GATA2 Is Associated with Familial Early-Onset Coronary Artery Disease

Jessica J. Connelly, Tianyuan Wang, Julie E. Cox, Carol Haynes, Liyong Wang, Svatı H. Shah, David R. Crosslin, A. Brent Hale, Sarah Nelson, David C. Crossman, Christopher B. Granger, Jonathan L. Haines, Christopher J. H. Jones, Jeffery M. Vance, Pascal J. Goldschmidt-Clermont, William E. Kraus, Elizabeth R. Hauser, Simon G. Gregory

The transcription factor GATA2 plays an essential role in the establishment and maintenance of adult hematopoiesis. It is expressed in hematopoietic stem cells, as well as the cells that make up the aortic vasculature, namely aortic endothelial cells and smooth muscle cells. We have shown that GATA2 expression is predictive of location within the thoracic aorta; location is suggested to be a surrogate for disease susceptibility. The GATA2 gene maps beneath the Chromosome 3q linkage peak from our family-based sample set (GENECARD) study of early-onset coronary artery disease. Given these observations, we investigated the relationship of several known and novel polymorphisms within GATA2 to coronary artery disease. We identified five single nucleotide polymorphisms that were significantly associated with early-onset coronary artery disease in GENECARD. These results were validated by identifying significant association of two of these single nucleotide polymorphisms in an independent case-control sample set that was phenotypically similar to the GENECARD families. These observations identify GATA2 as a novel susceptibility gene for coronary artery disease and suggest that the study of this transcription factor and its downstream targets may uncover a regulatory network important for coronary artery disease inheritance.

Introduction

Coronary artery disease (CAD) is the most common form of heart disease in the Western world. It affects more than 13 million Americans and is one of the leading causes of death in the United States [1]. CAD is a complex genetic disease; despite substantial evidence of a genetic contribution for CAD and its risk factors [2–10], the mode of inheritance does not follow Mendelian segregation. Complex genetic diseases are considered multifactorial in that they are characterized by the inheritance of multiple genetic variants acting in concert with environmental effects to promote the disease state.

Despite the obvious importance of environmental and behavioral risk factors, evidence of the genetic contribution to CAD is strong and consistent. The estimated relative risk of developing early-onset CAD in a first-degree relative is between 3.8 and 12.1, depending on the age of onset (AOO) of the proband, and higher risk correlates with earlier AOO [2,10]. Additionally, data from the Framingham Heart Study have shown an increased risk of incident cardiovascular disease in both age-sex-adjusted models (odds ratio [OR] = 1.55) and models adjusted for cardiovascular risk factors (OR = 1.45) for siblings of affected individuals [4].

Ten separate linkage screens have been performed in CAD sample sets to identify candidate genes that contribute to the genetic etiology of CAD, including our own GENECARD study [11]. Cumulatively, these screens have identified regions of linkage on Chromosomes 1, 2, 3, 4, 5, 7, 12, 13, 14, 16, 17, 19, and X [11–20], with genetic regions on 3q26–27 and 2q34–37 displaying cross-study CAD susceptibility [21]. Additionally, a recent genome screen by Bowden et al. [20] identified linkage to Chromosome 3q13 in a type 2 diabetes population subset by self-reported and clinically assessed CAD (The Diabetes Heart Study). This region is nearly identical to the Chromosome 3q13 linkage identified in the GENECARD study. Though a collection of CAD genetic regions have been identified, only a single gene mapping to Chromosome 13q, ALOX5AP, has been shown to contain variants that reproducibly predispose individuals to myocardial infarction, a specific manifestation of the cardiovascular disease phenotype [15,22–24].

Our goal is to characterize the underlying genetic mechanisms involved in the development of CAD. Through convergent analyses of linkage and expression data, we have identified the transcription factor GATA2 as having an increased potential to be involved with CAD susceptibility.

Received June 9, 2006; Accepted July 20, 2006; Published August 25, 2006

A previous version of this article appeared as an Early Online Release on July 20, 2006 (DOI: 10.1371/journal.pgen.0020139.eor).

Copyright: © 2006 Connelly et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: AOO, age of onset; APL, association in the presence of linkage; CAD, coronary artery disease; CADi, coronary artery disease index; LD, linkage disequilibrium; LOD, log of the odds; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism

* To whom correspondence should be addressed. E-mail: simon.gregory@duke.edu
Synopsis

Coronary artery disease (CAD) is the most common form of heart disease in the Western world and is one of the leading causes of death in the United States. CAD is inherited and is a complex genetic disease because it results from changes to multiple genes acting in concert with one another and the environment. The authors locate CAD susceptibility genes by convergence of techniques and identify variation within a gene of interest in an early-onset CAD population. If a specific variant is found more often in affected individuals or families than in controls, this can suggest that this gene variant is associated with disease. The authors have identified a gene, GATA2, which is located in a genomic region suspected to contain genes for CAD and displays expression patterns predictive of location of disease within human donor aortas. They have identified several GATA2 variants that segregate with CAD in a family-based early-onset CAD population and have further validated two of these associations in a separate young case-control sample affected with CAD. These data imply that the transcription factor GATA2 may play a role in CAD susceptibility and suggest that the study of GATA2 targets may uncover a set of GATA2-regulated genes important to CAD inheritance.

We utilized linkage data from our GENECARD genome-wide linkage study that sampled a cohort of families with at least two siblings with early-onset CAD (AOO in men <51 y and women <56 y) [25]. The most significant region of linkage, of the several regions that were identified [11], was localized to Chromosome 3q13 (multipoint log of the odds [LOD] = 3.50). GATA2 maps beneath the one-LOD down support interval for this linkage peak on Chromosome 3.

We also utilized gene expression patterns from Seo et al. [26] that identified GATA2 as one of the genes most predictive of the location within human donor aortas. Given that atherosclerotic disease has been shown to display increasing severity from proximal to distal locations within the aorta [27], these investigators theorized that differences in regional expression patterns within the aorta could be related to disease susceptibility.

Given the convergence of the results, we hypothesized that GATA2 may represent a newly identified susceptibility gene for CAD. Therefore, we investigated the haplotype structure of GATA2 to identify single nucleotide polymorphisms (SNPs) for genotyping within the gene in order to test the hypothesis that GATA2 is associated with CAD. We identified association in both a family-based sample (GENECARD) and in a validation dataset of nonfamilial CAD (CATHGEN). The results identify GATA2 as a novel CAD susceptibility gene and suggest that the study of this transcription factor and its downstream targets may undergo a regulatory network important for CAD.

Results

GATA2 SNP Selection and Genotyping

HapMap (http://www.hapmap.org) and Perlegen (http://genome.perlegen.com) databases were used to identify known SNPs within the GATA2 gene and 3,000 base pairs both upstream and downstream of the gene to account for putative promoter and downstream regulatory elements. A total of 12 SNPs were identified with a minor allele frequency (MAF) >10% in the Caucasian population (white Americans of European descent). LDSelect [28], which is used by SNPsselector [29], determined that these 12 SNPs, using an r² = 0.7, represent 12 linkage disequilibrium (LD) bins. All SNPs, with the exception of rs1806462, were genotyped in our GENECARD sample set (Figure 1, LD bin SNPs are black). Concurrently, we sought to determine whether additional coding SNPs within GATA2 could be identified. We sequenced each of the six GATA2 exons using a Caucasian sample consisting of 16 affected individuals and 16 individuals of unknown status. Five novel SNPs were identified within the 3’ untranslated region of the gene, whereas three previously validated SNPs were identified and two additional SNPs, which did not have a validated status in dbSNP (rs11708606 and rs10934857), were validated by our novel sequence (Table 1). We included rs10934857 in our analysis because LD calculations using the sequenced individuals suggested that this SNP may reside in a separate bin. The location of the five novel SNPs in relation to the 12 selected SNPs for our association studies is shown in Figure 1 (novel SNPs are grey). All 17 SNPs were genotyped in our association studies.

Single-Marker Family-Based Association in GENECARD

Genotyping was performed on the 12 a priori SNPs and the five novel SNPs in the GENECARD sample, which represents both the original genome linkage screen families as well as the follow-up collection set (Table 2, n = 1,101 families).
characteristics of this study group are defined in Tables 2 and 3 and elsewhere [11,25]. Pairwise LD between each SNP was measured using the Graphical Overview of Linkage Disequilibrium package separately in both Caucasian affected (GENECARD) probands with unaffected siblings) and unaffected (proband matched unaffected) individuals; no significant differences were seen between these two groups (unpublished data). We examined the LD structure of GATA2 in order to assess the quality of our tagSNP selection. The Haploview plot of these data in the unaffected Caucasian population (Figure 2) shows a weak block of LD in the 5′ end of the gene. The lack of haplotype block structure from our analysis was expected because of the prior selection of nonredundant haplotype tagging SNPs.

We employed the test for association in the presence of linkage (APL) to conduct association analysis at markers in GATA2 in order to make use of the large number of affected sibling pairs in the GENECARD sample, as well as to appropriately infer missing parental genotypes and to account for the correlation between transmission of parental marker alleles to multiple affected offspring due to linkage [30]. APL analysis of the 17 GATA2 SNPs identified five significant associations (p < 0.05) with early-onset CAD. We were not able to analyze the full seven SNP haplotypes in family-based APL analysis. Both of these haplotypes were present in the full family sample with remarkably strong associations (Table 5). These six pairs maintained strong association after Bonferroni correction (p < 0.0012 for both).

The pairwise haplotype results suggest that these haplotypes may be extended to include additional SNPs. We identified a single three-SNP haplotype and an independent four-SNP haplotype using the strongly associated haplotype pairs that shared a common SNP. These haplotypes were examined using the APL test. Both of these haplotypes were overtransmitted in GENECARD families (Table 5). We were not able to analyze the full seven SNP haplotypes in family-based APL analysis.

**Haplotype Analysis of GATA2 SNPs in GENECARD**

We have identified a region of association with CAD within the 3′ end of GATA2, which encompasses four SNPs. In order to more accurately identify the SNP(s) associated with early-onset CAD, we used the APL test to calculate the transmission frequencies of all possible haplotypes for pairwise combinations of the 17 SNP markers within GATA2 in the GENECARD sample. These frequencies were compared in a global test accounting for rare haplotypes [30] (Table S1). We observed six pairwise haplotypes of the 171 comparisons made, which involved seven SNPs with little to no LD (0.02 < r² < 0.28) with remarkably strong associations (Table 5). These six pairs maintained strong association after Bonferroni correction, with rs2713604__rs3803 and rs3803__rs2713594 exhibiting the strongest association (global Bonferroni corrected p < 0.0012 for both).

The pairwise haplotype results suggest that these haplotypes may be extended to include additional SNPs. We identified a single three-SNP haplotype and an independent four-SNP haplotype using the strongly associated haplotype pairs that shared a common SNP. These haplotypes were examined using the APL test. Both of these haplotypes were overtransmitted in GENECARD families (Table 5). We were not able to analyze the full seven SNP haplotypes in family-based APL analysis.

**Replicating Single-Marker Association of GATA2 SNPs in CATHGEN Case-Control Sample**

We identified five SNPs within GATA2, as well as six pairwise haplotypes and a single three-SNP and four-SNP haplotype that are significantly associated with early-onset CAD.

---

**Table 1. SNPs Identified by Sequencing the Six Exons of GATA2 in 32 Individuals Including Five Novel GATA2 SNPs**

| SNP          | Gene Location | Coding Change | NCBI35 bp Location | Alleles | MAF (%) |
|--------------|---------------|---------------|---------------------|---------|---------|
| RS1573858    | Exon 2        | PSP           | 129688558           | G/C     | 32.0    |
| RS2335052    | Exon 3        | A164T         | 129687649           | G/A     | 9.4     |
| RS11708606   | Intron 4      | —             | 129683504           | C/T     | 20.3    |
| RS10934857   | Exon 6        | 3′ UTR        | 129682360           | C/T     | 23.4    |
| RS3803       | Exon 6        | 3′ UTR        | 129682078           | C/T     | 20.3    |
| 3P0455b      | Exon 6        | 3′ UTR        | 129682377           | C/T     | 1.6     |
| 3P0456b      | Exon 6        | 3′ UTR        | 129682150           | C/T     | 1.6     |
| 3P0448b      | Exon 6        | 3′ UTR        | 129682014           | C/T     | 3.1     |
| 3P0459b      | Exon 6        | 3′ UTR        | 129681863           | C/T     | 1.6     |
| 3P0487b      | Exon 6        | 3′ UTR        | 129681678           | T/A     | 3.1     |

aMajor/minor allele; coding changes are defined by amino acid and location or by region within transcript.

DOI: 10.1371/journal.pgen.0020139.t001

---

**Table 2. GENECARD Sample Size**

| Sample                  | Number of families | Initial Genome Screen Families | Follow-Up Collection | All Families |
|-------------------------|---------------------|--------------------------------|----------------------|--------------|
| Number of families      | 420                 | 681                            | 1,101                |
| Individuals sampled     | 1,129               | 1,825                          | 2,954                |
| Affected individuals    | 1,005               | 1,255                          | 2,260                |
| Unaffected individuals  | 122                 | 481                            | 603                  |
| Unaffected > 60 y (M); 65 y (F) | 84                  | 192                            | 276                  |
| Affected sibling pairs  | 539                 | 427                            | 966                  |
| Discordant sibling pairs| 169                 | 467                            | 825                  |
| Discordant sibling pairs| 45                  | 56                             | 137                  |
| Parent–child trios      | 42                  | 35                             | 77                   |

DOI: 10.1371/journal.pgen.0020139.s001
CAD. We identified 656 cases and 410 controls meeting the study criteria (see Materials and Methods) from the CATHGEN study to validate these results in a phenotypically similar, nonfamilial young affected CAD case-control cohort. Baseline clinical characteristics in GENECARD probands (n = 1,101) and CATHGEN participants and unaffected controls are presented in Table 3. We genotyped the five SNPs significantly associated with disease status and the additional four SNPs that were identified in the haplotype analysis, in the CATHGEN cases and controls. Allelic association was examined using a multivariable logistic regression analysis. In order to test for a true genetic effect, we adjusted for race and sex or race, sex, and known CAD risk factors (see Materials and Methods). We identified significant associations in two of the five SNPs, rs2713604 and rs3803 (Table 6), with stronger association identified in the race, sex, and CAD risk factor–

![Pairwise LD between GATA2 SNPs](Figure 2. Pairwise LD between GATA2 SNPs)

LD was estimated in one unaffected Caucasian individual from each nonredundant GENECARD discordant sibling pair (n = 279). A similar pattern of LD was observed using the matched probands. DOI: 10.1371/journal.pgen.0020139.g002

### Table 3. Clinical Characteristics of GENECARD Probands and CATHGEN Participants

| Category | Characteristic | GENECARD Probands | CATHGEN Participants | Unaffected Controls |
|----------|----------------|-------------------|----------------------|---------------------|
|          |                | (n = 1,101)       | (n = 656)            | (n = 410)           |
| Mean age of onset (SD) | 43.9 (5.7)* | 46.4 (6.3) | N/A |
| Race | Caucasian | 89.0%* | 71.3% | 75.2% |
| | African American | 5.3%* | 22.0% | 19.9% |
| | American Indian | 4.0%* | 5.4% | 3.6% |
| | Asian | 0.5%* | N/A | N/A |
| | Other (unknown) | 1.3% | 1.2% | 1.3% |
| Sex | Male | 70.8%* | 79.2%* | 42.9% |
| | Ever smoked | 78.3%* | 69.1%* | 40.0% |
| | History of diabetes | 22.9%* | 33.3%* | 21.0% |
| | History of hypertension | 59.4%* | 69.6% | 61.7% |
| | History of myocardial infarction | 65.4%* | 52.1%* | N/A |
| | History of coronary artery bypass graft | 48.2%* | 39.7%* | N/A |
| Mean systolic blood pressure, mm Hg (SD) | 128.9 (20.4)* | 140.5 (24.3)* | 149.9 (23.6) |
| Mean diastolic blood pressure, mm Hg (SD) | 77.3 (12.5) | 78.3 (14.1) | 77.2 (13.6) |
| Total cholesterol, mg/dl (SD) | 212.3 (85.1)* | 192.9 (62.2) | 191.5 (48.7) |
| Low-density lipoprotein, mg/dl (SD) | 116.8 (87.4) | 109.8 (43.2) | 107.0 (35.8) |
| High-density lipoprotein, mg/dl (SD) | 35.2 (36.7)* | 40.5 (12.2)* | 52.3 (18.4) |
| Triglycerides, mg/dl (SD) | 216.6 (284.6) | 222.8 (260.8)* | 156.8 (125.5) |

*Significant difference (p < 0.05) between GENECARD probands and CATHGEN cases or CATHGEN cases and controls. Differences between the two groups were determined using a chi-square test (categorical variables) or a t-test (numerical variables).

DOI: 10.1371/journal.pgen.0020139.g003

SD, standard deviation; N/A, not applicable.
Table 4. GATA2 SNPs Are Associated with Early-Onset CAD in GENECARD

| SNP Combination | NCB135 bp Location | p-Valuea | Variance |
|-----------------|-------------------|----------|----------|
| RS6439129_3P0450  | 129695471          | 0.313    | 98.03    |
| RS6439129_3P0448  | 129695224          | 0.499    | 94.09    |
| RS2860228_3P0450  | 129692265          | 0.697    | 110.15   |
| RS1573858_3P0448  | 129688558          | 0.551    | 107.36   |
| RS2335052_3P0448  | 129687649          | 0.019    | 98.51    |
| RS26599689_3P0448 | 129685704          | 0.205    | 100.27   |
| RS2713603_3P0448  | 129683232          | 0.261    | 105.96   |
| RS2713604_3P0448  | 129683157          | 0.005    | 97.56    |
| RS2713605_3P0448  | 129683277          | 0.022    | 2.47     |
| RS10934857_3P0450 | 129682360          | 0.069    | 114.20   |
| RS3803_3P0450     | 129682150          | 0.653    | 13.81    |
| RS3803_3P0448     | 129682078          | 0.025    | 107.52   |
| RS4048_3P0450     | 129682014          | 0.745    | 36.17    |
| RS4046_3P0450     | 129681863          | 0.914    | 0.57     |
| RS4057_3P0450     | 129681878          | 0.548    | 29.02    |
| RS2713579_3P0450  | 129680802          | 0.004    | 106.48   |
| RS2713594_3P0450  | 129679198          | 0.109    | 126.21   |

p-Values ≤0.05 are indicated in red.

aAssociation in the Presence of Linkage (APL) test.

DOI: 10.1371/journal.pgen.0020139.t004

Table 5. Haplotype Analysis Identifies Significant Haplotypes Associated with Early-Onset CAD

| SNP Combination | GENECARD Global p-Value | GENECARD Global p-Value | CATHGEN Global p-Value |
|-----------------|-------------------------|-------------------------|------------------------|
| RS6439129_3P0448_3P0450 | 0.00019 | 0.00002 | 0.00001 |
| RS2713604_RS10934857_3P0448_3P0450 | 0.0001 | 0.00001 | 0.0001 |
| RS2713604_RS3803_RS10934857_3P0448_3P0450 | 0.0002 | 0.00002 | 0.00002 |
| RS2713604_RS3803_3P0448_3P0450 | 0.0001 | 0.00001 | 0.0001 |
| RS2713604_RS10934857_3P0448_3P0450 | 0.0008 | 0.00008 | 0.00008 |
| RS2713604_RS3803_3P0448_3P0450 | 0.0009 | 0.0009 | 0.0009 |
| RS2713604_RS10934857_3P0448_3P0450 | 0.0067 | 0.0067 | 0.0067 |
| RS2713604_RS3803_3P0448_3P0450 | ND | ND | ND |
| RS2713604_RS10934857_3P0448_3P0450 | 0.0512 | 0.0512 | 0.0512 |
| RS2713604_RS3803_3P0448_3P0450 | 0.2624 | 0.2624 | 0.2624 |

p-Values ≤0.05 are indicated in red.

aAdjustments made for race, sex, and history of hypertension, history of diabetes mellitus, body mass index, history of dyslipidemia, and smoking history.

DOI: 10.1371/journal.pgen.0020139.t005

Discussion

We identified a transcription factor, GATA2, through convergent analysis of linkage and expression data in an effort to define the underlying molecular mechanisms that lead to CAD. Genotyping and subsequent analysis of GATA2 tagging and novel SNPs in a family-based early-onset CAD sample identified five SNPs significantly associated with early-onset CAD. We validated the association of two of these SNPs, rs2713604 and rs3803, in an independent case-control dataset, as well as the direction of the association, thus identifying GATA2 as a susceptibility gene candidate for early-onset CAD.

The data suggest that polymorphisms in the 3’ end of GATA2 may increase susceptibility to developing CAD. We identified several novel SNPs within the sixth exon of GATA2 but did not identify association with early-onset CAD or LD between these SNPs and SNPs that were found to be associated with early-onset CAD. Although the functional relevance of our two most significantly associated SNPs remains unknown, one of these SNPs, rs3803, is conserved between multiple species (human, mouse, rat, and dog) and lies near one of three known
Polymorphisms and its downstream targets, is necessary in order to begin to understand the role this transcription factor plays in this disease.

Genetic and phenotypic heterogeneity are features of complex disease, particularly CAD. The pathobiology of CAD itself is sufficiently intricate in that the core of the disease is the formation of the atherosclerotic plaque. The pathways that are involved in lesion predisposition, formation, and disruption are numerous and can be modulated by the environment and other underlying genetic diseases (for example, diabetes or lipid disorders), as reviewed by Watkins and Farrall [34]. Additionally, the lack of a consistent clinical definition of CAD across studies further confounds genetic analysis. The detection of multiple, significant but non-overlapping chromosomal regions in ten genome screens is an indicator of the mixed genetic and phenotypic characteristics of this disease. We are encouraged by the recent data of Bowden et al. [20] that suggest a replication of the Chromosome 3q13 linkage evidence we initially reported, though phenotypic heterogeneity still remains. We also see evidence of this heterogeneity in our own dataset, both phenotypic [35] and in the linkage scores. We failed to detect evidence for linkage on Chromosome 3 in our follow-up dataset (maximum multipoint LOD score = 0.0 at GATA2), and though the result may be discouraging, the result is not surprising given the observed heterogeneity of linkage evidence across the ten previously published genome screens. Despite the evident heterogeneity in our CAD population, we detected an association signal with SNPs in GATA2 in two separate CAD datasets. We suggest that our validation of GATA2 warrants further study in our datasets, including an expansion to other racial groups such as Asians and African Americans, as well as in other studies to understand what role GATA2 might play in CAD development.

The results of our study are important for three main reasons. First, GATA2 is a transcription factor that is indispensable for all hematopoiesis [36]. It is essential for the development and differentiation of hematopoietic stem and progenitor cells [36–38]. It is expressed both in endothelial cells [37] and smooth muscle cells (J. Connelly, unpublished data), the two primary cell types comprising the aorta. It has been shown to regulate EDN1 transcription, a potent vasoconstrictor that is expressed only in endothelial cells [39]. It is also known to regulate several other

### Table 6. Replication of GATA2 SNPs Associated with Early-Onset CAD in the CATHGEN Case-Control Sample Set

| SNP          | Location (bp) | Minor | Controls | Cases | Allele MAF (%) | MAF (%) | p-Valuea (95% CI) | Genotype Odds Ratio | p-Valuea (95% CI) |
|--------------|---------------|-------|----------|-------|----------------|---------|-------------------|---------------------|-------------------|
| RS6439129    | 129695471     | A     | 31.8     | 32.6   | 0.126          | 1.3 (0.9, 1.8) | 0.174             | 1.2 (0.9, 1.5)     |
| RS2335052    | 12687649      | A     | 14.3     | 16.1   | 0.060          | 1.1 (0.8, 1.6) | 0.686             | 1.1 (0.8, 1.5)     |
| RS2713604    | 129683157     | T     | 30.5     | 33.3   | 0.011          | 1.5 (1.1, 2.1) | 0.029             | 1.3 (1.0, 1.7)     |
| 3P0455       | 129682377     | T     | 0.4      | 0.4    | 0.710          | 0.7 (0.2, 3.6) | 0.710             | 0.7 (0.2, 3.6)     |
| RS10934857   | 129682360     | T     | 27.1     | 25.0   | 0.062          | 0.7 (0.5, 1.0) | 0.093             | 0.8 (0.6, 1.0)     |
| RS3803       | 129682078     | A     | 21.8     | 19.2   | 0.028          | 0.7 (0.5, 1.0) | 0.052             | 0.8 (0.6, 1.0)     |
| 3P0448       | 129682014     | A     | 2.2      | 3.3    | 0.319          | 1.5 (0.7, 3.0) | 0.270             | 1.5 (0.7, 3.0)     |
| RS2713579    | 129680802     | G     | 31.6     | 33.0   | 0.164          | 1.3 (0.9, 1.7) | 0.208             | 1.2 (0.9, 1.5)     |
| RS2713594    | 129679198     | A     | 43.4     | 43.1   | 0.659          | 0.9 (0.7, 1.3) | 0.919             | 1.0 (0.8, 1.3)     |

p-Values ≤0.05 are indicated in red.

*aAdjustments made for race, sex, and history of hypertension, history of diabetes mellitus, body mass index, history of dyslipidemia, and smoking history.

bp, base pair, NCBI35; CI, confidence interval; MAF, minor allele frequency.

DOI: 10.1371/journal.pgen.0020139.t006

One of the more difficult aspects of genetic association analysis is the appropriate multiple comparison correction of the statistical significance of any given result. These corrections range from the most conservative Bonferroni correction, to false discovery rate approaches [33], to weighted corrections of combined data, to no correction at all. In our case we report uncorrected p-values; however, it should be noted that the results for the haplotype analysis would, in fact, survive a Bonferroni correction. We base our enthusiasm for GATA2 as a gene implicated in CAD susceptibility because we identified two separate significant markers in low LD in two independent populations. We believe that it is highly likely that they constitute two important independent markers for two reasons: they are physically linked within the same gene and in the same part of the gene and are significantly associated with disease in two independent populations. Further work to elucidate the function of GATA2 in CAD, including the CAD-associated polymorphisms and its downstream targets, is necessary in order to...
endothelial-specific genes, namely NOS3 [40], VWF [41], KDR [42], and PECAM1 [43], each of which can be linked to CAD [44–50]. The role GATA2 plays in hematopoiesis and endothelial cell function suggests that GATA2 may participate in endothelial progenitor cell potential and, thus, vascular disease propensity [51–55]. The effect of the CAD-associated polymorphisms on GATA2 function still remains to be elucidated. However, our data suggest that GATA2 may be functionally involved in the pathophysiology of CAD.

Second, our study results are important because of the strength of identifying significant associations with complex disease arising from a hypothesis-driven experimental design. We identified and genotyped tagging SNPs in a candidate gene that was identified through “genomic convergence” [56], a convergence of multiple lines of evidence. We used a family-based sample set that is heavily laden with a genetic burden for CAD, as well as a test of association (APL) that allows us to make full use of the GENECARD population while accounting for linkage. We selected a phenotypically similar case-control dataset to validate the GATA2 association we observed in GENECARD, which allowed us to identify two strong and separate associations with CAD. The use of convergence for candidate gene selection coupled with the thorough coverage of LD across a gene has allowed us to identify two significantly associated SNPs in two independent early-onset CAD samples.

Third, two other transcription factors have been implicated in cardiovascular phenotypes, MEF2A (myocardial infarction) [57–59] and USF1 (lipid traits) [60–62]. Complex genetic diseases, such as cardiovascular disease, are most likely the result of an accumulation of multiple small genetic changes that influence an individual’s ability to cope with biological and environmental effects. Transcription factors (and their cognate binding sites), which ultimately influence the expression of many downstream genes, may be important targets to characterize when considering how small genetic changes can influence multiple genetic outcomes. Slight changes to the level of a transcription factor in the cell can have a dramatic effect on the downstream targets of these factors. Hence, these types of genes will most likely be important in the dissection of complex human disease.

Our work suggests that common variants within GATA2 play a role in CAD, an important complex genetic disease. Identification of a transcription factor associated with CAD in two separate samples implies that GATA2-regulated genes may play a very significant role in CAD susceptibility and progression. These GATA2 target genes, which remain to be explored in this context, represent a rich set of candidate genes from which to dissect genetic contributions to coronary disease susceptibility.

Materials and Methods

Early-onset CAD family-based sample (GENECARD). GENECARD is a collaborative study involving investigators affiliated with the Duke Center for Human Genetics, the Duke University Center for Living, the Duke Clinical Research Institute, the Duke University Consortium for Cardiovascular Studies, and additional investigative sites of the GENECARD Study Network. The study is coordinated at Duke and located throughout five other international sites, and the study design has been previously reported [55]. In brief, collection of families began in March 1998 and was completed on 31 March, 2002. All study participants signed a consent form approved by the responsible institutional review board or local ethics committee.

The sample set used for the initial genome-wide linkage screen within the GENECARD project was composed of 493 affected sibling pairs in 420 families, where at least two siblings met the criteria for early-onset CAD. The characteristics of this study group are summarized in Table 2 and elsewhere [11,25]. We have expanded our collection to include an additional 681 families together with a larger number of unaffected family members. Unaffected participants were defined as siblings and relatives who have not been diagnosed with CAD and are older than 55 y of age (males) or older than 60 y of age (females). This additional collection has increased our sample size to 2,954 affected and unaffected individuals (Table 2).

Early-onset CAD case-control sample (CATHGEN). CATHGEN participants were recruited sequentially through the cardiac catheterization laboratories at Duke University Hospital (Durham, North Carolina, United States) with approval from the Duke Institutional Review Board. All participants undergoing catheterization were offered participation in the study and signed informed consent. Medical history and clinical data were collected and stored in the Duke Information System for Cardiovascular Care database maintained at the Duke Clinical Research Institute [65].

Controls and cases were chosen on the basis of the extent of CAD as measured by the CAD index (CADi). CADi is a numerical summary of coronary angiographic data that incorporate the extent and anatomical distribution of coronary disease [64]. CADi has been shown to be a better predictor of clinical outcome than extent of CAD [65]. Affected status was determined by the presence of significant CAD defined as a CADi ≥ 32 [66]. For patients older than 55 y of age, a higher CADi threshold (CADi ≥ 74) was used to adjust for the higher baseline extent of CAD in this group. Medical records were reviewed to determine the AOO of CAD, i.e., the age at first documented surgical or percutaneous coronary revascularization procedure, myocardial infarction, or cardiac catheterization meeting the above-defined CADi thresholds. The CATHGEN cases were stratified into a young affected group (AOO ≤ 55 y), which provides a consistent comparison for the GENECARD family study. Controls were defined as ≥60 y of age, with no CAD as demonstrated by coronary angiography and no documented history of cerebrovascular or peripheral vascular disease, myocardial infarction, or interventional coronary revascularization procedures. A comparison of clinical characteristics between GENECARD and CATHGEN probands and unaffected CATHGEN controls is presented in Table 3.

Novel SNP detection by GATA2 re-sequencing. GATA2 exons (six in total) were PCR amplified using standard conditions and sequenced using ABI Big Dye v3.1 and an ABI 3730 automated sequencer. Sequencing data were analyzed using Sequencher software (Gene Codes, Ann Arbor, Michigan, United States). All sequence amplicons were generated within 16 GENECARD affected individuals and 16 randomly ascertained individuals of unknown status. DNA derived from the blood of 17 Caucasian males and 15 Caucasian females was used for novel SNP discovery.

SNP genotyping. A minimal set of tagging SNPs with an MAF of >10% [28] was selected for genotyping in the GENECARD and CATHGEN samples to cover the predicted LD structure in GATA2 (~14 kilobases) using the SNPSelector program [29]. Additionally, five novel GATA2 SNPs identified by de novo sequencing were also genotyped in the GENECARD and CATHGEN collections. Genomic DNA for the GENECARD and CATHGEN samples was extracted from whole blood using the PureGene system (Gentra Systems, Minneapolis, Minnesota, United States). Genotyping in GENECARD was performed using the ABI PRISM Taqman SNP genotyping system (Applied Biosystems, Foster City, California, United States), which incorporates a standard PCR-based, dual fluor, allelic discrimination assay in a 384-well–plate format with a dual laser scanner. Allelic discrimination assays were purchased through Applied Biosystems, or, if the assays were not available, primer and probe sets were designed and purchased through Integrated DNA Technologies (Coralville, Iowa, United States). A total of 15 quality control samples—composed of six reference genotype controls in duplicate, two Centre d’Etude du Polymorphisme Humain pedigree individuals, and two interplate sample sets previously included in the 384-well plate. Genotyping in CATHGEN was performed using the Illumina BeadStation 500G SNP genotyping system (Illumina, San Diego, California, United States). Each Sentrix Array generates 1,536 genotypes for 96 individuals; within each individual array experiments, four quality controls were included. Two Centre d’Etude du Polymorphisme Humain pedigree individuals and two identical in-plate controls. Results of the Centre d’Etude du Polymorphisme Humain and quality-control samples were compared to identify...
possible sample plating errors and genotype calling inconsistencies. SNPs that showed mismatches on quality-control samples were reviewed by an independent genotyping supervisor for potential genotyping errors. All SNPs examined were successfully genotyped for 95% or more of the individuals in the study. Error rate estimates for SNPs meeting the quality control benchmarks were determined to be less than 0.2%.

Statistical analysis. All SNPs were tested for deviations from Hardy-Weinberg equilibrium in the affected and unaffected race-stratified groups. No such deviations were observed. Additionally, LD between pairs of SNPs was assessed using the Graphical Overview of Linkage Disequilibrium package [67] and displayed using Haploview [68]. Family-based association was tested using the APL test [30]. The APL test incorporates data from affected sibling pairs with available parental data and unaffected siblings in the analyses, effectively using all available information in the GENECARD families. The APL software appropriately accounts for the non-independence of affected siblings and calculates a robust estimate of the variance. APL results from markers with variance estimates of less than five are viewed as less reliable [69]. Allelic association in CATHGEN and the GENECARD probands from the United States was examined using multivariable logistic regression modeling adjusted for race and sex, and also for race, sex, and known CAD risk factors (history of hypertension, history of diabetes mellitus, body mass index, history of dyslipidemia, and smoking history) as covariates. These adjustments could hypothetically allow us to control for competing genetic pathways that might confound the true effect of the SNP on CAD, thereby allowing us to detect a separate CAD genetic effect. SAS 9.1 (SAS Institute, Cary, North Carolina, United States) was used for statistical analysis. APL software was used to identify haplotypes in GENECARD. The HaploStats package was used to identify and test for association of haplotypes in CATHGEN. HaploStats expands on the likelihood approach to account for ambiguity in case-control studies by using a generalized linear model to test for haplotype association, which allows for adjustment of nongenetic covariates [70]. This method derives a score statistic to test the null hypothesis of no association of the trait with the genotype. In addition to the global statistic, HaploStats computes score statistics for the components of the genetic vectors, such as individual haplotypes. The software MERLIN (Multipoint Engine for Rapid Likelihood Inference) was used for two-point and multipoint nonparametric linkage analysis [71].

Supporting Information

Table S1. Pairwise Haplotype Analysis within GENECARD Identities Six Significant SNP Pairs in GATA2

| Table S2. Replication of GATA2 SNPs Associated with Early-Onset CAD Adjusting for Race and Sex |
| Found at DOI: 10.1371/journal.pgen.0020139.s002 (19 KB XLS) |

| Table S3. Replication of Allele Direction of rs2713604 and rs3803 in GENECARD |
| Found at DOI: 10.1371/journal.pgen.0020139.s003 (15 KB XLS) |

| Table S4. The Risk and Protective Alleles of rs2713604 and rs3803 Do Not Occur on the Same Haplotype |
| Found at DOI: 10.1371/journal.pgen.0020139.s004 (15 KB XLS) |

Accession Numbers

The HUGO Gene Nomenclature Committee ([http://www.genenames.org](http://www.genenames.org)) accession numbers for the genes and gene products mentioned in this paper are ALOX5AP (436), EDN1 (3176), GATA2 (4171), KDRI (6307), MEPE2A (6903), NOS3 (7586), PECAM1 (8829), USF1 (12593), and VWF (12726).

The Online Mendelian Inheritance in Man ([http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM]) accession numbers for the genes and gene products mentioned in this paper are ALOX5AP (605700), EDN1 (131240), GATA2 (137250), KDR (191306), MEPE2A (606660), NOS3 (163729), PECAM1 (173445), USF1 (191523), and VWF (193400).

Acknowledgments

We thank the participants in the GENECARD and CATHGEN studies for their participation. We would also like to acknowledge the essential contributions of the following individuals to making this publication possible: Elaine Dowdy, the GENECARD Investigators Network; the CATHGEN Steering Committee Members (Chris Granger, Mike Sketch, Mark Donahue, Pascal Goldschmidt, Geoff Ginsburg, and Kristin Newby); Charlotte Nelson, Paul Hofmann, and Judy Stafford at the Duke Clinical Research Institute; Margaret Pericak-Vance, Eden Martin, Ren-Hua Chung, Julie Rombaut, Lamberton, and the staff at the Center for Human Genetics.

Author contributions. JJC, LW, SHS, JLH, JM, PJGC, WEK, ERH, and SGG conceived and designed the experiments. JJC and JEC performed the experiments. JJC, CH, SHS, DRC, ABH, SN, and ERH analyzed the data. TW, DCC, CBG, JLH, CJJH, PJGC, WEK, and ERH contributed reagents/materials/analysis tools. JJC, ERH, and SGG wrote the paper.

Funding. This work was supported by NIH grants HL073389 (ERH) and HL70542–03 (PJGC).

Competing interests. The authors have declared that no competing interests exist.

References

1. Thom T, Haase N, Rosamond W, Howard VJ, Rumsfeld J, et al. (2006) Heart disease and stroke statistics—2006 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 113: e85–151.
2. Shea S, Ottman R, Gabrieli C, Stein Z, Nichols A (1984) Family history as an independent risk factor for coronary artery disease. J Am Coll Cardiol 50: 945–953.
3. Ten Kate LP, Boman H, Daiger SP, Motulsky AG (1982) Familial aggregation of coronary heart disease and its relation to known genetic risk factors. Am J Cardiol 50: 945–953.
4. Muraliho JM, Pencina MJ, Nam BH, D’Agostino RB Sr., Wang, TJ, et al. (2005) Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. JAMA 294: 3117–3123.
5. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, et al. (2002) Heritability of death from coronary heart disease: A 36-year follow-up of 20,966 Swedish twins. J Intern Med 252: 247–254.
6. Lloyd-Jones DM, Nam BH, D’Agostino RB Sr., Levy D, Muraliho JM, et al. (2004) Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. A prospective study of parents and offspring. JAMA 291: 2204–2211.
7. Voss R, Cullen P, Schulze H, Assmann G (2002) Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. Int J Epidemiol 31: 1253–1262.
8. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, et al. (2004) Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. Lancet 364: 937–952.
9. Marenberg ME, Risch N, Berkmann LF, Flederus B, De Faire U (1994) Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med 330: 1041–1046.
susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. Hum Mol Genet 10: 2751–2765.

20. Bowden DW, Rudock M, Ziegler J, Lehtinen AB, Xu J, et al. (2006) Coincident linkage of type 2 diabetes, metabolic syndrome, and measures of subcutaneous fat in a genome scan of the diabetes heart study. Diabetes 55: 1985–1994.

21. Chiodini BD, Lewis CM (2005) Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. Arterioscler Thromb Vasc Biol 25: 1863–1868.

22. Kajimoto K, Shioji K, Ishida C, Iwanaga Y, Kokubo Y, et al. (2005) Validation of the association between the gene encoding 5-lipoxygenase-activating protein and myocardial infarction in a Japanese population. Circ J 69: 1029–1034.

23. Lohmueller E, Goeckendorf A, Mueller JC, Org T, Wichmann F, et al. (2005) ALOX5AP gene and the PDE4D gene in a central European population of stroke patients. Stroke 36: 731–736.

24. Martin ER, Bass MP, Hauser ER, Stenger JE, Pericak-Vance MA, et al. (2005) Selecting a population of stroke patients. Stroke 36: 731–736.

25. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

26. Helgadottir A, Manolescu A, Helgason A, Thorleifsson G, Thorsteinsdottir U, et al. (2004) Variable of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. Nat Genet 38: 68–74.

27. Hauser ER, Mooser V, Crossman DC, Haines JL, Jones CH, et al. (2005) Design of the genetics of early-onset cardiovascular disease (GENEcard) study. Am Heart J 145: 602–613.

28. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

29. Parks AL, Hingorani AD, Liang CF, Fatibene J, Lyon A, Monteith S, et al. (1999) A common variant of the endothelial nitric oxide synthase (Gln298–Asp) is a major risk factor for coronary artery disease in the UK. Circulation 100: 1515–1520.

30. Wei H, Fang L, Chowdhury SH, Gong, N., Xiong Z, et al. (2004) Platelet-endothelial cell adhesion molecule-1 gene polymorphism and its soluble level are associated with severe coronary artery stenosis in Chinese Singaporean. Clin Biochem 37: 1091–1097.

31. Feng L, Wei H, Chowdhury SH, Gong N, Song J, et al. (2005) Association of Leu125Val polymorphism of platelet endothelial cell adhesion molecule-1 (PECAM-1) gene and soluble PECAM-1 with coronary artery disease in Asian Indians. Indian J Med Res 121: 92–99.

32. Vasa M, Fichtlscherer S, Aicher A, Goetz HH, Urbich C, et al. (2001) Number and migratory activity of circulating endothelial progenitor cells inversely correlate with risk factors for coronary artery disease. Circ Res 89: E1–E7.

33. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach for multiple testing. J R Stat Soc B Methodol 57: 289–300.

34. Watkins H, Farrall M (2006) Genetic susceptibility to coronary artery disease: From promise to progress. Nat Rev Genet 7: 163–175.

35. Wang L, Fang C, Topol SE, Topol EJ, Wang Q (2005) Mutation of MEF2A in some 3q. Arterioscler Thromb Vasc Biol 23: 1863–1868.

36. Hauser MA, Li YJ, Takeuchi S, Walters R, Noureddine M, et al. (2003) Reduced number of circulating endothelial progenitor cells predicts future cardiovascular events: Proof of concept for the clinical importance of endogenous vascular repair. Circulation 111: 2981–2987.

37. Hauser MA, Li YJ, Takeuchi S, Walters R, Noureddine M, et al. (2003) Genomic convergence: Identifying candidate genes for Parkinson’s disease by combining serial analysis of gene expression and genetic linkage. Hum Mol Genet 12: 671–677.

38. Schmidt-Lucke C, Rossig L, Fichtlscherer S, Vasa M, Britten M, et al. (2005) Vascular biology redefined? Trends Biotechnol 23: 444–452.

39. Caplice NM, Doyle B (2005) Vascular progenitor cells: Origin and mechanisms of mobilization, differentiation, integration, and vasculogenesis. Stem Cells Dev 14: 122–139.

40. Zhang R, Min W, Sessa WC (1995) Functional analysis of the human endothelial cell adhesion molecule-1 gene polymorphism for association analyses using linkage disequilibrium. Am J Hum Genet 74: 196–199.

41. Xu H, Gregory SG, Hauser ER, Stenger JE, Pericak-Vance MA, et al. (2005) Circulating endothelial progenitor cells, vascular function, and cardiovascular risk. N Engl J Med 348: 595–600.

42. Minami T, Rosenberg RD, Aird WC (2001) Transforming growth factor- 

43. Gumina RJ, Kirschbaum NE, Piotrowski K, Newman PJ (1997) Characterization of the mechanisms of mobilization, differentiation, integration, and vasculogenesis. Stem Cells Dev 14: 122–139.

44. Kinlay S, Behrendt D, Wainstein M, Beltrame J, Fang JC, et al. (2005) Role of endothelin-1 in the activation of human atherosclerotic coronary vascular smooth muscle cells. Circ Res 96: 1371–1378.

45. Dimmeler S, Zeiher AM (2004) Vascular repair by circulating endothelial progenitor cells: The missing link in atherosclerosis? J Mol Med 82: 671–677.

46. Doyle B, Caplice N (2005) A new source of endothelial progenitor cells—

47. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

48. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

49. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

50. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

51. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

52. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

53. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

54. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

55. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

56. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.

57. Seo D, Wang T, Dressman H, Herderick EE, Iversen ES, et al. (2004) Gene expression phenotypes of atherosclerosis. Arterioscler Thromb Vasc Biol 24: 2442–2448.