Original Contribution

Identifying Genetic Susceptibilities to Diabetes-related Complications among Individuals at Low Risk of Complications: An Application of Tree-Structured Survival Analysis

Tina Costacou1, Yuefang Chang2, Robert E. Ferrell3, and Trevor J. Orchard1

1 Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA.
2 Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA.
3 Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Received for publication September 14, 2005; accepted for publication April 10, 2006.

The authors hypothesized that genetic predisposition to diabetes complications would be more evident among low-risk individuals and aimed to identify genes related to developing complications (confirmed distal symmetric polyneuropathy, overt nephropathy, or coronary artery disease) in low-risk groups. Participants in the Pittsburgh, Pennsylvania, Epidemiology of Diabetes Complications Study of childhood-onset type 1 diabetes, first seen in 1986–1988 (mean age, 28 years; diabetes duration, 19 years), were reexamined biennially for 10 years. For each complication, subgroups with the lowest disease risk were identified by using tree-structured survival analysis, and 15 candidate genes were compared between subjects with and without complications. In the group with the lowest incidence of confirmed distal symmetric polyneuropathy (n = 123), confirmed distal symmetric polyneuropathy risk increased fivefold for those with the eNOS GG genotype (p < 0.05). In the group with the lowest risk of overt nephropathy (n = 340), the ACE D polymorphism increased overt nephropathy risk twofold (p = 0.05), whereas a protective effect was observed for the LIPC CC genotype (p < 0.05). In the group with the lowest incidence of coronary artery disease (n = 331), the MTHFR CC genotype increased coronary artery disease risk threefold (p < 0.05). Tree-structured survival analysis may help identify genetic predispositions among individuals who, despite low risk, develop diabetes-related complications.

diabetes complications; diabetes mellitus; genetic predisposition to disease; statistics; survival analysis

Abbreviations: ACE, angiotensin-converting enzyme; eNOS, endothelial nitric oxide synthase; HDL, high density lipoprotein; LIPC, hepatic lipase; MTHFR, methylenetetrahydrofolate reductase; TSSA, tree-structured survival analysis.

Despite strong associations between glycemic control and microvascular complications in type 1 diabetes (1–4), there is still considerable variability in the risk of complications, suggesting that certain patient subgroups are particularly prone or resistant. Recent studies have clearly shown familial clustering of complications, in particular renal disease, providing further support for a genetic predisposition (5–7). However, as reviewed by Rich et al. (8), no consensus has been reached, at least as far as human lymphocyte antigen–associated genetic susceptibilities are concerned. We propose that the genetic component may be particularly important in subgroups at either end of the risk spectrum, as determined by known risk factors, because most individuals will develop complications on the basis of standard risk factors (e.g., glycemia, blood pressure, lipids). We also propose that studies in these extreme subgroups may reveal genetic markers of susceptibility that extend beyond the sole influence of the main risk factors. If this hypothesis holds true, then it would be easier to identify individuals with a genetic susceptibility to diabetes-related complications among those with a better...
overall risk profile (or, alternatively, to identify protective genes in those with a high-risk profile). This general approach we have previously found helpful; for example, we were able to identify three genetic markers for the development of overt nephropathy, by comparing those at high and low risk, that were not evident in the study population as a whole (9).

Our objective was to identify those individuals with a genetic susceptibility to three diabetes-related complications, namely, confirmed distal symmetric polyneuropathy, overt nephropathy, and coronary artery disease, among patients at seemingly low risk. To do so, we used the 10-year follow-up data from the Pittsburgh Epidemiology of Diabetes Complications Study, a prospective cohort of childhood-onset type 1 diabetes. Survival data have traditionally been analyzed by using Cox proportional hazards modeling (10). Newer regression techniques, such as tree-structured analysis, also known as recursive partitioning, provide one potential approach to identifying subgroups of patients characterized by the presence or absence of specific risk factors and who differ in disease risk (11–13). Tree-based classification does not require the assumption that the ratios of the hazards for the groups being compared are constant. We used this classification approach with an extension to survival data (tree-structured survival analysis (TSSA)) to facilitate the search for genetic factors whose influence may be apparent in only those at high or low overall risk. These analyses, which focus on the effect of genetic factors, are primarily considered hypothesis generating and need to be replicated despite the plausibility of the candidate genes studied.

MATERIALS AND METHODS

The population selected for study was a historical, prospective cohort based on incident cases of childhood-onset (children <17 years of age) type 1 diabetes diagnosed or seen within 1 year of diagnosis at Children’s Hospital of Pittsburgh, Pennsylvania, between 1950 and 1980 (14, 15). Although this population is clinic based, it has been shown to be representative of the type 1 diabetes population of Allegheny County, Pennsylvania (16). This cohort was first examined for the Pittsburgh Epidemiology of Diabetes Complications study between 1986 and 1988, when mean age was 28 years and diabetes duration was 19 years. Subjects were subsequently reexamined biennially for 10 years. Prior to each of the scheduled clinic visits, participants were sent questionnaires concerning demographic, health care, self-care, and medical history information. The University of Pittsburgh Institutional Review Board approved the study protocol.

Definition of complications and risk factors

Confirmed distal symmetric polyneuropathy was determined by a study-trained physician using the Diabetes Control and Complications Trial clinical examination protocol (17) and was diagnosed by the presence of at least two of the following states: symptoms and/or signs consistent with distal symmetric polyneuropathy and absent (or decreased) tendon reflexes, all in the absence of other known etiologies, in addition to the presence of a vibratory threshold above the age-specific normal range using the Vibratron II tester (Physitemp Instruments Inc., Clifton, New Jersey) and a forced-choice methodology (18, 19). Of 658 participants, 64 (9.7 percent) who did not return for a follow-up examination and 187 with prevalent confirmed distal symmetric polyneuropathy were excluded from analyses. Of the remaining 407, 370 had complete information on covariates examined and were thus included in analyses for the prediction of distal symmetric polyneuropathy incidence.

Overt nephropathy was defined as an albumin excretion rate of >200 mg/minute on at least two of three timed urine specimens, or, in the absence of urine, a serum creatinine level of >2 mg/dl, or renal failure or renal transplant. Follow-up overt nephropathy status could not be determined for 58 (8.8 percent) participants who did not return for a clinical examination. Another 171 were diagnosed with overt nephropathy at study entry and were thus excluded from analyses for the prediction of incident renal disease, along with 37 whose information on covariates examined was not complete.

Coronary artery disease was determined by study physician–diagnosed angina or myocardial infarction confirmed by Q waves on electrocardiogram or hospital records (Minnesota codes 1.1 or 1.2), or angiographic stenosis ≥50 percent, coronary artery bypass surgery, angioplasty, or ischemic electrocardiogram changes (Minnesota codes 1.3, 4.3, 4.2, 5.1, 5.3, 7.1) during the follow-up period of the study. Follow-up coronary artery disease status could not be determined for three participants (0.5 percent) who did not return for a follow-up examination. Fifty-two prevalent coronary artery disease cases as well as 64 persons lacking information on covariates examined were excluded from the analyses for the prediction of coronary artery disease incidence.

The proportion of participants included from analyses because of failure to return for a follow-up examination, and thus inability to ascertain disease status, varies by complication. This variation relates to the definition of each complication and the ability to determine disease status without a subject actually returning for a clinical examination. Thus, if a participant free of complications at baseline was unable to return for a follow-up visit but provided survey information indicating the presence of a specific condition (e.g., heart attack), that condition would be diagnosed, whereas the presence of other complications (e.g., overt nephropathy, which requires urine samples) could not be identified.

Blood pressure was measured with a random zero sphygmomanometer, according to the Hypertension Detection and Follow-up Program protocol, after a 5-minute rest (20). Hypertension was defined as blood pressure ≥140/90 mmHg or use of antihypertensive medication. Stable glycosylated hemoglobin was measured by ion exchange chromatography (Isolab, Akron, Ohio) and subsequently by automated high-performance liquid chromatography (Diamat, BioRad, Hercules, California). Readings obtained with the two methods are almost identical (r = 0.95). High density lipoprotein (HDL) cholesterol was determined by a precipitation technique (heparin and manganese chloride) with a modification (21) of the Lipid Research Clinics method (22). Cholesterol and triglycerides were measured enzymatically (23, 24).
Low density lipoprotein cholesterol levels were calculated from measurements of the levels of total cholesterol, triglycerides, and HDL cholesterol by using the Friedewald equation (25). Smoking status (ever/never) was obtained via self-report.

Glucose disposal rate (insulin sensitivity) was estimated by using a regression equation 
\[ (24.31 - (12.22 \times \text{waist-to-hip ratio}) - (3.29 \times \text{hypertension}) - (0.57 \times \text{glycosylated hemoglobin})) \]
declared from hyperinsulinemic euglycemic clamp studies on 24 study subjects chosen to represent the full spectrum of insulin resistance as identified by insulin resistance risk factors (26). White blood cell count was obtained by using a counter S-plus IV (Coulter Electronics, Inc., Hialeah, Florida) and fibrinogen by using a biuret colorimetric procedure and a clotting method.

For the 1999–2004 Pittsburgh Epidemiology of Diabetes Complications Study grant period, 15 genetic polymorphisms were proposed for study. All had suspected potential roles in the pathogenesis of complications through linkage to blood pressure, lipoprotein metabolism, glucose metabolism, glycocalyxation, or inflammation. Of these 15 polymorphisms, associations were found for four with the three diabetes complications examined (genetic polymorphisms not presented in the paper are included in appendix table 1). High-molecular-weight genomic DNA was isolated from ethylenediamine-tetraacetic acid anticoagulant whole blood by the method of Miller et al. (27). The angiotensin-converting enzyme (ACE) insertion/deletion polymorphism and the methylenetetrahydrofolate reductase (MTHFR) polymorphism were determined by the methods of Tiret et al. (28), Frosst et al. (29), Nakayama et al. (30), Hibi et al. (31), and Couture et al. (32). Genotypes were assigned by direct comparison with controls of known genotype.

**Statistical analyses**

The association of candidate genes with type 1 diabetes complications was evaluated by comparing the genotypes in the lowest risk groups generated by TSSA (12, 13). TSSA, a nonparametric statistical method, extends tree-based methods such as the classification and regression tree (11). The main objective of a tree-based method is to classify subjects into subgroups according to their characteristics and outcome so that, within each subgroup, the risk factors are similar but, between subgroups, the outcomes are distinct. Briefly, a tree model is built starting with the complete sample of subjects as one group, the root node. Subjects are then partitioned to two subgroups (or nodes) based on a predictor variable that separates the subgroups most in the outcome. The process is then repeated in each node, with a predictor being selected each time until no further dividing is worthwhile. TSSA extends the tree construction methodology to survival data taking into consideration time to event, so that within-group homogeneity in risk factors and between-group separation in survival experience are optimized.

TSSAs were conducted for the prediction of three diabetes-related complications: confirmed distal symmetric polyneuropathy, overt nephropathy, and coronary artery disease. Baseline risk factors used in the analyses for all three outcomes included diabetes duration, age at diabetes onset, serum creatinine, fibrinogen, having ever smoked, total cholesterol, triglycerides, systolic blood pressure, white blood cell count, and estimated glucose disposal rate, based on earlier analyses (9, 33, 34). Behavioral and socioeconomic risk factors were not considered. Subjects who were free from the complication at baseline and for whom all the risk factors were available were included in the TSSA (in total, 567 participants free at least one complication at baseline were included in these analyses). Once the subgroups were identified from TSSA for each complication, the relationship between the candidate gene and complication was examined in the lowest risk group. DNA was determined in 443 (78.1 percent) participants; however, two genotyping failures occurred for the eNOS, three for the ACE, 15 for the LIPC, and 28 for the MTHFR genotypes. Of the 15 genes considered, associations were found for four polymorphisms; results for the latter are reported here. Clearly, this candidate gene approach to studying three different complications inevitably leads to multiple comparisons. Whereas we present results on the basis of \( p < 0.05 \) on the grounds that this is a hypothesis-generating analysis only, a corrected Bonferroni level of significance would be \(<0.003\) given 15 genes. None of the four polymorphisms reported reached this level of significance.

In the current study, the TSSA was conducted by using S-Plus programs written by Dr. Mark Segal (Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California). Descriptive statistical analyses (Student’s \( t \) tests and chi-square tests for univariate associations between baseline risk factors and complication status during 10 years of follow-up) as well as estimates of risk were generated by using SAS version 8.2 software (SAS Institute, Inc., Cary, North Carolina). All reported \( p \) values are two sided. Because hazard ratios were calculated using groups of participants already characterized by their baseline risk attributes, adjustment was performed for diabetes duration only.

**RESULTS**

Baseline characteristics of participants by the incidence of complications during 10 years of follow-up are shown in table 1, whereas duration-adjusted hazard ratios for each of the four polymorphisms are shown in table 2. With the exception of protection against renal disease incidence among those with the LIPC CC genotype, no significant genetic associations were observed in the overall cohort by diabetes complication.

Figure 1 presents the TSSA results for confirmed distal symmetric polyneuropathy. As shown, 370 individuals were free of this complication at baseline; 114 (30.8 percent) of them developed it during the follow-up period. Starting with the group of 370, the tree first classified subjects by splitting on estimated glucose disposal rate, then age at diabetes onset, and then duration of diabetes (for those whose age at diabetes onset was \(<8.4\) years), or white blood cell count for individuals whose age at diabetes onset was \(>8.4\) years. The lowest incidence of confirmed distal symmetric polyneuropathy is observed in the group with the highest glucose disposal rate and the highest white blood cell count.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Confirmed distal symmetric polyneuropathy incidence‡</th>
<th>Overt nephropathy incidence‡</th>
<th>Coronary artery disease incidence‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean follow-up time (years)</td>
<td>No (n = 256)</td>
<td>8.7 (2.7)</td>
<td>6.3 (3.1)*</td>
</tr>
<tr>
<td>Age (years)</td>
<td>No (n = 343)</td>
<td>23.7 (6.9)</td>
<td>28.9 (6.7)*</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>No (n = 114)</td>
<td>7.9 (4.1)</td>
<td>9.4 (3.9)*</td>
</tr>
<tr>
<td>Diabetes duration (years)</td>
<td>No (n = 49)</td>
<td>15.8 (6.3)</td>
<td>19.6 (7.0)*</td>
</tr>
<tr>
<td>Male sex (%, no.)</td>
<td>No (n = 22)</td>
<td>49.6 (127)</td>
<td>49.1 (56)</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>No (n = 34)</td>
<td>23.3 (3.5)</td>
<td>23.9 (2.7)</td>
</tr>
<tr>
<td>Waist-to-hip ratio</td>
<td>No (n = 99)</td>
<td>0.8 (0.07)</td>
<td>0.82 (0.07)</td>
</tr>
<tr>
<td>Glycosylated hemoglobin (%)</td>
<td>No (n = 19)</td>
<td>10.1 (1.8)</td>
<td>10.6 (1.9)*</td>
</tr>
<tr>
<td>Estimated glucose disposal rate (mg/kg per minute)</td>
<td>No (n = 20)</td>
<td>8.5 (1.4)</td>
<td>7.6 (2.0)*</td>
</tr>
<tr>
<td>Ever smoked (%), no.)</td>
<td>No (n = 20)</td>
<td>27.0 (69)</td>
<td>41.2 (47)*</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>No (n = 20)</td>
<td>109.4 (11.5)</td>
<td>113.3 (15.6)*</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg)</td>
<td>No (n = 20)</td>
<td>70.6 (9.6)</td>
<td>73.5 (10.6)*</td>
</tr>
<tr>
<td>Hypertension (%), no.)</td>
<td>No (n = 20)</td>
<td>5.1 (13)</td>
<td>19.3 (22)*</td>
</tr>
<tr>
<td>HDL§ cholesterol (mmol/liter)</td>
<td>No (n = 20)</td>
<td>55.5 (12.3)</td>
<td>55.3 (12.6)</td>
</tr>
<tr>
<td>LDL§ cholesterol (mmol/liter)</td>
<td>No (n = 20)</td>
<td>106.0 (28.3)</td>
<td>120.5 (36.5)*</td>
</tr>
<tr>
<td>Non-HDL cholesterol (mmol/liter)</td>
<td>No (n = 20)</td>
<td>125.0 (35.6)</td>
<td>141.1 (41.4)*</td>
</tr>
<tr>
<td>Total cholesterol (mmol/liter)</td>
<td>No (n = 20)</td>
<td>180.4 (36.0)</td>
<td>196.5 (41.6)*</td>
</tr>
<tr>
<td>Triglycerides (mmol/liter)¶</td>
<td>No (n = 20)</td>
<td>92.1 (69.5)</td>
<td>105.1 (88.7)*</td>
</tr>
<tr>
<td>Fibrinogen (mmol/liter)§</td>
<td>No (n = 20)</td>
<td>276.3 (78.7)</td>
<td>282.2 (92.0)</td>
</tr>
<tr>
<td>White blood cell count × 10⁹/mm³</td>
<td>No (n = 20)</td>
<td>6.1 (1.6)</td>
<td>6.6 (1.8)*</td>
</tr>
<tr>
<td>Serum creatinine (mg/dl)¶</td>
<td>No (n = 20)</td>
<td>0.9 (0.4)</td>
<td>1.0 (0.9)</td>
</tr>
<tr>
<td>Overt nephropathy (%), no.)</td>
<td>No (n = 20)</td>
<td>11.7 (30)</td>
<td>27.2 (31)*</td>
</tr>
<tr>
<td>eNOS Glu 298 Asp (E298D)§</td>
<td>No (n = 20)</td>
<td>44.6 (99)</td>
<td>55.0 (61)</td>
</tr>
<tr>
<td>ACE I/D§</td>
<td>No (n = 20)</td>
<td>25.0 (56)</td>
<td>25.5 (28)</td>
</tr>
<tr>
<td>LIPC C(-514)T§</td>
<td>No (n = 20)</td>
<td>60.9 (131)</td>
<td>63.2 (67)</td>
</tr>
<tr>
<td>MTHFR C677T§</td>
<td>No (n = 20)</td>
<td>42.3 (88)</td>
<td>44.8 (47)</td>
</tr>
</tbody>
</table>

* p < 0.05.
† Values are presented as mean (standard deviation) for continuous variables and percentage (number) for categorical variables.
‡ DNA was determined in 443 (78.1%) of the 567 participants included in these analyses; 335 (111 incident cases) for confirmed distal symmetric polyneuropathy, 344 (44 incident cases) for overt nephropathy, and 421 (81 incident cases) for coronary artery disease.
§ HDL, high density lipoprotein; LDL, low density lipoprotein; eNOS Glu 298 Asp (E298D), endothelial nitric oxide synthase gene Glu 298 Asp polymorphism; ACE I/D, angiotensin-converting enzyme insertion/deletion polymorphism; LIPC C(-514)T, hepatic lipase gene C(-514)T polymorphism; MTHFR C677T, methylenetetrahydrofolate reductase gene C677T polymorphism.
¶ Logarithmically transformed before statistical testing.

Genetic Susceptibility to Diabetes Complications

polyneuropathy was observed in the second group; that is, among patients whose estimated glucose disposal rate was more than 6 mg/kg per minute, whose age at diabetes onset was ≤8.4 years, and whose duration of diabetes was less than approximately 24 years. In this lowest-risk group, a nominally significantly greater risk of confirmed distal symmetric polyneuropathy was observed for the eNOS GG genotype, with a diabetes duration–adjusted hazard ratio of 4.86 (95 percent confidence interval: 1.04, 22.72; p < 0.05) (table 3).

A total of 392 subjects were free of overt nephropathy at baseline (figure 2), and 49 (12.5 percent) of them developed this complication during the follow-up period. Once again, the tree first classified subjects by splitting on estimated glucose disposal rate, then age at diabetes onset, and, for

Am J Epidemiol 2006;164:862–872
those whose age at diabetes onset was greater than 7 years, white blood cell count. The first group, that is, subjects whose estimated glucose disposal rate was >6.58, had the lowest incidence of overt nephropathy (8.8 percent), where there was some indication that the D allele of the ACE polymorphism was associated with increased risk of overt nephropathy incidence (table 3; diabetes duration–adjusted hazard ratio = 2.11, 95 percent confidence interval: 1.00, 3.22).

### TABLE 2. Duration-adjusted Cox proportional hazards ratios for predicting the incidence of complications during 10 years of follow-up (1986–1988 to 1996–1998), The Pittsburgh Epidemiology of Diabetes Complications Study

<table>
<thead>
<tr>
<th>Gene variant*</th>
<th>Confirmed distal symmetric polyneuropathy incidence</th>
<th>Overt nephropathy incidence</th>
<th>Coronary artery disease incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard ratio 95% confidence interval</td>
<td>Hazard ratio 95% confidence interval</td>
<td>Hazard ratio 95% confidence interval</td>
</tr>
<tr>
<td>eNOS Glu 298 Asp (E298D)</td>
<td>GG vs. TG/TT 1.32 0.90, 1.92</td>
<td>0.75 0.41, 1.38</td>
<td>1.14 0.74, 1.78</td>
</tr>
<tr>
<td>ACE (I/D)</td>
<td>D vs. IIID 0.92 0.60, 1.42</td>
<td>1.38 0.74, 2.57</td>
<td>1.11 0.66, 1.84</td>
</tr>
<tr>
<td>LIPC C(-514)T</td>
<td>CC vs. CT/TT 1.17 0.79, 1.75</td>
<td>0.53 0.29, 0.97</td>
<td>0.86 0.55, 1.36</td>
</tr>
<tr>
<td>MTHFR C677T</td>
<td>CC vs. CT/TT 0.82 0.55, 1.21</td>
<td>1.35 0.74, 2.48</td>
<td>0.83 0.52, 1.33</td>
</tr>
</tbody>
</table>

* ACE I/D, angiotensin-converting enzyme insertion/deletion polymorphism; eNOS Glu 298 Asp (E298D), endothelial nitric oxide synthase gene Glu 298 Asp polymorphism; MTHFR C677T, methylenetetrahydrofolate reductase gene C677T polymorphism; LIPC C(-514)T, hepatic lipase gene C(-514)T polymorphism.
4.48; \( p = 0.05 \). In addition, the protective effect of the \text{LIPC} CC \text{ genotype} (table 3; diabetes duration–adjusted hazard ratio = 0.39, 95 percent confidence interval: 0.18, 0.88; \( p < 0.05 \)) remained.

Ninety-nine of 539 subjects free of coronary artery disease at the baseline examination (figure 3) developed this complication during the 10-year follow-up. This time, diabetes duration emerged as the most consequential predictor, followed by white blood cell count and either serum creatinine (among those whose white blood cell count was \( < 8 \times 10^3/\text{mm}^2 \)) or HDL cholesterol (in individuals with a high white blood cell count). Patients in the first group, that is, whose diabetes duration was \( \leq 21.19 \) years, exhibited the lowest incidence of coronary artery disease (6.3

### TABLE 3. Duration-adjusted Cox proportional hazards ratios for predicting the incidence of complications in the lowest-risk complication groups, as identified by tree-structured survival analysis, The Pittsburgh Epidemiology of Diabetes Complications Study (10 years of follow-up: 1986–1988 to 1996–1998)

<table>
<thead>
<tr>
<th>Gene variant*</th>
<th>Confirmed distal symmetric polyneuropathy incidence</th>
<th>Overt nephropathy incidence</th>
<th>Coronary artery disease incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>eNOS Glu 298 Asp (E298D)</td>
<td>Hazard ratio 4.86, 95% confidence interval 1.04–22.72</td>
<td>Hazard ratio 0.76, 95% confidence interval 0.36–1.63</td>
<td>Hazard ratio 0.79, 95% confidence interval 0.31–2.01</td>
</tr>
<tr>
<td>ACE (I/D)</td>
<td>Hazard ratio 2.14, 95% confidence interval 0.64–7.16</td>
<td>Hazard ratio 2.11, 95% confidence interval 1.00–4.48</td>
<td>Hazard ratio 1.53, 95% confidence interval 0.59–3.96</td>
</tr>
<tr>
<td>LIPC C(-514)T</td>
<td>CC vs. CT/TT Hazard ratio 0.76, 95% confidence interval 0.21–2.78</td>
<td>Hazard ratio 0.39, 95% confidence interval 0.18–0.88</td>
<td>Hazard ratio 1.09, 95% confidence interval 0.40–2.96</td>
</tr>
<tr>
<td>MTHFR C677T</td>
<td>CC vs. CT/TT Hazard ratio 1.91, 95% confidence interval 0.50–7.27</td>
<td>Hazard ratio 1.62, 95% confidence interval 0.75–3.49</td>
<td>Hazard ratio 2.95, 95% confidence interval 1.02–8.50</td>
</tr>
</tbody>
</table>

* eNOS Glu 298 Asp (E298D), endothelial nitric oxide synthase gene Glu 298 Asp polymorphism; ACE I/D, angiotensin-converting enzyme insertion/deletion polymorphism; LIPC C(-514)T, hepatic lipase gene C(-514)T polymorphism; MTHFR C677T, methylenetetrahydrofolate reductase gene C677T polymorphism.

FIGURE 2. Tree-structured survival analysis classification for overt nephropathy (ON), The Pittsburgh Epidemiology of Diabetes Complications Study, 10-year follow-up (1986–1988 to 1996–1998). Age at onset (years); eGDR, estimated glucose disposal rate (mg/kg per minute); WBC, white blood cell count \( \times 10^3/\text{mm}^2 \).
percent); thus, we chose this group to look for significant genetic associations. The CC genotype of the MTHFR C677T gene polymorphism was associated with an approximately threefold risk of coronary artery disease (table 3; diabetes duration–adjusted hazard ratio = 2.95, 95 percent confidence interval: 1.02, 8.50; \( p < 0.05 \)).

**DISCUSSION**

TSSA is an exploratory statistical method of categorizing individuals into meaningful prognostic subsets. Being non-parametric, this technique requires fewer modeling assumptions. Moreover, because it is designed to categorize persons into risk groups based on survival, it allows for further comparisons within these selected strata. However, tree methods are not without limitations. An important drawback is the inability to assess a covariate’s importance because the method relies on post hoc tests; highly correlated covariates may also affect factor selection in recursive partitioning (12).

In this study, we used TSSA to identify individuals at seemingly low risk of developing diabetes-related complications and then looked for significant genetic associations that might help explain case development within these lowest-risk subgroups. Indeed, genetic risk factors that were generally not apparent in the total study population emerged within the lowest-risk subgroups.

Interestingly, with the exception of diabetes duration, risk factors chosen for confirmed distal symmetric polyneuropathy (estimated glucose disposal rate, age at diabetes onset, white blood cell count) also seemed to predict incidence of overt nephropathy, although slightly different cutoffs were used each time. This finding is consistent with our previous hypothesis that insulin resistance underlies overt nephropathy susceptibility (9). However, although insulin sensitivity alone was adequate to identify the lowest incidence rate of overt nephropathy, for confirmed distal symmetric polyneuropathy, younger age at diabetes onset and shorter diabetes duration were also required. Duration of diabetes, white blood cell count, serum creatinine, and HDL cholesterol concentration were selected as the best predictors of coronary artery disease incidence, and the lowest incidence rate was observed in the subgroup with a shorter duration of diabetes.

The prognostic value of the factors selected by TSSAs has been previously demonstrated in other populations as well. Cardiovascular risk factors (35), insulin resistance (36), and inflammatory markers (36, 37) have all been implicated in the development of diabetic neuropathy. Reduced insulin sensitivity has been reported among individuals with diabetes and microalbuminuria (38, 39), whereas elevated white blood cell count has been shown to predict the development of renal disease in diabetics (40, 41). Decreased HDL cholesterol concentration (42, 43) and increased levels of inflammatory markers (37, 42) are recognized predictors of cardiovascular disease incidence. Furthermore, a variety of studies have shown that renal dysfunction increases the risk of cardiovascular disease (43–45).

*Am J Epidemiol* 2006;164:862–872
Although the 10-year incidence of confirmed distal symmetric polyneuropathy was approximately 31 percent in the study cohort, the rate was only 8.9 percent in the lowest-risk group (group 2), as assessed by TSSA. In this subgroup, the GG genotype of the eNOS Glu 298 Asp polymorphism was significantly associated with an almost fivefold risk of confirmed distal symmetric polyneuropathy.

Nitric oxide synthases produce nitric oxide, known to play an important role in regulation of the cardiovascular, central, and peripheral nervous systems (46). The G to T polymorphism at position 894 of the eNOS gene is the only polymorphism identified to date that changes the eNOS protein sequence, suggesting that genetic variation at this site may alter eNOS activity or regulation (47). Synthesis of nitric oxide is thought to be deficient in diabetes (48, 49), and chronic treatment of nondiabetic rats with a nitric oxide synthase inhibitor has been associated with nerve conduction velocity deficits similar to those observed in diabetes (50, 51). In experimental studies, nitric oxide has also been associated with neuropathy (52, 53).

The insertion/deletion polymorphism in intron 16 of the ACE gene determines most of the interindividual variation in ACE levels (54). Despite some disagreement (55–57), this polymorphism has been associated with diabetic nephropathy (9, 58–62) as well as insulin resistance (63–65). An association similar to the one observed in this report for the D allele has been previously described by our group; we used an alternative approach to compare low- and high-risk groups (i.e., comparing those who develop overt nephropathy early (short duration and before other complications) with those developing this complication late (long duration and after other complications)) (9). Furthermore, a recently published meta-analysis suggests that treatment with ACE inhibitors may significantly reduce albuminuria (66), providing further support for a role of the ACE gene in renal disease.

The human LIPC polymorphism that we examined is one of four identified polymorphisms of the human LIPC gene (67). Allelic variation in the genes encoding LIPC accounts for 25 percent of total interindividual variation in plasma HDL cholesterol levels (68). In a report from the Bogalusa Heart Study, the -514T variant of the hepatic lipase gene was associated with higher levels of HDL cholesterol longitudinally since childhood (69). The same variant was also associated with significant variations in the lipoprotein profile of male and female participants of the Framingham Offspring Study, where increases were noted in large HDL fractions among carriers of the -514 allele (32). Furthermore, LIPC appears to relate to markers of cardiovascular disease in type 1 diabetes (70). In the present analyses, a protective effect of the CC genotype of the LIPC C(-514)T polymorphism was observed for those at the lowest risk of overt nephropathy. We (9), among other groups (71–73), have shown that serum lipid levels are associated with overt nephropathy incidence.

With regard to coronary artery disease, the incidence rate in the lowest-risk group, as derived from TSSA, was 6.3 percent compared with an overall rate of 18.4 percent. An increased risk associated with the CC genotype in the MTHFR C677T gene polymorphism was noted. Recently, a number of studies (74–77) have supported a role of the C677T variant in the pathophysiology of cardiovascular disease. Such an association seems likely because of its influence on homocysteine levels (78), although, in this cohort, homocysteine levels were not related to either the MTHFR genotype or coronary artery disease incidence (data not shown).

In conclusion, genetic risk factors may be better identified in subgroups characterized by the absence of major traditional risk factors. It is important to note that, with the exception of the LIPC CC genotype, none of the other polymorphisms were associated with complication outcome in the total study population; associations with gene polymorphisms emerged in the lowest-risk subgroups only. This approach provides the potential for further insight into genetic variations and their role in the development of complications. It may be helpful in formulating prevention strategies by focusing attention on the newly defined high-risk population subgroups as well as addressing standard risk factors. However, the present genetic associations were derived essentially from hypothesis-generating analyses and need to be evaluated and replicated in other populations before acceptance.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant DK34818.

The authors thank the Pittsburgh Epidemiology of Diabetes Complications Study staff.

Conflict of interest: none declared.

REFERENCES


### APPENDIX TABLE 1. Other candidate genes for the study of type 1 diabetes complications in The Pittsburgh Epidemiology of Diabetes Complications Study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Rationale</th>
</tr>
</thead>
</table>