SUPPLEMENTAL MATERIAL

A common connexin-40 gene promoter variant affects connexin-40 expression in human atria and is associated with atria fibrillation

by Robert C Wirka, Shamone Gore, David R. Van Wagoner, Dan E. Arking, Steven A. Lubitz, Kathryn L. Lunetta, Emelia J. Benjamin, Alvaro Alonso, Patrick T. Ellinor, John Barnard, Mina K. Chung, and Jonathan D. Smith

Methods

Study Subjects

Cleveland Clinic Atrial Tissue Bank. Atrial appendage tissues were obtained from 61 patients undergoing coronary artery bypass grafting, valve surgery or the Maze surgical AF ablation procedure. Supplemental Table 1 describes the clinical characteristics of these subjects.

Cleveland Clinic Lone Atrial Fibrillation GeneBank (CCAF). The discovery lone AF case cohort consisted of individuals from CCAF, which enrolled subjects >18 years of age with a history of AF and no significant coronary artery disease (≤50% coronary artery stenosis by coronary angiography, a normal stress test or no presumed coronary artery disease by clinical criteria; stress testing or coronary angiography was required for inclusion in males >age 50 years or females >age 55 years,) and normal left-ventricular function (left ventricular ejection fraction≥50%). Subjects were excluded from the study if they had significant valvular disease (>2+ mitral regurgitation, >2+ tricuspid regurgitation, >2+ aortic insufficiency, >2+ pulmonic insufficiency or any mitral,

tricuspid, aortic or pulmonic stenosis), coronary artery disease (>50% coronary artery stenosis, prior myocardial infarction, prior percutaneous coronary intervention, or coronary artery bypass surgery), a history of current congenital heart disease (except isolated patent foramen ovale), or a history of or current structural heart disease.

Illumina Controls. Healthy population subjects were identified from the Illumina iControl Database. Search criteria for the iControl Database were (1) Caucasian subjects of any age and any gender; (2) genotyping platform= Hap550 or Hap610. These criteria yielded 3180 subjects from studies 64, 65, 66 and 67 (www.illumina.com) carried out in New York, NY and Philadelphia, PA. Two of these studies contained only adults and two studies contained only children; thus, the age distribution of the population controls was bimodal, with one peak at a mean age of 50.2 years and another peak at a mean age of 9.2 years.

Atherosclerosis Risk in Communities (ARIC) study. [1] Cases of early-onset AF were identified prospectively from the ARIC cohort. Cases were subjects aged less than 66 years with AF documented by electrocardiography, or AF documented on hospital discharge or death certificate in the absence of clinical evidence of structural heart disease. Referent subjects from the ARIC study had no history of AF at baseline or follow up, and no history of MI, heart failure or valve disease at baseline.

The Massachusetts General Hospital Atrial Fibrillation Study (MGH)[2] prospectively enrolled subjects with early-onset AF defined as an age of onset less than 66 years and AF documented by electrocardiography. Subjects with structural heart disease, hyperthyroidism, myocardial infarction (MI) or heart failure were excluded from the MGH cohort.

Framingham Heart Study (FHS). [3-5] Referent subjects for use with the MGH cohort were selected from the Original and Offspring cohorts of the FHS, and included unrelated subjects 18-74 years of age and no history of AF at blood draw or in follow up, and no history of myocardial infarction, heart failure or valve disease at baseline.

Preparation of genomic DNA, total RNA and cDNA from human atrial tissue and blood

Atrial genomic DNA (gDNA) and total RNA were prepared from 61 atrial tissue samples using Qiagen kits. cDNA was produced with the iScript Select cDNA Synthesis Kit (BioRad) using an oligo$(dT)_{20}$ primer. Buffy coat gDNA was prepared from blood samples of 500 subjects in the LAFGB using the MasterPure DNA Purification Kit for Blood Version II (Epicentre Biotechnologies).

Genotyping of human atrial tissue genomic DNA

Atrial gDNA was genotyped for rs11552588, a perfect proxy for the promoter A SNP[6, 7] by the TaqMan (Applied Biosystems) method in 61 human atrial tissue samples. Genotype data for the promoter B SNP rs10465885 and 6 additional SNPs, shown in Figure 1a, were obtained from Illumina HumanHap 550 BeadChip data on the same samples.

Genotyping of human blood DNA

Genotypes for the promoter B SNP and another perfect proxy for the promoter A SNP (rs12408178, r^2=1) were obtained from genome wide SNP genotyping arrays. Genotyping platforms for each case-control cohort are outlined in Supplemental Table 2.

Because the Affymetrix 5.0 array used for the FHS subjects did not contain the promoter A or B SNPs, imputed values were used in the analyses of the MGH/FHS cohort. Imputation was performed using the program Mach with the CEU Hapmap Phase 2 panel. The imputed SNPs used in the analyses had good imputation quality scores (for example the promoter B SNP had an observed/expected variance of 1.019), which was expected due to the existence of good proxies on the Affymetrix 5.0 array. No SNPs deviated from Hardy Weinberg equilibrium using a p-value cutoff of 0.005.

Measurement of Cx40 mRNA levels

cDNAs from the 61 human atrial tissue samples were analyzed for total *Cx40* and transcript A expression on Illumina Human Ref-8 v2 Expression Beads Chips. The expression data was log2 transformed and quantile normalized. Standard quality control and probe filtering were performed. 31 samples with sufficient RNA remaining were subjected to quantitative real-time polymerase chain reaction (qPCR) for measurement of *Cx40* transcript A and B levels. PCR primers for specific amplification of *Cx40* transcript B were upper primer 5'-GGAGAACACAGACAGGCAGAG-3' and lower primer 5'-CCAGCACGAGCATACGGAATA-3'. PCR primers for specific amplification of *Cx40* transcript A were upper primer 5'-AAAAAGCGTGGGCAGTTGGAG-3' and lower primer 5'-CCAGCACGAGCATACGGAATA-3'. The VIC-labeled TaqMan probe for both *Cx40* transcripts was 5'-AGGAAGCTCCAATCGC-3'. A commercially designed FAM-labeled TaqMan assay was used to measure cardiac muscle alpha-actin (*ACTC1*) cDNA as an internal control. Multiplex PCR reactions were performed in an iCycler IQ real-time PCR instrument (BioRad) using the following thermal protocol: 95°C for 9

minutes, then 40 cycles of: 95°C for 15 seconds, 60°C for 1 minute, and 72°C for 45 seconds (image capture step). Raw data for *Cx40* transcript A or B and *ACTC1* reactions were extracted as threshold cycle numbers. *ACTC1*-adjusted *Cx40* transcript A or B threshold cycle values were calculated relative to the median threshold cycle value, and were used to calculate relative *Cx40* transcript A and *Cx40* transcript B expression levels using the $2^{-\Delta\Delta Ct}$ method[8]. The association of the promoter B SNP (rs10465885) genotype with *Cx40* transcript B levels was confirmed through a LOD score analysis using the residual sum of squares (rss) method, with the equation $LOD=(n/2) \times \log_{10}(rss_0/rss_{model})$, where 'n' refers to the number of samples[9].

Quantitative sequencing for analysis of allelic expression imbalance

Sample selection - Genotypes at the 8 SNPs within the *GJA5* gene region (shown in Fig. 1A) were determined in 61 atrial tissue samples, and unphased genotypes at each SNP from the 61 individuals were assembled into 122 phased haplotypes using PHASE v.2.1.1.[10, 11]. For each tissue sample used in the allelic expression imbalance assay, the phased haplotypes were analyzed to determine the allelic relationship between the Cx40 promoter SNPs and the indicator SNPs within the GJA5 coding region.

Cx40 transcript B - *Cx40* transcript B cDNA-specific PCR amplification primers were 5'-AGACAGGCAGAGGATTACAACACA-3' and 5'-GGGCCTCCATAGCTGTCATCA-3'. gDNA specific PCR amplification primers were 5'-GGGGCAAGAGCAGAATCCATAT-3' and 5'-GGGCCTCCATAGCTGTCATCA-3'.

Cx40 transcript A - *Cx40* transcript A cDNA-specific PCR amplification primers were 5'-GGTGGAAGAGGAACAACTGA-3' and 5'-CATACGGAATATGAAGAGGACA-

3’. gDNA-specific PCR amplification primers were 5’-GGTGGAAGAGGAACAACTGA-3’ and 5’-GAAATAGCGGGAGGGGTAAG-3’. Both gDNA and cDNA for each transcript were PCR amplified for 40 cycles and PCR products were purified using the QIAquick PCR Purification Kit according to the manufacturer’s instructions (Qiagen). The purified products were sequenced using internal primers (5’-TGGCAGTCAGCAAAGGAAGTAAAT-3’ for *Cx40* transcript B products, 5’- GAAGCAGCCAGAGTGTGAAGA-3’ for *Cx40* transcript A products) on an Applied Biosystems 3730xl DNA analyzer. Raw .ABI sequence files were analyzed using PeakPicker2[12], a program that calculates allelic expression ratios based on peak heights at the indicator SNP site in both cDNA and gDNA PCR products. PeakPicker2 was then used to normalize each allelic cDNA expression ratio to the average of the allelic gDNA ratios, assuming a gDNA allele ratio of 1:1. This resulted in a normalized allelic mRNA expression ratio for each sample.

Luciferase reporter gene assay

A promoter fragment ranging from -765bp to +94bp relative to the first base pair of exon 1B was PCR amplified from genomic DNA of patients homozygous for either the A or G allele at the *Cx40* transcript B promoter SNP, using the upper primer 5’-GGAAGCTTCTGACCCCATCTTTCCCCATAA-3’ and lower primer 5’-GCTCGAGTTGCTGCCTTGTGTTGTAATCCTC-3’. HindIII and XhoI sites are underlined in the upper and lower primers, respectively. Using the HindIII and XhoI restriction enzymes, the A and G allele-containing promoter fragments were then cloned into a promoter-less pXPII vector[13] directly upstream of the firefly luciferase gene. The

sequence of both the A and G allele-containing vectors was verified by direct sequencing to ensure that the *Cx40-B* promoter SNP site was the only site of sequence discordance between the two vectors. HL-1 murine atrial cardiomyocytes[14] were cultured at 37°C and 5% CO_2 in Claycomb medium (Sigma) supplemented with 10% fetal bovine serum, 0.1 mM norepinephrine, 2 mM L-glutamine and 100 U/mL each penicillin and streptomycin. HL-1 cells were cultured in antibiotic-free medium in 6-well plates for 24 hours prior to transfection with 5 µL Lipofectamine LTX (Invitrogen). Cells were transiently co-transfected with 1µg of either the A or G allele-containing promoter-luciferase plasmid and 1µg of the ß-galactosidase-containing plasmid pCH110. Additional cells were co-transfected with the empty pXPII vector and pCH110 as a background control. The promoter driven luciferase constructs were transfected in sextuplicate (one well of the A allele got contaminated and was excluded from the analysis). 48 hours later, cell lysates were assayed in duplicate for luciferase and ß-galactosidase activity with the Dual-Light reporter gene assay system (Tropix), and the mean background-subtracted (wells with no transfection) luciferase/ ß-galactosidase levels were calculated for each well.

Statistical Analyses

We used both one-way ANOVA and linear regression with the additive genotype model to correlate SNP genotypes with *Cx40* expression values. The linear regression model included covariates to correct for age, atrial rhythm at the time of surgery, and expression assay batch effects. The promoter A and B SNPs were tested for association with early onset lone AF (including only subjects with age at AF diagnosis $\leq$60 years) and lone AF (including subjects of all ages) in the case-control cohorts using logistic regression in a

log-additive genetic model. Cohorts from each center were analyzed individually and in a meta-analysis. For the analysis of the CCAF cohort, sex was included as a covariate and the EIGENSTRAT method of population stratification adjustment was used to help account for the different origins of the CCAF case and control sets. The meta-analyses used a fixed-effects model with invariance-variance weights of the genotype regression effects (log odds ratios). Power calculations used a minimum power of 0.80 and an alpha of 0.05. Statistical analyses were carried out using JMP 7, GraphPad PRISM, and R version 2.11 software. Meta-analyses were performed using the 'metagen' function of the R package Meta, version 1.6-0. Power calculations were performed using the htPower.cc function of the DGCgenetics software package, version 1.2.

References

1. Alonso A, Agarwal SK, Soliman EZ, Ambrose M, Chamberlain AM, Prineas RJ, Folsom AR. Incidence of atrial fibrillation in whites and African-Americans: the Atherosclerosis Risk in Communities (ARIC) Study. *Am Hear J.* 2009;158:111-117.

2. Ellinor PT, Yoerger DM, Ruskin JN, MacRae CA. Familial aggregation in lone atrial fibrillation. *Hum Genet.* 2005; 118: 179-84.

3. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB S, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* 2007; 165: 1328-1335.

4. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol.* 1979; 110: 281-290.

5. Dawber TR, Meadors GF, Moore FE,Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health.* 1951; 41: 279-281.

6. Juang JM, Chern YR, Tsai CT, Chiang FT, Lin JL, Hwang JJ, Hsu KL, Tseng CD, Tseng YZ, Lai LP. The association of human connexin 40 genetic polymorphisms with atrial fibrillation. *Int J Cardiol.* 2006; 27: 27.

7. Firouzi M, Ramanna H, Kok B, Jongsma HJ, Koeleman BP, Doevendans PA, Groenewegen WA, Hauer RN. Association of human connexin40 gene polymorphisms with atrial vulnerability as a risk factor for idiopathic atrial fibrillation. *Circ Res.* 2004; 95: e29-33.

8. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001; 25: 402-408.

9. Bhasin JM, Chakrabarti E, Peng DQ, Kulkarni A, Chen X, Smith JD. Sex Specific Gene Regulation and Expression QTLs in Mouse Macrophages from a Strain Intercross. *PLoS ONE.* 2008; 3: .

10. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001; 68: 978-989.

11. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 2003; 73: 1162-1169.

12. Ge B, Gurd S, Gaudin T, Dore C, Lepage P, Harmsen E, Hudson TJ, Pastinen T. Survey of allelic expression using EST mining. *Genome Res.* 2005; 15: 1584-1591.

13. Nordeen SK. Luciferase reporter gene vectors for analysis of promoters and enhancers. *BioTechniques.* 1988; 6: 454-458.

14. Claycomb WC, Lanson NA,Jr, Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A, Izzo NJ,Jr. HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci U S A.* 1998; 95: 2979-2984.

Supplemental Table 1. Characteristics of the 61 Atrial Tissue Donors

Mean Age (SD)	62.5 (11.2)
Sex	**N (%)**
Male	37 (60.7)
Female	24 (39.3)
Race	
European American	56 (91.8)
African American	3 (4.9)
Other	2 (3.2)
AF Burden	
No AF	9 (14.8)
Paroxysmal	9 (14.8)
Persistent	8 (13.1)
Permanent	31 (50.8)
Unknown	4 (6.5)
AF Type	
Lone AF	5 (9.6)
Mixed AF	47 (90.4)
AF on PreOp ECG	30 (49.2)
Hypertension	28 (45.9)
Diabetes	7 (11.5)
CHF	24 (39.3)
MI	13 (21.3)
Surgery Type	
Valve Repair/Replace	45 (73.8)
Maze	51 (83.6)
CABG	13 (21.3)
Medications	
Digitalis	23 (37.7)
Beta Blocker	39 (63.9)
Ca Channel Blocker	13 (21.3)
Class 1a AA	1 (1.6)
Class 1b AA	4 (6.6)
Class 1c AA	4 (6.6)
Class 3 AA	7 (11.5)
Amiodarone	4 (6.6)
ACE Inhibitor	36 (59.0)
Statin	22 (36.1)

SD, standard deviation; MI, myocardial infarction; CHF, congestive heart failure; AA, Antiarrhythmic

Supplemental Table 2. Characteristics of Lone AF Case-Control Study Cohorts

		Cleveland Clinic			ARIC		Massachusetts General Hospital & Framingham Heart Study	
		Cases	Controls	Controls	Cases	Controls	Cases	Controls
	Study	CCAF	iControlDB Study 64/65	iControlDB Study 66/67	ARIC	ARIC	MGH	FHS
	Genotyping Platform	Illumina 550/610	Illumina 550/610	Illumina 550/610	Affymetrix 6.0	Affymetrix 6.0	Affymetrix 6.0	Affymetrix 5.0*
All Ages Lone AF	Number	596	1423	1587	119	7395	375	1101
	Age at enrollment, mean ± SD	59.1 ± 10.7	50.2 ± 11.1	9.2 + 5.5†	52.9 + 5.4	53.7 + 5.6	53.4 ± 10.6	59.4 ± 9.8
	Sex, Male, n (%)	450 (75.5)	314 (22.1)	821 (51.7)	64 (53.8)	3283 (44.4)	304 (81)	495 (45)
Early Onset Lone AF	Number	384	1423	1587	45	7395	335	1067
	Age at AF diagnosis (cases) or enrollment (controls), mean ± SD	47.3 ± 9.9	50.2 ± 11.1	9.2 + 5.5†	56.7 + 2.7	53.7 + 5.6	44.3 ± 10.1	58.7 ± 10.0
	Sex, Male, n (%)	306 (79.7)	314 (22.1)	821 (51.7)	26 (57.8)	3283 (44.4)	276 (82.3)	477 (44.7)

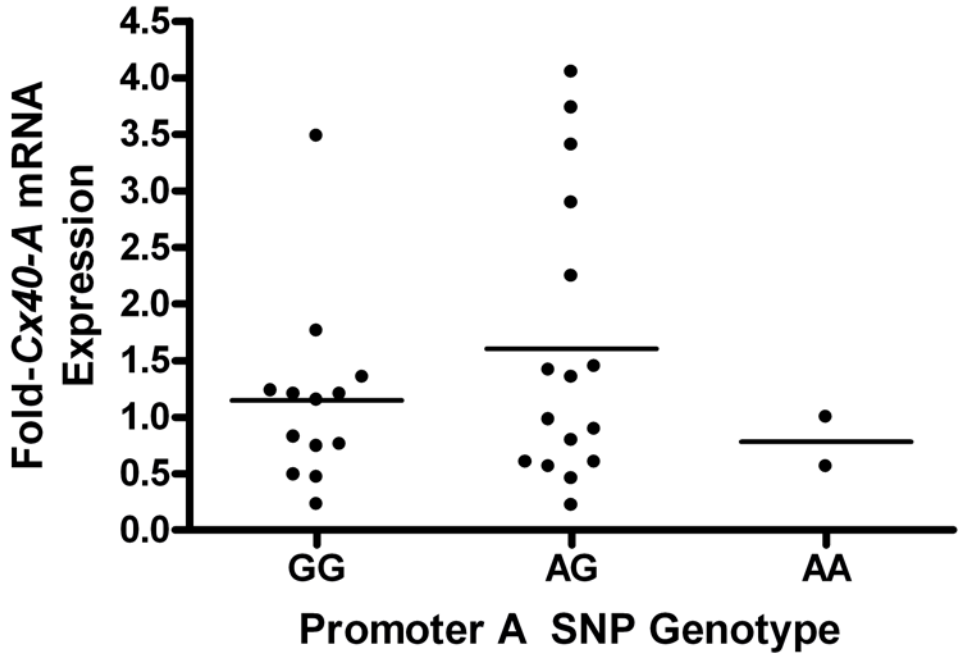
*Platform included the 50K Human Gene Focused Panel

† Two studies from the Illumina database contained pediatric patients

Supplemental Table 3. Association of promoter A and B SNPs with lone AF, all ages, additive genetic model

			Promoter A SNP			Promoter B SNP		
Study	**AF (n)**	**No AF (n)**	**MAF (%)**	**OR (95%CI)**	**P value**	**MAF (%)**	**OR (95%CI)**	**P value**
CCF	596	3010	24.1/22.7	1.01 (0.86-1.18)	0.946	50.4/46.9	1.09 (0.95-1.25)	0.214
MGH	375	1101	22.4/22.1	1.01 (0.79-1.29)	0.950	47.8/47.7	1.06 (0.86-1.29)	0.603
ARIC	119	7395	21.8/23.7	0.90 (0.65-1.22)	0.510	46.5/47.7	0.95 (0.73-1.24)	0.713
Meta	1090	11506	23.3/23.3	0.99 (0.88-1.12)	0.851	49.1/47.5	1.06 (0.95-1.17)	0.283

Supplemental Figure 1. Association of the promoter A SNP genotype with *Cx40* transcript A expression



Supplemental Figure 1. The promoter A genotype is not associated with *Cx40* transcript A expression (P=0.66). Relative levels of *Cx40* transcript A were quantified by real-time PCR in 31 atrial tissue samples.