Am. J. Hum. Genet. 70:593–603, 2002

Localization of a Susceptibility Gene for Common Forms of Stroke to 5q12

Solveig Gretarsdottir,[1] Sigurlaug Sveinbjörnsdottir,[2] Hjörtur H. Jonsson,[1] Finnbogi Jakobsson,[2] Elisabet Einarsdottir,[1] Uggi Agnarsson,[3] Dana Shkolny,[1] Gisli Einarsson,[2] Herdis M. Gudjonsdottir,[1] Einar M. Valdimarsson,[2] Olafur B. Einarsson,[1] Gudmundur Thorgeirsson,[3] Radinka Hadzic,[1] Sif Jonsdottir,[1] Sigridur Th. Reynisdottir,[1] Sigrun M. Bjarnadottir,[1] Thorunn Gudmundsdottir,[1] Gudrun J. Gudlaugsdottir,[3] Ramanjit Gill,[4] Klaus Lindpaintner,[4] Jesus Sainz,[1] Helgi H. Hannesson,[1] Gunnar Th. Sigurdsson,[1] Michael L. Frigge,[1] Augustine Kong,[1] Vilmundur Gudnason,[3] Kari Stefansson,[1] and Jeffrey R. Gulcher[1]

[1]deCODE Genetics, [2]National University Hospital, and [3]Icelandic Heart Association Heart Preventive Clinic, Reykjavik; and [4]F. Hoffmann–La Roche, Basel, Switzerland

Stroke is one of the most complex diseases, with several subtypes, as well as secondary risk factors, such as hypertension, hyperlipidemia, and diabetes, which, in turn, have genetic and environmental risk factors of their own. Here, we report the results of a genomewide search for susceptibility genes for the common forms of stroke. We cross-matched a population-based list of patients with stroke in Iceland with an extensive computerized genealogy database clustering 476 patients with stroke within 179 extended pedigrees. Linkage to 5q12 was detected, and the LOD score at this locus meets the criteria for genomewide significance (multipoint allele-sharing LOD score of 4.40, $P = 3.9 \times 10^{-6}$). A 20-cM region on 5q was physically and genetically mapped to obtain accurate marker order and intermarker distances. This locus on 5q12, which we have designated as "*STRK1*," does not correspond to known susceptibility loci for stroke or for its risk factors and represents the first mapping of a locus for common stroke.

Introduction

Stroke is a major health problem in western societies. It is the most common cause of disability, the second-most-common cause of dementia, and the third-most-common cause of death (Bonita 1992). Since it is more common in the elderly, the public health impact of stroke will increase in the next decades with growing life expectancy. Approximately one in four men and one in five women aged 45 years will have a stroke if they live to their 85th year (Bonita 1992). Strategies to diminish the impact of stroke include prevention and treatment with thrombolytic and, possibly, neuroprotective agents. The success of preventive measures will depend on the identification of risk factors and means to modulate their impact.

The clinical phenotype of stroke is complex but can be broadly divided into ischemic and hemorrhagic strokes. The majority (80%–90%) of strokes are ischemic—that is, they are caused by obstruction of blood flow through extra- or intracranial vessels (Mohr et al. 1978; Caplan 2000). The remainder (10%–20%) are hemorrhagic—that is, they result from ruptures of intracranial vessels. Ischemic stroke can be further subdivided into large-vessel occlusive disease, small-vessel occlusive disease, and cardiogenic stroke. For the purposes of this study, we have included transient ischemic attack (TIA) as a biological equivalent of ischemic stroke, even though TIA is not defined as a stroke (because the signs and symptoms, which are the same as those for stroke, last for a short period of time [i.e., <24 h; usually 5–10 min]). This is done because the same pathophysiological mechanisms are considered responsible for TIA and ischemic stroke (Caplan 2000).

The predominant risk factor for all types of stroke is hypertension (Thompson and Furlan 1997; Agnarsson et al. 1999). Hypertension is in itself a complex disease, as are the other known risk factors, diabetes and hyperlipidemia. In addition, there are environmental risk factors, such as smoking. Stroke is therefore considered to be a highly complex disease consisting of a group of heterogeneous disorders with multiple risk factors, both genetic and environmental.

The identification of genetic determinants of common diseases, such as stroke, that may result from the interplay of multiple genes and interactions between genes

Received May 22, 2001; accepted for publication December 12, 2001; electronically published February 6, 2002.

Address for correspondence and reprints: Dr. Kari Stefansson, deCODE Genetics, Sturlugata 8, IS 101-Reykjavik, Iceland. E-mail: kstefans@decode.is; or Dr. Jeffrey Gulcher, deCODE Genetics, Sturlugata 8, IS 101-Reykjavik, Iceland. E-mail: jgulcher@decode.is

0002-9297/2002/7003-0005$15.00

and environment has proven to be a difficult task. Studies of the genetic contribution to stroke have mainly focused either on rare Mendelian diseases in which stroke is a part of the phenotype or on finding association between stroke and possible candidate genes, such as genes contributing to hypertension or lipid metabolism. Several genes have been identified that play roles in the pathogenesis of rare stroke syndromes, such as *Notch3,* in cerebral autosomal dominant arteriopathy with subcortical infarctions and leukoencephalopathy (Tournier-Lasserve et al. 1993; Joutel et al. 1996); *Cystatin C,* in the Icelandic type of hereditary cerebral hemorrhage with amyloidosis (Palsdottir et al. 1988); *APP,* in the Dutch type of hereditary cerebral hemorrhage (Levy et al. 1990); and *KRIT1,* in hereditary cavernous angioma (Gunel et al. 1995; Laberge-le Couteulx et al. 1999; Sahoo et al. 1999).

To our knowledge, no genomewide search for stroke genes in patients with the common forms of stroke has ever been reported. Here we report the results of a genomewide search for susceptibility genes in common stroke by use of a broad but rigorous definition of the phenotype, including hemorrhagic stroke, ischemic stroke, and TIA. The result of this is the mapping of the first major locus reported in common stroke.

Subjects and Methods

Patients

An encrypted population-based list that contained 2,000 living Icelandic patients with stroke and was based on hospital *International Classification of Diseases, Ninth Revision* codes covering the period of time from 1993 to 1997 was run through our computerized genealogy database (Gulcher and Stefansson 1998; Gulcher et al. 2000), which covers the whole Icelandic nation. We excluded patients with subarachnoid hemorrhage or the Icelandic type of hereditary cerebral hemorrhage with amyloidosis. The distribution of stroke types in our study is similar to that reported in other white populations, with ~67% having ischemic strokes, 27% having TIAs, and 6% having hemorrhagic strokes (Caplan 2000). All patients underwent computerized tomography of the head, and the majority of patients underwent Doppler ultrasound of carotid arteries and echocardiography; Holter monitoring was frequently used.

We collected patients with stroke and/or TIAs by use of the criterion that the relationship between each patient and at least one additional patient was characterized by no more than six meiotic events (six meiotic events separate second cousins). Participating patients were more carefully phenotyped by the clinicians before their genotypes were generated. Patients with ischemic stroke and TIAs were classified according to the TOAST (Trial of Org 10172 in Acute Stroke Treatment) subclassification system (Adams et al. 1993). This system includes five categories: (1) large-artery atherosclerosis, (2) cardioembolism, (3) small-artery occlusion (lacune), (4) stroke of other determined etiology, and (5) stroke of undetermined etiology. The diagnoses were based on clinical features and on data from ancillary diagnostic studies. Patients classified as having large-artery atherosclerosis had clinical and brain-imaging findings of cerebral cortical dysfunction and either significant (>70%) stenosis (this is a stricter criteria than that used in TOAST, in which 50% stenosis is the cutoff) or occlusion of a major brain artery or branch cortical artery. Potential sources of cardiogenic embolism were excluded. The second category, cardioembolism, included patients with at least one cardiac source for an embolus and with potential large-artery sources of thromobosis and embolism having been eliminated. Patients with small-artery occlusion had one of the traditional clinical lacunar syndromes and no evidence for cerebral cortical dysfunction. A potential cardiac source of embolus and stenosis >70% in an ipsilateral extracranial artery was excluded. The fourth category, acute stroke of other determined etiology, included patients with rare causes of stroke and patients with two or more potential causes of stroke. If the causes of stroke could not be determined despite extensive evaluation, then patients were included in the fifth category, stroke of undetermined etiology. TOAST classification of patients with ischemic stroke and TIA whom we studied is presented in table 1. Apart from the proportion of large-vessel disease, which is lower in the population that we studied, the subtype distribution is similar to those reported in other studies (e.g., Caplan 2000). This is very likely due to the stricter stenosis criterion that we used to classify large-vessel disease.

The present study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Informed consent was obtained from all patients and their relatives whose DNA samples were used in the linkage analysis.

Genomewide Scan

A genomewide scan was performed on 476 patients and 438 of their relatives, by use of our framework marker set of 1,000 microsatellite markers. We have developed a microsatellite screening set that is based, in part, on the ABI Linkage Marker (version 2) screening set and the ABI Linkage Marker (version 2) intercalating set, in combination with 500 custom-made markers. All markers were extensively tested for robustness, ease of scoring, and efficiency in 4× multiplex PCRs. In our framework marker set, the average spacing between

markers was ~4 cM, with no gaps >10 cM. Marker positions were obtained from the Marshfield map (Center for Medical Genetics, Marshfield Medical Research Foundation), except for the region containing a three-marker putative inversion on chromosome 8 (Jonsdottir et al. 2000; Giglio et al. 2001; Yu et al. 2001). The PCR amplifications were prepared, run, and pooled on Perkin Elmer/Applied Biosystems 877 Integrated Catalyst Thermocyclers with a similar protocol for each marker. The reaction volume was 5 μl, and, for each PCR, 20 ng of genomic DNA was amplified in the presence of 2 pmol of each primer, 0.25 U Ampli*Taq* Gold, 0.2 mM dNTPs, and 2.5 mM $MgCl_2$ (buffer was supplied by manufacturer). Cycling conditions were 95°C for 10 min, followed by 37 cycles at 94°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min. The PCR products were supplemented with the internal size standard, and the pools were separated and detected on an Applied Biosystems model 377 Sequencer by use of Genescan version 3.0 peak-calling software. Alleles were automatically called with the TrueAllele program (Cybergenetics), and the DecodeGT program (deCODE Genetics) was used both to fractionate according to quality and to edit the called genotypes (Palsson et al. 1999). At least 180 Icelandic controls were genotyped for each marker to derive allele frequencies.

Statistical Methods for Linkage Analysis

In our analyses, we used multipoint, affected-only allele-sharing methods to assess the evidence for linkage. All results, including LOD and nonparametric linkage (NPL) scores, were obtained using the program Allegro (Gudbjartsson et al. 2000). We used the S_{pairs} scoring function (Whittemore and Halpern 1994; Kruglyak et al. 1996) and the exponential allele-sharing model (Kong and Cox 1997) to generate the relevant statistics with 1 df. When combining the family scores to obtain an overall score, instead of weighting the families equally (the default of Genehunter [Kruglyak et al. 1996]) or weighting the affected pairs equally, we used a weighting scheme that is halfway between the two in the log scale; the family weights that we used are the geometric means of the weights of the two schemes. Although not identical, this weighting scheme tends to yield results that are similar to those proposed by Weeks and Lange (1988) as an extension of a weighting scheme by Hodge (1984) that was designed for sibships. We computed the *P* value two different ways and here report the less significant result. The first *P* value was computed on the basis of large sample theory; the distribution of $Z_{lr} = \sqrt{2\,[\log_e(10)\mathrm{LOD}]}$ approximates a standard normal random variable under the null hypothesis of no linkage (Kong and Cox 1997). Because the normal approximation may not work well in some small-sample situations, we computed a second *P* value by comparing the observed LOD score with its complete data-sampling distribution under the null hypothesis (Gudbjartsson et al. 2000). When a data set consists of more than a few families, as is the case here, these two *P* values tend to be very similar. To ensure that the result was a true reflection of the information contained in the material, we considered a linkage result significant not only if the *P* value was $< 2 \times 10^{-5}$ (Lander and Kruglyak 1995) but also if the information content in the region was $\geqslant$85%. For the families in the present study, an information content of 85% corresponded to a marker density of approximately one marker per centimorgan. The information measure we used has been defined elsewhere (Nicolae 1999) and has been implemented in Allegro. This measure is closely related to a classical measure of information (Dempster et al. 1977), which has the property that it is between zero, if the marker genotypes are completely uninformative, and one, if the genotypes determine the exact amount of allele sharing by descent among the affected relatives.

After obtaining a significant allele-sharing LOD score,

Table 1

Subclassification of Patients with Stroke

	% Affected among	
Subtype	All Patients (n = 476)	Patients in Families with NPL >1 (n = 120)
Hemorrhagic	5	6
Ischemic:		
Large vessel[a]	13	13
Small vessel	16	13
Cardioembolic	23	28
Other cause	4	5
More than one subtype or unknown cause	39	35

[a] The definition of ischemic large-vessel disease that we used is stricter than that usually used in TOAST (see "Subject and Methods" section).

we attempted to understand the contribution of this susceptibility locus by fitting a range of parametric models to the data. Even when fitting parametric models, we performed affected-only analyses, in the sense that an individual is classified as either affected or as having unknown disease status. As a consequence, only ratios of penetrances are relevant. We fitted a range of single-locus dominant, additive, and multiplicative models (Risch 1990). With a complex disease such as stroke, none of these simple models are likely to be exactly true, and the effect of a gene and its variants can only be reliably determined after the at-risk variant (or variants) is identified. However, by the calculation of the corresponding contribution to the sibling recurrence-risk ratio, the fitted parametric models do provide some rough idea of how much the gene is contributing to the familial clustering of the disease.

We investigated the contributions, to the identified locus, of several subtypes of and risk factors for stroke. To do this, we utilized the complete family set and considered as affected only the patients with a particular subtype of or risk factor for stroke. In one particular case, to assess whether the LOD-score increase resulting from the subtraction of the 22 patients with hemorrhagic stroke would be likely to occur by chance, we selected 1,000 random sets of 22 patients whose status we then changed to unknown in an analysis. The *P* value we present is the fraction of the 1,000 simulations that produced, at the peak locus, a LOD-score increase that was equal to or greater than that which we observed by changing the affection status of the patients with hemorrhagic stroke to unknown.

Physical Mapping

To obtain correct marker order and sequence-ready contigs, we physically mapped a 20-cM region, on 5q, that was indicated in the genomewide scan. BAC contigs were generated by a method that combines the results of coincident primer-hybridization experiments with the mining of publicly available sequences. RPCI-11 human male BAC library segments 1 and 2 (Pieter de Jong, Children's Hospital Oakland Research Institute), containing ~200,000 clones with a 12× coverage of the genome, were arrayed using a 6 × 6 double-offset pattern on 23-cm × 23-cm membranes. Initially, hybridizations were performed with markers that were expected, on the basis of their locations in the Weizmann Institute of Science Unified Database for Human Genome Mapping, to be in the region of interest. We used 150 markers in the region (i.e., 31 polymorphic markers used in linkage and 120 markers generated from sequence-tagged sites), which were separated by, on average, 130 kb. The selected markers were used to generate two [^{32}P]-labeled probes: F, which contained the pooled forward primers, and R, which contained the pooled reverse primers. The coincident signals in both hybridizations were selected as positive clones. A set of overlapping clones was assembled through a combination of hybridization and BAC-fingerprint walking. Fingerprints of positive clones (FPCs) were analyzed using the FPC database developed at the Wellcome Trust Sanger Institute. Data from FPC contigs prebuilt with a cutoff of $3e^{-12}$ and from sequence data mining were integrated with the hybridization results. BACs in the region detected by data mining and hybridization were rearrayed. Small membranes (8 cm × 12 cm) were arrayed in 6 × 6 double-offset pattern and were individually hybridized with the markers of interest. A visual map was generated by combining the hybridization, fingerprinting, and sequence data. A total of 137 new markers were generated from BAC end sequences, and the process was repeated until the majority of gaps were closed. Estimates of contig lengths and of the distance between markers assigned to them were based on the FPC program.

Genetic Mapping

High-resolution genetic mapping was used to order contigs obtained by physical mapping and to determine their orientation. In addition to correct marker order, the high-resolution genetic map also provided better estimates of intermarker distances, both of which are important for an accurate linkage analysis (Halpern and Whittemore 1999; Daw et al. 2000). Data from 112 Icelandic nuclear families (sibships with genotypes for two to seven siblings and both parents) were analyzed together with the genotypes for nuclear families available within the stroke pedigrees. For the purpose of genetic mapping, the 112 families alone provide 588 meiotic events, and the inclusion of the data from the families with stroke yielded a map based on substantially more than 1,000 meiotic events. By comparison, the Marshfield genetic map (Center for Medical Genetics, Marshfield Medical Research Foundation) was constructed on the basis of 182 meiotic events. The large number of meiotic events within our families provides the ability to map markers to a resolution $\leqslant$1.0 cM. In evaluating one order of the markers versus another, by modifying the Allegro program, we computed the number of obligate crossovers for each order, and the order associated with a lower number of crossovers was preferred (Thompson 1987). Given an order, genetic distances between markers were estimated by implementing the expectation-maximization algorithm (Dempster et al. 1977) within the Allegro program. Combining the information from genetic mapping with the physical map resulted in a highly reliable order of markers and intermarker distances within this 20-cM region.

Results

We collected samples from a total of 476 patients, each of whom is related to (within and including six meiotic events) at least one other patient. Patients with hemorrhagic stroke clustered in families with ischemic stroke and TIA, and there were no families with a striking preponderance of either hemorrhagic stroke or further subtypes of ischemic stroke. Given this observation, we decided to study stroke as a broadly defined phenotype. The genome scan was performed with the 476 patients clustered into 179 families. The mean separation of affected pairs was 4.8 meiotic events. Figure 1, which displays four of the families, shows that several stroke subtypes, including hemorrhagic stroke, are found mixed together within the same pedigrees.

Figure 2 presents the allele-sharing LOD scores from the genome scan by use of the framework map. Three regions achieved a LOD score >1.0. Two of these regions were on 5q: one peak at ~69 cM, with a LOD score of 2.00, and a second peak at 99 cM, with a LOD score of 1.14. The third region is on 14q, at 55 cM, with a LOD score of 1.24.

The information for analysis of linkage at the 5q locus was increased by genotyping 45 additional markers over a 45-cM segment that contains both of the regions on 5q (fig. 3). Although the LOD score at the second peak decreased slightly, to ~1.05, the LOD score at the first peak increased to 3.39. However, close inspection of our results suggested not only that the Marshfield genetic map (Center for Medical Genetics, Marshfield Medical Research Foundation) lacks resolution (i.e., many markers were assigned to the same location) but also that there may be some errors in their order. When we followed the marker order of the Marshfield map and used the Allegro program and our data to estimate

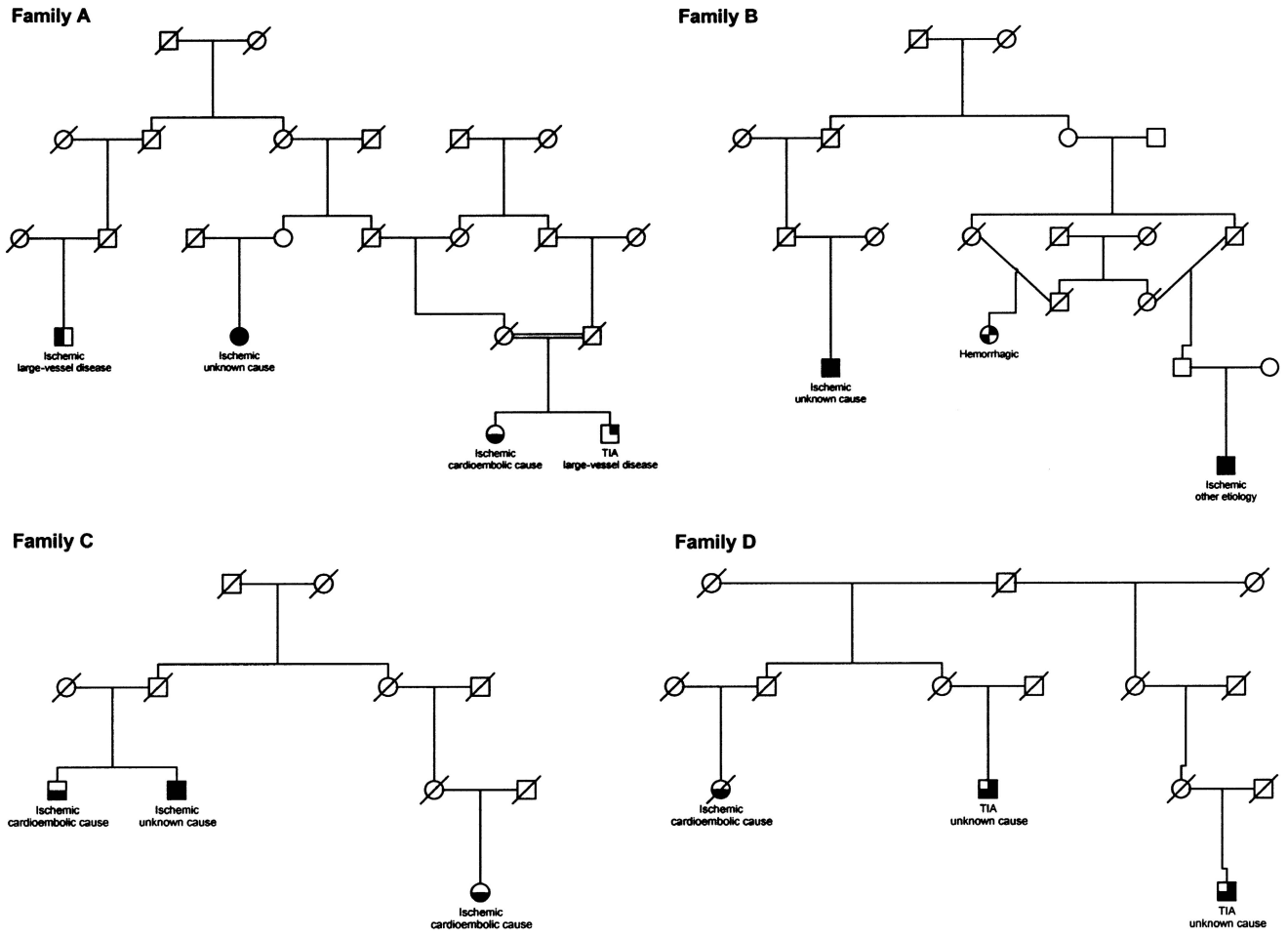

Figure 1 Four families with positive LOD scores. These families include patients with a variety of stroke subphenotypes, as defined by TOAST subclassification (as labeled underneath shaded symbols). Squares and circles represent males and females, respectively; slash marks through symbols indicate individuals who are deceased. Some sex indicators in the two upper generations of the pedigrees have been altered, and unaffected siblings of patients are not displayed, to protect the confidentiality of these families.

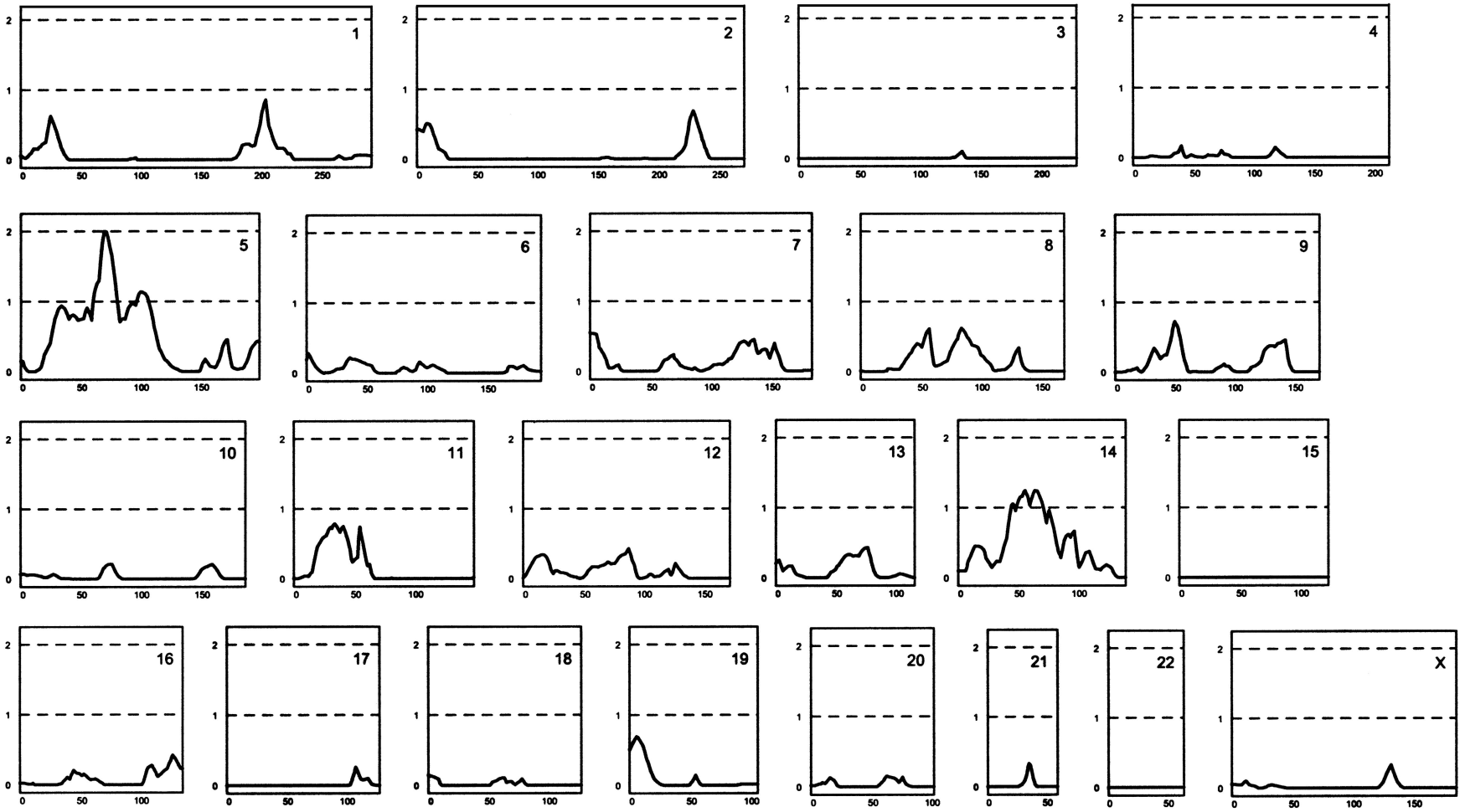

Figure 2 Genomewide scan for stroke-susceptibility loci. A framework map of ~1,000 markers was used. Each box represents a chromosome (indicated in the upper-right-hand corner of each box). The *X*-axis gives the genetic distance (in cM), and the *Y*-axis gives the LOD scores.

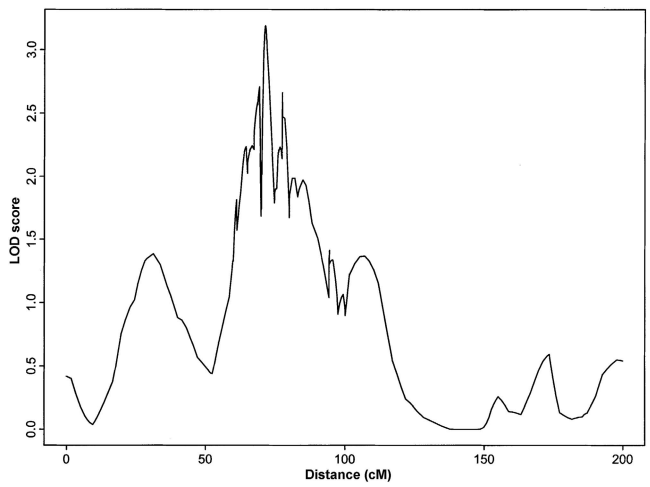

Figure 3 Dense mapping of stroke locus on chromosome 5, with 45 additional markers across the two peaks on 5q. The analysis used the marker order of the Marshfield map (Center for Medical Genetics, Marshfield Medical Research Foundation). The *X*-axis gives the genetic distance (in cM) along the chromosome, and the *Y*-axis gives the LOD score.

the genetic distances between markers, we found that our estimate of the genetic length of the region was substantially longer than that given in the Marshfield map. This indicates a problem with marker order because, in general, incorrect marker order leads to an increased number of apparent crossovers and increases the apparent genetic length. We improved the marker order and intermarker distances by constructing high-density physical and genetic maps over a 20-cM region between D5S474 and D5S2046 (fig. 4). It is worth noting that, although our final order and intermarker distances deviate from those of Marshfield, the overall genetic length for the region is similar.

Linkage analysis with genotypes from the higher-density markers by use of our marker order resulted in a LOD score of 4.40 ($P = 3.9 \times 10^{-6}$) on 5q12 at D5S2080. We designate this locus as "*STRK1*." With the addition of these extra markers, we were able to narrow the most promising region for the harboring of a stroke-susceptibility gene to a segment <6 cM, from D5S1474 to D5S398, as defined by a decrease of 1 in LOD score. Analyses with marker orders based on publicly available marker maps yielded lower LOD scores, 2.78–3.94, thereby highlighting the importance of accurate marker order when using multipoint analysis (fig. 5).

In an attempt to understand the contribution of this susceptibility locus to stroke, we fitted a range of parametric models to the data. The highest LOD score, 4.70, was obtained from a multiplicative model under the assumptions that the at-risk allele frequency was 27% and that there was a fivefold increase in risk for every at-risk allele carried. Under this model, the contribution of this gene to the sibling recurrence-risk ratio was 1.86. Seventy-five of the 179 family clusters yielded a positive LOD score; of these, 55 had LOD scores >0.1, and 5 had LOD scores >0.4. The four families displayed in figure 1 (i.e., families A–D) yielded LOD scores of 0.39, 0.40, 0.47, and 0.48, respectively. These results support the existence of a major stroke-susceptibility gene in this region.

The fractions of all patients in the study who have the various subtypes of stroke are listed in table 1. The fractions are also listed for those families with an NPL score >1 (within these families, there is more sharing among affected members of genetic material at the locus than was expected owing simply to their relationship). The families with more excess sharing at the locus do not show any substantial difference in phenotype pattern from the entire family set. Similar fractions are presented for the risk factors for stroke in each of the two family sets (table 2). Again, no substantial shift in the prevalence of the risk factors is obvious. To assess more directly the contribution of the various subtypes and risk factors to the peak locus, linkage runs were conducted in which only patients with the particular subtype or risk factor were considered as affected—that is, all other patients had their affection status changed to unknown for these runs. In each of these runs, the LOD scores were positive but were smaller than those in the run including all patients. These decreases in LOD score were consistent with the loss of power in the smaller sample sizes. We also conducted a run in which only patients with ischemic stroke were considered as affected. This run, which excluded the 22 patients with hemorrhagic stroke, had an increase in LOD score. The allele-sharing LOD score for this run increased to 4.86 at D5S2080. Although this 0.46 increase in LOD score suggests that *STRK1* is involved primarily in ischemic strokes and TIAs, the increase itself is not statistically significant, on the basis of simulations (one-sided $P = .09$). In summary, these results are consistent with a susceptibility gene at this locus that contributes to a broad spectrum of patients with stroke, the possible exception being patients with hemorrhagic stroke.

Discussion

In this study, we have successfully mapped a major locus for one of the most complex diseases known, by combining genealogy, a comprehensive population-based list of patients with broadly defined stroke, and allele-sharing methods. In any linkage or association study that uses multipoint marker analysis, a correct marker order and precise intermarker distances are important. Otherwise,

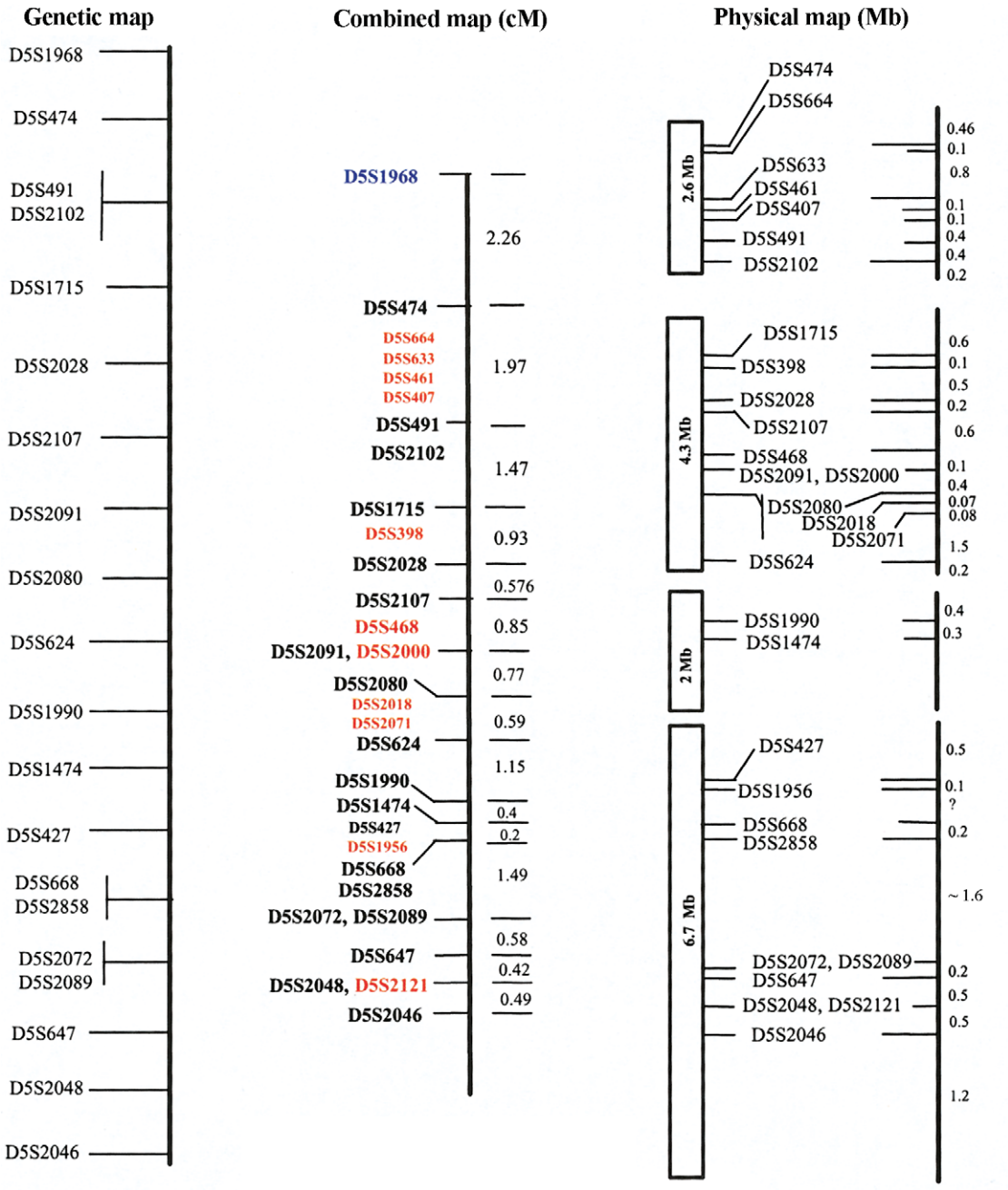

Figure 4 Genetic, physical, and combined maps for *STRK1*, on 5q, from D5S1968 to D5S2046. Markers assigned in both the genetic map and the physical map are displayed in black, markers derived only from physical-map information are displayed in red, and markers derived only from genetic-map information are displayed in blue. Marker distances (in cM) in the combined map were constructed by applying the estimation-maximization algorithm to the final marker order; marker distances (in Mb) in the physical map are estimations from the FPC program.

the apparent increase in information content is neutralized or reduced by the resulting misinformation. We found that a direct application of most public genetic and physical maps, which may have numerous inaccuracies or ambiguities, have a negative impact on the LOD score of this locus for stroke. While our work was in progress, an assembly of the current draft of the human genome—the University of California–Santa Cruz (UCSC) Human Genome Project Working Draft—was made available (Lander et al. 2001). This assembly merges together

overlapping fragments and orders and orients nonoverlapping fragments on the basis of mRNA, EST, paired plasmid reads, and other information. In the April 2001 freeze from UCSC (for which data was released in June 2001), 30 of our 31 linkage markers mapped to two contigs (the remaining marker was not mapped in this freeze). The marker order within the contigs was in agreement with our order, with the exception of two markers, D5S2858 and D5S668. However, in the latest release, in October 2001 (August 2001 freeze), several changes have occurred. Whereas the order of the two markers (D5S2858 and D5S668) has been changed and is now consistent with our order, two other pieces—one involving D5S2028–D5S2080 (>1 Mb) and the other involving D5S427 and D5S1956 (~200 kb)—are flipped and thus are inconsistent with what we believe to be the correct order. This indicates that there is still substantial uncertainty in the assembly of the public human sequence.

The types of stroke that are presented in this article do not reflect a rare stroke form or a form specific to Iceland. Rather, the diverse stroke phenotypes in Icelanders, as well as known risk factors for stroke, are similar to those of most other white populations (Sveinbjörnsdottir et al. 1998; Valdimarsson et al. 1998; Eliasson et al. 1999).

The known genetic factors contributing to common stroke may act indirectly, by increasing the risk of some predisposing conditions, such as diabetes, hyperlipidemias, and/or hypertension. It is also possible that there are genetic factors for stroke that do not influence susceptibilities to the known risk factors, as has been suggested by epidemiological studies for myocardial infarction (Shea et al. 1984; Friedlander et al. 1985; Myers et al. 1990). Epidemiological studies of the common forms of stroke have given conflicting results in regard to the role of family history. Some studies have shown that parental history predicts the risk of stroke

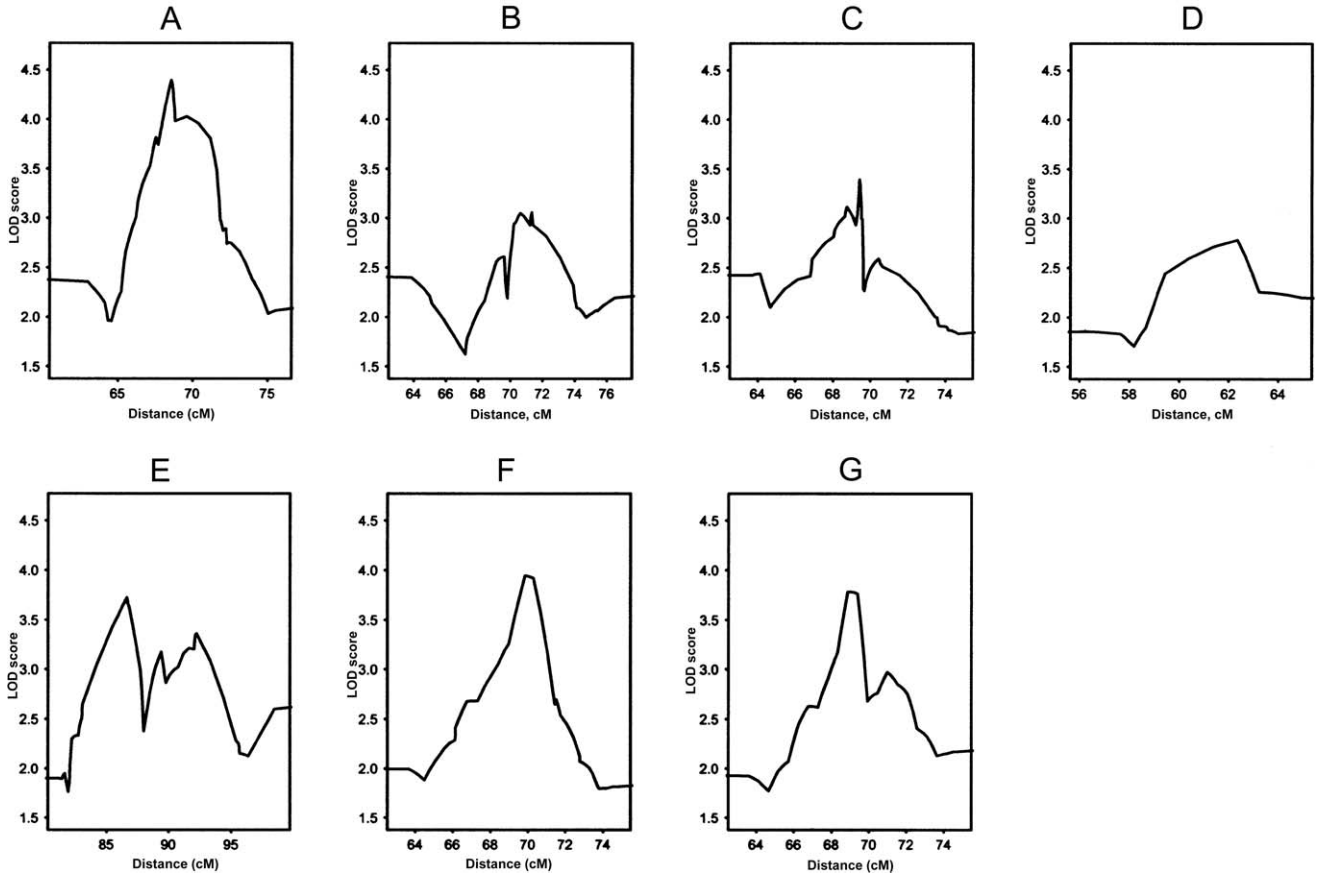

Figure 5 Linkage analyses with marker orders from our combined map (*A;* see fig. 4) and maps from Généthon (*B*), the Center for Medical Genetics, Marshfield Medical Research Foundation (*C;* where the Marshfield map has no resolution, our order was used), Stanford (*D;* radiation-hybrid map), the Weizman Institute Unified Database for Human Genome Mapping (*E*), Whitehead Institute Center for Genome Research (since the Whitehead map only gives order and not distances, it was run both with distances based on application of the estimation-maximization algorithm [*F*] and with equally spaced markers [*G*]).

Table 2

Prevalence of Risk Factors

Risk Factor	% [No.] Affected among	
	All Patients (n = 453)	Patients in Families with NPL >1 (n = 117)
Hypertension[a]	73 [329]	76 [89]
Diabetes[b]	14 [63]	15 [18]
Hypercholesterolemia[c]	24 [111]	21 [25]

Note.—For 23 patients, information on risk factors was unavailable.

[a] If patients (*a*) had measured blood-pressure values of SBP ⩾160 mmHg and/or DBP ⩾95 mmHg, (*b*) had a history of hypertension, or (*c*) had no history of hypertension but were being treated for hypertension.

[b] If patients (*a*) had nonfasting glucose levels ⩾10 mM, (*b*) had a history of diabetes, or (*c*) had no history but were being treated for diabetes.

[c] If patients (*a*) had total cholesterol ⩾7 mM or (*b*) were on lipid-lowering medication.

independently from conventional risk factors (Jousilahti et al. 1997; Liao et al. 1997), whereas others have failed to find evidence for such independent factors (Kiely et al. 1993; Lindenstrom et al. 1993; Graffagnino 1994). However, our work describes the first reported genome scan in search of genes that contribute to common forms of stroke. Our data suggest that the locus we have mapped contributes directly to stroke, rather than indirectly through known risk factors for stroke. This suggests that there may be biological pathways independent of the known risk factors that contribute to the pathogenesis of stroke. Regardless of what the mechanism is, the evidence presented supports a major genetic component in the pathogenesis of stroke in Iceland.

Acknowledgments

We thank the participating Icelandic patients with stroke and their families, whose contribution have made this study possible; the nurses at the Icelandic Heart Association, Hólmfrídur B. Thorsteinsdottir, and the members of deCODE's BAC physical-mapping and genotyping facilities, for excellent assistance; and Dr. Thorgeir Thorgeirsson, for valuable comments on the manuscript.

Electronic-Database Information

URLs for data in this article are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://research.marshfieldclinic.org/genetics/ (for genetic marker map)

Cybergenetics, http://www.cybgen.com/ (for TrueAllele program)

deCODE Genetics, http://www.decode.com/ (for DecodeGT program)

Généthon, http://www.genethon.fr/

UCSC Human Genome Project Working Draft ("Golden Path"), http://genome.ucsc.edu/

Unified Database for Human Genome Mapping, The, http://bioinformatics.weizmann.ac.il/udb/

Wellcome Trust Sanger Institute, The, http://www.sanger.ac.uk/

Whitehead Institute Center for Genome Research, http://www-genome.wi.mit.edu/

References

Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE (1993) Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. Stroke 24:35–41

Agnarsson U, Thorgeirsson G, Sigvaldason H, Sigfusson N (1999) Effects of leisure-time physical activity and ventilatory function on risk for stroke in men: the Reykjavik Study. Ann Intern Med 130:987–990

Bonita R (1992) Epidemiology of stroke. Lancet 339:342–344

Caplan LR (2000) Caplan's stroke: a clinical approach, 3d ed. Butterworth-Heinemann, Stoneham, MA, pp 1–556

Daw EW, Thompson EA, Wijsman EM (2000) Bias in multipoint linkage analysis arising from map misspecification. Genet Epidemiol 19:366–380

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Eliasson JH, Valdimarsson EM, Jakobsson F (1999) Case fatality after acute stroke at the Reykjavik Hospital in 1996–1997. Læknablaðið 85:517–525

Friedlander Y, Kark JD, Stein Y (1985) Family history of myocardial infarction as an independent risk factor for coronary heart disease. Br Heart J 53:382–387

Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. Am J Hum Genet 68:874–883

Graffagnino C (1994) The importance of family history in cerebrovascular disease. Stroke 25:1599–1603

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K (2000) Protection of privacy by third-party encryption in genetic research in Iceland. Eur J Hum Genet 8:739–742

Gulcher J, Stefansson K (1998) Population genomics: laying the groundwork for genetic disease modeling and targeting. Clin Chem Lab Med 36:523–527

Gunel M, Awad IA, Anson J, Lifton RP (1995) Mapping a gene causing cerebral cavernous malformation to 7q11.2-q21. Proc Natl Acad Sci USA 92:6620–6624

Halpern J, Whittemore AS (1999) Multipoint linkage analysis: a cautionary note. Hum Hered 49:194–196

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

Jonsdottir GM, Jonasson K, Gudbjartsson D, Kong A (2000)

Construction of a genetic marker map using genotypes from over 7500 Icelanders. Am J Hum Genet 67 Suppl 2:332

Jousilahti P, Rastenyte D, Tuomilehto J, Sarti C, Vartiainen E (1997) Parental history of cardiovascular disease and risk of stroke. A prospective follow-up of 14371 middle-aged men and women in Finland. Stroke 28:1361–1366

Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cecillion M, Marechal E, Maciazek J, Vayssiere C, Cruad C, Cabanis EA, Ruchoux MM, Wessenbach J, Bach JF, Bousser MG, Tournier-Lasserve E (1996) *Notch3* mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 383:707–710

Kiely DK, Wolf PA, Cupples LA, Beiser AS, Myers RH (1993) Familial aggregation of stroke: the Framingham Study. Stroke 24:1366–1371

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Laberge-le Couteulx S, Jung HH, Labauge P, Houtteville JP, Lescoat C, Cecillon M, Marechal E, Joutel A, Bach JF, Tournier-Lasserve E (1999) Truncating mutations in *CCM1,* encoding KRIT1, cause hereditary cavernous angiomas. Nat Genet 23:189–193

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Levy E, Carman MD, Fernandez-Madrid IJ, Power MD, Lieberburg I, van Duinen SG, Bots GT, Luyendijk W, Frangione B (1990) Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. Science 248: 1124–1126

Liao D, Myers R, Hunt S, Shahar E, Paton C, Burke G, Province M, Heiss G (1997) Familial history of stroke and stroke risk: the Family Heart Study. Stroke 28:1908–1912

Lindenstrom E, Boysen G, Nyboe J (1993) Risk factors for stroke in Copenhagen, Denmark. I. Basic demographic and social factors. Neuroepidemiology 12:37–42

Mohr JP, Caplan LR, Melski JW, Goldstein RJ, Duncan GW, Kistler JP, Pessin MS, Bleich HL (1978) The Harvard Cooperative Stroke Registry: a prospective registry. Neurology 28:754–762

Myers RH, Kiely DK, Cupples LA, Kannel WB (1990) Parental history is an independent risk factor for coronary artery disease: the Framingham Study. Am Heart J 120:963–969

Nicolae DL (1999) Allele sharing models in gene mapping: a likelihood approach. PhD thesis, Department of Statistics, University of Chicago, Chicago, p 126

Palsdottir A, Abrahamson M, Thorsteinsson L, Arnason A, Olafsson I, Grubb A, Jensson O (1988) Mutation in *cystatin C* gene causes hereditary brain haemorrhage. Lancet 2:603–604

Palsson B, Palsson F, Perlin M, Gudbjartsson H, Stefansson K, Gulcher J (1999) Using quality measures to facilitate allele calling in high-throughput genotyping. Genome Res 9:1002–1012

Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

Sahoo T, Johnson EW, Thomas JW, Kuehl PM, Jones TL, Dokken CG, Touchman JW, Gallione CJ, Lee-Lin SQ, Kosofsky B, Kurth JH, Louis DN, Mettler G, Morrison L, Gil-Nagel A, Rich SS, Zabramski JM, Boguski MS, Green ED, Marchuk DA (1999) Mutations in the gene encoding KRIT1, a Krev-1/rap1a binding protein, cause cerebral cavernous malformations (*CCM1*). Hum Mol Genet 8:2325–2333

Shea S, Ottman R, Gabrieli C, Stein Z, Nichols A (1984) Family history as an independent risk factor for coronary artery disease. J Am Coll Cardiol 4:793–801

Sveinbjörnsdottir S, Einarsson G, Magnusdottir S, Gudmundsson G, Jonsson JE (1998) Systematic registration of patients with stroke and TIA admitted to the National University Hospital, Reykjavik, Iceland, in 1997. Læknablaðið Suppl 36:54

Thompson DW, Furlan AJ (1997) Clinical epidemiology of stroke. Neurosurg Clin N Am 8:265–269

Thompson EA (1987) Crossover counts and likelihood in multipoint linkage analysis. IMA J Math Appl Med Biol 4:93–108

Tournier-Lasserve E, Joutel A, Melki J, Weissenbach J, Lathrop GM, Chabriat H, Mas JL, et al (1993) Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy maps to chromosome 19q12. Nat Genet 3:256–259

Valdimarsson EM, Sigurdsson G, Jakobsson F (1998) Etiology and treatment of cerebral ischemia at the Department of Neurology and Rehabilitation Medicine at Reykjavik City Hospital. Læknablaðið 84:921–927

Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. Am J Hum Genet 42:315–326

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001) Comparison of human genetic and sequence-based physical maps. Nature 409:951–953