Genome-Wide Association Scan for Diabetic Nephropathy Susceptibility Genes in Type 1 Diabetes

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

Published Version
doi:10.2337/db08-1514

Accessed
September 13, 2017 2:50:04 PM EDT

Citable Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4881556

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)
Genome-Wide Association Scan for Diabetic Nephropathy Susceptibility Genes in Type 1 Diabetes

Marcus G. Pezzolesi,1 G. David Poznik,1 Josyf C. Mychaleckyj,2 Andrew D. Paterson,3,4 Michelle T. Barati,5 Jon B. Klein,5 Daniel P.K. Ng,1,6 Grzegorz Placha,1,7 Luis H. Canani,1,8 Jacek Bochenksi,1 Daryl Waggott,9 Michael L. Merchant,5 Bozena Krolewski,1 Lucia Mirea,4,9 Krzysztof Waniec,1 Pisut Katavetin,1 Masahiko Kure,1 Pawel Wolkow,1,10 Jonathon S. Dunn,1 Adam Smiles,1 William H. Walker,1 Andrew P. Boright,11 Shelley B. Bull,4,9 the DCCT/EDIC Research Group,* Alessandro Doria,1 John J. Rogus,1 Stephen S. Rich,2 James H. Warram,1 and Andrzej S. Krolewski1

OBJECTIVE—Despite extensive evidence for genetic susceptibility to diabetic nephropathy, the identification of susceptibility genes and their variants has had limited success. To search for genes that contribute to diabetic nephropathy, a genome-wide association scan was implemented on the Genetics of Kidneys in Diabetes collection.

RESEARCH DESIGN AND METHODS—We genotyped ~360,000 single nucleotide polymorphisms (SNPs) in 820 case subjects (284 with proteinuria and 536 with end-stage renal disease) and 885 control subjects with type 1 diabetes. Confirmation of implicated SNPs was sought in 1,304 participants of the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) study, a long-term, prospective investigation of the development of diabetes-associated complications.

RESULTS—A total of 13 SNPs located in four genomic loci were associated with diabetic nephropathy with \( P < 1 \times 10^{-5} \). The strongest association was at the FRMD3 locus (4.1 single nucleotide polymorphism (SNP) in the FERM domain containing 3) locus (odds ratio \( [OR] \)) was 1.45, \( P = 5.0 \times 10^{-7} \). A strong association was also identified at the CARS locus (cysteinyl-tRNA synthetase) locus (OR = 1.36, \( P = 3.1 \times 10^{-5} \)). Associations between both loci and time to onset of diabetic nephropathy were supported in the DCCT/EDIC study (hazard ratio \( [HR] \) was 1.33, \( P = 0.02 \), and \( HR = 1.32, P = 0.01 \), respectively). We demonstrated expression of both FRMD3 and CARS in human kidney.

CONCLUSIONS—We identified genetic associations for susceptibility to diabetic nephropathy at two novel candidate loci near the FRMD3 and CARS genes. Their identification implicates previously unsuspected pathways in the pathogenesis of this important late complication of type 1 diabetes. Diabetes, 58:1403–1410, 2009

Diabetic nephropathy is the leading contributor to end-stage renal disease (ESRD) in the U.S. (1). Clinically, diabetic nephropathy is manifest as a progressive disease process that advances through characteristic stages. It begins with microalbuminuria (leakage of small amounts of albumin into the urine) and progresses to overt proteinuria. In a large proportion of these patients, renal function declines and continues to deteriorate until ESRD is reached, and replacement therapy is indicated (2–4). Overall, ESRD develops in ~20% of all patients with type 1 diabetes (5,6).

Despite evidence that genetic susceptibility plays a role in the development of diabetic nephropathy in type 1 diabetes (7–9), success in identifying the responsible genetic variants has been limited (10,11). This has been attributable, in part, to the small size of the DNA collections available to individual research groups and the narrow focus of the searches on candidate genes. Another challenge that has received little attention in previous studies is the possibility that successive stages of diabetic nephropathy are influenced by different genetic factors (12,13).

To conduct a statistically robust study that provides genome-wide coverage for detection of common variants that may have a small, but pathogenically significant, effect on risk of diabetic nephropathy in type 1 diabetes, the Genetics of Kidneys in Diabetes (GoKinD) collection was established (14). A genome-wide scan of this collection was supported by the Genetic Association Information Network (GAIN) initiative (15). This report presents 1) results of this genome-wide association scan in the GoKinD collection, 2) replication of the significant associations in this scan with time to onset of diabetes-associated complications (severe nephropathy) in the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) study, and 3) characterization of expression of the
identified candidate diabetic nephropathy genes in normal human cell lines.

RESEARCH DESIGN AND METHODS

Subjects for the GoKinD collection were recruited through two centers with different methods of ascertainment and recruitment (14). The George Washington University (GWU) Biostatistics Center coordinated the recruitment of volunteers (through mass media advertisement) living throughout the U.S. (excluding New England) and Canada to 1 of 27 clinical centers located across the U.S. and Canada. The Section of Genetics and Epidemiology at the Joslin Diabetes Center (JDC) recruited and examined patients of the Joslin Clinic from New England who were already enrolled in the Joslin Kidney Study on the Genetics of Diabetic Nephropathy, a clinic-based cohort study in which case subjects with diabetic nephropathy and a random sampling of eligible control subjects were identified and recruited (16).

A detailed description of the GoKinD collection has been published (14). Briefly, subjects enrolled in GoKinD had type 1 diabetes diagnosed before age 31, began insulin treatment within 1 year of their diagnosis, and were between 18 and 59 years of age at the time of enrollment. Participation in the DCCT/EDIC study was an exclusion criterion so that the two study populations would be independent. Case subjects with diabetic nephropathy had either persistent proteinuria, defined by a urinary albumin-to-creatinine ratio ≥300 μg/g in two of the last three measurements taken at least 1 month apart, or diabetic renal disease or renal transplant. Control subjects had type 1 diabetes for at least 15 years and normoalbuminuria, defined by an albumin-to-creatinine ratio <20 μg/g in two of the last three measurements taken at least 1 month apart (if a third measurement was required, a value <40 μg/g was necessary for inclusion), without ever having been treated with ACE inhibitors or angiotensin receptor blockers, and they were not being treated with antihypertensive medication at the time of recruitment into the study. For additional information regarding the definition of case and control subjects used in this analysis, refer to the report by Mueller et al. (14). In total, 1,879 subjects (935 case and 944 control subjects) were recruited into GoKinD. The GWU panel included 437 case subjects with diabetic nephropathy (58 with proteinuria and 379 with ESIRD) and 446 control subjects; the JDC panel included 498 case subjects with diabetic nephropathy (258 with proteinuria and 230 with ESIRD) and 498 control subjects. Further details are also provided in the supplementary information, which is available in an online appendix at http://diabetes.diabetesjournals.org/cgi/content/full/db08-1514/DC1.

Confirmation of our findings in the GoKinD collection was sought in genome-wide association data from the DCCT/EDIC study, a long-term, prospective investigation of the development of diabetes-associated complications (17,18). Of the original DCCT cohort recruited between 1983 and 1989, 1,575 subjects remained in the EDIC follow-up study. Participants in EDIC underwent baseline examinations between 1994 and 1995 and have since participated in annual follow-up examinations to assess the development or progression of complications. As of EDIC year 12 (2005), this cohort had 10–22 years of follow-up, and 132 cases of severe nephropathy (proteinuria or ESIRD) had been documented in 1,304 Caucasian DCCT/EDIC participants. The close correspondence of the closest SNPs to the lead SNPs in the GoKinD collection. Detailed clinical characteristics of the DCCT/EDIC study have been published (13,17,18). Additional details are also provided in the supplementary information.

Genotyping. The GoKinD collection was genotyped on an Affymetrix 5.0 500K single nucleotide polymorphism (SNP) array by the GAIN genotyping laboratory at the Eli and Edythe L. Broad Institute (Cambridge, MA). A description of study genotyping is available in the supplementary information. Additionally, two SNPs, rs698075 and rs6888747, were genotyped in the GoKinD collection using TaqMan (Applied Biosystems, Foster City, CA) technology by the genetics core of the Diabetes and Endocrinology Research Center at the JDC in accordance with the manufacturer’s protocols. DNA samples used for the genotyping of these SNPs in the GoKinD collection were obtained through the National Institute of Diabetes and Digestive and Kidney Diseases central repository (www.niddkreporatory.org).

SNP quality control. After internal quality control, the GAIN genotyping laboratory released genotypes for 467,144 SNPs. Several quality control metrics, including filters for minor allele frequency <0.01, rejection of Hardy-Weinberg assumptions (P > 10⁻⁵), and differential rates of missing data (by case/control status) were applied to these data. After reconciliation of SNPs eliminated by these analyses, the resulting data contained 359,193 autosomal SNPs. More details are available in the supplementary information and supplementary Table 1.

Sample quality control and SNP imputation. The application of quality control criteria reduced the GoKinD population from 1,879 to 1,705 individuals of European ancestry. Samples from the two GoKinD panels that constitute this sample are 379 GWU case subjects (49 with proteinuria and 330 with ESIRD), 413 GWU control subjects, 441 JDC case subjects (235 with proteinuria and 206 with ESIRD), and 472 JDC control subjects. The measured genotypes for these individuals were augmented by imputation of ungenotyped SNPs using two nearest-neighbors base pair regions flanking each of the lead SNPs in the GoKinD collection (where linkage disequilibrium, as measured by r², decayed to <0.20 for all lead SNPs). A total of 8,245 imputed HapMap SNPs across these four loci were included in our association analysis. Further details of sample quality control and imputation procedures are available in the supplementary information and supplementary Table 2.

Statistical analysis. Genotype data were collected under two separate ascertainment protocols (JDC and GWU panels) so that tests for latent residual stratification were performed. Cochran-Armitage trend tests of the JDC versus GWU control subjects, and JDC versus GWU case subjects, revealed an overdispersion in the test statistics for both control and case subjects compared with the complete null. The median genomic control parameters were estimated at λ = 1.13 for control subjects and λ = 1.097 for case subjects. Permutation control measures within the case and control subjects were included in a stratification significance of P < 10⁻⁵ for both case and control subjects. Therefore, the primary association analysis used in the study was a stratified test of association combining case-control tests of allele frequencies in JDC and GWU strata. Combined P values and odds ratios (ORs) were calculated using a Cochran-Mantel-Haenszel procedure. Homogeneity across strata was assessed using the Breslow-Day statistic. All genome-wide association analysis results are provided using the R package. Further details of quality control procedures, software, statistical analysis and adjustments, and cluster plots are available in the supplementary information. Data from our analysis of the GoKinD collection are available for specific SNPs and/or genes upon request.

Genotyping and analysis of the DCCT/EDIC data. Genotypes of the DCCT/EDIC study participants were generated with the Illumina Human1M Beadchip. Briefly, quality control measures resulted in 840,354 SNPs suitable for statistical analysis. Population substructure was assessed to ensure that all included samples were of European ancestry. Multivariate Cox proportional hazard analyses were performed on data from 1,304 Caucasian subjects using time to onset of severe nephropathy, defined by an albumin excretion rate (AER) >300 mg/24 h at least two consecutive examinations or dialysis/renal transplantation. Of these, 164 progressed to ESRD, with 110 with progressive renal failure (proteinuria >300 mg/24 h) as the outcome phenotype (n = 132). Among those with severe nephropathy, 116 subjects developed only proteinuria (AER >300 mg/24 h), whereas 16 progressed to ESRD. The DCCT cohort (primary prevention versus secondary intervention), treatment (intensive versus conventional), and interaction between cohort and treatment were used as covariates in the analysis of the effect of an independent additive SNP genetic factor. This model was examined for all associated loci in GoKinD and subsequently tested for both statistical significance and the same direction of effect for associated alleles (20–22).

Gene expression. The expression of candidate genes was examined in four primary human cell lines derived from cells that have been implicated in the pathogenesis of kidney complications (endothelial cells from the iliac artery, adult dermal fibroblasts, mesangial cells, and epithelial cells from proximal tubules) by quantitative real-time PCR. Sources of these cells, cell culture conditions, and protocols used in these experiments are available in the supplementary information.

RESULTS

Genome-wide association scans for genes associated with diabetic nephropathy in type 1 diabetes. The application of metrics for SNP and sample quality resulted in the analysis of 359,193 autosomal SNPs and 1,705 GoKinD samples of European ancestry (885 control subjects and 820 case subjects) (see RESEARCH DESIGN AND METHODS and the supplementary information). Clinical characteristics of the JDC and GWU panels are summarized in Table 1. Because different ascertainment protocols were used by the JDC and GWU, the resulting data were found to exhibit significant stratification. As a result, the primary association analyses were conducted using a stratified test of association.

Although no SNP achieved genome-wide significance (0.05/359,193 = 1.4 × 10⁻⁵), the primary association analysis identified 11 SNPs representing four distinct chromosomal regions with P < 1 × 10⁻⁵ (Fig. 1 and Table...
Finally, the region bounded by rs1411766/rs1742858 (OR 2), which were considered for replication. The strongest association with diabetic nephropathy occurred on chromosome 9q with rs10868025 (OR 2.10), which were considered for replication. The strongest associations were more strongly associated with diabetic nephropathy than the original associations. Of these, two SNPs in the 7p region (rs39075 and rs39076) were also associated with diabetic nephropathy than the original SNP in that region (rs39059) (Fig. 2A). Similarly, two imputed SNPs in the 7p region (rs39075 and rs39076) were also more strongly associated than the original SNP in that region (rs39059) (Fig. 2A). Both imputed SNPs were genotyped in the GoKinD samples, and the associations with the imputed data were confirmed (rs39075, P = 6.5 × 10^{-7}; and rs1888747, P = 6.3 × 10^{-7}) (Table 2).

If the etiology of diabetic nephropathy involves the interaction of a locus with the cumulative effect of hyperglycemia, the association of the locus with diabetic nephropathy can vary according to diabetes duration at diabetic nephropathy onset, such that it is strongest in early-appearing case subjects and diminishes in later ones, even reversing in direction in very late-appearing case subjects (23). We examined the SNPs in Table 2 according to diabetes duration by stratifying case and control subjects across tertiles of diabetes duration (at the onset of ESRD or at enrollment into GoKinD for proteinuria patients and control subjects). The strength of the associations was consistent across these strata (data not shown).

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>JDC GoKinD control subjects</th>
<th>JDC GoKinD case subjects</th>
<th>GWU GoKinD control subjects</th>
<th>GWU GoKinD case subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>472</td>
<td>441</td>
<td>413</td>
<td>379</td>
</tr>
<tr>
<td>Men/women</td>
<td>220/252</td>
<td>233/208</td>
<td>143/270</td>
<td>190/189</td>
</tr>
<tr>
<td>Age at diagnosis of type 1 diabetes (years)</td>
<td>12.9 ± 7.5</td>
<td>11.5 ± 6.7</td>
<td>12.9 ± 7.2</td>
<td>12.1 ± 6.6</td>
</tr>
<tr>
<td>Duration of type 1 diabetes (years)*</td>
<td>26.4 ± 8.0</td>
<td>27.0 ± 7.8</td>
<td>24.2 ± 7.3</td>
<td>25.9 ± 7.6</td>
</tr>
<tr>
<td>Age at examination (years)</td>
<td>39.4 ± 6.5</td>
<td>42.0 ± 7.1</td>
<td>37.1 ± 8.8</td>
<td>44.4 ± 6.5</td>
</tr>
<tr>
<td>Laser treatment (%)</td>
<td>22</td>
<td>84</td>
<td>10</td>
<td>85</td>
</tr>
<tr>
<td>A1C (%)†</td>
<td>7.5 ± 1.7</td>
<td>8.5 ± 1.6</td>
<td>7.4 ± 1.1</td>
<td>7.8 ± 1.4</td>
</tr>
<tr>
<td>ACR in control subjects (μg/mg)</td>
<td>7.0 ± 3.7</td>
<td>—</td>
<td>6.3 ± 3.6</td>
<td>—</td>
</tr>
<tr>
<td>Case subjects with proteinuria/ESRD</td>
<td>233/206</td>
<td>1,571 ± 1,534</td>
<td>1,282 ± 1,096</td>
<td>49/330</td>
</tr>
<tr>
<td>ACR in subjects with proteinuria (μg/mg)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ESRD duration (years)</td>
<td>7.5 ± 5.4</td>
<td>7.5 ± 5.4</td>
<td>7.2 ± 5.6</td>
<td>7.2 ± 5.6</td>
</tr>
<tr>
<td>Kidney transplant (%)‡</td>
<td>86</td>
<td>86</td>
<td>95</td>
<td>95</td>
</tr>
</tbody>
</table>

Clinical characteristics are the mean values ± SD for all Caucasian patients (n = 1,705) included in the current analysis after the application of quality control metrics (see supplemental information regarding sample quality control analysis and population substructure and ancestry analysis). *The duration of type 1 diabetes in control subjects and in subjects with proteinuria is based on the duration at examination. Among ESRD case subjects, this is based on the duration of type 1 diabetes at the onset of ESRD. All other clinical characteristics are based on measurements performed at examination. †Mean A1C values do not include data from case subjects that have undergone pancreas transplantation (11% of JDC case subjects and 58% of GWU case subjects); ‡percentages are of ESRD group.

2), which were considered for replication. The strongest association with diabetic nephropathy occurred on chromosome 9q with rs10868025 (OR = 1.45, P = 5.0 × 10^{-7}). This SNP is located near the 5’ end of the 4.1 protein ezrin, radixin, moesin (FERM) domain-containing 3 (FRMD3) gene.

Three additional genomic regions located on chromosomes 7p, 11p, and 13q were also associated with diabetic nephropathy. The rs39059 SNP (OR = 1.39, P = 5.0 × 10^{-6}) localizes to the first intron of CHN2 (β-chimerin) isoform 2 and upstream of an alternatively spliced CPVL (serine carboxypeptidase vitellogenic-like) transcript on chromosome 7p. The rs451041 SNP (OR = 1.45, P = 5.0 × 10^{-6}) is located on chromosome 11p in an intronic region of the CARS (cysteinyl-tRNA synthetase) gene. If the etiology of diabetic nephropathy involves the interaction of a locus with the cumulative effect of hyperglycemia, the association of the locus with diabetic nephropathy can vary according to diabetes duration at diabetic nephropathy onset, such that it is strongest in early-appearing case subjects and diminishes in later ones, even reversing in direction in very late-appearing case subjects (23). We examined the SNPs in Table 2 according to diabetes duration by stratifying case and control subjects across tertiles of diabetes duration (at the onset of ESRD or at enrollment into GoKinD for proteinuria patients and control subjects). The strength of the associations was consistent across these strata (data not shown).

**FIG. 1.** Summary of genome-wide association scan results in the GoKinD collection. The –log_{10} P values calculated using the Cochran-Mantel-Haenszel method (adjusting for sex and GoKinD subcollection [JDC/GWU]) across the entire genome are shown for the combined GoKinD collection. The horizontal dashed line corresponds to a –log_{10} P value = 5.0 (P = 1 × 10^{-5}). SNPs shown in green (n = 11) exceed this threshold (because of the resolution of this image, some of the SNPs located on chromosome 13 [n = 7] appear indistinguishable).
<table>
<thead>
<tr>
<th>Locus</th>
<th>Risk allele (non-risk allele)</th>
<th>Control subjects</th>
<th>Case subjects</th>
<th>P</th>
<th>Control subjects</th>
<th>Case subjects</th>
<th>P</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs39059</td>
<td>A(G)</td>
<td>0.61</td>
<td>0.69</td>
<td>8.8 \times 10^{-4}</td>
<td>0.60</td>
<td>0.67</td>
<td>1.7 \times 10^{-3}</td>
<td>5.0 \times 10^{-6}</td>
</tr>
<tr>
<td>rs39075†</td>
<td>G(A)</td>
<td>0.57</td>
<td>0.66</td>
<td>2.0 \times 10^{-4}</td>
<td>0.57</td>
<td>0.64</td>
<td>8.2 \times 10^{-4}</td>
<td>6.5 \times 10^{-7}</td>
</tr>
<tr>
<td>rs1888747†</td>
<td>G(C)</td>
<td>0.68</td>
<td>0.73</td>
<td>3.6 \times 10^{-3}</td>
<td>0.66</td>
<td>0.74</td>
<td>4.4 \times 10^{-5}</td>
<td>6.3 \times 10^{-7}</td>
</tr>
<tr>
<td>rs10868025</td>
<td>A(G)</td>
<td>0.60</td>
<td>0.67</td>
<td>1.9 \times 10^{-3}</td>
<td>0.56</td>
<td>0.66</td>
<td>7.2 \times 10^{-5}</td>
<td>5.0 \times 10^{-7}</td>
</tr>
<tr>
<td>rs739401</td>
<td>C(T)</td>
<td>0.49</td>
<td>0.55</td>
<td>4.7 \times 10^{-4}</td>
<td>0.49</td>
<td>0.55</td>
<td>3.6 \times 10^{-3}</td>
<td>6.4 \times 10^{-6}</td>
</tr>
<tr>
<td>rs451041</td>
<td>A(G)</td>
<td>0.46</td>
<td>0.54</td>
<td>6.9 \times 10^{-4}</td>
<td>0.48</td>
<td>0.56</td>
<td>1.3 \times 10^{-3}</td>
<td>3.1 \times 10^{-6}</td>
</tr>
<tr>
<td>rs1041466</td>
<td>A(G)</td>
<td>0.39</td>
<td>0.47</td>
<td>3.6 \times 10^{-3}</td>
<td>0.43</td>
<td>0.51</td>
<td>2.7 \times 10^{-4}</td>
<td>3.2 \times 10^{-6}</td>
</tr>
<tr>
<td>rs1411766/rs17412858‡</td>
<td>A(G)/G(A)</td>
<td>0.31</td>
<td>0.39</td>
<td>8.5 \times 10^{-4}</td>
<td>0.32</td>
<td>0.40</td>
<td>6.4 \times 10^{-4}</td>
<td>1.8 \times 10^{-6}</td>
</tr>
<tr>
<td>rs6492208/rs2391777§</td>
<td>T(C)/G(A)</td>
<td>0.55</td>
<td>0.62</td>
<td>8.7 \times 10^{-3}</td>
<td>0.56</td>
<td>0.65</td>
<td>1.9 \times 10^{-4}</td>
<td>6.1 \times 10^{-6}</td>
</tr>
<tr>
<td>rs7989848</td>
<td>A(G)</td>
<td>0.49</td>
<td>0.56</td>
<td>2.0 \times 10^{-3}</td>
<td>0.50</td>
<td>0.57</td>
<td>1.1 \times 10^{-3}</td>
<td>7.0 \times 10^{-6}</td>
</tr>
<tr>
<td>rs9521445</td>
<td>A(C)</td>
<td>0.47</td>
<td>0.54</td>
<td>2.1 \times 10^{-3}</td>
<td>0.47</td>
<td>0.55</td>
<td>4.2 \times 10^{-4}</td>
<td>2.9 \times 10^{-6}</td>
</tr>
</tbody>
</table>

The most strongly associated SNPs from the combined analysis of the GWU and JDC GoKinD panels are presented along with the risk allele frequencies and P values (calculated using the Cochran-Mantel-Haenszel method, adjusting for sex, between case and control subjects within each collection) for each separate collection. Combined P values and ORs were calculated using the Cochran-Mantel-Haenszel method. Chromosomal locations, SNP positions, and gene annotations are in reference to NCBI Build 36.1. A summary of the genotype frequencies for the most strongly associated SNPs in the GoKinD collection are presented in supplementary Table 3. †rs39075 and rs1888747 were identified through imputation and genotyped using Taqman assays in the GoKinD collection; ‡rs1411766 and rs17412858 were both genotyped on the Affymetrix array and are in complete linkage disequilibrium (r² = 1.0); §rs6492208 and rs2391777 were both genotyped on the Affymetrix array and are in complete linkage disequilibrium (r² = 1.0).
Additionally, if a locus influences mortality risk, the high mortality experienced by patients with ESRD would alter its association with diabetic nephropathy according to the duration of survival with ESRD and may mask the effect of a diabetic nephropathy risk allele or produce a false association. For this reason, we also analyzed the lead SNPs in Table 2 according to duration of ESRD. For each of these SNPs, the ORs were consistent across tertiles of ESRD duration (supplementary Table 4), a pattern consistent with the absence of survival bias. However, the current study is underpowered to formally exclude the presence of such effects.

**Confirmation of associated type 1 diabetic nephropathy SNPs in the DCCT/EDIC study.** Data from a genome-wide association scan of the DCCT/EDIC study were used to assess whether genome regions identified in the GoKinD collection were associated with advanced diabetic nephropathy in an independent collection. Among the 11 SNPs identified in GoKinD, eight were included on the Illumina array used in the DCCT/EDIC study (Table 3). The three SNPs not included on this platform, rs39059, rs739401, and rs9521445, were in strong linkage disequilibrium (\(r^2 \geq 0.87\)) with rs39075, rs451041, and rs7989848, respectively. Analysis of time to onset of severe nephropathy confirmed the significant associations with diabetic nephropathy in GoKinD for rs1888746 (FRMD3, \(P = 0.02\)), rs13289150 (FRMD3, \(P = 0.05\)), and rs451041 (CARS, \(P = 0.01\)).

**Analysis of candidate diabetic nephropathy gene expression.** Previous studies, as well as publicly available gene expression data (www.ncbi.nlm.nih.gov/geo), have shown that genes closest to the lead SNPs identified in GoKinD are expressed in a variety of human tissues, including kidney (24–26). To further test whether these candidate genes may be involved in the development of diabetic nephropathy, we examined their expression in cell lines relevant to this disease. The expression of CHN2, CPVL, FRMD3, and CARS was examined in four primary cell lines relevant to this disease. The expression of CHN2, CPVL, FRMD3, and CARS was examined in four primary cell lines relevant to this disease.

**FIG. 2.** Summary of genome-wide association results for the chromosome 7p, 9q, 11p, and 13q loci. A: Genome-wide association scan and imputed data for the chromosome 7p locus. rs39059 (solid red triangle) is located at position IVS1 + 21330 relative to exon 1 of CHN2 isoform 2 and in tight linkage disequilibrium with rs39075 (\(r^2 = 0.96\)), located at position IVS1 + 42572. rs39059 and rs39075 reside –69,318 and –90,540 kb, respectively, upstream of CPVL isoforms 1 and 2. A third alternate transcript (isoform 3) is predicted for CPVL and contains an exon that extends to intron 1 of CHN2. rs39059 and rs39075 are located at positions –20579 and –41801, respectively, relative to this transcript. A, SNPs genotyped on the Affymetrix array (n = 163); ●, imputed SNPs (n = 694). *Imputed SNP rs39075 was genotyped in the GoKinD samples to confirm the imputation. B: Genome-wide association scan and imputed data for the chromosome 9q locus. A total of 100 genotyped SNPs from the Affymetrix array data and 450 imputed SNPs are shown. rs10868025 (solid red triangle) is located at position –10829 relative to FRMD3’s transcription start site. rs10868025 is in complete linkage disequilibrium (\(r^2 = 1.0\)) and only 253 bp from imputed SNP rs13289150 (●, superimposed on rs10868025). rs1888747, located at position –2204, is in partial linkage disequilibrium (\(r^2 = 0.81\)) with rs10868025. *Imputed SNP rs1888747 was genotyped in the GoKinD samples to confirm the imputation. C: Genome-wide association scan and imputed data for the chromosome 11p locus. A total of 33 genotyped SNPs from the Affymetrix array data and 190 imputed SNPs are shown. rs739401 and rs451041 (solid red triangles) are in strong linkage disequilibrium (\(r^2 = 0.97\)). rs739401 is located in intron 16 (isoforms a and c)/17 (isoforms b and d) of the CARS gene (position IVS16 + 6977/IVS17 + 687). rs451041 is located in intron 4 (isoforms a and c)/5 (isoforms b and d), position IVS4 – 203/IVS5 – 203). D: Genome-wide association scan and imputed data for the chromosome 13q locus. A total of 68 genotyped SNPs from the Affymetrix array data and 268 imputed SNPs are shown. Seven lead SNPs (rs1041466, rs1117166, rs17412585, rs6492208, rs2391777, rs7989848, and rs9521445) from this region are indicated in red. rs1117166 and rs17412585 are in complete linkage disequilibrium (\(r^2 = 1.0\)). Similarly, rs6492208 and rs2391777 are in complete linkage disequilibrium (\(r^2 = 1.0\)). rs7989848 and rs9521445 are in strong linkage disequilibrium (\(r^2 = 0.87\)), whereas only modest linkage disequilibrium exists between all other SNP pairs (\(r^2 = 0.30–0.65\)). The two nearest genes are MYO1E and IRS2, located –384 kb centromeric and 120 kb telomeric of this region, respectively.
human cell lines: iliac artery endothelial cells, adult dermal fibroblasts, mesangial cells, and renal proximal tubule cells. Our data show that CARS expression was high in all four of the cell lines that we examined (Table 4). FRMD3 expression was also detected in each cell type, with its highest expression being observed in renal proximal tubule cells. Of the two candidate diabetic nephropathy genes located in chromosome 7p region, neither was detected in mesangial cells, whereas CPVL expression was greatest in proximal tubule cells.

**DISCUSSION**

In this report, we describe the results of a genome-wide association scan in the GoKinD collection to identify loci associated with risk of diabetic nephropathy in type 1 diabetes. The most significant associations were identified with variants located within four distinct chromosomal regions. Although the biology underlying these associations remains to be elucidated, they implicate CHN2/CPVL, FRMD3, CARS, and an intergenic region on chromosome 13q as novel genes/genetic regions involved in the pathogenesis of diabetic nephropathy. None of these loci overlap with previously reported associations between candidate genes and the development of any stage of diabetic nephropathy (10,11). Importantly, replication in a Cox proportional hazard analysis of the associations at the FRMD3 and CARS loci with time to the onset of severe nephropathy in the DCCT/EDIC study bolsters the significance of these two findings; that two studies having such different designs (one a case-control study and the other a prospective cohort study) yielded similar ORs strengthens confidence in this conclusion.

FRMD3 encodes the 4.1O protein, a structural protein with unknown function and a member of the 4.1 family of proteins (26). Members of the 4.1 protein family have well-characterized roles as cytoskeletal proteins, maintaining both cellular shape and form, in a variety of cell types, including mouse nephron (27,28). Although membership of the 4.10 protein in this family has recently been questioned, it does contain a FERM domain, which is a module that is integral in maintaining cell integrity through its interactions with transmembrane proteins and actin filaments (29,30). FRMD3 is detectable in adult ovaries as well as in fetal skeletal muscle, brain, and thymus (26). Our data extend the expression profile of FRMD3 to specifically include mesangial and proximal tubular cells. Interestingly, among 18 genes that contain FERM domains, including several members of the 4.1 protein family, we identified nominally significant associations with diabetic nephropathy for SNPs located in eight of these genes (supplementary Table 5), including FARF2 (FERM, Rho-GEF and pleckstrin domain protein 2; \( P = 3.0 \times 10^{-3} \)) and EPB41L2 (erythrocyte membrane protein band 4.1-like 2; \( P = 2.3 \times 10^{-3} \)). Although these findings require further study, including replication in additional cohorts, it is interesting to speculate that these data may point to the involvement of new, previously unsuspected pathways in the pathogenesis of diabetic nephropathy. The CARS gene encodes cysteinyl-tRNA synthetase, one of several aminoacyl-tRNA synthetases (ARSs) that have been identified in humans (31,32). ARSs are important regulators of intracellular amino acid concentrations and protein biosynthesis in both the cytoplasm and mitochondria.

**TABLE 3**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position (Mb)</th>
<th>Nearest gene(s)</th>
<th>Risk allele*</th>
<th>Frequency of risk allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs39075</td>
<td>7p</td>
<td>29.2</td>
<td>CPVL/CHN2</td>
<td>G</td>
<td>0.60</td>
</tr>
<tr>
<td>rs1888746‡</td>
<td>9q</td>
<td>85.3</td>
<td>FRMD3</td>
<td>C</td>
<td>0.70</td>
</tr>
<tr>
<td>rs13289150§</td>
<td>9q</td>
<td>85.4</td>
<td>FRMD3</td>
<td>A</td>
<td>0.62</td>
</tr>
<tr>
<td>rs451041</td>
<td>11p</td>
<td>3.0</td>
<td>CARS</td>
<td>A</td>
<td>0.51</td>
</tr>
<tr>
<td>rs1041466</td>
<td>13q</td>
<td>109.0</td>
<td>No gene</td>
<td>G</td>
<td>0.47</td>
</tr>
<tr>
<td>rs1411766</td>
<td>13q</td>
<td>109.1</td>
<td>No gene</td>
<td>T</td>
<td>0.61</td>
</tr>
<tr>
<td>rs6492208</td>
<td>13q</td>
<td>109.1</td>
<td>No gene</td>
<td>A</td>
<td>0.53</td>
</tr>
<tr>
<td>rs7989848</td>
<td>13q</td>
<td>109.1</td>
<td>No gene</td>
<td>A</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Data are from multivariate Cox proportional hazard analysis of time to onset of severe nephropathy. As of 2005, the number of severe nephropathy cases was 132 (vs. 1,172 censored). Chromosomal locations, SNP positions, and gene annotations are in reference to NCBI Build 36.1. *The risk alleles that are presented are in reference to those identified in the GoKinD collection; †one-sided \( P \) values (consistent with the current “best practices” for replication in GWA scans) (20–22) are used to test for the same direction of effect as in the GoKinD collection; ‡rs1888746 was genotyped on an Illumina array in DCCT/EDIC and is in complete linkage disequilibrium (\( r^2 = 1.0 \)) with rs1888747 (genotyped using a Taqman assay in GoKinD); §rs13289150 was genotyped on an Illumina array in DCCT/EDIC and is in complete linkage disequilibrium (\( r^2 = 1.0 \)) with rs10688025 (genotyped on an Affymetrix array in GoKinD).

### TABLE 4

Relative gene expression of novel candidate diabetic nephropathy susceptibility genes in primary human cell lines

<table>
<thead>
<tr>
<th>Cell line</th>
<th>CPVL ( C_t ) values</th>
<th>CHN2 ( C_t ) values</th>
<th>FRMD3 ( C_t ) values</th>
<th>CARS ( C_t ) values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endothelial cells, iliac</td>
<td>17.3 ± 0.7</td>
<td>12.5 ± 0.1</td>
<td>12.5 ± 0.3</td>
<td>8.4 ± 0.4</td>
</tr>
<tr>
<td>Fibroblasts</td>
<td>15.7 ± 0.1</td>
<td>n.e.</td>
<td>11.9 ± 0.7</td>
<td>6.5 ± 0.4</td>
</tr>
<tr>
<td>Mesangial cells</td>
<td>n.e.</td>
<td>n.e.</td>
<td>12.9 ± 0.5</td>
<td>7.2 ± 0.1</td>
</tr>
<tr>
<td>Proximal tubules</td>
<td>8.6 ± 0.1</td>
<td>13.1 ± 0.1</td>
<td>10.2 ± 0.2</td>
<td>8.0 ± 0.1</td>
</tr>
</tbody>
</table>

The relative gene expression of CPVL, CHN2, FRMD3, and CARS was determined in four primary human cell lines using real-time PCR. Relative gene expression of each gene was calculated in reference to a normalization control (\( \beta\)-actin) and is presented as the mean \( \Delta C_t \) (\( C_t \) value from each gene minus \( C_t \) value from the normalization control) and SD from three separate experiments. Mean \( \Delta C_t = 0 \) equals high relative gene expression (i.e., expression similar to \( \beta\)-actin), n.e., not expressed (mean \( \Delta C_t > 40.0 \)).
dria (a process facilitated by specialized mitochondria-specific and bifunctional ARSs). In the initial steps of protein translation, the function of these enzymes is to attach amino acids to their cognate tRNA molecules. To date, both autosomal dominant and recessive mutations in ARS-encoding genes have been identified only in neurodegenerative disease, including missense changes in glycylic tRNA synthetase (GARS) and both missense mutations and in-frame deletions in tyrosyl-tRNA synthetase (YARS) in Charcot-Marie-Tooth disease (32).

CARS has been implicated in cystinosis, an autosomal recessive renal tubule disorder caused by the accumulation of free cystine in cellular lysosomes (33,34). A recent study identified defects in lysosomal cystine transport as the primary cause of the disease (35). However, ESRD is prominent in this disorder, and such an outcome may be due to vulnerability of specific renal cells to damage by excess cystine. Interesting, in this light, is the observation that of all the associated SNPs, only those in the CARS locus were associated primarily with ESRD (supplementary Table 4). CARS is expressed in mesangial and proximal tubule cells. Further work is needed to characterize the role of CARS in the pathway that is involved in the development of ESRD in diabetes. Similar to the set of genes containing FERM domains, analysis of 21 ARS genes identified nominally significant associations with diabetic nephropathy for SNPs located in four members of this class of genes (supplementary Table 6), with the most significant association \( P = 9.1 \times 10^{-5} \) occurring at the TARS (threonyl-tRNA synthetase) locus.

Two additional loci were strongly associated with diabetic nephropathy in both panels of the GoKinD collection. Of the two genes located on chromosome 7p, CPVL, a carboxypeptidase that is highly expressed in the kidney and, more specifically, in proximal tubules, is a particularly interesting candidate gene. Other carboxypeptidases, such as ACE and bradykinin, are important regulators of renal hemodynamics and have previously been implicated in the pathogenesis of diabetic nephropathy (36,37). The last diabetic nephropathy–associated locus involves multiple SNPs within a 33 kb haplotype block on chromosome 13q. Previously, genomic deletions of this locus have been linked to congenital renal abnormalities (38). The two genes closest to the associated SNPs, MYO16 (myosin heavy-chain Myr 8) and IRS2 (insulin receptor substrate 2), are located –384 kb centromeric and 120 kb telomeric of this region, respectively. Although there is little linkage disequilibrium between the variants within this block and those in the vicinity of either MYO16 or IRS2, the multiple signals identified in this region give credence to the association detected in our analysis. Additional experiments are needed to characterize the nature of these associations further.

The findings presented in our study contribute to understanding the genetic susceptibility of diabetic nephropathy in type 1 diabetes. As has been reported for other complex genetic disorders, no single major gene that contributes to an increased risk of disease emerged (20,39). However, given the incomplete coverage of the genome by the genotyping platform and the suboptimal study design (prevalent rather than incident cases of ESRD), detection of any existing major gene effect was not guaranteed. For example, because most of the case subjects with ESRD had survived many years on dialysis or with a kidney transplant, a disease allele that not only increased susceptibility to diabetic nephropathy but also increased mortality in patients with ESRD could go undetected. Appreciably, the SNPs that we identified in the GoKinD collection were mortality neutral (supplementary Table 4). The optimal study design for detecting all disease loci, regardless of their effect on mortality, would be a large cohort of incident ESRD case subjects. Such a data set is currently unavailable.

There are other limitations to this study as well. The GoKinD collection is heavily weighted with case subjects with ESRD; thus, the small number of case subjects with proteinuria limited our ability to detect variants primarily associated with the risk of proteinuria. Second, because of the limited power of the DCCT/EDIC study and the need to contain inflation of the \( \alpha \)-error in seeking replication for multiple SNPs in this dataset, our replication efforts refrained from considering SNPs less significant than \( P = 1 \times 10^{-5} \). It is certainly possible that additional variants among those not meeting this threshold may truly be associated with diabetic nephropathy; however, given these limitations, these variants remain to be identified. Similarly, despite replication in the DCCT/EDIC cohort, we acknowledge that positive associations at both the FRMD3 and CARS loci require additional study to be certain of these findings. Third, although the locations of the variants confirmed in this study implicate both FRMD3 and CARS as novel genes involved in the pathogenesis of diabetic nephropathy, the underlying mechanisms of disease of these associations need to be elucidated. And, finally, although confirmation in DCCT/EDIC has been achieved for variations near FRMD3 and CARS, additional cohorts, particularly non-Caucasian, would be useful to further characterize the pathogenic role of these, and other, candidate genes identified in the GoKinD collection.

ACKNOWLEDGMENTS

This work was supported by the following grants from the National Institutes of Health (NIH): DK77592 (to A.S.K.), DK68386 (to the Genetics Core of the Diabetes and Endocrinology Research Center at the JDC), DK-62204 (to A.D.P.), and DK-077510 (to A.D.P.) and from the Foundation for NIH (FNIH): 06GAIN0 (to J.H.W.). We acknowledge JDC NIH Training Grant T32 DK007260-31 (to M.G.P.) and a CIHR (Canadian Institutes of Health Research) fellowship (to L.M.) and senior investigator award (to S.B.B.). We also acknowledge support from the Canadian Network of Centres of Excellence in Mathematics (to S.B.B.). The GoKinD study was conducted by the GoKinD Investigators and supported by the Juvenile Diabetes Research Foundation (JDRF) and by funding from the Centers for Disease Control (CDC; PL 105-33, 106-554, and 107-360 administered by the National Institute of Diabetes and Digestive and Kidney Diseases [NIDDK]). The GoKinD collection of DNA was genotyped through the GAIN program, with the support of the FNIH and NIDDK. The Gain database was accessed through the NCBI (National Center for Biotechnology Information). The DCCT/EDIC study is supported by contracts with the Division of Diabetes, Endocrinology, and Metabolic Diseases of the NIDDK and the General Clinical Research Centers Program, National Center for Research Resources. The content and conclusions presented in this manuscript do not necessarily reflect the opinions or views of the JDRF, CDC, FNIH, NIDDK, or NCBI.

No potential conflicts of interest relevant to this article were reported.
REFERENCES


