Mutat Res Fund Mol Mech Mutagen 803–805 (2017) 26–34

Contents lists available at ScienceDirect

Mutat Res Fund Mol Mech Mutagen

journal homepage: www.elsevier.com/locate/mut

Identification of intronic-splice site mutations in *GATA4* gene in Indian patients with congenital heart disease

Divya Bose[a], Vaigundan D.[a], Mitesh Shetty[a], Krishnappa J.[b], A.V.M. Kutty[a,*]

[a] Division of Genomics, Department of Cell Biology and Molecular Genetics, Sri Devaraj Urs Academy of Higher Education and Research, Tamaka, Kolar, Karnataka, India
[b] Department of Pediatrics, Sri Devaraj Urs Medical College, R. L. Jalappa Hospital and Research Centre, Tamaka, Kolar, Karnataka, India

ARTICLE INFO

Keywords:
Intronic variants
Branch point mutation
Congenital heart disease
Enhancer motif
Silencer motif

ABSTRACT

Congenital Heart Disease (CHD) is the most common birth defect among congenital anomalies that arise before birth. *GATA4* transcription factor plays an important role in foetal heart development. Mutational analysis of *GATA4* gene in CHD patients revealed five known heterozygous mutations (p.T355S, p.S377G, p.V380M, p.P394T and p.D425N) identified in exons 5 and 6 regions and fifteen intronic variants in the non-coding regions (g.76885T > C/Y, g.76937G > S, g.78343G > R, g.83073T > Y, g.83271C > A/M, g.83318G > K, g.83415G > R, g.83502A > C/M, g.84991G > R, g.85294C > Y, g.85342C > T/Y, g.86268A > R, g.87409G > A/R, g.87725T > Y, g.87813A > T/W). *In silico* analysis of these intronic variants identified two potential branch point mutations (g.83271C > A/M, g.86268A > R) and predicted effects of these on intronic splice sites as enhancer and silencer motifs. This study attempts to correlate the pattern of intronic variants of *GATA4* gene which might provide new insights to unravel the possible molecular etiology of CHD.

1. Introduction

Congenital Heart Disease (CHD) is the most common birth defect in humans affecting 1% of all live births in the first year of life and it is one of the major causes of morbidity and mortality in infants [1]. In India, nearly 1,80,000 children are born with CHD with the prevalence ranged from 8 to 10 of every 1000 live births [2]. CHD is a complex multifactorial disorder with genetic and environmental factors playing an important role in disease development [3]. It is categorized mainly into four groups namely septal defects, cyanotic heart disease, obstruction defects and hypoplasia, where genetic factors are implied in the etiology [4]. A group of highly conserved transcription factors such as *GATA4, NKX2-5, MYH6, NOTCH, TBX5, TBX20, ZIC3, TFAP2B* and others are involved in foetal heart development and regulation [5–7]. Among these *GATA4* is well studied and established gene known to cause sporadic and familial non-syndromic CHDs which include atrial septal defect (ASD), ventricular septal defect (VSD), patent ductus arteriosus (PDA), atrioventricular septal defect (AVSD), tetralogy of fallot (TOF) and pulmonary value stenosis (PS) [5–8].

GATA proteins (DNA binding proteins) are involved in cell differentiation, survival and proliferation of tissues that regulate gene expression. They recognize and bind to "*GATA*" consensus sequence of target genes [9]. Six members have been indentified in vertebrates *GATA* family. *GATA1, GATA2,* and *GATA3* are mainly involved in hematopoietic cell expression, whereas *GATA4, GATA5* and *GATA6* are expressed in heart, liver and gonadal tissues [10]. *GATA4* (Gene Id: 2626, OMIM: 600576) is a critical transcription factor expressed in embryo and adult cardiomyocytes and it promotes cardiac morphogenesis, survival and function of the heart [11]. Human *GATA4* gene maps to chromosome 8p23.1- p22 region, a hypermutable protein coding gene encodes 442 amino acids with two transcriptional activation domains (TAD1, 1–74 amino acids; TAD2, 130–177 amino acids), two zinc finger domains (ZF1, 215–240 amino acids; ZF2, 270–294 amino acids) and one nuclear localization signal (NLS, 254–324 amino acids) [12–14]. Mutations associated in *GATA4* gene had been reported in several cardiac diseases such as congenital heart disease, abnormal ventral folding and hypoplasia of ventricular myocardium [15,16]. Till date, more than 120 mutations have been reported in the *GATA4* gene. These mutations in *GATA4* are well-studied and reported from different countries in CHD patients and are implied as reasons for CHD in humans [17].

There are only few studies conducted in Indian population on CHD

Abbreviations: CHD, congenital heart disease; ASD, atrial septal defect; VSD, ventricular septal defect; PDA, patent ductus arteriosus; AVSD, atrioventricular septal defect; TOF, tetralogy of fallot; PS, pulmonary value stenosis; DORV, double outlet right ventricle; TAD, transcriptional activation domain; ZF, zinc finger domain; NLS, nuclear localization signal; EDTA, ethylenediaminetetraacetic acid; PolyPhen-2, polymorphism phenotyping version 2; SIFT, sorting intolerant from tolerant; ESE, exonic splicing enhancer; ESS, exonic splicing silencer; ISE, intronic splicing enhancer; ISS, intronic splicing silencer

* Corresponding author at: Faculty of Allied Health Sciences, Sri Devaraj Urs Academy of Higher Education and Research, Tamaka, Kolar, Karnataka, 563101, India.
E-mail address: kuttyavm@gmail.com (A.V.M. Kutty).

http://dx.doi.org/10.1016/j.mrfmmm.2017.08.001
Received 7 April 2017; Received in revised form 1 July 2017; Accepted 8 August 2017
Available online 18 August 2017
0027-5107/

Table 1
List of *GATA4* primers.

Exon	Forward primer (5′–3′)	Reverse Primer (5′–3′)	Annealing temperatue
Exon 2	GAACTCTCAGTGTCTGGGATTAG	GTGGCTCCAGCTAACTCTAAA	61 °C
Exon 3	TGACGGTGAATGATGGTTAGG	GGCCAGCAAAGTAGTTGAAAG	61 °C
Exon 4	CATCACACAGGTGCTCGATAAG	CCAAAGATGAAAGGACCGAGTA	61 °C
Exon 5	TGTAGCCCTCCGCAGATAA	GTCATGTCCACATCACCATCT	61 °C
Exon 6	TTCTGGGCAACCACAGTATC	AGTCCCATCAGCGTGTAAAG	61 °C
Exon 7	GCTCCTTCACTTCCAACATCTC	ACCCTCTCCCAGGAAATTAAAG	61 °C

Table 2
Cardiac septal defects of the subjects.

CHD condition	Number of subjects
ASD	22
VSD	12
PDA	3
ASD, VSD	4
ASD, PDA	9
ASD, VSD, PDA	4
VSD, DORV	1
ASD, PFO	1
VSD, TOF, PS	1
VSD, TOF	1
PS	1
PFO	1
Total Number of Patients	60

to identify the causes of this disease condition [18–20]. The present study was designed to look for mutations in the exonic and intronic regions of *GATA4* gene of patients with CHD.

2. Materials and methods

2.1. Subjects

Seventeen patients diagnosed with isolated CHDs at R. L. Jalappa Hospital and Narayana Hrudalaya heart centre from Kolar were recruited in this study. Pediatric cardiologists confirmed the CHD by two-dimensional echocardiography with color flow doppler. After getting the Institutional Ethics committee approval and written informed consent from the patients, whole peripheral blood samples were collected for further genetic study.

2.2. DNA extraction

Genomic DNA from all patients was isolated from blood lymphocytes collected in EDTA vacutainers using standard salting out method. Three ml of blood was mixed thoroughly and incubated for erythrocyte lysis. After erythrocyte lysis, the packed white blood cells pellet was kept for Proteinase-k digestion overnight in the presence of 20% SDS. The genomic DNA was precipitated by the addition of 5 M NaCl and the DNA was spun down and washed with alcohol to remove salts and eluted with Tris-EDTA buffer [21]. DNA concentration and purity was assessed by optical density ratios (260/280 nm) using UV–vis Spectrophotometer (Lambda 35, Perkin Elmer, Waltham, Massachusetts, USA) and stored at −20 °C until use.

2.3. Sequence analysis

The referential genomic DNA sequence of *GATA4* was retrieved from Genbank (Accession No. NC_000008) and sequence specific primer pairs were designed to amplify the coding exons and exon-intron boundary regions of *GATA4* gene with the help of Primer Quest tool, IDT DNA software (https://www.idtdna.com/) and listed in Table 1. Polymerase chain reaction (PCR) was carried out with these specific primers using C1000 Touch Thermal cycler (Bio-Rad Laboratories, Hercules, California, USA). Reactions contained 100 ng of genomic DNA, 10× PCR buffer, 10 mM dNTPs, 10 picomole of each primer, 1.5 mM $MgCl_2$, and 1 unit *Taq* DNA polymerase (Bangalore Genei, India) and the conditions followed with an initial denaturation at 95 °C for 5 min followed by 33 cycles of denaturation at 95 °C for 30 s, annealing at 61 °C for 30 s, 72 °C for 30 s and final extension at 72 °C for 10 min. The PCR products were analyzed on 1% agarose gel and purified with GeneJET PCR Purification kit (Thermo Fisher Scientific). Sequence analysis was performed for all the six exons with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) under ABI-3500 Genetic Analyzer (Applied Biosystems, Foster City, California, USA). DNA Sequences were analyzed by ABI Variant Reporter software version 1.1 (Applied Biosystems) using NG_008177 gene sequence as template. Variant sequences that were thus obtained were deposited in Clinvar.

2.4. Mutation prediction

Protein sequences of GATA4 protein from various species were

Table 3
Exonic variants in *GATA4* coding region.

Exon	Nucleotide change	Aminoacid change	dbSNP id	Mutation type	Mutation Prediction		
					Polyphen-2	SIFT	Mutation Taster
5	c.1064C > G	T355S	rs200167770	Missense	Benign	Tolerated	Polymorphism
5	c.1129A > G	S377G	rs3729856	Missense	Benign	Tolerated	Polymorphism
5	c.1138G > A	V380M	rs114868912	Missense	Benign	Tolerated	Polymorphism
6	c.1180C > A	P394T	rs200319078	Missense	Benign	Tolerated	Polymorphism
6	c.1273G > A	D425N	rs56208331	Missense	Probably damaging	Tolerated	Disease causing

Table 4A
Exonic and Intronic variants of *GATA4* gene in CHD patients.

	Gender	Sample Type	76885	83271	83502	85342	86268	87409	87813			
1	F	ASD	g.76885T > C	g.83271C > A								
2	F	ASD	g.76885T > C	g.83271C > A								
3	F	ASD, PFO	g.76885T > C	g.83271C > A								
4	M	ASD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
5	F	ASD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
6	M	ASD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
7	F	ASD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
8	F	ASD	g. 76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
9	F	ASD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
10	F	VSD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
11	M	VSD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
12	M	ASD, PDA	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
13	F	ASD, PDA	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
14	F	ASD, PDA	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
15	F	PFO	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W			
16	F	ASD,VSD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W	g.76937G > S		
17	M	ASD	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > W			
18	F	ASD	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > W			
19	M	ASD, PDA	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > W			
20	M	ASD,VSD	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > T			
21	M	VSD,TOF	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > T			
22	F	ASD	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > T			
23	NIL	ASD	g.76885T > C	g.83271C > A	g.83502A > C				g.87813A > T			
24	M	ASD,PDA	g.76885T > C	g.83271C > A					g.87813A > W	g.78343G > R		
25	F	ASD, VSD, PDA	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W		g.85294C > Y	
26	M	ASD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W	p.T355S	g.84991G > R	
58	M	VSD	g.76885T > C	g.83271C > A					g.87813A > W		p.P394T	g.84991G > R
59	F	VSD	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W		p.P394T	
27	M	ASD, PDA	g.76885T > C	g.83271C > A	g.83502A > M				g.87813A > W	p.T355S		
28	F	ASD	g.76885T > C	g.83271C > A		g.85342C > Y	g.86268A > R	g.87409G > R				
29	F	VSD	g.76885T > C	g.83271C > A		g.85342C > Y	g.86268A > R	g.87409G > R				
30	F	PDA	g.76885T > C	g.83271C > A		g.85342C > Y	g.86268A > R	g.87409G > R				
31	M	ASD,PDA	g.76885T > C	g.83271C > A		g.85342C > Y	g.86268A > R	g.87409G > R				
32	NIL	ASD,VSD,PDA	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R				
33	F	ASD	g.76885T > C	g.83271C > A		g.85342C > Y	g.86268A > R	g.87409G > A				
34	F	ASD	g.76885T > C	g.83271C > A		g.85342C > Y	g.86268A > R	g.87409G > R			g.85294C > Y	**p.V380M**
35	M	ASD	g.76885T > Y	g.83271C > M		g.85342C > Y	g.86268A > R	g.87409G > R		g.83415G > R	g.85294C > Y	**p.V380M**
36	M	VSD	g.76885T > C	g.83271C > A		g.85342C > T		g.87409G > A				
37	M	ASD	g.76885T > C	g.83271C > A		g.85342C > T		g.87409G > A				
38	M	ASD,PDA	g.76885T > C	g.83271C > A			g.86268A > R	g.87409G > R			p. P394T	
39	NIL	ASD, VSD	g.76885T > C			g.85342C > Y	g.86268A > G	g.87409G > R	g.87813A > W	g.76937G > S	p.S377G	
40	M	ASD,VSD	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W			
41	M	PDA	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W			
42	M	PDA	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W			
43	M	ASD	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W			
44	F	ASD	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W		g.83318G > K	
45	F	VSD	g.76885T > Y	g.83271C > M	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W		g.83415G > R	
46	F	VSD	g.76885T > Y	g.83271C > M	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > T		g.83415G > R	
47	F	VSD	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W			
48	M	PS	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W			
49	M	ASD,PDA	g.76885T > Y	g.83271C > M		g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > T			
50	M	ASD, VSD, PDA	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W		g.87725T > Y	
51	M	VSD, DORV	g.76885T > C	g.83271C > M			g.86268A > R		g.87813A > W	g.78343G > R	g.76937G > S	**g.83073T > Y**
52	F	TOF, VSD, PS	g.76885T > C	g.83271C > M	g.83502A > M		g.86268A > R		g.87813A > T	g.76937G > S	p.S377G	

(continued on next page)

Table 4A *(continued)*

	Gender	Sample Type	76885	83271	83502	85342	86268	87409	87813		
53	F	ASD,VSD,PDA	g.76885T > Y		g.83502A > M		g.86268A > R		g.87813A > W		p. D425N
54	M	VSD	g.76885T > C	g.83271C > M	g.83502A > M		g.86268A > R		g.87813A > T	g.76937G > S	p.S377G
56	F	ASD	g.76885T > C	g.83271C > M	g.83502A > M		g.86268A > R		g.87813A > W	g.76937G > S	p.D425N
55	M	ASD	g.76885T > C	g.83271C > M			g.86268A > G		g.87813A > T	g.76937G > C	p.S377G
57	F	VSD	g.76885T > C	g.83271C > A		g.85342C > Y		g.87409G > R	g.87813A > W		p.P394T
60	F	VSD	g.76885T > C	g.83271C > M	g.83502A > M		g.86268A > R			g.76937G > S	p.S377G

obtained from NCBI protein database and aligned using MUSCLE program (www.ebi.ac.uk/Tools/msa/muscle/). PolyPhen-2 (Polymorphism Phenotyping, http://genetics.bwh.harvard.edu/pph2/), SIFT (Sorting Intolerant From Tolerant, http://sift.jcvi.org/) and MutationTaster (www.mutationtaster.org) programs were utilized to validate the mutation prediction of variants.

2.5. In silico analysis

The intronic mutations, RNA splice sites and RNA branch points of *GATA4* gene were analyzed by using Human Splicing Finder (HSF) [22] software version 3.0 (www.umd.be/HSF3/), an online bioinformatics tool to predict the effects of mutations on splicing signals or splicing motifs either to enhance or suppress splicing such as Exonic Splicing Enhancers (ESEs), Exonic Splicing Silencers (ESSs), Intronic Splicing Enhancers (ISEs) and Intronic Splicing Silencers (ISSs).

3. Results

The patient population with CHD in this study included 22 ASD, 12 VSD, 9 ASD and PDA, 4 ASD and VSD, 4 ASD, VSD and PDA, 3 PDA, one each with PS, PFO, ASD with PFO, VSD with Double Outlet Right Ventricle (DORV), VSD with TOF as well as VSD, TOF with PS (Table 2). 10 samples without CHD were included as controls in the study.

Mutational analysis of CHD cases revealed twenty variants including five in the coding and fifteen in the non-coding region (Tables 3, 4A, and 4B). The data analysis of controls also presented intronic variants however the genotypic pattern were different from the cases under study group (Table 5).

There were five known heterozygous missense mutations (p.T355S, p.S377G, p.V380M, p.P394T and p.D425N) identified in exon 5 and exon 6 regions (Fig. 1A). Multiple sequence alignment of *GATA4* sequence from various species was aligned and only p.D425N showed the evolutionarily conserved throughout the species. Mutation prediction results showed that the p.T355S, p.S377G, p.V380 M and p.P394T variants were predicted to be benign, tolerated and polymorphism by PolyPhen-2, SIFT and MutationTaster. Only p.D425N variant was predicted to be probably damaging, tolerated and disease causing (Table 3).

Apart from these missense variants, there were 15 novel intronic variants (Fig. 1B) found in different CHD phenotypes as listed in Tables 4A and 4B. These intronic variants were analyzed for potential branch point and potential ISEs and ISSs. The results indicated two variants predicted which could be potential branch point sites. One potential branch point site at g.83271C > A/M variant with GCCTAAG motif of 83.82 consensus value in wild type and in addition, this mutation leads to formation of ISS motif 2 (AGAGTGCA) with 63.85 consensus value (Fig. 2A) and another potential branch point site at g.86268A > R variant with GGCTGAA motif of 82.24 consensus value in wild type. These intronic mutations at the potential branch point are predicted to be disrupted (Fig. 2B).

The g.76885T > C/Y intronic variant leads to formation of new ISE motif for SRp40; whereas the g.76937G > S and g.83073T > C/Y variants lead to the disruption of SRp40 and SRp55 ISE motifs respectively. On the other hand, g.78343G > A/R variant leads to the disruption of ISS motif 1 (CAAAACGT) and g.87813A > T/W leads to formation of ISS motif 2.

The g.83318G > K variant leads to the formation of SF2/ASF ISE motif and ISS Motif 2, whereas g.85294C > Y variant leads to the disruption of SRp40 ISE motif and formation of ISS motif 2. The g.84991G > R variant leads to the disruption of ISS motif 1 and formation of ISS motif 2 and the g.87725T > Y variant leads to the

Table 4B
Exonic and Intronic variants of *GATA4* gene in control group.

	Gender	Sample Type	76885	83271	83502	85342	86268	87409	87813	
1	F	Control1	g.76885T > C	g.83271C > A						
2	M	Control2	g.76885T > C	g.83271C > A	g.83502A > C					
3	F	Control3	g.76885T > C	g.83271C > A	g.83502A > M					
4	Nil	Control4	g.76885T > C		g.76937G > S		g.86268A > R			p.S377G
5	F	Control5	g.76885T > C	g.83271C > A		g.85342C > T	g.86268A > G			
6	M	Control6	g.76885T > C	g.83271C > A	g.83502A > M		g.86268A > R	g.87409G > R	g.87813A > T	
7	M	Control7	g.76885T > C	g.83271C > A			g.86268A > R	g.87409G > R		
8	M	Control8	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W	
9	F	Control9	g.76885T > C	g.83271C > A	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > T	g.78042G > R
10	Nil	Control10	g.76885T > Y	g.83271C > M	g.83502A > M	g.85342C > Y	g.86268A > R	g.87409G > R	g.87813A > W	g.83415G > R

Table 5
Genotypic pattern of *GATA4* gene at 85342, 86268, 87409 and 87813 positions in CHD patients (5A) and control group (5B).

A: CHD patients

S.No	Frequency	85342	86268	87409	87813
1.[a]	3	CC	AA	GG	AA
2.	22	CC	AA	GG	AT
3.	4	CC	AA	GG	TT
4.	7	CT	AG	AG	AA
5.	1	CT	AG	AA	AA
6.	2	TT	AA	AA	AA
7.	1	CC	AG	AG	AA
8.[a]	11	CT	AG	AG	AT
9.	1	CT	GG	AG	AT
10.	3	CC	AG	GG	AT
11.	2	CC	AG	GG	TT
12.	1	CC	GG	GG	TT
13.	1	CT	AA	AG	AT
14.	1	CC	AG	GG	AA

B: Control group

S.No	Frequency	85342	86268	87409	87813
1.[a]	3	CC	AA	GG	AA
2.	1	CC	AG	GG	AA
3.	1	TT	GG	GG	AA
4.	1	CC	AG	AG	TT
5.	1	CC	AG	AG	AA
6.[a]	2	CT	AG	AG	AT
7.	1	CT	AG	AG	TT

[a] Genotypic patterns are identical in both CHD patients and control group.

formation of SF2/ASF ISE motif and ISS motif 1.

The g.83415G > R variant leads to formation of SRp40 ISE motif and formation of ISS motif 3 and disruption of SF2/ASF2 ISE motif; g.85342C > T/Y leads to formation of SRp55 ISE motif and formation of ISS motif 2. The g.87409G > A/R leads to formation of SRp40 ISE motif and disruption of ISS motif 1. However, intronic variant g.83502A > C/M revealed no significance to the formation of potential branch point site or change of ISE and ISS (Table 6).

4. Discussion

Intronic variants and its consequence on the splicing mechanism can be better understood by using *in silico* analysis that includes different algorithms to identify various regulatory elements. In the present study, HSF3.0 *in silico* analysis showed that the intronic variant g.86268A > R predicted as potential branch point site. Consensus wild type score for the potential branch point ranges from 0 to 100 and the threshold value is defined to be 67. Scores above 67 is considered to be potential branch point. When a mutation occurs, if the wild type score is above 67 and the score variation (between wild type and mutant) decreases by more than 10%, it is considered that the mutation breaks the branch point. In the intronic variant, g.83271C > A/M apart from the loss of potential branch point, it also lead to the formation of ISS motif 2. Alterations in the branch point position might provoke serious defects in pre-mRNA splicing leading to exon skipping or intron inclusion or activation of cryptic splice sites [23–25].

Pre-mRNA splicing is an essential process in gene expression and majority of human genes undergo alternative splicing, leading to the formation of isoforms [26]. It is regulated by *cis*-elements (ESE, ESS, ISE and ISS) and *trans*-acting factors such as Serine and Arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoprotein (hnRNP) [27]. ESEs/ESSs will promote or inhibit inclusion of exon,

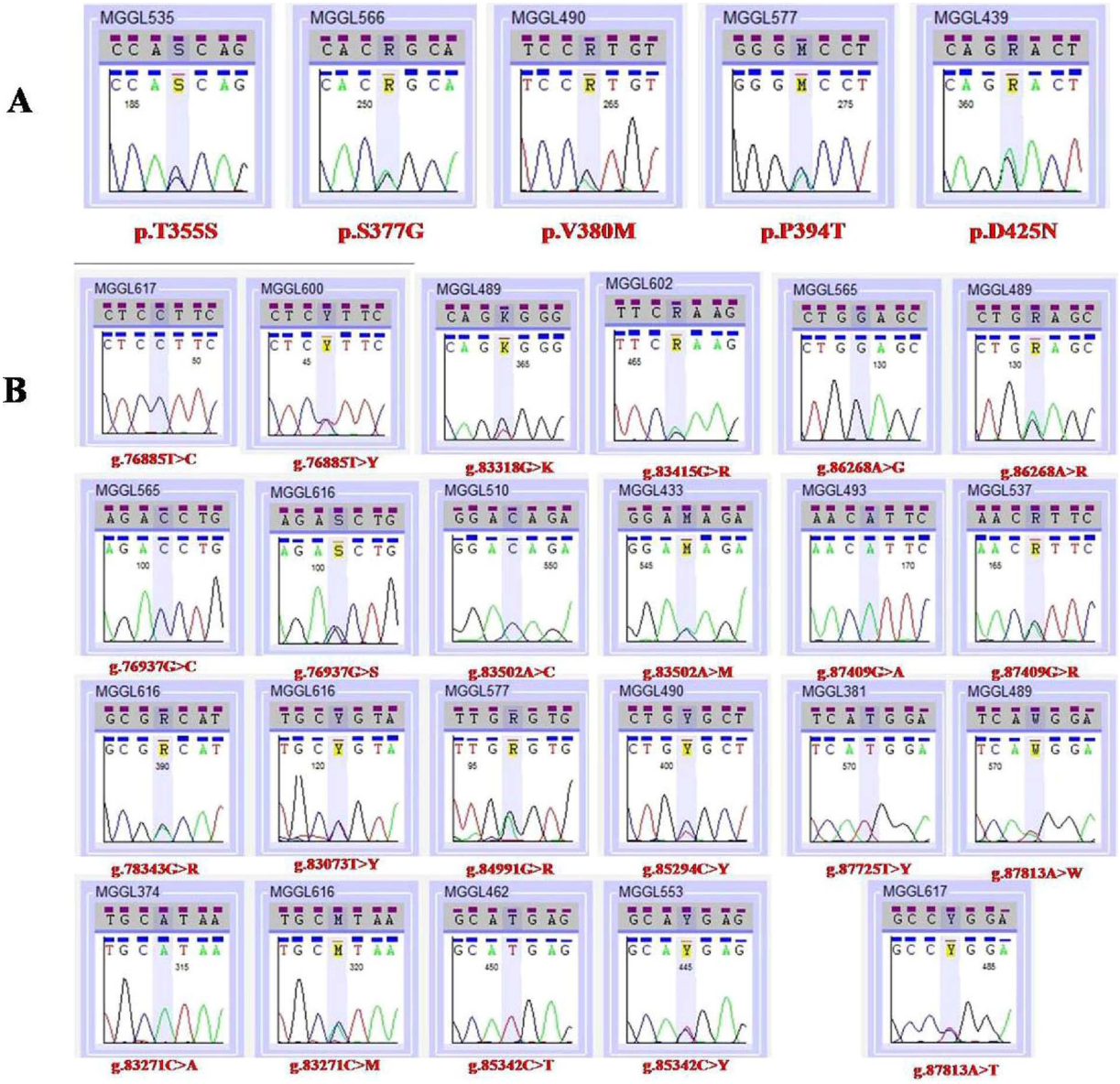

Fig. 1. (A) Sequence chromatogram indicating five heterozygous missense mutation (p.T355S, p.S377G, p.V380M, p.P394T and p.D425N). (B) Sequence chromatogram showing 15 novel intronic variants.

where as ISEs/ISSs will enhance or inhibit adjacent splice sites usage [27]. ESEs/ISEs are characterized as small RNA motifs that act as binding sites for SR proteins to enhance splicing by spliceosome recruitment [28]. ESSs/ISSs will bind to negative regulators including hnRNP family to oppose the positive effect of SR proteins and repress splicing [29–31].

SR proteins are RNA-binding proteins consists of C-terminal domain with highly conserved amino acid sequences (RS domain) and N-terminal RNA recognition domain (RRM domain) [32]. Presently, nine human SR proteins have been identified [33] and its characteristic feature is the capability to regulate 5′ splice site choice [32]. Among SR proteins, SF2/ASF, SRp40 and SRp55 were well-studied in alternative splicing. SR proteins act as splicing activators and can cooperate and compete together in splicing regulation [33].

SF2/ASF (SRSF1) was the first identified and most abundant protein which transfers between the nucleus and cytoplasm [34] and its consensus sequence is SRSASGA (S indicates G/C, R indicates Purine), rich with purine residues [33]. SRp40 (SRSF5) and SRp55 (SRSF6) are characterized with C-terminal (RS domain), RRM domain and RRM homology domain [35]. SRp40 yields the consensus sequence ACDGS (D indicates other than C, S indicates G/C) and SRp55 has consensus sequences USCGKM (S indicates G/C, K indicates U/G, M indicates A/C) [33,35]. SRp40 ESEs/ISEs are also activated by SRp55 because they are closely related in domain structures and molecular mass of 31.2 kD, 39.6 kD and 65% sequence identity [33]. At higher concentrations of SF2/ASF, it will help to use the sites which are proximal to 3′ splice site, whereas SRp40 and SRp55 will stimulate to use of distal sites and further expands the association of U1 snRNP with distal sites [32].

ISEs/ISSs are with primary importance in splicing regulation, due to the conserved sequences surrounding constitutive exons, around a

Heptamer position	Heptamer cDNA position	Branch Point position	Branch Point cDNA position	Branch Point motif	Consensus value (0-100)
44	+44	49	+49	CCCTCAG	97.27
46	+46	51	+51	CTCAGAG	73.89
54	+54	59	+59	GCCTAAG	83.82
117	+117	122	+122	ATCGGAC	72.36
133	+133	138	+138	CTTTCAG	70.07
152	+152	157	+157	AGCCCAG	79.41
206	+206	211	+211	GGCTCAA	87.17
211	+211	216	+216	AACTCAA	82.27
224	+224	229	+229	GCCTGAT	89.54
234	+234	239	+239	TGCCGAC	79.93
247	+247	252	+252	GACCCAG	79.74
254	+254	259	+259	TGCTCAG	91.17
269	+269	274	+274	AGCCGAG	74.48

A

Heptamer position	Heptamer cDNA position	Branch Point position	Branch Point cDNA position	Branch Point motif	Consensus value (0-100)
1002	+1002	1007	+1007	TGCTCAG	91.17
1013	+1013	1018	+1018	AGCTGAT	82.34
1020	+1020	1025	+1025	GCATCAC	72.65
1024	+1024	1029	+1029	CACCCAG	81.37
1026	+1026	1031	+1031	CCCAGAC	79.98
1033	+1033	1038	+1038	CCTTCAT	72.27
1039	+1039	1044	+1044	TGCCTAG	70.31
1091	+1091	1096	+1096	TCCTGAG	90.81
1098	+1098	1103	+1103	GGCTGAA	82.24
1136	+1136	1141	+1141	CCCTCCC	70.37
1139	+1139	1144	+1144	TCCCCAG	86.72
1153	+1153	1158	+1158	CTCCCAA	80.98
1154	+1154	1159	+1159	TCCCAAG	74.89
1161	+1161	1166	+1166	CCCTCAG	97.27
1182	+1182	1187	+1187	GCCTCAT	94.48

B

Fig. 2. *In silico* analysis of intronic variants by HSF3.0. (A) Potential branch point showing consensus value of g.83271C > A/M variant (B) Potential branch point showing consensus value of g.86268A > R variant.

distance of 150 bp or more in mammals [36]. Among ISSs, motif 2–(U/G)G(U/A)GGGG proved as a strong splicing silencer than motif 1(CTAGAGGT) and motif 3 (TCTCCCAA), where it predicted to wield too weak silencing effects for making any variation in splicing [28]. Another two groups identified Fox-1/2 binding motif (UGCAUG) with more conserved sequence, associated with tissue-specific splicing factors [37,38].

In the present study, the variants g.85342C > T/Y, g.86268A > G/R, g.87409G > A/R and g.87813A > T/W were observed in both cases and controls. Interestingly, these genotypic patterns observed in cases were not identical with the controls except two patterns as represented in Tables 4A and 4B. This observation suggested the possibility of other intronic variants presented responsible for the development of CHD. The other variants namely g.76937G > S, g.78343G > A/R and g.83073T > C/Y were seen in VSD with DORV condition and in PDA, apart from g.76885T > C/Y, g.83271C > A/M, g.86268A > R, and g.87813A > T/W variants, the g.87409G > A/R was picked up consistently in all cases. In one VSD, there were two missense heterozygous mutations (p.T355S and p.S377G) along with other intronic variants.

To the best of our knowledge, all these potential branch point mutations, mutations in ISEs and ISSs have not been presented in the literature. In the present study, we could not establish the role of these intronic variants to different clinical condition of CHDs and thus it could be attributed essentially to mutations in other transcription factors. These mutational studies in other genes or whole genome sequencing in CHD patients might shed light on the clinical spectrum of the disease.

5. Conclusion

In conclusion, this study expands the spectrum of *GATA4* intronic mutations in Indian population and provides additional information on molecular aberration seen in CHD patients.

Conflict of interest

The authors declare that there are no conflicts of interest.

Funding

This study was supported by Sri Devaraj Urs Academy of Higher Education and Research (SDUAHER), Kolar, Karnataka, India.

Acknowledgements

We thank the patients and their families for their kind support and participation in this study.

Table 6
Branch point analysis showing Potential branch points, Enhancer motifs (ISE) and Silencer motifs (ISS) for fifteen intronic variants.

Intron	Intronic mutations	Mutation frequency in subjects	Branch Point Analysis					
			Potential branch point		Enhancer motifs		Silencer motifs	
			Wild type	Mutant type	Wild type	Mutant type	Wild type	Mutant type
1	g.76885T > C	55/60	–	–	–	Formation of SRp40 binding motif 81.26 (78.08)	–	–
	g.76885T > Y	5/60						
1	g.76937G > C	2/60	–	–	SRp40 binding motif 79.70 (78.08)	Disruption of SRp40 binding motif	–	–
	g.76937G > S	7/60						
3	g.78343G > R	2/60	–	–	–	–	Motif 1 61.17 (60)	Disruption of Motif 1
3	g.83073T > Y	1/60	–	–	SRp55 binding motif 77.0(73.86)	Disruption of SRp55 binding motif	–	–
4	g.83271C > A	48/60	GCCTAAG 83.82 (67)	Disruption of GCCTAAG site	–	–	–	Formation of Motif 2 63.85 (60)
	g.83271C > M	10/60						
4	g.83318G > K	1/60	–	–	–	Formation of SF2/ASF binding motif 77.15 (72.98)	–	Formation of Motif 2 70.71 (60)
4	g.83415G > R	3/60	–	–	SF2/ASF2 binding motif 72.15 (72.98)	Disruption of SF2/ASF2 & Formation of SRp40 binding motifs 85.51 (78.08)	–	Formation of Motif 3 63.50 (60)
4	g.83502A > C	7/60	No significant mutation prediction					
	g.83502A > M	33/60						
4	g.84991G > R	2/60	–	–	–	–	Motif 1 70.15 (60)	Disruption of Motif 1 & Formation of Motif 2 66.40 (60)
5	g.85294C > Y	3/60	–	–	SRp40 binding motif 84.97(78.08)	Disruption of SRp40 binding motif	–	Formation of Motif 2 62.75 (60)
5	g.85342C > T	2/60	GCCGCAC 83.28 (67)	GCCGCAC −79.39	–	Formation of SRp55 binding motif 77.19 (73.86)	–	Formation of Motif 2 68.15 (60)
	g.85342C > Y	21/60						
5	g.86268A > G	2/60	GGCTGAA 82.24 (67)	Disruption of GGCTGAA site	–	–	–	–
	g.86268A > R	26/60						
6	g.87409G > A	3/60	–	–	–	Formation of SRp40 binding motif 78.98 (78.08)	Motif 1 65.06 (60)	Disruption of Motif 1
	g.87409G > R	21/60						
6	g.87725T > Y	1/60	–	–	–	Formation of SF2/ASF binding motif 92.18 (72.98)	–	Formation of Motif 1 72.64 (60)
6	g.87813A > T g.87813A > W	9/60 36/60	–	–	–	–	–	Formation of Motif 2 61.41 (60)

*Threshold value of ISEs and ISSs motif values are given in parentheses.

References

[1] J.I. Hoffman, S. Kaplan, The incidence of congenital heart disease, J. Am. Coll. Cardiol. 39 (2002) 1890–1900.
[2] T. Deeparani, M.R. Pillai, T. Elavazhagan, Detection of MTHFR C677T and A1298C gene polymorphism in congenital heart disease, Middle-East J. Sci. Res. 4 (2009) 127–132.
[3] J.M. Ko, Genetic syndromes associated with congenital heart disease, Korean Circ. J. 45 (2015) 357–361.
[4] M.E. Pierpont, C.T. Basson, D.W. Benson Jr., B.D. Gelb, T.M. Giglia, E. Goldmuntz, G. McGee, C.A. Sable, D. Srivastava, C.L. Webb, American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young, Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics, Circulation 115 (2007) 3015–3038.
[5] E. Goldmuntz, E. Geiger, D.W. Benson, NKX2.5 mutations in patients with tetralogy of Fallot, Circulation 104 (2001) 2565–2568.
[6] K.L. Clark, K.E. Yutzey, D.W. Benson, Transcription factors and congenital heart defects, Annu. Rev. Physiol. 68 (2006) 97–121.
[7] M. Maitra, M.K. Schluterman, H.A. Nichols, J.A. Richardson, C.W. Lo, D. Srivastava, V. Garg, Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development, Dev. Biol. 326 (2009) 368–377.
[8] A. Tomita-Mitchell, C.L. Maslen, C.D. Morris, V. Garg, E. Goldmuntz, GATA4 sequence variants in patients with congenital heart disease, J. Med. Genet. 44 (2007) 779–783.
[9] J.D. Molkentin, D.V. Kalvakolanu, B.E. Markham, Transcription factor GATA-4 regulates cardiac muscle-specific expression of the alpha-myosin heavy-chain gene, Mol. Cell. Biol. 14 (1994) 4947–4957.
[10] S. Pikkarainen, H. Tokola, R. Kerkelä, H. Ruskoaho, GATA transcription factors in the developing and adult heart, Cardiovasc. Res. 63 (2004) 196–207.
[11] S. Agnihotri, A. Wolf, D.M. Munoz, C.J. Smith, A. Gajadhar, A GATA4-regulated tumor suppressor network represses formation of malignant human astrocytomas, J. Exp. Med. 208 (2011) 689–702.
[12] E. Wang, S. Sun, B. Qiao, W. Duan, G. Huang, Y. An, S. Xu, Y. Zheng, Z. Su, X. Gu, L. Jin, H. Wang, Identification of functional mutations in GATA4 in patients with congenital heart disease, PLoS One 8 (2013) e62138.
[13] R.A. White, L.L. Dowler, L.M. Pasztor, L.L. Gatson, L.R. Adkison, S.V. Angeloni, D.B. Wilson, Assignment of the transcription factor GATA4 gene to human chromosome 8 and mouse chromosome 14: Gata4 is a candidate gene for Ds (disorganization), Genomics 27 (1995) 20–26.
[14] M.G. Posch, A. Perrot, K. Schmitt, S. Mittelhaus, E.M. Esenwein, B. Stiller, C. Geier, R. Dietz, R. Gessner, C. Ozcelik, F. Berger, Mutations in GATA4, NKX2.5, CRELD1, and BMP4 are infrequently found in patients with congenital cardiac septal defects, Am. J. Med. Genet. A 146A (2008) 251–253.
[15] W.T. Pu, T. Ishiwata, A.L. Juraszek, Q. Ma, S. Izumo, GATA4 is a dosage-sensitive regulator of cardiac morphogenesis, Dev. Biol. 275 (2004) 235–244.
[16] S.M. Reamon-Buettner, S.H. Cho, J. Borlak, Mutations in the 3′-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD), BMC Med. Genet. 8 (2007) 38.
[17] J. Chen, B. Qi, J. Zhao, W. Liu, R. Duan, M. Zhang, A novel mutation of GATA4 (K300T) associated with familial atrial septal defect, Gene 575 (2016) 473–477.
[18] S. Ramegowda, A. Kumar, M.R. Savitha, B. Krishnamurthy, N. Doddaiah, N.B. Ramachandra, Missense mutation G296S in GATA4 is not responsible for cardiac septal defects, Indian J. Hum. Genet. 13 (2007) 30–32.
[19] S.M. Dinesh, K. Lingaiah, M.R. Savitha, B. Krishnamurthy, D. Narayanappa, N.B. Ramachandra, GATA4 specific nonsynonymous single-nucleotide polymorphisms in congenital heart disease patients of Mysore India, Genet. Test. Mol. Biomark. 15 (2011) 715–720.
[20] S. Mattapally, S. Nizamuddin, K.S. Murthy, K. Thangaraj, S.K. Banerjee, c.620C > T mutation in GATA4 is associated with congenital heart disease in South India, BMC Med. Genet. 18 (16:7) (2015).
[21] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, Nucleic Acids Res. 16 (1988) 1215.
[22] F.O. Desmet, D. Hamroun, M. Lalande, G. Collod-Beroud, M. Claustres, C. Beroud, Human Splicing Finder: an online bioinformatics tool to predict splicing signals, Nucleic Acids Res. 37 (2009) e67.
[23] T. Hara, M. Ichihara, M. Takagi, A. Miyajima, Interleukin-3 (IL-3) poor-responsive inbred mouse strains carry the identical deletion of a branch point in the IL-3 receptor alpha subunit gene, Blood 85 (1995) 2331–2336.
[24] C. Maslen, D. Babcock, M. Raghunath, B. Steinmann, A rare branch-point mutation is associated with missplicing of fibrillin-2 in a large family withcongenital contractural arachnodactyly, Am. J. Hum. Genet. 60 (1997) 1389–1398.
[25] R. Scholl, J. Marquis, K. Meyer, D. Schumperli, Spinal muscular atrophy: position and functional importance of the branch site preceding SMN exon 7, RNA Biol. 4 (2007) 34–37.
[26] J.M. Johnson, J. Castle, P. Garrett-Engele, Z. Kan, P.M. Loerch, C.D. Armour, R. Santos, E.E. Schadt, R. Stoughton, D.D. Shoemaker, Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays, Science 302 (2003) 2141–2144.
[27] Z. Wang, C.B. Burge, Splicing regulation: from a parts list of regulatory elements to an integrated splicing code, RNA 14 (2008) 802–813.
[28] M. Sironi, G. Menozzi, L. Riva, R. Cagliani, G.P. Comi, N. Bresolin, R. Giorda, U. Pozzoli, Silencer elements as possible inhibitors of pseudoexon splicing, Nucleic Acids Res. 32 (2004) 1783–1791.
[29] C.D. Chen, R. Kobayashi, D.M. Helfman, Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat b-tropomyosin gene, Genes Dev. 13 (1999) 593–606.
[30] F. Del Gatto-Konczak, M. Olive, M.C. Gesnel, R. Breathnach, hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer, Mol. Cell. Biol. 19 (1999) 251–260.
[31] Y. Lee, D.C. Rio, Mechanisms and regulation of alternative pre-mRNA splicing, Annu. Rev. Biochem 84 (2015) 291–323.
[32] J. Valcárcel, M.R. Green, The SR protein family: pleiotropic functions in pre-mRNA splicing, Trends Biochem. Sci. 21 (1996) 296–301.
[33] H.X. Liu, M. Zhang, A.R. Krainer, Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins, Genes Dev. 12 (1998) 1998–2012.
[34] A.M. Zahler, K.M. Neugebauer, W.S. Lane, M.B. Roth, Distinct functions of SR proteins in alternative pre-mRNA splicing, Science 260 (1993) 219–222.
[35] S. Jeong, SR proteins: binders, regulators. and connectors of RNA, Mol. Cells 40 (2017) 1–9.
[36] R. Sorek, G. Ast, Intronic sequences flanking alternatively spliced exons are conserved between human and mouse, Genome Res. 13 (2003) 1631–1637.
[37] R.B. Voelker, J.A. Berglund, A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing, Genome Res. 17 (2007) 1023–1033.
[38] G.W. Yeo, E.L. Nostrand, T.Y. Liang, Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements, PLoS Genet. 3 (2007) e85.