ONLINE APPENDIX—Supplemental Methods

Description of the Go-DARTS Study. The DARTS (Diabetes Audit and Research in Tayside Scotland) Study was originally created to identify all patients with diabetes in a community using electronic record linkage of multiple data sources and to compare this method of case ascertainment with registers of diabetic patients derived from primary care (Morris et al., Br Med J 1997;315:524-528). Electronic capture-recapture linkage of records included data on all patients attending hospital diabetes clinics, all encashed prescriptions for diabetes related drugs and monitoring equipment, all patients discharged from hospital, patients attending a mobile unit for eye screening, and results for glycated haemoglobin and plasma glucose concentrations from the regional biochemistry database. Diabetes registers from primary care were from a random sample of eight Tayside general practices. A detailed manual study of relevant records for the 35,144 patients registered with these eight general practices allowed for validation of the case ascertainment. The setting of the study was the Tayside region of Scotland, population 391,274 on 1 January 1996. The main outcomes were: prevalence of diabetes; population of patients identified by different data sources; sensitivity and positive predictive value of ascertainment methods. Electronic record linkage identified 7,596 diabetic patients, giving a prevalence of known diabetes of 1.94% (0.21% insulin dependent diabetes, 1.73% non-insulin dependent): 63% of patients had attended hospital diabetes clinics, 68% had encashed diabetes related prescriptions, 72% had attended the mobile eye screening unit, and 48% had biochemical results diagnostic of diabetes. A further 701 patients had isolated hyperglycaemia (plasma glucose >11.1 mmol/l) but were not considered diabetic by general practitioners. Validation against the eight general practices (636 diabetic patients) showed electronic linkage to have a sensitivity of 0.96 and a positive predictive value of 0.95 for ascertainment of known diabetes. General practice lists had a sensitivity of 0.91 and a positive predictive value of 0.98. In conclusion, it was demonstrated that electronic record linkage was more sensitive than general practice registers in identifying diabetic subjects and identified an additional 0.18% of the population with a history of hyperglycaemia who might warrant screening for undiagnosed diabetes. DARTS databases are continuously updated from various sources, including clinic encounters, hospital biochemistry reports, and hospital discharge data. All clinical data are recorded according to a standard dataset, and all case records are validated by a team of research nurses who create a "cradle to the grave" electronic record. This automated electronic follow-up is manually validated through a continuous cycle of review by dedicated study clinicians. Incident cardiovascular events in this population have been described previously (Evans et al., Br Med J 2002;324:939-942). Following written informed consent from individuals registered on DARTS, blood samples for genetic studies have been collected, thereby creating a genetic sub-study known as Go-DARTS (Genetics of DARTS) Study. Rigorous compliance with NHS data protection and encryption standards is maintained and the study was approved by the local research ethics committee.

Study population. We studied 2,182 individuals with type 2 diabetes (T2D) participating to Go-DARTS. In our study cohort, smoking status is recorded as current smoker, non-smoker, or ex-smoker. Because ex-smoker status was likely to be unreliably reported and confounded because of its association with background cardiovascular events or risk, we dichotomized smoking status to simplify the modeling process. For "ever-smokers," a current smoker and/or ex-smoker code was recorded within 1 year of recruitment. "Never-smokers" were defined as individuals who had only a non-smoking code during this time. Ratio of total cholesterol to HDL-cholesterol, and systolic (SBP) and diastolic (DBP) blood pressures were determined as the average of values

recorded within 1 year of enrollment in the study and recorded as continuous variables. Mean arterial blood pressure (MAP) was determined as follows: MAP = [(DBPx2) + SBP]/3. This study was approved by the local research ethics committee.

Genetic analyses. The following gene variants were analyzed: *IL-6 rs1800795*, *MCP-1 rs1024611*, *ICAM-1 rs5498*, *selectin-E rs5361*, *MMP-3 rs679620*. Analyses were performed DNA was prepared from blood and stored in aliquots at –20°C. Genotypes were assessed by use of the TaqMan-based allelic discrimination assays using the following oligonucleotides: for *ICAM-1*, g probe 5'Fam-TCA CCC GCG AGG TGA CCG T-Tamra3', a probe 5'Tet-TCA CCC GCA AGG TGA CCG TG-Tamra3', forward primer 5'-CAG TGA CTG TCA CTC GAG ATC TTG A-3', reverse primer 5'-CGG CTC ACT CAC AGA GCA CA-3'; for *IL-6* allele 1 5'-GAA GGT GAC CAA GTT CAT GCT GCA ATG TGA CGT CCT TTA GCA TG-3', for *IL-6* allele 2 5'-GAA GGT CGG AGT CAA CGG ATT GCA ATG TGA CGT CCT TTA GCA TC-3'; for *IL-6* allele c 5'-GCA CTT TTC CCC CTA GTT GTG TCT T-3'; for *MCP-1* allele 1 5'-GAA GGT GAC CAA GTT CAT GCT CAG AAA AGA AAG TCT TCT GGA AAG TGA T-3', for *MCP-1* allele 2 5'-GAA GGT CGG AGT CAA CGG ATT AGA AAA GAA AGT CTT CTG GAA AGT GAC-3'; for *MCP-1* allele c 5'-CAG AAG TGG GAG GCA GAC AGC T-3'; for *Selectin-E* allele 1 5'-GAA GGT GAC CAA GTT CAT GCT GCC TGT ACC AAT ACA TCC TGC A-3', for *E-Selectin* allele 2 5'-GAA GGT CGG AGT CAA CGG ATT GCC TGT ACC AAT ACA TCC TGC C-3', for *selectin-E* allele c 5'-GGG TCA CAC TTG CAA GTG TAA TTA TTG AT-3'; for *MMP-3* allele 1 5'-GAA GGT GAC CAA GTT CAT GCT GAA ATA TCT AGA AAA CTA CTA CGA CCT CA-3', for *MMP-3* allele 2 5'-GAA GGT CGG AGT CAA CGG ATT AAT ATC TAG AAA ACT ACT ACG ACC TCG-3', for *MMP-3* allele c 5'-GTC CTT TCT CCT AAC AAA CTG TTT CAC AT-3'. Allelic discrimination assays were performed on an Applied Biosystems 7700 sequence detection system using procedures specified by the manufacturer (Applied Biosystems).

Creation of the multigenic model. The 5 inflammatory SNPs described above were used to create an Inflammatory Risk Score (IRS). According to our previous study, in which the association of these SNPs with history of ischemic stroke was studied (Flex et al., Stroke 2004;35:2270-2275), the following genotypes were considered "at risk": *IL-6 rs1800795* GG or GC, *MCP-1 rs1024611* GG, *ICAM-1 rs5498* EE, *sel-E rs5361* RR, *MMP-3 rs679620* 5A5A. Therefore, the IRS model was created as an ordinal variable, ranging from 0 to 5, based on the number of "at risk" genotypes concomitantly carried by a given individual. Subjects without "at risk" genotypes were assigned an IRS = 0. Subjects with 1 – 5 "at risk" genotypes were assigned an IRS from 1 to 5, respectively. Stratification of the population according to the IRS was done prior to the determination of the number of disease events and the execution of the statistical analyses.

Disease events. All individuals were followed up until their first non-fatal or fatal stroke after recruitment. The mean follow-up period was 6.2 ± 1.1 years. Non-fatal strokes were determined from both the hospital Scottish Morbidity Record and the DARTS program of nurse-facilitated validation. For fatal events, the date of death was ascertained from DARTS, with the cause of death obtained from the General Registrars Office. Deaths resulting from stroke were defined as ICD (*International Classification of Diseases*) revision 9 (ICD-9) codes I60 to I69 and 430 to 438.

Statistical analyses. Cox's regression was used to determine the association of genotype with the outcome after recruitment until the end of the study period. Initially, each genotype was tested in isolation, using a recessive model. In all cases, the null allele was coded 1. Survival functions were adjusted subsequently for age at recruitment in all the Cox regressions, although they were removed when age at death was modeled. Smoking status was coded 1 for ever-smokers and 0 for never-smokers. The significance of the effect modification of genotype with smoking status or other risk factors was formally assessed by constructing a 2-way interaction term. For testing for the effect of smoking by genotype, a composite variable was constructed (ever-smokers genotype+, ever-smokers genotype–, never-smokers genotype+, and never-smokers genotype–) and entered as indicator variables. Fully adjusted models were also considered in which ratio of total to HDL-cholesterol, mean arterial pressure, and years with diabetes were also included. ROC analysis was used to compare the predictive value of the genetic model investigated in this study with the predictive value of classical risk factors for stroke. STATA version 8 was used for all analyses.

Supplemental Table 1: Genotype distribution and allele frequencies of inflammatory gene variants in the Go-DARTS cohort

Variant	Genotypes			MAF	HWE (P)
IL-6	GG	GC	CC	0.411	Pass (0.07)
rs1800795	735(33.7%)	1099 (50.4%)	348 (16.0%)		
MCP-1	AA	AG	GG	0.271	PASS (0.8)
rs1024611	1159 (53.1%)	865 (39.6%)	158 (7.24%)		
ICAM-1	KK	KE	EE	0.439	PASS (0.8)
rs5498	686 (31.5%)	1071 (49.1%)	425 (19.5%)		
sel-E	SS	SR	RR	0.095	PASS (0.6)
rs5361	1790 (82.0%)	369 (16.9%)	23 (1.0%)		
MMP-3	6A6A	5A6A	5A5A		PASS (0.5)
rs679620	557 (25.5%)	1103 (50.7%)	522 (23.9%)		

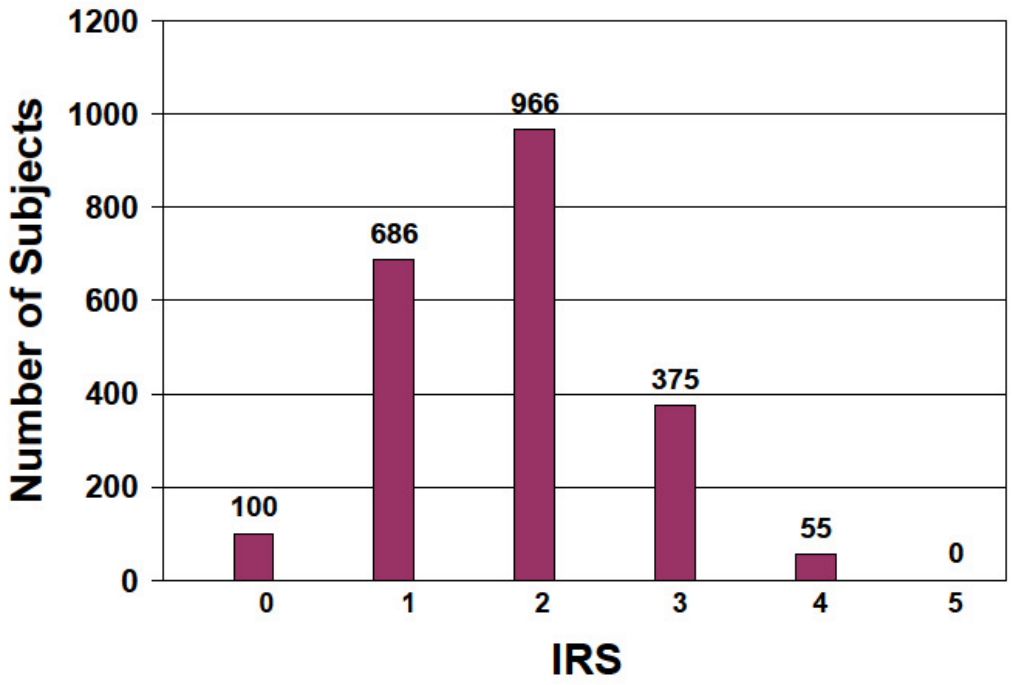
Supplemental Table 2: IRS genotypic model in subjects with and without previous cardiovascular diseases

	Hazard Ratio	95% CI	P
IRS in subjects **without** previous CV diseases	1.54	1.1-2.2	0.013
IRS in subjects **with** previous CV diseases	1.21	0.9-1.6	0.197

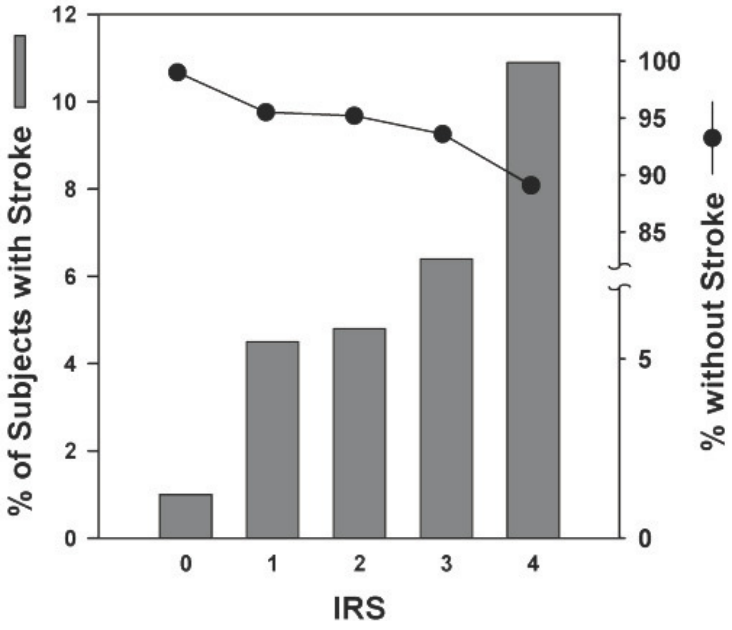
Analyses were performed on 2,123 genotyped individuals (1,403 without and 720 with previous cardiovascular diseases) with full covariates available for Cox's proportional hazards study

Analyses were corrected for age at enrolment, gender, smoking, blood pressure, and cholesterol ratio

Supplemental Fig. 1: Distribution of the Go-DARTS subjects according to the IRS.



Supplemental Fig. 2: Percentage of subjects with stroke (bars) and without stroke (line) according to the IRS.



Supplemental Fig. 3: ROC analysis of the predictive value of IRS compared to classical risk factors for stroke. The "non genetic model" included age, gender, smoking habit, blood pressure, lipid profile, body mass index, and previous cardiovascular diseases". In the "full model", IRS was added to classical risk factors. The area under the ROC curve was 0.7456 ± 0.0214 (CI 0.70358-0.78762) and 0.7563 ± 0.0216 (CI 0.71391-0.79862) for the two models, respectively. IRS adds 1.1% to the effectiveness of the model (P = n.s.).