A novel homozygous *TPM1* mutation in familial pediatric hypertrophic cardiomyopathy and *in silico* screening of potential targeting drugs

S.J. CARLUS[1], I.S. ALMUZAINI[2], M. KARTHIKEYAN[3], L. LOGANATHAN[3], G.S. AL-HARBI[1], F.H. CARLUS[4], A.H. AL-MAZROEA[1], M.M. MORSY[5], H.M. ABO-HADED[1,6], A.M. ABDALLAH[7], K.M. AL-HARBI[1]

[1]Cardiogenetics Unit, Pediatrics Department, College of Medicine, Taibah University, Al-Madinah, Kingdom of Saudi Arabia
[2]Department of Pediatric Cardiology, Al-Madinah Maternity and Children Hospital (MMCH), Al-Madinah, Kingdom of Saudi Arabia
[3]Department of Bioinformatics, Alagappa University, Karaikudi, Tamil Nadu, India
[4]Department of Zoology, Pachaiyappa's College Chennai-30, India
[5]Madina Cardiac Center, Al-Madinah, Kingdom of Saudi Arabia
[6]Pediatric Cardiology Unit, Department of Pediatrics, Faculty of Medicine, Mansoura University, Egypt
[7]College of Health Sciences, QU Health, Qatar University, Doha, Qatar

Abstract. – OBJECTIVE: Familial hypertrophic cardiomyopathy (HCM) is the most common genetic cardiac disease. While sarcomeric gene mutations explain many HCM cases, the genetic basis of about half of HCM cases remains elusive. Here we aimed to identify the gene causing HCM in a non-consanguineous Saudi Arabian family with affected family members and a history of sudden death. The impact of the identified mutation on protein structure and potential drug targets were evaluated *in silico*.

MATERIALS AND METHODS: Triplets (two HCM subjects and one patent ductus arteriosus (PDA) case) and unaffected parents were screened by targeted next-generation sequencing (NGS) for 181 candidate cardiomyopathy genes. *In silico* structural and functional analyses, including protein modeling, structure prediction, drug screening, drug binding, and dynamic simulations were performed to explore the potential pathogenicity of the variant and to identify candidate drugs.

RESULTS: A homozygous missense mutation in exon 1 of *TMP1* (assembly GRCh37-chr15: 63340781; G>A) was identified in the triplets [two HCM and one patent ductus arteriosus (PDA)] that substituted glycine for arginine at codon 3 (p.Gly3Arg). The parents were heterozygous for the variant. The mutation was predicted to cause a significant and deleterious change in the TPM1 protein structure that slightly affected drug binding, stability, and conformation. In addition, we identified several putative TPM1-targeting drugs through structure-based *in silico* screening.

CONCLUSIONS: *TPM1* mutations are a common cause of HCM and other congenital heart defects. To date, *TPM1* has not been associated with isolated PDA; to our knowledge, this is the first report of the homozygous missense variation p.Gly3Arg in *TPM1* associated with familial autosomal recessive pediatric HCM and PDA. The identified candidate *TPM1* inhibitors warrant further prospective investigation.

Key Words:
Pediatric familial hypertrophic cardiomyopathy, Targeted Gene sequencing, TPM1, Saudi Arabia, Consanguinity, Molecular docking, Molecular dynamics.

Introduction

Hypertrophic cardiomyopathy (HCM) is an inherited cardiac disease characterized by left ventricular hypertrophy that is associated with a number of potential clinical outcomes, including impaired diastolic function, heart failure, and sudden cardiac death (SCD). HCM is rare in the pediatric population but affects over 1 in 500 of the general adult population[1]. It is also one of the most common causes of SCD in young athletes[2,3].

HCM has mainly been considered an autosomal dominant disease, although some cases can be explained by *de novo* mutations and, less

Corresponding Author: S. Justin Carlus, Ph.D; e-mail: justincarlus@gmail.com

commonly, autosomal recessive inheritance. Pathogenic mutations are usually detected in eight genes encoding sarcomeric proteins, which generate the molecular force of myocyte contraction, with 50% of mutations occurring in cardiac myosin-binding protein C (*MYBPC3*) and beta myosin heavy chain (*MYH7*). Other HCM genes include cardiac troponin T2 (*TNNT2*), cardiac troponin I (*TNNI*3), alpha tropomyosin (*TPM1*), myosin regulatory light chain (*MYL2*), myosin essential light chain (*MYL3*), and cardiac alpha actin (*ACTC1*), which harbor a much lower frequency of pathogenic variants (1-5% each)[4]. Overall, mutations in over 70 genes have been reported to cause HCM, accounting for 50-60% of affected individuals but leaving the remainder without a known genetic basis[5]. TPM1 is a highly conserved actin-binding protein belonging to the tropomyosin family. The main function of the protein is as part of the troponin complex, regulating the calcium-dependent interaction of actin and myosin during muscle contraction. Specifically, targeting the myofilament molecules involved in muscle contraction could positively regulate Ca^{2+}-based signaling pathways, so TPM1 could be a drug target for the prevention and treatment of HCM[6].

Our surveys of the registry of a pediatric cardiology clinic at the Madinah Maternity and Children Hospital (MMCH), Al-Madinah, Saudi Arabia revealed a very high prevalence of cardiomyopathy in the pediatric heart patient population[7]. Saudi Arabia has one of the highest rates of consanguinity in the world, as marriages between first cousins and close relatives are widely accepted and frequently occur[8]. Consanguinity may be a risk factor for some congenital abnormalities[9]. Unfortunately, few studies have been conducted on Saudi Arabian cardiomyopathy patients, and the disease remains poorly understood in this population.

Here we aimed to identify novel mutation(s) and gene(s) responsible for familial HCM using targeted next-generation sequencing (NGS) technology in a Saudi Arabian family with affected family members. In doing so, we identified a novel homozygous missense mutation in *TMP1* in affected individuals. To provide functional insights, *in silico* analysis was used to predict the deleterious effect of the mutation and to understand its impact on protein structure and drug binding. Potential *TPM1* inhibitors were identified through *in silico* screening of a natural compound database.

Materials and Methods

Ethical Statement

This study was conducted fully in accordance with the ethical standards of the Taibah University Ethical Research Committee, the Ethical Review Board of Madinah Maternity and Children Hospital (MMCH), and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Study Participants and Clinical Evaluation

Peripheral blood samples were collected from five individuals: triplets [two HCM subjects and one patent ductus arteriosus (PDA) case] and unaffected normal parents from a non-consanguineous Saudi Arabian family attending the Department of Pediatric Cardiology, MMCH. The parents provided written informed consent for study participation. HCM was diagnosed according to published guidelines[10,11] using the family history, physical and clinical examination, electrocardiography (ECG), chest x-rays, and 2D-echocardiography. Doppler echocardiography was performed using a Hewlett-Packard 1500 or 5500 echocardiographic system. In adults, HCM is usually defined by a maximal LV wall thickness ≥15 mm on echocardiography, with a wall thickness of 13 to 14 mm considered borderline, particularly in the presence of other compelling information (e.g., family history of HCM). In children, an increased LV wall thickness is defined as a wall thickness ≥2 standard deviations above the mean (z-score ≥2) for age, sex, or body size. In familial cases, the diagnosis can be made using echocardiographic criteria as above and/or one of the following ECG abnormalities: LV hypertrophy (Romhilt-Estes score >4); Q-waves (duration >0.04 s and/or a depth >1/4 of ensuring R wave in at least two leads); or marked repolarization abnormalities (T wave inversion in at least two leads).

Cardiomyopathy Gene Panel, Targeted Gene Sequencing, and Bioinformatics Analysis

Targeted sequencing of 181 cardiomyopathy disease genes was performed on the five samples. The development of an in-house custom gene panel covering 181 cardiomyopathy genes, DNA extraction, targeted sequencing, and bioinformatics analysis are described in our previous publication[7].

Validation by Sanger Sequencing

Sanger sequencing was performed to confirm the presence of the variant. Primer sequences were designed using Primer3. The primer pairs used for PCR were forward: 5'-ACTTCCG-GACTGCTCCT-3' and reverse: 5'-GTGATGG-GTGTATCCCTTACG-3'. The amplicons were directly sequenced using BigDye chain termination chemistry on an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA, USA).

Sequence and Structure Data Collection

The *TPM1* gene was targeted to assess the identified variant's role on the protein structure. A full-length protein crystal structure was not available; therefore, a molecular modeling approach was used to build a new tertiary structure. The protein sequence was collected from the UniProtKB database (P09493) and the tertiary structure predicted using several computational tools, including SWISS-MODEL, Robetta, and I-TASSER[12]. After typical structure validations, a significant model was selected for further analysis. All *in silico* analyses were carried out on the CentOS v7.0 Linux platform with an Intel® Core™ i7-4770 CPU@3.40 GHz processor using the Schrödinger platform (Schrödinger, New York, NY, USA; 2018-4 package)[13].

Sequence Alignment and Secondary Structure Prediction

The protein sequence was used for sequence alignment using Geneious Pro software (Auckland, New Zealand). The mutant TPM1 protein sequences were prepared by manually editing the FASTA file. Both wild and mutant-type protein sequences were aligned by MAFFT alignment v.7.017 with the Blosum62 matrix[14]. The aligned sequence was further used for secondary structure prediction, with the results visualized to find secondary structure differences, namely a loss of loops and addition of a helix around the mutation site (Figure 1).

Protein Structure Modeling

The best TPM1 protein structure was predicted using Robetta[15], the mutation lying in the third residue and therefore not predicted to significantly impact the structure. Molecular modeling did not reveal any major difference in the early section of the protein (Figure 2A). Hence, an *ab initio* modeling protocol was utilized to model the TPM1 protein.

Molecular Dynamics Simulation

The TPM1 protein was subjected to molecular dynamics simulation using Desmond v2.3, (Schrödinger)[16], an approach that helps to identify structural changes in wildtype and mutant proteins. The system was also utilized to analyze the root mean square deviation (RMSD) and the root mean square fluctuation (RMSF). The wildtype and mutant protein structures were processed separately for molecular dynamics simulation to observe the effect of the mutation on the protein structure, hypothesizing that there would be no significant difference between the wild and mutant-type protein structure such that these proteins may be considered as drug targets due to their conserved nature. The protein structure was considered as a system with ions added in the single point charge (SPC) solvent model. An orthorhombic box was placed around the molecule to add solvent and ions. Approximately 22 sodium ions at 0.15 M concentration were added to neutralize the whole system. The NVT ensemble was set with the thermostat using the Nose-Hoover chain algorithm at 300 K and 1 bar pressure at

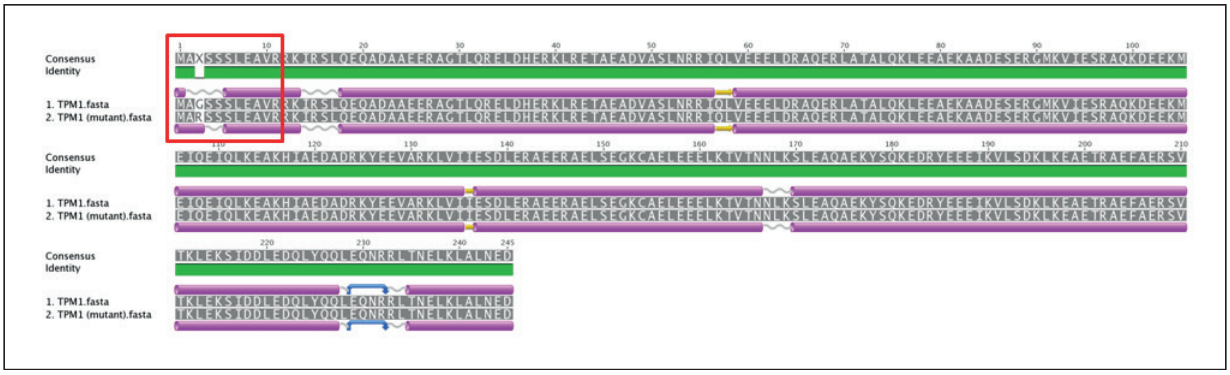

Figure 1. Secondary structure prediction of wild type and mutated TPM1 protein.

1.0 ps and 2.0 ps relaxation time, respectively. Finally, the whole system was submitted for a 20 ns simulation. After completion of the simulation, the whole trajectory file was utilized to analyze the RMSD and RMSF with respect to the initial structure.

Molecular Docking Studies

The predicted tertiary structure of TPM1 was used for molecular docking studies using a virtual screening workflow. To identify potential TPM1 inhibitors, we selected a natural small molecule database, Life Chemicals[17]. The 2D structures of the screened compounds were retrieved from the Life Chemicals web server. The potential activities of the screening compounds were compared with known drugs listed in the Human Gene Database[18].

Protein Preparation

The predicted tertiary structure of TPM1 was prepared using the Protein Preparation Wizard in Maestro (Schrödinger). In the preprocessing step, hydrogens, bond length, missing side chains, and loops were added. Hydrogen bonds were optimized, and the final structure was minimized. In order to obtain a proper structure, the obtained protein was used for restrained energy minimization with an RMSD cutoff of 0.3 nm using OPLS3e[19]. The minimized protein structure of TPM1 was then used in downstream analyses.

Active Site Prediction And Grid Generation

The refined protein structure was used for active site prediction using Sitemap v4.9 (Schrödinger). The active site of the target protein consists of only a single coiled-coil domain. A hydrophobic rich region is essential for molecular docking analysis. Therefore, we focused on identifying more hydrophobic sites and hydrogen bond acceptor-rich and donor-rich sites, which better define the active site of the protein. Of the five top ranked sites in the protein, the best site was selected based on the site score and the area coverage (Figure 2B). The identified best site was used to define the grid box around the active site of the protein using the Receptor Grid Generation Panel in Glide (Schrödinger).

Ligand Preparation

The database compounds and standard drug molecules were used for ligand preparation. All 2D structure files were imported for conversion to 3D structures with generated conformers[20]. The ligands were prepared following the standard protocol. Approximately 288,651 compounds were used for 3D conversion and minimization of chemical compounds using Ligprep in Maestro (Schrödinger). Conformers were generated using a rapid torsion angle search approach followed by minimization of each generated structure using the OPLS-3e force field, with 30 implicit GB/SA solvent models[21]. Thirty-two conformers were generated for all the compounds based on the rotatable bonds present in the ligand molecules.

Virtual Screening Workflow

A structure-based virtual screening workflow was used to screen the large chemical compound database to identify possible lead molecules. The compounds in the Life Chemicals database (288,651) were screened using the active site of the TPM1 complex using the Glide-VSW module. The virtual screening workflow is a well-defined protocol that filters chemical compounds by screening and docking. Initially, the QikProp filter was applied with Lipinski's Rule, followed by the high-throughput virtual screening (HTVS), standard precision (SP),

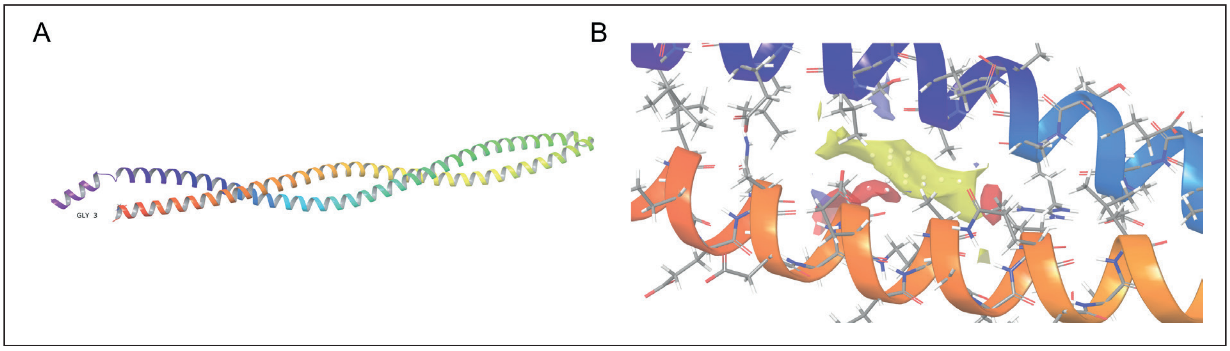

Figure 2. A, Modeled tertiary structure of the TPM1 protein with labeled mutation site. **B**, The predicted active site consists of hydrophobic surfaces and hydrogen bond acceptor and donor-rich regions in the TPM1 protein.

and finally extra precision (XP) docking programs to screen potential ligands. Lipinski's rule in QikProp helps to eliminate non-drug-like compounds[22]. Glide HTVS and SP docking use a series of hierarchical filters to find the best possible ligand-binding locations in the defined receptor grid space. Further, the poses generated by Glide SP were again refined by Glide XP[23]. At each stage of HTVS, SP, and XP, the top 10% compounds were retained for the next stage[24]. In this way, database compounds were screened individually with the predicted protein structure and the results were scrutinized to find potential inhibitors. Top-hit compounds were ranked according to the Glide XP score, Glide energy, and interacting protein residues with the compound. The Glide XP score represents the total g-score or final score to identify the best ligand. Best pose identification was calculated through the non-bonded interactions, while the Glide energy was a modified Coulomb-van der Waals interaction energy score. The Glide energy is useful for comparing the binding affinities of different ligands. The top ten ligands from the Life Chemicals database were further used in downstream analyses.

Binding Energy Calculation

Binding energy calculations were conducted for the top ten protein-ligand complexes and with standard drug molecules using the Prime/MM-GBSA approach[25]. MM-GBSA is a post scoring approach that helps to evaluate molecular docking and validate the accuracy of different compounds. This method is very helpful for predicting the binding energy of different sets of ligand poses to the receptor. The following equations were used to calculate the binding energy:

$$\Delta Gbind = \Delta E + \Delta Gsolv + \Delta GSA \ldots\ldots\ldots (1)$$

$$\Delta E = Ecomplex - Eprotein - Eligand \ldots\ldots (2)$$

where Ecomplex, Eprotein, and Eligand are the minimized energies of the protein-inhibitor complex, protein, and inhibitor, respectively; ΔGsolv is the generalized Born electrostatic solvation energy of the complex; and ΔGSA is a non-polar contribution to the solvation energy due to the surface area[26].

Absorption, Distribution, Metabolism, Excretion (ADME) Prediction

Absorption, distribution, metabolism, excretion (ADME) is an essential approach for analyzing the physicochemical properties of drugs. Early data collection on ADME[27] properties provides useful insights into drug formulations and clinical suitability. The identified best lead compounds were studied for their ADME properties using the QikProp module (Schrödinger). The required principle and physiochemical properties of drug compounds were predicted from the defined dataset and the acceptability of the known and screened compounds was evaluated based on Lipinski's rule of five[22].

Results

Clinical Assessment of the HCM Family

The triplets were born of non-consanguineous Saudi parents. The father was 36 years old and the mother was 32 years old. The father had no significant medical history and the mother had type 1 diabetes mellitus. Both parents had normal echocardiograms. There was a family history of sudden cardiac death on the mother's side; two cousins died aged 20 and 27 years from HCM. The family pedigree was highly suggestive of a recessive pattern of inheritance (Figure 3A). The triplets were six years old and born preterm (31 weeks' gestation).

At birth, subject 1 weighed 1.5 kg and was admitted to the neonatal intensive care unit (NICU) and ventilated due to low birth weight, respiratory distress, and need for respiratory support because of frequent apnea. Echocardiography revealed left ventricular hypertrophy, no left ventricular outflow tract (LVOT) obstruction, and normal left ventricular (LV) systolic functions. The primary diagnosis was HCM with disease onset since birth. Other clinical symptoms were seizures, brain hemorrhage, cerebral palsy, and bronchial asthma.

Subject 2 weighed 1.3 kg at birth and was admitted to the NICU and ventilated due to low birth weight, respiratory distress, and for respiratory support because of frequent apnea. Echocardiography revealed a tiny patent ductus arteriosus (PDA) but no manifestation of HCM. A brain CT was normal. Other clinical symptoms were bronchial asthma and diabetic ketoacidosis, which developed at three years of age.

Subject 3 weighed 900 g at birth and was admitted to the NICU and ventilated due to low birth weight, respiratory distress, and need for respiratory support because of frequent apnea. Echocardiography revealed mild LV septal hypertrophy but without LVOT obstruction. Brain CT was normal.

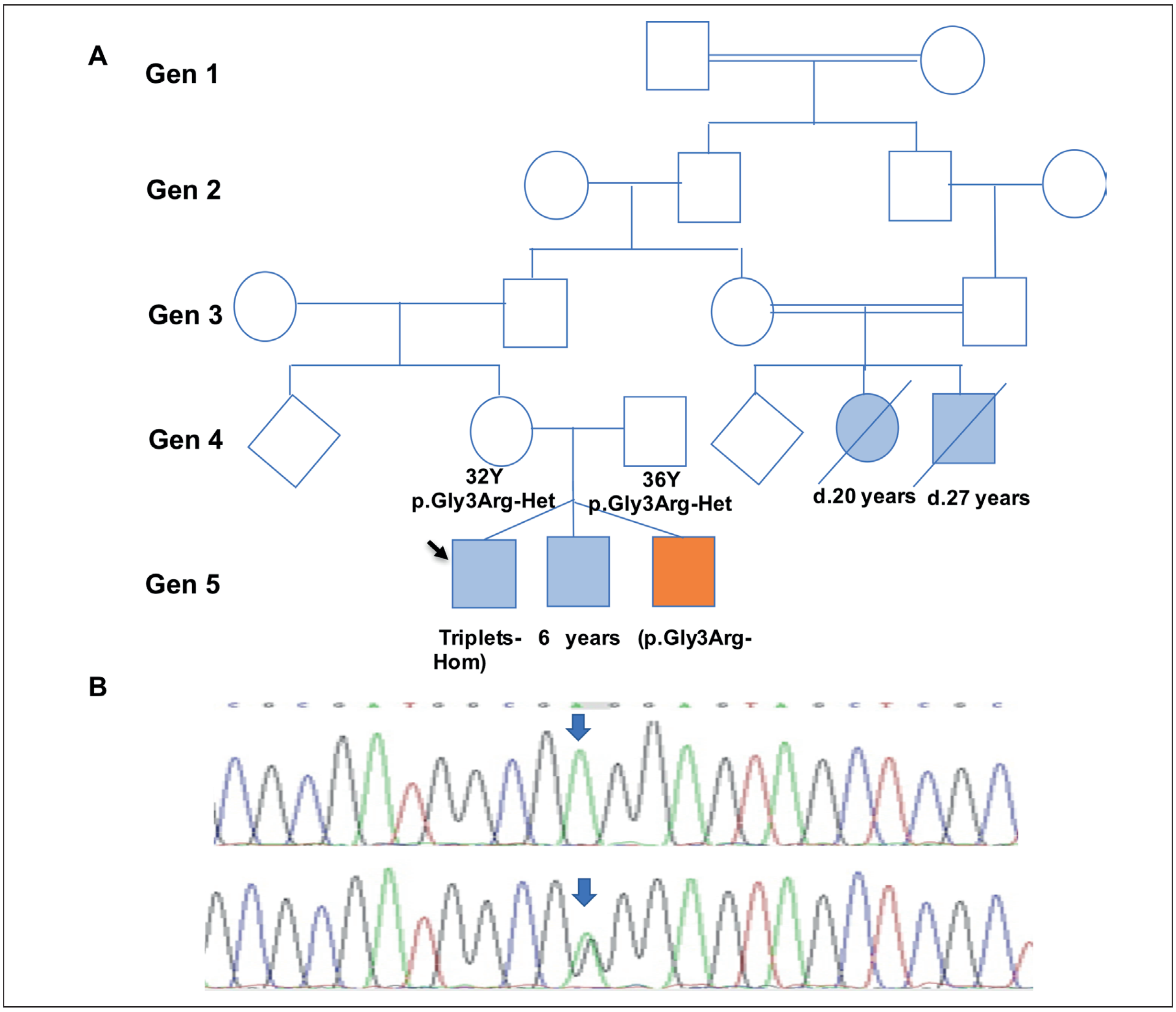

Figure 3. A, Family pedigree, with proband indicated with an arrow. Males are depicted as squares, females as circles, unknown gender depicted as diamonds, and deceased individuals with a diagonal line. Blue and brown shading represent individuals who are clinically affected triplets (two HCM subjects and one PDA case), while no shading indicates unaffected individuals. Het indicates a heterozygous TPM1 p.Gly3Arg mutation and Hom indicates the homozygous TPM1 p.Gly3Arg mutation. **B**, Homozygous missense mutation in exon 1 of TPM1 (chr15: 63340781; G>A) resulting in the substitution of glycine for arginine at codon 3 (p.Gly3Arg) in affected HCM and PDA subjects. The mutation was present in the heterozygous state in unaffected parents.

Genetic Analysis and Systematic Prioritization of Candidate Variants

Targeted sequencing of 181 cardiomyopathy disease genes was performed on the triplets (two HCM subjects and one PDA case) and their parents' samples. A mean of 0.6 Gb of sequence was generated per individual, of which 94.2% of bases had ≥Q30 quality and over 98% of reads aligned to the human genome. Bioinformatics analysis detected a homozygous missense mutation in exon 1 of the *TMP1* gene (on assembly GRCh37-chr15: 63340781; G>A; ENSG00000140416; rs397516490) in all triplets (HCM subjects and one PDA case), which resulted in a glycine for arginine substitution at codon 3 (p.Gly3Arg) (Figure 3B). The variant was heterozygous in the parental samples.

This variant has not been described in the ExAC Aggregation Consortium (ExAC) database (>60,000 samples), the 1000 Genomes (1KG) database (2,500 samples), the Exome Sequencing Project (ESP) database (6,500 samples), or in the Atlas of Cardiac Genetic Variation (https://www.cardiodb.org/acgv/acgv_variant.php). However, while the variant has been reported in ClinVar (RCV000036633.3), its pathogenicity has not been clinically demonstrated. The *in silico* variant was predicted to be "probably damaging" by PolyPhen-2, "damaging" by MetaLR, "damaging" by SIFT (low confidence), "disease causing" by MutationTaster2, and "pathogenic" by REVEL[28].

Protein Structure Prediction

The tertiary protein structure of TPM1 was modeled on the Robetta server. The domain, predicted by the Ginzu domain prediction algorithm in Robetta, consisted of 1-245 amino acids with a confidence score of 0.997. The domain has a predicted coiled-coil structure with two polypeptide chains. The predicted structure was validated with the Procheck tool[29], and the Ramachandran plot showed that 99.6% of residues were in favored/allowed regions and 0.4% were in disallowed regions (Figure 4A). The amino acids located in the disallowed regions were optimized during energy minimization of the modeled protein.

Molecular Dynamics Simulation

The structural stability of the predicted model of TPM1 was analyzed using Desmond software. The molecular dynamics simulation revealed that the predicted model was stable and compact. A 20 ns simulation run was used to plot RMSD and RMSF data generated through the trajectory files. The RMSD plot showed that the modeled protein was stable with an RMSD value of 6-5 Å. The mutated protein deviated little from the wild type protein structure (~1 Å) (Figure 4B). The RMSF plot helps to visualize the relative changes in residue fluctuation, with a large fluctuation observed at the start and end residues of the protein; the fluctuation was higher at the end of the protein. The active site residues Gln 20, Arg27, Arg202, Phe205, Ser209, and Ser216 were considered as interacting residues and have a leading role in ligand binding. The RMSF plot clearly shows that fluctuation of these residues was under 3.0 Å (Figure 4C), with the wild type protein having slightly higher fluctuations than the mutant-type structure. Finally, the molecular dynamics study confirmed that the structural stability and conformational changes persisted throughout the simulation period.

Structure-Based Virtual Drug Screening

A virtual drug screening approach can help to identify potential inhibitors of wildtype and mutant protein structures. The refined TPM1 protein structure with a defined active site was used as the receptor to dock database compounds. 228 compounds were identified as top hit molecules from virtual screening. Of these, we selected the top ten compounds with good Glide gscores and non-covalent interactions, which were then visualized using both 2D (Figure 5A and B) and 3D (Figure 6) binding poses in the active site. Notably, the top five Life Chemical compounds (F0760-0534, F1899-0458, F1691-3363, F1278-0497, and F2258-0953) had Glide gscores between -5.381 and -4.677 kcal/mol with Glide energies ranging from -44.991 to -31.682 kcal/mol. From these results, we propose that these five small molecules may act as a potential TPM1 inhibitors. Furthermore, the observed binding energy of the top hit molecule was -58.57 to -39.30 kcal/mol, denoting strong binding affinity towards TPM1. Several hydrogen bond interactions were observed between the ligand and protein. Specifically, Arg27, Gln20, Ser209, and Arg202 formed hydrogen

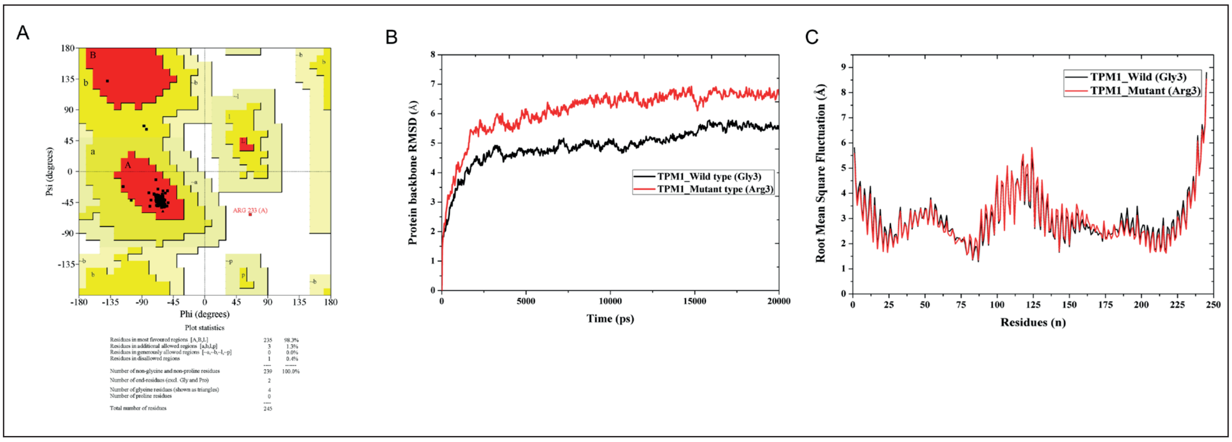

Figure 4. *A*, The Ramachandran plot values for the predicted protein structure model of TPM1. **B**, The RMSD of the modeled TPM1 protein from a 20 ns simulation. **C**, RMSF plot of the wild and mutant structure model of TPM1 from a 20 ns simulation.

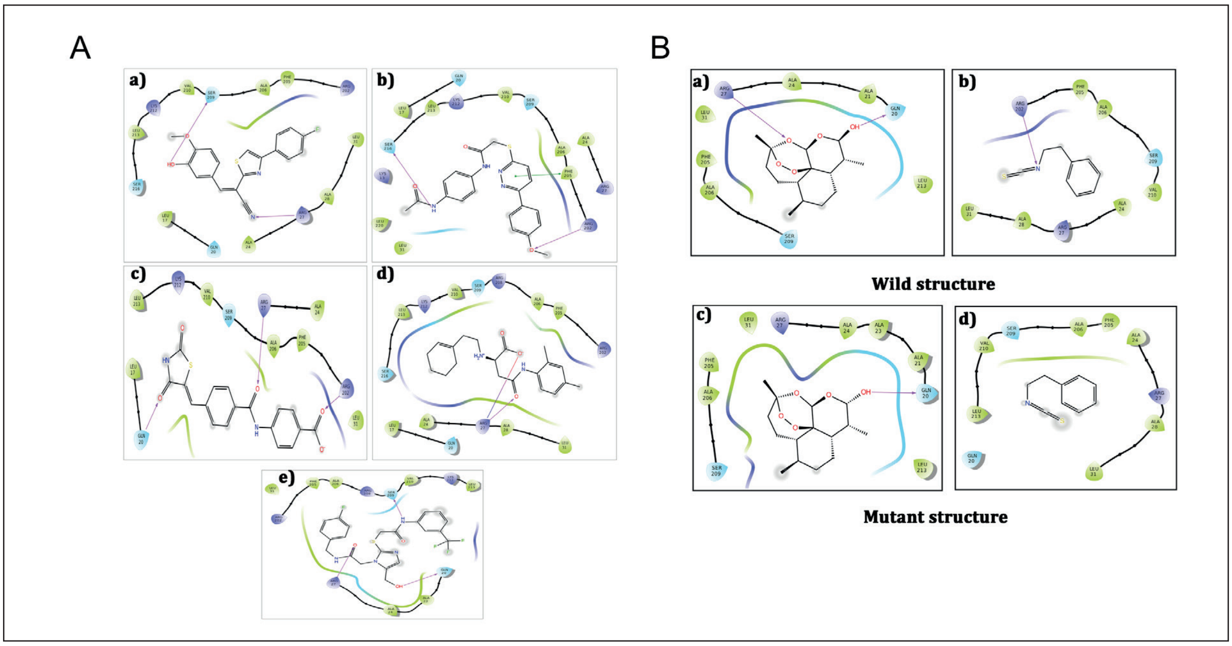

Figure 5. A, The 2D interaction pattern of the top five hit compounds. a) F0760-0534, b) F1899-0458, c) F1691-3363, d) F1278-0497, and e) F2258-0953. **B**, The 2D interaction pattern of the standard drugs with TPM1. a, c) CID-6918483; b, d) CID-16741.

bonds with the ligand molecules. The 2D and 3D interactions clearly show the binding poses of the top five compounds (Figures 5 and 6).

Similar activity and binding affinity were observed in the mutated protein structure. The Glide gscores ranged from -5.354 to -3.471 kcal/mol, and the Glide energy and binding energy ranged from -42.391 to -31.471 kcal/mol and -56.52 to -39.87 kcal/mol, respectively. The glide gscores were very similar for the top four compounds, but compound F2258-0953 had a very low Glide gscore and binding energy. Similarly, the binding affinities of the mutated protein were similar to wildtype. Therefore, the docking scores and in-

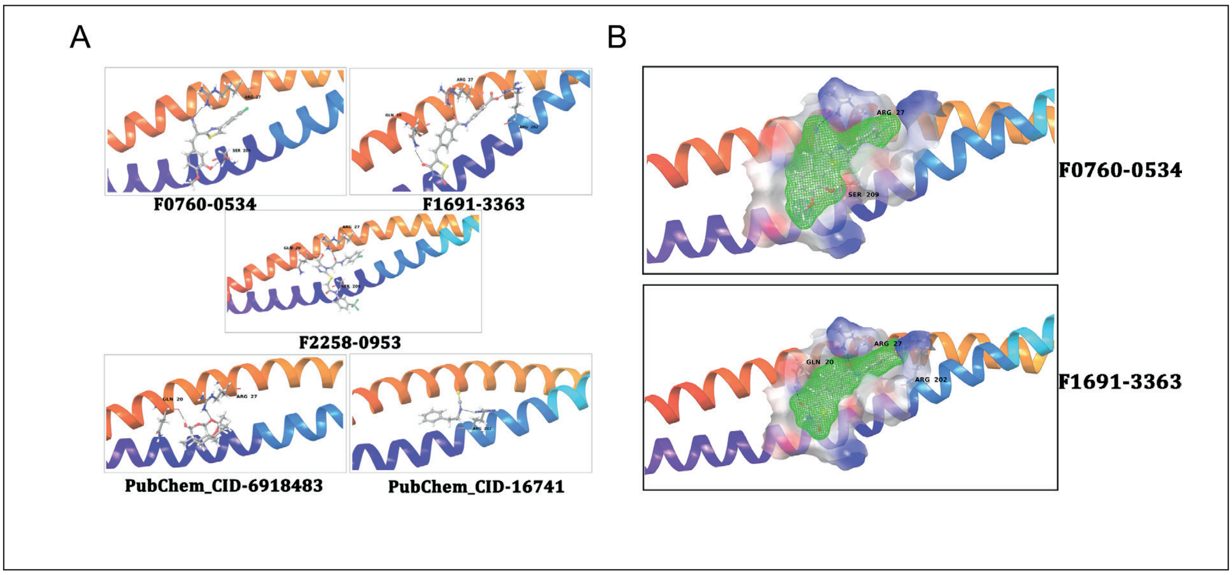

Figure 6. A, The 3D interaction pattern and binding poses of the screened compounds and standard drugs in the TPM1 active site. **B**, The electrostatic surface of the top hit Life Chemicals compounds (green surface) in the active site of TPM1.

teractions between compounds and wildtype and mutant structures were similar, suggesting that the Gly3Arg mutation does not significantly affect drug binding and affinity towards the TPM1 protein active site.

Evaluation of "Standard" Drugs

The "standard" drugs predicted to target TPM1, namely dihydroartemisinin (PubChem ID: 6918483), and phenethyl isothiocyanate (PubChem ID: 16741) were docked individually to both wildtype and mutated TPM1 proteins. These drugs are predicted to have activity against TPM1 based on the evidence provided from the Human Genome Database and, as expected, the molecular docking studies showed that the standard drugs had good binding affinity towards the TPM1 active site. The Glide gscores were comparatively lower than the compounds identified from screening (Tables I and II) at -3.169 and -2.183 kcal/mol in wildtype and -3.300 and -2.372 kcal/mol in mutant proteins for dihydroartemisinin and phenethyl isothiocyanate, respectively. There were significant differences in Glide gscore, glide energy, binding energy, and number of interacting residues between the two standard drugs and top hit compounds. The standard drugs only formed three hydrogen bonds with Gln20, Arg27, and Arg202. Figure 6 shows the binding poses of the top compounds with the compounds highlighted inside the green surfaces and the binding site indicated by electrostatic surfaces. The binding site is in a hydrophobic cavity with electropositive (blue) and electronegative regions (red).

Pharmacokinetic Properties

The ADME or pharmacokinetic properties of the identified compounds were estimated using QikProp. All five top hit compounds were within an acceptable range (shown in parentheses below) and could be identified as drug-like compounds. The molecular weight of the compounds ranged from 344.4 to 496.4 g/mol (130 to 725 g/mol). The log *Kp* is a useful assessment of potential compound penetration through skin, and the predicted skin permeability QPlogKp ranged from -1.629 to -5.155 (-8.0 to -1.0). The predicted aqueous solubility QPlogS was between -4.169 to -6.424 (-6.5 to 0.5), and the predicted perceptible Caco-2 cell permeability of the compounds was 45.728 to 652.512 nm/sec (<25 poor permeability; >500 high permeability). The predicted IC_{50} values for blockage of HERG K^+ channels were -4.0 to -6.4 (<-5 is a "good" value). The predicted blood/brain partition coefficients were -2.4 to 0.8 (-3.0 to 1.2). The percentage human oral absorption was predicted for the identified compounds using a quantitative multiple linear regression model. All compounds had >50% oral absorption (range 50-100%, where <25% is low and >80% is high) (Table III). Further, we assessed whether the candidate compounds violated the "rule of five" prediction, and they did not violate any rule (mol_MW < 500, QPlogPo/w < 5, donorHB ≤ 5, accptHB ≤ 10). Compounds that satisfy the rule of five can be considered drug-like compounds[30], so these compounds could be potential inhibitors of TPM1.

Table I. Molecular docking results for the identified top hit and standard compounds for the wildtype protein.

Compound ID	Glide gscore (kcal/mol)	Glide energy (kcal/mol)	ΔG bind energy (kcal/mol)	Interacting residues
F0760-0534	-5.381	-44.171	-50.42	Arg27, Ser209
F1691-3363	-5.229	-41.199	-41.50	Gln20, Arg27, Arg202
F2258-0953	-5.132	-40.200	-40.05	Gln20, Arg27, Ser209
F1278-0497	-4.856	-31.682	-47.64	Arg27
F1899-0458	-4.770	-44.014	-58.57	Arg202, Phe205, Ser216
F1089-0033	-4.764	-38.405	-43.10	Phe205
F2096-0085	-4.798	-44.991	-52.91	Ser209, Lys212
F2256-0264	-4.711	-37.176	-46.26	Gln20, Arg202
F1847-0028	-4.678	-35.794	-45.18	Arg27, Ser209
F2539-0318	-4.677	-34.193	-39.30	Arg27, Arg202
PubChem_CID-6918483	-3.169	-19.030	-29.26	Gln20, Arg27
PubChem_CID-16741	-2.183	-18.207	-24.65	Arg202

Table II. Molecular docking results for the identified top hit leads with mutated TPM1.

Compound ID	Glide gscore (kcal/mol)	Glide energy (kcal/mol)	ΔG bind energy (kcal/mol)	Interacting residues
F0760-0534	-5.354	-33.900	-51.48	Arg27, Ser209
F1691-3363	-5.074	-42.391	-39.91	Arg202, Ser216
F2258-0953	-4.443	-36.600	-39.87	Arg27, Arg202
F1278-0497	-4.379	-31.248	-47.26	Arg27
F1899-0458	-3.471	-40.960	-56.52	Arg27, Arg202, Phe205
PubChem_CID-6918483	-3.300	-21.956	-32.18	Gln20
PubChem_CID-16741	-2.372	-18.128	-27.94	-

Discussion

HCM is a global disease[31] and is considered one of the most common inherited heart disorders, with an estimated prevalence of over 1 in 500[1]. Identifying novel mutations and genes responsible for familial HCM is important, both for providing treatment and prevention strategies and for triggering clinical and genetic surveillance of family members.

Saudi Arabia has one of the highest rates of consanguinity in the world, as marriages between first cousins and close relatives are common and widely accepted[8]. There has been no decrease in the prevalence of consanguinity over a generation, with the tradition of marrying within the family still a preferred practice despite the awareness that certain genetic disorders occur at a higher frequency in cousin marriages[32]. Consanguinity is a risk factor for some congenital abnormalities[9]. In Al-Madinah, in Western Saudi Arabia, the prevalence of cardiomyopathies is very high the pediatric heart patient population[6]. We exploited the information obtainable from a pedigree to investigate the genetic cause of familial HCM in a set of triplets (two HCM subjects and one PDA case) using a 181-gene targeted NGS approach to identify potential pathogenic mutations and genes responsible for inherited HCM.

The triplets were born of non-consanguineous Saudi parents and were born preterm. Subject 1 was diagnosed clinically with severe heart failure and an HCM phenotype; subject 2 was diagnosed with PDA and developed diabetic ketoacidosis at age 3; and subject 3 was diagnosed clinically with a mild HCM phenotype. Their mother and father did not have a clinically relevant history, but there was a family history of sudden cardiac death on the mother's side; two cousins died at age 20 and 27 years from HCM. The family pedigree was highly suggestive of a recessive pattern of inheritance. Bioinformatics analysis detected a homozygous missense mutation in exon 1 of *TMP1* in all triplets that substituted glycine for arginine at codon 3 (p.Gly3Arg). The variant was heterozygous in the parental samples, and there were no disease-causing mutations in the other 180 genes tested.

TPM1 mutations account for ~3% of cases of HCM[33-36]. TPM1 is essential for normal heart development and contractile function[37]. The sarcomeric *TPM1* gene is considered one of commonest causes of HCM [33], DCM[38], left ventricular noncompaction[39], and congenital heart defects (CHDs)[37].

TPM1 p.Gly3Arg was first reported in ClinVar (RCV000036633.3) but the pathogenicity of this variant has not been clinically demonstrated. The mutation was in a highly conserved region of

Table III. The predicted ADME properties of the identified compounds using QikProp.

Compound ID	Mol weight (Da)	QPlogKp	QPlogS	Qplog HERG	QPlogBB	QPPCaco	Percentage of HOA	ROF
F0760-0534	352.3	-2.149	-6.424	-6.044	-0.856	652.512	100	0
F1691-3363	368.3	-5.155	-4.169	-4.052	-2.434	45.728	50.106	0
F2258-0953	496.4	-1.629	-6.010	-5.705	-1.022	400.964	100	0
F1278-0497	344.4	-4.700	-5.097	-4.429	-0.877	87.911	67.064	0
F1899-0458	408.4	-2.089	-6.262	-7.220	-1.389	483.378	96.643	0

the gene. *In silico* predictions of the variant were "probably damaging" by PolyPhen-2, "damaging" by MetaLR, "damaging" by SIFT (low confidence), "disease causing" by MutationTaster2, and "pathogenic" by REVEL. Based on the above evidence and following ACMG classification guidelines[40], we classified the *TPM1* p.Gly3Arg variant as "likely pathogenic".

Our *in silico* studies provide new knowledge on the structural impact of the p.Gly3Arg mutation and drug binding behavior at the active site of the mutated protein. The tertiary protein structure of TPM1 was predicted and validated using Ramachandran plots and molecular dynamics simulations. The Procheck result showed that 98.3% of residues fell within the most favored regions and confirmed that the structure was very reliable for drug development. The molecular dynamics simulation study revealed strong structural stability. The predicted model was used to perform structure-based virtual drug screening to identify potential TPM1 inhibitors. Several potent compounds were identified that had comparable predicted efficacy and high binding affinity, with performance characteristics better than standard drugs. Docking studies revealed that Arg27, Gln20, Ser209, and Arg202 were responsible for the high binding affinity of the identified compounds. The RMSD and RMSF plot support the evidence observed from docking studies, with residue fluctuation and structural deviation contributing to the variation in docking scores and interactions of the wildtype and mutant structures. The binding poses and conformational changes varied only slightly between wildtype and mutant structures. Furthermore, the pharmacokinetic properties of the identified screened compounds were within acceptable ranges.

Our results predict that the homozygous missense mutation in exon 1 of *TPM1* (p.Gly3Arg) is pathogenic could lead to autosomal recessive pediatric HCM and sudden death. Our data support the use of the targeted NGS to identify the mutation(s) and novel gene(s) responsible for familial DCM of unknown genetic cause. Drug binding properties play a major role in determining the potency and affinity of the drug molecule to the target protein active site. The wildtype and mutant structures had some differences in their structural conformation and binding poses, but this did not significantly affect the binding properties of our new candidate drugs[41,42]. Thus, our findings have clear implications for disease diagnosis and management and may be useful in the future for pharmacogenomics and personalized medicine approaches in HCM.

Conclusions

TPM1 mutations are a frequent cause of HCM, DCM, and congenital heart defects. To date, TPM1 has not been associated with isolated PDA; here we describe the first case of this association. By predicting a reliable tertiary structure, we identified potential TPM1 inhibitors through structure-based virtual screening and that had acceptable drug-like properties and pharmacokinetic profiles. To our knowledge, this is the first report of the homozygous missense variation p.Gly3Arg in *TPM1* in association with a familial autosomal recessive pediatric HCM and PDA phenotype. Our findings suggest that *TPM1* p.Gly3Arg is associated with life-threatening, recessively inherited pediatric HCM and PDA, and further *in vitro* and *in vivo* investigations of the identified compounds are warranted.

Funding

This research was supported by the Strategic Technologies Programs of the National Plan for Science, Technology and Innovation (MAARIFAH), Kingdom of Saudi Arabia. Project No: 12-MED3174-05, through the Science and Technology Unit (STU), Taibah University, Al Madinah Al Munawwarah, Kingdom of Saudi Arabia.

Conflict of Interests

All authors declare that they have no conflicts of interest.

References

1) Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. J Am Coll Cardiol 2015; 65: 1249-1254.
2) de Noronha SV, Sharma S, Papadakis M, Desai S, Whyte G, Sheppard MN. Aetiology of sudden cardiac death in athletes in the United Kingdom: a pathological study. Heart 2009; 95: 1409-1414.
3) Maron BJ, Doerer JJ, Haas TS, Tierney DM, Mueller FO. Sudden deaths in young competitive athletes: analysis of 1866 deaths in the United States, 1980-2006. Circulation 2009; 119: 1085-1092.
4) Sabater-Molina M, Pérez-Sánchez I, Hernández Del Rincón JP, Gimeno JR. Genetics of hypertrophic cardiomyopathy: a review of current state. Clin Genet 2018; 93: 3-14.
5) Silas JC, Al Harbi KM. Gene database for the development of genetic testing for hypertrophic cardiomyopathy. BMC Genomics 2014; 15: P65-P65.
6) Tardiff JC, Carrier L, Bers DM, Poggesi C, Ferrantini C, Coppini R, Maier LS, Ashrafian H, Huke S, van

der Velden J. Targets for therapy in sarcomeric cardiomyopathies. Cardiovasc Res 2015; 105: 457-470.

7) Carlus SJ, Almuzaini IS, Karthikeyan M, Loganathan L, Al-Harbi GS, Abdallah AM, Al-Harbi KM. Next-generation sequencing identifies a homozygous mutation in ACADVL associated with pediatric familial dilated cardiomyopathy. Eur Rev Med Pharmacol Sci 2019; 23: 1710-1721.

8) el-Hazmi MA, al-Swailem AR, Warsy AS, al-Swailem AM, Sulaimani R, al-Meshari AA. Consanguinity among the Saudi Arabian population. J Med Genet 1995; 32: 623-626.

9) Abdulrazzaq YM, Bener A, al-Gazali LI, al-Khayat AI, Micallef R, Gaber T. A study of possible deleterious effects of consanguinity. Clin Genet 1997; 51: 167-173.

10) Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P, Hagege AA, Lafont A, Limongelli G, Mahrholdt H, McKenna WJ, Mogensen J, Nihoyannopoulos P, Nistri S, Pieper PG, Pieske B, Rapezzi C, Rutten FH, Tillmanns C, Watkins H. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). Eur Heart J 2014; 35: 2733-2779.

11) Kampmann C, Wiethoff CM, Wenzel A, Stolz G, Betancor M, Wippermann CF, Huth RG, Habermehl P, Knuf M, Emschermann T, Stopfkuchen H. Normal values of M mode echocardiographic measurements of more than 2000 healthy infants and children in central Europe. Heart 2000; 83: 667-672.

12) Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008; 9: 40.

13) Schrödinger Release 2018-4: Maestro, Schrödinger, LLC, New York, NY, 2018.

14) Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 2012; 28: 1647-1649.

15) Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004; 32: W526-W531.

16) Schrödinger Release 2018-4: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2018.

17) Loganathan L, Muthusamy K, Jayaraj JM, Kajamaideen A, Balthasar JJ. insights on tankyrase protein: A potential target for colorectal cancer. J Biomol Struct Dyn 2019; 37: 3637-3648.

18) Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. The GeneCards Suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016; 54: 1.30.31-31.30.33.

19) Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J Chem Theory Comput 2016; 12: 281-296.

20) Greenwood JR, Calkins D, Sullivan AP, Shelley JC. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J Comput Aided Mol Des 2010; 24: 591-604.

21) Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA, Harder ED. OPLS3e: extending force field coverage for drug-like small molecules. J Chem Theory Comput 2019; 15: 1863-1874.

22) Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 2004; 1: 337-341.

23) Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 2006; 49: 6177-6196.

24) Loganathan L, Natarajan K, Muthusamy K. Computational study on cross-talking cancer signalling mechanism of ring finger protein 146, AXIN and Tankyrase protein complex. J Biomol Struct Dyn 2019: 1-13.

25) Cappel D, Hall ML, Lenselink EB, Beuming T, Qi J, Bradner J, Sherman W. Relative binding free energy calculations applied to protein homology models. J Chem Inf Model 2016; 56: 2388-2400.

26) Beard H, Cholleti A, Pearlman D, Sherman W, Loving KA. Applying physics-based scoring to calculate free energies of binding for single amino acid mutations in protein-protein complexes. PLoS One 2013; 8: e82849.

27) Chatterjee A, Cutler SJ, Doerksen RJ, Khan IA, Williamson JS. Discovery of thienoquinolone derivatives as selective and ATP non-competitive CDK5/p25 inhibitors by structure-based virtual screening. Bioorg Med Chem 2014; 22: 6409-6421.

28) Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 2016; 99: 877-885.

29) Laskowski RA MMW, Thornton JM. PROCHECK: validation of protein-structure coordinates. in International Tables of Crystallography, Crystallog-

raphy of Biological Macromolecules,. Vol Volume F. Rossmann, M.G., Arnold, E., Eds.: Kluwer Academic Publishers: Dordrecht, The Netherlands; 2001: 722-725.

30) Schrödinger Release 2018-4: QikProp, Schrödinger, LLC, New York, NY, 2018.

31) Maron BJ. Hypertrophic cardiomyopathy: an important global disease. Am J Med 2004; 116: 63-65.

32) Warsy AS, Al-Jaser MH, Albdass A, Al-Daihan S, Alanazi M. Is consanguinity prevalence decreasing in Saudis?: a study in two generations. Afr Health Sci 2014; 14: 314-321.

33) Watkins H, Anan R, Coviello DA, Spirito P, Seidman JG, Seidman CE. A de novo mutation in alpha-tropomyosin that causes hypertrophic cardiomyopathy. Circulation 1995; 91: 2302-2305.

34) Nakajima-Taniguchi C, Matsui H, Nagata S, Kishimoto T, Yamauchi-Takihara K. Novel missense mutation in alpha-tropomyosin gene found in Japanese patients with hypertrophic cardiomyopathy. J Mol Cell Cardiol 1995; 27: 2053-2058.

35) Yamauchi-Takihara K, Nakajima-Taniguchi C, Matsui H, Fujio Y, Kunisada K, Nagata S, Kishimoto T. Clinical implications of hypertrophic cardiomyopathy associated with mutations in the alpha-tropomyosin gene. Heart 1996; 76: 63-65.

36) Coviello DA, Maron BJ, Spirito P, Watkins H, Vosberg HP, Thierfelder L, Schoen FJ, Seidman JG, Seidman CE. Clinical features of hypertrophic cardiomyopathy caused by mutation of a "hot spot" in the alpha-tropomyosin gene. J Am Coll Cardiol 1997; 29: 635-640.

37) England J, Granados-Riveron J, Polo-Parada L, Kuriakose D, Moore C, Brook JD, Rutland CS, Setchfield K, Gell C, Ghosh TK, Bu'Lock F, Thornborough C, Ehler E, Loughna S. Tropomyosin 1: Multiple roles in the developing heart and in the formation of congenital heart defects. J Mol Cell Cardiol 2017; 106: 1-13.

38) Olson TM, Kishimoto NY, Whitby FG, Michels VV. Mutations that alter the surface charge of alpha-tropomyosin are associated with dilated cardiomyopathy. J Mol Cell Cardiol 2001; 33: 723-732.

39) Probst S, Oechslin E, Schuler P, Greutmann M, Boyé P, Knirsch W, Berger F, Thierfelder L, Jenni R, Klaassen S. Sarcomere gene mutations in isolated left ventricular noncompaction cardiomyopathy do not predict clinical phenotype. Circ Cardiovasc Genet 2011; 4: 367-374.

40) Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015; 17: 405-424.

41) Miceli LA, Teixeira VL, Castro HC, Rodrigues CR, Mello JF, Albuquerque MG, Cabral LM, de Brito MA, de Souza AMT. Molecular docking studies of marine diterpenes as inhibitors of wild-type and mutants HIV-1 reverse transcriptase. Mar Drugs 2013; 11: 4127-4143.

42) Raghav PK, Kumar R, Kumar V, Raghava GPS. Docking-based approach for identification of mutations that disrupt binding between Bcl-2 and Bax proteins: Inducing apoptosis in cancer cells. Mol Genet Genomic Med 2019; 7: e910.