SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
DISEASE GENETICS
CARDIOVASCULAR DISEASES

Received
3 December 2013

Accepted
4 December 2013

Published
8 January 2014

Correspondence and requests for materials should be addressed to E.E. (elahe.elahi@gmail.com; elaheelahi@ut.ac.ir)

* Current address: Dept. of Cardiology & Radiology, Stanford School of Medicine, Stanford, CA 94305-5454.

Mutation in *ST6GALNAC5* identified in family with coronary artery disease

Kolsoum InanlooRahatloo[1,2]*, Amir Farhang Zand Parsa[3], Klaus Huse[2], Paniz Rasooli[1], Saeid Davaran[4], Matthias Platzer[2], Marcel Kramer[5], Jian-Bing Fan[6], Casey Turk[6], Sasan Amini[6], Frank Steemers[6], Kevin Gunderson[6], Mostafa Ronaghi[6] & Elahe Elahi[1,7]

[1]School of Biology, College of Science, University of Tehran, Tehran, Iran, [2]Genome Analysis, Leibniz Institute for Age Research–Fritz Lipmann Institute, Jena, Germany, [3]Department of Interventional Cardiology, Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran, [4]Tehran Heart Center, Tehran University of Medical Sciences, Tehran, Iran, [5]Integrated Research and Treatment Center, Center for Sepsis Control and Care, Jena University Hospital, Jena, Germany, [6]Illumina, San Diego, California, USA, [7]Dept. of Biotechnology, College of Science, University of Tehran, Tehran, Iran.

We aimed to identify the genetic cause of coronary artery disease (CAD) in an Iranian pedigree. Genetic linkage analysis identified three loci with an LOD score of 2.2. Twelve sequence variations identified by exome sequencing were tested for segregation with disease. A p.Val99Met causing mutation in *ST6GALNAC5* was considered the likely cause of CAD. *ST6GALNAC5* encodes sialyltransferase 7e. The variation affects a highly conserved amino acid, was absent in 800 controls, and was predicted to damage protein function. *ST6GALNAC5* is positioned within loci previously linked to CAD-associated parameters. While hypercholesterolemia was a prominent feature in the family, clinical and genetic data suggest that this condition is not caused by the mutation in *ST6GALNAC5*. Sequencing of *ST6GALNAC5* in 160 Iranian patients revealed a candidate causative stop-loss mutation in two other patients. The p.Val99Met and stop-loss mutations both caused increased sialyltransferase activity. Sequence data from combined Iranian and US controls and CAD affected individuals provided evidence consistent with potential role of *ST6GALNAC5* in CAD. We conclude that *ST6GALNAC5* mutations can cause CAD. There is substantial literature suggesting a relation between sialyltransferase and sialic acid levels and coronary disease. Our findings provide strong evidence for the existence of this relation.

Coronary artery disease (CAD) is a leading cause of death worldwide[1]. Its most severe outcome is myocardial infarction (MI). Atherosclerosis which causes accumulation of plaques within coronary arteries is the major cause of CAD. CAD is a paradigmatic complex disorder caused by multiple physiologic, genetic, and environmental factors. These factors include family history of CAD or MI, smoking, advanced age, male gender, high plasma low-density lipoprotein (LDL) cholesterol, high plasma triglycerides, high blood pressure, and obesity[1]. The genetic component of CAD is evidenced by family clustering and results of twin studies; estimates of heritability range from 30% to 60%[2,3]. For CAD, there is a trend of decreased heritability with increased age of group studied, and this predicts that genetic investigations on early onset CAD and MI patients may be more fruitful[4].

Risk factors that contribute to atherosclerosis and CAD were first identified in epidemiologic studies[5]. Subsequently, candidate gene approaches[6], animal model studies[7], and genome-wide association (GWA) studies[8,9] were used to identify CAD-relevant genes and loci. The GWA studies have identified numerous loci that potentially contribute to the disease, but each of these exhibit only a modest effect (Odds Ratio < 1.3)[8]. Pedigree based genome-wide linkage analysis is an alternative approach, suitable for identification of sequence variations with large contributions and unknown pathways relevant to disease etiology. With respect to CAD, *MEF2A* encoding Myocyte-specific enhancer factor 2A and *LRP6* encoding LDL receptor-related protein 6 were identified as causative genes by pedigree based linkage analyses[10,11]. Some experiments have supported the potential role of *LRP6*[12], but the contribution of *MEF2A* to CAD remains controversial[13–15]. Here, we present results of genetic analysis on an early onset CAD pedigree. We tried to the best of our ability within the scope of the present research to address the caveats of using pedigree based genetic analysis for identification of CAD relevant genes[16]. Our results strongly suggest that *ST6GALNAC5* is the causative gene in the pedigree studied. A second mutation in *ST6GALNAC5* was observed in two other individuals upon sequencing all exons of the gene in 160 additional CAD patients. The mutation in one of the patients was also present in his CAD affected sibling, but absent in two

unaffected siblings. Sequencing of the coding exons in 100 elderly Iranian controls did not reveal putative disease causing variations. Comparison of numbers of individuals who carry rare sequence variations in *ST6GALNAC5* that cause amino acid changes in combined control cohorts from Iran and the United States (900 individuals) and in combined patient cohorts from the two populations (310 patients) revealed that the frequency of such variations is higher among patients ($P = 0.003$). Expression of *ST6GALNAC5* in human derived heart cells which was previously evidenced in microarray based tissue transcriptome studies (accession numbers in EMBI-EBI (http://www.ebi.ac.uk/expressionprofiler/): E-GEOD-2240, E-GEOD-40231, E-GEOD-3526) was here confirmed by cDNA synthesis[17]. It was shown in an *in vitro* assay that both of the CAD associated mutations caused enhanced enzymatic activity.

Results

Genetic analysis. The occurrence of CAD in pedigree CAD-105 was suggestive of Mendelian inheritance, but did not definitively distinguish between autosomal dominant and autosomal recessive modes of inheritance. Parametric and non-parametric analyses of genome-wide SNP genotyping data on the eight available individuals of the pedigree showed that highest logarithm of odds (LOD) scores were confined to two close regions on chromosome 1 and one region on chromosome 19 (Fig. S1). Each region spanned 3.5 to 9.0 centimorgans, and together they contained over 200 annotated protein coding genes (Fig. S1; Table S2). The highest LOD score (2.2) was obtained under an autosomal dominant mode of inheritance. Other than these scores on chromosomes 1 and 19, a score higher than 1.0 was not obtained under any of the models tested anywhere on the genome. Although results of linkage analysis were not definitive, the locus on chromosome 1p31 was most promising because previous studies had linked this locus to peripheral arterial occlusive disease[18], elevated apolipoprotein B (apoB) levels[19], and elevated triglyceride levels in families with premature CAD and MI[20].

The Illumina TruSeq® Exome Enrichment platform was chosen for exome sequencing of DNAs of affected individuals III-1 and III-2. Over eight gigabases of high quality sequences for each subject were generated. During development of this platform, it was shown that targeted libraries with up to 11 bp staggered substitutions, 15 bp consecutive substitutions and 15 bp indels compared to the probes, are efficiently enriched without much loss of coverage in the targeted regions (Online Data Supplements text; Fig. S2–S5). The ability of the probes to enrich sequencing libraries that vary significantly in homology (>80–85% homology) enables enrichment of highly polymorphic regions like HLA, and demonstrates that a wide variety of variants can be discovered using the Exome enrichment assay.

Two hundred eighty three sequence variations distributed in 259 genes that met the selection criteria were identified in the exome sequence data of patients III-1 and III-2 (Table S3). Six of the variations were within five genes positioned within or very close to the peak linkage loci. These variations and those assessed to have potential functional relevance to CAD and CAD risk factors were further pursued; these constituted 12 variations distributed in 11 genes (Table S4). Exome data reported all the variations in the heterozygous state in both patients, and all caused amino acid changes. The validity of the variations in patients III-1 and III-2 was confirmed by direct Sanger sequencing (Table S5, Fig. S6A). The SIFT (http://sift.jcvi.org) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) bioinformatics tools predicted that two of the 12 variations would be damaging to protein function. Sequencing of PCR amplicons that included the variations in other members of the pedigree showed that only c.G295A in *ST6GALNAC5* that causes p.Val99Met segregated with disease status (Fig. 1A). Sequencing was performed on the DNAs of all pedigree members used in the linkage analysis as well as in the DNAs of one affected (III-6) and one unaffected individual (II-4) which became available after the linkage analysis had been performed. The variation in *ST6GALNAC5* was observed in all seven living affected family members and absent in the three elderly unaffected individuals. All affecteds were heterozygous carriers, except III-3 who harbored the mutation in the homozygous state. In the later stages of the study, when younger individuals of the pedigree (average age = 30.2 yrs.; range 19–47 yrs.) became available, these were screened and many were found to be heterozygous or homozygous carriers of the mutation (Table S1-A, Fig. 1A). It was considered that the absence of CAD in the younger individuals with the mutation may be due to their age, and that they are at risk of CAD as they grow older. *ST6GALNAC5* encodes sialyltransferase 7e, also known as ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5 and α2,6-sialyltransferase 5. Amino acid p.Val99 that was changed by the mutation in pedigree CAD-105 individuals is highly conserved in paralogous and orthologous proteins; it remains unchanged in species as evolutionarily distant as humans and zebra fish (Table S6). The c.G295A nucleotide change in *ST6GALNAC5* was not observed in 800 ethnically matched control individuals without cardiac disorders. It was concluded that the sequence variation that caused p.Val99Met in sialyltransferase 7e is the likely cause of early onset CAD in pedigree CAD-105 and that CAD inheritance in the pedigree is autosomal dominant. The conclusion is based on results of whole genome linkage and exome sequencing, bioinformatics prediction of deleterious effect of the variation on protein function, segregation of the mutation with disease status within the family, evolutionary conservation of the altered amino acid in sialyltransferase 7e, and absence of the sequence variation in 800 ethnically matched healthy individuals.

To obtain confirmatory evidence for the contribution that *ST6GALNAC5* may make to CAD, all exons of the gene were sequenced in 100 Iranian control individuals and in 160 confirmed unrelated Iranian CAD patients. In addition to a variation in the 5′-NC region, eight variations that affected codons were observed, three in both controls and patients, one only in controls, and four only in patients (Table S7). The variations observed in both controls and patients or only in controls all created synonymous codons and are, therefore, unlikely to be associated with CAD. All four variations found only in patients caused amino acid changes, although only the one that caused p.*337Qext*20 among these was predicted by bioinformatics tools to damage protein function (Fig. S6B). This stop-loss variation was observed in two unrelated patients. One of the patients who harbored the p.*337Qext*20 mutation had one CAD affected sibling and two unaffected siblings; genetic analysis of members of his family showed segregation of the mutation with disease status (Fig. 1B). CAD-relevant phenotypic features of the siblings in this family are shown in Table S1-B. Age at diagnosis of CAD in the two affecteds was relatively early (51 and 54 years). The number of affected and unaffected individuals in the family is low, but there was no clear difference in CAD risk factors between the two affected and two unaffected individuals at their present age. The causative nucleotide variation was absent in the DNAs of 800 control individuals. The variation causes addition of 20 amino acids to the COOH-terminus of the encoded protein. The large catalytic domain of sialyltransferase 7e is positioned at the COOH-terminus, suggesting the mutation may alter catalytic activity[21]. Based on conservation during evolution, four conserved motifs (sialylmotifs) ranging in size from 4 to 48 amino acids in the C-terminus of vertebrate sialyltransferases are defined. The last amino acid in the most C-terminal sialylmotif (in ST6GALNAC5) is residue 301. The wild type protein has 35 additional amino acids and ends at residue 336[21,22]. The p.*337Qext*20 mutation increases the number of amino acids after the last sialylmotif from 35 to 55 amino acids, and the increased length may have allosteric effects that affect the unknown functions of the conserved motifs. Additionally, the added amino acids (Q G M S M P D C N P R Y S L H Q T P R H **Stop**) include one cysteine and three proline

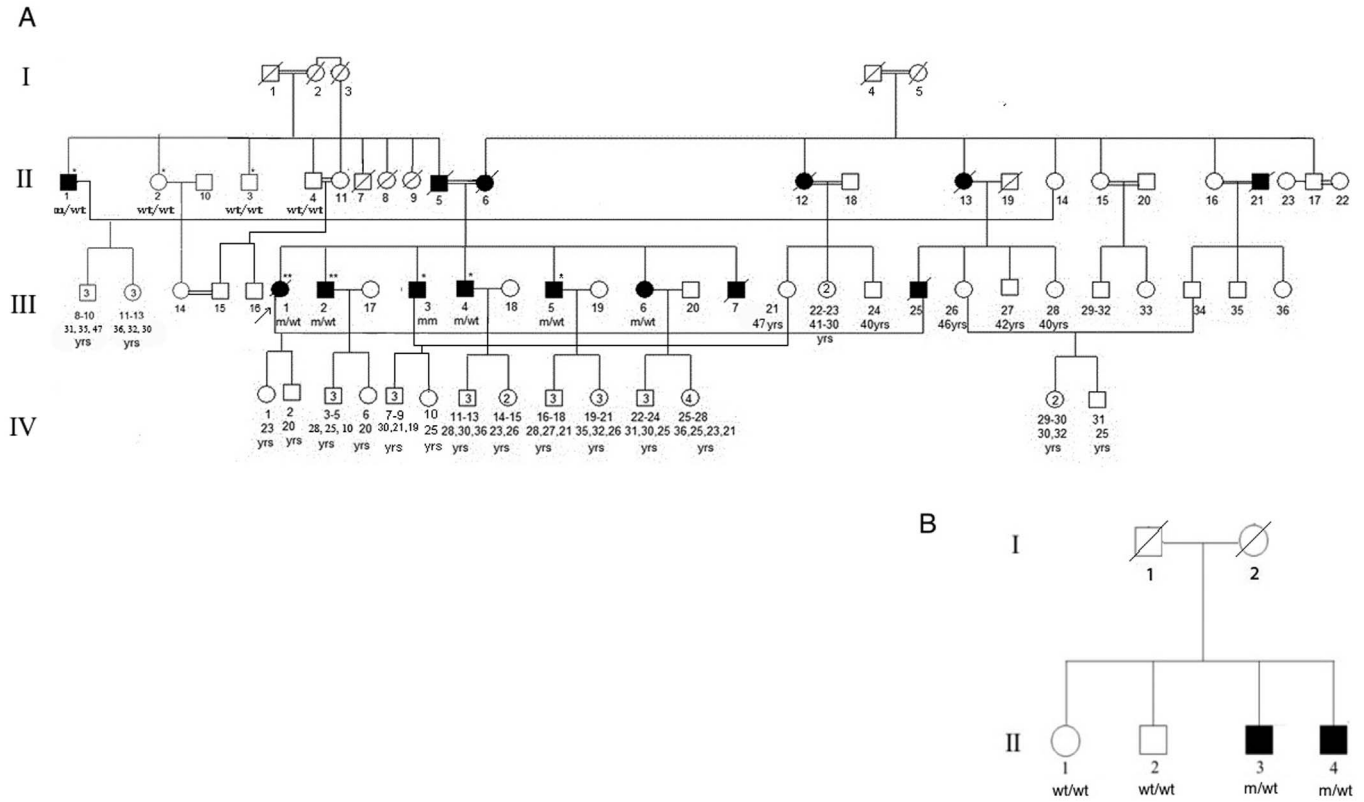

Figure 1 | CAD pedigrees with mutations in *ST6GALNAC5*. (A), CAD-105 pedigree with mutation that causes p.V99M. (B), CAD family with mutation that causes p.*337Qext*20. Known *ST6GALNAC5* genotypes and the present age of some individuals are presented. The five generation I individuals of CAD-105 are all reported to be blood relatives with one another. CAD status and cause of death in individuals of generation I of CAD-105 pedigree and in parents of the other family are unknown. For other individuals: ●,■,CAD affected; ○,□,not CAD affected; *, included in linkage analysis; **, included in linkage and exome sequence analysis; m, mutated *ST6GALNAC5* allele; wt, wild type *ST6GALNAC5* allele. In part A, individuals in generations II and III are not numbered consecutively so that their pedigree ID corresponds with their DNA ID or with other records on the individuals.

residues, which are amino acids that very often have substantial effects on protein structure. With respect to sialyltransferases, cysteine residues within the conserved motifs are thought to participate in the formation of disulfide bonds[21,23]. The DiANNA software (http://clavius.bc.edu/~clotelab/DiANNA/) that predicts disulfide bond formation within proteins based on a neural network approach was used to predict whether the mutated C-terminus would change disulfide bond patterns within ST6GALNAC5[24]. It was predicted that whereas residue 96 would bind with residue 259 in the wild type protein (score of 0.99846 on a scale of 0 to 1), residue 96 would likely bind with residue 344 in the mutated protein (score of 0.99954). This changed pattern of disulfide bond formation may affect enzyme activity.

Finally, *ST6GALNAC5* sequence data (kindly provided by Dr. Leslie G. Biesecker) on 150 CAD patients and 800 control individuals from the United States were compared (Table S8). Five and two variations that caused amino acid changes were found, respectively, in the control and patient groups. One of the variations (p.N101D) of the control group was found in seven individuals, suggesting it is a fairly common polymorphism. Among the remaining variations each of which was observed in only one individual, one variation of the control group and one in the patient group were predicted to damage protein function. Considering that rare DNA sequence variations that cause amino acid changes are more likely to be associated with disease than those that do not cause amino acid changes, we compared the numbers of controls and patients with such variations (Table S9). Comparisons were made within the Iranian cohort, the US cohort, and the combined cohorts of both populations. Whereas the difference did not reach statistical significance in either population alone, it was significant ($P = 0.003$) when combined controls of both populations were compared to combined patients of both populations. Analysis revealed that the Odds Ratio (OR) was 5.93 (95% CI: 1.77–19.85). This suggests that, at the 95% confidence level, individuals in the case group are at least 1.77 times more likely to carry a rare *ST6GALNAC5* sequence variation that will cause an amino acid change than individuals in the control group. A comparable analysis on rare DNA sequence variations that cause amino acid changes predicted to be damaging revealed that the difference in numbers among controls and patients of the combined populations nearly reaches statistical significance ($P = 0.054$) (Table S9). It is notable that *ST6GALNAC5* is positioned within loci previously linked to CAD-associated parameters[18–20]. Furthermore, as discussed below, information available in the literature on sialyltransferases and sialic acid levels in coronary disease patients is consistent with the proposal that defects in sialyltransferase 7e may affect CAD status.

Effect of p.Val99Met and p.*337Qext*20 mutations on enzymatic activity of sialyltransferase 7e. Sialyltransferase activity of wild type and mutated sialyltransferase 7e were compared in protein extracts derived from COS-7 cells transfected with pcDNA3.3- *ST6GALNAC5*, pcDNA3.3- *ST6GALNAC5*- p.Val99Met, and pcDNA3.3- *ST6GALNAC5*-p.*337Qext*20. Expression of exogenous wild type and mutated *ST6GALNAC5* in the COS-7 cells 24 hours post transfection was confirmed by RT-PCR and Western blotting (Fig. S7). The spectrophotometric enzyme assay that was used is based on measurement of inorganic phosphate released from the nucleotide monophosphate that remains after sialic acid is transferred from CMP-sialic acid to a donor molecule[25]. Mixed model statistical analysis showed a significant difference in enzyme activity between

extracts of cells transfected with wild type gene and mutated gene through the range of protein concentrations tested ($P < 0.001$ for mutation p.Val99Met and $P < 0.05$ for mutation p.*337Qext*20; Fig. 2). The variations caused approximately a two-fold increase in enzyme activity at the protein concentrations tested. The concentrations of ST6GALNAC5 protein in the various extracts as measured by an ELISA assay were very similar (average 17.9 $\pm$ 0.9 pg ST6GALNAC5/ng total protein), meaning that the standard deviation of ST6GALNAC5 concentration is relatively small compared to its mean. The coefficient of variation (CV) is only 5%. This suggests that the difference in enzyme activity was not due to differences in the amount of ST6GALNAC5 protein (Table S10).

Discussion

We have shown by genetic analysis that a heterozygous mutation in *ST6GALNAC5* that causes p.Val99Met is the likely cause of CAD in affected individuals of the CAD-105 pedigree. Another heterozygous mutation that causes p.*337Qext*20 may also be the cause of CAD in the Iranian patients who carry this mutation. Both mutations were shown to cause increased sialyltransferase activity in an *in vitro* enzyme assay, presumably reflecting a similar enhanced activity *in vivo*. Analysis of sequence data on control and CAD patients from the United States led to statistical evidence for potential contribution of *ST6GALNAC5* to CAD. The numbers of individuals with rare DNA sequence variations that cause amino acid changes were not significantly different in the controls and patients of this study group. However, when the Iranian and US cohorts were combined (900 controls, 310 patients), the difference reached statistical significance ($P = 0.003$), and the OR was 5.93 (95% CI: 1.77–19.85). Yet larger numbers of controls and patients need to be sequenced in order to attain a more accurate assessment on the contribution of *ST6GALNAC5* to CAD at population levels. We are aware that the genetic analysis and *in vitro* studies do not definitively reveal a causal relation between the gene and CAD, but the results in combination with previous literature are highly suggestive and justify performance of large scale screenings and creation of animal models.

The role of sialyltransferase 7e with respect to CAD at the molecular level was not shown here. The LDL level of CAD-105 patients for whom clinical data is available was higher than that of three elderly unaffected pedigree members. However, clinical and genetic data gathered on younger members of the pedigree suggest that a tight association between *ST6GALNAC5* genotype and LDL levels does not exist (Tables S1-A and S1-C). While hypercholesterolemia is a prevalent phenotype in the pedigree, the LDL concentrations (83, 52, and 85 mg/dl) of three individuals who have mutations (III-9, IV-8, and IV-10, respectively), including one who is homozygous for the mutation (III-9), are within the normal range. Additionally, one individual (III-21) with a very high native cholesterol level (262 mg/dl) is homozygous for the wild type *ST6GALNAC5* allele. The average LDL concentration of individuals homozygous for the mutated allele (164.3 mg/dl) is very similar to that of heterozygous carriers (167.6 mg/dl). It therefore appears that the *ST6GALNAC5* mutation is not the cause of high LDL levels in the pedigree. Finally, while LDL levels in all three individuals older than 70 years are higher than the normal range, none of these are affected with CAD. The sum of clinical and genetic data suggest that hypercholesterolemia is not the major cause of CAD in pedigree CAD-105, although in some individuals it may be a co-factor that amplifies the effects of the *ST6GALNAC5* mutation by increasing predisposition for disease. The available data also suggest that sialyltransferase 7e influences

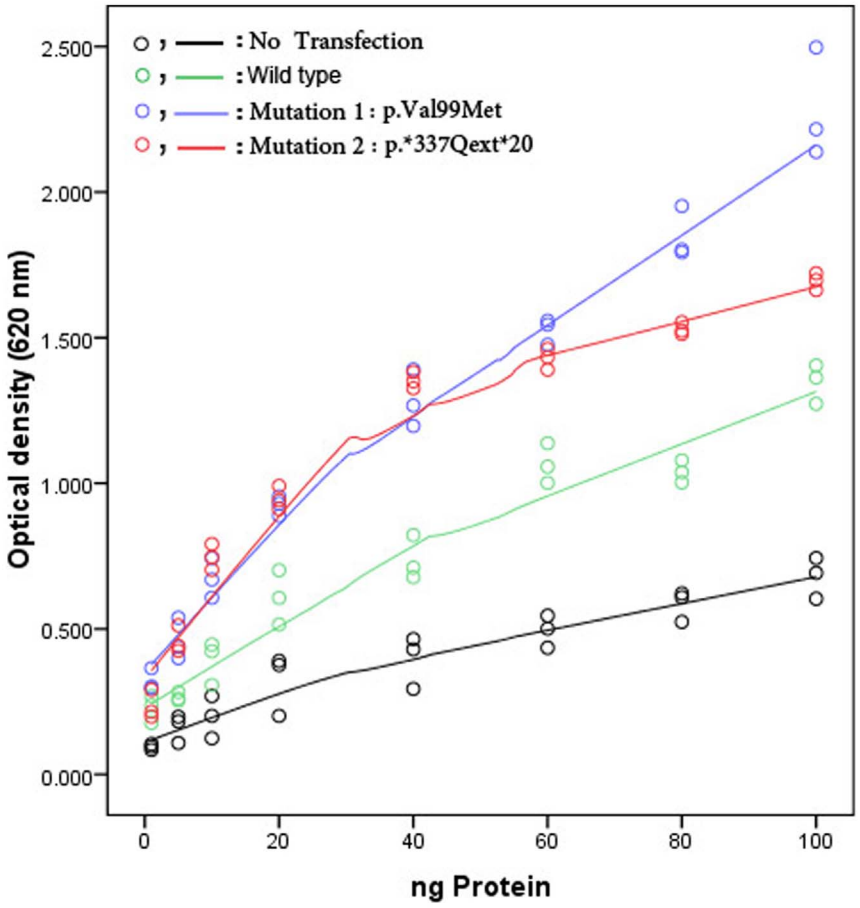

Figure 2 | Comparison of enzymatic activity of wild type and mutated sialyltransferase 7e. Optical density has a linear correlation with sialyltransferase activity. Circles show results of three repeated experiments. Continuous lines show fit of data by Loess method.

factors other than those commonly associated with CAD (Table S1-C).

Sialyltransferases comprise evolutionarily related glycotransferases that catalyze the transfer of acetylated derivatives of neuraminic acid, also known as sialic acids to termini of carbohydrate chains in glycoproteins and glycolipids[21,26]. They are transmembrane proteins preferentially positioned within the Golgi network such that their large catalytic COOH-terminal domain is oriented within the lumen of the organelle[21,26]. Soluble forms of the enzyme are presumed to be proteolytic products of the membrane bound forms[26]. Glycotransferases including sialyltransferases may also be expressed on cell surfaces[27,28]. With respect to the vascular wall, sialyltransferase activity has been observed in membrane preparations containing the Golgi apparatus isolated from human aortic intima[29]. Expression of *ST6GALNAC5* in murine (accession numbers in EMBI-EBI (http://www.ebi.ac.uk/expressionprofiler/): E-GEOD-775, E-GEOD-1479, and E-GEOD-24940) and human (E-GEOD-2240, E-GEOD-40231) heart cells, including human coronary artery cells (E-GEOD-3526), has been shown in several microarray experiments[17]. Vertebrate sialyltransferases are grouped into four families and approximately twenty sub-families, and sialyltransferase 7e encoded by *ST6GALNAC5* is a member of family ST6GALNAC[21,30]. The biological functions of the various sialyltransferases are presently not well elucidated. Their functions are expected to be multiple and important, because terminal glycosylation patterns affect multiple phenomena including cell recognition, proliferation, adhesion, and differentiation[31,32]. Elevated sialyltransferase activity has been associated with atherosclerosis in various studies[33–35]. Recently, gene expression analyses showed that α2,6-sialyltransferase 5 (alternate name for sialyltransferase 7e) is a mediator of breast cancer metastasis to the brain and causes increased adhesion to brain endothelial cells and migration across the blood-brain barrier[36]. As compared to sialyltransferases, the amount of literature available on sialic acid levels with relation to coronary diseases is larger[37].

To the best of our knowledge, the earliest reports on a link between sialic acid and coronary disease, specifically myocardial infarction, were published approximately three decades ago[38,39]. Later, positive associations between serum sialic acid concentrations and mortality from cardiovascular diseases[40], asymptomatic carotid atherosclerosis[41], and coronary heart disease[42,43] were reported. The associations with sialic acid levels in the large population based or case-cohort studies were shown to be independent of potential confounding factors, albeit a small correlation with some of the factors including triglyceride levels was assessed[42,43].

Several possibilities can be considered for the dominant effect of mutations in sialytransferase 7e. Shaping a molecular phenotype might be an indirect mechanism caused by out-competing other enzymes acting on the same substrate[44]. Furthermore, relative sialic acid levels in substrates of different sialyltransferase enzymes may influence manifestation of a disease phenotype. *ST6GAL1* which encodes another sialyltransferase, beta-galactoside alpha-2,6 sialyltransferase, was recently reported to show significant association with CAD in a large association study including approximately 2000 patients of European or Eastern Asian descent[45]. Although sialyltransferase and sialic acid levels are not among the commonly recognized risk factors for CAD, there is substantial literature that suggests a relation between these molecules and coronary disease. We feel that the significance of our findings lies in providing strong evidence for the existence of this relation.

That mutations in *ST6GALNAC5* can contribute to CAD is unlikely to be unrelated to the reported correlation between sialic acid levels and cardiovascular diseases. The reason for this correlation is not known. A proposal has been made based on reported atherogenic consequences of decreased sialylation of various biological components[46–51]. It was suggested that increased serum sialic acid in patients may function as substrate for resialylation of the sialic deficit vascular membrane, circulating cells, and macromolecules in the patients[46]. Alternatively, the increased sialic acid levels may reflect a correlation with coagulation factors such as plasma fibrinogen which is a sialic acid carrier in plasma and which has been linked to cardiovascular morbidity[41,52]. There is a well-established association between serum sialic acid levels and the acute phase response, explained by the fact that many acute phase proteins are glycoproteins that are sialylated[53]. As the acute phase response is an inflammatory response, increased basal levels of sialic acid in prospective coronary heart disease patients may reflect part of an inflammatory response that occurs during the atherosclerosis process. The process of inflammation is now considered to have central importance in atherosclerotic disease, and a sustained acute phase response may be associated with cardiovascular disease[54,55]. Interactions between sialic-acid-containing ligands on leukocytes and selectins on the endothelium of coronary vessels may facilitate passage of these cells through the endothelial layer and be involved in mediation of inflammation[56]. This would be reminiscent of the effect of sialyltransferase 7e in enhancing adhesion of breast cancer cells to brain endothelial cells[36]. It is expected that proper control of sialyltransferases, and possibly also of sialidases, may be important for maintenance of appropriate sialylation levels of molecules involved in coronary functions. Our findings have clinical implications because measurements of sialic acid or sialyltransferase enzyme levels could be considered for identification of individuals at risk for developing CAD. Furthermore, it is quite reasonable to consider that an increased *ST6GALNAC5* activity can be targeted by drugs that act to specifically inhibit the enzyme, and that development of such inhibitors might lead the way to novel CAD therapies and prevention protocols.

Methods

This research was performed in accordance with the declaration of Helsinki and with approval of the ethics board of the University of Tehran. Participants, including CAD-105 pedigree members, control individuals and additional CAD patients, consented to participate after being informed of the nature of the research.

Subjects. A highly inbred Iranian pedigree (CAD-105) with multiply affected young onset CAD members was identified (Fig. 1A, Table 1, Table S1-A). Family members reported that there was no other disease in their pedigree. Recruitment of pedigree members and information gathering occurred in multiple stages. Initially, almost all members of the left branch of the pedigree through generation III (including I-1, I-2, II-1 to II-3, II-5 to II-7, III-1 to III-5, and III-7) were identified. Individuals on whom linkage analysis and exome sequencing were performed were all among those identified at this stage. Later, more affected and unaffected pedigree members were identified, and genetic analysis was performed on all of these who were available. In all, the pedigree included fourteen affected members, seven of whom were deceased. The proband died during the course of the study, making a total of eight dead affected individuals (Fig. 1A).

Among pedigree members initially identified, five living affected siblings in generation III (III-1 to III-5) and their affected uncle (II-1) had undergone coronary artery bypass graft surgeries and two of these had suffered myocardial infarction (MI). Average age at diagnosis of these individuals was 46 years; occurrence of MI in the female proband of the pedigree (III-1) at the age of 38 years was particularly notable. A sibling aged 54 years (III-6) who is also female presents with angina pectoris. The mother of these two women (II-6) had survived a MI that occurred when she was 68 years old; she died eight years later due to unspecified causes. The remaining six affected individuals (II-5, II-12, II-13, II-21, III-7, and III-25) had no knowledge of their condition prior to experiencing MI and these all died shortly after the attack. The pedigree includes three available unaffected individuals (II-2 to II-4) who are over 72 years old.

There are also additional members who have not been diagnosed with CAD, but many of these, particularly those of generation IV, are relatively young and some may become affected at later ages. The ages of the 17 males and 14 females of generation IV range, respectively, from 10–30 years (average 25) and 20–36 years (average 26.9).

CAD risk factors hypertension, diabetes mellitus and smoking were absent in affected as well as unaffected individuals. Being aware that CAD may later occur in presently unaffected young individuals (average age = 30.2 yrs.; range = 19–47 yrs.), phenotypic features were initially compared between affected individuals and only the unaffected individuals who were over 70 years old (Table 1, Table S1-A). Average systolic blood pressure of affected individuals was lower than that of unaffected individuals, possibly due to significantly older age of the unaffecteds; the CAD individuals are not under medication for blood pressure. Although all affected individuals use medication (Atrovastatin) for control of cholesterol levels, their average LDL and triglycerides levels are higher than those of the unaffected

Table 1 | Comparison of averages of phenotypic features of CAD affected/*ST6GALNAC5* mutation carriers and CAD unaffected/*ST6GALNAC5* mutation non-carriers in CAD-105 pedigree*

Phenotypic feature	CAD affected *ST6GALNAC5*: Mut/Wt (±SD)**	CAD unaffected *ST6GALNAC5*: Wt/Wt (±SD)*
Age at diagnosis	46.6 (±5.7)	-
Present age	53.4 (±8.5)	75.0 (±3.0)
LDL (mg/dl)$^{\delta}$	184.3 (±6.1)	154.0 (±9.0)
Triglycerides (mg/dl)$^{\delta}$	137.3 (±38.1)	86.3 (±5.1)
HDL (mg/dl)$^{\delta}$	42.4 (±2.4)	41.0 (±1.0)
Fasting blood glucose (mg/dl)$^{\delta}$	87.1 (±14.4)	94.7 (±17.9)
Systolic BP (mm Hg)$^{\beta}$	117.1 (±4.9)	143.3 (±15.3)
Diastolic BP (mm Hg)$^{\beta}$	85.7 (±5.3)	90.0 (±0.0)
BMI (kg/m^2)	24.8 (±1.1)	24.7 (±1.5)

*Data on younger individuals recruited late in the study not used in calculations; **(±standard deviation); δ Measured after 12 hour fast; β average of four measurements taken at five minute intervals in the lying position using a mercury sphygmomanometer; Mut, mutated allele; Wt, wild type allele; LDL, low-density lipoprotein cholesterol; HDL, high-density lipoprotein cholesterol; BP, blood pressure; BMI, body mass index.

individuals. The patients all started use of Atrovastatin after surgery, initially at 10 mg/day; dosage has increase to 20 mg/day in some patients. HDL levels in affecteds and unaffecteds were comparable. None of the pedigree members were obese and 26 was the highest body mass index among these individuals. The five patients who underwent bypass surgery had no knowledge of their atherosclerosis status prior to surgery, and they have no records of previous laboratory tests. However, results of laboratory tests on available children (III-9–III-11, III-21, IV-3, IV-4, and IV-6–IV-10) of four affected individuals (II-1, II-12, III-2, and III-3) who are young and who are not under medication suggest that native LDL levels in some members of the pedigree are indeed very high (Table S1-A). The average LDL level of the eleven children was 160.6 (±74.0) mg/dl. The LDL levels of two sons (IV-7 and IV-9) of a CAD affected father (III-3) who are aged 30 years and 19 years were, respectively, 279 mg/dl and 175 mg/dl. Similarly, the LDL level of the 20 year old daughter (IV-6) of CAD affected father III-2 was 200 mg/dl. The levels in her 25 (IV-4) and 28 (IV-3) years old brothers were, respectively, 127 and 122 mg/dl, which are at the high end of the normal range. The LDL levels of two of the offsprings (III-10 and III-11) of affected individual II-1 were 173 and 209 mg/dl. One child (III-9) of II-1 and two children (IV-8 and IV-10) of III-3 had LDL levels well within the normal range (83, 52, and 85 mg/dl, respectively).

In addition to CAD-105 pedigree members, 800 Iranian control individuals over the age of 50 years who self-reported to be healthy and not affected with any disease and 160 confirmed unrelated Iranian CAD patients were recruited from the Department of Interventional Cardiology of Imam Khomeiny Hospital associated with Tehran University of Medical Sciences. *ST6GALNAC5* sequences derived from whole genome sequencing performed in the United States on 150 CAD affected individuals and 800 individuals diagnosed not to be affected with CAD were kindly provided by Dr. Leslie G. Biesecker (Genetic Diseases Research Branch, National Human Genome Research Institute, USA).

Genome-wide linkage analysis. Genome-wide SNP genotyping was carried out on DNA samples of eight individuals of the CAD-105 pedigree using HumanCytoSNP-12v1-0_D BeadChips and the iScan reader (GEO accession no.: GSE42137). The individuals included six CAD affected and two CAD unaffected individuals (Fig. 1-A). MERLIN was used for linkage analysis[57]. Parametric and nonparametric analyses were performed.

Genome wide exome sequencing and identification of potential CAD causing variations. Exome sequencing was performed on the DNAs of the proband (III-1) and another affected individual (III-2). Genomic DNA was isolated from blood samples by standard methods. DNA libraries were enriched using the TruSeq® Exome Enrichment kit (Illumina, San Diego, CA, USA) and subsequently sequenced on an Illumina HiSeq® 2000 system. The TruSeq Exome assay targets 62 Mb of protein coding and regulatory untranslated regions of the genome. Base calling was performed by the Illumina pipeline with default parameters. Development protocols and features of the TruSeq® Exome Enrichment kit are described for the first time in the Online Data Supplements. Briefly, for evaluation of this kit, a study was designed to test the efficiency of Exome library enrichment using oligonucleotide probes varying in homology to their cognate targets. Capture probes with varying degrees of artificial deletions, insertions, and consecutive and staggered substitutions as compared to the hg19 reference genome were designed and prepared.

Exome sequence reads were mapped to the human reference genome UCSC NCBI37/hg19 using ELANDv2 software (Illumina). Variant detection was performed with CASAVA software (version 1.8.1; Illumina), and candidate variants were filtered to have a CASAVA quality threshold of 10. CASAVA filtered out duplicate reads and reads without matched pairs. In addition to CASAVA, variants were analyzed using Enlis Genomics (http://www.enlis.com) and NextBio (http://www.nextbio.com/b/nextbio.nb) analysis softwares, again with reference to human genome reference sequence NCBI37/hg19. Variants were systematically filtered to identify novel copy number (CNVs) and sequence variations that were absent in the NCBI dbSNP v130 and 1000 genomes databases and in available control exome sequences, that resulted in amino acid changes or affected splicing, and that were present in both patients. Available control exomes included the 60 whole-exome sequence data available within the Enlis Genomics data set and 15 other exome sequences sequenced along with the CAD-105 patients, but derived from healthy Iranians or Iranians affected with unrelated disorders. The identified variants were submitted to Genome Trax™ (www.biobase-international.com) and OMIM (www.ncbi.nml.nih./gov/omim) for identification of genes with potential functional relevance to CAD and CAD risk factors. Variations in the genes thus identified and variations within the linked loci were analyzed with the SIFT and PolyPhen-2 *in silico* bioinformatics tools to predict those that would be damaging to protein function. The validity of the selected variations in patients III-1 and III-2 was confirmed by direct Sanger sequencing. Finally, segregation with disease status was assessed by sequencing of PCR amplicons that included the variations in all available pedigree members. The amplicon that contained the potential disease causing mutation in *ST6GALNAC5* was screened by sequencing of 800 ethnically matched control individuals. All exons of *ST6GALNAC5* were sequenced in 100 Iranian control individuals and in 160 confirmed Iranian CAD patients. A potential disease causing variation (p.*337Qext*20) found among these patients was also screened by sequencing in 800 controls.

Creation of *ST6GALNAC5*–containing vectors and measurement of sialyltransferase enzyme activity. *ST6GALNAC5* cDNA was PCR amplified from a human heart Multiple_Tissue_cDNA_Panel (Clontech, Heidelberg, Germany); the reverse primer was designed to encode the FLAG-tag. The amplified fragment was cloned into pcDNA3.3 (Invitrogen, Karlsruhe, Germany) using the Topo TA Cloning system (Invitrogen), and pcDNA3.3- *ST6GALNAC5* was created. PcDNA3.3 allows constitutive expression of recombinant genes under the cytomegalovirus promoter. The vector was amplified in TOP10 *E. coli* (Invitrogen) and site directed mutagenesis on the recovered plasmid was performed using the QuickChange site-directed mutagenesis kit (Agilent Technology, Karlsruhe, Germany) to create plasmid pcDNA3.3- *ST6GALNAC5*-p.Val99Met carrying the c.G295A mutation (NM_030965.1). Overlap extension PCR was performed to create plasmid pcDNA3.3- *ST6GALNAC5*- p.*337Qext*20 carrying the c.T1009C mutation. These vectors were also amplified in and recovered from TOP10 *E. coli* cells. *ST6GALNAC5* sequences in the three constructs were verified by Sanger sequencing. The vectors were transfected into African green monkey kidney derived COS-7 cells (ATCC, Rockville, MD, USA) and *ST6GALNAC5* expression in transfected cells was shown by RT-PCR and by Western blotting. In the Western blot experiments, monoclonal mouse M2-anti-FLAG (Sigma-Aldrich, Munich, Germany), anti-human sialyltransferase 7e rabbit (abcam, Cambridge, MA, USA), or anti-human lamin B goat (Santa Cruz Biotechnology, Santa Cruz, CA, USA) primary antibody and appropriate secondary anti-IgG antibody coupled to horseradish peroxidase were used. Detection was performed using the enhanced chemiluminescence (ECL) Western blotting detection system (Invitrogen). Sialyltransferase enzyme activity in protein extracts of COS-7 cells transfected with pcDNA3.3- *ST6GALNAC5*, pcDNA3.3-*ST6GALNAC5*- p.Val99Met, and pcDNA3.3- *ST6GALNAC5*- p.*337Qext*20 was measured using the Sialyltransferase Activity Kit (R&D Systems, Wiesbaden-Nordenstadt, Germany). The amount of ST6GALNAC5 protein in the extracts was measured using an ELISA kit (antibodies-online Inc., Atlanta, GA, USA) designed for detection of human ST6GALNAC5. Methods in detail are presented in the Online Data Supplements.

1. Wang, Q. Molecular genetics of coronary artery disease. *Curr. Opin. Cardiol.* **20**,182–188 (2005).
2. Lusis, A. J., Mar, R. & Pajukanta, P. Genetics of atherosclerosis. *Annu. Rev. Genomics. Hum. Genet.* **5**,189–218 (2004).

3. Marenberg, M. E., Risch, N., Berkman, L. F., Floderus, B. & de Faire, U. Genetic susceptibility to death from coronary heart disease in a study of twins. *New. Engl. J. Med.* **330**, 1041–1046 (1994).
4. Nora, J. J., Lortsher, R. H., Spangler, R. D., Nora, A. H. & Kimberling, W. J. Genetic-epidemiologic study of early-onset ischemic heart disease. *Circulation.* **61**, 503–508 (1980).
5. Schildkraut, J. M., Myers, R. H., Cupples, L. A., Kiely, D. K. & Kannel, W. B. Coronary risk associated with age and sex of parental heart disease in the Framingham Study. *Am. J. Cardiol.* **64**, 555–559 (1989).
6. Hegele, R. A. Plasma lipoproteins: genetic influences and clinical implications. *Nat. Rev. Genet.* **10**, 109–121 (2009).
7. Ghazalpour, A., Wang, X., Lusis, A. J. & Mehrabian, M. Complex inheritance of the 5-lipoxygenase locus influencing atherosclerosis in mice. *Genetics.* **173**, 943–951 (2006).
8. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
9. Holdt, L. M. & Teupser, D. Recent studies of the human chromosome 9p21 locus, which is associated with atherosclerosis in human populations. *Arterioscler. Thromb. Vasc. Biol.* **32**, 196–206 (2012).
10. Wang, L., Fan, C., Topol, S. E., Topol, E. J. & Wang, Q. Mutation of *MEF2A* in an inherited disorder with features of coronary artery disease. *Science.* **302**, 1578–1581 (2003).
11. Mani, A. *et al. LRP6* mutation in a family with early coronary disease and metabolic risk factors. *Science.* **315**, 1278–1282 (2007).
12. Fujino, T. *et al.* Low-density lipoprotein receptor-related protein 5 (*LRP5*) is essential for normal cholesterol metabolism and glucose-induced insulin secretion. Proc. *Natl Acad Sci USA.* **100**, 229–234 (2003).
13. Weng, L. *et al.* Lack of MEF2A mutations in coronary artery disease. *J. Clin. Invest.* **115**, 1016–1020 (2005).
14. Guella, I. *et al.* Association and functional analyses of MEF2A as a susceptibility gene for premature myocardial infarction and coronary artery disease. *Circ Cardiovasc Genet* **2**, 165–172 (2009).
15. Liu, Y. *et al.* Variants in Exon 11 of MEF2A Gene and Coronary Artery Disease: Evidence from a Case-Control Study, Systematic Review, and Meta-Analysis. *PLoS ONE*, **7**, 1 (2012).
16. Altshuler, D. & Hirschhorn, J. N. MEF2A sequence variants and coronary artery disease: a change of heart? *J. Clin. Invest.* **115**, 831–833 (2005).
17. Barth, A. S. *et al.* Functional profiling of human atrial and ventricular gene expression. *Eur. J. Phy.* **450**, 201–208 (2005).
18. Gudmundsson, G. *et al.* Localization of a Gene for Peripheral Arterial Occlusive Disease to Chromosome 1p31. *Am. J. Hum. Genet.* **70**, 586–592 (2002).
19. Allayee, H. *et al.* Locus for elevated apolipoprotein B levels on chromosome 1p31 in families with familial combined hyperlipidemia. *Circ. Res.* **90**, 926–931 (2002).
20. Seidelmann, S. B., Li, L., Shen, G. Q., Topol, E. J. & Wang, Q. Identification of a novel locus for triglyceride on chromosome 1p31–32 in families with premature CAD and MI. *J. Lipid. Res.* **49**, 1034–1038 (2008).
21. Harduin-Lepers, A. Comprehensive Analysis of Sialyltransferases in Vertebrate Genomes. *Glycobiology. Insights.* **2**, 29–61 (2010).
22. Audry, M. *et al.* Current trends in the structure-activity relationships of sialyltransferases. *Glycobiology* **21**, 716–726 (2011).
23. Data, A. K., Chammas, R. & Paulson, J. C. Conserved cysteines in the sialyltransferase sialylmotifs form an essential disulfide bond. *J. Biol. Chem.* **276**, 15200–15207 (2001).
24. Ferre, F. & Clote, P. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res.* Jul 1; 34(Web Server issue): W182–5 (2006).
25. Wu, Z. L., Ethen, C. M., Prather, B., Machacek, M. & Jiang, W. Universal phosphatase-coupled glycosyltransferase assay. *Glycobiology.* **21**, 727–733 (2011).
26. Broquet, P., Baubichon-Cortay, H., George, P. & Louisot, P. Glycoprotein sialyltransferases in eucaryotic cells. *Int. J. Biochem.* **23**, 385–389 (1991).
27. Appeddu, P. A. & Shur, B. D. Molecular analysis of cell surface β-1, 4-galactosyltransferase function during cell migration. *Proc. Natl. Acad. Sci. USA.* **91**, 2095–2099 (1994).
28. Kaufmann, M. *et al.* Identification of an α2, 6-sialyltransferase induced early after lymphocyte activation. *Int. Immunol.* **11**, 731–738 (1999).
29. Gracheva, E. V. *et al.* Sialyltransferase activity of human plasma and aortic intima is enhanced in Atherosclerosis. *Biochem. Biophys. Acta.* **1586**, 123–128 (2002).
30. Okajima, T. *et al.* Molecular cloning of brain-specific GD1alpha synthase (*ST6GalNAc V*) containing CAG/Glutamine repeats. *J. Biol. Chem.* **274**, 30557–30562 (1999).
31. Schauer, R. Chemistry, metabolism, and biological functions of sialic acids. *Adv. Carbohydr. Chem. Biochem.* **40**, 131–234 (1982).
32. Varki, A. Sialic acids as ligands in recognition phenomena. *FASEB. J.* **11**, 248–255 (1997).
33. Scialla, S. J., Speckart, S. F., Haut, M. J. & Kimball, D. B. Alterations in platelet surface sialyltransferase activity and platelet aggregation in a group of cancer patients with a high incidence of thrombosis. *Cancer. Res.* **39**, 2031–2035 (1979).
34. Wu, K. K. & Ku, C. S. L. Stimulation of platelet surface sialyltransferase activity by platelet aggregating agents. *Thromb. Res.* **13**, 183–192 (1978).
35. Hagg, S. *et al.* Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. *PloS. Genet.* **5**, e1000754; *DOI: 10.1371/pgen1000754* (2009).
36. Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature.* **459**, 1005–1009 (2009).
37. Gopaul, K. P. & Crook, M. A. Sialic acid: A novel marker of cardiovascular disease? *Clin. Biochem.* **39**, 667–681 (2006).
38. Succari, M., Foglietti, M. J. & Percheron, F. Perchlorosoluble glycoproteins and myocardial infarct: modifications of the carbohydrate moiety. *Pathol Biol.* **30**, 151–154 (1982).
39. Hrncir, Z., Pidrman, V., Tichy, M. & Hamet, A. Serum sialic acid in acute myocardial infarction in a dynamic follow up. *Vnitr. Lek.* **21**, 436–439 (1975).
40. Lindberg, G., Råstam, L., Gullberg, B. & Eklund, G. A. Serum sialic acid concentration predicts both coronary heart disease and stroke mortality: multivariate analysis including 54385 men and women during 20.5 years follow-up. *Int. J. Epidemiol.* **21**, 253–257 (1992).
41. Råstam, L. *et al.* Association between serum sialic acid concentration and carotid atherosclerosis measured by B-mode ultrasound. *Int. J. Epidemiol.* **25**, 953–958 (1996).
42. Knuiman, M. W., Watts, G. F. & Divitini, M. L. Is sialic acid an independent risk factor for cardiovascular disease? A 17-year follow-up study in Busselton, Western Australia. *Ann. Epidemiol.* **14**, 627–632 (2004).
43. Watts, G. F., Crook, M. A., Haq, S. & Mandalia, S. Serum sialic acid as an indicator of change in coronary artery disease. *Metabolis.* **44**, 147–148 (1995).
44. Tsuchida, A. *et al.* Synthesis of disialyl Lewis a (Le^a) structure in colon cancer cell lines by a sialyltransferase, ST6GalNAc VI, responsible for the synthesis of α-series gangliosides. *J. Biol. Chem.* **278**, 22787–22794 (2003).
45. Saade, S. *et al.* Large scale association analysis identifies three susceptibility loci for coronary artery disease. *PLoS ONE.* **6**, e29427. doi:10.1371/journal.pone.0029427 (2011).
46. Lindberg, G. Resialylation of sialic acid deficit vascular endothelium, circulating cells and macromolecules may counteract the development of atherosclerosis: A hypothesis. *Atherosclerosis.* **192**, 243–245 (2007).
47. Görög, P. & Born, G. V. Uneven distribution of sialic acids on the luminal surface of aortic endothelium. *Br. J. Exp. Pathol.* **64**, 418–424 (1983).
48. Görög, P. & Born, G. V. R. Increase uptake of circulating LDL and fibrinogen by arterial walls after removal of sialic acids from their endothelial surface. *Br. J. Exp. Path.* **63**, 447–451 (1982).
49. Tertov, V. V., Sobenin, I. A. & Orekhov, A. N. Characterisation of desialylated low-density lipoprotein which cause intracellular lipid accumulation. *Int. J. Tissue. Reac.* **14**, 155–162 (1992).
50. Ruelland, A. *et al.* LDL sialic acid content in type 2 diabetic patients. *Clinica. Chimica. Acta.* **259**, 191–193 (1997).
51. Mandic, R., Opper, C., Krappe, J. & Wesemann, W. Platelet sialic acid as a potential pathogenic factor in coronary heart disease. *Thromb. Res.* **106**, 137–141 (2002).
52. Kario, K. & Matsuo, T. Relation between sialic acid concentrations and the haemostatic system in the elderly. *BMJ.* **306**, 1650–1651 (1993).
53. Crook, M. A., Tutt, P., Simpson, H. & Pickup, J. C. Serum sialic acid and acute phase proteins in type 1 and type 2 diabetes mellitus. *Clin. Chim. Acta.* **219**, 131–138 (1993).
54. Paoletti, R., Gotto, A. M. & Hajjar, D. P. Atherosclerosis: Evolving vascular biology and clinical implications, Inflammation in atherosclerosis and implications for therapy. *Circulation.* **109**, (Suppl III):III20–6; *DOI: 10.1161/01.CIR.0000131514.71167.2e* (2004).
55. The CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetic.* **45**, 25–33 (2013).
56. Rosen, S. D. Ligands for L-selectin: homing, inflammation, and beyond. *Annu. Rev. Immunol.* **22**, 129–156 (2004).
57. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).

Acknowledgments

We thank the patients and their family members for consenting to participate in this study, and Dr. Mehdi Yaseri for advice in statistical analysis. We acknowledge the Iran National Science Foundation, the Research Division of the University of Tehran, and the Leibniz Institute for Age Research–Fritz Lipmann Institute for funding this research. The TruSeq® Exome Enrichment technology summarized in this paper would not have been possible without the efforts of many dedicated individuals. We would like to thank Kerri York for the initial enrichment technology development experiments. Melissa Won and Weihua Chang developed the variant capture probes. Jean Lozach and Jerry Kakol designed the probes and wrote data analysis scripts, and Tobias Mann performed the variant probe data analysis. We also like to thank people at Illumina for their valuable contributions in reversible terminator sequencing technology development (formerly Solexa), and the biochemistry development group for the Exome enrichment technology development. Illumina, Genome Analyzer, TruSeq, Nextera, and Infinium are registered trademarks or trademarks of Illumina, Inc. All other brand names contained herein are the property of their respective owners.

Author contributions

K.I.R. performed the linkage analysis, the analysis on the exome data of family CAD-105, all the experiments regarding *ST6GALNAC5*, and provided invaluable comments on the text of the manuscript; A.Z.P. introduced patients and provided clinical data; K.H. and M.P. provided materials, read the manuscript, and gave suggestions; S.A. performed data analysis for development of TruSeq® Exome Enrichment technology; P.R. prepared DNA of 160 CAD patients; S.D. introduced CAD patients; J.B.F. supervised microarray experiments for linkage analysis; M.K. provided technical advice on experiments relating to *ST6GALNAC5*; C.T. performed the Exome Enrichment and technology development experiments; F.S. and K.G. conceived the TruSeq® Exome Enrichment technology; F.S. wrote the text that describes development protocols and features of the TruSeq® Exome Enrichment technology; M.R. conceived the collaboration; E.E. conceived the genetic study, supervised the project, and wrote the manuscript except the text that describes development protocols and features of the TruSeq® Exome Enrichment technology. All authors read and approved the final manuscript.

Additional information

Funding: Funding was provided by Iran National Science Foundation, Research Division of the University of Tehran, and Leibniz Institute for Age Research–Fritz Lipmann Institute.

Author information: Genome-wide SNP genotyping data of eight individuals using HumanCytoSNP-12v1-0_D BeadChips Illumina platform deposited at GEO; accession no.: GSE42137.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: J.B.F., C.T., S.A., F.S., K.G. and M.R. declare having competing financial interests. The remaining authors do not have competing financial interests.

How to cite this article: InanlooRahatloo, K. *et al.* Mutation in *ST6GALNAC5* identified in family with coronary artery disease. *Sci. Rep.* **4**, 3595; DOI:10.1038/srep03595 (2014).