Diabetologia (2007) 50:990–999
DOI 10.1007/s00125-007-0606-3

ARTICLE

Genetic risk factors for diabetic nephropathy on chromosomes 6p and 7q identified by the set-association approach

K. Kaňková · A. Stejskalová · L. Pácal · S. Tschoplová · M. Hertlová · D. Krusová · L. Izakovičová-Hollá · M. Beránek · A. Vašků · S. Barral · J. Ott

Received: 17 July 2006 / Accepted: 4 December 2006 / Published online: 8 March 2007

Abstract
Aims/hypothesis In the present study we investigated potential associations of a set of 45 single nucleotide polymorphisms (SNP) in 20 candidate genes on eight chromosomes with diabetic nephropathy (DN) in type 2 diabetes mellitus. We aimed to compare two methodological approaches suitable for analysing susceptibility to complex traits: single- and multi-locus analyses.
Materials and methods The study comprised a total of 647 subjects in one of three groups: diabetes with or without DN, or no diabetes. Genotypes were detected by PCR-based methodology (PCR only, PCR plus RFLP, or allele-specific PCR). Haplotypes were inferred in silico. Set association (tested using SUMSTAT software) was used for multilocus analysis.
Results After correction for multiple comparisons, only one SNP, in the gene encoding the receptor of advanced glycation end products, *AGER* 2184A/G (gene symbol formerly known as *RAGE*) showed a significant association with DN (p=0.0006) in single-locus analysis. In multi-locus analysis, six SNPs exhibited a significant association with DN: four SNPs on chromosome 6p (*AGER* 2184A/G, *LTA* 252A/G, *EDN1* 8002G/A and *AGER* -429T/C) and two SNPs on chromosome 7q (*NOS3* 774C/T and *NOS3* E298D), omnibus p=0.033. Haplotype analysis revealed significant differences between DN and control groups in haplotype frequencies on chromosome 6 (p=0.0002); however, there were no significant difference in the frequencies of the *NOS3* haplotypes on chromosome 7. Logistic regression analysis identified SNPs *AGER* 2184A/G and *NOS3* 774C/T, together with diabetes duration and HbA_{1c}, as significant predictors of DN. Testing for interactions between SNPs on chromosomes 6 and 7 did not provide significant evidence for epistatic interaction.
Conclusions/interpretation Using the set-association approach we identified significant associations of several SNPs on chromosomes 6 and 7 with DN. The single- and multi-locus analyses represent complementary methods.

Keywords Diabetic nephropathy · Endothelin · Haplotype · LTA · Lymphotoxin α · Nitric oxide synthase · NOS3 · RAGE · Receptor of advanced glycation end products · Set association

Abbreviations
DN diabetic nephropathy
FDR false discovery rate

Electronic supplementary material The online version of this article (doi:10.1007/s00125-007-0606-3) contains supplementary material, which is available to authorised users.

K. Kaňková (✉) · A. Stejskalová · L. Pácal · S. Tschoplová · L. Izakovičová-Hollá · M. Beránek · A. Vašků
Faculty of Medicine, Department of Pathophysiology, Masaryk University Brno, Komenskeho nam. 2, 662 43 Brno, Czech Republic
e-mail: kankov@med.muni.cz

M. Hertlová
Department of Internal Medicine, University Hospital Brno-Bohunice, Brno, Czech Republic

D. Krusová
Department of Internal Medicine, St Anne's Faculty Hospital, Brno, Czech Republic

S. Barral · J. Ott
Rockefeller University, Laboratory of Statistical Genetics, New York, NY, USA

HWE	Hardy–Weinberg equilibrium
LD	linkage disequilibrium
LTA	lymphotoxin α
OR	odds ratio
SNP	single nucleotide polymorphism
VNTR	variable number of tandem repeats

Introduction

Diabetic nephropathy (DN) is a serious long-term consequence of diabetes, affecting roughly one-third of both type 1 and type 2 diabetic patients. Complex interplay between haemodynamic and metabolic factors, initiated by hyperglycaemia, results in the renal morphological and functional alterations that characterise DN [1]. Although sustained hyperglycaemia has a profound effect on the kidneys, hyperglycaemia-independent factors (blood pressure, obesity, etc.) also play an important role.

It has been established that the onset and progression of DN are genetically determined [2, 3]. Ethnic differences in prevalence, uneven incidence over the time-course of the disease (culminating in the second decade of diabetes duration) [4], familial clustering in different populations for both type 1 and 2 diabetes [5–13], and the results of segregation analyses [14, 15] suggest the existence of susceptibility genes for DN in addition to those leading to diabetes. Linkage analyses have thus far been inconclusive [16–19]. Nevertheless, results indicate that DN has the characteristics of a complex trait. Association studies investigating functional and positional candidates using case-control or family-based trio designs, led to the identification of numerous susceptibility loci [20]; however, these results were difficult to replicate in different populations and study settings. The majority of conventional association studies investigated the isolated effects of one or a small number of markers, i.e. marginal associations.

Several approaches have recently been developed that test markers jointly [21, 22]. A novel paradigm lies in searching for statistical interactions between loci, which might have little or no marginal effects individually but contribute jointly to the trait. This is all the more justified given the lack of uniform criteria for significance in multi-locus or genome-wide association studies and the inter-population genetic heterogeneity, which can influence marginal associations to a greater extent than the interactions, and thus partially explain the difficulty in repeating positive associations.

The aim of our study was to investigate a possible association of a set of single nucleotide polymorphisms (SNPs) in functional candidate genes with DN in type 2 diabetes using two approaches: (1) traditional 'single-locus' analyses by subsequent correction for multiple comparisons; and (2) one of the recently proposed 'multi-locus' methods, the so-called set-association [23]. Although this study is still far from being genome-wide, and the selection of SNPs was hypothesis-driven, to the best of our knowledge it is one of the first to consider the anticipated oligo-/multigenic pattern of DN and to simultaneously assess possible associations of multiple interacting markers. Furthermore, comparison of two alternative methods can demonstrate whether the newly proposed multi-locus approach offers any substantial advantages over the classical one.

Subjects and methods

Subjects A total of 647 unrelated subjects (286 men/361 women) were enrolled into the study and assigned to one of three groups: type 2 diabetes with DN (n=235; age 63.5± 13.9 years); diabetes without DN (n=184; age 62.6± 11.4 years), and no diabetes (n=228; 59.1±15.5). The clinical characteristics of the study subjects are shown in Electronic supplementary material (ESM) Table 1. All diabetic subjects were followed in diabetic outpatient clinics; those with DN were also followed in nephrology units. The non-diabetic group consisted of subjects with fasting glycaemia within the normal range, as assessed by their general practitioners during routine medical check-ups. All subjects were from the same geographical area and were of the same ethnic (white, central European) origin. The study was performed in accordance with the principles of the Declaration of Helsinki and approved by the Ethical Committee of Medical Faculty, Masaryk University Brno. Informed consent was obtained from all patients prior to their inclusion in the study.

Definition of phenotype The presence of DN was assessed on the basis of the patients' medical history of diabetes with established glucose-lowering treatment and periodical measurement of AER and GFR. DN cases were subjects with minimally persistent microalbuminuria (and subsequent more advanced stages of DN, i.e. overt proteinuria with variable decline of renal functions or end-stage renal disease). Repeated detection (at least two successive measurements within 3–6 months) of an AER of ≥30 mg/day or ≥20 μg/min was used as the minimal inclusion criterion. Negative results of periodical screening for microalbuminuria and normal plasma creatinine in the presence of diabetes were used as classification criteria for diabetic non-DN subjects.

Description of gene variants and genotyping methods The panel of genetic markers comprised 45 SNPs in 20 genes on eight chromosomes (1, 4, 6, 7, 16, 17, 19 and 20; for the

complete list of SNPs see ESM Table 2). The selection of genes was pathway-based (i.e. direct association study); their products are involved in the pathogenesis of diabetic complications, and majority of them have previously been studied for marginal association. The proteins coded for by the genes studied included components of the renin-angiotensin system, other haemodynamic factors, antioxidant enzymes, cytokines, growth factors, receptors and extracellular matrix remodelling enzymes. The selection of particular SNPs was based on: (1) their population frequency; (2) their known functional or regulatory impact; and/or (3) a previously described association with DN. Figure 1 shows the basic characteristics of the SNPs studied.

DNA was isolated from peripheral blood leucocytes by a standard method using proteinase K. Genotyping was performed by means of PCR followed by either: (1) direct electrophoretic separation of PCR fragments (insertion/deletion, variable number of tandem repeats (VNTR) polymorphisms or allele-specific PCR); or (2) RFLP with subsequent electrophoresis. Details of the methodology are provided in ESM Table 3.

Linkage Disequilibrium Analyzer 1.0 [24] and GOLD software [25] were used for analysis of linkage disequilibrium (LD). Lewontin's standardised disequilibrium coefficient, D' (rather than r^2, because the latter is more sensitive to the variation in allele frequency [26]), was used to describe LD among SNPs.

Statistical analysis Deviations of genotype distribution from Hardy–Weinberg equilibrium (HWE) were tested by χ^2 tests. Differences in allele frequencies between case and control subjects were tested by likelihood ratio χ^2 tests for 2×2 tables (two alleles, case vs control subjects). The effect of multiple comparisons was assessed by calculating the overall false discovery rate (FDR) at a 5% significance level according to the Benjamini–Hochberg method [27].

Set-association analysis was performed as described previously [23], using the SUMSTAT program (http://www.genemapping.cn/sumstat.html, last accessed in January 2007). Briefly, for each of the 45 SNPs (SNP_i) a single statistic (s_i)—the absolute mean difference in genotype codes between affected and unaffected groups—was calculated. SNPs were ordered according to their s_i value (i.e. $s_{(1)} \geq s_{(2)} \geq s_{(3)}$ ...), irrespective of their genomic location. Sums with increasing numbers of SNP markers (i.e. n) were formed, starting with the markers ranked highest $S_{(n=1)}=s_{(1)}$; $S_{(n=2)}=s_{(1)}+s_{(2)}$, and so on. The primary interest was to find the n of SNPs in S that reflects the association of this set of markers with the disease. The significance levels (empirical p values; p_n) associated with the n-th S were determined by permutation testing (20,000 permutations). As the number n of SNP markers in S increases, the pattern of p values changes; following an initial decline to a minimum (min-p_n), it increases again. The smallest empirical significance level was considered our statistic of interest, and its significance level ($p_{\min}$, representing the whole genome significance), was determined again by a permutation test (20,000 permutations).

Estimation of haplotype frequencies was performed by the Bayesian-based algorithm (PHASE software. Available from http://depts.washington.edu/ventures/UW_Technology/Express_Licenses/PHASEv2.php, last accessed in January 2007) [28]. The significance of differences in estimated haplotype frequencies between case and control subjects in this study was assessed empirically via permutation testing (5,000 permutations). In addition, haplotype-specific effects were analysed using inferred haplotype pairs by computing χ^2 statistics for a 9×2 table.

Power analysis was performed post hoc for a given sample size to assess the power of detecting an allele frequency in cases different from that in control subjects for the best associated marker in single and multilocus analyses. Calculations were done with the PAWE program

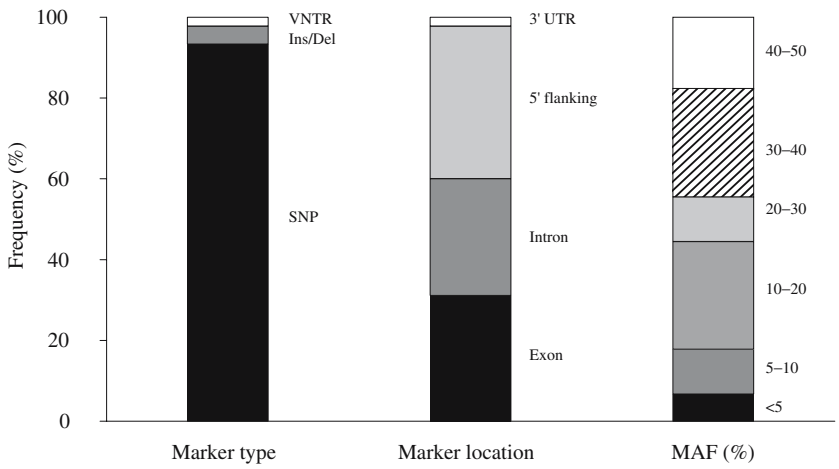

Fig. 1 Basic characteristics of the polymorphisms studied. The Marker location column shows the frequency of particular polymorphisms according to their position in a gene, and the MAF column shows the frequency of the polymorphisms according to their minor allele frequency (MAF). *Ins/Del* insertion/deletion; *UTR* untranslated region

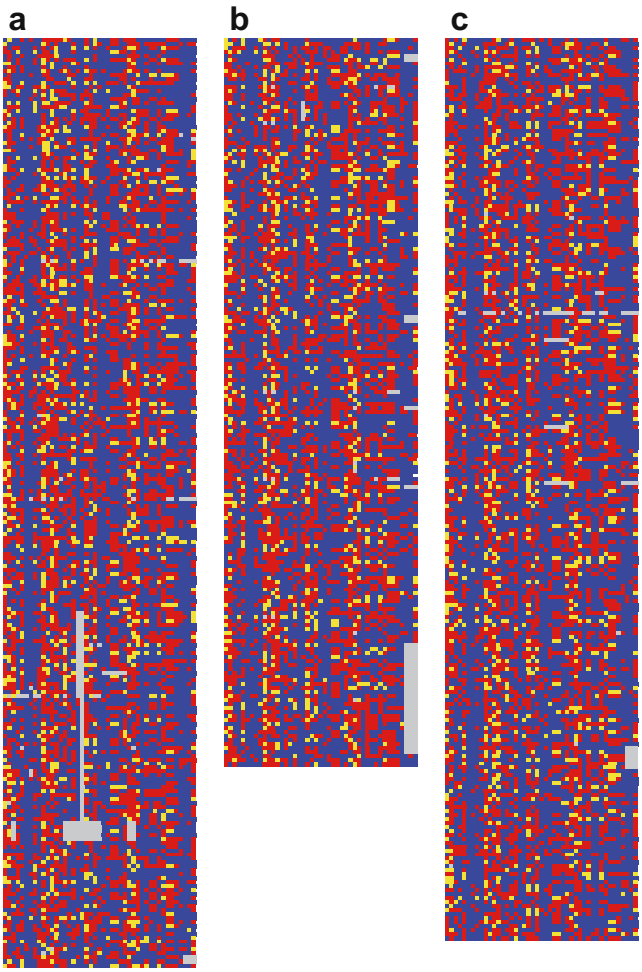

Fig. 2 Visual genotype in the three groups: **a** diabetes with DN, **b** diabetes without DN, **c** no diabetes. The columns represent the 45 SNPs, and the rows represent the subjects. *Blue* homozygous genotype for the common allele; *red* heterozygous genotype; *yellow* homozygous genotype for the rare allele; *grey* missing data

(http://linkage.rockefeller.edu/pawe/, last accessed in January 2007).

In order to identify genetic as well as non-genetic variables that may contribute to predicting the disease phenotype, we carried out a forward stepwise logistic regression, a sequential procedure of adding one input variable at a time to build up a regression model in which the dependent variable (i.e. presence or absence of DN) is represented as the linear combination of independent variables (several clinical parameters and genotypes of 45 SNPs). Genotype codes were used as quantitative variables (AA=0, AB=1, BB=2). These codes were equivalent to counts of B alleles and reflected the allelic contribution of a given SNP such that each SNP represented one variable in the logistic regression.

Testing for gene–gene interactions was performed by conducting tests for contingency tables with subsequent correction for multiple comparisons (the overall FDR at 5% significance level). For a given pair of SNPs, we computed χ^2 statistics in 3×3 tables (three genotypes at one SNP vs three genotypes at the other SNP) as follows: (1) for the case group only (χ^2_{cases} with 4 *df*); (2) for the control group only (χ^2_{controls} with 4 *df*); and (3) for pooled case and control groups ($\chi^2_{\text{cases+controls}}$ with 4 *df*). To test if an interaction between a particular marker pair might be different between case and control subjects, we computed a new approximate χ^2 statistic as $\chi^2 = [(\chi^2_{\text{cases}} + \chi^2_{\text{controls}}) - \chi^2_{\text{cases+controls}}]$ with 4 *df* and its associated *p* value.

Results

Genotyping was nearly complete (98.63%); missing genotypes were due to either consistent PCR dropout or depletion of template DNA. Figure 2 shows the 'visual genotype' in each of the three groups. Data analysis employed the following strategies: (1) testing for departure from HWE and for an eventual stratification in both control groups, with the aim of pooling control subjects; (2) classical comparison of genotype and allele frequencies of each SNP between case and control subjects; (3) set-association with subsequent refinement of the results by haplotype analysis; (4) analysis of the relative contribution of genetic variables together with non-genetic covariates to the observed phenotype by stepwise logistic regression; and (5) testing the existence of epistatic gene-gene interactions between SNP markers located on different chromosomes.

HWE was not significantly violated in any of the three groups for any of the SNPs studied. Significant population stratification, ascertained by comparison of genotype and allele frequencies and assessment of the departure from HWE, was not detected in either of the control groups; hence, the two original control groups (DM non-DN and non-diabetic subjects, originally serving the purpose of distinguishing between eventual associations of markers with diabetes itself vs its complications) were pooled for all subsequent analyses.

Single-locus approach The genotype and allele frequencies of particular SNPs in DN and control subjects were compared marker by marker (for allelic frequencies see ESM Table 4). After correction for multiple comparisons (45 single tests) by calculating the FDR (α=0.05), there were no significant differences for any SNPs except for *AGER* 2184A/G, which showed significant differences in allele frequencies between case and control subjects (p= 0.0006), that is, a marginal association of the 2184G allele with DN.

Set-association approach Using the SUMSTAT program, we tested whether a set of several SNPs might be synergistically associated with DN. This method selects the 'best' set of n SNPs, whose sum statistic (S) is associated with the highest significance. This leads to the inference that this entire set of SNPs might be interacting in some way to increase disease risk or, alternatively, that all SNPs contribute independently to disease risk. ESM Table 4 and Fig. 3 shows the combined statistics (s_i) for each SNP, and sums (S_n) with increasing numbers of SNP markers (Fig. 3a), and changes of p_n values corresponding to the sums formed by increasing numbers of SNP markers (Fig. 3b). Six SNPs exhibited a significant set-association with DN: *NOS3* E298D, *AGER* 2184A/G, *LTA* 252A/G, *EDN1* 8002G/A, *NOS3* 774C/T and *AGER* -429T/C, overall p=0.033 (permutation test, 20,000 permutations). Of these, four SNPs were localised in genes on chromosome 6 (*EDN1*, *LTA* and *AGER*) and two were localised in a gene on chromosome 7 (*NOS3*).

Post hoc power analysis assessed the power of detecting difference in allele frequency between control subjects and cases in a given sample size for the best associated marker in the single- and the multi-locus analyses. We considered the power of detecting an allele frequency in cases to be different from that in control subjects. For our best associated marker, SNP *AGER* 2184A/G, the frequency of the G allele in the 412 control subjects was 0.15. The power needed to detect a significant difference in the allele frequency (f) in the 235 case individuals was computed as follows. We assumed significance levels of α=0.05 and α=0.05/45=0.001 (for 45 SNPs) and 1% error rates that one allele was erroneously identified as the other allele. Statistical powers to detect *AGER* 2184G allele frequencies 0.20; 0.23 or 0.25 in the case group were (A) 0.61; 0.94 and 0.99, respectively, for the uncorrected significance levels (α=0.05) and (B) 0.14; 0.58 and 0.84, respectively, for Bonferonni-corrected significance levels (α=0.001). Clearly, our samples had sufficient power to detect the frequency of 0.23 actually observed in the DN group, even with a Bonferroni-corrected significance level of 0.001.

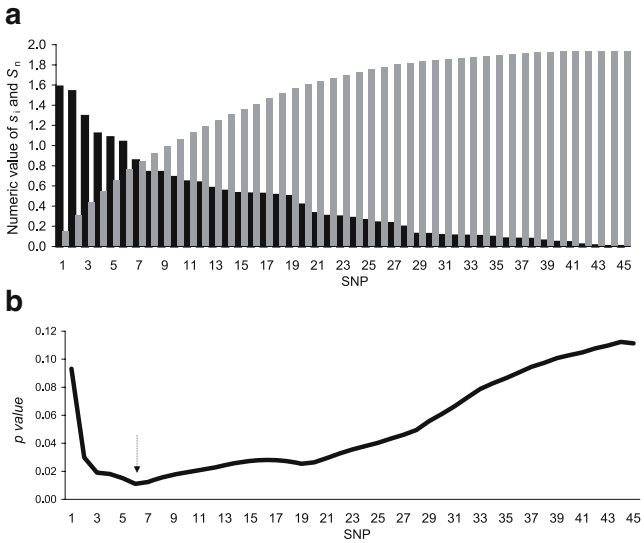

Fig. 3 **a** Comparison of combined statistics, $s_i[t_i \times u_i]$ for each SNP (*black bars*) and sums, S_n, with increasing number of terms (*grey bars*). **b** Changes of p_n values corresponding to the sums formed by increasing number of terms. The arrow indicates statistics of interest

Haplotype and multi-locus genotype pattern analysis of chromosome 6 Estimation of haplotype frequencies was carried out with the PHASE program, using unphased genotypes of SNPs in the *EDN1*, *LTA* and *AGER* genes, spanning the region 6p24-6p21.3 (see Fig. 4). A total of 14 haplotypes were inferred, eight of them with a frequency >1% (the remainder were combined into a single group denoted 'rare'). Haplotype frequencies (see Table 1, upper part) were compared between affected and unaffected subjects by permutation testing. A significant difference in haplotype frequencies with omnibus p=0.0002 was ascertained (5,000 permutations).

To analyse haplotype-specific effects on the trait (i.e. presence of DN) we assigned pairs of haplotypes to individuals. For each of the haplotypes identified we counted how many individuals appeared to have this particular haplotype. We computed the χ^2 statistic for this 9×2 table. Odds ratios (ORs), 95% CIs and associated Fisher p values for each of the haplotypes inferred are included in Table 1. We have identified three specific haplotypes, 2122 (AACG), 1122 (GACG) and 2211 (AGTA) with ORs>1.5 (ORs of 2.28, 1.80 and 1.64, respectively) exhibiting a significant association with the disease (p=0.0152, p=0.0237 and p=0.0191, respectively).

The results of pairwise LD analysis between all studied markers located on the 6p chromosome (including those contained in the disease haplotypes) in the control group indicated the most consistent LD between the two *AGER* SNPs ($D'\geq 0.85$, $p<1\times10^{-4}$) (see Fig. 5a–c). This was in agreement with our previous findings [29]. Conversely, LD between *EDN1* and the remaining markers was low ($D'\leq$ 0.34). Considering the large distance of *EDN1* from the rest of the 'set-associated' markers on chromosome 6 (~19.2 Mb) and consequent likely recombination, we concluded that *EDN1* 8002G/A is probably not a part of the same putative haplotype block and that its association could be considered as independent of the remaining SNPs on this chromosome.

More detailed analysis of previously set-associated SNPs on chromosome 6 revealed highly significant differences in their haplotype distribution. Essentially, all haplotypes containing more than one minor allele were over-represented in the DN group.

Haplotype analysis of chromosome 7 Haplotype analysis for the markers on chromosome 7 was performed as described for chromosome 6. There are another five SNPs located in the *NOS3* gene in addition to those found to be significant by the SUMSTAT program (*NOS3* E298D and 774C/T). Since *NOS3* appears to be an interesting candidate gene previously studied for linkage as well as association with DN, we performed a haplotype analysis of this region, including all seven SNPs in the *NOS3* gene (see Table 1). A total of 54 haplotypes were inferred, 13 of which had a frequency >1%. Haplotype distribution did not exhibit significant differences between affected and unaffected subjects (omnibus significance $p=0.274$, 5,000 permutation samples, 5,000 iterations). Major allele frequencies of both set-associated *NOS3* polymorphisms (i.e. E298D and 774C/T) were higher in the DN group (see ESM Table 4). Table 1 (lower part) suggests that haplotypes containing major alleles in these positions are slightly more common (however statistically insignificantly) in the DN group. This accounts for the general over-representation (and association) of both 298E and 774C alleles with DN without significant difference in haplotype distribution.

LD was generally very high across the whole *NOS3* gene (Fig. 5d–f), with almost complete LD between SNPs 774C/T and E298D ($D' \geq 0.92$, $p < 1 \times 10^{-4}$). The information content did not increase with simultaneous detection of both SNPs, which indicates that they are likely to be redundant when capturing *NOS3* gene variability.

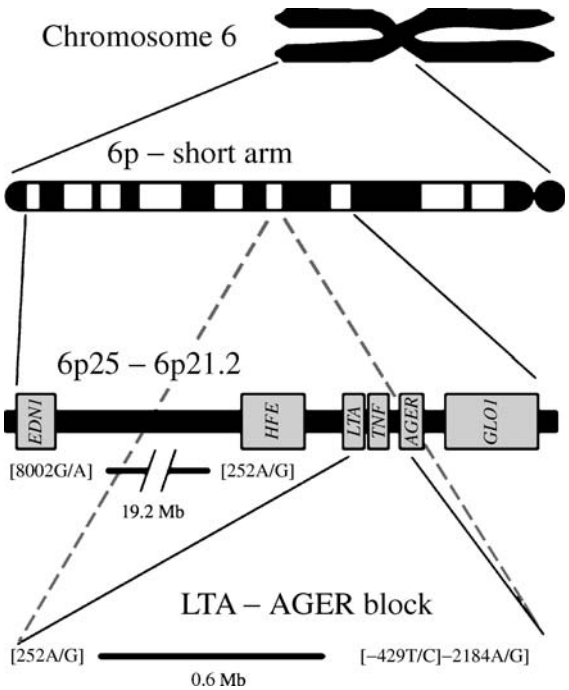

Fig. 4 Details of the location and mutual distances of all markers located on the short arm of the chromosome 6 and those constituting the risk haplotype for DN

Logistic regression Using forward stepwise logistic regression, we attempted to analyse relationships between disease status and genetic and non-genetic variables, i.e. to identify predictors of clinical outcome. The presence or absence of DN was considered an outcome variable, and the following input variables were sequentially added into the model: age, sex, diabetes duration, systolic and diastolic blood pressures, HbA_{1c} and the genotypes of 45 SNPs. Table 2 shows variables that remained significantly included into the model. HbA_{1c}, diabetes duration and genotypes of the *AGER* 2184A/G and *NOS3* 774C/T SNPs were the strongest predictors of DN. *TGFB1* R25P and *AGER* 2245G/A also appeared to be good indicators of disease status. The overall joint effect of all variables remaining significantly in the model was $\chi^2=57.01$, $df=10$, $p<1\times10^{-6}$.

Gene–gene interactions The SUMSTAT algorithm does not allow interactions among associated markers to be specifically tested for. We re-analysed our data using the set-association algorithm once again, this time omitting the most significant marker (*AGER* 2184A/G), to determine whether and how the results would change. Without this SNP, we were not able to identify any significant set of associated markers among the remaining 44 SNPs (all $p_n>0.05$, $p_{min}>0.05$, 20,000 permutations). Apparently, the *AGER* 2184A/G SNP, which exhibits both single-and multi-locus associations, is a driving marker of the set-association, suggesting a non-additive mode of putative inter-locus interaction.

We evaluated potential multiplicative interactions between four SNPs in the *EDN1*, *LTA* and *AGER* genes, located on chromosome 6, and all seven SNPs in the *NOS3* gene, located on chromosome 7, and tested whether pairwise interactions are significantly different between case and control subjects. For each of the 28 possible SNP pairs, we computed the appropriate χ^2 statistic with 4 *df* and its associated *p* value. After correction for multiple testing, none of the SNP pairs showed significant differences (data not shown). Therefore, although the results of the set-association analysis suggested a non-additive mode of interaction, we were not able to prove that genotypes at chromosomes 6 and 7 epistatically affect the genetic susceptibility to DN.

Discussion

Complex diseases, including diabetes and its complications, are presumably products of interactions among genetic variability in multiple susceptibility genes on different chromosomes. Disease occurs when a particular combination(s) of susceptibility genotypes are present in a given individual exposed to environmental factors; however, each

Table 1 Haplotype frequency estimates (using the PHASE program) for chromosomes 6 and 7, and the results of comparisons of haplotype frequencies

	Haplotype	DM+DN (n=235)		Control subjects (n=412)		OR (95% CI)[b]
		Frequency	SE	Frequency	SE	
Chromosome 6						
1	1111	0.381	0.009	0.483	0.006	0.70 (0.56–0.88)
2	1211	0.181	0.009	0.164	0.006	1.12 (0.79–1.58)
3	2111	0.128	0.008	0.126	0.006	0.86 (0.58–1.29)
4	1122	0.106	0.006	0.061	0.004	1.80 (1.09–2.96)[b]
5	2211	0.079	0.008	0.047	0.005	1.64 (1.10–2.45)[b]
6	1222	0.050	0.006	0.043	0.004	1.25 (0.78–2.01)
7	2122	0.035	0.006	0.016	0.003	2.28 (1.19–4.36)[b]
8	2222	0.019	0.005	0.013	0.002	1.45 (0.67–3.12)
9	rare	0.020	–	0.047	–	–
					p=0.0001[a]	
Chromosome 7						
1	1121111	0.362	0.006	0.332	0.006	–
2	2112112	0.135	0.005	0.100	0.004	
3	1121112	0.122	0.005	0.121	0.005	
4	2121221	0.098	0.004	0.124	0.004	
5	2221221	0.065	0.002	0.071	0.002	
6	1121221	0.048	0.005	0.063	0.004	
7	2121111	0.023	0.004	0.017	0.003	
8	1121222	0.022	0.004	0.024	0.003	
9	2112111	0.022	0.004	0.027	0.003	
10	1112112	0.018	0.003	0.017	0.002	
11	2121112	0.011	0.003	0.011	0.002	
12	1121121	0.010	0.002	0.021	0.003	
13	2221121	0.001	0.001	0.010	0.001	
14	rare	0.064	–	0.063	–	
					p=0.274[a]	

Haplotypes were inferred from genotypes of SNPs in telomere to centromere orientation (chromosome 6) or in 5′ to 3′ orientation (chromosome 7), i.e. *EDN1* 8002G/A → *LTA* 252A/G → *AGER* −429T/A and 2184A/G or –786T/C → −691C/T → 4a/5bVNTR → 5495A/G → 774C/T → 894G/T → 11G/T, respectively. The nucleotide in the first position of a particular substitution was denoted as 1 (major allele), the second nucleotide as 2 (minor allele). Haplotypes were ordered according to decreasing haplotype frequency in the DM + DN group; those with frequencies less than 1% in both groups were pooled together as 'rare'

[a] Global differences in haplotype frequency profiles between particular groups were tested by permutation testing (PHASE, 5,000 permutation samples, 5,000 iterations). ORs and 95% CIs were assessed by computing χ^2 statistics for the 9×2 table

[b] OR>1.5 with significant association with DN (Fisher p<0.05)

DM type 2 diabetes mellitus

individual locus is likely to have a low or no marginal effect on its own. In the present study we used the set-association approach, a tool specifically developed to tackle this problem, to ascertain whether any SNPs of the total set of 45 markers in different candidate genes for DN are synergistically associated with the phenotype. The results can be summarised as follows. A set of six SNPs in four different genes localised on two chromosomes, namely, *AGER* -429T/C and 2184A/G, *LTA* 252A/G and *EDN1* 8002G/A (chromosome 6) and *NOS3* 774C/T and E298D (chromosome 7), exhibited significant joint association with DN Only one of these SNPs, *AGER* 2184A/G, showed a significant marginal association with DN in a single-locus analysis. This SNP appeared to be a decisive [driving] marker among the group of set-associated SNPs, since omitting the *AGER* 2184A/G SNP completely abolished the association of the remaining SNPs. In addition, frequencies of the *EDN1-LTA-AGER* haplotypes were significantly different between case and control subjects, whereas a similar situation was not observed for the *NOS3* haplotypes, and significant associations hold only for the two neighbouring SNPs *NOS3* 774C/T and E298D. Furthermore, together with diabetes duration and HbA_{1c}, the SNPs *AGER* 2184A/G and *NOS3* 774C/T proved to be the strongest genetic predictors of DN. Finally, we have not proved any significant epistatic interactions between SNP markers on chromosomes 6 and 7, despite the decisive role of SNP *AGER* 2184A/G among set-associated markers.

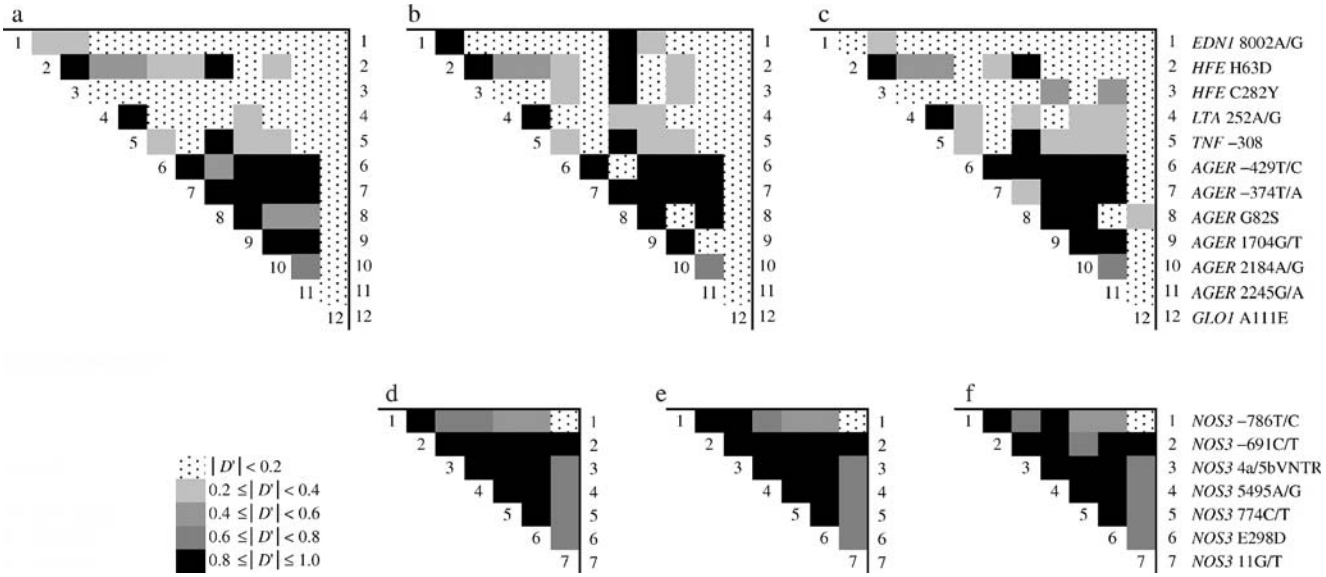

Fig. 5 Plot of pairwise linkage disequilibrium (D') values for the SNPs studied on chromosome 6 (**a**–**c**) and chromosome 7 (**d**–**f**). $|D'|$ values are shaded in greyscale according to the key. Values shown are for the whole set of subjects (**a**, **d**), diabetic subjects with DN (**b**, **e**) and control subjects (**c**, **f**)

Ascertained associations might either directly reflect causality of substitutions or, alternatively, suggest the existence of causal markers in their vicinity. In a previous study that focused solely on the *AGER* gene, we identified the so-called $RAGE_2$ haplotype (containing the −429C/2184G variants) as a disease marker for DN [29]. There is increasing evidence supporting the role of receptor of advanced glycation end products (RAGE) in DN pathogenesis [30]. While the *EDN1* locus is quite distant from the other two, *LTA* and *AGER* are both located within a relatively short stretch of 6p21.3 that harbours genes of the MHC-III. In the case of *LTA* itself, many studies have implicated the SNP 252A/G in susceptibility to various diseases. The 252G allele associated with DN in our study has recently been associated with susceptibility to myocardial infarction in Japanese genome-wide association study comprising over 2,000 individuals genotyped for over 65,000 SNPs [31]. The study showed that the transcriptional activity of haplotypes containing the 252G allele was 1.5-fold higher than that of the other haplotypes. This finding is in agreement with functional evidence that the 252G allele is associated with higher *LTA* [32] and *TNF* [33] expression (only 1.2 kB separate the polyadenylation site of *LTA* and the transcription start site of *TNF*), as well as increased *LTA* transcriptional activity [34]. TNF-α undoubtedly contributes to renal diabetic pathology—

Table 2 Results of forward stepwise logistic regression

	Parameter	β	SE	*t* ratio	*p* value
1	Constant[a]	2.834	2.966	0.956	0.339
2	Diabetes duration	0.092	0.034	2.732	0.006
3	*AGER* 2184A/G	1.480	0.441	3.354	0.001
4	*NOS3* 774C/T	−0.977	0.444	−2.204	0.028
5	HbA_{1c}	0.446	0.154	2.892	0.004
6	*TGFB1* R25P	1.186	0.602	1.968	0.049
7	*AGER* 2245G/A	−1.299	0.602	−2.159	0.031
8	Diastolic BP	−0.043	0.023	−1.838	0.066
9	*FGF2* −834T/A	−2.135	1.310	−1.629	0.103
10	*MTHFR* 677C/T	−0.649	0.382	−1.697	0.090
11	*ACE* I/D	−0.560	0.372	−1.507	0.132
			χ^2=57.01, *df*=10, $p<1\times10^{-6}$		

The presence (or absence) of DN was considered as the outcome. The following parameters were considered as predictor variables: age, sex, diabetes duration, systolic and diastolic BP, HbA_{1c} and the genotypes of 45 SNPs

[a] This is a constant of the equation indicating the predicted log odds for the dependent variable (case/control) when all the coefficients for the different independent variables are set to zero

TNF-α levels are correlated with urinary albumin excretion in patients with type 2 diabetes at an early stage of nephropathy [35, 36], as well as in rodents with streptozotocin-induced diabetes [37], and TNF-α is a potent inducer of *AGER* expression in endothelial cells [38]. Taken together, although the telomeric part of the MHC III appears to be an interesting region in the context of DN, more detailed mapping and functional studies are warranted to clarify potential causality of the particular markers.

NOS3 is a well-documented functional candidate for DN susceptibility because of the involvement of nitric oxide in DN pathogenesis [39, 40]. One of the first indications that genes within chromosome 7q35 (including *NOS3*) could influence genetic susceptibility to DN came from the non-parametric linkage study in Pima Indians [16]. Several positive associations of some of the numerous *NOS3* SNPs with renal disease were documented in type 1 [41] and type 2 diabetes [42], as well as in non-diabetic subjects [43, 44], while other studies did not find an association [45]. The two 'set-associated' SNPs in our study—silent substitution 774C/T in exon 6 altering the third base of the codon (both encoding an Asp) and non-synonymous substitution 894G/T alleles in exon 7 (probably not functional though since the resulting amino acid difference at position 298 Glu to Asp in the oxygenase domain is a conservative one)—are the most probably neutral genetic markers of a yet unidentified functional *NOS3* or neighbouring gene polymorphism in a linkage disequilibrium with the two SNPs.

Although a relatively large body of data supports (predominantly paracrine) action of the endothelin-1 in renal pathology [46, 47] and specifically in DN [48] authors are not aware of any study investigating the eventual association between *EDN1* locus and DN. The *EDN1* gene was studied for possible linkage with end-stage renal disease in non-diabetic African Americans (with negative results) [49] and for association with renal impairment in whites [50]. The latter study, comprising over 7,000 individuals, demonstrated significant association between one *EDN1* haplotype and diminished GFR and lower creatinine clearance. Association of the SNP 8002G/A represents, according to our knowledge, the first indication that *EDN1* gene might contribute to susceptibility to DN and will certainly prompt further and more detailed study of *EDN1* in connection with DN.

In summary, conducting a direct association study of 45 SNPs in several pathway-based candidate genes and DN in type 2 diabetes mellitus by means of multi-locus analysis we identified several SNPs in *EDN1*, *LTA*, *AGER* and *NOS3* gene jointly associated with susceptibility to DN in a yet unidentified mode of interaction. Certain haplotypes on chromosome 6 confer the susceptibility to DN. SNP *AGER* 2184 exhibited the strongest association with DN-marginal on its own and decisive among the set of associated markers. Contributions of others are likely to be minor. The set association did not bring substantial methodological improvement, more likely, both methods were complementary. However, such a statement based on a single application of the method is certainly preliminary. We consider chromosomal regions associated in this study worth further detailed analysis.

Acknowledgements Financial support: MSM 141100002 from the Ministry of Education, Youth and Physical Education of the Czech Republic (K. Kaňková), grants 303/02/D127 (K. Kaňková), 310/06/0827 (L. Izakovičová-Hollá) and 303/05/P523 (M. Beránek) from the Czech Science Foundation and MH44292 from the US National Institute of Mental Health (J. Ott).

Duality of interest None declared.

References

1. Wolf G (2004) New insights into the pathophysiology of diabetic nephropathy: from haemodynamics to molecular pathology. Eur J Clin Invest 34:785–796
2. Jacobsen PK (2005) Preventing end stage renal disease in diabetic patients-genetic aspect (part I). J Renin Angiotensin Aldosterone Syst 6:1–14
3. Boright AP, Paterson AD, Mirea L et al (2005) Genetic variation at the ACE gene is associated with persistent microalbuminuria and severe nephropathy in type 1 diabetes: the DCCT/EDIC Genetics Study. Diabetes 54:1238–1244
4. Krolewski AS, Warram JH, Christlieb AR, Busick EJ, Kahn CR (1985) The changing natural history of nephropathy in type I diabetes. Am J Med 78:785–794
5. The Diabetes Control and Complications Trial Research Group (1997) Clustering of long-term complications in families with diabetes in the diabetes control and complications trial. Diabetes 46:1829–1839
6. Borch-Johnsen K, Norgaard K, Hommel E et al (1992) Is diabetic nephropathy an inherited complication? Kidney Int 41:719–722
7. Canani LH, Gerchman F, Gross JL (1999) Familial clustering of diabetic nephropathy in Brazilian type 2 diabetic patients. Diabetes 48:909–913
8. Faronato PP, Maioli M, Tonolo G et al (1997) Clustering of albumin excretion rate abnormalities in Caucasian patients with NIDDM. The Italian NIDDM Nephropathy Study Group. Diabetologia 40:816–823
9. Quinn M, Angelico MC, Warram JH, Krolewski AS (1996) Familial factors determine the development of diabetic nephropathy in patients with IDDM. Diabetologia 39:940–945
10. Seaquist ER, Goetz FC, Rich S, Barbosa J (1989) Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. N Engl J Med 320:1161–1165
11. Freedman BI, Tuttle AB, Spray BJ (1995) Familial predisposition to nephropathy in African-Americans with non-insulin-dependent diabetes mellitus. Am J Kidney Dis 25:710–713
12. Pettitt DJ, Saad MF, Bennett PH, Nelson RG, Knowler WC (1990) Familial predisposition to renal disease in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia 33:438–443
13. Harjutsalo V, Katoh S, Sarti C, Tajima N, Tuomilehto J (2004) Population-based assessment of familial clustering of diabetic nephropathy in type 1 diabetes. Diabetes 53:2449–2454

14. Fogarty DG, Hanna LS, Wantman M, Warram JH, Krolewski AS, Rich SS (2000) Segregation analysis of urinary albumin excretion in families with type 2 diabetes. Diabetes 49:1057–1063
15. Imperatore G, Knowler WC, Pettitt DJ, Kobes S, Bennett PH, Hanson RL (2000) Segregation analysis of diabetic nephropathy in Pima Indians. Diabetes 49:1049–1056
16. Imperatore G, Hanson RL, Pettitt DJ, Kobes S, Bennett PH, Knowler WC (1998) Sib-pair linkage analysis for susceptibility genes for microvascular complications among Pima Indians with type 2 diabetes. Pima Diabetes Genes Group. Diabetes 47:821–830
17. Imperatore G, Knowler WC, Nelson RG, Hanson RL (2001) Genetics of diabetic nephropathy in the Pima Indians. Curr Diab Rep 1:275–281
18. Moczulski DK, Rogus JJ, Antonellis A, Warram JH, Krolewski AS (1998) Major susceptibility locus for nephropathy in type 1 diabetes on chromosome 3q: results of novel discordant sib-pair analysis. Diabetes 47:1164–1169
19. Vardarli I, Baier LJ, Hanson RL et al (2002) Gene for susceptibility to diabetic nephropathy in type 2 diabetes maps to 18q22.3–23. Kidney Int 62:2176–2183
20. Liu Y, Freedman BI (2005) Genetics of progressive renal failure in diabetic kidney disease. Kidney Int Suppl Dec 99:S94–S97
21. Hoh J, Ott J (2003) Mathematical multi-locus approaches to localizing complex human trait genes. Nat Rev Genet 4:701–709
22. Heidema AG, Boer JM, Nagelkerke N, Mariman EC, van der A DL, Feskens EJ (2006) The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. BMC Genet 7:23
23. Hoh J, Wille A, Ott J (2001) Trimming, weighting, and grouping SNPs in human case-control association studies. Genome Res 11:2115–2119
24. Ding K, Zhou K, He F, Shen Y (2003) LDA-a java-based linkage disequilibrium analyzer. Bioinformatics 19:2147–2148
25. Abecasis GR, Cookson WO (2000) GOLD-graphical overview of linkage disequilibrium. Bioinformatics 16:182–183
26. Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311–322
27. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc [Ser B] 57:289–300
28. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989
29. Kankova K, Stejskalova A, Hertlova M, Znojil V (2005) Haplotype analysis of the RAGE gene: identification of a haplotype marker for diabetic nephropathy in type 2 diabetes mellitus. Nephrol Dial Transplant 20:1093–1102
30. Yamamoto Y, Doi T, Kato I et al (2005) Receptor for advanced glycation end products is a promising target of diabetic nephropathy. Ann N Y Acad Sci 1043:562–566
31. Ozaki K, Ohnishi Y, Iida A et al (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654
32. Messer G, Spengler U, Jung MC et al (1991) Polymorphic structure of the tumor necrosis factor (TNF) locus: an NcoI polymorphism in the first intron of the human TNF-beta gene correlates with a variant amino acid in position 26 and a reduced level of TNF-beta production. J Exp Med 173:209–219
33. Abraham LJ, French MA, Dawkins RL (1993) Polymorphic MHC ancestral haplotypes affect the activity of tumour necrosis factor-alpha. Clin Exp Immunol 92:14–18
34. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP (2003) In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. Nat Genet 33:469–475
35. Navarro JF, Mora C, Maca M, Garca J (2003) Inflammatory parameters are independently associated with urinary albumin in type 2 diabetes mellitus. Am J Kidney Dis 42:53–61
36. Moriwaki Y, Yamamoto T, Shibutani Y et al (2003) Elevated levels of interleukin-18 and tumor necrosis factor-alpha in serum of patients with type 2 diabetes mellitus: relationship with diabetic nephropathy. Metabolism 52:605–608
37. Kalantarinia K, Awad AS, Siragy HM (2003) Urinary and renal interstitial concentrations of TNF-alpha increase prior to the rise in albuminuria in diabetic rats. Kidney Int 64:1208–1213
38. Mukherjee TK, Mukhopadhyay S, Hoidal JR (2005) The role of reactive oxygen species in TNFalpha-dependent expression of the receptor for advanced glycation end products in human umbilical vein endothelial cells. Biochim Biophys Acta 1744:213–223
39. Prabhakar SS (2004) Role of nitric oxide in diabetic nephropathy. Semin Nephrol 24:333–344
40. Komers R, Anderson S (2003) Paradoxes of nitric oxide in the diabetic kidney. Am J Physiol Renal Physiol 284:F1121–F1137
41. Zanchi A, Moczulski DK, Hanna LS, Wantman M, Warram JH, Krolewski AS (2000) Risk of advanced diabetic nephropathy in type 1 diabetes is associated with endothelial nitric oxide synthase gene polymorphism. Kidney Int 57:405–413
42. Ksiazek P, Wojewoda P, Muc K, Buraczynska M (2003) Endothelial nitric oxide synthase gene intron 4 polymorphism in type 2 diabetes mellitus. Mol Diagn 7:119–123
43. Asakimori Y, Yorioka N, Taniguchi Y et al (2002) T(−786)→C polymorphism of the endothelial nitric oxide synthase gene influences the progression of renal disease. Nephron 91:747–751
44. Noiri E, Satoh H, Taguchi J et al (2002) Association of eNOS Glu298Asp polymorphism with end-stage renal disease. Hypertension 40:535–540
45. Rippin JD, Patel A, Belyaev ND, Gill GV, Barnett AH, Bain SC (2003) Nitric oxide synthase gene polymorphisms and diabetic nephropathy. Diabetologia 46:426–428
46. Kohan DE (1997) Endothelins in the normal and diseased kidney. Am J Kidney Dis 29:2–26
47. Bruzzi I, Remuzzi G, Benigni A (1997) Endothelin: a mediator of renal disease progression. J Nephrol 10:179–183
48. Jandeleit-Dahm K, Allen TJ, Youssef S, Gilbert RE, Cooper ME (2000) Is there a role for endothelin antagonists in diabetic renal disease? Diabetes Obes Metab 2:15–24
49. 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
50. 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