

Microarray analysis of multiple candidate genes and associated plasma proteins for nephropathy secondary to type 2 diabetes among Chinese individuals

S. C. Lim · J. J. Liu · H. Q. Low · N. G. Morgenthaler · Y. Li · L. Y. Yeoh · Y. S. Wu ·
S. K. Goh · C. Y. Chionh · S. H. Tan · Y. C. Kon · P. C. Soon · Y. M. Bee ·
T. Subramaniam · C. F. Sum · K. S. Chia

Received: 11 January 2009 / Accepted: 25 March 2009 / Published online: 5 May 2009
© Springer-Verlag 2009

Abstract

Aims/hypothesis Evolving research suggests that common and rare alleles jointly constitute the genetic landscape of complex disease. We studied the association between 43 pathway-related candidate genes with ‘intermediate pheno-

Electronic supplementary material The online version of this article (doi:10.1007/s00125-009-1368-x) contains supplementary material, which is available to authorised users.

S. C. Lim (✉) · L. Y. Yeoh · Y. S. Wu · S. K. Goh ·
T. Subramaniam · C. F. Sum
Department of Medicine, Alexandra Hospital,
378 Alexandra Road,
Singapore 159964, Republic of Singapore
e-mail: Su_chi_lim@alexhosp.com.sg

J. J. Liu · H. Q. Low · Y. Li
Human Genetics, Genome Institute of Singapore,
Singapore, Republic of Singapore

N. G. Morgenthaler
Research Department, B.R.A.H.M.S. Biotechnology Center,
Hennigsdorf/Berlin, Germany

C. Y. Chionh · S. H. Tan
Department of General Medicine, Tan Tock Seng Hospital,
Singapore, Republic of Singapore

Y. C. Kon · P. C. Soon
Department of Endocrinology, Tan Tock Seng Hospital,
Singapore, Republic of Singapore

Y. M. Bee
Department of Endocrinology, Singapore General Hospital,
Singapore, Republic of Singapore

K. S. Chia
Centre for Molecular Epidemiology,
National University of Singapore,
Singapore, Republic of Singapore

type’ (i.e. corresponding plasma protein) and diabetic nephropathy in a customised microarray of 1,536 SNPs.

Methods In this case–control study of type 2 diabetic Chinese individuals with and without diabetic nephropathy, cases ($n=545$) were defined on the basis of a spot urinary albumin/creatinine ratio (ACR) >113 mg/mmol; the value for controls ($n=503$) was ACR <3.3 mg/mmol. Genotyping was performed using Illumina GoldenGate assay.

Results No single nucleotide polymorphism (SNP) remained significant in single locus analysis after correction for multiple testing. Therefore, we explored the best ~1% SNPs. Of these 13 SNPs, four clustered to a 5′ end NADPH oxidase homologue 4 (*NOX4*) haplotype (GGCC frequency=0.776) with estimated OR for diabetic nephropathy of 2.05 (95% CI 1.04–4.06) (heterozygous) and 2.48 (1.27–4.83) (homozygous) ($p=0.0055$). The haplotype was correlated with plasma Cu/Zn superoxide dismutase (SOD) concentration, suggesting increased oxidative burden. Endothelin-1 SNP (rs1476046G>A, frequency=0.252) was correlated with plasma C-terminal pro-endothelin-1 concentrations with an estimated OR for diabetic nephropathy of (heterozygous) 1.26 (0.96–1.66) and (homozygous) 1.87 (1.13–3.12) ($p=0.0072$). Nitric oxide synthase 1 (*NOS1*) 5′ haplotype (TGTC frequency=0.38) also revealed a suggestive association with diabetic nephropathy: heterozygous 1.26 (0.95–1.67), homozygous 1.57 (1.04–2.35) ($p=0.0073$). A rare NADPH oxidase homologue 1 (*NOX1*)-coding non-synonymous SNP (Arg315His, frequency=0.006) was found exclusively among cases.

Conclusions/interpretation Our preliminary observations suggest that common haplotypes from *NOX4* and endothelin-1 SNP correlated with plasma Cu/Zn SOD and C-terminal pro-endothelin-1 concentrations, respectively, and might have conferred diabetic nephropathy susceptibility. Common *NOS1* and

rare *NOX1* variants also revealed a suggestive association with diabetic nephropathy. Future studies to validate our observation are needed.

Keywords Candidate genes · Diabetic nephropathy · Microarray

Abbreviations

ACR	Albumin/creatinine ratio
LD	Linkage disequilibrium
MAF	Minor allele frequency
NOS	Nitric oxide synthase
NOX	NADPH oxidase homologue
Q–Q	Quantile–quantile
SNP	Single nucleotide polymorphism
SOD	Superoxide dismutase

Introduction

Diabetic nephropathy is a major vascular complication of long-standing type 2 diabetes. While metabolic (e.g. glycaemic burden) and haemodynamic (e.g. blood pressure) factors are important in the pathogenesis of diabetic nephropathy [1], accumulating data (estimates of heritability, distinct familiar clustering and disparate ethnic susceptibility) suggest that genetic determinants are also important [2]. However, the exact identity of susceptibility genes in diabetic nephropathy have remained elusive.

Numerous diabetic nephropathy candidate genes have been investigated in association studies and a few genes (e.g. angiotensin-converting enzyme) have shown promise [3]. However, like most complex diseases, replication of these positive observations has been difficult [4]. In type 1 diabetes, a panel of 115 candidate genes has recently been studied using a family-based approach (transmission disequilibrium test) [5]. The investigators noted that the results were non-conclusive due to its relative small sample size (total 72 families) and hence limited statistical power. In type 2 diabetes, preliminary ‘hypothesis-free’, genome-wide association studies using relatively low resolution first generation microarray (81,315 single nucleotide polymorphism [SNP] loci) have been attempted. These investigators adopted a two-stage approach, i.e. full genotyping in the first 94 case–control ‘training pairs’, with follow-on genotyping of promising SNPs in a larger validation cohort [6]. A separate team of investigators analysed DNA pooled from some 105 cases (thereby precluding genotype-based analysis at genome scale) [7]. In these studies, *ELMO1* and *PVT1* genes were novel candidates conferring susceptibility to diabetic nephropathy, respectively. Very recently, association of *PVT1* with diabetic nephropathy was also reported

in patients with type 1 diabetes [8], making it a promising candidate gene awaiting further replication. Notably, these studies were performed in Japanese individuals or American Indians, whose distribution of genetic variation may differ (especially the latter) from Chinese individuals [9].

In the field of diabetic nephropathy, very few studies have investigated the relationship between genotype and ‘intermediate phenotype’ (i.e. corresponding plasma protein concentrations). Given the less complex search space, a relationship between genotype and intermediate phenotype may be more readily demonstrable and may support a likely association between genotype and final phenotype (i.e. diabetic nephropathy). In addition, recent insights suggest that the genetic landscapes of complex diseases are likely to be a composite map of common variants with small effect size and rare variants with large effect size [10]. However, commercial arrays have preferentially incorporated common alleles, i.e. low-frequency alleles are often under-represented [11]. Given the current technical limitation, studying population-specific low-frequency founder alleles is only feasible using a candidate gene approach. Therefore, in a case–control study of diabetic patients with and without diabetic nephropathy, we selected SNPs with evenly distributed allele frequency (i.e. common and rare) from 43 pathway-related, putative candidate genes and studied their association with intermediate phenotype (i.e. corresponding plasma protein concentrations) and diabetic nephropathy in a customised microarray of 1,536 SNPs among 1,048 Chinese patients. To the best of our knowledge, such a ‘candidate gene-wide association’ approach has not been adopted before for diabetic nephropathy among Chinese.

Methods

Participants We recruited 1048 participants with type 2 diabetes from three secondary care hospitals and two primary care clinics for a case–control study. Diabetes was diagnosed according to the American Diabetes Association criteria [12]. Most of the cases, i.e. patients with diabetic nephropathy were recruited from the secondary care hospitals, whereas most of the controls, i.e. diabetic patients without nephropathy came from the primary care clinics. Participants attending the primary care clinics and who developed diabetic nephropathy would be referred to the secondary care hospital for further management. Thus, controls were chosen to represent the exposure experience of the source population giving rise to the cases. The phenotype of these participants has been described in detail elsewhere [13]. Briefly, cases were participants with overt diabetic nephropathy ($n=545$) as defined by the presence of proteinuria ≥ 1.0 g/day (equivalent to spot urinary albumin/creatinine ratio [ACR] ≥ 113 mg/mmol or persistently

elevated serum creatinine). The controls consisted of unrelated individuals ($n=503$) with ACR ≤ 3.3 mg/mmol and consistently normal serum creatinine. To minimise confounding by population admixture, only Chinese patients were enrolled. Measurement of standard laboratory biochemistry was performed as described in Electronic supplementary material (ESM).

Ethics approval This study followed the recommendations of the Declaration of Helsinki and was approved by the Domain Specific Review Board, National Health Group, Republic of Singapore. Written, informed consent was obtained from all participants.

Procedure Selection of candidate genes (for details, see ESM and ESM Table 1) was based on current understanding of the pathophysiology of diabetic nephropathy. At the planning phase of the study, some of the widely used public domain databases (e.g. HapMap) on genetic variation of these candidate genes were at an early stage of release. Nevertheless, variants regarded as likely haplotype tag SNPs for Han Chinese were preferentially selected for optimal coverage. All haplotype tag SNPs chosen would have fulfilled the minimal pairwise-tagging threshold of $r^2 \geq 0.8$. Approximately 10 kb 5' upstream of the putative transcription start site (to capture the gene regulatory region) and 5 kb 3' downstream of the last exon were included in the gene region considered for SNP selection. To optimise our SNP map, we also applied other strict inclusion and exclusion criteria to the SNP selection process (see ESM).

All SNPs were genotyped using customised (GoldenGate; Illumina, San Diego, CA, USA) microarray assay. Subsequently, to confirm the allele assignment for an interesting but low-frequency NADPH oxidase homologue 1 (*NOX1*) SNP (rs2071756G>A, frequency 0.006), all samples were re-genotyped using another platform, namely TaqMan assay (Applied Biosystems, Carlsbad, CA, USA). The allele call was completely concordant. Polymorphic SNPs that met the quality control criteria of the assay were included for statistical analysis.

To follow up on potentially interesting SNPs (i.e. the best ~1% according to p values on single locus analysis), we investigated genotype–intermediate phenotype relationship by correlating endothelin-1 SNP rs1476046G>A with plasma C-terminal pro-endothelin-1 concentrations (i.e. akin to the measurement of NT-proBNP instead of mature BNP in the diagnosis of heart failure [14, 15]) among the 868 individuals (448 cases, 420 controls) whose plasma samples were available and deemed suitable for laboratory measurement. Similarly, the relationship between nitric oxide synthase 1 (*NOS1*) (neuronal) and *NOX4* haplotypes was also correlated with plasma Cu/Zn superoxide dis-

mutase (SOD). This was done because *NOS1* and *NOX4* are associated with oxidative stress and Cu/Zn SOD is a powerful enzymatic scavenger of superoxide [16]. Hence, elevated plasma Cu/Zn SOD (in response to oxidative stress) could serve as a surrogate maker for oxidative burden.

Cases by definition had renal impairment, which was known to affect endothelin and Cu/Zn SOD function. Indeed, we observed vast differences in plasma C-terminal pro-endothelin-1 and Cu/Zn SOD concentrations between cases and controls (see Results). Therefore, we stratified our analysis of genotype–intermediate phenotype by diabetic nephropathy status, i.e. cases or controls, since pooling them together may not be biologically justifiable.

Statistical analysis Data are expressed as mean \pm SD. All statistical analyses were calculated using STATA version 9.0 (StataCorp, College Station, TX, USA), Haploview version 4.0 (<http://hydra.usc.edu/gxe>, accessed 1 April 2009) and PHASE version 2.1 (<http://stephenslab.uchicago.edu/software.html>, accessed 1 April 2009). Continuous variables were tested for normality of distribution. Student's t test and ANOVA (or Kruskal–Wallis test for non-normal distribution) were used to compare continuous variables between two or more groups respectively.

We used χ^2 analysis to test for deviation of genotype distribution from Hardy–Weinberg equilibrium ($p < 0.05$) and to determine whether frequency distribution of allele or genotype differed between cases and controls. The Cochran–Armitage trend test (3×2 tables formed by cross-classifying participants by genotype and disease status) was also used to test for association between genotype and diabetic nephropathy. For genotypes with low minor allele frequency (MAF) (homozygous minor genotype count < 5 in either cases or controls), the Monte Carlo permutation test was performed. Bonferroni method was used for statistical correction of multiple hypothesis testing. False discovery rate analysis was also performed to identify SNPs that may reveal potentially important associations with diabetic nephropathy. Quantile–quantile (Q–Q) plots were produced by plotting the ranked values of the test statistic against the expected order statistic, $F^{-1}[i/(N+1)]$, under the global null hypothesis that no true association exists [17]. The Q–Q plot also reveals cryptic relatedness or population stratification, if present. The power of the study was estimated using QUANTO freeware, version 5.1 (<http://hydra.usc.edu/gxe>, accessed 1 April 2009).

The computer program PHASE was used to construct the haplotype from genotype. Phase-inferred haplotypes were tested for association with diabetic nephropathy using a trend test. Linkage disequilibrium (LD) between the two SNPs was also estimated using D' and r^2 . A value of $r^2 > 0.80$ was considered to indicate a significantly strong LD

between the SNPs. Finally, regularised regression analysis of variable-sized sliding windows haplotypes developed by our co-authors was also performed to further optimise power for detection of susceptibility genes [18]. Briefly, in this novel statistical approach, the maximum size of a sliding window was determined by local haplotype diversity and samples size. Subsequently, the problem of multiple degrees of freedom in the haplotype test was managed by regularised regression analysis. Using both simulated and experimental data, this approach was found to be more efficient and effective than other currently available methods.

Results

Participant characteristics Clinical characteristics of the study groups are summarised in Table 1. Cases and controls were similar in distribution of sex, age, duration of diabetes and HbA_{1c}. Not surprisingly, systolic BP, diastolic BP and BMI were higher among the cases. As expected, more cases than controls suffered from retinopathy (23% non-proliferative, 56% proliferative vs 16% and 14% respectively; $p < 0.001$), since diabetic nephropathy is strongly associated with retinopathy.

SNP selection and association A total of 1,536 SNPs was selected to represent the 43 candidate genes (ESM Table 1). Eventually, only 932 of the 1,048 individuals recruited (cases 487, controls 445) and 914 SNPs (59.5%) were available for final analysis (for details on samples and SNP

elimination based on quality control considerations, see ESM). The MAF distribution of the 914 SNPs included in final analysis is shown in ESM Fig. 1 and suggested that low-frequency and common alleles were well represented in the final analysis.

Using the most conservative Bonferroni method for statistical correction of multiple hypothesis testings, the corrected global, experiment-wise p value would be 5.4×10^{-5} (i.e. $0.05 \div 914$ SNPs tested). In our present study, none of the SNPs or haplotypes revealed such robust statistical evidence. Similarly, false discovery rate analysis did not suggest any particular SNP with an important association with diabetic nephropathy (details not shown). The Q–Q plot of the Cochran–Armitage test statistics for individual SNPs revealed that there was no cryptic relatedness or appreciable population substructure (ESM Fig. 2). On the other hand, the Q–Q plot did not suggest any strong evidence of association with diabetic nephropathy, either. Therefore, our data should be considered as preliminary and at best indicative (but non-conclusive). Having said so, we explored the relationship between diabetic nephropathy and the best ~1% (i.e. 13) SNPs in our study. Variants from four gene regions (*NOX4*, endothelin-1, *NOS1* and *NOX1*) were potentially interesting. Besides coding non-synonymous rs2071756C>T (Arg315His) from *NOX1* (X chromosome), all are intronic SNPs.

NOX4 Of these 13 SNPs, four (rs614128G>C, rs490934G>C, rs3017887C>A and rs553635C>T) spanning ~15 kb in chromosome 11, 11q14.2–q21) clustered to form a 5' end *NOX4* haplotype block. The LD patterns of these four SNPs are shown in ESM Table 2 and ESM Fig. 3. The haplotype GGCC (frequency 0.776) had an estimated OR for diabetic nephropathy of 2.05 (95% CI 1.04–4.06) (heterozygous) and 2.48 (1.27–4.83) (homozygous) ($p = 0.0055$) (Table 2). Consistent with other reports, we observed that plasma Cu/Zn SOD concentrations in cases was much higher than in controls (29.9 ± 45.7 vs 8.9 ± 18.6 nmol/L, $p < 0.001$) [19]. Interestingly, homozygosity of this *NOX4* haplotype (and component SNPs) was also associated with increased plasma Cu/Zn SOD concentration among cases, suggesting increased oxidative burden (Table 3).

Endothelin-1 Suggestive association was also observed for rs1476046G>A (Table 4), with an estimated OR for diabetic nephropathy of (heterozygous) 1.26 (0.96–1.66) and (homozygous minor allele) 1.87 (1.13–3.12) ($p = 0.0072$ for trend). Plasma C-terminal pro-endothelin-1 concentrations among cases was much higher than among controls (84.9 ± 50.3 vs 48.0 ± 16.8 pmol/L, $p < 0.001$). Interestingly, the minor allele of rs1476046G>A correlated with the decrement in plasma C-terminal pro-endothelin-1 concentration among controls ($p = 0.014$ for trend) (Table 5).

Table 1 Clinical characteristics of study participants whose genotype data met the quality control criteria

Characteristic	Study groups		
	Controls	Cases	p value
n	445	487	
Male sex (%)	57	55	0.57
Age (years)	61±10	63±10	0.02
Systolic BP (mmHg)	134±15	146±21	<0.001
Diastolic BP (mmHg)	77±8	80±10	0.001
BMI (kg/m ²)	25.1±3.7	25.8±4.5	0.001
HbA _{1c} (%)	7.8±1.3	8.3±3.0	0.07
Duration (years)	15±6	15±7	0.34
Retinopathy (%)			
Normal	70	21	<0.001
Non-proliferative	16	23	
Proliferative	14	56	

Unless otherwise stated, values are mean±SD

Table 2 Distribution of endothelin-1, *NOS1* and *NOX4* haplotype frequencies among cases and controls

Gene	Frequency	Case	Control	χ^2	<i>p</i> value	OR (one copy vs none)	OR (two copies vs none)
<i>NOX4</i>							
GGCC	0.777	0.803	0.749	7.723	0.0055	2.05 (1.04–4.06)	2.48 (1.27–4.83)
CCAT	0.197						
GGAT	0.017						
Endothelin-1							
AA	0.132	0.156	0.098	17.03	0.00020	1.69 (1.22–2.33)	2.65 (0.94–7.52)
CA	0.150						
AG	0.669						
<i>NOS1</i>							
TGTC	0.379	0.408	0.347	7.197	0.0073	1.26 (0.95–1.67)	1.57 (1.04–2.35)
CGCC	0.314						
CACG	0.283						

OR values given with 95% CI

Linear regression analysis suggested that every copy of the minor allele (A) was associated with an approximately 3.60 ± 1.35 pmol/L reduction in plasma C-terminal pro-endothelin-1 concentrations. This relationship was absent among cases. In addition, variable-sized sliding windows haplotypes analysis revealed that an endothelin-1 5' haplotype AA (frequency: 0.129, formed by rs3087459A>C and rs1476046G>A, spanning ~3.6 kb in chromosome 6) appeared to confer increased susceptibility to diabetic nephropathy (OR 1.71 [1.28–2.27], $p=0.00020$) (Table 2). Pairwise LD between the two endothelin-1 SNPs, however, was modest ($D'=0.66$, $r^2=0.27$) (ESM Fig. 4). The endothelin-1 haplotype also showed borderline significant correlation ($p=0.081$) with plasma C-terminal pro-endothelin-1 concentrations in controls (Table 5).

NOS1 Four other alleles (rs527590C>T, rs693534G>A, rs3782219C>T and rs9658255C>G, spanning ~17 kb in chromosome 12, 12q24.2–q24.31) clustered to form a haplotype over the 5' ends of *NOS1*. The LD patterns of these four SNPs are shown in ESM Table 3 and ESM Fig. 5. The haplotype TGTC (frequency 0.38) appeared to confer increased susceptibility to diabetic nephropathy, with an OR of (heterozygous) 1.26 (95% CI 0.95–1.67) and (homozygous) 1.57 (1.04–2.35) ($p=0.0073$ for trend) (Table 2). However, *NOS1* genotype and haplotype did not show any correlation with plasma Cu/Zn SOD concentration (ESM Table 4).

NOX1 A rare coding non-synonymous SNP from *NOX1* (rs2071756G>A, R315H, MAF 0.006) was found exclusively

Table 3 Plasma SOD concentrations (nmol/l) per *NOX4* SNPs and haplotype among cases and controls

SNP/haplotype per study group	<i>n</i>	SOD per genotype (nmol/l)			
		Homozygous major	Heterozygous	Homozygous minor	<i>p</i> value
rs614128G>C					
Cases	448	32.8±52.2	21.5±22.5	56.7±62.7	0.006
Controls	420	7.6±8.3	11.4±29.0	5.8±5.2	0.110
rs490934G>C					
Cases	448	33.0±52.2	22.0±23.6	56.2±64.5	0.011
Controls	420	7.6±8.3	11.5±29.2	6.1±5.2	0.109
rs3017887C>A					
Cases	448	33.5±53.0	21.5±22.3	48.0±61.4	0.012
Controls	420	7.6±8.3	11.1±28.7	7.4±7.9	0.178
rs553635C>T					
Cases	448	33.3±52.9	21.7±22.4	48.0±61.4	0.015
Controls	420	7.6±8.3	11.2±28.8	7.4±7.9	0.164
Haplotype GGCC					
Cases	448	33.4±53.3	21.5±22.0	53.0±61.5	0.006
Controls	420	7.6±8.4	11.1±28.6	7.1±7.7	0.172

Values are mean±SD
p values are for trend

Table 4 Distribution of genotypes and allele frequencies of endothelin-1 SNP rs1476046G>A among cases and controls

Endothelin-1 variants and allele frequencies	Study groups		χ^2	<i>p</i> value
	Controls	Cases		
rs1476046G>A				
AA	0.06	0.10	7.47	0.024
GA	0.39	0.42		
GG	0.55	0.48		
Frequency				
A frequency	0.252	0.309	7.258	0.0071
G frequency	0.748	0.691		

among cases only (allele frequency 0.016 [$p=0.046$] and 0.007 [$p=0.12$] for men and women respectively) (ESM Table 5). Using the online resource PolyPhen (<http://coot.embl.de/PolyPhen>, accessed 1 April 2009) to bio-informatically predict whether rs2071756G>A may have functional impact on the corresponding protein [20], we noted that rs2071756G>A may be benign (i.e. be well tolerated).

Other SNPs The remaining three SNPs, rs865716A>G from *SCARB1*, rs12720136C>T from *EDNRB* and rs2799103A>G from *TGFB2*, demonstrated a provisionally interesting association with diabetic nephropathy and are shown in ESM Table 5. Using Quanto 5.1, we estimated that our study has ~80% power to detect a relative risk of ≥ 1.6 conferred by any susceptibility allele with a frequency of ≥ 0.25 at $\alpha < 5.4 \times 10^{-5}$.

Discussion

To the best of our knowledge, this is the first systematic survey of SNPs (both common and rare) from 43 highly probable diabetic nephropathy candidate genes along several putative biological pathways in type 2 diabetes Chinese individuals. We observed that common variants

from *NOX4* and endothelin-1 were associated with differential plasma Cu/Zn SOD and C-terminal pro-endothelin-1 concentrations (i.e. intermediate phenotype) respectively. We also found preliminary indications that these variants might potentially be associated with diabetic nephropathy, although the possibility that these were chance observations could not be confidently ruled out. Similarly, the *NOS1* 5' haplotype and *NOX1* (a low-frequency coding non-synonymous SNP) may be potentially interesting diabetic nephropathy candidate genes among Chinese. Our preliminary observations will need to be replicated in other populations.

Growing evidence suggests that NADPH oxidase-derived reactive oxygen species might play an important role in the initiation and progression of diabetic nephropathy [21]. NADPH oxidase is an enzyme complex with the following subcomponents: cytochrome b-245, alpha polypeptide (CYBA, also known as p22^{phox}), cytochrome b-245, beta polypeptide (CYBB, also known as gp91^{phox} [renal homologue is known as NADPH oxidase homologue 4, NOX4]), neutrophil cytosolic factor 1 (NCF1, also known as p47^{phox}), neutrophil cytosolic factor 2 (NCF2, also known as p67^{phox}), neutrophil cytosolic factor 4 (NCF4, also known as p40^{phox}) and GTPase-Rac1 [22]. NOX4 has been described in renal cells such as tubular epithelial cells and glomerular mesangial cells [23]. In the vasculature of streptozotocin-induced *ApoE*^{-/-} mice, hyperglycaemia was associated with enhanced *Nox4* gene expression and increased oxidative stress that was reversed by SOD [24]. Hence, *NOX4* is an attractive candidate gene for diabetic nephropathy. Our group and others have previously reported an inconsistent association between candidate SNPs of *CYBA* and diabetic nephropathy [11]. However, to the best of our knowledge, *NOX4* has not been studied as a potential diabetic nephropathy candidate gene. Hence, the suggestive relationship observed by us between *NOX4* and diabetic nephropathy could be novel. Moreover, we observed that homozygosity of the risk allele/haplotype was associated with increased Cu/Zn SOD concentrations among cases (Table 3). This suggests that the risk allele might be associated with increased oxidative burden and

Table 5 Plasma C terminal pro-endothelin-1 concentrations per endothelin-1 SNP rs1476046G>A genotypes and related endothelin-1 haplotype AA (formed by rs3087459A>C and rs1476046G>A) among study groups

SNP/haplotype per study groups	<i>n</i>	Plasma C terminal pro-endothelin-1 per genotype (pmol/l)			
		Homozygous major	Heterozygous	Homozygous minor	<i>p</i> value
rs1476046G>A					
Cases	448	86.5±47.7	81.3±50.6	92.9±59.5	0.309
Controls	420	48.6±15.9	47.2±18.1	39.5±14.0	0.014
Haplotype AA					
Cases	448	84.8±50.0	83.9±50.5	97.9±56.7	0.655
Controls	420	48.9±17.1	44.4±15.3	40.5±7.3	0.081

Values are mean±SD
p values are for trend

hence a corresponding elevation of plasma Cu/Zn SOD as ‘response to injury’ rescue mechanism. Interestingly, the *NOX4* haplotype associated with diabetic nephropathy in our study stretched over the gene region that captured the 5′ putative regulatory motifs, first and second exons (ESM Fig. 3). Taken together, genetic variation in *NOX4* might be associated with differential oxidative burden, thereby modulating diabetic nephropathy susceptibility.

Endothelin-1, the predominant isoform in cardiovascular and renal systems, is the most powerful endogenous vasoconstrictor with profibrotic and proinflammatory effects [25]. It has been found to affect three pivotal aspects of renal physiology: (1) vascular and mesangial tone; (2) sodium and water excretion; and (3) cell proliferation and matrix formation [26]. In vitro experiments using rat mesangial cells revealed that hyperglycaemia could stimulate endothelin-1 promoter activity and gene expression [27]. A recent study on twins suggested that the heritability of plasma endothelin-1 concentration was approximately 0.58 [28]. Therefore, endothelin-1 is important in the pathogenesis of diabetic nephropathy and a promising candidate gene. However, the relationship between endothelin-1 gene and renal impairment appears to be inconsistent. Thus Freedman et al. [29] reported negative, whereas Kanková et al. [30] and Pinto-Sietsma et al. [31] reported positive association, suggesting the need for further studies [32]. Notably, these studies did not attempt to correlate endothelin-1 gene with plasma endothelin-1 concentrations. In our study, we observed that endothelin-1 minor allele rs1476046G>A was associated with decreased plasma C-terminal pro-endothelin-1 concentrations (intermediate phenotype) among controls (Table 5). This relationship, however, could not be established among cases (possibly due to confounding by multiple medications and other co-morbidities). We speculated that the absence of specific genotype-mediated attenuation in plasma C-terminal pro-endothelin-1 concentrations among cases may contribute to diabetic nephropathy susceptibility (given that endothelin-1 overactivity could be detrimental). Interestingly, the endothelin-1 5′ haplotype (formed by rs3087459A>C and rs1476046G>A [ESM Fig. 4]) that was associated with diabetic nephropathy stretched over the gene region where a putative functional insertion/deletion SNP (rs10478694A/–) in the 5′ untranslated region within the first exon has been described (rs10478694A/– was not genotyped in our study due to technical limitation of assay platform) [33]. rs10478694A/– has been associated with hypertension [34], orthostatic intolerance [35] and differential endothelin-1 gene expression in vitro [36]. However, the LD pattern of rs10478694A/– with surrounding genetic variants is presently unknown. Hence, we were unable to estimate whether rs3087459A>C or rs1476046G>A could be legitimate proxies for rs10478694A/–. Taken together,

rs1476046G>A (and related haplotype) could be functional (i.e. associated with differential plasma C-terminal pro-endothelin-1 concentration) and might confer susceptibility to diabetic nephropathy.

Animal studies suggest that nitric oxide synthase 1 (NOS1) is the dominant isoform in the generation of nitric oxide in diabetic nephropathy [37]. Recent observations in streptozotocin-induced diabetic rats have also revealed ongoing overactivity of NOS1 secondary to loss of normal control by renal macula densa [38]. Overactive NOS1 may contribute to renal hyperfiltration, which has been postulated to result in flow-mediated vascular injury [39]. Hence, *NOS1* is a highly relevant biological candidate gene for diabetic nephropathy. As far as we know, little is known about *NOS1* gene and diabetic nephropathy. A *NOS1* intronic microsatellites marker (NOS1B) has been reported to be associated with end-stage renal disease in African-American families (more pronounced among non-diabetic end-stage renal disease patients) [29]. In our study, the *NOS1* haplotype that revealed suggestive association with diabetic nephropathy stretched over the exon containing the putative transcription start site, and the surrounding 5′ regulatory consensus sequence for *cis*-acting transcription factors (e.g. transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha) [TFAP2A], nuclear respiratory factor 1 [NRF-1] and nuclear factor kappaB [NFκB]) [40] (ESM Fig. 5). Nevertheless, we did not observe any differential plasma Cu/Zn SOD concentrations associated with the candidate *NOS1* haplotype. This could be due to the underlying complex and contradictory role of nitric oxide in the natural history of diabetic nephropathy [41]. Opposing forces (modified by flux in metabolic control) gain dominance at different phases of the disease thereby confounding the relationship between *NOS1* genotype and intermediate phenotype. Taken together, *NOS1* appeared to be a potentially promising diabetic nephropathy candidate gene. Hence, replication of our observation and follow-up functional genetic studies are eagerly awaited.

A *NOX1* low-frequency coding non-synonymous SNP (rs2071756G>A, R315H, MAF 0.006) was found exclusively among cases (ESM Table 5). This allele may represent the evolving group of recently recognised population-specific (i.e. founder), low-frequency and intermediate effect size variant associated with complex disease [10]. It has been suggested that several such rare variants may jointly make an appreciable contribution to population-attributable risk (i.e. the ‘common disease, rare variant’ hypothesis) [42]. In fact, it is believed that both common and rare alleles are probably complementary variants that jointly constitute the genetic landscape of complex disease like diabetic nephropathy [10]. Biologically, NOX-derived ROS is implicated in the pathogenesis of diabetic nephropathy, although the exact NOX isoform responsible is still

being debated [22]. Moreover, a recent genome scan among West African sib pairs affected by type 2 diabetes also revealed possible association between *NOX1* and renal function (measured using creatinine clearance) [43]. Therefore, further study on this SNP is warranted.

Our study has the following strengths: first, in addition to genetic association, we investigated the relationship between genotype, intermediate phenotype (i.e. corresponding plasma protein using novel assays) and disease. Within the framework of Mendelian randomisation [44], our result would have provided a reliable estimate of the impact of circulating C-terminal pro-endothelin-1 and Cu/Zn SOD levels on diabetic nephropathy. Second, in keeping with the modern theory of genetic architecture of complex traits, we systematically studied common and low-frequency SNPs from 43 pathway-guided candidate genes based on state-of-the-art understanding of diabetic nephropathy pathogenesis (ESM Fig. 1). As far as we know, our study strategy (i.e. the study of rare and common alleles simultaneously) is unique for diabetic nephropathy. Third, to ensure accuracy of allele assignment, we genotyped the low-frequency *NOX1* allele reported to be associated with diabetic nephropathy using two independent platforms (i.e. Illumina GoldenGate assay and Applied Biosystems TaqMan). Fourth, to optimise haplotype analysis, we adopted the novel variable-size sliding window haplotype analysis method developed by our co-authors to maximise study power. Fifth, we recruited participants with ‘extreme phenotype’ (established and severe renal impairment vs completely normal renal function in spite of long-standing diabetes) to enhance the efficiency of candidate gene discovery. Sixth, and finally, our study is probably the largest case–control study of this kind for diabetic nephropathy in Chinese to date.

There are several, important limitations to our study. First, although our sample size is one of the largest to date, it is still insufficient to detect weak association (i.e. $OR < 1.6$). Most irrefutable disease-susceptibility variants identified so far have allelic ORs between 1.1 and 1.5 [45]. Hence, our study could be under-powered. Second, although the 43 candidate genes were carefully chosen, recent ‘hypothesis-free’ genome-wide association studies of complex disease have often revealed novel genes that were previously unsuspected [46]. Therefore, due to ignorance of the cause of the disease, we may have missed unsuspected diabetic nephropathy candidate genes. Third, our choice of genetic variation was limited by technical feasibility of our genotyping platform. We did not study copy number variation and SNPs may not be adequate proxies for that [47]. Last, it is possible that we did not adequately address the issue of confounding by population admixture in genetic association. Nevertheless, we did try to minimise population stratification by limiting recruitment to Singaporean Chinese, most of whom are

descendants of immigrants from Southern China. Moreover, Q–Q plot did not reveal systematic deviation of test statistic (F) from what was expected, suggesting there was no cryptic relatedness or population stratification (ESM Fig. 2) [48].

In conclusion, our preliminary observations suggest that common haplotypes from *NOX4* and endothelin-1 SNP correlated with plasma Cu/Zn SOD and C-terminal pro-endothelin-1 concentrations respectively and might confer diabetic nephropathy susceptibility. Common *NOS1* and rare *NOX1* variants also revealed suggestive association with diabetic nephropathy. Future studies are needed to validate our preliminary observations.

Acknowledgements This work was supported by the Biomedical Research Council (BMRC 05/1/37/19/419), A*STAR and the Republic of Singapore. The funding agencies had no role in the research presented in the paper and the researchers were fully independent in pursuing this research. We would like to thank our colleagues at the Ang Mo Kio and Tao Payoh Polyclinics, and the National Health Group Polyclinic, Singapore for their help in recruiting study volunteers.

Duality of interest N. G. Morgenthaler is employed by B.R.A.H.M.S. AG. This company is involved in the development of in vitro diagnostic products, including C-terminal pro-endothelin-1 and Cu/Zn SOD. The authors declare that there is no other duality of interest associated with this manuscript.

References

- Schrijvers BF, de Vriese AS, Flyvbjerg A (2004) From hyperglycaemia to diabetic kidney disease: the role of metabolic, hemodynamic, intracellular factors and growth factors/cytokines. *Endocr Rev* 25:971–1010
- Liu Y, Freedman BI (2005) Genetics of progressive renal failure in diabetic kidney disease. *Kidney Int Suppl* 99:S94–S97
- Rich SS (2006) Genetics of diabetes and its complications. *J Am Soc Nephrol* 17:353–360
- Colhoun HM, McKeigue PM, Davey Smith G (2003) Problems of reporting genetic associations with complex outcomes. *Lancet* 361:865–872
- Ewens KG, George RA, Sharma K, Ziyadeh FN, Spielman RS (2005) Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test. *Diabetes* 54:3305–3318
- Shimazaki A, Kawamura Y, Kanazawa A et al (2005) Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. *Diabetes* 54:1171–1178
- Hanson RL, Craig DW, Millis MP et al (2007) Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. *Diabetes* 56:975–983
- Millis MP, Bowen D, Kingsley C, Watanabe RM, Wolford JK (2007) Variants in the plasmacytoma variant translocation gene (PVT1) are associated with end-stage renal disease attributed to type 1 diabetes. *Diabetes* 56:3027–3032
- International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
- Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40:695–701
- Bhangale TR, Rieder MJ, Nickerson DA (2008) Estimating coverage and power for genetic association studies using near-complete variation data. *Nat Genet* 40:841–843

12. American Diabetes Association (2006) Diagnosis and classification of diabetes mellitus. *Diabetes Care* 29:S43–S48
13. Lim SC, Goh SK, Lai YR et al (2006) Relationship between common functional polymorphisms of the p22phox gene (−930A > G and +242C > T) and nephropathy as a result of type 2 diabetes in a Chinese population. *Diabet Med* 23:1037–1041
14. Bettencourt P, Januzzi JL Jr (2008) Amino-terminal pro-B-type natriuretic peptide testing for inpatient monitoring and treatment guidance of acute destabilized heart failure. *Am J Cardiol* 101:67–71
15. Papassotiriou J, Morgenthaler NG, Struck J, Alonso C, Bergmann A (2006) Immunoluminometric assay for measurement of the C-terminal endothelin-1 precursor fragment in human plasma. *Clin Chem* 52:1144–1151
16. Zelko IN, Mariani TJ, Folz RJ (2002) Superoxide dismutase multigene family: a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radic Biol Med* 33:337–349
17. Clayton DG, Walker NM, Smyth DJ et al (2005) Population structure, differential bias and genomic control in a large-scale, case–control association study. *Nat Genet* 37:1243–1246
18. Li Y, Sung WK, Liu JJ (2007) Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. *Am J Hum Genet* 80:705–715
19. Pawlak K, Pawlak D, Mysliwiec M (2005) Cu/Zn superoxide dismutase plasma levels as a new useful clinical biomarker of oxidative stress in patients with end-stage renal disease. *Clin Biochem* 38:700–705
20. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30:3894–3900
21. Li JM, Shah AM (2003) ROS generation by nonphagocytic NADPH oxidase: potential relevance in diabetic nephropathy. *J Am Soc Nephrol* 14:S221–S226
22. Bedard K, Krause KH (2007) The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol Rev* 87:245–313
23. Orient A, Donkó A, Szabó A, Leto TL, Geiszt M (2007) Novel sources of reactive oxygen species in the human body. *Nephrol Dial Transplant* 22:1281–1288
24. Ding H, Hashem M, Triggle C (2007) Increased oxidative stress in the streptozotocin-induced diabetic apoE-deficient mouse: changes in expression of NADPH oxidase subunits and eNOS. *Eur J Pharmacol* 561:121–128
25. Richter CM (2006) Role of endothelin in chronic renal failure—developments in renal involvement. *Rheumatology* 45:iii36–iii38
26. Naicker S, Bhoola KD (2001) Endothelins: vasoactive modulators of renal function in health and disease. *Pharmacol Ther* 90:61–88
27. Hargrove GM, Dufresne J, Whiteside C, Muruve DA, Wong NC (2000) Diabetes mellitus increases endothelin-1 gene transcription in rat kidney. *Kidney Int* 58:1534–1545
28. Lillie EO, Mahata M, Khandrika S et al (2007) Heredity of endothelin secretion: human twin studies reveal the influence of polymorphism at the chromogranin A locus, a novel determinant of endothelial function. *Circulation* 115:2282–2291
29. Freedman BI, Yu H, Anderson PJ, Roh BH, Rich SS, Bowden DW (2000) Genetic analysis of nitric oxide and endothelin in end-stage renal disease. *Nephrol Dial Transplant* 15:1794–1800
30. Kanková K, Stejskalová A, Pácal L et al (2007) Genetic risk factors for diabetic nephropathy on chromosomes 6p and 7q identified by the set-association approach. *Diabetologia* 50:990–999
31. Pinto-Sietsma SJ, Herrmann SM, Schmidt-Petersen K et al (2003) Role of the endothelin-1 gene locus for renal impairment in the general nondiabetic population. *J Am Soc Nephrol* 14:2596–2602
32. Rossi GP, Pitter G (2006) Genetic variation in the endothelin system: do polymorphisms affect the therapeutic strategies? *Ann N Y Acad Sci* 1069:34–50
33. Diefenbach K, Arjomand-Nahad F, Meisel C et al (2006) Systematic analysis of sequence variability of the endothelin-1 gene: a prerequisite for association studies. *Genet Test* 10:163–168
34. Stevens PA, Brown MJ (1995) Genetic variability of the ET-1 and the ETA receptor genes in essential hypertension. *J Cardiovasc Pharmacol* 26(Suppl 3):S9–S12
35. Winker R, Garland EM, Rüdiger HW et al (2005) Influence of an insertion variant in the 5'UTR of the endothelin-1 gene on orthostatic intolerance. *Am J Med Sci* 330:166–171
36. Popowski K, Sperker B, Kroemer HK et al (2003) Functional significance of a hereditary adenine insertion variant in the 5'-UTR of the endothelin-1 gene. *Pharmacogenetics* 13:445–451
37. Komers R, Lindsley JN, Oyama TT, Allison KM, Anderson S (2000) Role of neuronal nitric oxide synthase (NOS1) in the pathogenesis of renal hemodynamic changes in diabetes. *Am J Physiol Renal Physiol* 279:F573–F583
38. Thomson SC, Deng A, Komine N, Hammes JS, Blantz RC, Gabbai FB (2004) Early diabetes as a model for testing the regulation of juxtaglomerular NOS I. *Am J Physiol Renal Physiol* 287:F732–F738
39. O'Bryan GT, Hostetter TH (1997) The renal hemodynamic basis of diabetic nephropathy. *Semin Nephrol* 17:93–100
40. Hall AV, Antoniou H, Wang Y et al (1994) Structural organization of the human neuronal nitric oxide synthase gene (NOS1). *J Biol Chem* 269:33082–33090
41. Komers R, Anderson S (2003) Paradoxes of nitric oxide in the diabetic kidney. *Am J Physiol Renal Physiol* 284:F1121–F1137
42. Iyengar SK, Elston RC (2007) The genetic basis of complex traits: rare variants or “common gene, common disease”? *Methods Mol Biol* 376:71–84
43. Chen G, Adeyemo AA, Zhou J et al (2007) A genome-wide search for linkage to renal function phenotypes in West Africans with type 2 diabetes. *Am J Kidney Dis* 49:394–400
44. Davey Smith G, Ebrahim S (2004) Mendelian randomisation: prospects, pitfalls and limitations. *Int J Epidemiol* 33:30–42
45. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177–182
46. Frayling TM, McCarthy MI (2007) Genetic studies of diabetes following the advent of the genome-wide association study: where do we go from here? *Diabetologia* 50:2229–2233
47. Stranger BE, Forrest MS, Dunning M et al (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–853
48. McCarthy MI, Abecasis GR, Cardon LR et al (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9:356–369