Online Supplemental Material to

Chromosome 2q31.1 is associated with ESRD in women with type 1 diabetes

Niina Sandholm, Amy Jayne McKnight, Rany M. Salem, Eoin P. Brennan, Carol Forsblom, Valma Harjutsalo, Ville-Petteri Mäkinen, Gareth J. McKay, Denise M. Sadlier, Winfred W. Williams, Finian Martin, Nicolae Mircea Panduru, Lise Tarnow, Jaakko Tuomilehto, Karl Tryggvason, Gianpaolo Zerbini, Mary E. Comeau, Carl D. Langefeld for the FIND Consortium, Catherine Godson, Joel N. Hirschhorn, Alexander P. Maxwell, Jose C. Florez and Per-Henrik Groop on behalf of the FinnDiane Study Group and the GENIE Consortium.

Table of Contents

Corresponding author:

Per-Henrik Groop
Division of Nephrology
Department of Medicine
Helsinki University Central Hospital
Biomedicum Helsinki
Haartmaninkatu 8, P.O. Box 63
FIN-00014 University of Helsinki
Finland
Tel. +358 9 191 25459
Fax +358 9 191 25452
E-mail per-henrik.groop@helsinki.fi

Full Methods

Patients: Details of study participants have been described previously.[1] The discovery cohort consisted of 3,652 subjects with T1D from the Finnish Diabetic Nephropathy (FinnDiane) study, which is a Finnish nationwide multicenter study. The replication was performed on three independent cohorts that participate in the GENIE collaboration: All Ireland-Warren 3-Genetics of Kidneys in Diabetes UK (UK-ROI) Collection with 1,830 genotyped T1D subjects, Genetics of Kidneys in Diabetes US Study (GoKinD US) with 1,792 subjects and an Italian cohort from the Milan region comprising of 397 T1D subjects. Replication was also considered to include patients from the Steno Diabetes Center, an additional group of the UK-ROI subjects, a Swedish collection from the Stockholm and Umeå regions, and the RomDiane study, but the number of women with ESRD was deemed too low in any of these cohorts in order to be able to perform robust statistical analysis of genotype differences (N = 3 - 29 female ESRD cases; minor allele count < 10 within female ESRD cases).

We compared the T1D subjects with ESRD with the T1D controls that had no signs of diabetic kidney disease despite long duration of diabetes. ESRD was defined as the need for chronic dialysis treatment or having received a kidney transplant, and the minimum duration of T1D was 10 years. We excluded all T1D patients who were known to have ESRD due to any non-diabetic cause. In FinnDiane, 85% of the ESRD cases in FinnDiane had laser treatment (9% had missing data) and only six ESRD patients had no retinopathy. In UK-ROI and GoKinD US all ESRD cases had retinopathy as inclusion criteria. Controls were defined as T1D subjects with stable normal urinary albumin excretion and a long diabetes duration of at least 15 years; normal albumin excretion was defined as albumin excretion rate

(AER) < 20 µg/min or <30 mg/24h or a urinary albumin to creatinine ratio (ACR) < 2.5 mg/mmol for men and ACR < 3.5 mg/mmol for women in overnight, 24-hour or spot urine collections, respectively. Overt proteinuria was diagnosed in some cases using dipsticks. All the subjects were diagnosed with T1D before the age of 35 years.

Genotyping: Genotyping of the FinnDiane subjects was done with the Illumina 610Quad chip and genotype calling, quality control and imputation procedures have been described earlier.[1] In brief, genotypes with low genotyping quality or low minor allele frequency (<0.01) were discarded, as well as samples with low genotyping quality or cryptic relatedness, and geographical outliers based on principal component analysis. 3,546 samples and 549,530 SNPs remained after quality control. Imputation was based on the HapMap II CEU population and resulted in ~2.4 million SNPs across the autosomal genome.

Genotyping of the replication cohorts: UK-ROI and GoKinD US were included as part of the GENIE collaboration where genotyping was performed using the Illumina Omni1 Quad Array for UK-ROI whereas existing genotype data for GoKinD US was downloaded from the dbGAP (phs000018.v2.p1, retrieved June 2012) with updated genotypes from Affymetrix 500K array.[1] The SNP chosen for replication was selected from the GWAS data, where the quality control and imputation procedures were the same as described above. Quality control resulted in 1,726 UK-ROI samples and 1,595 GoKinD US samples. The DNA samples from Italy, Steno Diabetes Center, UK-ROI replication cohort, Sweden (Stockholm and Umeå regions), and the RomDiane study were genotyped using Sequenom IPLEX assays (Sequenom Inc, San Diego, CA). Only UK-ROI, GoKinD US and the Italian cohort had minor allele count of ten or higher within the female ESRD cases and

thus sufficient number of cases to be included in the statistical analysis.

Family Investigation of Nephropathy and Diabetes (FIND) Study, patients and genotyping: The FIND cohort consisting of 885 samples from European Americans, 1,460 samples from African Americans, 889 samples from American Indians, and 1,535 samples from Mexican Americans were genotyped using the Affymetrix SNP 6.0 GeneChip at the Genotyping Core Facility at Affymetrix (Santa Clara, California). Samples were plated by ethnic group, randomizing by case/control status; 217 pairs of duplicate samples we included for quality control. Genotype calls used the Birdsuite algorithm as implemented in the GCOS software (Affymetrix, Santa Clara, California). Samples with call rates greater than 95% were subject to additional quality control procedures for sample and SNP heterozygosity, sample and SNP missingness, gender verification, expected and unexpected relatedness, and population substructure analysis via principal components analysis. After trimming, 342 cases and 404 controls of European American ancestry, 979 cases and 304 controls of African American ancestry, 538 cases and 319 controls of American Indian ancestry, and 779 cases and 594 controls of Mexican American ancestry were included in a final meta-analysis.

Statistical analysis: The association analysis that compared ESRD cases with T1D controls with normal AER despite long duration of T1D was performed with the PLINK 1.07 software[2] using logistic regression. Estimated allele dosages were employed rather than the most likely genotypes in order to account for the uncertainty arising from the imputation process. Women and men were analyzed separately. The association analyses were adjusted for age, T1D duration, and the ten first principal components obtained with the EIGENSTRAT software.[3] The quantile-quantile (QQ) plots of the both analyses showed good adherence to the diagonal line of expected *P*-values and very little

excess genomic inflation was observed (λ=1.034 for women, λ=1.045 for men). The same methods were used for the statistical analysis of UK-ROI and GoKinD US. Because of the small number of women with ESRD in the Italian cohort, we utilized Fisher's exact test of association, which is more robust for small sample numbers. Consequently, the Italian replication cohort was not adjusted for any covariates. The robustness of the results for the Italian cohort was tested by repeating the association analysis with logistic regression as described above and the results were essentially the same: *P*=0.56 and OR=1.20 with Fisher's exact test, *P*=0.77 and OR=1.12 using logistic regression. Meta-analysis of the four cohorts was performed using a fixed effect model based on the standard errors and *P*-values, implemented with the METAL software.[4]

Power calculations: Power calculations were performed with Genetic Power Calculator.[5] The prevalence of ESRD in the non-exposed group (P_0) was estimated to be 0.1. Genotype relative risk (RR) was approximated with the formula $RR = OR/((1-P_0)+P_0 \times OR)$.[6] We used as the OR in the power calculations the lower 95% CI of the OR in women in FinnDiane. Study-specific allele frequency was employed for each cohort, and rs4972593 was assumed to be in full linkage (D'=1) with the background causal variant in each study.

Association analysis in FIND: Association analysis was calculated in the FIND study separately for men and women and for all the included study cohorts (African Americans, American Indians, European Americans and Mexican Americans). The minimum count of ten minor allele homozygotes was required for the analysis. Association was calculated assuming an additive genetic model.

Transcription factor binding sites (TFBS) and regulatory function: We looked for the TFBSs

directly created or deleted due to rs4972593 using MatInspector (Release professional 8.06, August 2012; Matrix Family Library Version 8.4 (June 2011)) from the Genomatix software suite (Genomatix Software, GmbH, Munich, Germany). The flanking region 5 kbp up- and downstream of rs4972593 was downloaded from NCBI SNP data base (http://www.ncbi.nlm.nih.gov/projects/SNP/) and MatInspector was used to detect estrogen responsive elements (V$EREF family) within this region. Furthermore, we sought for evidence of the regulatory function of the SNPs using the RegulomeDB database, which annotates SNPs with known and predicted regulatory elements in the intergenic regions. The annotation is based on regions of DNAase hypersensitivity, binding sites of transcription factors, and promoter regions that have been biochemically characterized to regulate transcription.[7]

Renal gene expression: Disease and gender-associated gene expression in published human diabetic nephropathy microarray datasets was determined using Nephromine (www.nephromine.org). Selected datasets comprised microdissected renal biopsies from diabetic nephropathy patients versus living donor/minimal change disease patients.[8,9] We performed a total of ten various gender- and disease specific association look-ups (**Supplemental Table 8**). Therefore the *P*-values were adjusted for multiple testing according to ten performed tests; *P*-value of 0.005 was required for significance after adjustment.

eQTL gene expression: We studied if rs4972593 was associated with the gene expression level of any of the genes within a 1Mbp region up- and downstream in the HapMap3 lymphoblastoid cell lines[10] using the Genevar user interface (http://www.sanger.ac.uk/resources/software/genevar/). The analysis included all the SNPs in full LD (r^2=1) with rs4972593 in the HapMap2 CEU samples and with data on HapMap3 eQTL in Genevar; three SNPs filled the criteria (rs530673, rs4972590, rs4972591). The

P-value threshold for statistical significance after multiple testing was $P<0.00063$ based on $\alpha=0.05$ significance level, 10 studied genes and 8 included HapMap populations. As the selected SNPs were in strong LD, we did not adjust for the number of SNPs.

Supplemental Table 1: GWAS top loci with $P<10^{-5}$ in FinnDiane and *in silico* replication in UK-ROI and GoKinD US.

				FinnDiane			UK-ROI		GoKinD US		Meta	
SNP (A1/A2)	Chr	Bp	Genes	Frq	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Women												
rs4972593 (A/T)	2	174171100	*SP3 - CDCA7*	0.11	2.39 (1.75 - 3.25)	3.0×10^{-8}	1.16 (0.75 - 1.8)	0.50	2.07 (1.24 - 3.46)	0.005	1.92 (1.53 - 2.4)	1.7×10^{-8}
rs530673 (G/C)	2	174162256	*SP3 - CDCA7*	0.11	2.38 (1.75 - 3.24)	3.5×10^{-8}	1.18 (0.76 - 1.81)	0.47	2.08 (1.24 - 3.47)	0.005	1.91 (1.53 - 2.4)	1.8×10^{-8}
rs4963667 (C/T)	12	26622226	*ITPR2*	0.23	0.50 (0.37 - 0.66)	1.7×10^{-6}	0.73 (0.45 - 1.18)	0.20	0.82 (0.56 - 1.2)	0.30	0.62 (0.5 - 0.76)	5.1×10^{-6}
rs9325969 (G/T)	22	47601109	*FAM19A5 - LOC100128946*	0.39	1.69 (1.35 - 2.11)	5.6×10^{-6}	0.85 (0.6 - 1.2)	0.36	0.89 (0.68 - 1.16)	0.39	1.19 (1.02 - 1.39)	0.02
rs10926584 (C/T)	1	240071458	*WDR64 - EXO1*	0.44	1.65 (1.33 - 2.04)	6.0×10^{-6}	1.02 (0.72 - 1.44)	0.92	1.01 (0.78 - 1.32)	0.92	1.28 (1.1 - 1.49)	1.2×10^{-3}
Men												
rs10871288 (A/G)	16	83408355	*USP10 - CRISPLD2*	0.19	1.89 (1.46 - 2.44)	1.2×10^{-6}	0.9 (0.61 - 1.33)	0.61	1.29 (0.79 - 2.12)	0.31	1.47 (1.21 - 1.79)	1.1×10^{-4}
rs2655224 (A/T)	3	13516713	*HDAC11*	0.42	1.63 (1.33 - 2.01)	2.9×10^{-6}	0.96 (0.7 - 1.32)	0.79	0.71 (0.48 - 1.04)	0.08	1.25 (1.06 - 1.46)	6.2×10^{-3}

Only the SNPs with the lowest *P*-value are shown for each independent signal, except for the locus between *SP3* and *CDCA7* genes where both SNPs with genome-wide statistical significance are shown. SNP (A1/A2): rs number and alleles A1 and A2. A1 is the minor allele in FinnDiane. Chr: Chromosome number. Bp: basepair position, according to human genome built 36. Genes: The gene name if the SNP is within the gene, or two flanking genes if the SNP is intergenic. Frq: A1 allele frequency in the FinnDiane subjects. Meta: Meta-analysis of FinnDiane, UK-ROI and GoKinD US.

Supplemental Table 2: Patient clinical characteristics and association with rs4972593 in FinnDiane.

	ESRD			Controls			Association with rs4972593					
	Women	Men		Women	Men		All		Women		Men	
	N = 258	N = 387	*P*	N = 935	N = 656	*P*	Effect	P_{SNP}	Effect	P_{SNP}	Effect	P_{SNP}
Age at T1D onset (years)	11.3 ± 6.7	13.7 ± 7.6	***	14.7 ± 7.8	15.8 ± 8.9	**	-0.05	ns	0.27	ns	-0.51	ns
Age (years)	44.4 ± 8.9	47.7 ± 8.6	***	42.9 ± 11.1	43 ± 11.9		0.12	ns	-0.34	ns	0.65	ns
T1D duration (years)	33.1 ± 8.8	34 ± 8.3		28.3 ± 9.6	27.2 ± 9.2	*	0.26	ns	-0.60	ns	1.13	ns
Transplanted (N (%))	123 (47.7)	193 (49.9)		0 (0)	0 (0)							
AHT medication (N (%))	229 (88.8)	361 (93.3)	*	204 (21.9)	179 (27.3)	*						
Lipid-lowering medication (N (%))	99 (38.4)	162 (41.9)		114 (12.2)	99 (15.1)							
BMI (kg/m^2)	24 ± 3.9	25.1 ± 4.3	**	25.1 ± 3.8	25.4 ± 3.1		0.75	4.0×10^{-5}	0.84	0.001	0.67	0.009
SBP (mmHg)	151 ± 24	153 ± 25		133 ± 18	136 ± 16	***	-0.34	ns	-0.87	ns	0.20	ns
DBP (mmHg)	82 ± 13	84 ± 12	*	78 ± 9	79 ± 10	*	0.63	ns	0.58	ns	0.71	ns
HbA_{1c} (%)	8.8 ± 1.8	8.7 ± 1.7		8.1 ± 1.2	8 ± 1.2		0.11	ns	0.22	0.01	-0.02	ns
Total cholesterol (mmol/l)	5.4 ± 1.3	5.1 ± 1.1	*	4.9 ± 0.8	4.7 ± 0.9	**	0.09	ns	0.10	ns	0.07	ns
Triglycerides (mmol/l)	1.5 (1.0 – 1.8)	1.9 (1.1 – 2.3)	***	1.0 (0.7 – 1.1)	1.1 (0.7 – 1.3)	***	0.04	ns	0.01	ns	0.08	ns
HDL cholesterol (mmol/l)	1.4 ± 0.5	1.2 ± 0.4	*	1.6 ± 0.4	1.4 ± 0.4	***	-0.03	ns	-0.07	ns	-0.00	ns
Association with ESRD, adjusted							0.43	6.8×10^{-5}	0.75	3.0×10^{-6}	0.03	ns

Data are mean ± SD or N (%) or mean (interquartile range). AHT= antihypertensive; BMI= body mass index; SBP= systolic blood pressure; DBP= diastolic blood pressure; HDL= high density lipoprotein. *P* = *P*-value for the T-test that the continuous phenotype varies between men and women within ESRD cases or within controls. Association with rs4972593: Association between rs4972593 and the continuous phenotypes, calculated with linear regression. All: Both men and women are included in the regression models. Analysis is stratified by gender and ESRD status by analyzing the groups separately and combining the results with fixed effects meta-analysis. Women: Regression models include the women only, analysis is stratified by ESRD status. Men: Regression models include men only, analysis is stratified by ESRD status. Effect: effect size estimate (beta) for the association given as the change in the continuous phenotype per one rs4972593 risk allele. P_{SNP} = *P*-value for the association between rs4972593 and the continuous phenotype. Association with ESRD, adjusted: Association between rs4972593 and ESRD after adjustment for the covariates that were significantly associated with the SNP. BMI and/or HbA_{1C} was missing for 83 women and for 71 men. * $P<0.05$; ** $P<0.01$; *** $P<0.001$; ns = *P*-value ≥ 0.05.

Supplemental Table 3: Association results stratified by gender for rs4972593 and previously reported susceptibility loci for ESRD in T1D.

			Women		Men		All	
Locus	**SNP**	**Study**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**
CDCA7 - SP3	rs4972593	FinnDiane	3.0×10^{-8}	2.39 (1.75 - 3.25)	0.77	0.95 (0.69 - 1.31)	2.9×10^{-4}	1.51 (1.21 - 1.88)
		GENIE	1.7×10^{-8}	1.92 (1.53 - 2.4)	0.95	1.01 (0.79 - 1.28)	6.7×10^{-5}	1.39 (1.18 - 1.64)
AFF3	rs7583877	FinnDiane	0.0024	1.43 (1.14 - 1.79)	6.6×10^{-4}	1.47 (1.18 - 1.82)	7.0×10^{-6}	1.44 (1.23 - 1.68)
		GENIE	2.8×10^{-4}	1.34 (1.15 - 1.58)	3.2×10^{-4}	1.34 (1.14 - 1.57)	4.9×10^{-7}	1.33 (1.19 - 1.48)
		GENIE ESRD*	6.2×10^{-5}	1.35 (1.17 - 1.57)	1.4×10^{-5}	1.35 (1.18 - 1.54)	4.8×10^{-9}	1.34 (1.22 - 1.48)
RGMA - MCTP2	rs12437854	FinnDiane	0.038	1.82 (1.04 - 3.19)	0.021	1.97 (1.12 - 3.46)	0.0011	1.94 (1.31 - 2.86)
		GENIE	0.0038	1.8 (1.21 - 2.68)	0.0027	1.82 (1.23 - 2.69)	1.6×10^{-5}	1.82 (1.39 - 2.39)
		GENIE ESRD*	0.013	1.59 (1.1 - 2.29)	5.4×10^{-4}	1.77 (1.28 - 2.44)	7.6×10^{-6}	1.72 (1.36 - 2.18)
EPO	rs1617640	FinnDiane	0.87	1.02 (0.82 - 1.26)	0.33	1.11 (0.91 - 1.35)	0.39	1.07 (0.92 - 1.23)
		GENIE	0.34	0.93 (0.79 - 1.09)	0.91	1.01 (0.86 - 1.18)	0.50	0.96 (0.86 - 1.07)

Association results for loci that have reached genome-wide statistical significance ($P<5\times10^{-8}$) for association with ESRD in T1D in this or in previous studies: *AFF3* and *RGMA - MCTP2* (Sandholm et al.)[1], EPO (Tong et al.)[11]. Study: GENIE includes the GENIE GWAS discovery cohorts FinnDiane, UK-ROI and GoKinD US; GENIE ESRD*: ESRD phenotype defined as ESRD vs. non-ESRD as in the original publication.[1]

Supplemental Table 4: Association analysis results for rs4972593 in FIND.

				Women					Men				
Study	SNP	A1/A2	Freq	Cases	Controls	Power	*P*	OR	Cases	Controls	Power	*P*	OR
AA	rs4972590	A/G	0.35	570	239	NA	0.58	0.94	408	74	NA	0.42	0.85
AA	rs530673	C/G	0.62	570	239	NA	0.29	0.88	408	74	NA	0.07	0.69
AI	rs4972590	A/G	0.03	323	238	NA	NA	NA	214	81	NA	NA	NA
AI	rs530673	C/G	0.13	323	238	NA	NA	NA	215	81	NA	NA	NA
EA	rs4972590	A/G	0.16	165	236	0.85	0.88	0.97	177	168	NA	NA	NA
EA	rs530673	C/G	0.17	165	236	0.86	0.84	0.96	177	168	NA	NA	NA
MA	rs4972590	A/G	0.09	411	412	0.93	NA	NA	365	179	NA	NA	NA
MA	rs530673	C/G	0.14	413	414	0.98	0.35	0.87	366	180	NA	NA	NA

Association analysis results for the four FIND studies. AA: African Americans, AI: American Indians, EA: European Americans, MA: Mexican Americans. Two proxies for rs4972593 were found in the FIND Study: rs4972590 and rs530673. A1/A2: A1 is the reference allele and in LD with the minor A allele of rs4972593 (HapMap2 CEU). Freq is the A1 allele frequency. Cases: number of cases; Controls: number of controls. Power: Statistical power to detect association of OR 1.75 with α=0.05 significance level. Power was not estimated for AA study because no proxies were found for rs4972593 in the YRI HapMap2 population; in other studies full linkage (D'=1) was assumed between rs4972593 and the proxies. OR>1 indicates A1 as a risk allele for ESRD. NA is given for analyses that were not performed due to too low minor allele count or high genotype missingness.

Supplemental Table 5: In silico prediction of TFBSs that are lost or created due to rs4972592 minor **A** allele.

TFBS ID	Effect	From		To	Strand	Core Similarity	Matrix similarity	TFBS description	TFBS family
V$CREB/XBP1.01	lost	174462848	-	174462868	+	1.000	0.882	X-box-binding protein 1	cAMP-responsive element binding proteins
V$DMRT/DMRT5.01	new	174462844	-	174462864	-	0.808	0.800	Doublesex and mab-3 related transcription factor 5	DM domain-containing transcription factors
V$EBOX/NMYC.01	lost	174462851	-	174462863	+	1.000	0.967	N-Myc	E-box binding factors
V$HAND/SCX.01	lost	174462847	-	174462867	-	0.941	0.938	Tendon-specific bHLH transcription factor scleraxis	Twist subfamily of class B bHLH transcription factors
V$HESF/DEC2.01	lost	174462850	-	174462864	-	1.000	0.983	Basic helix-loop-helix protein known as Dec2 or Sharp2	Vertebrate homologues of enhancer of split complex
V$HESF/DEC2.01	lost	174462851	-	174462865	+	1.000	0.965	Basic helix-loop-helix protein known as Dec2 or Sharp2	Vertebrate homologues of enhancer of split complex
V$HIFF/ARNT.01	lost	174462849	-	174462865	+	1.000	0.973	AhR nuclear translocator homodimers	Hypoxia inducible factor, bHLH/PAS protein family
V$HIFF/HRE.02	lost	174462850	-	174462866	-	1.000	0.977	Hypoxia-response elements	Hypoxia inducible factor, bHLH/PAS protein family
V$OCT1/POU3F3.01	lost	174462855	-	174462869	+	1.000	0.834	POU class 3 homeobox 3 (POU3F3), OTF8	Octamer binding protein

TFBSs were predicted with Genomatix software package. Effect: effect of the minor A allele on the TFBS. From - To: Predicted TFBS bp location, NCBI build 37; rs4972592 position is 174,462,854 on the same build. Core Similarity: Sequence similarity with the core sequence of a TFBS matrix, which is defined as the (usually 4) highest conserved positions of the TFBS matrix. Matrix similarity: Sequence similarity with the TFBS defining matrix.

Supplemental Table 6: *In silico* predicted estrogen responsive elements within **5** kbp up- or downstream of rs4972593.

TFBS ID	TFBS description	From bp	To bp	Distance (kbp)	Strand	Core Similarity	Matrix Similarity	Sequence
V$ESRRB.01	Estrogen-related receptor beta	174,458,480	174,458,448	-4.4	-	1	0,942	agtaaAGGTcactcttgctatgc
V$ESRRA.02	Estrogen-related receptor alpha (secondary DNA binding preference)	174,459,984	174,459,962	-2.9	-	1	0,928	tttgggGGTCacatccaagaaat
V$ESRRA.02	Estrogen-related receptor alpha (secondary DNA binding preference)	174,460,830	174,460,810	-2.0	-	1	0,942	ctatggGGTCaacatttttagat
V$ERR.01	Estrogen related receptor	174,461,494	174,461,516	-1.4	+	1	0,915	gtccAAGGacaaacagctagtca
V$ER.04	Estrogen response elements, IR3 sites	174,461,682	174,461,660	-1.2	-	1	0,889	tcctcagGTCAttaggtcataag
V$ESRRA.01	Estrogen-related receptor alpha	174,461,826	174,461,804	-1.1	-	1	0,888	tttcAAGGtaatataaatcctgt
V$ESRRA.01	Estrogen-related receptor alpha	174,464,202	174,464,224	1.3	-	1	0,88	ttccAAGGtaattcaatggggaa
V$ESRRA.02	Estrogen-related receptor alpha (secondary DNA binding preference)	174,466,672	174,466,694	3,8	+	1	0,943	ttgagaGGTCatacttttttgt

TFBSs were predicted with Genomatix software package. From bp: Start of the ERE in NCBI build 37. To bp: End of the ERE in NCBI build 37. rs4972593 position is 174,462,854 on the same build. Distance (kbp): Distance between ERE and rs4972593 in kbp. Core Similarity: Sequence similarity with the core sequence of a TFBS matrix, which is defined as the (usually 4) highest conserved positions of the TFBS matrix. Matrix similarity: Sequence similarity with the TFBS defining matrix. Sequence: TFBS matching sequence; red color indicates highly conserved bases (consensus index vector > 60); capital letters indicate the core TFBS sequence.

Supplemental Table 7: Evidence of DNA features and regulatory elements overlapping SNPs in high linkage with rs4972593 (r2>0.8 in CEU) according to the RegulomeDB, ENCODE project.[7]

SNP	BP	Distance	r^2	*P* GWAS	Score	Experiment	Overlapping feature	Motif
rs530673	174162256	-8,8 kbp	1	3.5×10^{-8}	3a	Protein binding, ChIP-Seq	GATA2 binding region in HUVEC cells (umbilical vein endothelial cells)	
						Motifs, PWM	SMAD4 binding motif	
						Chromatin structure, DNase-Seq	DNase hypersensitivity peak in A549 cells, indicating chromatin accessibility	
						Histone modifications, ChIP-seq	H3k27ac histone mark on 7 cell lines (often found near active regulatory elements).	
						Histone modifications, ChIP-seq	Multiple histone marks in multiple cell types.	
rs4972593	174171100	-	-	3.0×10^{-8}	6			
rs4972590	174168185	-2,9 kbp	1	NA	6			
rs4972591	174168287	-2,8 kbp	1	NA	NA			

Evidence is specified only for SNPs with RegulomeDB scores 1-3 corresponding to high – moderate evidence. BP: Base pair position, human genome b36. Distance, kb: Distance to rs4972593 in kbp. r^2: r^2 correlation measure of LD in the CEU population (HapMap Rel 27, Phase II+III, Feb09, CEU, dbSNP b126). *P* GWAS: *P*-value in the discovery GWAS in women. Score: RegulomeDB score for evidence of regulative activity on the scale of 1-6. 3a: TF binding + any motif + DNase peak, "Less likely to affect binding". 6: "Minimal binding evidence". Experiment: RegulomeDB experiment type. Overlapping feature: Short explanation of the experimental results overlapping the SNP position. Motif: image of the protein binding motif sequence.

Supplemental Table 8: eQTL association of rs4972593 and SNPs in LD with rs4972593.

SNP	Pop	Gene	rho	P	Pemp	Pbonf
rs530673	LWK	*GPR155*	-0.218	0.049	0.050	1
rs530673	JPT	*PDK1*	-0.239	0.030	0.030	1
rs4972590	MKK	*CDCA7*	-0.174	0.041	0.045	1

All the eQTL associations with uncorrected $P<0.05$. Three SNPs in LD with rs4972593 were found in the lymphoblastoid cell line eQTL data base (rs530673, rs4972590, rs4972591). Pop: HapMap3 population; LWK: Luhya in Webuye, Kenya; JPT: Japanese in Tokyo, Japan; MKK: Maasai in Kinyawa, Kenya. Pemp: empirical *P*-value based on 10,000 permutations. Pbonf: *P*-value after Bonferroni correction for multiple testing.

Supplemental Table 9: Differential expression of *SP3* in kidney biopsies

Type	Ref	Comparison name	Direction	*P*-value	P_{COR}	Fold change	Rank
DN	Schmid	Group: Diabetic Nephropathy vs. Minimal Change Disease and Control	over	0.008	0.08	1.347	987
DN	Woroniecka	Tubulointerstitium: Diabetic Nephropathy vs. Healthy Living Donor	over	0.013	ns	1.626	1547
DN	Woroniecka	Glomeruli: Diabetic Nephropathy vs. Healthy Living Donor	over	0.056	ns	1.291	2337
Sex	Schmid	Diabetic Nephropathy: Sex	over	0.23	ns	1.112	4407
Sex	Woroniecka	Glomeruli: Sex	under	0.004	0.04	-1.454	31 (top 3‰)
Sex	Woroniecka	Healthy Living Donor Glomeruli: Sex	under	0.009	0.09	-1.432	180 (top 2%)
Sex	Woroniecka	Diabetic Nephropathy Glomeruli: Sex	under	0.12	ns	-1.375	1287
Sex	Woroniecka	Tubulointerstitium: Sex	under	0.14	ns	-1.239	2499
Sex	Woroniecka	Diabetic Nephropathy Tubulointerstitium: Sex	under	0.40	ns	-1.074	6483
Sex	Woroniecka	Healthy Living Donor Tubulointerstitium: Sex	under	0.41	ns	-1.062	6611

Type: analysis type, either comparison of differential expression in DN vs. non-DN biopsies ("DN"), or in men vs. women ("Sex"). Ref: Schmid *et al.*[8] or Woroniecka KI *et al.*[9] Comparison: name in the Nephromine data base (www.nephromine.org). Direction: over/under-expression according to the fold change. *P*-value: *P*-value for the differential expression, unadjusted for multiple testing. P_{COR}: *P*-value after adjustment for multiple testing (10 performed look-ups), P_{COR}=*P*-value × 10. Fold change: positive, if expression is higher in samples defined as cases (DN or men). Rank: gene ranking based on the P-value within the comparison. Rank is given according to the direction of the fold change for *SP3*. For the comparison "Glomeruli: Sex, under-expression" the minimum, median, and maximum fold changes were -6.11, -1.01 and 10.83, respectively, and the inter-quartile range was -1.08 – 1.05. A total 12,561 genes were included for this comparison.

Supplemental Table 10: Genotype quality for rs4972593 and rs530673.

		Imputed SNP		Genotyped SNP					
					P-HWE				
		Quality	Rsq	Call rate	All	control women	case women	control men	case men
FinnDiane	rs4972593	0.99	0.95						
	rs530673	0.99	0.97						
UK-ROI	rs4972593	0.99	0.98						
	rs530673	0.996	0.99						
GoKinD US	rs4972593	0.85	0.48						
	rs530673	0.85	0.48						
Italy	rs4972593			0.95	0.85	0.68	1	1	1

Both SNPs were imputed in FinnDiane, UK-ROI and GoKinD US. Rs4972593 was directly genotyped for the Italian study. For the imputed SNPs, the MACH Quality measure (the average posterior probability for the most likely genotype) and Rsq value (estimate of the squared correlation between imputed and true genotypes; According to the MACH tutorial, a cut-off of 0.30 will flag most of the poorly imputed SNPs, but only a small number (<1%) of well imputed SNPs) are given. For the genotyped SNP, call rate and *P*-value for Hardy Weinberg disequilibrium (*P*-HWE) are given.

Supplemental Table 11: FinnDiane physicians and nurses at each center participating in the collection of the FinnDiane patients.

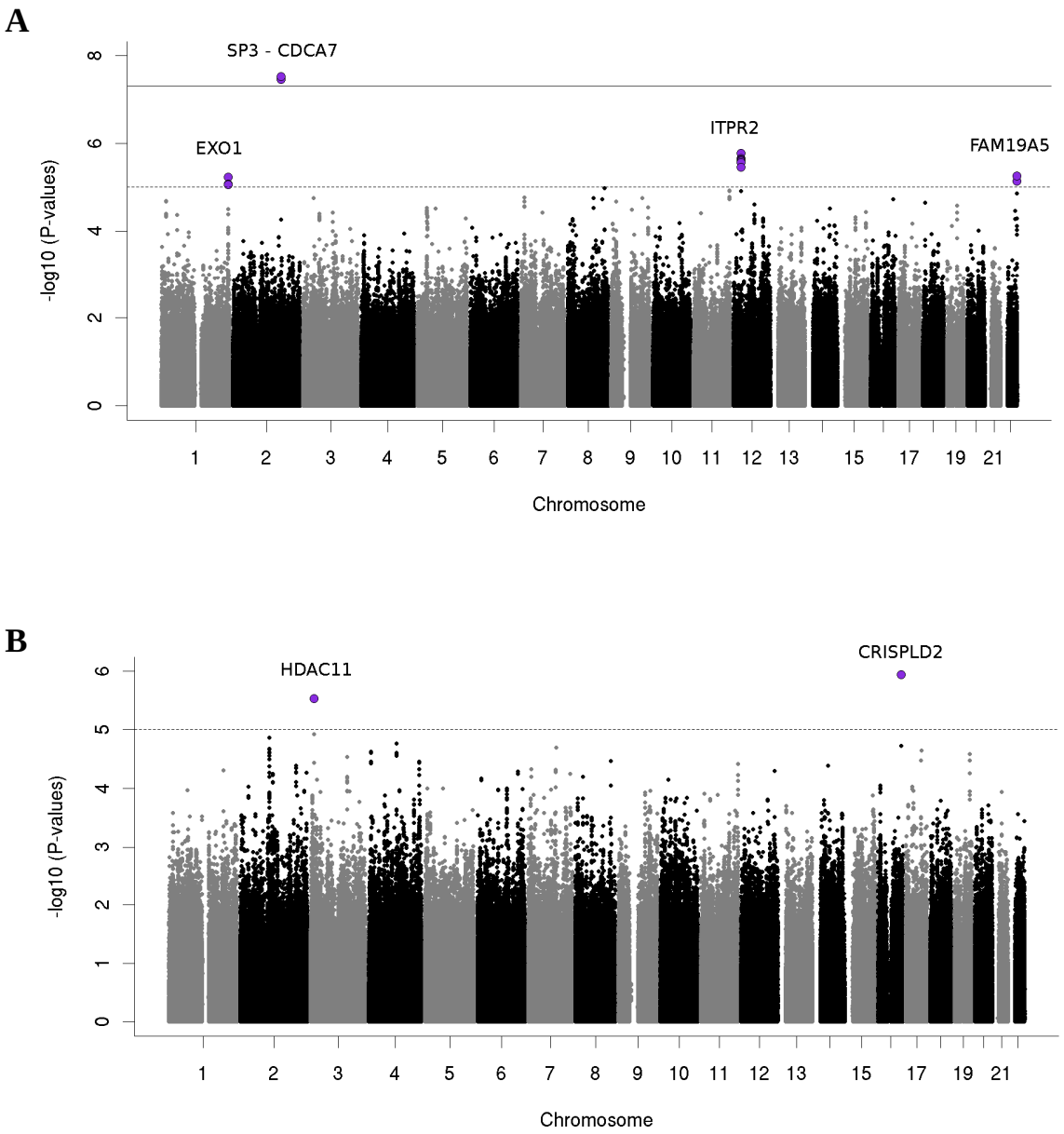
FinnDiane Study Centers	Physicians and nurses
Anjalankoski Health Centre	S. Koivula, T. Uggeldahl
Central Finland Central Hospital, Jyväskylä	T. Forslund, A. Halonen, A. Koistinen, P. Koskiaho, M. Laukkanen, J. Saltevo, M. Tiihonen
Central Hospital of Åland Islands, Mariehamn	M. Forsen, H. Granlund, A-C. Jonsson, B. Nyroos
Central Hospital of Kanta-Häme, Hämeenlinna	P. Kinnunen, A. Orvola, T. Salonen, A. Vähänen
Central Hospital of Länsi-Pohja, Kemi	H. Laukkanen, P. Nyländen, A. Sademies
Central Ostrabothnian Hospital District, Kokkola	S. Anderson, B. Asplund, U. Byskata, P. Liedes, M. Kuusela, T. Virkkala
City of Espoo Health Centre	
Espoonlahti	A. Nikkola, E. Ritola
Tapiola	M. Niska, H. Saarinen
Samaria	E. Oukko-Ruponen, T. Virtanen
Viherlaakso	A. Lyytinen
City of Helsinki Health Centre	
Puistola	H. Kari, T. Simonen
Suutarila	A. Kaprio, J. Kärkkäinen, B. Rantaeskola
Töölö	P. Kääriäinen, J. Haaga, A-L. Pietiläinen
City of Hyvinkää Health Centre	S. Klemetti, T. Nyandoto, E. Rontu, S. Satuli-Autere
City of Vantaa Health Centre	
Korso	R. Toivonen, H. Virtanen
Länsimäki	R. Ahonen, M. Ivaska-Suomela, A. Jauhiainen
Martinlaakso	M. Laine, T. Pellonpää, R. Puranen
Myyrmäki	A. Airas, J. Laakso, K. Rautavaara
Rekola	M. Erola, E. Jatkola
Tikkurila	R. Lönnblad, A. Malm, J. Mäkelä, E. Rautamo
Heinola Health Centre	P. Hentunen, J. Lagerstam

FinnDiane Study Centers	Physicians and nurses
Helsinki University Central Hospital, Department of Medicine, Division of Nephrology	A. Ahola, M. Feodoroff, D. Gordin, O. Heikkilä, K Hietala, J. Kytö, S. Lindh, M. Parkkonen, K. Pettersson-Fernholm, M. Rosengård-Bärlund, A. Sandelin, A-R Salonen, L. Salovaara, M. Saraheimo, T. Soppela, A. Soro-Paavonen, L. Thorn, N. Tolonen, J. Tuomikangas, T. Vesisenaho, J. Wadén
Herttoniemi Hospital, Helsinki	V. Sipilä
Hospital of Lounais-Häme, Forssa	T. Kalliomäki, J. Koskelainen, R. Nikkanen, N. Savolainen, H. Sulonen, E. Valtonen
Iisalmi Hospital	E. Toivanen
Jokilaakso Hospital, Jämsä	A. Parta, I. Pirttiniemi
Jorvi Hospital, Helsinki University Central Hospital	S. Aranko, S. Ervasti, R. Kauppinen-Mäkelin, A. Kuusisto, T. Leppälä, K. Nikkilä, L. Pekkonen
Jyväskylä Health Centre, Kyllö	K. Nuorva, M. Tiihonen
Kainuu Central Hospital, Kajaani	S. Jokelainen, P. Kemppainen, A-M. Mankinen, M. Sankari
Kerava Health Centre	H. Stuckey, P. Suominen
Kirkkonummi Health Centre	A. Lappalainen, M. Liimatainen, J. Santaholma
Kivelä Hospital, Helsinki	A. Aimolahti, E. Huovinen
Koskela Hospital, Helsinki	V. Ilkka, M. Lehtimäki
Kotka Heath Centre	E. Pälikkö-Kontinen, A. Vanhanen
Kouvola Health Centre	E. Koskinen, T. Siitonen
Kuopio University Hospital	E. Huttunen, R. Ikäheimo, P. Karhapää, P. Kekäläinen, M. Laakso, T. Lakka, E. Lampainen, L. Moilanen, L. Niskanen, U. Tuovinen, I. Vauhkonen, E. Voutilainen
Kuusamo Health Centre	T. Kääriäinen, E. Isopoussu
Kuusankoski Hospital	E. Kilkki, I. Koskinen, L. Riihelä
Laakso Hospital, Helsinki	T. Meriläinen, P. Poukka, R. Savolainen, N. Uhlenius
Lahti City Hospital	A. Mäkelä, M. Tanner
Lapland Central Hospital, Rovaniemi	L. Hyvärinen, S. Severinkangas, T. Tulokas
Lappeenranta Health Centre	P. Linkola, I. Pulli
Lohja Hospital	T. Granlund, M. Saari, T. Salonen
Loimaa Health Centre	A. Mäkelä, P. Eloranta
Länsi-Uusimaa Hospital, Tammisaari	I-M. Jousmaa, J. Rinne
Malmi Hospital, Helsinki	H. Lanki, S. Moilanen, M. Tilly-Kiesi
Mikkeli Central Hospital	A. Gynther, R. Manninen, P. Nironen, M. Salminen, T. Vänttinen

FinnDiane Study Centers	Physicians and nurses
Mänttä Regional Hospital	I. Pirttiniemi, A-M. Hänninen
North Karelian Hospital, Joensuu	U-M. Henttula, P. Kekäläinen, M. Pietarinen, A. Rissanen, M. Voutilainen
Nurmijärvi Health Centre	A. Burgos, K. Urtamo
Oulankangas Hospital, Oulainen	E. Jokelainen, P-L. Jylkkä, E. Kaarlela, J. Vuolaspuro
Oulu Health Centre	L. Hiltunen, R. Häkkinen, S. Keinänen-Kiukaanniemi
Oulu University Hospital	R. Ikäheimo
Päijät-Häme Central Hospital	H. Haapamäki, A. Helanterä, S. Hämäläinen, V. Ilvesmäki, H. Miettinen
Palokka Health Centre	P. Sopanen, L. Welling
Pieksämäki Hospital	V. Javtsenko, M. Tamminen
Pietarsaari Hospital	M-L. Holmbäck, B. Isomaa, L. Sarelin
Pori City Hospital	P. Ahonen, P. Merensalo, K. Sävelä
Porvoo Hospital	M. Kallio, B. Rask, S. Rämö
Raahe Hospital	A. Holma, M. Honkala, A. Tuomivaara, R. Vainionpää
Rauma Hospital	K. Laine, K. Saarinen, T. Salminen
Riihimäki Hospital	P. Aalto, E. Immonen, L. Juurinen
Salo Hospital	A. Alanko, J. Lapinleimu, P. Rautio, M. Virtanen
Satakunta Central Hospital, Pori	M. Asola, M. Juhola, P. Kunelius, M-L. Lahdenmäki, P. Pääkkönen, M. Rautavirta
Savonlinna Central Hospital	E. Korpi-Hyövälti, T. Latvala, E. Leijala
South Karelia Central Hospital, Lappeenranta	T. Ensala, E. Hussi, R. Härkönen, U. Nyholm, J. Toivanen
Tampere Health Centre	A. Vaden, P. Alarotu, E. Kujansuu, H. Kirkkopelto-Jokinen, M. Helin, S. Gummerus, L. Calonius, T. Niskanen, T. Kaitala, T. Vatanen
Tampere University Hospital	I. Ala-Houhala, T. Kuningas, P. Lampinen, M. Määttä, H. Oksala, T. Oksanen, K. Salonen, H. Tauriainen, S. Tulokas
Tiirismaa Health Centre, Hollola	T. Kivelä, L, Petlin, L. Savolainen
Turku Health Centre	I. Hämäläinen, H. Virtamo, M. Vähätalo
Turku University Central Hospital	K. Breitholz, R. Eskola, K. Metsärinne, U. Pietilä, P. Saarinen, R. Tuominen, S. Äyräpää
Vaajakoski Health Centre	K. Mäkinen, P. Sopanen
Valkeakoski Regional Hospital	S. Ojanen, E. Valtonen, H. Ylönen, M. Rautiainen, T. Immonen
Vammala Regional Hospital	I. Isomäki, R. Kroneld, M. Tapiolinna-Mäkelä
Vaasa Central Hospital	S. Bergkulla, U. Hautamäki, V-A. Myllyniemi, I. Rusk

Supplemental Figure 1: Manhattan plots of the gender specific GWASs for A) women and B) men.

The x-axis indicates the chromosomal position, and y-axis gives the statistical significance as $-\log_{10}(P\text{-value})$. The solid horizontal line indicates the threshold for the genome-wide statistical significance ($P<5\times10^{-8}$), and the dashed horizontal line indicates a suggestive *P*-value threshold $P<10^{-5}$. Signals above this line are marked with purple dots and listed in Supplemental Table 2 as well.



References

1. Sandholm N, Salem RM, McKnight AJ, Brennan EP, Forsblom C, Isakova T, McKay GJ, Williams WW, Sadlier DM, Makinen VP, Swan EJ, Palmer C, Boright AP, Ahlqvist E, Deshmukh HA, Keller BJ, Huang H, Ahola AJ, Fagerholm E, Gordin D, Harjutsalo V, He B, Heikkila O, Hietala K, Kyto J, Lahermo P, Lehto M, Lithovius R, Osterholm AM, Parkkonen M, Pitkaniemi J, Rosengard-Barlund M, Saraheimo M, Sarti C, Soderlund J, Soro-Paavonen A, Syreeni A, Thorn LM, Tikkanen H, Tolonen N, Tryggvason K, Tuomilehto J, Waden J, Gill GV, Prior S, Guiducci C, Mirel DB, Taylor A, Hosseini SM, DCCT/EDIC Research Group, Parving HH, Rossing P, Tarnow L, Ladenvall C, Alhenc-Gelas F, Lefebvre P, Rigalleau V, Roussel R, Tregouet DA, Maestroni A, Maestroni S, Falhammar H, Gu T, Mollsten A, Cimponeriu D, Ioana M, Mota M, Mota E, Serafinceanu C, Stavarachi M, Hanson RL, Nelson RG, Kretzler M, Colhoun HM, Panduru NM, Gu HF, Brismar K, Zerbini G, Hadjadj S, Marre M, Groop L, Lajer M, Bull SB, Waggott D, Paterson AD, Savage DA, Bain SC, Martin F, Hirschhorn JN, Godson C, Florez JC, Groop PH, Maxwell AP: New susceptibility loci associated with kidney disease in type 1 diabetes. [Electronic version]. *PLoS Genet* 8: e1002921, 2012

2. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: A tool set for whole-genome association and population-based linkage analyses. [Electronic version]. *Am J Hum Genet* 81: 559-575, 2007

3. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. [Electronic version]. *Nat Genet* 38: 904-909, 2006

4. Willer CJ, Li Y, Abecasis GR: METAL: Fast and efficient meta-analysis of genomewide association scans. [Electronic version]. *Bioinformatics* 26: 2190-2191, 2010

5. Purcell S, Cherny SS, Sham PC: Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. [Electronic version]. *Bioinformatics* 19: 149-150, 2003

6. Zhang J, & Yu KF: What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. [Electronic version]. *JAMA* 280: 1690-1691, 1998

7. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M: Annotation of functional variation in personal genomes using RegulomeDB. [Electronic version]. *Genome Res* 22: 1790-1797, 2012

8. Schmid H, Boucherot A, Yasuda Y, Henger A, Brunner B, Eichinger F, Nitsche A, Kiss E, Bleich M, Grone HJ, Nelson PJ, Schlondorff D, Cohen CD, Kretzler M, European Renal cDNA Bank

(ERCB) Consortium: Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. [Electronic version]. *Diabetes* 55: 2993-3003, 2006

9. Woroniecka KI, Park AS, Mohtat D, Thomas DB, Pullman JM, Susztak K: Transcriptome analysis of human diabetic kidney disease. [Electronic version]. *Diabetes* 60: 2354-2369, 2011

10. Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET: Patterns of cis regulatory variation in diverse human populations. [Electronic version]. *PLoS Genet* 8: e1002639, 2012

11. Tong Z, Yang Z, Patel S, Chen H, Gibbs D, Yang X, Hau VS, Kaminoh Y, Harmon J, Pearson E: Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. [Electronic version]. *Proceedings of the National Academy of Sciences* 105: 6998, 2008