LETTERS

Variants conferring risk of atrial fibrillation on chromosome 4q25

Daniel F. Gudbjartsson[1], David O. Arnar[2], Anna Helgadottir[1], Solveig Gretarsdottir[1], Hilma Holm[2], Asgeir Sigurdsson[1], Adalbjorg Jonasdottir[1], Adam Baker[1], Gudmar Thorleifsson[1], Kristleifur Kristjansson[1], Arnar Palsson[1], Thorarinn Blondal[1], Patrick Sulem[1], Valgerdur M. Backman[1], Gudmundur A. Hardarson[1], Ebba Palsdottir[1], Agnar Helgason[1], Runa Sigurjonsdottir[2], Jon T. Sverrisson[3], Konstantinos Kostulas[4], Maggie C. Y. Ng[5], Larry Baum[5], Wing Yee So[5], Ka Sing Wong[5], Juliana C. N. Chan[5], Karen L. Furie[6], Steven M. Greenberg[6], Michelle Sale[6], Peter Kelly[6], Calum A. MacRae[7], Eric E. Smith[6], Jonathan Rosand[6], Jan Hillert[4], Ronald C. W. Ma[5], Patrick T. Ellinor[7], Gudmundur Thorgeirsson[2], Jeffrey R. Gulcher[1], Augustine Kong[1], Unnur Thorsteinsdottir[1] & Kari Stefansson[1]

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia in humans and is characterized by chaotic electrical activity of the atria[1]. It affects one in ten individuals over the age of 80 years, causes significant morbidity and is an independent predictor of mortality[2]. Recent studies have provided evidence of a genetic contribution to AF[3–5]. Mutations in potassium-channel genes have been associated with familial AF[6–10] but account for only a small fraction of all cases of AF[11,12]. We have performed a genome-wide association scan, followed by replication studies in three populations of European descent and a Chinese population from Hong Kong and find a strong association between two sequence variants on chromosome 4q25 and AF. Here we show that about 35% of individuals of European descent have at least one of the variants and that the risk of AF increases by 1.72 and 1.39 per copy. The association with the stronger variant is replicated in the Chinese population, where it is carried by 75% of individuals and the risk of AF is increased by 1.42 per copy. A stronger association was observed in individuals with typical atrial flutter. Both variants are adjacent to *PITX2*, which is known to have a critical function in left–right asymmetry of the heart[13–15].

We conducted a genome-wide association study with the use of the Illumina Hap300 BeadChip on an Icelandic population with AF and/or atrial flutter (AFl); 316,515 single-nucleotide polymorphisms (SNPs) satisfying our quality criteria (Supplementary Information) were tested individually for association with AF or AFl in a sample of 550 patients and 4,476 controls from Iceland. Three strongly correlated SNPs, all located within a single linkage disequilibrium (LD) block on chromosome 4q25, were the only SNPs found to be significant on a genome-wide basis after the 316,515 SNPs tested had been accounted for ($P < 0.05/316{,}515 = 1.58 \times 10^{-7}$): rs2200733 (odds ratio (OR) = 1.75; $P = 1.6 \times 10^{-10}$), rs2220427 (OR = 1.75; $P = 1.9 \times 10^{-10}$) and rs2634073 (OR = 1.60; $P = 2.1 \times 10^{-9}$). These results and all other results based on the Icelandic population were adjusted for the relatedness of individuals. The two most significant SNPs, rs2200733 and rs2220427, are perfect proxies for one another in the CEPH CEU HapMap[16] data set and are close to being perfect proxies for one another in the Icelandic data set ($D' = 1$, $r^2 = 0.999$); therefore only rs2200733 will be referred to in the following discussion. The correlation of rs2634073 with rs2200733 is weaker in the Icelandic data set ($D' = 0.95$, $r^2 = 0.605$). On further study of the Illumina Hap300 SNPs in the vicinity of the first three SNPs and conditioning on the association with rs2200733, an association with a new SNP, rs10033464, was identified (OR = 1.42; $P = 0.0024$). After the association with rs2200733 and rs10033464 had been accounted for, the association with rs2634073 was no longer significant ($P = 0.30$). Henceforth, all association results for rs2200733 T and rs10033464 T, including those presented in Table 1, are based on a comparison with the wild-type haplotype, which carries neither of the two at-risk alleles, rather than on a comparison with the major alleles of each SNP separately. Specifically, ORs for rs2200733 T and rs10033464 T are each computed conditionally (see Methods Summary) and could be interpreted as the estimated relative risk of each variant compared with the wild type. The at-risk alleles T of rs2200733 and T of rs10033464 have estimated population allelic frequencies of 12.05% and 8.53% in Iceland, respectively, and are never observed together on the same chromosome, in either the Icelandic data set or the CEU HapMap data set. A third SNP, rs13143308, which has a minor allele that corresponds completely to chromosomes carrying either the T allele of rs2200733 or the T allele of rs10033464, was identified through the CEU HapMap data set. Figure 1 demonstrates the haplotype structure over the key SNPs of the associated region. Sets of SNPs that are perfect proxies of each of these three key SNPs in the CEU HapMap samples are provided in Supplementary Table 1, and relative locations are shown in Fig. 2. We emphasize that the SNPs named should be considered representatives of the haplotypes defined by the SNPs to which they are equivalent and are chosen primarily for the sake of convenience.

A microsatellite marker, D4S406, located in the same LD block as the two SNPs was identified. In Iceland, three of the four shortest alleles of D4S406 (−8, −4 and −2) combine to form a near-perfect surrogate for the T allele of rs2200733 ($D' = 0.995$, $r^2 = 0.98$), and the two shortest remaining alleles (−6 and 0) form a good surrogate for the T allele of rs10033464 ($D' = 0.98$, $r^2 = 0.75$; Supplementary Table 2). None of the remaining (longer) alleles of D4S406 are associated with AF/AFl after the effect of the short alleles had been accounted for. For replication of the original observation in

[1]deCODE genetics, Sturlugata 8, 101 Reykjavik, Iceland. [2]Division of Cardiology, Department of Medicine, Landspitali University Hospital, 101 Reykjavik, Iceland. [3]Department of Medicine, Akureyri Regional Hospital, 600 Akureyri, Iceland. [4]Department of Neurology, Karolinska Institutet at Karolinska University Hospital, Huddinge S-141 86, Sweden. [5]Department of Medicine and Therapeutics, Prince of Wales Hospital, Chinese University of Hong Kong, Shatin, Hong Kong. [6]Department of Neurology, [7]Cardiology Division and Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA.

Table 1 | Analysis of the association of rs2200733 and rs10033464 on chromosome 4q25 with AF/AFl

Sample (cases/controls)	rs2200733 T* Frequency§	OR (95% CI)	P	rs10033464 T*† Frequency§	OR (95% CI)	P	Comparison P‡	Joint PAR
Iceland‖								
Discovery (550/4,476)	0.191 (0.114)	1.84 (1.54–2.21)	2.0×10^{-11}	0.110 (0.080)	1.42 (1.13–1.77)	0.0024	0.041	0.216
Replication (2,251/13,238)	0.166 (0.108)	1.64 (1.49–1.81)	2.7×10^{-23}	0.108 (0.080)	1.40 (1.24–1.58)	8.2×10^{-8}	0.028	0.176
Combined (2,801/17,714)	0.171 (0.110)	1.68 (1.53–1.83)	1.9×10^{-30}	0.108 (0.080)	1.40 (1.25–1.55)	9.4×10^{-9}	0.0025	0.180
Other European ancestry								
Sweden (143/738)	0.179 (0.098)	2.01 (1.38–2.93)	0.00027	0.172 (0.111)	1.65 (1.14–2.41)	0.0087	0.41	0.272
United States (636/804)	0.229 (0.139)	1.84 (1.51–2.23)	9.8×10^{-10}	0.105 (0.083)	1.30 (1.00–1.69)	0.052	0.026	0.232
Combined¶	– (–)	1.88 (1.58–2.23)	1.2×10^{-12}	– (–)	1.41 (1.13–1.75)	0.0019	0.027	0.237
All European ancestry								
Combined¶	– (–)	1.72 (1.59–1.86)	3.3×10^{-41}	– (–)	1.39 (1.26–1.53)	6.9×10^{-11}	0.00019	0.206
Hong Kong								
Hong Kong (333/2,836)	0.605 (0.528)	1.42 (1.16–1.73)	0.00064	0.190 (0.218)	1.08 (0.84–1.39)	0.55	0.0099	0.346

Each row contains the results from a joint analysis of two variants, rs2200733 T and rs10033464 T†. The numbers of cases and controls are shown for each case-control study and for each variant the allelic frequencies of the variant in cases and controls, the OR with a 95% CI and two-sided *P* values are shown. In addition a *P* value for comparing the effect of the two variants and their joint PAR is reported. For example, the first row indicates that, for the initial Icelandic discovery samples, rs2200733 T has an estimated OR of 1.84 (95% CI 1.54–2.21, $P = 2.0 \times 10^{-11}$) versus the wild type (rs2200733 C, rs13143308 G, rs10033464 G haplotype), and rs2200733 T has an estimated OR of 1.42 (95% CI 1.13–1.77, $P = 0.0024$) versus the wild type.

* Results of comparing rs2200733 T and rs10033464 T with the wild-type rs2200733 C, rs13143308 G, rs10033464 G haplotype.

† In the Swedish and US samples rs10033464 T was tagged by the rs2200733 C, rs13143308 T haplotype.

‡ *P* value for comparing the ORs of rs2200733 T and rs10033464 T.

§ Shown as cases (controls).

‖ The association analysis was adjusted for the relatedness of some of the individuals.

¶ For the combined study populations of European decent, the PAR was calculated by using the average, unweighted control frequency of the populations, whereas the OR and *P* value were estimated by using the Mantel–Haenszel model.

Iceland the D4S406 genotypes were used to provide information when SNP genotypes were not available.

In an attempt to replicate our original discovery we analysed an additional Icelandic sample consisting of 2,251 AF/AFl patients and 13,238 controls (Table 1). The association of both SNPs with AF/AFl was replicated in this sample (OR = 1.64, $P = 2.7 \times 10^{-23}$ for rs2200733; OR = 1.40, $P = 8.2 \times 10^{-8}$ for rs10033464) and both achieve genome-wide significance in the combined Icelandic samples (OR = 1.68, $P = 1.9 \times 10^{-30}$ for rs2200733; OR = 1.38, $P = 9.4 \times 10^{-9}$ for rs10033464). We also typed all the 18 Hap300 Illumina SNPs in the region around our signal in 404 of the additional AF cases and 2,036 of the additional controls. None of these SNPs remained significant after the association with rs2200733 and rs10033464 had been accounted for (Supplementary Table 3).

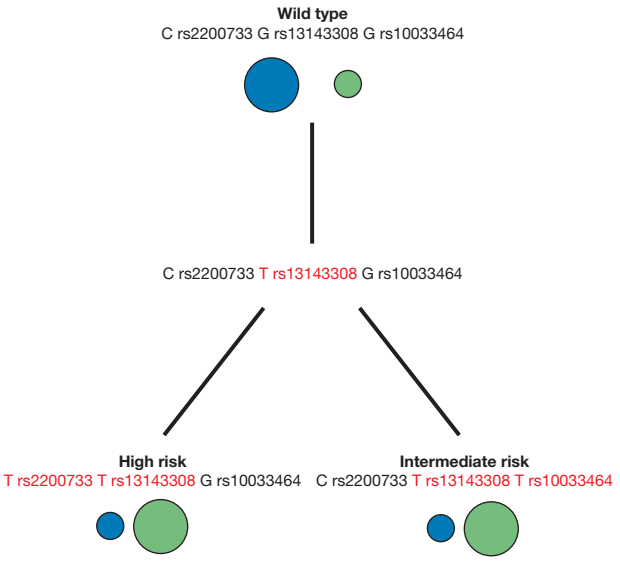

Figure 1 | Diagram of the haplotype structure at the associated region. Each edge in the graph corresponds to one mutation. The areas of the blue circles are proportional to the haplotype frequencies of the haplotypes in Iceland, and the areas of the green circles are proportional to the haplotype frequencies in Hong Kong. Note that the intermediary haplotype, shown in the middle of the graph, no longer exists with certainty in either of the two populations (its estimated frequency is less than 0.2% which is indistinguishable from genotyping errors).

In further attempts to replicate our results, we tested these variants for an association with AF in two populations of European ancestry, one from Sweden, consisting of 143 cases and 738 controls, and the other from the United States, consisting of 636 cases and 804 controls (Table 1). The association with rs2200733 was strongly replicated in both populations (OR = 2.01, $P = 0.00027$ in Sweden; OR = 1.84, $P = 9.8 \times 10^{-10}$ in the United States). The association with rs10033464 is weaker but was nonetheless replicated in the Swedish population (OR = 1.65, $P = 0.0087$) and was nearly significant in the US population (OR = 1.30, $P = 0.052$). When combined with the Icelandic samples, the association with rs2200733 was unequivocal (OR = 1.72, $P = 3.3 \times 10^{-41}$), and the significance of rs10033464 was well beyond the threshold of genome-wide significance (OR = 1.39, $P = 6.9 \times 10^{-11}$). If the multiplicative model is assumed, the population attributable risk (PAR) of the two variants combined is about 20% in populations of European ancestry.

Finally, we attempted to replicate these signals in a Han Chinese population from Hong Kong consisting of 333 cases and 2,836 controls. The association with rs2200733 T was significantly replicated (OR = 1.42, $P = 0.00064$), but the association with rs10033464 T was not significant, although the direction of association was consistent with that in the European samples (OR = 1.08, $P = 0.55$; Table 1). The T allele of rs2200733 is much more frequent in the Chinese population (the allelic frequency in controls is 0.528) than in those of European descent (allelic frequency in controls 0.098–0.139; Fig. 1), which is reflected in a greater joint PAR of about 35%, even though the estimated risk is less. The LD block containing the two variants is more fragmented in the Chinese CHB and Japanese JPT HapMap samples than in the CEU HapMap samples (Fig. 2). We therefore analysed several markers in the Hong Kong population that were in perfect LD with rs2200733 in the CEU samples but in imperfect LD in the CHB and JPT samples (Supplementary Table 4). These markers had a weaker apparent association with AF than with rs2200733, suggesting that the functional variants driving the association is located in the roughly 20-kilobase (kb) region around the original rs2200733 variant and defined by the SNPs that remain equivalent to rs2200733 in the CHB and JPT samples (red in Fig. 2).

For the initial Icelandic discovery samples, rs2200733 had a significantly higher OR than rs10033464 ($P = 0.041$). This held true in the replication samples, and overall there is a significant difference in the risks associated with the two variants ($P = 0.00019$ in the combined European samples and $P = 0.0099$ in Hong Kong). When genotype-specific ORs were studied, some deviation from the multiplicative model is detectable in the combined data set ($P = 0.018$ for

European samples; see Supplementary Table 5). Estimated risks of heterozygous carriers relative to non-carriers were similar, but homozygous carriers of rs2200733 T and rs10033464 T have estimated risks that were, respectively, higher and lower than predicted by a multiplicative model. A similar trend was seen in the Hong Kong samples, although the sample size was too small to have power to detect such deviations with significance. In the combined populations of European descent, the observed OR for individuals homozygous for rs2200733 T was 3.64 in comparison with individuals homozygous for the wild-type haplotype, and 1.77 for the Chinese population, showing that these variants are important components in any predictive modelling of AF.

The age at diagnosis of AF/AFl for the Icelandic samples is correlated with the two SNPs (diagnosis occurs 2.28 years earlier per T allele of rs2200733 and 1.10 years earlier per T allele of rs10033464; joint $P = 1.29 \times 10^{-6}$). The effect of the age at diagnosis was also evaluated by measuring the strength of association while stratifying by age at diagnosis. The association of the two variants is strongest in those diagnosed at a younger age, although the risk remains significant even in those diagnosed after reaching 80 years of age (Table 2). Information on age at diagnosis of AF was not available for the Swedish samples. The US samples consisted of two main groups: younger patients with either lone AF or AF and hypertension, and older AF cases who were mostly patients with haemorrhagic and ischaemic stroke. In both populations there is a clear trend towards a stronger association in younger AF cases than in older cases. Our analysis of the data did not suggest any differential association by sex (Table 2).

AF1 often accompanies AF, but it can occur in isolation[17]; we observed a strong association between the variants and the small subset ($N = 116$) of the AF1 Icelandic patients (OR = 2.60, 95% confidence interval (CI) = 1.83–3.68, $P = 7.5 \times 10^{-8}$ for rs2200733; OR = 1.94, 95% CI = 1.26–3.00, $P = 0.0028$ for rs10033464). Indeed, for rs2200733, the OR for these definite AFl cases is significantly higher than that for the cases with an AF phenotype ($P = 0.0026$), and is close to significantly higher for rs10033464 ($P = 0.084$). Our results suggest that while these traits share genetic risk factors, AFl is less influenced by phenocopies than AF.

Neither variant showed an association with obesity, hypertension or myocardial infarction in the Icelandic samples, all known risk factors for AF (observed OR < 1.1 in all instances; Supplementary Table 6). Although these negative results do not exclude the possibility that the new variants associate with these phenotypes, they do suggest, along with the high risk in US lone AF and earlier age at onset in carriers, that the new variants are not affecting the risk of AF through these known risk factors.

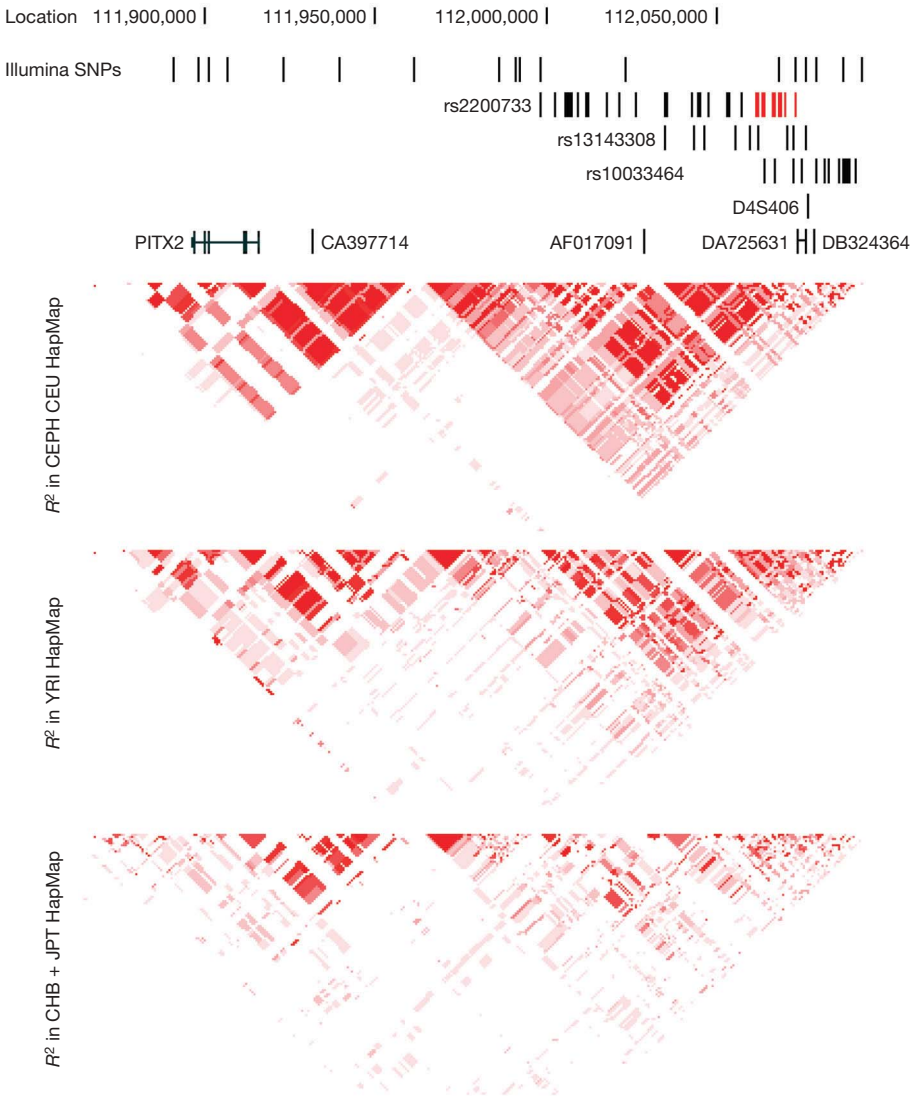

Figure 2 | Overview of a 200-kb genomic neighbourhood of rs2200733 and rs10033464. Predicted ESTs, the locations of the three main classes of equivalent SNPs in the CEU HapMap samples and an overview of the LD structure of the region in the CEU, CHB and JPT HapMap samples are shown. The SNPs that remain equivalent to rs2200733 in the CHB and JPT HapMap samples are shown in red.

Table 2 | Association by age at diagnosis in Iceland and by AF sub-phenotype in the United States

Sample (cases/controls)	Male (%)	Age (yr)	OR (95% CI)		P	Sex P
			rs2200733*	rs10033464*†		
Iceland‡						
Diagnosis at age ≤60 yr (510/17,714)	77.8	50.7 ± 8.4	2.12 (1.77–2.54)	1.69 (1.34–2.12)	6.3×10^{-18}	0.82
Diagnosis at age 60–70 yr (654/17,714)	66.2	65.6 ± 2.9	1.88 (1.60–2.21)	1.44 (1.18–1.77)	6.7×10^{-15}	0.58
Diagnosis at age 70–80 yr (958/17,714)	58.9	75.0 ± 2.8	1.60 (1.39–1.84)	1.23 (1.03–1.47)	7.5×10^{-11}	0.96
Diagnosis at age >80 yr (679/17,714)	47.4	85.6 ± 4.2	1.20 (1.01–1.43)	1.31 (1.08–1.60)	0.0044	0.36
United States						
Lone AF (251/804)	81.7	46.1 ± 11.5	2.32 (1.80–2.99)	1.68 (1.19–2.37)	1.2×10^{-10}	0.46
AF + hypertension (67/804)	74.6	54.5 ± 10.2	2.23 (1.43–3.48)	1.66 (0.90–3.04)	0.0010	0.54
Other AF (318/804)	52.8	75.2 ± 11.3	1.44 (1.12–1.84)	0.97 (0.69–1.37)	0.015	0.85

Each row contains the results from a joint analysis of two variants, rs2200733 T and rs10033464 T*. The numbers of cases and controls, the percentage of male cases and the age (mean ± s.d.) for cases are shown for each case-control study. The OR, with a 95% CI, and *P* values are shown for each variant. In addition a joint *P* value is shown for the combined effect of the two variants, as is a joint *P* value for testing whether there is a difference in the allelic frequency of the variants between the sexes within each subgroup of patients.
* Results of comparing rs2200733 T and rs10033464 T with the wild-type rs2200733 C, rs13143308 G, rs10033464 G haplotype.
† In the US samples, rs10033464 T was tagged by the rs2200733 C, rs13143308 T haplotype.
‡ The association analysis was adjusted for the relatedness of some of the individuals.

There is no known gene present in the LD block containing rs2200733 and rs10033464 (Fig. 2). The LD block contains one spliced expressed sequence tag (EST) (DA725631) and two single-exon ESTs (DB324364 and AF017091). Reverse-transcriptase-mediated polymerase chain reaction of complementary DNA libraries from various tissues did not detect the expression of these ESTs (Supplementary Information). The *PITX2* gene located in the adjacent upstream LD block is the gene closest to the risk variants. The protein encoded by this gene, the paired-like homeodomain transcription factor 2, is an interesting candidate for AF/AFl because it is known to be important in cardiac development by directing the asymmetric morphogenesis of the heart[13]. In a mouse knockout model, Pitx2 was shown to suppress a default pathway for sinoatrial node formation in the left atrium[14,15]. There is very little mRNA expression of *PITX2* in all easily accessible tissues, such as blood and adipose tissue, hampering the study of correlation between genotypes and expression levels. The next gene upstream of *PITX2* is *ENPEP*, an aminopeptidase responsible for the breakdown of angiotensin II in the vascular endothelium[18]. This gene is expressed more widely but the variants associated with AF showed no correlation with its expression in blood or adipose tissue (Supplementary Information). No other annotated genes are located within a 400-kb region upstream and 1.5-megabase regions downstream of the associated variants.

Thus, we have identified two variants on chromosome 4q25 that are strongly associated with AF in three distinct populations of European descent. The stronger variant also replicates well in a Chinese population in which it is much more common, and it has higher PAR than in populations of European descent. This association is particularly compelling in younger patients and in those with lone AF, but it is also present in older patients with more commonly encountered forms of AF. Although the mechanism for this association is unknown, our results provide a foundation for further studies on the molecular underpinnings of AF.

METHODS SUMMARY

Subjects. The Icelandic cases consisted of all patients diagnosed with AF and/or AFl at the two largest hospitals in the country from 1987 to 2005. The Swedish cases were recruited from 1996 to 2002 as a part of a continuing genetic epidemiology study, the South Stockholm Ischaemic Stroke Study. The cases in the United States were a mixture of stroke patients with a diagnosis of AF and younger consecutive patients with lone AF or AF with a coexisting diagnosis of hypertension. The Hong Kong cases were a collection of stroke and diabetes patients with a diagnosis of AF. The diagnosis of AF was confirmed by a 12-lead electrocardiogram in all study populations.

The Icelandic controls were chosen at random from individuals who had participated in other genetic studies at deCODE, excluding first-degree relatives of patients and controls (Supplementary Table 7). The Swedish controls were recruited from the same region as patients from blood donors (in 2001) and healthy volunteers (1990–94). The US controls were recruited from a large primary care practice and from patients participating in a study of haemorrhagic stroke. The Hong Kong controls were individuals without a diagnosis of AF.

Association analysis. In the genome-wide association stage, Icelandic case and control samples were assayed with Infinium HumanHap300 SNP chips (Illumina), containing 317,511 SNPs, from which 316,515 were polymorphic and satisfied our quality criteria.

A likelihood procedure described previously[19] was used for the association analyses. Allele-specific OR was calculated on the assumption of a multiplicative model[20]. Results from multiple case-control groups were combined by using a Mantel–Haenszel model[21]. In all tables, *P* values for both rs2200733 and rs10033464 were computed on the basis of comparison with the wild-type rs2200733 C, rs13143308 G, rs10033464 G haplotype carrying neither of the at-risk alleles.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 April; accepted 11 June 2007.
Published online 1 July 2007.

1. Go, A. S. *et al.* Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *J. Am. Med. Assoc.* **285,** 2370–2375 (2001).
2. Miyasaka, Y. *et al.* Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation* **114,** 119–125 (2006).
3. Arnar, D. O. *et al.* Familial aggregation of atrial fibrillation in Iceland. *Eur. Heart J.* **27,** 708–712 (2006).
4. Fox, C. S. *et al.* Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring. *J. Am. Med. Assoc.* **291,** 2851–2855 (2004).
5. Ellinor, P. T., Yoerger, D. M., Ruskin, J. N. & MacRae, C. A. Familial aggregation in lone atrial fibrillation. *Hum. Genet.* **118,** 179–184 (2005).
6. Chen, Y. H. *et al.* KCNQ1 gain-of-function mutation in familial fibrillation. *Science* **299,** 251–254 (2003).
7. Yang, Y. *et al.* Identification of a KCNE2 gain-of-function mutation in patients with familial atrial fibrillation. *Am. J. Hum. Genet.* **75,** 899–905 (2004).
8. Xia, M. *et al.* A Kir2.1 gain-of-function mutation underlies familial atrial fibrillation. *Biochem. Biophys. Res. Commun.* **332,** 1012–1019 (2005).
9. Olson, T. M. *et al.* Kv1.5 channelopathy due to KCNA5 loss-of-function mutation causes human atrial fibrillation. *Hum. Mol. Genet.* **15,** 2185–2191 (2006).
10. Hong, K., Bjerregaard, P., Gussak, I. & Brugada, R. Short QT syndrome and atrial fibrillation caused by mutation in KCNH2. *J. Cardiovasc. Electrophysiol.* **16,** 394–396 (2005).
11. Ellinor, P. T. *et al.* Mutations in the long QT gene, KCNQ1, are an uncommon cause of atrial fibrillation. *Heart* **90,** 1487–1488 (2004).
12. Ellinor, P. T., Petrov-Kondratov, V. I., Zakharova, E., Nam, E. G. & MacRae, C. A. Potassium channel gene mutations rarely cause atrial fibrillation. *BMC Med. Genet.* **7,** 70 (2006).
13. Franco, D. & Campione, M. The role of Pitx2 during cardiac development. Linking left–right signaling and congenital heart diseases. *Trends Cardiovasc. Med.* **13,** 157–163 (2003).
14. Faucourt, M., Houliston, E., Besnardeau, L., Kimelman, D. & Lepage, T. The pitx2 homeobox protein is required early for endoderm formation and nodal signaling. *Dev. Biol.* **229,** 287–306 (2001).
15. Mommersteeg, M. T. *et al.* Molecular pathway for the localized formation of the sinoatrial node. *Circ. Res.* **100,** 354–362 (2007).
16. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
17. Waldo, A. L. The interrelationship between atrial fibrillation and atrial flutter. *Prog. Cardiovasc. Dis.* **48,** 41–56 (2005).

18. Zini, S. *et al.* Identification of metabolic pathways of brain angiotensin II and III using specific aminopeptidase inhibitors: predominant role of angiotensin III in the control of vasopressin release. *Proc. Natl Acad. Sci. USA* **93,** 11968–11973 (1996).
19. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nature Genet.* **35,** 131–138 (2003).
20. 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,** 227–233 (1987).
21. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22,** 719–748 (1959).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the patients and their family members whose contribution made this work possible; the nurses at Noatun (deCODE's sample recruitment center), personnel at the deCODE core facilities, and M. Shea for the ongoing enrolment of patients at Massachusetts General Hospital; and A. Plourde and S. Makino for technical assistance.

Author Contributions D.F.G., D.O.A., A.H., S.G., P.T.E., J.R. U.T. and K.S. wrote the first draft of the paper. D.O.A., H.H., R.S., J.T.S. and G.T. collected and diagnosed the Icelandic AF samples. Ko.K. and J.H. collected and diagnosed the Swedish samples. K.L.F., S.M.G., M.S., P.K., C.A.M., E.E.S., J.R. and P.T.E. collected and diagnosed the US samples. M.C.Y.N., L.B., W.Y.S., K.S.W., J.C.N.C. and R.C.W.M collected and diagnosed the Hong Kong samples. A.H., S.G., A.S., A.J., A.B., T.B., V.M.B., G.A.H. and E.P. performed genotyping and experimental work. D.F.G., G.T., A.P., P.S., A.H. and A.K. analyzed the data. D.F.G., D.O.A., A.H., S.G., Kr.K., J.R., J.H., R.C.W.M., P.T.E, G.T, J.R.G., A.K., U.T. and K.S. planned, supervised and coordinated the work. All authors contributed to the final version of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.F.G. (daniel.gudbjartsson@decode.is) or K.S. (kstefans@decode.is).

METHODS

Icelandic study population. This study initially included all the patients consenting to participation who were diagnosed with AF and/or AFl (ICD (International Classification of Diseases) 10 diagnosis I48 and ICD 9 diagnosis 427.3) at Landspitali University Hospital in Reykjavik, the only tertiary referral centre in Iceland, and at Akureyri Regional Hospital, the second largest hospital in the country, from 1987 to 2005. All diagnoses were confirmed by a 12-lead electrocardiogram (ECG), which was read manually by a cardiologist. All cases were included, regardless of whether or not the patients had clinical symptoms, except those diagnosed only immediately after open cardiac surgery.

A set of 550 cases were successfully genotyped in accordance with our quality control criteria in a genome-wide SNP genotyping effort, using the Infinium II assay method and the Sentrix HumanHap300 BeadChip (Illumina). The age at diagnosis for this initial group of 550 patients (370 males and 180 females) was 72.5 ± 11.0 years (mean ± s.d.) and the range was 34.7–96.2 years. The validation group of 2,273 patients (1,359 males and 913 females) had an age at diagnosis of 70.5 ± 13.0 years and the range was 16.8–100.6. The AF/AFl-free controls (2,201 males and 2,275 females at the initial genome-wide screening aged 61.5 ± 15.8 years and 5,654 males and 7,597 females at the validation stage aged 61.9 ± 18.4 years) used in this study consisted of controls randomly selected from the Icelandic genealogical database and individuals from other ongoing related genetic studies at deCODE. Controls with first-degree relatives (siblings, parents or offspring) with AF/AFl, or a first-degree control relative, were excluded from the analysis.

The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Written informed consent was obtained from all patients, relatives and controls. Personal identifiers associated with medical information and blood samples were encrypted with a third-party encryption system as described previously[22].

Swedish study population. Patients with ischaemic stroke or transient ischaemic attack attending the stroke unit or the stroke outpatient clinic at Karolinska University Hospital, Huddinge unit, in Stockholm, Sweden, were recruited from 1996 to 2002 as part of an ongoing genetic epidemiology study, the South Stockholm Ischaemic Stroke Study (SSISS). The study was approved by the Bioethics Committee of Karolinska Institutet (Dnr 286/96 and 08/02). A diagnosis of AF in the Swedish samples was based on a 12-lead ECG. The fraction of males in the Swedish AF cases was 46.2% and the age at stroke diagnosis for the Swedish AF cases was 74.4 ± 8.7 years.

The Swedish controls used in this study were population-based controls recruited from the same region in central Sweden as the patients, representing the general population in this area. The individuals were either blood donors (recruited in 2001) or healthy volunteers (collected in 1990–94) recruited by the clinical chemistry department at the Karolinska University Hospital to represent a normal reference population. The fraction of males in the Swedish controls was 59.7% and the age at recruitment for the Swedish controls was 43.1 ± 12.3 years.

Study population in the United States. US subjects were enrolled in ongoing case-control and cohort studies at Massachusetts General Hospital between January 1998 and July 2006. All aspects of these studies have been approved by the local Institutional Review Board. Subjects who were enrolled in the case-control study consisted of patients hospitalized with acute ischaemic or haemorrhagic stroke confirmed by computed tomography or magnetic resonance imaging, admitted to a single acute-care hospital. Of the 328 haemorrhagic stroke patients recruited, 78 were diagnosed with AF and were used as cases for the current study; the remaining 250 were used as controls. A total of 170 ischaemic stroke patients had a diagnosis of AF and were treated as cases, but no ischaemic stroke patients were treated as controls. Patients were excluded for primary subarachnoid haemorrhage and for intracerebral haemorrhage secondary to head trauma, tumour, vascular malformation, or vasculitis. A total of 624 stroke-free controls were recruited from a large, primary care practice (more than 18,000 patients) serving the hospital's catchment area as well as the hospital's Anticoagulation Management Service; 70 of the 624 individuals collected as controls were diagnosed with AF and treated as cases for the purposes of the current study. Of all individuals used as controls, 50.9% were males and their age was 67.4 ± 12.3 years. All subjects or an accompanying informant provided informed consent for participation in genetic studies and were interviewed prospectively about medical history, medications and social and family history. The presence or absence of AF was documented prospectively through interviews and from a review of medical records.

The second part of the US subjects consisted of consecutive patients with lone AF or AF with a coexisting diagnosis of hypertension referred to the arrhythmia service who provided written informed consent for participation in genetic studies. Inclusion criteria were AF documented by ECG, and an age of 65 years or less. The exclusion criteria were structural heart disease as assessed by echocardiography, rheumatic heart disease, hyperthyroidism, myocardial infarction or congestive heart failure. Each patient underwent a physical examination and a standardized interview to identify past medical conditions, medications, symptoms and possible triggers for the initiation of AF. All patients were evaluated by 12-lead ECG, echocardiogram and laboratory studies. ECGs and echocardiograms were interpreted by using standard criteria.

Study population in Hong Kong. All subjects in the Hong Kong study population were of southern Han Chinese ancestry residing in Hong Kong. The cases consisted of 217 individuals (49.1% male, aged 68.1 ± 9.6 years) selected from the Prince of Wales Hospital Diabetes Registry[23], and 116 subjects (30.2% male, aged 76.1 ± 10.9 years) from the Stroke Registry[24]. All subjects were diagnosed by ECG as having AF. The controls consisted of 2,836 subjects without evidence of AF. Informed consent was obtained for each participating subject. This study was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong.

22. Grant, S. F. *et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature Genet.* **38,** 320–323 (2006).
23. Yang, X. *et al.* Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong Diabetes Registry. *Diabetes Care* **30,** 65–70 (2007).
24. Baum, L. *et al.* Methylenetetrahydrofolate reductase gene A222V polymorphism and risk of ischemic stroke. *Clin. Chem. Lab. Med.* **42,** 1370–1376 (2004).