1   A novel GATA-binding protein 4 gene variation associated with familial atrial septal

2   defect

3

4   Yongchao Yang[1,4†], Yu Xia[1,4†], Yueheng Wu[1,3,4], Shufang Huang[3,4], Yun Teng[1],

5   Xiaobing Liu[1], Ping Li[2,3], Jimei Chen[1,4], Jian Zhuang[1,4*]

6   [1]Department of Cardiac Surgery, Guangdong Cardiovascular Institute, Guangdong

7   General Hospital, Guangdong Academy of Medical Sciences, Guangzhou,

8   Guangdong, China, 510100

9   [2]Department of Obstetrics and Gynecology, Guangdong General Hospital, Guangdong

10   Academy of Medical Science, Guangzhou, Guangdong, China, 510100

11   [3]Prenatal Diagnosis Center, Guangdong General Hospital, Guangdong Academy of

12   Medical Science, Guangzhou, Guangdong, China, 510100

13   [4]Guangdong Provincial Key Laboratory of South China Structural Heart Disease,

14   Guangdong General Hospital, Guangdong Academy of Medical Science, Guangzhou,

15   Guangdong, China, 510100

16   † Yongchao Yang and Yu Xia contributed equally to this article and share co-first

17   authorship.

18   * Corresponding author

19   **Running Title:** Novel mutant GATA4 associated with ASD

20   **Keywords**. atrial septal defect; variation; GATA-binding protein 4; familial.

21   **Corresponding author:**

22   Jian Zhuang, Department of Pediatric Cardiovascular Surgery, Guangdong

23   Cardiovascular Institute, Guangdong General Hospital, Guangdong Academy of

24   Medical Sciences, 96 Dongchuan Road, Guangzhou 510100

25   Email: zhuangjiangenetics@163.com

26

27   **Abstract**

28   Atrial septal defect (ASD) is the most common congenital heart defect. Part of ASD

29   exhibits familial predisposition, but the genetic mechanism remains largely unknown.

30   In the current study, we use multiple methods to identify and confirm the gene

31   associated with a familial ASD. Chromosomal microarray analyses, whole exome

32   sequencing, Sanger sequencing, multiple bioinformatics programs, in silico protein

33   structure modeling and molecular dynamics simulation were performed to predict the

34   pathogenic of the variant gene. Dual-Luciferase reporter gene assay was performed to

35   evaluate the influence of downstream target gene of the target variation. A novel,

36   heterozygous, missense variant GATA-binding protein 4 (*GATA4*):c.958C>T,

37   p.R320W was identified. An autosomal dominant inheritance pattern with incomplete

38   penetrance was observed in the family. Multiple prediction indicate the variant in

39   *GATA4* to be deleterious. Molecular dynamics simulation further revealed that the

40   variation of p.R320W could prevent the zinc finger of *GATA4* from interacting with

41   the DNA. Dual-Luciferase reporter assay demonstrated a significant decrease in

42   transcriptional activity (0.90 ± 0.099 vs 1.50 ± 0.079, $p = 0.001$) of the variant *GATA4*

43   compared with the wild type. We believe the novel variation of *GATA4* (c.958C>T,

44   p.R320W) with a pattern of incomplete inheritance that may be highly associated with

45   this familial ASD. The finding enriched our knowledge of variations that may

46   associated with ASD.

47 **Introduction**

48 Congenital heart defect (CHD) is one of the most common life-threatening birth

49 defects, with an estimated incidence of 7 to 9 per one thousand and affecting 100,000

50 to 150,000 newborns in China each year.[1-3] There are more than 400 identified variant

51 genes, including *NKX2-5*, *MYH6*, *GATA4*, *ZIC3*, and *ELN*, associated with CHD and

52 accounting for about 10% of CHD.[4-7] The ostium secundum atrial septal defect (ASD),

53 comprising of approximately 30 to 40% of CHD, is one of the most common subtypes

54 of interatrial communication,[8] and often associated with other cardiac and/or

55 extra-cardiac anomalies and genetic syndromes.[9]

56 The understanding of the genetic pathogenesis mechanisms of ASD has

57 significantly improved in recent decades. Pathogenic genomic copy number variants

58 (CNVs) and gene variants have been identified in ASD. Rare CNVs at recurrent loci,

59 such as 22q11.2, 5q35.1, 8p23.1, and 18q11.2, and more than ten genes, including as

60 *NKX2-5*, *GATA4, GATA6*, *TBX5*, *TBX20*, and *CITED2*, have been associated with

61 sporadic or inherited ASD.[10-12]

62 *GATA4*, a zinc finger transcription factor that contains seven exons located on

63 chromosome 8p23.1-p22, plays an important role in early stage of embryonic heart

64 development.[13,14] It is comprising of 442 amino acids with four conserved domains –

65 transcription activation domain 1(TAD1, amino acids (aa) 1 to 74), transcription

66 activation domain 2 (TAD2, aa 130 to 177), N-terminal zinc finger (ZF1, aa 217 to

67 241), and C-terminal zinc finger (ZF2, aa 271 to 295).[14,15] The nuclear localization

68 signal (NLS, aa 271 to 325) allows GATA4 to be imported to the nucleus via the

69   nuclear pore complex.[14] Small changes in the level of GATA4 protein expression can

70   dramatically influence cardiac development and embryonic survival. Presently, there

71   are 150 variant sites of *GATA4* that have been identified in association with different

72   subtypes of CHD, including atrial septal defects, ventricular septal defects, tetralogy

73   of Fallot, and atrial fibrillation.[16-19] There are 104 missense or nonsense variants of

74   *GATA4* reported to be associated with these CHD in Human Gene Mutation Database

75   (HGMD) (http://www.hgmd.cf.ac.uk/ac/all.php).

76   However, as a genetically heterogeneous disease, such a large number of

77   variations still cannot fully cover the variations of ASD. Here, we use multiple

78   methods to identify and confirm the variant gene associated with a familial ASD.

79

80   **Materials and methods**

81   **Study subjects**

82   A family with 8 members were enrolled in this study (II-2, III-1, III-2, III-3, IV-3,

83   IV-4, IV-5 and IV-6, Figure 1). All enrolled members underwent a complete physical

84   examination. Clinical data, including medical records, electrocardiograms, and

85   echocardiography were systematically reviewed. The study protocol (protocol number

86   20140829) was approved by the Research Ethics Committee of Guangdong General

87   Hospital, Guangdong Academy of Medical Sciences, Guangdong, China. Informed

88   written consent (informed consent form number 20150424) was obtained from all

89   family members. Approximately 6.0 mL of peripheral blood was collected from each

90   of the study participants. DNA was extracted from the peripheral blood lymphocytes

91   using modified salting-out precipitation method by Gentra Puregene blood kit

92   (QIAGEN, Santa Clara, CA, USA).

93

94   **Chromosomal microarray analysis and CNV evaluation and validation**

95   Two hundred fifty nanograms (ng) of DNA was amplified from III-1, III-2, IV-3, and

96   IV-4 (Fig. 1), then labeled and hybridized to the CytoScan HD array platform

97   (Affymetrix, USA) according to the manufacturer's protocol. The array was designed

98   specifically for cytogenetic research, offering more than 2,700,000 markers across the

99   whole genome, including 750,000 SNP probes and 1,950,000 probes to detect CNVs

100   (Cyto-arrays). Data were visualized and analyzed with the Chromosome Analysis

101   Suite (ChAS) software package (Affymetrix, USA) with a minimal cutoff of 20

102   consecutive markers in a length of 25-kb for CNVs calling. All of the segments were

103   monitored for the degree of overlap with previously identified common CNVs,

104   annotated by Database of Genomic Variants (DGV). All of the reported CNVs are

105   based on NCBI human genome build 37 (hg 19).

106   Detected CNVs meeting the following criteria were selected for further analysis:

107   (1) deletions greater than or equal to 50 kb and duplications greater than or equal to

108   50 kb; (2) without recurrence in the normal populations that have been cataloged in

109   DGV; and (3) possessing less than 50% overlap with known segmental duplications.

110   Following the American College of Medical Genetics and Genomics' (ACMG)

111   standards and guidelines for the interpretation of CNVs, the remaining CNVs were

112   classified into three categories—pathogenic (P), variants of uncertain significance

113 (VOUS), and benign (B). VOUS was further divided into three parts—likely

114 pathogenic (LP), likely benign (LB), and no sub-classification (NS). For this study,

115 only genes that function in a dominant manner that are within P and LP CNVs were

116 investigated. All of the annotated CNVs were experimentally validated by real-time

117 quantitative PCR (qPCR).

118

119 **Whole exome sequencing and variant analysis**

120 To systematically search for disease-causing genes, exome sequencing in three

121 affected individuals and two unaffected individuals (parents of proband) from the

122 family with a history of ASD was performed using the Agilent Sure Select Human All

123 Exon V5 Kit on the Illumina HiSeq 2000 platform by Novogene Bioinformatics

124 Technology Co., Ltd. One and 0.5 µg of genomic DNA from the proband was used to

125 construct the exome library. The genomic DNA was sheared into fragments with a

126 length of 180 to 280 bp by sonication and hybridized for enrichment according to the

127 manufacturer's protocol. The library enriched for target regions was sequenced on the

128 Illumina HiSeq 2000 platform to get paired-end reads with a read length of 100 bp.

129 The average sequencing depth of 57.36X provided enough depth to exactly call

130 variants at 97.4% of the targeted exome.

131   The human reference genome was obtained from the University of California

132 Santa Cruz (UCSC) database (build 37.1, version hg19, http://genome.ucsc.edu/), and

133 sequence alignment was performed using the Burrows-Wheeler Alignment tool.

134 High-quality alignment was required to guarantee variant calling accuracy (greater

135   than 0). Picard (http://sourceforge.net/projects/picard/) was employed to mark

136   duplicates resulting from PCR amplification. Genome Analysis Toolkit (GATK) Indel

137   Realigner and GATK Realigner Target Creator were performed to do realignment

138   around the indels. GATK Base Recalibrator was performed to do base quality score

139   recalibration. GATK Variant Filtration was performed to make the raw callsets

140   suitable for meaningful analysis. Exome CNV was performed for CNV detection.

141   Sequence Alignment/Map (SAM) tools were used to perform variant calling and

142   identify single nucleotide polymorphisms (SNPs) or indels. After the analysis-ready

143   Binary Alignment/Map (BAM) alignment result was obtained, Annotate Variation

144   (ANNOVAR) was performed to annotate SNPs and indels. All candidate variants

145   were filtered against the Single Nucleotide Polymorphism Database (dbSNP142,

146   http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi), 1000 Genomes Project

147   (2016 April release, http://www.1000genomes.org/), Exome Aggregation Consortium

148   (ExAC, http://exac.broadinstitute.org/) and NHLBI Exome Sequencing Project (ESP)

149   6500 to remove the polymorphism loci. Sorting Intolerant from Tolerant (SIFT) and

150   Polymorphism Phenotyping version 2 (PolyPhen-2), Mutation Taster, and Combined

151   Annotation Dependent Depletion (CADD) were performed to predict whether an

152   amino acid substitution affects the function of the protein.

153   In addition to the standard variant quality controls, six independent filters were

154   applied to facilitate detection of possible causal variants among the enrolled ASD

155   patients. Variants were filtered by: (1) a minor allele frequency (MAF) of less than 1%

156   in east Asian population; (2) CADD score of greater than 20; (3) the variants must be

157   pathogenic; (4) highly expressed in the heart or associated with cardiac development;

158   (5) genotype–phenotype matched under the assumption of complete penetrance.

159

160   **Sanger sequencing confirmation**

161   Direct Sanger sequencing was performed with ABI 3500 sequencer (Applied

162   Biosystems, Foster City, CA, USA) to confirm potential causative variants in the

163   family. Primer sequences for pathogenic variant in the *GATA4* gene (NM_002052.4)

164   were   designed   as   follows: 5'-CAATGCCTGCGGCCTCTAC-3'   and

165   5'-AGGAAGAAGACAAGGGAGGACTG-3'. Once a variant was confirmed, all

166   family members were screened to analyze variant segregation within the family.

167

168   **Conservation analysis across species, in silico protein structure modeling and**

169   **molecular dynamics simulation**

170   The protein sequences of GATA4 in 11 species from *Drosophila melanogaster* to

171   *Homo sapiens* were aligned using Clustal X (version 1.81) software.

172   The wild type GATA4 protein from the previous homology modeling study was

173   used as the initial structure in molecular dynamics (MD) simulations. Then, the

174   energy minimized and equilibrated structure was used to acquire the R320W mutant

175   protein by using UCSF Chimera.[20] Finally, 10 ns MD simulations were performed for

176   both wild type and mutant proteins of GATA4. All preparation and simulations were

177   performed in AmberTools15 (Amber 2015, University of California, San Francisco).

178   The Leap module was employed to assign AMBER ff14SB force field for protein and

179   zinc (Zn^{2+}) ions. Zn^{2+} ions with coordinate atoms were constrained to be tetrahedron

180   using the cationic dummy atom (CaDA) approach of Pang et al.[21] TIP3P water model

181   was used and the box was set to 10 Å. Counter ions Cl^- were added in order to

182   neutralize charges of the system. A 30 ns of NPT ensemble MD simulation was

183   performed using Lengevin dynamics method[22] to control temperature with collision

184   frequency of 1.0 ps^{-1}. A SHAKE algorithm was applied to constrain bonds involving

185   hydrogen atoms. The van der Waals cutoff was kept 10 Å and long range electrostatic

186   interactions were treated using the Particle Mesh Ewald (PME)[23] method. Atomic

187   coordinates were saved after every 500 steps. RMSD, RMSF, radial gyration (Rg),

188   solvent accessible surface area (SASA) and secondary structure analyses were carried

189   out to study the effect that mutagenesis would have on the structure and functions of

190   GATA4 protein.

191

192   **Plasmids and site-directed mutagenesis**

193   The full-length wild type cDNA of the human *GATA4* gene was amplified by PCR

194   using PrimeSTAR® HS DNA Polymerase (Takara, Liaoning, CHN) and primers

195   (5'-GGGGTACCATGTATCAGAGCTTGGCCATGGCC-3'   and

196   5'-CCGCTCGAGTTACGCAGTGATTATGTCCCC-GTGA-3'). PCR fragments were

197   double digested by endonucleases Kpnl and Xhol (Thermo Fisher, Shanghai, CHN).

198   The digested product was fractionated by using 1.5% agarose gel electrophoresis,

199   purified by using the E.Z.N.A® Gel Extraction Kit (OMEGA, Norcross, GA, USA),

200   and then subcloned into pcDNA3.1 (Promega, Beijing, CHN) to construct the

201   recombinant eukaryotic expression vector WT-pcDNA3.1-hGATA4.

202   The *GATA4* variant c.958C>T (p.R320W) was introduced into a wildtype *GATA4*

203   clone using a QuikChange Site-Directed Mutagenesis Kit (Stratagene, Agilent

204   Technologies, CA, USA) and primers

205   (5'-ATCCAAACCAGAAAATGGAAGCCCAAGAACC-3' and

206   5'-GGTTCTTGGGCTTCCATTTTCTGGTTTGGAT-3'). The clones were sequenced

207   to confirm the expected variant and exclude other variants.

208

209   **Dual-Luciferase assays**

210   293T cells were transiently transfected with 400 ng brain natriuretic peptide

211   (*BNP*)-luciferase reporter plasmid and internal control reporter plasmid pGL4.75

212   [hRluc/CMV] (Promega, Southampton, UK) in combination with 100 ng of wild type

213   *GATA4*, mutant type *GATA4* p.R320W, or pcDNA3 plasmids using Lipofectamine^{TM}

214   2000 (Invitrogen, Cat.) according to the manufacturer's protocol. Luciferase activity

215   was measured 48 hours after transfection using the Dual-Glo Luciferase Assay

216   (Promega, Southampton, UK) according to the manufacturer's protocol. Mean

217   luciferase activity was calculated after normalization to Renilla luciferase activity.

218   The experiments were performed and repeated at least three times respectively.

219   Finally, the expression level of protein GATA4 were evaluated by western blot

220   analysis.

221

222   **Statistical Analysis**

223   Statistical analysis was performed with SAS-PC (9.3v, SAS Institute, Inc, Cary, NC).

224   Results are expressed as mean ± Standard Deviation (SD). Logarithmic

225   transformation was applied to the data for the luciferase assay to achieve approximate

226   normality. Comparisons between two groups were performed with the chi-square test.

227   Statistical significance was determined by two-way analysis of variance and

228   subsequent pairwise comparisons were performed. A p value of less than 0.05 was

229   considered to be statistically significant.

230

231   **Data availability**

232   The authors state that all data necessary for confirming the conclusions presented in

233   the article are represented fully within the article.

234

235   **Results**

236   **Clinical Characteristics**

237   The proband of this family is the second child, who was diagnosed with ostium

238   secundum atrial septal defect by ultrasound (IV-4) (Fig. 1). After reviewing the family

239   history, we identified the first child (IV-3) (Fig. 1) was also affected by the same

240   defect and had been cured by transcatheter closure four years prior to our interview.

241   The mother of proband was not detected with any cardiac disorders by

242   echocardiogram (III-2) (Figure 1). The maternal grandmother (II-2) (Fig. 1) indicated

243   upon interview that she had undergone cardiac surgery due to ASD when she was

244   approximately 20 years of age (the detail medical records were missing). She also

245   informed us that her mother (I-2) (Fig. 1) may have had a cardiac disorder as she was

246   unable to perform manual labor and died due to an unidentified cause at

247   approximately 40 years of age. The mother of the proband had two abnormal

248   pregnancies with a spontaneous miscarriage at approximately 8 weeks during the first

249   pregnancy (IV-1) (Fig. 1) and a termination due to the detection of fetal bradycardia at

250   18th week during the second pregnancy (IV-2) (Fig. 1). Conventional G-banded

251   cytogenetic analysis was performed for patients II-2, III-1, III.2, IV-3 and IV-4, but no

252   clinically significant result was found.

253

254   **No significant CNV was found associated with ASD**

255   CMA was performed on DNA samples from members III-1, III-2, IV-3, and IV-4.

256   There was only one subject (III-1) with detection of a 1.641 Mb duplication at

257   Yq11.223 (chrY: 24,148,853-25,790,030). This duplication only contents one disease

258   causing gene of *DAZ1* according to OMIM. The deletion of this gene may be related

259   to spermatogenesis, but it is unknown if its duplication has any significant clinical

260   annotation.

261

262   *GATA4* **was located as target variant gene for this familial ASD**

263   Exome analysis was performed on five DNA samples from members II-2, III-1, III-2,

264   IV-3 and IV-4. A total of 19,816,364 pairs of sequenced reads with the average read

265   length of 125 bp were generated by exome sequencing. Approximately 98.76%

266   (19,569,878) of sequenced reads passed the quality assessment and were mapped to

267   99.81% of the human reference genome. There were more than 20,000 indels,

268   167,000 SNPs and 10 CNVs found in each subject. After filtering, more than 1500

269   variants including SNP and indels were shared by II-2, IV-3 and IV-4, and finally, a

270   heterozygous missense variant *GATA4*: NM_002052: exon4: c.C958T: p.R320W

271   (CADD_Phred score: 35, SIFT score: 0.0, Polyphen2_HVAR score: 0.999,

272   Polyphen2_HDIV score: 1.0, Mutation Taster predict: Disease causing), was selected

273   as our target pathogenic variant detected in IV-3, IV-4, III-2 and II-2.

274   In order to validate the target variant, we performed the Sanger sequencing in all

275   relevant subjects (II-2, III-1, III-2, III-3, IV-3, IV-4, IV-5 and IV-6) (Fig. 2A). The

276   heterozygous *GATA4* p.R320W variant was found in subjects II-2, III.2, IV-3 and IV-4

277   but not in subjects III-1, III-3, IV-5 and IV-6 (Table 1).

278

279   **The variant site of *GATA4* is evolutionarily conserved and plays an important**

280   **role in the region of nuclear localization signal**

281   A cross-species alignment of the *GATA4* amino acid sequences revealed that the

282   altered amino acid arginine (CGG) at position 320 was completely conserved

283   evolutionarily (Fig. 2B). The identified variant is located in the region of the nuclear

284   localization signal (Fig. 2C), which plays an important role in the nuclear

285   translocation of *GATA4*.[14]

286   After structure preparation and MD simulation, we found the amino acid

287   structure of *GATA4* in position of 214-322 mainly contains two zinc binding regions

288   linked by a random coil, with the variant site located at one end of coil. In wild type

289   *GATA4*, the random coil is far from the other zinc binding region without any

290   connection. The structure has an "open loop", with an angle of 167° between the

291   two-helix structures of zinc binding regions (Fig. 3A). In the mutant *GATA4*,

292   connections via hydrogen bonds and hydrophobic interactions between the random

293   coil and the opposite side of zinc binding region were formed. The structure has a

294   "closed loop", with an angle of about 73° (Fig. 3C). This was likely due to the change

295   of electric charge from the variation of Arg320 to Trp320 where Trp320 formed the

296   hydrophobic interaction with Leu227 and hydrogen bonds with Arg230. The variant

297   also forced a change of the random coil in segments of 256 to 259 into a helical

298   structure (Fig. 3B, 3D).

299

300   **Variation of *GATA4* significantly decrease the expression of *BNP***

301   Previous studies have revealed that *GATA4* is an upstream transcriptional regulator of

302   several genes expressed in different signaling pathway during cardiac development,

303   including genes that encode atrial natriuretic peptide (*ANP*), brain natriuretic peptide

304   (*BNP*), and β-myosin heavy chain (*MHC*).[13] Therefore, the functional effect of the

305   *GATA4* variant may be reflected by biochemical analysis of the transcriptional activity

306   of the *BNP* promoter in cells transfected with mutant *GATA4* in contrast to its wild

307   type counterpart.

308   In the dual-luciferase reporter assay (Fig. 4A), we found wild type *GATA4* can

309   significantly increase the transcription of *BNP* in comparison with nature control type

310   (1.50 ± 0.079 vs 1.0 ± 0.064, $p = 0.001$), which in accordance to previous knowledge

311   that *GATA4* is activating transcription factor for *BNP*. The mutant *GATA4* displayed a

312   significant decrease in transcriptional activity in comparison to wild type (0.90 $\pm$

313   0.099 vs 1.50 $\pm$ 0.079, *p* = 0.001). There is no significantly difference in protein levels

314   between wild type and mutant GATA4 by analyzing of Western blot (Fig. 4B).

315

316   **Discussion**

317   We interviewed and evaluated members of a family with at least 3 patients diagnosed

318   with ASD in the clinic. After a complete examination including karyotyping, CMA,

319   WES and Sanger sequencing for the whole family, we found a novel, heterozygous,

320   missense variation of *GATA4*, c.C958T:p.R320W, in 3 patients with ASD and one

321   unaffected carrier (the mother). This finding was not consistent with traditional

322   autosomal dominant Mendelian inheritance. Interestingly, E. D'Amato et al[24] reported

323   a heterozygous missense variant *GATA4* c.1512C>T, p.Arg319Trp by HGMD in two

324   children from an Italian family with pancreatic agenesis and ASD, also with the

325   inheritance of incomplete penetrance. There is no completely reasonable explanation

326   for this type of inheritance, it may also be related to complicated gene–gene or gene–

327   environment interactions. With the childbearing history of two terminal pregnancies,

328   the pleiotropism of the gene may explain why the proband's mother carried the same

329   *GATA4* variant but had distinct clinical phenotypes. This phenomenon may highlight

330   the fact that even in a familial CHD, the underlying genetic etiology can be complex.

331   There is no record of this variant in the databases of 1000G, ESP6500, ExAC

332   and NCBI. However, the bioinformatics programs across all prediction algorithms,

333   including PolyPhen-2, SIFT, Mutation Taster, and CADD, etc. According to the

334   conclusions of Philips AS, our finding of the *GATA4* variant p.R320W may be

335   pathogenic.[14] The algorithms demonstrated that four amino acids, Arg282, Arg283,

336   Arg317, Arg319, played crucial roles in nuclear localization of *Gata4* in murine cell

337   line models. Coincidentally, we identified the same variant site of Arg320 (the same

338   site with murine Arg319) in this family. Given that this variant site is highly

339   evolutionarily conserved across 11 species, the variant may lead to a pathogenic

340   change of function after translated into protein. Our prediction of molecular

341   architecture of *GATA4* protein shows the influence of the variant on changing the

342   stability of conformation and structure, which may decrease or inhibit its function in

343   transcription. Therefore, it is very likely that dysfunctional *GATA4* contributes to ASD

344   in this family.

345   As a key transcription factor, *GATA4* regulates transcription of many genes

346   involved heart development, including *MHC*, *ANP*, *BNP* and endothelial nitric oxide

347   synthesis (*eNOS*),[13.25] and *GATA4* also plays an essential role in cardiac adaptive

348   responses, including myocyte survival, angiogenesis, and hypertrophy in response to

349   exercise.[13, 25-27] In our dual-luciferase reporter assay, we also confirmed that *GATA4* is

350   a transcriptional factor of BNP. The variation of *GATA4* p.R320W significantly

351   decreased its transcriptional activity on downstream cardiac genes of *BNP* in 293T

352   cells, indicating that this variant could contribute to the pathogenesis of ASD.

353   Our study demonstrates the fact that CMA and WES are reliable techniques in

354   detection of pathogenic variants in familial ASD. Detailed clinical genetic data will

355   facilitate genetic diagnosis and counseling when evaluating the prognosis of newborns

356   and the future risks for other family members with new pregnancies. Our future work

357   will focus on exploring the change of relevant signaling pathway caused by the

358   variation of *GATA4* and the prevalence of this variant site in population of ASD.

359

360   **Conclusion**

361   In conclusion, we identified a novel variation of *GATA4* (c.958C>T, p.R320W) in

362   familial ASD. The identified variant caused impaired biological function of the

363   protein in vitro, suggesting that it is likely to play a role in the pathogenesis of ASD.

364   Reasonable use of a variety of variant detection techniques will help us to identify

365   pathogenic variants in CHD patients, especially in familial disease. Elucidation of the

366   genetic basis of CHD has valuable clinical implications and will continue to expand

367   our understanding of the pathogenesis of CHD.

368 **Conflict of Interest**

369 The authors declare that there are no conflicts of interest.

370

371 **Acknowledgments**

Reference

[1]van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, et al. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J Am Coll Cardiol. 2011; 58:2241-7.

[2]Bernier PL, Stefanescu A, Samoukovic G, Tchervenkov CI. The challenge of congenital heart disease worldwide: epidemiologic and demographic facts. Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu. 2010; 13:26-34.

[3]Hu Z, Yuan X, Rao K, Zheng Z, Hu S. National trend in congenital heart disease mortality in China during 2003 to 2010: a population-based study. J Thorac Cardiovasc Surg. 2014; 148:596-602.

[4]Gelb BD. Recent advances in understanding the genetics of congenital heart defects. Curr Opin Pediatr. 2013;25:561-6.

[5]Ellesøe SG, Johansen MM, Bjerre JV, et al. Familial Atrial Septal Defect and Sudden Cardiac Death: Identification of a Novel NKX2-5 Mutation and a Review of the Literature. Congenit Heart Dis. 2016; 11(3):283-90.

[6]Theis JL, Zimmermann MT, Evans JM, et al. Recessive MYH6 mutations in hypoplastic left heart with reduced ejection fraction. Circ Cardiovasc Genet. 2015; 8:564–571

[7]El Malti R, Liu H, Doray B, et al. A systematic variant screening in familial cases of congenital heart defects demonstrates the usefulness of molecular genetics in this field. Eur J Hum Genet. 2016; 24:228–236.

[8]Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll

405 Cardiol. 2002; 39:1890-900.

406 [9]Jin SC, Homsy J, Zaidi S, et al. Contribution of rare inherited and de novo variants

407 in 2,871 congenital heart disease probands. Nat Genet. 2017; 49:1593-1601.

408 [10]Pires R, Pires LM, Vaz SO, et al. Screening of copy number variants in the

409 22q11.2 region of congenital heart disease patients from the São Miguel Island,

410 Azores, revealed the second patient with a triplication. BMC Genet. 2014;15:115.

411 [11]Fahed AC, Gelb BD, Seidman JG, Seidman CE. Genetics of congenital heart

412 disease: the glass half empty. Circ Res. 2013;112:707-20.

413 [12]Lalani SR, Belmont JW. Genetic basis of congenital cardiovascular

414 malformations. Eur J Med Genet. 2014;57:402-13.

415 [13]Molkentin JD. The zinc finger-containing transcription factors GATA-4, -5, and

416 -6. Ubiquitously expressed regulators of tissue-specific gene expression. J Biol Chem.

417 2000;275:38949-52.

418 [14]Philips AS, Kwok JC, Chong BH. Analysis of the signals and mechanisms

419 mediating nuclear trafficking of GATA-4. Loss of DNA binding is associated with

420 localization in intranuclear speckles. J Biol Chem. 2007;282:24915-27.

421 [15]Reamon-Buettner SM, Borlak J. GATA4 zinc finger mutations as a molecular

422 rationale for septation defects of the human heart. J Med Genet. 2005;42:e32.

423 [16]Yang YQ, Gharibeh L, Li RG, Xin YF, Wang J, Liu ZM, Qiu XB, Xu YJ, Xu L,

424 Qu XK, Liu X, Fang WY, Huang RT, Xue S, Nemer G. GATA4 loss-of-function

425 mutations underlie familial tetralogy of fallot. Hum Mutat. 2013;34:1662-71.

426 [17]Chen Y, Han ZQ, Yan WD, Tang CZ, Xie JY, Chen H, Hu DY. A novel mutation

427   in GATA4 gene associated with dominant inherited familial atrial septal defect. J

428   Thorac Cardiovasc Surg. 2010;140:684-7.

429   [18]Yang YQ, Li L, Wang J, Liu XY, Chen XZ, Zhang W, Wang XZ, Jiang JQ, Liu X,

430   Fang WY. A novel GATA4 loss-of-function mutation associated with congenital

431   ventricular septal defect. Pediatr Cardiol. 2012;33:539-46.

432   [19]Posch MG, Boldt LH, Polotzki M, Richter S, Rolf S, Perrot A, Dietz R, Ozcelik C,

433   Haverkamp W. Mutations in the cardiac transcription factor GATA4 in patients with

434   lone atrial fibrillation. Eur J Med Genet. 2010;53:201-3.

435   [20]Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera–A visualization

436   system for exploratory research and analysis. J Comput Chem. 2004; 25:1605-12.

437   [21]Pang, Y.-P. Novel zinc protein molecular dynamics simulations: Steps toward

438   antiangiogenesis for cancer treatment. Journal of Molecular Modeling. 1999,

439   5:196-202.

440   [22]Jean-Paul Ryckaert, Giovanni Ciccotti, Herman J. C. Berendsen. Numerical

441   integration of the cartesian equations of motion of a system with constraints:

442   Molecular Dynamics of N-alkanes. Journal of Computational Physics, 1977,

443   23:327-41.

444   [23]Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N log (N)

445   method for Ewald sums in large systems. Journal of Chemical Physics, 1993,

446   98:10089-92.

447   [24]D'Amato E, Giacopelli F, Giannattasio A, D'Annunzio G, Bocciardi R, Musso M,

448   Lorini R, Ravazzolo R. Genetic investigation in an Italian child with an unusual

449   association of atrial septal defect, attributable to a new familial GATA4 gene mutation,

450   and neonatal diabetes due to pancreatic agenesis. Diabet Med. 2010;27:1195-200.

451   [25]Heineke J, Auger-Messier M, Xu J, Oka T, Sargent MA, York A, Klevitsky R,

452   Vaikunth S, Duncan SA, Aronow BJ, Robbins J, Crombleholme TM, Molkentin JD.

453   Cardiomyocyte GATA4 functions as a stress-responsive regulator of angiogenesis in

454   the murine heart. J Clin Invest. 2007;117:3198-210.

455   [26]Bisping E, Ikeda S, Kong SW, Tarnavski O, Bodyak N, McMullen JR, Rajagopal

456   S, Son JK, Ma Q, Springer Z, Kang PM, Izumo S, Pu WT. Gata4 is required for

457   maintenance of postnatal cardiac function and protection from pressure

458   overload-induced heart failure. Proc Natl Acad Sci U S A. 2006;103:14471-6.

459   [27]Liang Q, De Windt LJ, Witt SA, Kimball TR, Markham BE, Molkentin JD. The

460   transcription factors GATA4 and GATA6 regulate cardiomyocyte hypertrophy in

461   vitro and in vivo. J Biol Chem. 2001;276:30245-53.

462   **Figure legends**

463   **Figure 1. Pedigree of the nuclear and extended family showing affected and**

464   **unaffected members with ASD.**

465   Subjects I1 and II1 died naturally. Subject I2 was a suspected positive. Subjects II2,

466   IV3, and IV4 were affected patients. Subject III2 was a carrier. Subject IV1,

467   miscarried at the 8th week of pregnancy. Subject IV2, mid-trimester aborted for fetal

468   bradycardia. The arrow indicates the proband patient. '?' indicates the suspected

469   patient. The oblique line represents a deceased subject.

470

471   **Figure 2. Heterozygous missense variation of *GATA4* in the family and the**

472   **conservation analysis of the variant site.**

473   (A): The arrow indicates the heterozygous nucleotides of C/T. Subject III-1 shows the

474   normal individual. Subjects IV-3, IV-4, III-2 and II-2 show the missense variant. The

475   rectangle denotes the nucleotides comprising a codon of *GATA4*. (B) Alignment of

476   multiple GATA4 amino acid sequences across species. The altered arginine at amino

477   acid position 320 (p.R320) of GATA4 is completely conserved evolutionarily among

478   various species. The yellow column shows the R320 site. The fluorescent blue shows

479   unconserved sites in different species around the R320 site. (C) Schematic diagram of

480   GATA4 protein. TAD1: transcription activation domain 1 (amino acid 1-74); TAD2:

481   transcription activation domain 2 (aa 130–177); ZF1: N-terminal zinc finger (aa 217–

482   241); ZF2: C-terminal zinc finger (aa 271–295); NLS: nuclear localization signal

483   region (aa 307–325). Red bar shows NLS and the red letters are mutant amino acids

484   (R320>W).

485

486   **Figure 3. The structural difference between wild type and mutant *GATA4*.**

487   (A and B): wild type; (C and D): mutant GATA4. Colorful cartoon depicting the

488   protein of GATA4. Red indicates the α-Helix while yellow indicates the β-Sheet.

489   Green shows a random coil. Sphere and stick of blue shows the variant site. Wheat

490   shows the related residues. Dotted line in (D) shows the new hydrogen bonds. Blue

491   dotted line in (A) and (C) shows the angle of two α-helices.

492

493   **Figure 4. Diminished transcriptional activity of *GATA4* caused by the variation.**

494   (A) 293T cells were transfected with 100 ng of wild type, mutant GATA4 or pcDNA3

495   plasmids and 400 ng of BNP luciferase reporter. The result showed the transcriptional

496   activity of mutant *GATA4* significantly decreased in comparison to wild type. NC:

497   negative control; WT: wild type; MUT: mutant. Data is displayed as mean ± SD; NC

498   was set as 1.0; * represent a statistically significant difference with $p < 0.01$.

499   (B) The protein levels of wild type and mutant GATA4 were evaluated by western

500   blot analysis after 293T cells were transfected with 100 ng of either type of plasmid.

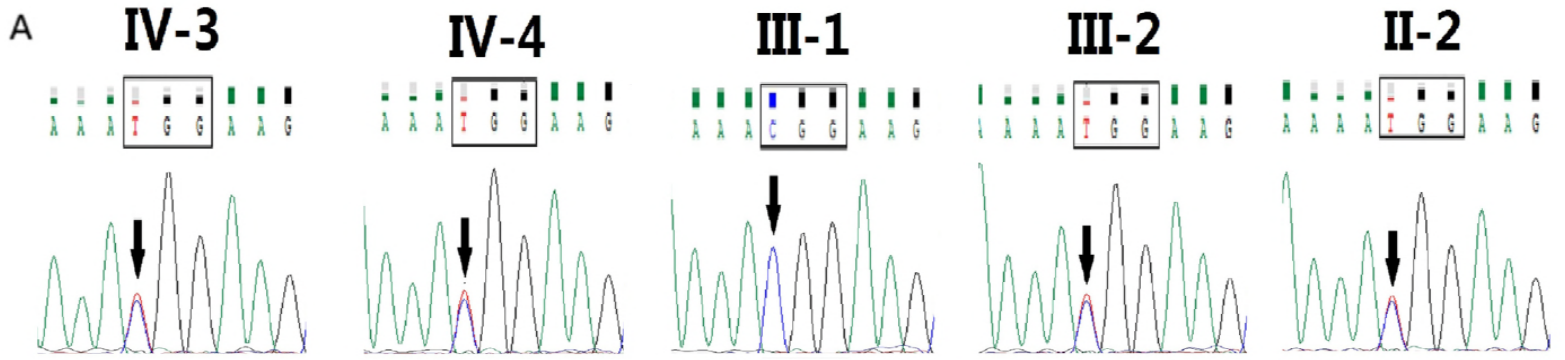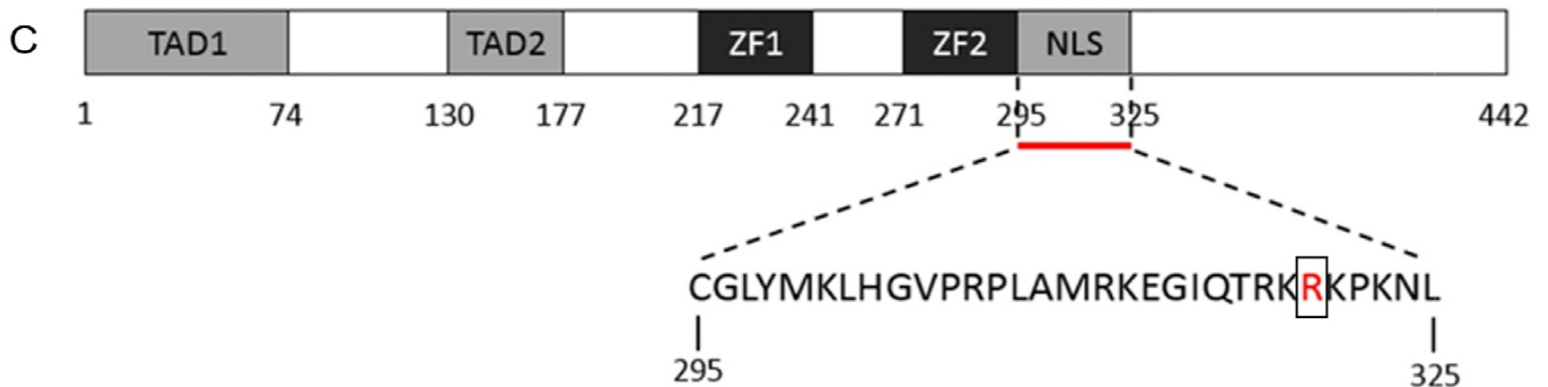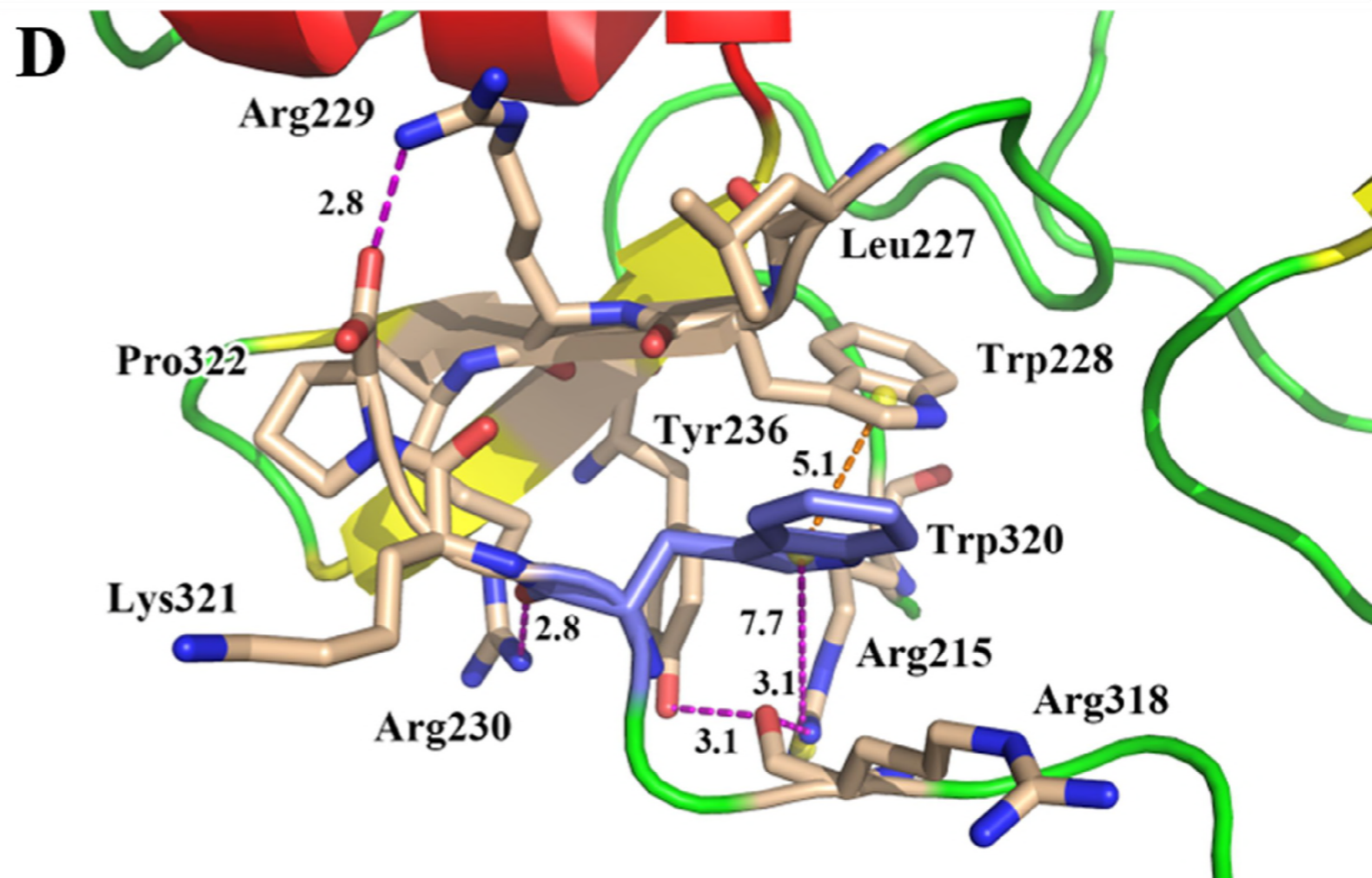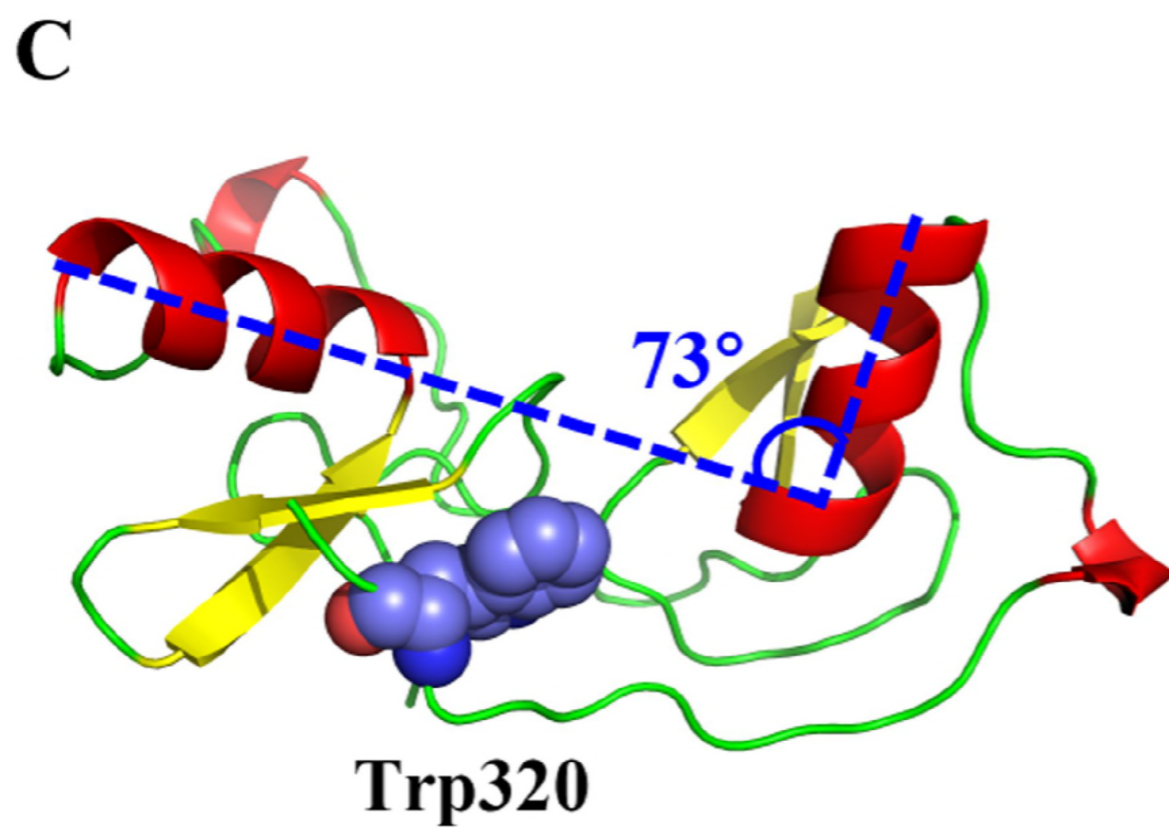501   No significantly differences were observed between the two groups.

A

IV-3   IV-4   III-1   III-2   II-2
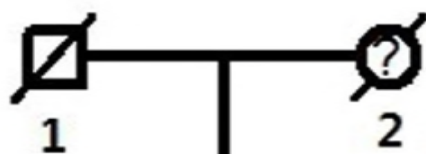
B

Homo Sapiens_GATA4	307	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	325
Pan troglogdytes_GATA4	273	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	291
Macaca mulatta_GATA4	308	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	326
Canis lupus familiaris_GATA4	307	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	325
Bos taurus_GATA4	307	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	325
Mus musculus_GATA4	306	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	324
Rattus norvegicus_GATA4	306	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	324
Gallus gallus_gata4	277	L	A	M	R	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	295
Danio rerio_GATA4	260	L	A	M	K	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	278
Xenopus tropicalis_GATA4	276	L	A	M	K	K	E	G	I	Q	T	R	K	R	K	P	K	N	L	294
Drosophila melanogaster_GATA4	262	L	A	M	R	K	D	G	I	Q	T	R	K	R	K	P	K	K	T		280

C

TAD1		TAD2		ZF1		ZF2	NLS		
1	74	130	177	217	241	271	295	325	442

CGLYMKLHGVPRPLAMRKEGIQTRK R KPKNL

295   325

A

BNP

B

GATA4

Actin

NC   WT   MUT