since schizophrenia has been linked to 8p in other populations. *NRG1* provides a way of unifying a large body of evidence coming from many directions, suggesting that multiple neurotransmitter systems and their receptors are involved in schizophrenia by representing a common denominator upstream of neurotransmitter expression and activation. Furthermore, it may provide support for the view that schizophrenia is caused by disregulation of synaptic plasticity in the adult.

Acknowledgments

We thank the participating patients and their families, and we thank Hjördís Pálsdóttir, Hallbera Leifsdóttir, Þuríður Þórðardóttir, Guðrun Jóhannesdóttir, and Tómas Þór Ágústsson for assisting with the sample collection.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Children's Hospital Oakland, BACPAC Resources, http://www.chori.org/bacpac/ (for RCPI-11 human BAC library)
deCODE Genetics, http://www.decode.com/nrg1/markers/ (for SNPs and microsatellite markers in the *NRG1* locus sequence)
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for *NRG1* [AF491780 and TPA BK000383])
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *NRG1* [MIM 142445], schizophrenia [MIM 181500], and SCZD6 [MIM 603013])
Phred/Phrap/Consed home page, http://bozeman.mbt.washington.edu/index.html
Polyphred, http://www.codoncode.com/polyphred/

References

Bilder RM, Corcoran R, Frith CD (1996) Neuropsychology and neurophysiology in schizophrenia. Curr Opin Psychiatry 9:57–62
Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, et al (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet 20:70–73
Braff DL, Geyer MA (1990) Sensorimotor gating and schizophrenia: human and animal model studies. Arch Gen Psychiatry 47:181–188
Brzustowicz LM, Honer WG, Chow EW, Little D, Hogan J, Hodgkinson K, Bassett AS (1999) Linkage of familial schizophrenia to chromosome 13q32. Am J Hum Genet 65:1096–1103
Cameron JS, Dryer L, Dryer SE (2001) β-Neuregulin-1 is required for the in vivo development of functional Ca^{2+}-activated K^+ channels in parasympathetic neurons. Proc Natl Acad Sci USA 98:2832–2836
Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC, Gottesman II, Farmer AE, McGuffin P, Reveley AM, Murray RM (1999) Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Arch Gen Psychiatry 56:162–168
Carlsson M, Carlsson A (1990) Interactions between glutamatergic and monoaminergic systems within the basal ganglia—implications for schizophrenia and Parkinson's disease. Trends Neurosci 13:272–276
Chen X, Levine L, Kwok PY (1999) Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9:492–498
Chu GC, Moscoso LM, Sliwkowski MX, Merlie JP (1995) Regulation of the acetylcholine receptor subunit gene by recombinant ARIA: an in vitro model for transynaptic gene regulation. Neuron 14:329–339
Clayton D, Jones H (1999) Transmission/disequilibrium tests for extended marker haplotypes. Am J Hum Genet 65:1161–1169
Creese I, Burt DR, Snyder SH (1976) Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. Science 192:481–483
Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38
Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB (2000) Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc Natl Acad Sci USA 97:10483–10488
Erickson SL, O'Shea KS, Ghaboosi N, Loverro L, Frantz G, Bauer M, Lu LH, Moore MW (1997) ErbB3 is required for normal cerebellar and cardiac development: a comparison with ErbB2-and heregulin-deficient mice. Development 124:4999–5011
Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927
Fischbach GD, Rosen KM (1997) ARIA: a neuromuscular junction neuregulin. Annu Rev Neurosci 20:429–458
Freedman R, Hall M, Adler LE, Leonard S (1995) Evidence in postmortem brain tissue for decreased numbers of hippocampal nicotinic receptors in schizophrenia. Biol Psychiatry 38:22–33
Gao XM, Sakai K, Roberts RC, Conley RR, Dean B, Tamminga CA (2000) Ionotropic glutamate receptors and expression of N-methyl-D-aspartate receptor subunits in subregions of human hippocampus: effects of schizophrenia. Am J Psychiatry 157:1141–1149
Garcia RA, Vasudevan K, Buonanno A (2000) The neuregulin receptor ErbB-4 interacts with PDZ-containing proteins at neuronal synapses. Proc Natl Acad Sci USA 97:3596–3601
Gassmann M, Casagranda F, Orioli D, Simon H, Lai C, Klein R, Lemke G (1995) Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. Nature 378:390–394
Gerber DJ, Sotnikova TD, Gainetdinov RR, Huang SY, Caron MG, Tonegawa S (2001) Hyperactivity, elevated dopaminergic transmission, and response to amphetamine in M1 muscarinic acetylcholine receptor-deficient mice. Proc Natl Acad Sci USA 98:15312–15317
Gerlai R, Pisacane P, Erickson S (2000) Heregulin, but not

ErbB2 or ErbB3, heterozygous mutant mice exhibit hyperactivity in multiple behavioral tasks. Behav Brain Res 109: 219–227

Glickstein SB, Schmauss C (2001) Dopamine receptor functions: lessions from knockout mice. Pharmacol Ther 91: 63–83

Goff DC, Coyle JT (2001) The emerging role of glutamate in the pathophysiology and treatment of schizophrenia. Am J Psychiatry 158:1367–1377

Goring HH, Terwilliger JD, Blangero J (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 69:1357–1369

Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, et al (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. Am J Hum Genet 70:593–603

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

Gurling HM, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS, Read T, Murphy P, Blaveri E, McQuillin A, Petursson H, Curtis D (2001) Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. Am J Hum Genet 68:661–673

Hawley M, Kidd K (1995) HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J Hered 86:409–411

Holt DJ, Herman MM, Hyde TM, Kleinman JE, Sinton CM, German DC, Hersh LB, Graybiel AM, Saper CB (1999) Evidence for a deficit in cholinergic interneurons in the striatum in schizophrenia. Neuroscience 94:21–31

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, et al (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nature Genet 26:163–175

Huang YZ, Won S, Ali DW, Wang Q, Tanowitz M, Du QS, Pelkey KA, Yang DJ, Xiong WC, Salter MW, Mei L (2000) Regulation of neuregulin signaling by PSD-95 interacting with ErbB4 at CNS synapses. Neuron 26:443–455

Ibrahim HM, Hogg AJ Jr, Healy DJ, Haroutunian V, Davis KL, Meador-Woodruff JH (2000) Ionotropic glutamate receptor binding and subunit mRNA expression in thalamic nuclei in schizophrenia. Am J Psychiatry 157:1811–1823

Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD, Meyer J, Zambuto CT, Schmitt K, Matise TC, Harkavy Friedman JM, Hampe C, Lee H, Shore D, Wynne D, Faraone SV, Tsuang MT, Cloninger CR (1998) NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. Am J Med Genet 81:282–289

Kendler KS, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Easter SM, Webb BT, Zhang J, Walsh D, Straub RE (1996) Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 153:1534–1540

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

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241–247

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

Lau LF, Huganir RL (1995) Differential tyrosine phosphorylation of N-methyl-D-aspartate receptor subunits. J Biol Chem 270:20036–20041

Levinson DF, Wildenauer DB, Schwab SG, Albus M, Hallmayer J, Lerer B, Maier W, et al (1996) Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. Am J Med Genet 67:580–594

Liddle PF, Carpenter WT, Crow T (1994) Syndromes of schizophrenia: classic literature. Br J Psychiatry 165:721–727

Liu X, Hwang H, Cao L, Buckland M, Cunningham A, Chen J, Chien KR, Graham RM, Zhou M (1998) Domain-specific gene disruption reveals critical regulation of neuregulin signaling by its cytoplasmic tail. Proc Natl Acad Sci USA 95: 13024–13029

Long JC, Williams RC, Urbanek M (1995) An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 56:799–810

Mohn AR, Gainetdinov RR, Caron MG, Koller BH (1999) Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. Cell 98:427–436

O'Donovan MC, Owen MJ (1999) Candidate-gene association studies of schizophrenia. Am J Hum Genet 65:587–592

O'Neill MF, Shaw G (1999) Comparison of dopamine receptor antagonists on hyperlocomotion induced by cocaine, amphetamine, MK-801 and the dopamine D1 agonist C-APB in mice. Psychopharmacology 145:237–250

Ozaki M, Sasner M, Yano R, Lu HS, Buonanno A (1997) Neuregulin-beta induces expression of an NMDA-receptor subunit. Nature 390:691–694

Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin JL, Kimberland M, Babb R, Vourlis S, Chen HM, Lalioti M, Morris MA, Karayiorgou M, Ott J, Meyers D, Antonarakis SE, Housman D, Kazazian HH (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet 60: 252–260

Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

Shaw SH, Kelly M, Smith AB, Shields G, Hopkins PJ, Loftus J, Laval SH, Vita A, De Hert M, Cardon LR, Crow TJ, Sherrington R, DeLisi LE (1998) A genome-wide search for schizophrenia susceptibility genes. Am J Med Genet 81: 364–376

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Spitzer R, Endicott J (eds) (1977) The schedule for affective disorders and schizophrenia, lifetime version. 3rd ed. New York State Psychiatric Institute, New York
Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria: rationale and reliability. Arch Gen Psychiatry 35: 773–782
Steiner H, Blum M, Kitai ST, Fedi P (1999) Differential expression of ErbB3 and ErbB4 neuregulin receptors in dopamine neurons and forebrain areas of the adult rat. Exp Neurol 159:494–503
Tsuang MT, Stone WS, Faraone SV (2001) Genes, environment and schizophrenia. Br J Psychiatry Suppl 40:18–24
Wang JY, Miller SJ, Falls DL (2001) The N-terminal region of neuregulin isoforms determines the accumulation of cell surface and released neuregulin ectodomain. J Biol Chem 276: 2841–2851

nature genetics

The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke

Anna Helgadottir[1], Andrei Manolescu[1], Gudmar Thorleifsson[1], Solveig Gretarsdottir[1], Helga Jonsdottir[1], Unnur Thorsteinsdottir[1], Nilesh J Samani[2], Gudmundur Gudmundsson[1], Struan F A Grant[1], Gudmundur Thorgeirsson[3], Sigurlaug Sveinbjornsdottir[3], Einar M Valdimarsson[3], Stefan E Matthiasson[3], Halldor Johannsson[3], Olof Gudmundsdottir[1], Mark E Gurney[1], Jesus Sainz[1], Margret Thorhallsdottir[1], Margret Andresdottir[1], Michael L Frigge[1], Eric J Topol[4], Augustine Kong[1], Vilmundur Gudnason[5], Hakon Hakonarson[1], Jeffrey R Gulcher[1] & Kari Stefansson[1]

Reprinted from nature genetics, march 2004

nature genetics

ARTICLES

The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke

Anna Helgadottir[1], Andrei Manolescu[1], Gudmar Thorleifsson[1], Solveig Gretarsdottir[1], Helga Jonsdottir[1], Unnur Thorsteinsdottir[1], Nilesh J Samani[2], Gudmundur Gudmundsson[1], Struan F A Grant[1], Gudmundur Thorgeirsson[3], Sigurlaug Sveinbjornsdottir[3], Einar M Valdimarsson[3], Stefan E Matthiasson[3], Halldor Johannsson[3], Olof Gudmundsdottir[1], Mark E Gurney[1], Jesus Sainz[1], Margret Thorhallsdottir[1], Margret Andresdottir[1], Michael L Frigge[1], Eric J Topol[4], Augustine Kong[1], Vilmundur Gudnason[5], Hakon Hakonarson[1], Jeffrey R Gulcher[1] & Kari Stefansson[1]

We mapped a gene predisposing to myocardial infarction to a locus on chromosome 13q12–13. A four-marker single-nucleotide polymorphism (SNP) haplotype in this locus spanning the gene *ALOX5AP* encoding 5-lipoxygenase activating protein (FLAP) is associated with a two times greater risk of myocardial infarction in Iceland. This haplotype also confers almost two times greater risk of stroke. Another *ALOX5AP* haplotype is associated with myocardial infarction in individuals from the UK. Stimulated neutrophils from individuals with myocardial infarction produce more leukotriene B4, a key product in the 5-lipoxygenase pathway, than do neutrophils from controls, and this difference is largely attributed to cells from males who carry the at-risk haplotype. We conclude that variants of ALOX5AP are involved in the pathogenesis of both myocardial infarction and stroke by increasing leukotriene production and inflammation in the arterial wall.

Cardiovascular diseases (CVD) are the leading causes of death and disability in the developed world[1], with an increasing prevalence due to the aging of the population and the obesity epidemic. More than 1 million deaths in the US alone were caused by myocardial infarction and stroke in 2003 (ref. 2). Some of the processes underlying myocardial infarction are now understood: it is generally attributed to atherosclerosis with arterial wall inflammation that ultimately leads to plaque rupture, fissure or erosion[3,4]. This process is known to involve diapedesis of monocytes across the endothelial barrier; activation of neutrophils, macrophage cells and platelets; and release of a variety of cytokines and chemokines[5,6], but the genetic basis of the process has not yet been deciphered.

Two different approaches have been used to search for genes associated with myocardial infarction. SNPs in candidate genes have been tested for association and have, in general, not been replicated or confer only a modest risk of myocardial infarction. Case-control association studies have identified several proinflammatory genes with variants that are associated with either an increased risk of myocardial infarction or a protective effect[7–9]. Four genome-wide scans in families with myocardial infarction have yielded several loci with formidable linkage peaks, but the gene(s) underlying these loci have not yet been identified[10–14]. In addition, one large pedigree study identified a deletion mutation of a transcription factor gene, *MEF2A*, with autosomal dominant transmission[14]. This is an interesting cause of myocardial infarction, but the prevalence of this or other mutations in *MEF2A* outside this family remains to be determined.

Here we report a genome-wide scan of 296 multiplex Icelandic families including 713 individuals with myocardial infarction. Through suggestive linkage to a locus on chromosome 13q12–13, we identified the gene (*ALOX5AP*) encoding FLAP and found that a four-SNP haplotype in the gene confers a nearly two times greater risk of myocardial infarction and stroke. FLAP is a regulator[15] of a crucial pathway in the genesis of leukotriene inflammatory mediators, which are implicated in atherosclerosis both in a mouse model[16] and in human studies[17,18]. Males had the strongest association to the at-risk haplotype, and male carriers of the at-risk haplotype also had significantly greater production of leukotriene-B4 (LTB4), supporting the idea that proinflammatory activity has a role in the pathogenesis of myocardial infarction. We confirmed the association of *ALOX5AP* with myocardial infarction in an independent cohort of British individuals with another haplotype. These results indicate that *ALOX5AP* is the first specific gene isolated that confers substantial population-attributable risk (PAR) of the complex traits of both myocardial infarction and stroke.

[1]deCODE genetics, Sturlugata 8, Reykjavik, Iceland. [2]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. [3]National University Hospital, Reykjavik, Iceland. [4]Cleveland Clinic Foundation, Cleveland, Ohio, USA. [5]Icelandic Heart Association, Reykjavik, Iceland. Correspondence should be addressed to K.S. (kstefans@decode.is).

Published online 8 February 2004; doi:10.1038/ng1311

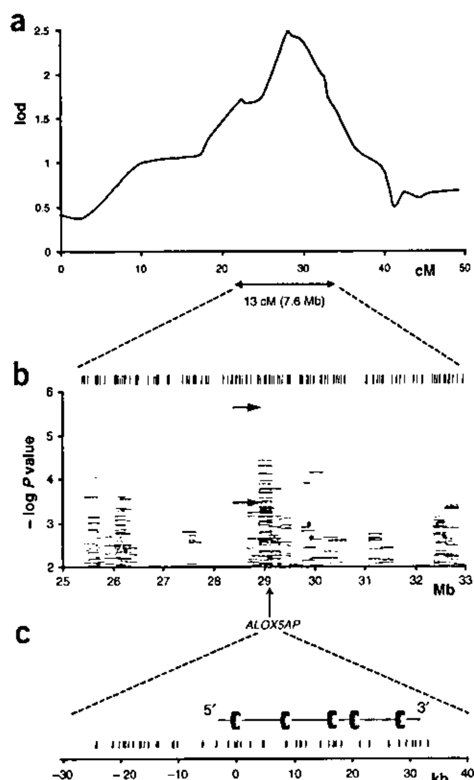


Figure 1 Schematic view of the chromosome 13 linkage region showing *ALOX5AP*. (**a**) The linkage scan for females with myocardial infarction and the one-lod drop region that includes *ALOX5AP*. (**b**) Microsatellite association for all individuals with myocardial infarction: single-marker association (black dots) and two-, three-, four- and five-marker haplotype association (black, blue, green and red horizontal lines, respectively). The blue and red arrows indicate the location of the most significant haplotype association across *ALOX5AP* in males and females, respectively. (**c**) *ALOX5AP* gene structure, with exons shown as colored cylinders, and the locations of all SNPs typed in the region. The green vertical lines indicate the position of the microsatellites (**b**) and SNPs (**c**) used in the analysis.

RESULTS

Linkage analysis

We carried out a genome-wide scan in search of myocardial infarction susceptibility genes using a framework set of 1,068 microsatellite markers. The initial linkage analysis included 713 individuals with myocardial infarction who fulfilled the World Health Organization (WHO) MONICA research criteria[19] and were clustered in 296 extended families. We repeated the linkage analysis for individuals with early onset, for males and for females separately. A description of the number of affected individuals and families in each analysis is provided in **Supplementary Table 1** online, and the corresponding allele-sharing lod scores are given in **Supplementary Figure 1** online. None of these analyses yielded a locus of genome-wide significance. The most promising lod score (2.86) was observed on chromosome 13q12–13 for linkage with females with myocardial infarction at the peak marker *D13S289* (**Supplementary Fig. 1** online). This locus also had the most promising lod score (2.03) for individuals with early-onset myocardial infarction. After we increased the information on identity-by-descent sharing to over 90% by typing an additional 14 microsatellite markers in a 30-cM region around *D13S289*, the lod score for the association in females dropped to 2.48 ($P = 0.00036$), and the lod score remained highest at *D13S289* (**Fig. 1a**). In an independent linkage study of males with ischemic stroke or transient ischemic attack (TIA), we observed linkage to the same locus with a lod score of 1.51 at the same peak marker (**Supplementary Fig. 2** online), further suggesting that a cardiovascular susceptibility factor might reside at this locus.

Microsatellite association study

The 7.6-Mb region that corresponds to a drop of 1 in lod score in the female–myocardial infarction linkage analysis contains 40 known genes (**Supplementary Table 2** online). To determine which gene in this region was most likely to contribute to myocardial infarction, we typed 120 microsatellite markers in the region and carried out a case-control association study using 802 unrelated (separated by at least three meioses) individuals with myocardial infarction and 837 population-based controls. We also repeated the association study for each of the three phenotypes that were used in the linkage study: individuals with early onset, males and females with myocardial infarction. In addition to testing each marker individually, we also tested haplotypes based on these markers for association. To limit the number of haplotypes tested, we considered only haplotypes spanning less than 300 kb that were over-represented among the affected individuals.

The haplotype with the strongest association to myocardial infarction ($P = 0.00004$) covered a region that contains two known genes: *ALOX5AP* (**Fig. 1b**) and a gene with an unknown function called highly charged protein (*D13S106E*). The haplotype association in this region for females with myocardial infarction was less significant ($P = 0.0004$) than for all individuals with myocardial infarction, and the most significant haplotype association was observed for males with myocardial infarction ($P = 0.000002$). The haplotype associated with males with myocardial infarction was the only haplotype that retained significant association after adjusting for all haplotypes tested.

FLAP, together with 5-lipoxygenase (5-LO), is a regulator of the leukotriene biosynthetic pathway that has recently been implicated in the pathogenesis of atherosclerosis[16–18]. Therefore, *ALOX5AP* was a good candidate for the gene underlying the association with myocardial infarction.

Screening for SNPs in *ALOX5AP* and LD mapping

To determine whether variations in *ALOX5AP* significantly associate with myocardial infarction and to search for causal variations, we sequenced *ALOX5AP* in 93 affected individuals and 93 controls. The sequenced region covers 60 kb containing *ALOX5AP*, including the five known exons and introns, the 26-kb region 5′ to the first exon and the 7-kb region 3′ to the fifth exon. We identified 144 SNPs, of which we excluded 96 from further analysis owing to either a low minor allele frequency or complete correlation (redundancy) with other SNPs. **Figure 1c** shows the distribution of the 48 SNPs chosen for genotyping, relative to exons, introns and the 5′ and 3′ flanking regions of *ALOX5AP*. We identified only one SNP in a coding sequence (exon 2), which did not lead to an amino acid substitution. The locations of the 48 SNPs in the National Center for Biotechnology Information human genome assembly build 34 are listed in **Supplementary Table 3** online. In addition to the SNPs, we typed a polymorphism consisting of a monopolymer A repeat in the *ALOX5AP* promoter region[20].

The linkage disequilibrium (LD) block structure defined by the 48 genotyped SNPs is shown in **Figure 2**. Strong LD was detected across the *ALOX5AP* region, although at least one historical recombination seems to have occurred, dividing the region into two strongly correlated LD blocks.

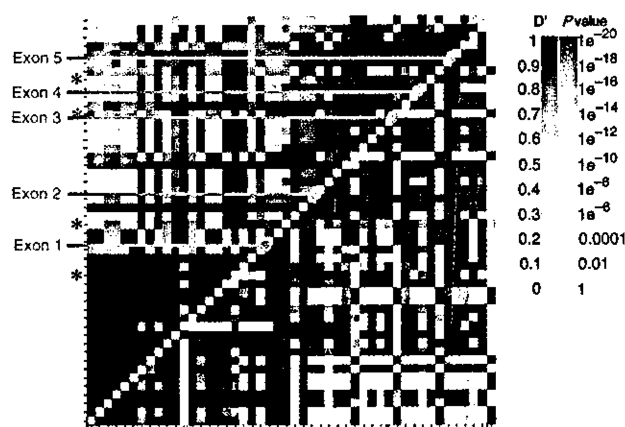


Figure 2 Pairwise LD between SNPs in a 60-kb region encompassing *ALOX5AP*. The markers are plotted equidistantly. Two measures of LD are shown: D' in the upper left triangle and P values in the lower right triangle. Colored lines indicate the positions of the exons of *ALOX5AP*, and the green stars indicate the location of the markers of the at-risk haplotype HapA. Scales for both measures of the LD strength are provided on the right.

Haplotype association with myocardial infarction

In a case-control association study, we genotyped the 48 selected SNPs and the monopolymer A repeat marker in a set of 779 unrelated individuals with myocardial infarction and 624 population-based controls. We tested each of the 49 markers individually for association with the disease. Three SNPs, one located 3 kb upstream of the first exon and the other two 1 kb and 3 kb downstream of the first exon, showed nominally significant association to myocardial infarction (**Supplementary Table 4** online). After adjusting for the number of markers tested, however, these results were not significant. We then searched for haplotypes associated with the disease using the same cohorts. We limited the search to haplotype combinations constructed from two, three or four SNPs and tested only haplotypes that were over-represented in the individuals with myocardial infarction. The resulting P values were adjusted for all the haplotypes we tested by randomizing the affected individuals and controls.

Several haplotypes were significantly associated with the disease at an adjusted significance level of $P < 0.05$ (**Supplementary Table 5** online). We observed the most significant association with a four-SNP haplotype spanning 33 kb, including the first four exons of *ALOX5AP* (**Fig. 1c**), with a nominal P value of 0.0000023 and an adjusted P value of 0.005. This haplotype, called HapA, has a haplotype frequency of 15.8% (carrier frequency 29.1%) in affected individuals versus 9.5% (carrier frequency 18.1%) in controls (**Table 1**). The relative risk conferred by HapA compared with other haplotypes constructed from the same SNPs, assuming a multiplicative model, was 1.8 and the corresponding PAR was 13.5%. HapA was present at a higher frequency in males (carrier frequency 30.9%) than in females with myocardial infarction (carrier frequency 25.7%; **Table 1**). All other haplotypes that were significantly associated with an adjusted P value less than 0.05 were highly correlated with HapA and should be considered variants of that haplotype (**Supplementary Table 5** online).

Association of HapA with stroke and PAOD

Because of the high degree of comorbidity among myocardial infarction, stroke and peripheral arterial occlusive disease (PAOD), with most of these cases occurring on the basis of an atherosclerotic disease, we wanted to determine whether HapA was also associated with stroke or PAOD. We typed the SNPs defining HapA for these cohorts. We removed first- and second-degree relatives and all known cases of myocardial infarction and tested for association in 702 individuals with stroke and 577 individuals with PAOD (**Table 1**). We observed a significant association of HapA with stroke, with a relative risk of 1.67 ($P = 0.000095$). In addition, we determined whether HapA was primarily associated with a particular subphenotype of stroke and found that both ischemic and hemorrhagic stroke were significantly associated with HapA (**Supplementary Table 6** online). Finally, although HapA was more frequent in the PAOD cohort than in the population controls (**Table 1**), this was not significant. Similar to the stronger association of HapA with males with myocardial infarction than with females with myocardial infarction, HapA also showed stronger association with males than with females with stroke and PAOD (**Table 1**).

Haplotype association in a British cohort

In an independent study, we determined whether variants in *ALOX5AP* also affected the risk of myocardial infarction in a population outside Iceland. We typed SNPs defining HapA in a cohort of 753 individuals from the UK who had sporadic myocardial infarction and in 730 British population controls. The affected individuals and controls were from three separate study cohorts recruited in Leicester and Sheffield. We found a slightly higher frequency of HapA in affected individuals versus controls (16.8% versus 15.1%, respectively), but the results were not statistically significant. As in the Icelandic population, HapA was more common in males with myocardial infarction (carrier frequency 31.7%) than in females with myocardial infarction (carrier frequency 28.0%). When we typed an additional nine SNPs, distributed across *ALOX5AP*, in the British cohort and searched for other haplotypes that might be associated with myocardial infarction, two SNPs showed association to myocardial infarction with a nominally significant P value (data not shown). Moreover, three- and four-SNP haplotype combinations were associated with higher risk of myocardial infarction in the British cohort, and we observed the most signifi-

Table 1 Association of HapA with myocardial infarction, stroke and PAOD

Phenotype (n)	Frequency	RR	PAR	P value	P value[a]
Myocardial infarction (779)	0.158	1.80	0.135	0.0000023	0.005
Males (486)	0.169	1.95	0.158	0.00000091	ND
Females (293)	0.138	1.53	0.094	0.0098	ND
Early onset (358)	0.139	1.53	0.094	0.0058	ND
Stroke (702)[b]	0.149	1.67	0.116	0.000095	ND
Males (373)	0.156	1.76	0.131	0.00018	ND
Females (329)	0.141	1.55	0.098	0.0074	ND
PAOD (577)[b]	0.122	1.31	0.056	0.061	ND
Males (356)	0.126	1.36	0.065	0.057	ND
Females (221)	0.114	1.22	0.041	0.31	ND

[a] P value adjusted for the number of haplotypes tested. [b] Excluding known cases of myocardial infarction.

Shown is HapA of *ALOX5AP* and the corresponding number of affected individuals (n), the haplotype frequency in affected individuals, the relative risk (RR), PAR and P values. HapA is defined by the SNPs SG13S25, SG13S114, SG13S89 and SG13S32 (**Supplementary Table 5** online). The same controls ($n = 624$) were used for the association analysis in myocardial infarction, stroke and PAOD as well as for the analysis of males, females and individuals with early onset. The frequency of HapA in the control cohort is 0.095. ND, not done.

Table 2 Association of HapB with myocardial infarction in British individuals

Phenotype (*n*)	Frequency	RR	PAR	*P* value	*P* value[a]
Myocardial infarction (753)	0.075	1.95	0.072	0.00037	0.046
Males (549)	0.075	1.97	0.072	0.00093	ND
Females (204)	0.073	1.90	0.068	0.021	ND

[a] *P* value adjusted for the number of haplotypes tested using 1,000 randomization tests.

Shown are the results for HapB that shows the strongest association in the British myocardial infarction cohort. HapB is defined by the SNPs SG13S377, SG13S114, SG13S41 and SG13S35, which have the alleles A, A, A and G, respectively. In all three phenotypes shown, the same set of 730 British controls was used and the frequency of HapB in the control cohort is 0.040. Number of affected individuals (*n*), haplotype frequency in affected individuals, relative risk (RR) and PAR are indicated. ND, not done.

cant association for a four-SNP haplotype with a nominal *P* value of 0.00037 (**Table 2**). We call this haplotype HapB. The haplotype frequency of HapB was 7.5% in the individuals with myocardial infarction (carrier frequency 14.4%) compared with 4.0% (carrier frequency 7.8%) in controls, conferring a relative risk of 1.95 (**Table 2**). This association of HapB remained significant after adjusting for all haplotypes tested, using 1,000 randomization steps, with an adjusted $P = 0.046$. No other SNP haplotype had an adjusted *P* value <0.05. The two at-risk haplotypes, HapA and HapB, are mutually exclusive; there are no instances in which the same chromosome carries both haplotypes.

More LTB4 in individuals with myocardial infarction

To determine whether individuals with a past history of myocardial infarction had greater activity of the 5-LO pathway than controls, we measured production of LTB4 (a key product of the 5-LO pathway) in blood neutrophils isolated from Icelandic individuals with myocardial infarction and controls before and after stimulation with the calcium ionophore ionomycin. We detected no difference in LTB4 production in resting neutrophils from individuals with myocardial infarction versus controls. In contrast, LTB4 generation by neutrophils stimulated with ionomycin was substantially greater in individuals with myocardial infarction than in controls after 15 and 30 min, respectively (**Fig. 3a**). Moreover, the observed difference in release of LTB4 was largely accounted for by male carriers of HapA (**Fig. 3b**), whose cells produced significantly more LTB4 than cells from controls ($P = 0.0042$; **Supplementary Table 7** online). There was also a heightened LTB4 response in males who did not carry HapA, but this difference was of borderline significance (**Supplementary Table 7** online). This could be explained by additional variants in *ALOX5AP* that have not been uncovered, or in other genes belonging to the 5-LO pathway, that may account for upregulation of the LTB4 response in some individuals without the *ALOX5AP* at-risk haplotype. We did not detect differences in LTB4 response in females (**Supplementary Table 7** online), but because of the small sample size, this result is not conclusive. The elevated levels of LTB4 production in stimulated neutrophils from male carriers of the at-risk haplotype suggest that the disease-associated variants of *ALOX5AP* heighten the response of FLAP to factors that stimulate inflammatory cells.

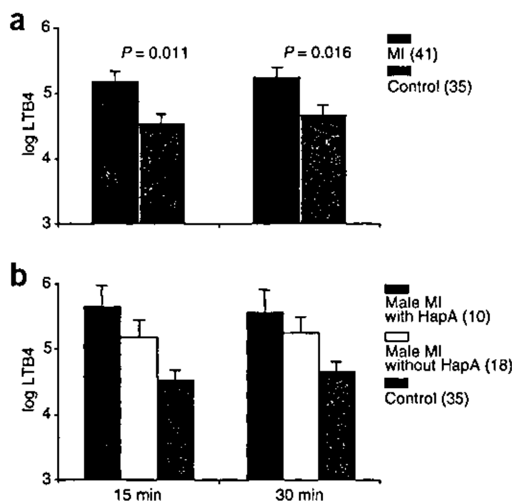


Figure 3 LTB4 production of ionomycin-stimulated neutrophils from individuals with myocardial infarction ($n = 41$) and controls ($n = 35$). The log-transformed (mean ± s.d.) values measured at 15 and 30 min in stimulated cells are shown. (**a**) LTB4 production in individuals with myocardial infarction (MI) and controls. The difference in the mean values between affected individuals and controls was tested using a two-sample *t*-test of the log-transformed values. (**b**) LTB4 production in males with myocardial infarction carrying HapA (red bars) and not carrying HapA (white bars). Mean values of controls (blue bars) are included for comparison. Males with HapA produced the highest amounts of LTB4 ($P < 0.005$ compared with controls). Data for females are shown in **Supplementary Table 7** online.

DISCUSSION

Our results show that variants of *ALOX5AP* encoding FLAP are associated with greater risk of myocardial infarction and stroke. In our Icelandic cohort, a haplotype that spans *ALOX5AP* is carried by 29.1% of all individuals with myocardial infarction and almost doubles the risk of myocardial infarction. We then replicated these findings in an independent cohort of individuals with stroke. Furthermore, stimulated neutrophils from individuals with myocardial infarction had greater production of LTB4, one of the key products of the 5-LO pathway. When we examined this in the context of the at-risk haplotype, however, the gain of function was largely attributed to male carriers of the at-risk haplotype, who also had the strongest association with the *ALOX5AP* haplotype. Another haplotype spanning *ALOX5AP* was associated with myocardial infarction in a British cohort. Although the pathogenic variants responsible for the effects associated with the disease haplotypes are unknown, the greater production of LTB4 observed in ionomycin-stimulated neutrophils from male carriers of the at-risk haplotype suggests that the disease-associated variants increase the response of FLAP to factors that stimulate inflammatory cells.

We observed suggestive linkage to chromosome 13q12–13 with several different phenotypic groups, including females with myocardial infarction, individuals of both sexes with early-onset myocardial infarction and males with ischemic stroke or TIA. But we observed the strongest haplotype association for males with myocardial infarction or stroke. Therefore, the linkage signal in females with myocardial infarction and in individuals with early-onset myocardial infarction is not explained by the at-risk haplotype that we identified, and we expect that there may be other unidentified variants or haplotypes in *ALOX5AP*, or in other genes in the linkage region, that may confer risk of these cardiovascular phenotypes. These variants are probably rarer than HapA with relatively high penetrance, higher in women than in men.

FLAP has an important role in the initial steps of leukotriene biosynthesis[15], which is largely confined to leukocytes and can be

triggered by a variety of stimuli. In this biosynthetic pathway, unesterified arachidonic acid is converted to LTA4 by the action of 5-LO and its activating protein FLAP. The unstable epoxide LTA4 is further metabolized to LTB4 or LTC4 by LTA4 hydrolase and LTC4 synthase, respectively. In addition, LTA4 can be exported to neighboring cells that are devoid of 5-LO activity and become subject to transcellular leukotriene biosynthesis[21–23]. The leukotrienes have a variety of proinflammatory effects[24,25]. LTB4 activates leukocytes, leading to chemotaxis and increased adhesion of leukocytes to vascular endothelium, release of lysosomal enzymes such as myeloperoxidase and production of superoxide anions[25]. The cysteinyl-containing leukotrienes (LTC4 and its metabolites LTD4 and LTE4) increase vascular permeability in postcapillary venules and are potent vasoconstrictors of coronary arteries[26–28].

The importance of the 5-LO pathway is well established in asthma, and drugs inhibiting this pathway have been developed for treating asthma. The role of the 5-LO pathway in the pathogenesis of atherosclerosis has recently received attention. A study of postmortem pathologic specimens showed an increase in the expression of members of the 5-LO pathway, including 5-LO and FLAP, in atherosclerotic lesions at various stages of development in the aorta, coronary arteries and carotid arteries[18]. Furthermore, 5-LO was localized to macrophages, dendritic cells, foam cells, mast cells and neutrophilic granulocytes, and the number of cells expressing 5-LO was markedly greater in advanced lesions[18]. The leukocytes positive for 5-LO accumulated at distinct sites that are most prone to rupture[29], such as the shoulder regions below the fibrous cap of the atherosclerotic lesion[18]. A 5-LO promoter variant is associated with abnormal carotid artery intima-media thickness and heightened inflammatory biomarkers[30]. In addition, antagonists of LTB4 block the development of atherosclerosis in apo-E-deficient and LDRL-deficient mice[31], and a congenic mouse strain with a heterozygous deficiency of 5-LO shows resistance to atherosclerosis[16], further supporting the idea that greater activity of the 5-LO pathway has a role in predisposition to atherosclerosis.

Our data also show that the at-risk haplotype of *ALOX5AP* has higher frequency in all subgroups of stroke, including ischemic stroke, TIA and hemorrhagic stroke. HapA confers significantly higher risk of myocardial infarction and stroke than it does of PAOD. This could be explained by differences in the pathogenesis of these diseases. Unlike individuals with PAOD, who have ischemic legs because of atherosclerotic lesions that are responsible for gradually diminishing blood flow to the legs, individuals with myocardial infarction and stroke have suffered acute events, with disruption of the vessel wall suddenly decreasing blood flow to regions of the heart and the brain.

We did not find association between HapA and myocardial infarction in a British cohort, but we did find significant association between myocardial infarction and a different *ALOX5AP* variant. The existence of different haplotypes of the gene conferring risk to myocardial infarction in different populations is not unexpected. It is not unreasonable to assume that a common disease like myocardial infarction is associated with many different mutations or sequence variations and that the frequencies of these disease-associated variants may differ between populations. It would also not be unexpected for the same mutation to arise on different haplotypic backgrounds.

Our work suggests that *ALOX5AP* has an important role in the pathogenesis of myocardial infarction and stroke in humans. Our study, together with others, may provide the necessary background to launch therapeutic trials to determine whether pharmacological inhibition of FLAP will prevent the development of myocardial infarction and stroke.

METHODS

Study population. We recruited the individuals in the study from a registry of over 8,000 individuals, which includes all individuals who had myocardial infarctions before the age of 75 in Iceland from 1981 to 2000. This registry is a part of the WHO MONICA Project[19]. Diagnoses of all individuals in the registry follow strict diagnostic rules based on signs, symptoms, electrocardiograms, cardiac enzymes and necropsy findings.

We used genotypes from 713 individuals with myocardial infarction and 1,741 of their first-degree relatives in the linkage analysis. For the microsatellite association study of the locus associated with myocardial infarction, we used 802 unrelated (no first- or second-degree relatives) individuals with myocardial infarction (233 females, 624 males and 302 with early onset) and 837 population-based controls. The females studied were post-menopausal. Over 90% of the individuals were taking aspirin or other nonsteroidal anti-inflammatory drugs. For the SNP association study in and around *ALOX5AP*, we genotyped 779 unrelated individuals with myocardial infarction (293 females, 486 males and 358 with early onset). The control group for the SNP association study was population-based and comprised of 624 unrelated males and females 20–90 years of age whose medical history was unknown. The stroke and PAOD cohorts used in this study have previously been described[32–34]. For the stroke linkage analysis, we used genotypes from 342 males with ischemic stroke or TIA that were linked to at least one other male within and including six meioses in 164 families. For the association studies, we analyzed 702 individuals with all forms of stroke (329 females and 373 males) and 577 individuals with PAOD (221 females and 356 males). Individuals with stroke or PAOD who also had myocardial infarction were excluded. Controls used for the stroke and PAOD association studies were the same as used in the myocardial infarction SNP association study.

The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. We obtained informed consent from all study participants. Personal identifiers associated with medical information and blood samples were encrypted with a third-party encryption system as previously described[35].

Statistical analysis. We carried out a genome-wide scan as previously described[33], using a set of 1,068 microsatellite markers. We used multipoint, affected-only allele-sharing methods[36] to assess the evidence for linkage. All results were obtained using the program Allegro[37] and the deCODE genetic map[38]. We used the S_{pairs} scoring function[39,40] and the exponential allele-sharing model[36] to generate the relevant 1-degree-of-freedom statistics. When combining the family scores to obtain an overall score, we used a weighting scheme that is halfway on a log scale between weighting each affected pair equally and weighting each family equally. In the analysis, all genotyped individuals who were not affected were treated as 'unknown'. Because of concern with small-sample behavior, we usually computed corresponding *P* values in two different ways for comparison and report the less significant one. The first *P* value was computed based on large sample theory, $Z_{lr} = \sqrt{(2 \log_e (10) \text{ lod})}$, and is distributed approximately as a standard normal distribution under the null hypothesis of no linkage[36]. A second *P* value was computed by comparing the observed lod score with its complete data sampling distribution under the null hypothesis[37]. When a data set consisted of more than a handful of families, these two *P* values tended to be very similar. The information measure we used, which is implemented in Allegro, is closely related to a classical measure of information and has a property that is between 0 (if the marker genotypes are completely uninformative) and 1 (if the genotypes determine the exact amount of allele sharing by descent among the affected relatives)[41,42].

For single-marker association studies, we used Fisher's exact test to calculate two-sided *P* values for each allele. All *P* values are unadjusted for multiple comparisons unless specifically indicated. We present allelic rather than carrier frequencies for microsatellites, SNPs and haplotypes. To minimize any bias due to the relatedness of the individuals who were recruited as families for the linkage analysis, we eliminated first- and second-degree relatives. For the haplotype analysis we used the program NEMO[32], which handles missing genotypes and uncertainty with phase through a likelihood procedure, using the expectation-maximization algorithm as a computational tool to estimate haplotype frequencies. Under the null hypothesis, the affected individuals and controls were assumed to have identical haplotype frequencies. Under the alternative

hypotheses, the candidate at-risk haplotype was allowed to have a higher frequency in the affected individuals than in controls, and the ratios of frequencies of all other haplotypes were assumed to be the same in both groups. Likelihoods were maximized separately under both hypotheses, and a corresponding 1-degree-of-freedom likelihood ratio statistic was used to evaluate statistical significance[32]. Although we only searched for haplotypes that increased the risk, all reported P values are two-sided unless otherwise stated. To assess the significance of the haplotype association corrected for multiple testing, we carried out a randomization test using the same genotype data. We randomized the cohorts of affected individuals and controls and repeated the analysis. This procedure was repeated up to 1,000 times, and the P value we present is the fraction of replications that produced a P value for a haplotype tested that was lower than or equal to the P value we observed using the original affected individual and control cohorts.

For both single-marker and haplotype analysis, we calculated relative risk (RR) and PAR assuming a multiplicative model[43,44] in which the risk of the two alleles of haplotypes a person carries multiply. We calculated LD between pairs of SNPs using the standard definition of D' (ref. 45) and R^2 (ref. 46). Using NEMO, we estimated frequencies of the two marker allele combinations by maximum likelihood and evaluated deviation from linkage equilibrium by a likelihood ratio test. When plotting all SNP combinations to elucidate the LD structure in a particular region, we plotted D' in the upper left corner and the P value in the lower right corner. In the LD plots we present, the markers are plotted equidistantly rather than according to their physical positions.

Identification of DNA polymorphisms. We identified new polymorphic repeats (dinucleotide or trinucleotide repeats) with the Sputnik program. We subtracted the lower allele of the CEPH sample 1347-02 (CEPH genomics repository) from the alleles of the microsatellites and used it as a reference. We detected SNPs in the gene by PCR sequencing exonic and intronic regions from affected individuals and controls. We also detected public polymorphisms by BLAST search of the National Center for Biotechnology Information SNP database. We genotyped SNPs using a method for detecting SNPs with fluorescent polarization template-directed dye-terminator incorporation[47] and TaqMan assays (Applied Biosystems).

Isolation and activation of peripheral blood neutrophils. We drew 50 ml of blood from each of 41 individuals with myocardial infarction and 35 age- and sex-matched controls into vacutainers containing EDTA. All blood was drawn at the same time in the early morning after 12 h of fasting. We isolated neutrophils using Ficoll-Paque PLUS (Amersham Biosciences).

We collected the red cell pellets from the Ficoll gradient and then lysed red blood cells in 0.165 M ammonium chloride for 10 min on ice. After washing them with phosphate-buffered saline, we counted neutrophils and plated them at 2×10^6 cells ml^{-1} in 4-ml cultures of 15% fetal calf serum (GIBCO BRL) in RPMI-1640 medium (GIBCO BRL). We then stimulated cells with maximum effective concentration of ionomycin (1 μM). At 0, 15, 30, 60 min after adding ionomycin, we aspirated 600 μl of culture medium and stored it at −80 °C for the measurement of LTB4 release as described below. We maintained cells at 37 °C in a humidified atmosphere of 5% carbon dioxide–95% air. We treated all samples with indomethasine (1 μM) to block the cyclooxygenase enzyme.

Ionomycin-induced release of LTB4 in neutrophils. We used the LTB4 Immunoassay (R&D systems) to quantify LTB4 concentration in supernatant from cultured ionomycin-stimulated neutrophils. The assay we used is based on the competitive binding technique in which LTB4 present in the testing samples (200 μl) competes with a fixed amount of alkaline phosphatase–labeled LTB4 for sites on a rabbit polyclonal antibody. During the incubation, the polyclonal antibody becomes bound to a goat antibody to rabbit coated onto the microplates. After washing to remove excess conjugate and unbound sample, a substrate solution was added to the wells to determine the bound enzyme activity. We stopped the color development and read the absorbance at 405 nm. The intensity of the color is inversely proportional to the concentration of LTB4 in the sample. Each LTB4 measurement using the LTB4 Immunoassay was done in duplicate.

British study population. We recruited three separate British cohorts as described previously[48,49]. The first two cohorts comprised 549 individuals from among those who were admitted to the coronary care units of the Leicester Royal Infirmary, Leicester (July 1993–April 1994), and the Royal Hallamshire Hospital, Sheffield (November 1995–March 1997), and satisfied the WHO criteria for acute myocardial infarction in terms of symptoms, elevations in cardiac enzymes or electrocardiographic changes[50]. We recruited 532 control individuals in each hospital from adult visitors of individuals with noncardiovascular disease on general medical, surgical, orthopedic and obstetric wards to find subjects representative of the source population from which the affected individuals originated. Individuals who reported a history of coronary heart disease were excluded.

In the third cohort, we recruited 204 individuals retrospectively from the registries of three coronary care units in Leicester. All had suffered a myocardial infarction according to WHO criteria before the age of 50 years. At the time of participation, individuals were at least 3 months from the acute event. The control cohort comprised 198 individuals with no personal or family history of premature coronary heart disease, matched for age, sex and current smoking status with the cases. We recruited control individuals from three primary care practices located in the same geographical area. In all cohorts, individuals were white of Northern European origin. Local research ethics committees approved all the studies, and individuals provided written informed consent for use of samples in genetic studies of coronary artery disease.

URLs. The Sputnik program is available at http://espressosoftware.com/pages/sputnik.jsp. The National Center for Biotechnology Information SNP database is available at http://www.ncbi.nlm.nih.gov/SNP/index.html.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the affected individuals and their families whose contribution made this study possible and the nurses at the Icelandic Heart Association, personnel at the deCODE core facilities, T. Jonsdottir, F. Runarsson, E. Palsdottir, J. Kostic, K. Channer, R. Steeds, R. Singh and P. Braund for their contributions. N.J.S. is supported by the British Heart Foundation.

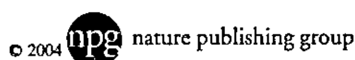
COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 7 January; accepted 26 January 2004
Published online at http://www.nature.com/naturegenetics/

1. Bonow, R.O., Smaha, L.A., Smith, S.C. Jr., Mensah, G.A. & Lenfant, C. World Heart Day 2002: the international burden of cardiovascular disease: responding to the emerging global epidemic. *Circulation* **106**, 1602–1605 (2002).
2. *Heart Disease and Stroke Statistics, 2003 Update* (American Heart Association, Dallas, Texas, 2002).
3. Lusis, A.J. Atherosclerosis. *Nature* **407**, 233–241 (2000).
4. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).
5. Stratford, N., Britten, K. & Gallagher, P. Inflammatory infiltrates in human coronary atherosclerosis. *Atherosclerosis* **59**, 271–276 (1986).
6. Poole, J.C. & Florey, H.W. Changes in the endothelium of the aorta and the behaviour of macrophages in experimental atheroma of rabbits. *J. Pathol. Bacteriol.* **75**, 245–251 (1958).
7. Topol, E.J. *et al.* Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. *Circulation* **104**, 2641–2644 (2001).
8. Ozaki, K. *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
9. Yamada, Y. *et al.* Prediction of the risk of myocardial infarction from polymorphisms in candidate genes. *N. Engl. J. Med.* **347**, 1916–1923 (2002).
10. Broeckel, U. *et al.* A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat. Genet.* **30**, 210–214 (2002).
11. Francke, S. *et al.* A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum. Mol. Genet.* **10**, 2751–2765 (2001).
12. Harrap, S.B. *et al.* Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. *Arterioscler. Thromb. Vasc. Biol.* **22**, 874–878 (2002).
13. Pajukanta, P. *et al.* Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland. *Am. J. Hum. Genet.* **67**, 1481–1493 (2000).
14. Wang, L., Fan, C., Topol, S.E., Topol, E.J. & Wang, Q. Mutation of MEF2A in an inherited disorder with features of coronary artery disease. *Science* **302**, 1578–1581 (2003).

15. Dixon, R.A. *et al.* Requirement of a 5-lipoxygenase-activating protein for leukotriene synthesis. *Nature* **343**, 282–284 (1990).
16. Mehrabian, M. *et al.* Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91**, 120–126 (2002).
17. Brezinski, D.A., Nesto, R.W. & Serhan, C.N. Angioplasty triggers intracoronary leukotrienes and lipoxin A4. Impact of aspirin therapy. *Circulation* **86**, 56–63 (1992).
18. Spanbroek, R. *et al.* Expanding expression of the 5-lipoxygenase pathway within the arterial wall during human atherogenesis. *Proc. Natl. Acad. Sci. USA* **100**, 1238–1243 (2003).
19. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J. Clin. Epidemiol.* **41**, 105–14 (1988).
20. Koshino, T. *et al.* Novel polymorphism of the 5-lipoxygenase activating protein (FLAP) promoter gene associated with asthma. *Mol. Cell. Biol. Res. Commun.* **2**, 32–35 (1999).
21. Sala, A., Bolla, M., Zarini, S., Muller-Peddinghaus, R. & Folco, G. Release of leukotriene A4 versus leukotriene B4 from human polymorphonuclear leukocytes. *J. Biol. Chem.* **271**, 17944–17948 (1996).
22. Dahinden, C.A., Clancy, R.M., Gross, M., Chiller, J.M. & Hugli, T.E. Leukotriene C4 production by murine mast cells: evidence of a role for extracellular leukotriene A4. *Proc. Natl. Acad. Sci. USA* **82**, 6632–6636 (1985).
23. Fiore, S. & Serhan, C.N. Formation of lipoxins and leukotrienes during receptor-mediated interactions of human platelets and recombinant human granulocyte/macrophage colony-stimulating factor-primed neutrophils. *J. Exp. Med.* **172**, 1451–1457 (1990).
24. Ford-Hutchinson, A.W. Leukotriene B4 in inflammation. *Crit. Rev. Immunol.* **10**, 1–12 (1990).
25. Samuelsson, B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science* **220**, 568–575 (1983).
26. Burke, J.A., Levi, R., Guo, Z.G. & Corey, E.J. Leukotrienes C4, D4 and E4: effects on human and guinea-pig cardiac preparations in vitro. *J. Pharmacol. Exp. Ther.* **221**, 235–241 (1982).
27. Roth, D.M. & Lefer, A.M. Studies on the mechanism of leukotriene induced coronary artery constriction. *Prostaglandins* **26**, 573–581 (1983).
28. Wargovich, T., Mehta, J., Nichols, W.W., Pepine, C.J. & Conti, C.R. Reduction in blood flow in normal and narrowed coronary arteries of dogs by leukotriene C4. *J. Am. Coll. Cardiol.* **6**, 1047–1051 (1985).
29. Falk, E., Shah, P.K. & Fuster, V. Coronary plaque disruption. *Circulation* **92**, 657–671 (1995).
30. Dwyer, J.H. *et al.* Arachidonate 5-Lipoxygenase Promoter Genotype, Dietary Arachidonic Acid, and Atherosclerosis. *N. Engl. J. Med.* **350**, 29–37 (2004).
31. Aiello, R.J. *et al.* Leukotriene B4 receptor antagonism reduces monocytic foam cells in mice. *Arterioscler. Thromb. Vasc. Biol.* **22**, 443–449 (2002).
32. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.* **35**, 131–138 (2003).
33. Gretarsdottir, S. *et al.* Localization of a susceptibility gene for common forms of stroke to 5q12. *Am. J. Hum. Genet.* **70**, 593–603 (2002).
34. Gudmundsson, G. *et al.* Localization of a gene for peripheral arterial occlusive disease to chromosome 1p31. *Am. J. Hum. Genet.* **70**, 586–592 (2002).
35. Gulcher, J.R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur. J. Hum. Genet.* **8**, 739–742 (2000).
36. Kong, A. & Cox, N.J. Allele-sharing models: LOD scores and accurate linkage tests. *Am. J. Hum. Genet.* **61**, 1179–1188 (1997).
37. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* **25**, 12–13 (2000).
38. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
39. Whittemore, A.S. & Halpern, J. A class of tests for linkage using affected pedigree members. *Biometrics* **50**, 118–127 (1994).
40. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
41. Nicolae, D. *Allele Sharing Models in Gene Mapping: A Likelihood Approach* (University of Chicago, 1999).
42. Dempster, A., Laird, N.M. & Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39**, 1–38 (1977).
43. Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337–346 (1992).
44. Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51** Pt 3, 227–233 (1987).
45. Lewontin, R.C. The interaction of selection and linkage. ii. Optimum models. *Genetics* **50**, 757–782 (1964).
46. Hill, W.G. & Robertson, A. The effects of inbreeding at loci with heterozygote advantage. *Genetics* **60**, 615–628 (1968).
47. Chen, X., Zehnbauer, B., Gnirke, A. & Kwok, P.Y. Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proc. Natl. Acad. Sci. USA* **94**, 10756–10761 (1997).
48. Steeds, R., Adams, M., Smith, P., Channer, K. & Samani, N.J. Distribution of tissue plasminogen activator insertion/deletion polymorphism in myocardial infarction and control subjects. *Thromb. Haemost.* **79**, 980–984 (1998).
49. Brouilette, S., Singh, R.K., Thompson, J.R., Goodall, A.H. & Samani, N.J. White cell telomere length and risk of premature myocardial infarction. *Arterioscler. Thromb. Vasc. Biol.* **23**, 842–846 (2003).
50. Nomenclature and criteria for diagnosis of ischemic heart disease. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. *Circulation* **59**, 607–609 (1979).



To order reprints, please contact:

In the Americas: Tel 212 726 9233; Fax 212 679 0843; reprints@natureny.com

Europe/UK/ROW: Tel + 44 (0)1256 329242; Fax + 44 (0) 1256 354018; reprints@nature.com

Japan & Korea: Tel +81 3 3267 8751; Fax +81 3 3267 8746; reprints@naturejpn.com

Printed by The Sheridan Press

Familial Risk of Lung Carcinoma in the Icelandic Population

Steinn Jonsson, MD
Unnur Thorsteinsdottir, PhD
Daniel F. Gudbjartsson, PhD
Hjortur H. Jonsson, MSc
Kristleifur Kristjansson, MD
Sigurdur Arnason, MD
Vilmundur Gudnason, MD, PhD
Helgi J. Isaksson, MD
Jonas Hallgrimsson, MD
Jeffrey R. Gulcher, MD, PhD
Laufey T. Amundadottir, PhD
Augustine Kong, PhD
Kari Stefansson, MD, PhD

Context The dominant role of tobacco smoke as a causative factor in lung carcinoma is well established; however, an inherited predisposition may also be an important factor in the susceptibility to lung carcinoma.

Objective To investigate the contribution of genetic factors to the risk of developing lung carcinoma in the Icelandic population.

Design, Setting, and Participants Risk ratios (RRs) of lung carcinoma for first-, second-, and third-degree relatives of patients with lung carcinoma were estimated by linking records from the Icelandic Cancer Registry (ICR) of all 2756 patients diagnosed with lung carcinoma within the Icelandic population from January 1, 1955, to February 28, 2002, with an extensive genealogical database containing all living Icelanders and most of their ancestors since the settlement of Iceland. The RR for smoking was similarly estimated using a random population-based cohort of 10541 smokers from the Reykjavik Heart Study who had smoked for more than 10 years. Of these smokers, 562 developed lung cancer based on the patients with lung cancer list from the ICR.

Main Outcome Measures Estimation of RRs of close and distant relatives of patients with lung carcinoma and comparison with RRs for close and distant relatives of smokers.

Results A familial factor for lung carcinoma was shown to extend beyond the nuclear family, as evidenced by significantly increased RR for first-degree relatives (for parents: RR, 2.69; 95% confidence interval [CI], 2.20-3.23; for siblings: RR, 2.02; 95% CI, 1.77-2.23; and for children: RR, 1.96; 95% CI, 1.53-2.39), second-degree relatives (for uncles/aunts: RR, 1.34; 95% CI, 1.15-1.49; and for nephews/nieces: RR, 1.28; 95% CI, 1.10-1.43), and third-degree relatives (for cousins: RR, 1.14; 95% CI, 1.05-1.22) of patients with lung carcinoma. This effect was stronger for relatives of patients with early-onset disease (age at onset ≤60 years) (for parents: RR, 3.48; 95% CI, 1.83-8.21; for siblings: RR, 3.30; 95% CI, 2.19-4.58; and for children: RR, 2.84; 95% CI, 1.34-7.21). The hypothesis that this increased risk is solely due to the effects of smoking was rejected for all relationships, except cousins and spouses, with a single-sided test of the RRs for lung carcinoma vs RRs for smoking.

Conclusions These results underscore the importance of genetic predisposition in the development of lung carcinoma, with its strongest effect in patients with early-onset disease. However, tobacco smoke plays a dominant role in the pathogenesis of this disease, even among those individuals who are genetically predisposed to lung carcinoma.

JAMA. 2004;292:2977-2983 www.jama.com

Lung carcinoma is the leading cause of death from cancer among men and women in many Western countries.[1] Mortality due to lung carcinoma in the United States exceeds the death rate from breast, prostate, and colon cancer combined.[2] Treatment results for lung carcinoma have remained disappointing and only marginal gains have been made during the last 30 to 40 years. Five-year survival is now approaching 14% given the best available diagnostic and treatment modalities.[3]

The dominant role of tobacco smoke as a causative factor in lung carcinoma is well established. Most studies report that more than 90% of patients with lung carcinoma are smokers.[1] Previous epidemiological case-control studies have shown an approximately 2-fold increase in the development of lung carcinoma in first-degree relatives of patients with lung carcinoma, after controlling for confounding factors, such as smoking and age, suggesting a genetic predisposition.[4-7]

Similar risk has also been observed for relatives of patients with lung carcinoma in larger registry-based studies utilizing the Utah Population and Cancer Registry Database[8,9] and the Swedish Family-Cancer Database.[10-12]

See also pp 2984 and 3026.

Author Affiliations: Departments of Medicine and Pathology, Landspitali-University Hospital (Drs S. Jonsson, Isaksson, and Hallgrimsson), deCODE Genetics (Drs Thorsteinsdottir, Gudbjartsson, Kristjansson, Arnason, Gulcher, Amundadottir, Kong, and Stefansson, and Mr H. Jonsson), and Icelandic Heart Association (Dr Gudnason), Reykjavík, Iceland.
Corresponding Authors: Kari Stefansson, MD, PhD, and Unnur Thorsteinsdottir, PhD, deCODE Genetics, Sturlugata 8, 101 Reykjavík, Iceland (kari.stefansson@decode.is and unnur.thorsteinsdottir@decode.is).

©2004 American Medical Association. All rights reserved.

These registry-based studies are more meaningful as they are less prone to sampling bias, resulting from proband identification and oversampling of families with several affected members.[13] However, none of these larger studies were controlled for smoking. It is important to control for smoking for 2 reasons. First, it is possible that the increased incidence of lung carcinoma in first-degree relatives is due to shared environment (second-hand smoke or other environmental factors), as demonstrated by increased lung cancer risk for spouses of patients with lung cancer in 1 of the Swedish studies.[12] Second, the familiality of lung cancer could be entirely due to the familiality of nicotine addiction and smoking.

In our study, we estimated the familiality of lung carcinoma in the Icelandic population by linking together records from the Icelandic Cancer Registry (ICR)[14,15] of all cases of lung carcinoma diagnosed in Iceland from January 1, 1955, to February 28, 2002, with a nationwide genealogical database containing all living Icelanders and the majority of their ancestors since the settlement of Iceland in 870 AD. This allowed us to examine all relationships among all of the lung carcinoma cases registered in the ICR and to estimate risk for lung carcinoma development beyond first-degree relatives of patients with lung carcinoma, thus reducing the effects of shared environment. Furthermore, by using information on smoking history from the Reykjavik Heart Study,[14] we estimated the familiality of smoking, and compared the risk ratio (RR) of lung carcinoma with the RR of smoking to examine whether there is a genetic component to the risk of lung carcinoma.

METHODS

Study Population

The study population included all patients diagnosed with lung carcinoma in Iceland from January 1, 1955, to February 28, 2002. These cases were all registered in the ICR.[14,15] Lung carcinoma was defined as a malignant neoplasm of epithelial origin according to the World Health Organization histological classification.[16] Carcinoid tumors as well as tumors of lymphoid and mesenchymal origin were excluded from our analysis. Information in the ICR includes year of diagnosis, year of death, Systematized Nomenclature of Medicine code, *International Classification of Diseases, 10th Revision (ICD-10)*, and mode of lung carcinoma verification. During this 47-year period, 2756 patients with lung carcinoma were identified (1504 men and 1252 women). Histological and cytological verification was available for 2516 patients with lung carcinoma; the remaining 240 patients were diagnosed clinically.

A random collection of 10541 adult smokers from the Icelandic population was obtained from the Icelandic Heart Association. These were individuals who had been randomly selected to take part in a nationwide study of cardiovascular risk factors (the Reykjavik Heart Study) during the years 1967 to 2002 and had answered a questionnaire on entry, which included information about their smoking habits. All individuals who had smoked for more than 10 years were defined as smokers. Of the 10541 smokers in the study, 562 developed lung carcinoma. Because we had smoking information only on a small proportion of all patients with lung carcinoma and their relatives, we could not calculate lung carcinoma RR directly, taking smoking into account. Instead, we used the random sample of smokers to estimate the familiality of smoking.

All data were encrypted through a process approved by the Data Protection Commission of Iceland before being sent to our laboratory for analysis.[17] The study was approved by the National Bioethics Committee of Iceland and the Data Protection Commission of Iceland.

Genealogical Database

We have built a computerized database of genealogical information in Iceland, including the names of all 284000 living Icelanders and their deceased ancestors.[18] Currently, more than 685000 individuals are registered in the database. Control groups were assembled to match the patients with lung carcinoma group according to year of birth, sex, and number of ancestors within the database in the preceding 5 generations. The Data Protection Commission of Iceland reversibly encrypted the data along with the genealogical database, before making it available to our laboratory.[17]

RR Calculation

To evaluate familial risk of lung carcinoma in the Icelandic population, we calculated RRs of close and distant relatives of the probands.[18] The RR for relatives of patients with lung carcinoma were defined as the risk of lung carcinoma in the relatives of affected individuals divided by the prevalence in the general population. In other words, if P denotes the event in which the proband is affected and R denotes the event in which the relative is affected, the RR is defined as

$$\frac{P(R|P)}{P(R)}.$$

When calculating the risk of lung carcinoma in relatives, we restricted our analyses to relatives born during the period covering the lifespan of the group of patients in question. We used the same restriction according to year of birth in estimating the risk in the general population for the given RR.

The RR of smoking was evaluated in a similar way as the RR of lung carcinoma using the list of the 10541 smokers and the Icelandic genealogical database. The RR for smoking together with the RR for lung carcinoma allows for a statistical test on the effects of smoking on lung carcinoma.

Statistical Analysis

Let r be the number of relatives of probands (counting multiple times individuals who are relatives of multiple probands[19]), a the number of relatives of probands that are affected (again possibly counting the same individual more

 ©2004 American Medical Association. All rights reserved.

than once), n the size of the population, and x the number of affected individuals in the population. If $P(R)$ and $P(R|P)$ can reasonably be assumed to be constant in the population, then respectively x/n and a/r are estimates of these probabilities. Given the estimates, RR is consistently estimated by

$$\frac{a/r}{x/n}.$$

Assuming the population may be split into N subpopulations, within each of which $P(R)$ and $P(R|P)$ can reasonably be assumed to be constant, although they may vary between subpopulations, and assuming RR is the same in all subpopulations, it is consistently estimated by any weighted sum of the estimates for the N subpopulations. We chose to select weights such that the efficiency of the estimator is at maximum for RR equal to 1. Making the simplifying assumption that the relatives are independent (although this assumption is obviously wrong, it only affects efficiency, not validity), the optimal weight for group j is

$$w_j = \frac{x_j r_j}{n_j - x_j}$$

(this is the inverse of the variance of the estimate for RR in subpopulation j), where the meaning of a, r, x, and n is the same as above, restricted to the subpopulation j, except that all affected individuals in the population are still taken as probands and not just the individuals in the subpopulation. Given these weights, our estimate of RR is

$$\frac{\sum_{j=1}^{N} w_j \frac{a_j/r_j}{x_j/n_j}}{\sum_{j=1}^{N} w_j} = \frac{\sum_{j=1}^{N} \frac{a_j n_j}{n_j - x_j}}{\sum_{j=1}^{N} \frac{r_j x_j}{n_j - x_j}}.$$

In our analysis, potential differences in $P(R)$ and $P(R|P)$ between subpopulations stem from time-dependent censoring of affection statuses and possibly sex-specific differences. Therefore, we have taken j to run over groups of relatives born in the same 5-year period and of the same sex. The patients with lung carcinoma in our analysis were born between the years 1868 and 1977, yielding 44 subpopulations.

In the case of smoking, our list is only a random sample of all the smokers. By applying the same method to estimate RR with this partial list, a_j/r_j is an underestimate of $P(R|P)$ and x_j/n_j is an underestimate of $P(R)$. However, since these estimates should be off by the same factor, $(a_j/r_j)/(x_j/n_j)$ continues to be a valid estimate of RR.

Because a person can both be a proband and a relative of 1 or more other probands, a_j does not have a binominal distribution. In general, for stratum j, a_j/r_j can be considered as a weighted average of many unbiased but correlated estimates of $P(R|P)$. It follows that $(a_j/r_j)/(x_j/n_j)$ is a ratio of 2 unbiased estimates and a consistent estimate of RR. Our overall estimate of RR is a weighted average of the estimates obtained from the various strata and is itself a consistent estimate. However, appropriate simulations, instead of purely analytical calculations, are needed to study its sampling variation. To assess the significance of the RR obtained for a given group of patients, we compared their observed values with the RR computed for 1000 independently drawn and matched groups of control individuals. Each patient was matched to a specific control individual in each control group. The control individuals were drawn at random from the genealogical database and had the same year of birth, the same sex, and the same number of ancestors recorded in the database, as did the patients to whom they were matched. A reported $P=.05$ for the RR would indicate that 50 of the 1000 matched control groups had values as large or larger than that for the patient's relatives or spouses. When none of the values computed for the control groups were larger than the value for the patient's relatives or spouses, we report $P<.001$. Using a variance stabilizing square root transform, an approximate confidence interval (CI) may be constructed based on the control distribution.[19]

Relationship Between RR of Smoking and Lung Carcinoma Under Certain Assumptions

We show that, assuming that the familial clustering of lung carcinoma is entirely explained by the familial clustering of smoking, the RR of smoking must be greater than that of lung carcinoma. Mathematically, when we say "the familial clustering of lung carcinoma is entirely explained by the familial clustering of smoking," we mean that the 4 random variables, proband lung carcinoma status, proband smoking status, relative smoking status, and relative lung carcinoma status, form a Markov Chain. For example, this means that relative lung carcinoma status is conditionally independent of proband lung carcinoma status, given the smoking status of either the proband or the relative.

Let P_{LC}, P_S, R_S, and R_{LC} denote the events that the proband has lung carcinoma, the proband smokes, the relative smokes, and the relative has lung carcinoma, respectively. Given that these events are all positively correlated and if we make the Markov assumption described above, then

(1) $P(R_{LC}|P_{LC}) \leq P(R_{LC}|P_S)$

and

(2) $P(P_S|R_{LC}) \leq P(P_S|R_S)$.

We want to prove that

(*) $[P(R_{LC}|P_{LC})/P(R_{LC})] \leq [P(R_S|P_S)/P(R_S)]$.

Because of (1), to prove (*), it is sufficient to show that

(**) $[P(R_{LC}|P_S)/P(R_{LC})] \leq [P(R_S|P_S)/P(R_S)]$.

Applying Bayes' Rule, the left-hand side of (**) can be rewritten as

(3) $P(P_S|R_{LC})/P(P_S)$

and the right-hand side of (**) can be rewritten as

(4) $P(P_S|R_S)/P(P_S)$.

It follows from (2) that (3) $\leq$ (4). Hence, (**) and (*) hold. It is also worth noting that equality holds in (*) if and only if (1) and (2) are both equalities. The latter is true if and only if

©2004 American Medical Association. All rights reserved.

$P(P_S|P_{LC}) = 1$ and $P(R_S|R_{LC}) = 1$. In other words, equality holds in (*) if and only if an individual must smoke to get lung carcinoma.

RESULTS

When the 2756 patients with lung carcinoma were matched to the Icelandic genealogical database, 274 affected sibling pairs, 296 affected avuncular pairs, and 724 affected cousin pairs were observed.

Estimates of the RR for relatives of the 2756 patients are shown in TABLE 1. Parents, siblings, and children (first-degree relatives) had RRs of 2.69 (95% CI, 2.20-3.23), 2.02 (95% CI, 1.77-2.23), and 1.96 (95% CI, 1.53-2.39), respectively. The RRs for uncles/aunts and nephews/nieces (second-degree relatives) and for cousins (third-degree relatives) were less than that of first-degree relatives but were also significantly increased. The RR for spouses was also significantly increased, although less than that for first-degree relatives.

To determine whether the risk of developing lung carcinoma is greater for relatives of patients with early-onset vs late-onset disease, we calculated the RR for relatives of patients diagnosed with lung carcinoma at 60 years or younger (Table 1). For all groups of relatives analyzed, the risk was greater for relatives of patients with early-onset disease than for relatives of all patients with lung carcinoma. Thus, the risk for second-degree relatives (RR, 1.96; 95% CI, 1.35-2.78, for uncles/aunts; and RR, 1.94; 95% CI, 1.32-2.72, for nephews/nieces) of patients with early-onset disease is similar to the risk for children and siblings (RR, 1.96; 95% CI, 1.53-2.39; and RR, 2.02; 95% CI, 1.77-2.23, respectively) of all patients with lung carcinoma.

All 4 major histological types of lung carcinoma (adenocarcinoma and small cell, large cell, and squamous cell carcinoma) are significantly associated with smoking, and the risk of developing lung carcinoma increases with number of cigarettes smoked and the duration of smoking. However, the strength of this relationship varies between the histological types with adenocarcinoma displaying the weakest overall relationship to smoking.[20,21] Due to this difference, we calculated the risk of lung carcinoma development for relatives and spouses for adenocarcinoma separately from the other major histological types of lung carcinoma (ie, small cell, large cell, and squamous cell carcinomas) (TABLE 2). No significant difference in lung carcinoma risk was detected between relatives and spouses of patients with lung carcinoma from these 2 histological groups. However, the risk for spouses of patients with adenocarcinoma of the lung was only half of that of spouses of the combined group of

Table 1. Estimation of Risk Ratio of Lung Carcinoma for Relatives and Spouses of Icelandic Patients With Lung Carcinoma

	All Lung Carcinoma Patients (n = 2756)*			Lung Carcinoma Patients, Age at Onset ≤60 y (n = 793)†		
Relationship	No. of Relatives‡	RR (95% CI)	P Value§	No. of Relatives‡	RR (95% CI)	P Value§
Parents	4874	2.69 (2.20-3.23)	<.001	1489	3.48 (1.83-8.21)	<.001
Siblings	11 081	2.02 (1.77-2.23)	<.001	3130	3.30 (2.19-4.58)	<.001
Children	8748	1.96 (1.53-2.39)	<.001	2416	2.84 (1.34-7.21)	.007
Uncles/aunts	20 348	1.34 (1.15-1.49)	<.001	6779	1.96 (1.35-2.78)	<.001
Nephews/nieces	29 134	1.28 (1.10-1.43)	<.001	8577	1.94 (1.32-2.72)	<.001
Cousins	49 275	1.14 (1.05-1.22)	.002	18 006	1.32 (1.03-1.69)	.02
Spouses	2909	1.75 (1.29-2.33)	<.001	895	1.91 (0.71-5.68)	.06

Abbreviations: CI, confidence interval; RR, risk ratio.
*All patients with lung carcinoma who were registered in the Icelandic Cancer Registry (ICR) from 1955-2002.
†All patients with lung carcinoma who were registered in the ICR from 1955-2002 with age at onset of 60 years or younger.
‡Numbers given for relatives are unique counts.
§Reported $P<.001$ indicates that none of the 1000 matched control groups had RR values as large or larger than that for the patients' relatives or spouses.

Table 2. Estimation of Risk Ratio of Lung Carcinoma for Relatives and Spouses of Icelandic Patients With Adenocarcinoma vs Small Cell, Large Cell, and Squamous Cell Carcinoma

	Adenocarcinoma (n = 829)*			Small Cell, Large Cell, and Squamous Cell Carcinoma (n = 1118)*		
Relationship	No. of Relatives†	RR (95% CI)	P Value	No. of Relatives†	RR (95% CI)	P Value
Parents	1559	2.12 (1.01-3.80)	.02	2064	2.47 (1.36-4.55)	<.001
Siblings	3565	2.61 (1.77-3.66)	<.001	4758	2.18 (1.64-2.74)	<.001
Children	2672	1.65 (0.67-3.22)	.13	3507	1.81 (0.91-3.56)	.05
Uncles/aunts	6978	1.67 (1.16-2.36)	.004	8881	1.91 (1.40-2.40)	<.001
Nephews/nieces	9962	1.57 (1.05-2.19)	.02	12 885	1.77 (1.35-2.22)	<.001
Cousins	18 437	1.10 (1.05-1.33)	.31	22 593	1.22 (1.06-1.48)	.004
Spouses	857	1.08 (0.20-4.15)	.38	1198	2.23 (1.14-4.09)	.008

Abbreviations: CI, confidence interval; RR, risk ratio.
*Patients with adenocarcinoma, small cell, large cell, and squamous cell carcinoma who were registered in the Icelandic Cancer Registry from 1955-2002.
†Numbers given for relatives are unique counts.

©2004 American Medical Association. All rights reserved.

small cell, large cell, and squamous cell lung carcinoma. Although this difference was large, it was not significant as the CI for the spouses of patients with adenocarcinoma lung cancer was wide due to low number of spouses in that cohort.

It has been proposed that nicotine addiction (smoking) is at least in part inherited. We thus calculated the risk of smoking for relatives and spouses of smokers using a random list of 10541 individuals who had smoked at least 1 package of cigarettes per day for more than 10 years. As shown in TABLE 3, the risk of having smoked for more than 10 years is significant for first-, second-, and third-degree relatives of smokers. The risk was, however, highest for spouses of smokers (RR, 2.39; 95% CI, 2.28-2.48), suggesting that in addition to genetic factors, environmental factors and/or nonrandom mating have a substantial effect on smoking habits.

Prolonged exposure to tobacco smoke precedes the development of lung carcinoma in the vast majority of patients with lung carcinoma. We demonstrate mathematically that if the familiality of lung carcinoma is entirely explained by the familiality of smoking, the risk for smoking (Table 3) must be higher than that of lung carcinoma (Table 1). Therefore, if the RR of lung carcinoma is actually higher than the RR of smoking, it would be a rejection of the null hypothesis that lung carcinoma is entirely due to smoking. Based on a single-sided test of the RRs for lung carcinoma vs RRs for smoking, the null hypothesis was rejected beyond the nuclear family (TABLE 4). This was evident by significantly higher RRs for lung carcinoma than for smoking for all relationships except for cousins. In contrast, the RR for smoking of spouses was significantly higher than the RR for lung carcinoma.

Taken together, our data on the nationwide evaluation of lung carcinoma familiality in Iceland demonstrates that heritable factors are indeed involved in the etiology of lung carcinoma. Furthermore, this genetic predisposition goes beyond the predisposition to smoking.

COMMENT

We investigated the role of genetic factors in the development of lung carcinoma by linking together information on all lung carcinoma cases diagnosed within the Icelandic population from January 1, 1955, to February 28, 2002, with an extensive genealogical database covering all Icelanders living during this time and most of their ancestors. Using these data, we found that there is a familial predisposition to the development of lung carcinoma, as RR estimates for first-, second-, and third-degree relatives of patients with lung carcinoma were all significantly increased. This effect was strongest for relatives of patients with early-onset lung carcinoma, in accordance with previous articles.[22] Significantly increased RR for spouses of patients with lung carcinoma also indicates the presence of shared environmental factors and/or nonrandom mating.

The nationwide genealogy database used in our study provided a means for uncovering the familial component by revealing more connections between patients, missed in most other populations. The first-degree relatives (siblings, children, and parents) of patients with lung carcinoma (early- and late-onset) are at a 2- to 3.5-fold increased risk of developing lung carcinoma than the general population. However, members of a nuclear family share environment, as evidenced by the 1.75-fold risk of lung carcinoma development in spouses. Thus, this RR increase in first-degree relatives of patients with lung carcinoma is the result of a combination of environmental, genetic factors, or both. Using genealogy, our study goes further than other reported studies by demonstrating that this familial factor extends beyond the nuclear family as evidenced by significantly increased RR for second- and third-degree relatives of patients with lung carcinoma. In the more distant relationships, shared environmental factors are likely to be of less significance, providing a stronger evidence for genetic factors given that RR is in excess.

We had smoking information only for a proportion of our nationwide cohort of patients with lung carcinoma and therefore could not estimate RR di-

Table 3. Estimation of Smoking Risk Ratio for Relatives and Spouses of Smokers (n = 10 541)*

Relationship	No. of Relatives†	RR (95% CI)‡
Parents	14 343	1.94 (1.70-2.17)
Siblings	30 299	1.42 (1.38-1.46)
Children	27 374	1.52 (1.31-1.70)
Uncles/aunts	42 898	1.16 (1.08-1.22)
Nephews/nieces	70 144	1.17 (1.12-1.23)
Cousins	86 249	1.14 (1.12-1.16)
Spouses	10 946	2.39 (2.28-2.48)

Abbreviations: CI, confidence interval; RR, risk ratio.
*Smokers were defined as those individuals who had smoked at least 1 package of cigarettes per day for more than 10 years and were randomly gathered by the Icelandic Heart Association.
†Numbers given for relatives are unique counts.
‡For all comparisons, $P<.001$.

Table 4. Single-Sided Comparison of Risk Ratio of Lung Carcinoma for Relatives and Spouses of Icelandic Patients With Lung Carcinoma With the Risk Ratio of Smoking for Relatives and Spouses of Icelandic Smokers

	All Lung Carcinoma Patients (n = 2756)		Lung Carcinoma Patients, Age at Onset ≤60 y (n = 793)	
Relationship	Δ RR of Lung Carcinoma − Δ RR of Smoking	P Value*	Δ RR of Lung Carcinoma − Δ RR of Smoking	P Value*
Parents	0.75	.003	1.55	.08
Siblings	0.60	<.001	1.87	.002
Children	0.44	.04	1.32	.10
Uncles/aunts	0.18	.02	0.80	.01
Nephews/nieces	0.11	.007	0.77	.01
Cousins	−0.011	.61	0.18	.08
Spouses	−0.64	>.99	−0.48	.74

Abbreviation: RR, risk ratio.
*For $P<.001$, for no pair of controls was the RR for smoking higher than the RR for lung carcinoma.

©2004 American Medical Association. All rights reserved.

rectly taking smoking into account. However, we demonstrated mathematically that single-sided comparison of the RR for smoking to that of lung carcinoma in relatives and spouses of smokers and patients with lung carcinoma, respectively, can be used to determine whether lung carcinoma is entirely due to smoking. When that comparison is applied, the risk for lung carcinoma is significantly higher than the risk for smoking beyond the first-degree relatives of patients with lung carcinoma, demonstrating that increased risk for relatives of patients with lung carcinoma is not solely due to smoking. In contrast, this effect for spouses is opposite (the RR for smoking is higher than for lung carcinoma). These results suggest that the increased risk for lung carcinoma among spouses may be solely due to tobacco exposure. Furthermore, and more importantly, these data also demonstrate that the increased risk for close and distant relatives of patients with lung carcinoma is not solely due to tobacco smoke exposure. Similar conclusion was also reached in a study in which survival models were applied in a case-control analysis of lung carcinoma (ie, the familial aggregation of lung carcinoma could not be fully explained by the familial aggregation of smoking).[23] Based on previous theoretical analysis by Khoury et al,[24] it is unlikely that other unknown environmental factors could explain fully the increased familial risk in lung carcinoma, implying an underlying genetic predisposition in lung carcinoma.

When we compared the risk of lung carcinoma for spouses and relatives of patients with adenocarcinoma to that of spouses and relatives of patients with other histological types of lung carcinoma, the greatest difference (more than half, although not significant) was observed between the spouses of these 2 groups. This suggests a weaker environmental influence for adenocarcinoma than for the 3 other major histology types of lung carcinoma. These data concur with epidemiological studies that have demonstrated a weaker association between smoking and adenocarcinoma vs other histological types of lung carcinoma.[20,21]

Comparison of the concordance of cancer between monozygotic and dizygotic pairs of twins has been used to quantify the extent to which an observed familial pattern is due to genetic or shared environmental factors.[25] However, these studies are limited because twins are rare and few twin registries go far enough back in time for cancer assessment.[26] The largest of these studies have suggested a limited heritability of lung carcinoma, although none reached statistical significance.[25]

In previous epidemiological studies on lung carcinoma using segregation analysis, a codominant model of inheritance best fitted the data, suggesting that a rare major autosomal gene plays a role in the development of lung carcinoma.[27] Other studies have suggested that a number of low-penetrance, high-frequency polymorphisms are likely to account for a proportion of lung carcinoma risk.[28] Polymorphisms in these genes could explain individual differences in susceptibility to tobacco carcinogens and are likely to include genes involved in decreasing or increasing the activity of carcinogens (eg, *CYP1A*, *CYP2E*, and *GSTM1*) and genes involved in monitoring and repairing tobacco carcinogen-induced DNA damage (eg, p53 and *ERCC1*).[29-31] Our results of RR calculation cannot discriminate between different models of inheritance. Recently, a major lung cancer susceptibility locus was mapped to chromosome 6q23-25 using multigenerational densely-affected families.[32] The characteristics of this locus are consistent with a dominant or codominant major locus. Information gained from epidemiological and genetical studies such as our study may be of particular importance in allowing for risk stratification with respect to lung carcinoma. Further information gained from linkage and association studies may give additional value in this respect.

In conclusion, to our knowledge, this study is the first population-based study using a comprehensive and extensive genealogy database, taking into account the effects of smoking, which demonstrates a familial nature of lung carcinoma that strongly suggests a genetic predisposition to the disease. However, although the results presented here support a role for genetics in the risk of lung carcinoma, it should be emphasized that tobacco smoke plays a dominant role in the pathogenesis of this disease, even among those individuals who are genetically predisposed to lung carcinoma.

Author Contributions: Drs Jonsson and Stefansson had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: S. Jonsson, Thorsteinsdottir, Kristjansson, Arnason, Hallgrimsson, Gulcher, Amundadottir, Stefansson.
Acquisition of data: S. Jonsson, Isaksson.
Analysis and interpretation of data: S. Jonsson, Thorsteinsdottir, H. Jonsson, Kong, Gudbjartsson.
Drafting of the manuscript: S. Jonsson, Thorsteinsdottir.
Critical revision of the manuscript for important intellectual content: S. Jonsson, Thorsteinsdottir, H. Jonsson, Kong, Gudbjartsson, Kristjansson, Arnason, Isaksson, Hallgrimsson, Gulcher, Amundadottir, Stefansson.
Statistical analysis: H. Jonsson, Kong, Gudbjartsson.
Obtained funding: Stefansson.
Administrative, technical, or material support: S. Jonsson, Thorsteinsdottir, Kristjansson, Arnason, Isaksson, Gulcher, Amundadottir, Stefansson.
Study supervision: S. Jonsson, Thorsteinsdottir.
Funding/Support: All of the work, data generation, and analysis of this study was supported by deCODE Genetics.
Role of the Sponsor: deCODE Genetics participated in the design and conduct of the study, the collection, analysis, and interpretation of the data, and the preparation, review, and approval of the manuscript.
Independent Statistical Analysis: Kristjan Jonasson, PhD, Associate Professor, Department of Mathematics, Faculty of Science, University of Iceland, was given access to the complete data, including genealogical data and lung cancer and smoking data, after coding of personal identification numbers. Dr Jonasson completed a thorough check of the methods and data analysis, and confirmed that the results reported in the submitted manuscript are both statistically correct and in accordance with the data.
Acknowledgment: We thank the Icelandic Cancer Registry for providing us with the list of patients with lung carcinoma.

REFERENCES

1. Mannino DM, Ford E, Giovino GA, Thun M. Lung cancer deaths in the United States from 1979 to 1992: an analysis using multiple-cause mortality data. *Int J Epidemiol.* 1998;27:159-166.
2. Schiller JH, Harrington D, Belani CP, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med.* 2002;346:92-98.
3. Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1999. *CA Cancer J Clin.* 1999;49:8-31.
4. Amos CI, Xu W, Spitz MR. Is there a genetic basis for lung cancer susceptibility? *Recent Results Cancer Res.* 1999;151:3-12.
5. Ooi WL, Elston RC, Chen VW, Bailey-Wilson JE, Rothschild H. Increased familial risk for lung cancer. *J Natl Cancer Inst.* 1986;76:217-222.
6. Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. *J Natl Cancer Inst.* 1963;30:289-312.

©2004 American Medical Association. All rights reserved.

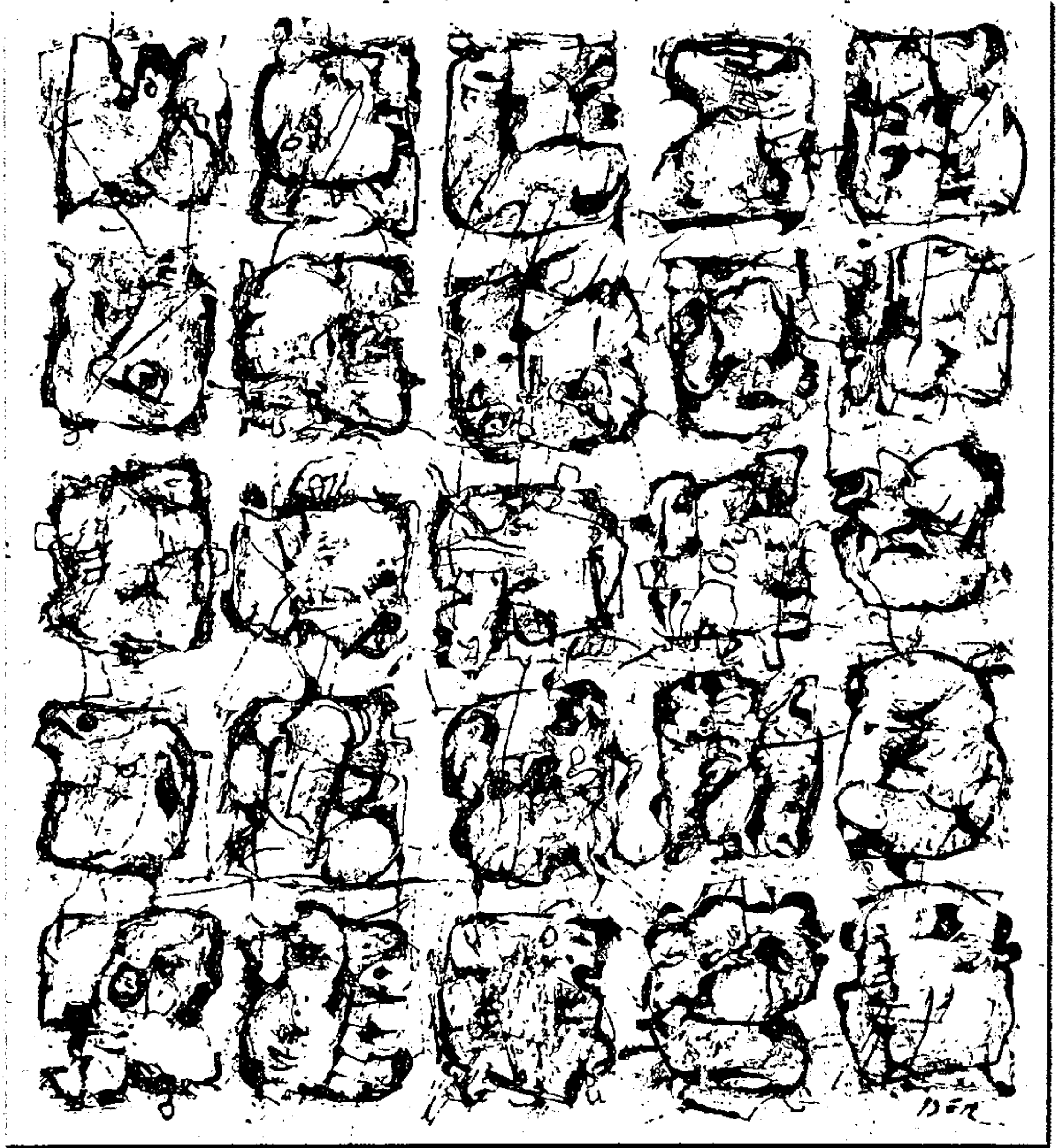
7. Samet JM, Humble CG, Pathak DR. Personal and family history of respiratory disease and lung cancer risk. *Am Rev Respir Dis*. 1986;134:466-470.
8. Cannon-Albright LA, Thomas A, Goldgar DE, et al. Familiality of cancer in Utah. *Cancer Res*. 1994;54:2378-2385.
9. Thomas A, Cannon-Albright L, Bansal A, Skolnick MH. Familial associations between cancer sites. *Comput Biomed Res*. 1999;32:517-529.
10. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer*. 2002;99:260-266.
11. Dong C, Hemminki K. Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families. *Int J Cancer*. 2001;92:144-150.
12. Hemminki K, Jiang Y. Cancer risks among long-standing spouses. *Br J Cancer*. 2002;86:1737-1740.
13. Guo SW. Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. *Am J Hum Genet*. 1998;63:252-258.
14. Tulinius H, Sigfusson N, Sigvaldason H, Bjarnadottir K, Tryggvadottir L. Risk factors for malignant diseases: a cohort study on a population of 22,946 Icelanders. *Cancer Epidemiol Biomarkers Prev*. 1997; 6:863-873.
15. Parkin DM, Shelan SL, Ferlay J, Raymond L, Young J. *Cancer Incidence in Five Continents*. Vol VII. Lyon, France: IARC Scientific Publication; 1997.
16. Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E. Histologic typing of lung and pleural tumours. In: *International Histological Classification of Tumours, World Health Organization*. 3rd ed. Berlin, Germany: Springer; 1999.
17. Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur J Hum Genet*. 2000;8:739-742.
18. Gulcher J, Stefansson K. Population genomics: laying the groundwork for genetic disease modeling and targeting. *Clin Chem Lab Med*. 1998;36:523-527.
19. Sveinbjornsdottir S, Hicks AA, Jonsson T, et al. Familial aggregation of Parkinson's disease in Iceland. *N Engl J Med*. 2000;343:1765-1770.
20. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer*. 2001;31:139-148.
21. Zatloukal P, Kubik A, Pauk N, Tomasek L, Petruzelka L. Adenocarcinoma of the lung among women: risk associated with smoking, prior lung disease, diet and menstrual and pregnancy history. *Lung Cancer*. 2003;41:283-293.
22. Gauderman WJ, Morrison JL. Evidence for age-specific genetic relative risks in lung cancer. *Am J Epidemiol*. 2000;151:41-49.
23. Mack W, Langholz B, Thomas DC. Survival models for familial aggregation of cancer. *Environ Health Perspect*. 1990;87:27-35.
24. Khoury MJ, Beaty TH, Liang KY. Can familial aggregation of disease be explained by familial aggregation of environmental risk factors? *Am J Epidemiol*. 1988;127:674-683.
25. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343:78-85.
26. Hemminki K, Lonnstedt I, Vaittinen P, Lichtenstein P. Estimation of genetic and environmental components in colorectal and lung cancer and melanoma. *Genet Epidemiol*. 2001;20:107-116.
27. Sellers TA, Bailey-Wilson JE, Elston RC, et al. Evidence for mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Inst*. 1990;82:1272-1279.
28. Spitz MR, Wei Q, Li G, Wu X. Genetic susceptibility to tobacco carcinogenesis. *Cancer Invest*. 1999; 17:645-659.
29. Wu X, Zhao H, Amos CI, et al. p53 Genotypes and haplotypes associated with lung cancer susceptibility and ethnicity. *J Natl Cancer Inst*. 2002;94:681-690.
30. Bosken CH, Wei Q, Amos CI, Spitz MR. An analysis of DNA repair as a determinant of survival in patients with non-small-cell lung cancer. *J Natl Cancer Inst*. 2002;94:1091-1099.
31. Wei Q, Cheng L, Amos CI, et al. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J Natl Cancer Inst*. 2000;92:1764-1772.
32. Bailey-Wilson JE, Amos CI, Pinney SM, et al. A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am J Hum Genet*. 2004;75: 460-474.

It has never been my object to record my dreams, just to realize them.
—Man Ray (1890-1976)

©2004 American Medical Association. All rights reserved.

nature genetics

VOLUME 36 NUMBER 11 NOVEMBER 2004
www.nature.com/naturegenetics



Spindle checkpoint in cancer
Legionnaires' disease genome
Selection for human fertility

nature genetics

Recombination rate and reproductive success in humans

Augustine Kong[1], John Barnard[2], Daniel F Gudbjartsson[1], Gudmar Thorleifsson[1], Gudrun Jonsdottir[1], Sigrun Sigurdardottir[1], Bjorgvin Richardsson[1], Jonina Jonsdottir[1], Thorgeir Thorgeirsson[1], Michael L Frigge[1], Neil E Lamb[3], Stephanie Sherman[3], Jeffrey R Gulcher[1] & Kari Stefansson[1]

Intergenerational mixing of DNA through meiotic recombinations of homologous chromosomes during gametogenesis is a major event that generates diversity in the eukaryotic genome. We examined genome-wide microsatellite data for 23,066 individuals, providing information on recombination events of 14,140 maternal and paternal meioses each, and found a positive correlation between maternal recombination counts of an offspring and maternal age. We postulated that the recombination rate of eggs does not increase with maternal age, but that the apparent increase is the consequence of selection. Specifically, a high recombination count increased the chance of a gamete becoming a live birth, and this effect became more pronounced with advancing maternal age. Further support for this hypothesis came from our observation that mothers with high oocyte recombination rate tend to have more children. Hence, not only do recombinations have a role in evolution by yielding diverse combinations of gene variants for natural selection, but they are also under selection themselves.

We recently constructed a high-resolution genetic map that highlights the variation in recombination rates between the sexes and across the genome[1]. We confirmed a previous observation[2] that recombination rates among mothers can differ substantially and observed that even gametes of one mother have different recombination rates; a gamete with a high recombination count in one chromosome tends also to have high recombination counts in other chromosomes[1]. Here, we focused on determining whether recombination rate is related to the age of the mother. The chiasma frequency of mouse oocytes is reported to decrease as the mouse ages[3]. It was suggested that a reduction in crossing over leading to formation of univalents might explain age-dependent nondisjunction. As chiasma formation occurs prenatally, however, the 'production line' hypothesis was proposed. This hypothesis states that there is a gradient in the fetal ovary, so that the first-formed oocytes have a higher chiasma frequency than those formed later, and that oocytes are ovulated in the same order that they enter meiosis. Attempts to validate this model have been equivocal[4,5]; however, studies in the mouse suggest that the last-formed oocytes are also the last to be ovulated[6].

In humans, a number of studies have been done to estimate recombination counts using genetic data from families (that is, parent-offspring transmissions), but none has provided convincing evidence that the recombination count in an ooctye is correlated with maternal age. A reported decrease in recombination with increasing maternal age using the Venezuelan Reference Pedigree[7] could not be replicated by further analysis using the same data source[8]. Most earlier studies were based on small sample sizes and were not genome-wide investigations[9–11]. Two genome-wide studies[1,2] did not detect a statistically significant age effect. Suspecting that the failures of previous studies to detect an effect were due to the lack of power, we carried out a large study using two primary resources: a genetic database with genotypic data on ~1,000 microsatellite markers typed in 70,000 individuals and a genealogy database covering the entire Icelandic nation. We used these to construct a data set consisting of 5,463 families, with 23,066 individuals genotyped (average yield >800 genotypes per person) and providing information on 14,140 maternal and paternal meioses each. These are nuclear families with two or more siblings and at least one parent genotyped (**Table 1**). Our genealogy database provides the birth years of the individuals, rounded up to the nearest five years to protect privacy[12]. We calculated the approximate age of the mother at the birth of every child and determined the total number of children a mother had, regardless of genotype status. We chose families in which the mothers were

Table 1 Count of families according to the number of genotyped children and the status of parental genotyping

	Number of genotyped children				
Parent genotyped	2	3	4	5 or more	Total
Father only	342	178	58	31	609
Mother only	885	345	102	59	1,391
Father and mother	2,059	956	307	141	3,463
Total	3,286	1,479	467	231	5,463

[1]deCODE Genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. [2]Department of Biostatistics and Epidemiology, Cleveland Clinic Foundation, Cleveland, Ohio, USA. [3]Department of Human Genetics, Emory University, Atlanta, Georgia, USA. Correspondence should be addressed to A.K. (kong@decode.is) or K.S. (kstefans@decode.is).

Published online 3 October 2004; doi:10.1038/ng1445