviewed the results for this SNP in publicly available CKDGen results for European ancestry. We found weak nominal association of this SNP with quantitative blood urea nitrogen (beta(SE)= -0.005 (0.0023), P=0.03) in 211K samples but no association with CKD as a binary trait(N=388K), nor eGFR(N=484K) or UACR (N=510K) as continuous traits, Supplemental Results Table S8.

Normal Human Kidney Cell Types and Other Tissues: We assessed HSD17B14 protein expression in sections of nephrectomized normal human kidney by immunohistochemical staining. HSD17B14 was expressed in proximal tubules (Figure 4A and 4B, brown staining), with little evidence of staining in glomeruli. Cell type expression estimates from combined snRNA-seq data from adult nephrectomies of 3 healthy and 3 early diabetic kidney disease patients corroborated the higher expression within proximal tubule cells (Figure 5A and 5B, clusters PCT-1/2/3) with lower fraction of cells expressing the *HSD17B14* gene, and lower average per cell expression, in other kidney cell types. A second snRNA-seq data set for a single healthy adult confirmed the higher % of cells and level of expression in proximal tubule segments (Supplemental Figure S6). Publicly available bulk RNA-seq data sets confirmed relatively high *HSD17B14* expression in kidney across multiple data sets. Kidney *HSD17B14*

gene expression was ranked 1st of 37 tissues in the Human Protein Atlas (mean transcripts/million=53), 3/36 tissues in FANTOM5 (mean tags/million=43) and 10/53 tissues in GTEx V7 (median transcripts/million=35) (Supplemental Results and Figure S4)[45].

HSD17B14 Expression in Human Diabetic Kidney Disease: We compared normal and DKD kidney tissue by immunofluorescence for HSD17B14 and KIM-1, a marker of tubular injury and de-differentiation. In normal kidney sections from non-DKD individuals, no KIM-1-expressing proximal tubules were evident as expected, but HSD17B14 was highly expressed in the same cells (Figure 4C, arrows). In contrast, we found markedly reduced HSD17B14 expression in DKD, whereas KIM-1 was upregulated in proximal tubules (Figure 4D, arrowheads). Similarly, we found evidence for reduction of *HSD17B14* gene expression in DKD. Using previously published Affymetrix U133 microarray data from renal biopsies of 17 diabetic nephropathy patients and 31 healthy controls[44], we found that bulk mRNA *HSD17B14* expression levels in tubulointerstitial cells were significantly lower in the diabetic nephropathy group compared to controls (t-test, p=0.017, Figure 5C). We recapitulated this in a second data set (Figure 5D) discussed in the next section.

***HSD17B14* Gene Expression in Multiple Human Kidney Pathologies:** We analyzed *HSD17B14* gene expression in 433 micro-dissected human kidney tubule samples obtained from subjects with diabetes (DM) or hypertension (HTN) and subjects with DKD or hypertensive kidney disease, as measured by RNA-seq (Figure 5D)[47]. The median gene expression in the DKD group was lower than that in non-overlapping groups of samples from the same study: Controls, CKD, DM, HTN samples, and differed significantly between groups (Anova

P=0.00054). The expression of *HSD17B14* was strongly positively correlated with eGFR (cor=0.27, $p=3x10^{-12}$) among all disease samples (Figure 5E), and negatively correlated with degree of tubulointerstitial fibrosis (cor=-0.56, $p=2x10^{-16}$, Figure 5F). Analogous plots for glomeruli from the same study are shown in Supplemental Figure S7. A replication data set that included samples with mixed kidney pathologies confirmed the lower *HSD17B14* expression in other CKD[44]. Mean expression was lower in CKD associated with hypertension, IgA nephropathy, rapidly progressing glomerulonephritis, and lupus at P<0.05 (Supplemental Figure S8).

Association of the Common Missense Variant R130W (rs35299026) with *HSD17B14* Gene Expression: Since the common SNP missense variant R130W was individually negatively associated and protective against progression to T1D ESKD (Table 3, P=0.0077, MAF=0.052), we tested whether the coding variant was also associated with *HSD17B14* gene expression using a published data set, Table 5. We found that the minor A (W amino acid) allele of the SNP that was associated with protection against progression of T1D ESRD, was also associated with decreased gene expression in the tubule, and with less confidence, with decreased gene expression in glomeruli.

HSD17B14 Expression in Mouse Models of Kidney Injury: We assessed HSD17B14 expression in mouse kidneys by co-immunostaining HSD17B14 and KIM-1 in cortical and outer medullary proximal tubules, in tissue sections of sham-operated control kidneys (Figure 4E). These data were compared to Day 2 of post-ischemia reperfusion injury mouse model kidneys (Figure 4F). In addition, we evaluated the protein expression on day 14 after induction of murine

aristolochic acid nephropathy (AAN) (Figure 4G). The KIM-1 positive tubular cells, which exhibit undifferentiated characteristics including loss of apical brush border, had markedly reduced HSD17B14 staining in both post IRI and AAN kidneys. There was a strong reduction in *HSD17B14* mRNA levels post-ischemia injury at day 2 compared to normal mouse control, and this decline persisted over 21 days. The mRNA analysis confirmed the protein expression reduction in proximal tubule after IRI at day 2 and 21 in two mice (Figure 4H).

DISCUSSION

To the best of our knowledge, this is the first genome-wide screen of genes that are associated with longitudinal progressive loss of kidney function to ESKD in study participants with advanced DKD. Using data collected under the JDRF DNCRI, we identified hydroxysteroid 17-beta dehydrogenase 14 (*HSD17B14*) as a novel gene associated with protection against onset of ESKD in individuals with type 1 diabetes. The gene encodes an enzyme that is known to convert estradiol (E2) to estrone (E1) and is a member of an enzyme family that controls the relative balance of estrogen and androgen substrates, with secondary functions such as fatty acid metabolism. The overall effect of the minor and rare genetic coding variants appeared to protect carriers against more rapid terminal loss of renal function, as measured by delayed time to onset of ESKD. Since the genetic analysis focused on protein coding variants, we have greater initial confidence that we have identified an important, or sole, disease-associated gene, unlike reports from GWAS studies. In humans, the gene is highly expressed in proximal tubule cells, with lower levels of expression in other kidney cell types, consistent with previously reported patterns of protein expression for this enzyme, namely strong expression in kidney and particularly in the epithelial cells of proximal and distal tubule sections, while Bowman's capsule and glomerular epithelium remained relatively unstained[48]. Paradoxical to the presumed disruptive effect of the rare gene variants, we found that the gene was down-regulated in advanced DKD, with a suggestion from other data that it is also down-regulated in CKD states associated with other kidney pathologies. By comparing normal and DKD kidney tissue by immunofluorescence, we found markedly decreased proximal tubular expression of HSD17B14 in DKD specimens suggesting that the expression of the gene is positively correlated with protein expression. Cell-specific expression of HSD17B14 and KIM-1 (a marker of kidney injury[49]) was inversely

correlated such that tubule cells expressing HSD17B14 lacked KIM-1 reactivity, while injured positive KIM-1 cells showed attenuated HSD17B14 expression, an observation which was consistent with our mouse model studies. Hence the downregulation of HSD17B14 appears to be tightly associated with injury and de-differentiation of the proximal tubules in both mouse kidney disease model and human kidney diseases. Finally, our findings are strikingly similar to another member of this gene family, *HSD17B13*, which encodes an analogous enzyme and was very recently discovered to play a role in liver metabolism, in which a loss of function gene variant protects against chronic liver disease and fibrosis, yet whose active substrate in liver function also remains unknown[50].

Currently, we do not know the mechanism for the association of this gene with loss of kidney function in DKD, but prima facie, our results pose an intriguing paradox. The gene and encoded protein were downregulated in the proximal tubules in advanced DKD, yet the genetic variants that protected against loss of further kidney function are presumed to disrupt wild type protein function. This relationship between clinical outcome and human molecular expression was corroborated by the observation that the minor allele of the common coding missense SNP in the gene (rs35299026) was associated with less rapid loss of kidney function and protection against ESKD onset, but was also associated with decreased expression of *HSD17B14*, most significantly in the tubule cells. There are, however, important caveats to this line of reasoning. The human kidney tissues we studied may not have carried protective variants. Also, directly equating loss of gene activity through disruption of protein function with overall decrease in gene expression may be simplistic and overlook the fact that the rare protein variants were almost all heterozygous leaving a functional gene haplocopy, while down-regulation of gene

expression affects both alleles. Additionally, the loss of gene expression may be wholly or partly a consequence of the advanced diabetic environmental insult to the kidney, increasing the fraction of cells undergoing fibrosis and profoundly altering individual cellular programming. Preserved function in the non-fibrotic cells may be the determinant of risk hidden amongst the gross molecular and structural changes[47].

Our resequencing to develop a broader catalog of variants in the *HSD17B14* gene locus identified three variants with putative loss or gain of function or within the canonical intron-exon flanking splice site boundary. The only frameshift variant (p.A249CfsTer55) was predicted to result in an elongated mutated C-terminal tail, and the carrier was found to have had the longest duration since diabetes diagnosis in the cohort, suggesting the possibility that the participant acquired addition genetic protection from progression to ESKD. However, the splice variants did not appear to result in unusual protection or risk for their carriers. Whilst overall, the *HSD17B14* gene variants carried by our cohort participants were aggregately associated with protection, we were unable to definitively specify the direction of effect for individual rare variants, because of the very small number of carriers and uncertainty around their estimated effect size. Association tests using our predicted Deleterious risk set resulted in a more significant protective statistical result than the most permissive risk set of any missense or splice variant, suggesting that this Deleterious set did capture more of the larger effect size variants. Remarkably, the initial set of variants on the array resulted in a comparable association and protection in the sequencing data to the Deleterious set. This fortuitous occurrence of variants of relatively strong joint effect on the original array no doubt aided the initial exome-wide discovery of the gene. We did not detect any sex-specific differences in the gene effects although direct comparison between the sex strata

was imperfect because the sets of variants seen differed by sex because of the distribution of rare alleles.

HSD17B14 is the last member of the hydroxysteroid 17β dehydrogenase enzyme family which, with varying affinities and kinetics, interconvert 17-keto and 17β-hydroxy steroids, thus catalyzing conversion between estrone (E1) and estradiol (E2), androstenedione (A) and testosterone (T), and 5α-androstanedione and dihydrotestosterone. HSD17B14 preferentially oxidizes estradiol (E2) to estrone (E1), and T to A although conversion rates are low compared to other family members suggesting the possibility of other functional roles[48,51]. Other HSD17B members are known to be involved in fatty acid metabolism although in different compartments and at different biochemical pathway steps[52–54]. By analogy, it is possible that HSD17B14 also has multiple sex steroid and fatty acid metabolic functions, but the fatty acid substrate(s) remains unknown. Of particular interest is the fact that the HSD superfamily is a key target for drug development of steroid hormonal-stimulated diseases. Inhibitors of HSD17B14 are under active development[55,56].

The *HSD17B14* gene locus has not previously been identified in any genome-wide scan of kidney related traits, which now approach sample sizes of 1 million individuals for some traits in CKDGen[57,58], nor in the largest T1DKD studies[10,11]. These previous studies focused on common variants whereas our exome-centric tests included rare variants. Additionally, the analyses of eGFR and other traits in CKDGen studies are largely measuring natural population variation, and the CKD binary trait case-control analysis was more likely to identify genetic loci specific for generalized early stage renal impairment. Despite the larger samples sizes and exhaustive

phenotype testing in the very recent type 1 diabetes genetic study paper from this consortium[11], the analyses employed more pragmatic but less stringent phenotype/study group definitions and would have been underpowered to identify similar disease stage genes. We caution that to further replicate or extend these results, simple permissive ascertainment of cases and controls, even in participants with DKD, may not be an appropriate strategy. In the Joslin and FinnDiane cohorts with proteinuria, the cumulative risk of ESKD after 15 years was about 50% [2], suggesting that the misclassification rate of controls versus cases could be extremely high in unselected proteinuria versus ESKD comparisons.

Our study has multiple strengths. We defined a precise outcome phenotype to be tested in multiple incidence cohorts ascertained under a strict set of criteria and restricted to a particular stage in the natural history of kidney function loss in type 1 diabetes. We utilized the intensive follow-up data in diabetes clinics to identify the time-to-event from cohort baseline. The samples were uniformly genotyped, sequenced, and underwent QC at centralized laboratory and analysis centers. We performed cellular studies using immunohistochemistry to verify changes in expression of the encoded protein in human specimens and tested the gene in mouse models of kidney injury. However, there are also limitations to our study. Although we demonstrated consistency of effects and replication of our top gene, our sample size was limited and there are undoubtedly additional genes associated with this phenotype still to be discovered. We used a standard weighting function for the aggregated rare and minor allele variant analysis which places disproportionately more weight on rare variants in the collapsed variant statistics to reflect an expectation of larger effect on the phenotype with lower frequency alleles. This weighting function was a necessary pre-specified assumption and a different optimized function closer to

the true effect size distribution might yield more significant association results. Our study groups were ascertained on advanced type 1 DKD so the generalizability of these findings to type 2 DKD and other CKD pathologies remains an open question, although the reduction of expression in other CKD pathologies provides preliminary data that this gene may have a role in kidney function loss in these other pathologies also. The genetic variants we identified were not confirmed by RNA-Seq in the carrier participants and therefore their functional role in these individuals remains predicted rather than proven. Our functional experiments on HSD17B14 suggested a possible role for the encoded gene, but more extensive molecular and cellular work is needed to establish the exact mechanism and role of the gene during kidney function loss.

AUTHOR CONTRIBUTIONS

Study conception: M.P., S.S.R., J.V.B., S.H., P.R., P-H.G, J.H.W., A.S.K. Study design: J.C.M., A.S.K., A.G. Data acquisition, analysis, or interpretation: J.C.M., E.V., T.I., T.S.A., C.D., R.G.M., B.G., J.C., I.G.S., S.O-G., E.S., A.M.S., J.K.H., D-A.T, A.D.P., C.F., H.A.K., K.O., M.G.P., N.S., T.C., T.J.O., G.L.K., R.K., B.E.K., K.S., J.H.W. Manuscript writing: J.C.M., A.S.K. Manuscript review and editing: J.V.B., R.K., T.I. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The Pittsburgh Epidemiology of Diabetes Complications (EDC) study gratefully acknowledges the financial support of the Rossi Memorial Fund.

DATA SHARING

The summary statistics for the whole-exome gene-based scan are available in an Open Science Framework (https://osf.io) project with doi 10.17605/OSF.IO/9NEXD; and at the AMP-Type 1 Diabetes Knowledge Portal (http://www.type1diabetesgenetics.org/), under Renal, JDRF Diabetic Nephropathy Collaborative Research Initiative datasets. Genotypic data by individual cannot be shared because of study consent restrictions, and due to European Union and national regulations regarding protection of individual genetic data.

DISCLOSURES

We acknowledge the following conflicts of interest:

P.-H.G is an advisory board member for AbbVie, Astellas, Astra Zeneca, Bayer, Boehringer-Ingelheim, Eli Lilly, Janssen, Medscape, MSD, Mundipharma, Novartis, Novo Nordisk, Sanofi, and has received lecture honoraria from Astellas, Astra Zeneca, Boehringer-Ingelheim, Eli Lilly, Elo Water, Genzyme, MSD, Novartis, Novo Nordisk, PeerVoice, Sanofi, and SCIARC. J.V.B reports being co-founder of Goldfinch Bio and is co-inventor with T.I. on KIM-1 patents assigned to Partners Healthcare. He is a consultant for, Aldeyra, Cerespir, Merck, Mitobridge, and PTC. Dr. Bonventre also reports ownership equity in Dicerna, DXNow, Goldfinch, Goldilocks, Innoviva, Medibeacon, Medssenger, Rubius, Sensor-Kinesis, Sentien, Theravance and Verinano.

FUNDING

This study was supported by JDRF grant (3-SRA-2018-529-M-B) to J.C.M., J.V.B., R.K., and A.S.K; JDRF grant (17-2013-8) for DNCRI subproject “Search for genes determining time to onset of ESRD in type 1 diabetes patients with proteinuria” to S.H., P.R., P.-H.G., and A.S.K.; JDRF grants (6-2010-550, 1-2008-1018) to A.S.K; National Institutes of Health grant DK-041526 to A.S.K. and the Joslin Diabetes Research Center grant P30 DK036836. The FinnDiane study was funded by JDRF (17-2013-7), Novo Nordisk Foundation (NNF OC0013659), Academy of Finland (275614, 299200, and 316664), Folkhälsan Research Foundation (2019),

Wilhelm and Else Stockmann Foundation (2018), and Helsinki University Hospital Research Funds (TYH2018207) to E.V., N.S., C.F., V.H., and P.-H.G. The EDC study was supported by NIDDK grant DK34818. The WESDR study was funded by NEI EY016379. J.V.B. was also supported by NIH-NIDDK grants: R37 DK39773, R01 DK072381, and NCATS/NIDDK UG TR002155. A.G. was supported by NIA Claude D. Pepper Older American Independence Centers grant AG08808.

SUPPLEMENTAL MATERIAL
This article contains the following supplemental material available online.

1. Supplemental Methods
    1.1. Additional Cohort Information
    1.2. Amplicon Design for Resequencing
    1.3. Samples Included for Sequencing
    1.4. Amplicon Resequencing
    1.5. Bioinformatics Pipeline for Sequence Analysis
    1.6. Post-Sequencing Quality Control
    1.7. Annotation Files
    1.8. Statistical Genetic Models
    1.9. Genomic Control Adjustment of Gene-based Whole Exome Screen

2. Supplemental Results
    2.1 Supplemental Protein Modeling Results
    2.2 Comparison of Results from the Variant Risk Sets
    2.3 Expression of HSD17B14 in public bulk RNA-seq data sets

    Supplemental Results Table S1. Top 10 genes for association with survival against ESKD in type 1 diabetes patients with advanced DKD

    Supplemental Results Table S2. Results of array gene-based testing of HSD17B14 by cohort and case-control group including SKAT results

    Supplemental Results Table S3. Top 10 genes after meta-analysis of the type 1 diabetes cohort and case-control results

    Supplemental Results Table S4. Technical sequencing variant quality statistics of predicted LOF/GOF/splice site variants

    Supplemental Results Table S5. Characteristics of carriers of predicted LOF/GOF/splice site variants

    Supplemental Results Table S6. *HSD17B14* coding and splice variants in gnomAD

    Supplemental Results Table S7. Sensitivity of association results for survival time to ESKD by HSD17B14 variant risk set, genotypes from resequencing

    Supplemental Results Table S8. CKDGen Consortium results at rs35299026 (R130W) from publicly available GWAS data sets

3. Supplemental Figures

    Supplemental Figure S1. Quantile-quantile (QQ) plots for the 5-cohort discovery, gene-based, whole-exome scan.

    Supplemental Figure S2. Amplicon design in the *HSD17B14* genomic region for the resequencing project.

Supplemental Figure S3. Distribution of maximum duration since diagnosis of T1D cohort carrying LOF variants.

Supplemental Figure S4. Tetrameric organization of the wild-type HSD17B14 protein.

Supplemental Figure S5. Bulk non-diseased tissue expression of *HSD17B14* in multiple human public data sets.

Supplemental Figure S6. Single nucleus RNA-seq results for *HSD17B14* gene expression in normal, undiseased kidney tissue from a single adult nephrectomy.

Supplemental Figure S7. Variation in *HSD17B14* expression in kidney tissue from patients from 4 human disease states, and undiseased controls, measured by RNA-seq.

Supplemental Figure S8. *HSD17B14* comparative gene expression in multiple chronic kidney disease pathologies.

FIGURE LEGENDS

Figure 1. Study design and primary gene-based genome scan results. (A) The six T1DKD cohorts and two case-control studies showing participation in each phase of the study: primary whole exome scan (Scan), replication (Replic), meta-analysis, and resequencing (Reseq). The first five cohorts (Joslin, FinnDiane, Steno, INSERM, EDC) were used for the primary genome screen resulting in the initial exome scan meta-analysis result (15,449 genes), while the smaller extension cohort, WESDR, was also included with the replication case-control studies of the primary gene result (*HSD17B14)* in the overall meta-analysis of the top gene. (B) Manhattan plot for the primary burden test P-values. The gene *HSD17B14* was the most significant in the primary scan (5 cohorts) and improved with the inclusion of the WESDR extension cohort and the two replication case-control studies in the overall meta-analysis (+WESDR +Replication). The experiment-wise corrected P-threshold of $1.6x10^{-6}$ is shown as a red dashed line.

Figure 2. Genomic positions (build hg38), alleles, and risk set of all nonsynonymous (protein coding change) and splice site variants discovered by resequencing. The 'Missense' class contained any nonsynonymous or splice site donor or acceptor genetic variant; the 'LOF/GOF' class contained only variants predicted to cause loss or gain of function or splice site disruption; the 'Deleterious' class contained all LOF/GOF and any other missense variant that was predicted by SIFT and PolyPhen to be deleterious. The Genome Array line shows the variants present in the original Illumina HumanCoreExome array genome scan. LOF/GOF variants are in red text. The filled circles beneath the variants are shaded red if their effect direction was to increase risk (-ve effect), and shaded blue if they decreased risk (+ve effect).

Figure 3. Clinical consequences and protein modeling of loss of function p.A249CfsTer55 frameshift variant in *HSD17B14*. (A) Joslin type 1 diabetes cohort, distribution of maximum time since type 1 diabetes diagnosis either because of incident ESKD (green color) or censored last clinical visit with serum creatinine measurement for GFR estimation (blue color). The time to ESKD for the participant carrying the frameshift (diagnosed at age 2 yrs.) is shown with the red dotted line and has the maximum survival time since diagnosis in the entire cohort (65.7 yrs.), both incident ESKD and censored subgroups at last follow-time. (B) Schematic diagram of the change in the protein C-terminal tail as a result of the frameshift, blue beads - wild type sequence, red beads - mutated. The first blue residue is Ala249 which mutated to Cys249, other key residues indicated; the entire 22aa tail mutated to 54aa. Q9BPX1 is the UniProt accession number. (C) Homotetrameric organization of wild-type HSD17B14. Each subunit is colored separately. The core of each subunit is shown in molecular surface representation while the C-terminal fragments 249-270 are shown in ball-and-stick representation. C-terminal fragments of only two subunits are visible on this panel (green and purple) while the C-terminal fragments of other two subunits are symmetrically located on the opposite non-visible side. (D) Interactions between the C-terminal fragment 249-270 and the tetramer core, colored by the electrostatic potential of its surface: blue – positive, red – negative, white – hydrophobic. The C-terminal fragment is shown in ball-and-stick representation with white carbon atoms and sulfur in yellow. Key residues mentioned in the text are labelled, those from the adjacent subunit are labelled with B in parenthesis. (E) The entrance to the active site in a subunit of the tetramer. The NAD cofactor molecule, shown in magenta ball-and-stick model, can be seen through the entrance.

Figure 4. HSD17B14 expression in nephrectomized normal human kidney tissue, human diabetic kidney disease (DKD) biopsy specimens, and mouse models of kidney injury. (A and B) HSD17B14 expression was assessed by immunohistochemical staining. HSD17B14 (brown staining) is primarily expressed in the proximal tubules of the normal human kidney sections from nephrectomized kidneys. (C and D) Differential expression of HSD17B14 (arrows) and KIM-1 (arrowheads) was examined by immunofluorescence analysis in normal kidney tissue (C) and diabetic kidney disease biopsy tissue (D). HSD17B14= red, KIM-1= green, DAPI = blue. White scale bar = 20 μm. (E – G) Immunofluorescence staining of HSD17B14 in sham (E) compared to post-ischemic reperfusion injury (IRI) model kidney at day 2 post-ischemia (F) and 14 days after aristolochic acid (AAN, aristolochic acid nephropathy) (G) to establish expression levels in tubules with the kidney tissue markers. HSD17B14= red, Endomucin (EMCN, endothelial cells) = green, KIM-1(injured tubule) = white, DAPI = blue. (H) Relative mRNA levels by RT-PCR in the IRI model at sequential time points. HSD17B14 transcript expression by RT-PCR normalized to normal kidney, in day 2 (d2) and day 21 (d21) post-ischemia in two replicates each. Bars: 20 μm.

Figure 5. *HSD17B14* gene expression in normal and diseased human states. (A and B) Cell type clusters from 3 control and 3 early diabetic nephropathy renal cortex nephrectomies, visualized using UMAP projection of snRNA-seq count data. (A) shows the inferred clusters: Pod - podocyte; PCT-1/2/3 - three proximal convoluted tubule clusters; CFH-1/2 - two clusters expressing novel CFH marker pattern; TAL-1/2 - two thick ascending limb clusters; ATL - ascending thin limb; DCT - distal convoluted tubule; CNT - connecting tubule; PC-1/2 - two principal cell clusters; ICA-1/2 - two collecting duct intercalated type A clusters; ICB -

intercalated type B; Mes - mesangium; SMC+P - smooth muscle cells + pericytes; Endo - endothelial; Endo(Fen) - fenestrated endothelial; Endo(Lym) - lymphatic endothelial; Leuk - leukocytes. (B) Percentage of cells of each cluster with non-zero expression of *HSD17B14* corresponding to (A). Relative expression level in each cell color coded from grey (no expression) to blue (high expression). (C) Log2 expression of bulk *HSD17B14* mRNA measured using U133 2.0 Affymetrix arrays in tubulointerstitium of renal biopsies from 17 diabetic nephropathy and 31 healthy controls (data from NephroSeq 3.0). (D) bulk RNA-Seq expression of *HSD17B14* in log2(transcripts per million) in 5 disease states: Con: Control, non-diseased; CKD: chronic kidney disease, eGFR<60; DKD: diabetic kidney disease from type 2 diabetic patients; DM: type 2 diabetes without DKD; HTN: hypertension. (E and F) Corresponding to samples in (D) the same expression of *HSD17B14* plotted against eGFR (E) and % of fibrosis (F) with fitted linear relationships shown.

References

1. Libby P, Nathan DM, Abraham K, Brunzell JD, Fradkin JE, Haffher SM, et al.: Report of the National Heart, Lung, and Blood Institute-National Institute of Diabetes and Digestive and Kidney Diseases Working Group on cardiovascular complications of type 1 diabetes mellitus. *Circulation* 111: 3489–3493, 2005
2. Rosolowsky ET, Skupien J, Smiles AM, Niewczas M, Roshan B, Stanton R, et al.: Risk for ESRD in Type 1 Diabetes Remains High Despite Renoprotection. *J Am Soc Nephrol* 22: 545–53, 2011
3. Hovind P, Tarnow L, Rossing P, Jensen BR, Graae M, Torp I, et al.: Predictors for the development of microalbuminuria and macroalbuminuria in patients with type 1 diabetes: inception cohort study. *BMJ* 328: 1105, 2004
4. Costacou T, Orchard TJ: Cumulative kidney complication risk by 50 years of type 1 diabetes: The effects of sex, age, and calendar year at onset. *Diabetes Care* 41: 426–433, 2018
5. Krolewski AS, Gohda T, Niewczas MA: Progressive renal decline as the major feature of diabetic nephropathy in type 1 diabetes. *Clin Exp Nephrol* 18: 571–583, 2014
6. 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* 91: 1300–1311, 2017
7. Frodsham SG, Yu Z, Lyons AM, Agarwal A, Pezzolesi MH, Dong L, et al.: The familiality of rapid renal decline in diabetes. *Diabetes* 68: 420–429, 2019
8. Skupien J, Smiles AM, Valo E, Ahluwalia TS, Gyorgy B, Sandholm N, et al.: Variations in risk of end-stage renal disease and risk of mortality in an international study of patients with type 1 diabetes and advanced nephropathy. *Diabetes Care* 42: 93–101, 2019
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* 58: 1403–1410, 2009
10. 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. *J Am Soc Nephrol* 28: 557–574, 2017
11. Salem RM, Todd JN, Sandholm N, Cole JB, Chen WM, Andrews D, et al.: Genome-Wide Association Study of Diabetic Kidney Disease Highlights Biology Involved in Glomerular Basement Membrane Collagen. *J Am Soc Nephrol* 30: 2000–2016, 2019
12. Van Zuydam NR, Ahlqvist E, Sandholm N, Deshmukh H, Rayner NW, Abdalla M, et al.: A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes* 67: 1414–1427, 2018
13. Guan M, Keaton JM, Dimitrov L, Hicks PJ, Xu J, Palmer ND, et al.: Genome-wide association study identifies novel loci for type 2 diabetes-Attributed end-stage kidney disease in African Americans. *Hum Genomics* 13: 21, 2019
14. Bakris GL, Molitch M: Microalbuminuria as a risk predictor in diabetes: The continuing saga. *Diabetes Care* 37: 867–875, 2014
15. Caramori ML, Fioretto P, Mauer M: The need for early predictors of diabetic nephropathy risk: Is albumin excretion rate sufficient? *Diabetes* 49: 1399–1408, 2000
16. Pezzolesi MG, Krolewski AS: Diabetic Nephropathy: Is ESRD Its Only Heritable Phenotype? *J Am Soc Nephrol* 24: 1505–1507, 2013

17. Ahlqvist E, Van Zuydam NR, Groop LC, McCarthy MI: The genetics of diabetic complications. *Nat Rev Nephrol* 11: 277–287, 2015
18. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al.: Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337: 1190–1195, 2012
19. Wang D, Rendon A, Wernisch L: Transcription factor and chromatin features predict genes associated with eQTLs. *Nucleic Acids Res* 41: 1450–1463, 2013
20. Li B, Leal SM: Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *Am J Hum Genet* 83: 311–321, 2008
21. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X: Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89: 82–93, 2011
22. Abifadel M, Varret M, Rabès JP, Allard D, Ouguerram K, Devillers M, et al.: Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34: 154–156, 2003
23. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson P V., Bjornsson S, et al.: A mutation in APP protects against Alzheimer‘s disease and age-related cognitive decline. *Nature* 488: 96, 2012
24. Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, et al.: Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* 6: 5897, 2015
25. Orchard TJ, Dorman JS, Maser RE, Becker DJ, Drash AL, Ellis D, et al.: Prevalence of complications in IDDM by sex and duration. Pittsburgh Epidemiology of Diabetes Complications Study II. *Diabetes* 39: 1116–1124, 1990
26. Klein BEK, Klein R, McBride PE, Cruickshanks KJ, Palta M, Knudtson MD, et al.: Cardiovascular Disease, Mortality, and Retinal Microvascular Characteristics in Type 1 Diabetes. *Arch Intern Med* 164: 1917, 2004
27. Keenan HA, Costacou T, Sun JK, Doria A, Cavellerano J, Coney J, et al.: Clinical factors associated with resistance to microvascular complications in diabetic patients of extreme disease duration: The 50-year medalist study. *Diabetes Care* 30: 1995–1997, 2007
28. Mueller PW, Rogus JJ, Cleary PA, Zhao Y, Smiles AM, Steffes MW, et al.: Genetics of Kidneys in Diabetes (GoKinD) study: A genetics collection available for identifying genetic susceptibility factors for diabetic nephropathy in type 1 diabetes. *J Am Soc Nephrol* 17: 1782–1790, 2006
29. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, et al.: zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 28: 2543–5, 2012
30. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM: Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867–73, 2010
31. Zhang J, Kobert K, Flouri T, Stamatakis A: PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620, 2014
32. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760, 2009
33. Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, et al.: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491–501, 2011

34. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al.: The Ensembl Variant Effect Predictor. *Genome Biol* 17: 1, 2016
35. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, et al.: A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009
36. Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 55: 997–1004, 1999
37. Zaykin D V.: Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J Evol Biol* 24: 1836–1841, 2011
38. Kishi S, Brooks CR, Taguchi K, Ichimura T, Mori Y, Akinfolarin A, et al.: Proximal tubule ATR regulates DNA repair to prevent maladaptive renal injury responses. *J Clin Invest* 129: 4797–4816, 2019
39. Yang L, Besschetnova TY, Brooks CR, Shah J V., Bonventre J V.: Epithelial cell cycle arrest in G2/M mediates kidney fibrosis after injury. *Nat Med* 16: 535–543, 2010
40. McNicholas S, Potterton E, Wilson KS, Noble MEM: Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr Sect D Biol Crystallogr* 67: 386–394, 2011
41. Qiu C, Huang S, Park J, Park YS, Ko YA, Seasock MJ, et al.: Renal compartment–specific genetic variation analyses identify new pathways in chronic kidney disease. *Nat Med* 24: 1721–1731, 2018
42. Wilson PC, Wu H, Kirita Y, Uchimura K, Ledru N, Rennke HG, et al.: The single-cell transcriptomic landscape of early human diabetic nephropathy. *Proc Natl Acad Sci U S A* 116: 19619–19625, 2019
43. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al.: Comprehensive Integration of Single-Cell Data. *Cell* 177: 1888-1902.e21, 2019
44. Ju W, Nair V, Smith S, Zhu L, Shedden K, Song PXK, et al.: Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci Transl Med* 7: 316ra193, 2015
45. Wu H, Uchimura K, Donnelly EL, Kirita Y, Morris SA, Humphreys BD: Comparative Analysis and Refinement of Human PSC-Derived Kidney Organoid Differentiation with Single-Cell Transcriptomics. *Cell Stem Cell* 23: 869-881.e8, 2018
46. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al.: Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* [Internet] 531210, 2019 Available from: http://biorxiv.org/content/early/2019/08/13/531210.abstract
47. Beckerman P, Qiu C, Park J, Ledo N, Ko YA, Park ASD, et al.: Human Kidney Tubule-Specific Gene Expression Based Dissection of Chronic Kidney Disease Traits. *EBioMedicine* 24: 267–276, 2017
48. Sivik T, Vikingsson S, Gréen H, Jansson A: Expression patterns of 17β-hydroxysteroid dehydrogenase 14 in human tissues. *Horm Metab Res* 44: 949–956, 2012
49. Han WK, Bailly V, Abichandani R, Thadhani R, Bonventre J V.: Kidney Injury Molecule-1 (KIM-1): A novel biomarker for human renal proximal tubule injury. *Kidney Int* 62: 237–244, 2002
50. Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al.: A protein-truncating HSD17B13 variant and protection from chronic liver disease. *N Engl J Med* 378: 1096–1106, 2018
51. Lukacik P, Keller B, Bunkoczi G, Kavanagh K, Hwa Lee W, Adamski J, et al.: Structural

and biochemical characterization of human orphan DHRS10 reveals a novel cytosolic enzyme with steroid dehydrogenase activity. *Biochem J* 402: 419–427, 2007

52. Yang SY, He XY, Miller D: HSD17B10: A gene involved in cognitive function through metabolism of isoleucine and neuroactive steroids. *Mol Genet Metab* 92: 36–42, 2007
53. Venkatesan R, Sah-Teli SK, Awoniyi LO, Jiang G, Prus P, Kastaniotis AJ, et al.: Insights into mitochondrial fatty acid synthesis from the structure of heterotetrameric 3-ketoacyl-ACP reductase/3R-hydroxyacyl-CoA dehydrogenase. *Nat Commun* 5: 4805, 2014
54. Hiltunen JK, Kastaniotis AJ, Autio KJ, Jiang G, Chen Z, Glumoff T: 17B-hydroxysteroid dehydrogenases as acyl thioester metabolizing enzymes. *Mol Cell Endocrinol* 489: 107–118, 2019
55. Braun F, Bertoletti N, Möller G, Adamski J, Steinmetzer T, Salah M, et al.: First Structure-Activity Relationship of 17β-Hydroxysteroid Dehydrogenase Type 14 Nonsteroidal Inhibitors and Crystal Structures in Complex with the Enzyme. *J Med Chem* 59: 10719–10737, 2016
56. Song C, Burgess S, Eicher JD, O'Donnell CJ, Johnson AD, Huang J, et al.: Causal effect of plasminogen activator inhibitor type 1 on coronary heart disease. *J Am Heart Assoc* 6: 2017
57. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al.: A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* 51: 957–972, 2019
58. Teumer A, Li Y, Ghasemi S, Prins BP, Wuttke M, Hermle T, et al.: Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun* 10: 4130, 2019

Table 1. Type 1 Diabetic Kidney Disease Study Groups

Macroalbuminuria Cohorts	**Joslin**	**FinnDiane**	**Steno**	**INSERM**	**EDC**	**WESDR**
[US State,] Country	MA, USA	Finland	Denmark	France	PA, USA	WI, USA
Recruitment period	1993-2002	1998-2001	1993-99	1993-98	1986-88	1980-82
Patients w/Exome array data	614	783	414	257	144	N/A [a]
Age at cohort entry, median yrs (IQR)	35.0 (29.0, 35.8)	37.9 (31.3, 46.4)	40.9 (34.4, 49.0)	41.2 (32.4, 50.1)	31.7 (26.2, 38.0)	34.5 (27.5, 43.2)
Male, N (%)	339 (55.2)	474 (60.5)	251 (60.6)	154 (59.9)	82 (56.9)	100 (64.1)
Estimated GFR at cohort entry, median mL/min/1.73m^2 (IQR)	81.1 (62.9, 106.2)	71.7 (60.4, 83.3)	71.5 (51.8, 94.5)	66.1 (45.7, 89.0)	88.4 (66.9, 113.6)	96.9 (74.9, 114.4)
Time to ESKD, median yr (IQR)	6.6 (4.3,9.9)	6.4 (4.1, 9.9)	7.0 (4.4, 11.1)	5.4 (3.0,10.2)	10.0 (5.3, 13.3)	6.0 (5.1, 18.8)
Number of ESKD events	354	447	132	99	63	20
Case-control Groups	**Joslin-Fresenius ESKD Cases**	**Joslin Medalist Controls**	**GWU-GoKinD Cases**	**GWU-GoKinD Controls**		
[US State], Country	USA	MA, USA	USA	USA		
Patients w/Exome array data	946	610	126	142		
Age, median yrs (IQR)	50.2 (46.0, 57.0) [b]	N/A	43.0 (38.0, 48.0)	41.5 (36.0, 47.0)		
Male, N (%)	527 (55.7)	280 (45.9)	58 (46.0)	57 (40.1)		
Estimated GFR, median mL/min/1.73m^2 (IQR)	N/A	80.9 (70.7, 91.2)	47.8 (34.7, 66.4)	87.4 (77.2, 102.5)		
Number w/ESKD	946	0	126	0		

[US State,] is omitted for USA cohorts or groups that were recruited in multiple states.
Recruitment period is the range in years for the 25th and 75th percentiles of the collection.
a For reasons of study timing and logistics, WESDR analyses were performed using only genotypes from resequencing.
b In Joslin-Fresensius ESKD Cases, Age is the Age at ESKD onset.

Table 2. Results of gene-based testing of *HSD17B14* by cohort and case-control group

	Events / N	Variants	Burden Beta (SE)	Burden P-value
Discovery Cohorts				
Joslin	354 / 614	3	-0.051 (0.030)	0.091
FinnDiane	447 / 783	5	-0.045 (0.012)	1.5×10^{-4}
Steno	132 / 414	3	-0.218 (0.081)	0.0072
INSERM	99 / 257	4	0.011 (0.043)	0.80
EDC	63 / 144	4	-0.078 (0.052)	0.13
Discovery Meta-analysis (n=5)	1095 / 2212	5	-0.045 (0.010)	8.6×10^{-6}
Non-Discovery Cohort				
WESDR	20 / 160	3	-0.097 (0.10)	0.35
Cohort Meta-analysis (n=6)	***1115 / 2372***	***6***	***-0.046 (0.010)***	***6.3×10^{-6}***
Replication Case-Control	**N Cases / Controls**		**Burden Beta (SE)**	**Burden P-value**
Joslin-Fresenius vs Joslin Medalists	946 / 610	5	-0.041 (0.019)	0.052
GWU-GoKinD Cases vs Controls	126 / 142	5	-0.057 (0.041)	0.17
Case-Control Meta-analysis (n=2)	***1072 / 752***	***6***	***-0.045 (0.019)***	***0.017***
Overall Meta-Analysis				***3.3×10^{-7}***

Results used the variants present on the Illumina Infinium Human CoreExome Bead Array. Analysis for all cohorts except WESDR used array genotyping data; WESDR genotypes were derived from resequencing (described later). Variants column shows the number of variants tested in that cohort or case-control analysis of *HSD17B14*. The number varies because of batch quality control. Integers in parentheses in the Cohorts or Case-Control groups (n) are the number of studies included. Events are ESKD diagnosed clinically or by proxy from eGFR < 10mL/min/1.73m2. Betas are the log values of the Hazard (HR) /Odd Ratios (OR) and measure the mean effect of the aggregated rare or less common alleles weighted by the discovery weight function beta(1,25). Weights for rare variants (MAF< 0.005) are close to 25, while those for common SNPs (MAF > 0.1) are near 0. SE are the standard errors of the beta estimates. The exome-wide gene discovery and case-control results were corrected for genomic control within each cohort or case-control study.

Table 3. Association of individual coding and splice site variants in *HSD17B14* with survival time to ESKD in 5 type 1 diabetic incidence cohorts with proteinuria

Position	Location	REF	ALT	Consequence	P-val	MAC	MAF	Beta (SE)	Dir
LOF/GOF/Splice site									
48813244	Exon 9	A	AC	A249CfsTer55	0.17	1	$2.2x10^{-4}$	-1.4 (1.0)	-
48835804	Exon 2 +1 donor	C	G	Splicing	0.25	1	$2.2x10^{-4}$	-0.8 (0.7)	-
48835844	Exon 2 -1 acceptor	C	T	Splicing	0.50	1	$2.2x10^{-4}$	1.3 (2.0)	+
Deleterious									
48815123	Exon 6	G	A	R130W	0.0077	234	0.052	-0.3 (0.1)	-----
48834302	Exon 3	C	A	D62Y	0.033	9	0.0020	-1.0 (0.5)	--
48836333	Exon 3	G	A	R27C	0.064	2	$5.8x10^{-4}$	-2.1 (1.1)	-
48836347	Exon 3	C	T	G22E	0.61	2	$5.7x10^{-4}$	-0.6 (1.2)	-+
48836366	Exon 1	C	A	G16W	0.27	16	0.0045	0.6 (0.5)	+--+
48836387	Exon 1	C	T	G9R	0.10	1	$2.2x10^{-4}$	-1.4 (0.9)	-
Missense									
48813201	Exon 9	C	T	V263M	0.42	1	$2.2x10^{-4}$	1.5 (1.8)	+
48813206	Exon 9	G	T	T261N	0.64	1	$2.2x10^{-4}$	-4.4 (9.6)	-
48813307	Exon 7	C	T	R159Q [a]	0.99	1	$2.2x10^{-4}$	-1.4 (301.0)	-
48831700	Exon 5	G	T	L113M [b]	0.52	1	$2.3x10^{-4}$	-1.4 (2.2)	-
48831714	Exon 5	C	T	R108H	0.25	1	$2.4x10^{-4}$	2.8 (2.5)	+
48831732	Exon 5	G	A	T102I	0.21	1	$2.4x10^{-4}$	-2.5 (2.0)	-
48831756	Exon 5	G	C	P94R [b]	0.94	1	$2.3x10^{-4}$	12.6 (169.0)	+
48832675	Exon 4	C	G	A90P	0.61	1	$2.2x10^{-4}$	0.7 (1.5)	+
48832699	Exon 4	G	A	R82C	0.53	1	$2.2x10^{-4}$	-1.4 (2.3)	-
48834281	Exon 3	C	T	V69M	0.08	3	$6.7x10^{-4}$	-1.4 (0.8)	---
48834301	Exon 3	T	C	D62G [b]	0.20	6	0.0013	-0.7 (0.6)	-+-
48834320	Exon 3	C	T	A56T	0.49	80	0.018	0.1 (0.2)	++-+-
48835841	Exon 2	T	C	N31D	0.08	1119	0.25	0.1 (0.06)	++++-

Shaded rows are the variants that were present in the post-QC Illumina exome-wide array scan.
Deleterious variants: LOF or GOF + splice donor or acceptor sites AND deleterious missense sites (SIFT=deleterious AND PolyPhen=probably_damaging).
Missense: LOF or GOF + splice donor or acceptor sites AND all missense variants
MAC is the minor allele count. MAF is the minor allele frequency. Dir is the direction of the variant effect in the cohorts where the variant was seen. Beta gives the estimated effect size for the rare or uncommon allele at each variant. Consequence was predicted by VEP with LOFTEE plugin for the canonical full-length transcript. Individual results were corrected using the genomic control parameters from the whole exome scan within each cohort. N=2239 total samples.
a This missense variant was predicted to be present in a secondary transcript (ENST00000595764) only.
b These just missed inclusion in the Deleterious set and were predicted to be deleterious (SIFT) and possibly_damaging (PolyPhen)

Table 4. Survival time to ESKD association results for variants in *HSD17B14* by variant risk set and gender, genotypes from resequencing

Variant Risk Set	Gender Stratum	N	Variants Tested	Burden Beta (SE)	Burden P-value	Gender Interaction P-value
LOF/GOF/Splice	Both	2239	3	-0.75 (0.49)	0.13	Not Tested
Deleterious	Both	2239	9	-0.032 (0.009)	0.00036	
	Female	1003	6	-0.032 (0.015)	0.035	
	Male	1236	7	-0.034 (0.012)	0.0040	0.90
Missense	Both	2239	22	-0.018 (0.006)	0.0055	
	Female	1003	14	-0.021 (0.010)	0.029	
	Male	1236	16	-0.016 (0.009)	0.065	0.72
GWAS Variants	Both	2239	6	-0.044 (0.010)	$1.4x10^{-5}$	
	Female	1003	5	-0.044 (0.015)	0.0043	
	Male	1236	6	-0.047 (0.014)	0.00081	0.87
Common, MAF >0.01 [a]	Both	2239	3	0.024 (0.053)	0.65	
	Female	1003	3	0.011 (0.080)	0.90	
	Male	1236	3	0.004 (0.072)	0.96	0.95
Rare, MAF <0.01 [a]	Both	2239	19	-0.65 (0.22)	0.0033	
	Female	1003	19	-0.62 (0.36)	0.086	
	Male	1236	19	-0.69 (0.29)	0.016	0.88

The variant risk sets were: LOF/GOF/Splice: loss or gain or function or in a splice donor/accepter site; Deleterious: LOF/GOF variants plus those predicted by PolyPhen and SIFT to be deleterious; Missense: Deleterious variants plus any other non-synonymous variants; GWAS Variants: any Missense variant present on the original array after QC. Results were corrected using the genomic control parameters from the whole exome scan within each cohort. Burden P-value was derived from a score test. MAF is the minor allele frequency of the variant.

a For MAF strata, the variants were equally weighted; for all other risk sets, variants were weighted based on their estimated MAF using a beta(1,25) function.

Note: Sample sizes (N), variants seen, and distribution of genotype missing data differs from the sample sets used for the whole exome gene-based scan and are not directly comparable.

Table 5. Kidney phenotype associations of the common missense SNP R130W (rs35299026)

Phenotype	Effect (SD)	P-value	Minor Allele Direction
Tubule *HSD17B14*	-0.74 (0.215)	0.00081	Decreased gene expression
Glomerulus *HSD17B14*	-0.54 (0.22)	0.019	Decreased gene expression
Time to T1D ESKD	-2.9 (1.1)	0.0008	Protective against ESKD

The association results are coded for the minor allele (MAF=0.052), genome allele A, protein amino acid allele W.