Data in Brief 7 (2016) 607–613

Contents lists available at ScienceDirect

Data in Brief

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

Data article

Prediction and visualization data for the interpretation of sarcomeric and non-sarcomeric DNA variants found in patients with hypertrophic cardiomyopathy

Irene Bottillo [a,*], Daniela D'Angelantonio [a], Viviana Caputo [b], Alessandro Paiardini [c], Martina Lipari [a], Carmelilia De Bernardo [a], Silvia Majore [a], Marco Castori [a], Elisabetta Zachara [d], Federica Re [d], Paola Grammatico [a]

[a] *Medical Genetics, Department of Molecular Medicine, Sapienza University, San Camillo-Forlanini Hospital, Rome, Italy*
[b] *Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy*
[c] *Department of Biochemical Sciences, Sapienza University of Rome, Rome, Italy*
[d] *Cardiomyopathies Unit, Division of Cardiology and Cardiac Arrhythmias, San Camillo-Forlanini Hospital, Rome, Italy*

ARTICLE INFO

Article history:
Received 8 December 2015
Received in revised form
16 February 2016
Accepted 1 March 2016
Available online 10 March 2016

ABSTRACT

Genomic technologies are redefining the understanding of genotype–phenotype relationships and over the past decade, many bioinformatics algorithms have been developed to predict functional consequences of single nucleotide variants. This article presents the data from a comprehensive computational workflow adopted to assess the biomedical impact of the DNA variants resulting from the experimental study "Molecular analysis of sarcomeric and non-sarcomeric genes in patients with hypertrophic cardiomyopathy" (Bottillo et al., 2016) [1]. Several different independently methods were employed to predict the functional consequences of alleles that result in amino acid substitutions, to study the effect of some DNA variants over the splicing process and

DOI of original article: http://dx.doi.org/10.1016/j.gene.2015.11.048
* Correspondence to: Medical Genetics, Department of Molecular Medicine, Sapienza University, San Camillo-Forlanini Hospital Circonvallazione Gianicolense, 87 - 00152 Rome, Italy. Tel.: +39 06 58704622; fax: +39 06 5870 4657.
E-mail address: i.bottillo@gmail.com (I. Bottillo).

http://dx.doi.org/10.1016/j.dib.2016.03.004

to investigate the impact of a sequence variant with respect to the evolutionary conservation.

Specifications Table

Subject area	Biology
More specific subject area	In silico predictions of DNA variants
Type of data	Tables, figures
How data was acquired	Prediction tools: SIFT, Polyphen HDIV, Polyphen HVAR, Provean, LRT, Mutation Taster, Mutation Assessor, FATHMM, RadialSVM, LR, CADD, HSF, GERP++, PhyloP placental, PhyloP veterbrate, SiPhyMolecular Modeling
Data format	Processed, filtered and analyzed
Experimental factors	Genomic DNA from peripheral blood was tested by next generation sequencing on Ion Torrent PGM (ThermoFisher, Carlsbad, CA, USA) with a custom cardiomyopathy panel
Experimental features	The identified rare (Minor Allele Frequency $\leq 0{,}01$) non-synonymous DNA changes were subjected to different *in silico* predictions
Data source location	Rome, Italy
Data accessibility	These data are with this article

Value of the data

- These data delineate a prompt informatic pipeline for the prioritization of the most likely pathogenetic DNA variants in a clinical context.
- These data are supportive for the researchers to evaluate the prevalence of sarcomeric and non-sarcomeric gene variants in hypertrophic cardiomyopathy.
- The described computational strategy is helpful to researchers for the rapid interpretation of Variants of Unknown Significance (VUS) implicated in rare, common and complex diseases.

1. Data

Here we report the in silico predictions data of the non-synonymous changes found in 41 HCM patients and in 3 HCM-related cases [1] (Table 1).

2. Experimental design, materials and methods

2.1. Analysis of the nucleotides' evolutionary conservation

Nucleotide-specific estimates of evolutionary constraint were explored by (i) GERP++ (Genomic Evolutionary Rate Profiling); (ii) PhyloP placental; (iii) PhyloP veterbrate and (iv) SiPhy.

2..2. Analysis of the splicing variants

The analysis of intronic variants leading to splicing defects was tested by Human Splicing Finder (HSF) 3.0.

Table 1
Results of the in silico predictions of the non-synonymous changes found in 41 HCM patients and in 3 HCM-related cases. Deleterious predictions are in bold.

Mutation	Chr	Start	End	Ref	Alt	SIFT	Polyphen2 HDIV	Polyphen2 HVAR	Provean	LRT	Mutation Taster	Mutation Assessor	FATHMM	RadialSVM	LRT	CADD phred	Uniprot	Domain	X-ray	Model Template	% id	Prosa Z-Score	Possible effect	HSF (altered site)	GERP++ RS	phyloP46way placental	phyloP100way vertebrate	SiPhy 29way
MYBPC3 NM_000256:c.T1664C:p.M555T	11	47363668	47363668	A	G	T (1)	B (0)	B (0)	N (4,48)	.	N (0,997)	N (-2,74)	T (-0,1)	T (-1,057)	T (0,041)	0,033	Q14896	Ig-like C2-Type	-	2DLT	65	-8.58	No predicted deleterious effect	.	**4,22**	1,401	**7,388**	**14,08**
RYR2 NM_001035:c.A3380G:p.E1127G	1	237730032	237730032	A	G	**D (0,05)**	**D (0,979)**	**P (0,675)**	**D (-5,34)**	**D (0)**	**D (1)**	L (1,06)	T (-0,43)	T (-0,504)	T (0,279)	**20,2**	Q92736	SPRY	-	3TOY	28	-5.23	**Loss of n-wixin interaction with Trp1145**	.	**5,29**	**1,995**	**7,302**	**15,238**
TNPO NM_003276:c.A1037G:p.H346R	12	98927072	98927072	A	G	T (0,42)	B (0,001)	B (0,001)	N (0,22)	N (0,816)	N (1)	N (0,345)	T (1,7)	T (-0,979)	T (0,032)	3,347	.	.	.	.	.	.	.	.	-1,03	-0,111	-0,041	8,773
MYH6 NM_002471:c.643-5C>T	14	23873602	23873602	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**branch point**	.	.	.	.
RBM20 NM_001134363:c.G3373A:p.E1125K	10	112583294	112583294	G	A	**D (0,01)**	**P (0,911)**	B (0,293)	N (-1,80)	.	**D (0,992)**	L (1,355)	T (-0,91)	T (-0,499)	T (0,22)	**17,99**	.	.	.	.	.	.	.	.	**6,16**	**2,937**	5,214	**13,979**
MYH6 NM_002471:c.2928+5G>A	14	23862870	23862870	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**donor site**	.	.	.	.
CSRP3 NM_003476:c.T10C:p.W4R	11	19213986	19213986	A	G	T (0,08)	**P (0,888)**	**P (0,657)**	**D (-3,25)**	**D (0)**	**A (1)**	L (1,78)	T (-0,1)	T (-0,44)	T (0,244)	**25,9**	.	.	.	.	.	.	.	.	**6,02**	**2,304**	**7,415**	**16,203**
LAMA4 NM_001105206:c.A4937G:p.E1646G	6	112438986	112438986	T	C	**D (0)**	**D (1)**	**D (0,997)**	**D (-6,11)**	**D (0)**	**D (1)**	**M (2,43)**	T (-1,07)	**D (0,289)**	**D (0,582)**	**26,9**	Q16363	Laminin G-like	-	3ASI	24	-6.42	**Loss of interaction between two adjacent domains**	.	**5,6**	**2,141**	**7,17**	**15,783**
LAMA4 NM_001105206:c.G1565C:p.R522T	6	112486465	112486465	C	G	T (0,07)	B (0,324)	B (0,174)	N (-1,24)	N (0,006)	N (1)	L (1,04)	T (2,94)	T (-0,978)	T (0,008)	8,741	.	.	.	.	.	.	.	.	-1,86	-0,523	-0,169	9,422
NEBL NM_006393:c.G604A:p.G202R	10	21157673	21157673	C	T	T (0,47)	B (0,001)	B (0,001)	N (-0,48)	N (0,002)	**D (1)**	N (0,69)	T (0,61)	T (-1,042)	T (0,025)	10,5	.	.	.	.	.	.	.	.	**4,17**	**1,541**	1,543	8,948
RYR2 NM_001035:c.945D>C>T	1	237868504	237868504	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MYBPC3 NM_000256:c.C1112A:p.P371Q	11	47365154	47365154	G	T	**D (0)**	**D (1)**	**D (0,999)**	**D (-7,05)**	.	**D (1)**	**M (2,48)**	T (0,96)	T (-0,415)	T (0,296)	**31**	Q14896	Ig-like C2-Type	1PD6	.	.	.	.	**Possible decreased protein stability**	.	**4,99**	**2,602**	**8,974**	**18,474**
GLA NM_000169:c.A644G:p.N215S	X	100653930	100653930	T	C	**D (0,01)**	B (0,048)	B (0,004)	**D (-4,39)**	**D (0)**	**D (1)**	.	**D (-7,51)**	**D (1,044)**	**D (0,982)**	**16,95**	P06280	Melibiase	4NXS	.	.	.	**Loss of a glycosylated site**	.	**5,97**	**2,018**	**8,04**	**15,388**
DSP NM_004415:c.G137A:p.G46D	6	7542285	7542285	G	A	T (0,17)	B (0,03)	B (0,007)	N (-0,25)	N (0,867)	N (1)	N (0)	T (-0,45)	T (-1,037)	T (0,069)	6,748	.	.	.	.	.	.	.	.	0,722	-0,071	0,42	3,809
TNNRC2 NM_006440:c.G1150A:p.G384S	22	19868177	19868177	C	T	T (0,29)	B (0,367)	B (0,03)	**D (-3,01)**	**D (0,001)**	**D (1)**	**M (2,15)**	T (0,1)	T (-0,941)	T (0,116)	12,53	Q17RE6	.	.	1ZDL	84	-11.53	No predicted deleterious effect	.	**4,35**	1,411	2,985	8,752
JPH2 NM_020433:c.G1536C:p.W512C	20	42744779	42744779	C	G	T (0,18)	**D (1)**	**D (0,997)**	N (-1,95)	U (0,004)	**D (1)**	L (1,845)	T (0,1)	T (-0,508)	T (0,315)	12,45	.	.	.	.	.	.	.	.	3,12	1,441	**4,931**	**13,328**
SCN5A NM_001099405:c.A5605T:p.I1869F	3	38592204	38592204	T	A	**D (0)**	**D (0,998)**	**D (0,986)**	**D (-2,56)**	**D (0)**	**D (1)**	**M (2,005)**	**D (-3,79)**	**D (0,918)**	**D (0,892)**	**17,96**	Q14524	Interaction with FGF13	4DCK	.	.	.	**Potential loss of interaction with FGF13**	.	**3,7**	0,887	1,661	10,421
FHL2 NM_001450:c.G109T:p.A37S	2	106002865	106002865	C	A	T (0,14)	**P (0,62)**	B (0,225)	N (-1,46)	**D (0)**	**D (0,999)**	L (0,98)	**D (-2,32)**	T (-0,375)	T (0,359)	13,95	Q14192	Zinc Finger C4-type	2M1U	.	.	Q14192	**Steric clashes with T34 and P46**	.	**4,66**	**2,811**	3,772	11,58
DSC2 NM_004949:c.C1787T:p.A596V	18	28654750	28654750	G	A	**D (0,03)**	**P (0,882)**	B (0,345)	**D (-2,85)**	N (0,006)	**D (1)**	**M (2,995)**	T (-0,34)	T (-0,14)	T (0,354)	14,64	.	.	.	.	.	.	.	.	**5,27**	**2,614**	**5,918**	**16,176**
PKP2 NM_001005242:c.G76A:p.D26N	12	33049590	33049590	C	T	T (0,06)	**D (0,992)**	**P (0,874)**	N (-1,08)	N (0,004)	**D (0,999)**	L (0,895)	**D (-1,72)**	**D (0,033)**	T (0,451)	**33**	.	.	.	.	.	.	.	.	**4,07**	**2,081**	3,56	**15,41**
MYOM1 NM_003803:c.C2131T:p.R711C	18	3135623	3135623	G	A	**D (0)**	**D (1)**	**D (1)**	**D (-7,47)**	**D (0)**	**D (1)**	**H (3,89)**	T (0,15)	**D (0,396)**	**D (0,528)**	**26,8**	P52179	Fn3	.	1XSY	40	-4.69	**Loss of inter-residues contacts**	.	**4,09**	1,498	**6,747**	**15,443**
RAF1 NM_002880:c.G624A:C125T:p.A42I	3	12660096	12660097	GC	AT	T (0,15)	B (0,44)	B (0,08)	N (-0,15)	**D (0)**	**D (1)**	N (0,55)	T (-0,9)	T (-0,708)	T (0,248)	**20,8**	.	.	.	.	.	.	.	.	**6,17**	**2,941**	3,851	**20,879**
TNNT2 NM_00100:1430:c.A252T:p.R84S	1	201334750	201334750	T	A	**D (0)**	**D (0,968)**	**P (0,712)**	**D (-2,99)**	**D (0,001)**	**D (1)**	L (1,2)	**D (-5,31)**	**D (0,887)**	**D (0,92)**	**18,22**	.	.	.	.	.	.	.	.	**3,53**	0,733	0,317	4,525
MYOM1 NM_003803:c.4485-6T>C	18	3079346	3079346	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**branch point**	.	.	.	.
RBM20 NM_001134363:c.A59G:p.D20G	10	112404271	112404271	A	G	**D (0)**	**P (0,941)**	B (0,355)	N (-0,94)	N (0,832)	N (0,884)	N (0,345)	**D (-2,93)**	T (-0,205)	**D (0,575)**	**22,2**	.	.	.	.	.	.	.	.	**4,23**	**1,54**	1,69	9,721
CAV3 NM_001234:c.C233T:p.T78M	3	8787330	8787330	C	T	T (0,15)	**D (0,973)**	**P (0,738)**	N (-0,83)	**D (0)**	**D (0,999)**	L (1,67)	**D (-3,08)**	**D (0,466)**	**D (0,71)**	**16,45**	.	.	.	.	.	.	.	.	**4,63**	**2,386**	3,58	**16,203**
DSG2 NM_001943:c.A208G:p.I70V	18	29098892	29098892	A	G	T (0,42)	B (0,001)	B (0,005)	N (-0,93)	U (0,001)	**D (0,716)**	N (0,67)	T (0,22)	T (-0,944)	T (0,11)	10,79	.	.	.	.	.	.	No predicted deleterious effect	.	**4,05**	0,919	**4,397**	10,003
DSP NM_004415:c.G5218A:p.E1740K	6	7581641	7581641	G	A	T (0,19)	**D (0,962)**	B (0,173)	N (-1,58)	**D (0)**	**D (1)**	L (1,1)	T (-0,85)	T (-0,875)	T (0,14)	11,6	.	.	.	.	.	.	.	.	**5,09**	1,496	3,14	**17,082**
MYOM1 NM_003803:c.G2132A:p.R711H	18	3135622	3135622	C	T	**D (0)**	**D (1)**	**D (1)**	**D (-6,67)**	**D (0)**	**D (1)**	**H (3,89)**	T (0,15)	**D (0,366)**	**D (0,57)**	**36**	P52179	Fn3	.	1XSY	40	-4.67	**Loss of inter-residues contacts**	.	**5,95**	**2,821**	**7,818**	**20,381**
SCN5A NM_001099405:c.C5837T:p.S1946F	3	38591972	38591972	G	A	**D (0)**	**D (0,98)**	**P (0,641)**	**D (-2,82)**	N (0,931)	N (0,899)	N (0,345)	**D (-3,81)**	**D (0,336)**	**D (0,743)**	13,71	.	.	.	.	.	.	.	.	3,08	1,282	2,36	5,503
DSC2 NM_004949:c.C2328G:p.I776M	18	28649040	28649040	G	C	T (0,27)	B (0,008)	B (0,023)	N (-0,99)	N (0,155)	N (1)	N (-0,295)	T (0,36)	T (-1,066)	T (0,064)	8,255	.	.	.	.	.	.	.	.	-0,947	0,087	0,3	3,159
LAMA4 NM_001105206:c.T3482C:p.M1161T	6	112455744	112455744	A	G	T (0,16)	B (0)	B (0,002)	N (-1,49)	N (0,021)	**D (1)**	L (0,895)	T (-1,01)	T (-0,94)	T (0,152)	12,5	Q16363	Laminin G-like	-	2W35	23	-6.42	No predicted deleterious effect	.	**3,53**	0,495	**5,708**	9,835

LAMP2 NM_001122606.c.C418G p.L140V	X	119582963	119582963	G	C	T (0,8)	B (0,012)	B (0,022)	N (0,25)	N (0,04)	N (1)	N (-0,05)	T (1,45)	T (-1,031)	T (0,027)	1,336	.	.	.	.	.	.	.	.	.	-6,74	-1,237	-4,427	1,415		
DSG2 NM_001943.c.G2147A p.G716E	18	29122628	29122628	G	A	T (0,98)	B (0,132)	B (0,06)	N (-1,68)	N (0,043)	N (1)	N (0,55)	T (0,28)	T (-1,012)	T (0,094)	11,87	.	.	.	.	.	.	.	.	.	4,14	0,789	1,617	9,789		
LAMA4 NM_001105206.c.4665+8G>T	6	112441678	112441678	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	IIE site	.	.	.	.		
MYBPC3 NM_000256.c.2309-2A>G	11	47359347	47359347	T	C	.	.	.	.	.	D (1)	.	.	.	.	16,83	.	.	.	.	.	.	.	.	acceptor site	4,66	1,961	7,501	13,428		
RYR2 NM_001035.c.G14544A p.R4854Q	1	237617852	237617852	G	A	T (0,07)	**D (0,997)**	**P (0,821)**	**D (-2,94)**	U (0)	**D (1)**	**M (2,575)**	**D (-3,96)**	**D (1,018)**	**D (0,925)**	**27,2**	Q92736	RYDR_ITPR	.	4IKQ	93	-9,43	**Loss of n-cation interaction with Tyr462**	.	**5,62**	**2,645**	**9,804**	**19,666**		
ABCC9 NM_005691.c.G2200A p.V734I	12	22017410	22017410	C	T	T (0,45)	B (0,004)	B (0,009)	N (-0,10)	N (0,044)	**D (1)**	N (0,75)	**D (-3,27)**	T (-0,808)	**D (0,642)**	13,11	O60706	ABC_tran	.	2CBZ	36	-8,25	No predicted deleterious effect	.	**4,97**	**2,575**	2,757	**18,028**		
CALR3 NM_145046.c.A1036G p.I346V	19	16590061	16590061	T	C	T (0,29)	**D (0,883)**	**P (0,672)**	N (0,10)	N (0,042)	N (0,999)	N (0,63)	T (0,99)	T (-1,063)	T (0,085)	11,06	Q96L12	P8035848	.	3PO5	61	-1,93*	No predicted deleterious effect	.	2,67	0,28	1,313	6,573		
TNNT2 NM_001001430.c.264+7G>A	1	201334731	201334731	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	donor site	.	.	.	.		
MYH7 NM_000257.c.G4402A p.E1468K	14	23886479	23886479	C	T	**D (0)**	**D (0,999)**	**D (0,999)**	**D (-2,94)**	.	**D (1)**	**M (3,495)**	T (-1,48)	**D (0,727)**	**D (0,768)**	**36**	.	.	.	.	.	.	.	.	.	**5,37**	**2,793**	**7,314**	**19,297**		
ILK NM_001014795.c.G1086C p.G362H	11	6631386	6631386	G	C	**D (0,01)**	**D (0,998)**	**D (0,929)**	**D (-3,99)**	**D (0)**	**D (1)**	**M (2,83)**	**D (-1,68)**	**D (0,434)**	**D (0,693)**	**15**	Q13418	Pkinase_tyr	3REP	.	.	.	No predicted deleterious effect	.	3,17	0,795	3,198	10,671		
LAMA4 NM_001105206.c.G1967A p.S656N	6	112476142	112476142	C	T	T (0,34)	B (0,001)	B (0,002)	N (-0,54)	N (0,1)	**D (1)**	L (0,895)	T (2,59)	T (-0,995)	T (0,009)	9,958	.	.	.	.	.	.	.	.	.	3,39	0,749	0,577	4,097		
MYH6 NM_002471.c.G68A p.R23H	14	23876365	23876365	C	T	**D (0,05)**	**D (0,998)**	**P (0,813)**	**D (-2,81)**	.	**D (1)**	**M (2,71)**	**D (-2,05)**	**D (0,596)**	**D (0,714)**	**17,94**	P13533	Myosin rotor	4D81	.	.	.	**Loss of a salt-bridge with E20**	.	**3,53**	**1,974**	**7,464**	**15,232**		
TPM1 NM_001018008.c.C134T p.A45V	15	63349185	63349185	C	T	**D (0)**	**D (1)**	**D (0,993)**	**D (-3,35)**	N (0,004)	**D (1)**	**H (3,855)**	T (-1,28)	**D (0,843)**	**D (0,793)**	**36**	.	.	.	.	.	.	.	.	.	**5,57**	**2,628**	**7,818**	**18,541**		
PKP2 NM_001005242.c.C2299A p.R767S	12	32949101	32949101	G	T	**D (0,01)**	B (0,434)	B (0,219)	**D (-3,22)**	N (0,004)	**D (1)**	**M (2,585)**	T (0,78)	T (-0,636)	T (0,201)	**17,8**	Q99959	Arm repeat 8	.	1XM9	36	-4,71	**Loss of salt-bridge with E852**	.	**5,06**	**2,508**	3,303	**17,583**		
DSP NM_004415.c.C9516T p.T1319I	6	7540239	7540379	C	T	T (0,26)	**D (0,998)**	**D (0,987)**	N (-1,03)	**D (0)**	**D (1)**	L (1,1)	**D (-2,86)**	**D (0,392)**	**D (0,725)**	**15,02**	.	.	.	.	.	.	.	.	.	**5,27**	**2,495**	**4,343**	**18,553**		
ABCC9 NM_005691.c.2238-1G>A	12	22015989	22015989	C	T	.	.	.	.	.	**D (1)**	.	.	.	.	**24,1**	.	.	.	.	.	.	.	.	acceptor site	**5,44**	**2,723**	**4,703**	**18,618**		
DTNA NM_001198958.c.C2095T p.R699C	6	7580379	7580379	C	T	.	.	.	N (-1,03)	.	.	.	.	.	.	**15,02**	.	.	.	.	.	.	.	.	.	.	.	.	.		
LMNA NM_001257374.c.1363-7T>C	1	156108272	156108272	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
ABCC9 NM_005691.c.816+11G>A	12	22068591	22068591	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	donor site	.	.	.	.		
JPH2 NM_020433.c.G1513A p.G505S	20	42744802	42744802	C	T	T (0,71)	B (0,019)	B (0,002)	N (0,13)	U (0,122)	N (1)	L (1,39)	T (0,26)	T (-1,061)	T (0,039)	1,421	.	.	.	.	.	.	.	.	.	1,03	-0,002	0,214	3,534		
JUP NM_002230.c.909+6C>T	17	39923625	39923625	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
RAF1 NM_002880.c.G185A p.A62OT	3	12626102	12626102	C	T	**D (0,05)**	**P (0,724)**	B (0,132)	**D (-2,76)**	**D (0)**	**D (1)**	L (1,33)	T (-0,9)	T (-0,456)	T (0,286)	**15,77**	.	.	.	.	.	.	.	.	.	**5,51**	**2,868**	**4,487**	**19,61**		
SCN5A NM_000335.c.C1810G p.L604V	3	38645283	38645283	G	C	T (1)	**D (0,992)**	**P (0,866)**	N (0,17)	N (0,461)	**D (0,565)**	N (0,66)	**D (-2,59)**	T (-0,143)	**D (0,528)**	5,963	.	.	.	.	.	.	.	.	.	3,28	**2,164**	0,885	3,132		
ABCC9 NM_005691.c.2867-5T>C	12	21998771	21998771	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
MYOM1 NM_003803.c.T1996g p.S67A	18	3215023	3215023	A	C	T (1)	B (0)	B (0,003)	N (-0,13)	N (0,726)	N (1)	N (-0,345)	T (1,12)	T (-1,002)	T (0,043)	0,006	.	.	.	.	.	.	.	.	.	-0,951	-0,851	-1,373	11,115		
TNNT2 NM_001001432.c.A755T p.K252I	1	201330414	201330414	T	A	**D (0)**	**D (0,997)**	**D (0,969)**	**D (-6,86)**	**D (0)**	**D (1)**	**M (2,565)**	**D (-3,13)**	**D (0,877)**	**D (0,851)**	**22,3**	P45379	Coiled coil	1J1D	.	.	.	No predicted deleterious effect	.	**3,73**	**1,552**	**7,699**	12,581		
SCN5A NM_000335.c.G2614A p.D872N	3	38627355	38627355	C	T	T (0,46)	**D (1)**	**D (0,993)**	N (-0,42)	N (0,099)	N (1)	N (-0,505)	**D (-4,89)**	**D (0,308)**	**D (0,851)**	12,06	Q14524	Na-channel	.	4DXW	29	-1,93	**Loss of negative charge at Na-channel**	.	3,26	**2,169**	-0,24	8,027		
BAG3 NM_004281.c.G463A p.A155T	10	121429645	121429645	G	A	T (0,81)	B (0)	B (0,001)	N (0,56)	N (0,061)	N (1)	N (0,205)	T (-0,74)	T (-1,06)	T (0,094)	3,627	.	.	.	.	.	.	.	.	.	1,69	0,515	2,03	0,918		
MYPN NM_032578.c.C3335T p.P1112L	10	69959174	69959174	C	T	.	**D (1)**	**D (0,999)**	**D (-5,37)**	**D (0)**	**D (1)**	N (0,61)	T (-0,3)	T (-0,646)	T (0,24)	**25,9**	Q86TC9	Ig-like	.	2DM3	70	-6,29	**Possible decreased protein stability**	.	**5,38**	**2,813**	**5,719**	**14,201**		
CAV3 NM_001234.c.G221A p.R74H	3	8787318	8787318	G	A	**D (0,01)**	**D (1)**	**D (0,99)**	**D (-4,23)**	**D (0)**	**D (1)**	**M (3,105)**	**D (-3,41)**	**D (1,032)**	**D (0,907)**	**19,29**	.	.	.	.	.	.	.	.	.	**4,63**	**2,386**	**9,181**	**16,203**		
MYH7 NM_000257.c.G676A p.A226T	14	23900850	23900850	C	T	T (0,21)	**D (0,989)**	**D (0,948)**	N (-1,82)	.	**D (1)**	**M (3,24)**	**D (-3,67)**	**D (0,932)**	**D (0,906)**	**20,1**	P12883	Myosin head	4PTH	.	.	.	**Clashes with residues A335-F336**	.	3,28	**1,84**	1,119	11,973		
DSP NM_004415.c.G6188A p.R2063Q	6	7583683	7583683	G	A	T (0,54)	**D (0,998)**	**P (0,872)**	N (-1,91)	**D (0)**	N (0,639)	L (1,245)	T (0,73)	T (-0,751)	T (0,267)	**19,86**	.	.	.	.	.	.	.	.	.	**4,98**	**2,459**	3,613	**18,609**		
DSC2 NM_004949.c.C955T p.R319C	18	28669437	28669437	G	A	**D (0)**	**P (0,862)**	B (0,266)	**D (-4,68)**	N (0,451)	N (1)	**M (3,2)**	T (0,56)	T (-0,589)	T (0,247)	10,26	Q02487	Cadherin	.	3K5R	39	-6,36	**Possible disulfide bond with C197**	.	-1,58	-0,645	-0,446	**14,629**		
MYH7 NM_000257.c.G334A p.E111K	14	23889434	23889434	C	T	**D (0,04)**	**D (1)**	**D (0,999)**	**D (-2,69)**	.	**D (1)**	**M (3,07)**	**D (-1,71)**	**D (0,695)**	**D (0,773)**	**26,6**	.	.	.	.	.	.	.	.	.	**5,13**	**2,678**	**7,521**	**18,771**		
MYH6 NM_002471.c.G622A p.D208N	14	23873940	23873940	C	T	T (0,1)	B (0,006)	B (0,063)	N (-1,25)	.	**D (0,967)**	N (0,47)	T (-0,57)	T (-0,796)	T (0,152)	**16,84**	.	.	.	.	.	.	.	.	.	3,42	**1,909**	2,163	**13,909**		
MYH6 NM_002471.c.G4139A p.R1398Q	14	23857530	23857530	C	T	**D (0,01)**	B (0,222)	B (0,076)	**D (-2,81)**	.	**D (1)**	**M (3,425)**	T (-1,27)	**D (0,174)**	**D (0,54)**	**15**	.	.	.	.	.	.	.	.	.	**4,64**	**2,292**	**7,528**	**17,892**		
DSG2 NM_001943.c.G2033C p.G678A	18	29122514	29122514	G	C	T (1)	B (0,003)	B (0,001)	N (-0,45)	N (0,928)	N (1)	N (-0,805)	T (0,4)	T (-1,015)	T (0,04)	5,823	.	.	.	.	.	.	.	.	.	0,162	-0,176	-0,487	0,709		
PKP2 NM_001005242.c.C1445T p.T482M	12	32994073	32994073	G	A	T (0,12)	B (0,198)	B (0,032)	N (-1,40)	N (0,002)	N (1)	L (1,56)	T (-1,07)	T (-0,835)	T (0,225)	3,327	Q99959	.	.	3TT9	.	.	No predicted deleterious effect	.	2,01	0,068	0,502	8,816		
MYH7 NM_000257.c.G428A p.R143Q	14	23901922	23901922	C	T	**D (0,04)**	**D (0,998)**	**P (0,888)**	N (-2,39)	.	**D (1)**	L (1,65)	**D (-2,28)**	**D (0,494)**	**D (0,691)**	**32**	P12883	Myosin head	4PTH	.	.	.	**Loss of a salt-bridge with Asp159**	.	**3,61**	**2,015**	**4,65**	**15,807**		
MYBPC3 NM_000256.c.A649G p.S217G	11	47371330	47371330	T	C	**D (0,01)**	**P (0,907)**	**P (0,605)**	**D (-2,69)**	.	**D (0,905)**	L (1,59)	T (0,43)	T (-0,445)	T (0,28)	13,31	.	.	.	.	.	.	.	.	.	**3,67**	0,801	**4,944**	10,479		
MYH7 NM_000257.c.C2167T p.R723C	14	23895023	23895023	G	A	T (0,25)	**D (1)**	**D (0,995)**	**D (-4,82)**	.	**D (1)**	**M (2,06)**	**D (-2,34)**	**D (0,39)**	**D (0,732)**	**19,53**	P12883	Myosin head coiled coil	4PTH	.	.	.	**Loss of interaction between two adjacent domains**	.	**3,94**	**2,531**	1,866	**13,979**		
TNPO NM_002276.c.G1696A p.D566N	12	98927731	98927731	G	A	**D (0)**	B (0,079)	B (0,088)	N (-0,75)	N (0,002)	**D (1)**	N (0,695)	T (-0,01)	T (-0,875)	T (0,157)	**15,84**	P42166	Four-stranded coiled coil	.	2V0X	79	-5,07	No predicted deleterious effect	.	**5,08**	**1,539**	3,006	10,994		
MYL2 NM_000432.c.634T p.S212C	12	111356967	111356967	C	A	T (0,08)	**D (1)**	**D (0,996)**	**D (-2,86)**	**D (0)**	**D (1)**	L (0,895)	T (-0,96)	**D (0,122)**	**D (0,553)**	14,45	.	.	.	.	.	.	.	.	.	**5,33**	**2,491**	**7,197**	**17,805**		
VCL NM_003373.c.A1507G p.H636R	10	75860740	75860740	A	G	**D (0,04)**	**D (0,989)**	**D (0,954)**	**D (-3,09)**	**D (0)**	**D (1)**	**M (2,005)**	T (0,59)	T (-0,43)	T (0,309)	**21,6**	P18206	N-terminal globular head	.	1ST6	95	-13,63	**Possible misfolding of four-helix bundle**	.	**5,04**	**2,016**	**8,659**	**15,063**		
ACTN2 NM_001103.c.C1484T p.T495M	1	236911044	236911044	C	T	T (0,06)	**D (1)**	**D (1)**	N (-2,14)	**D (0)**	**D (1)**	L (1,3)	T (0,66)	T (-0,703)	T (0,252)	**23,5**	P35609	Spectrin	1HCI	.	.	.	No predicted deleterious effect	.	**5,73**	**2,854**	**5,842**	**20,27**		
SCN5A NM_001099405.c.C2074A p.Q692K	3	38639408	38639408	G	T	**D (0,04)**	B (0)	B (0,002)	N (-1,11)	N (0,372)	N (0,958)	L (1,245)	**D (-1,68)**	**D (0,062)**	**D (0,746)**	10,67	.	.	.	.	.	.	.	.	.	1,95	0,211	-0,035	11,668		
DMD NM_004017.c.G1634A p.R545Q	X	31164440	31164440	C	T	T (0,36)	**D (0,999)**	**D (0,986)**	N (0,10)	U (0,006)	**D (1)**	N (0,2)	T (0,89)	T (-0,654)	T (0,261)	**28,5**	.	.	.	.	.	.	.	.	.	**5,53**	**2,553**	**6,826**	**18,451**		
DMD NM_004011.c.C3164T p.T1055I	X	31854848	31854848	G	A	T (0,32)	B (0,235)	B (0,144)	N (-0,73)	U (0,003)	**D (1)**	N (0,695)	T (1,33)	T (-1,081)	T (0,065)	10,65	.	.	.	.	.	.	.	.	.	**4,72**	**2,36**	3,932	**13,433**		

* Quality of the modeled comparable to an NMR structure, given its sequence identity and length

Chr: chromosome; **Ref**: reference nucleotide; **Var**: variant nucleotide; **X-ray**: PDB Code of the experimentally determined protein structure, when available; **Model Template**: PDB Code of the experimentally determined nearest homologous protein structure; **% id**: % identity between target sequence to be modeled and nearest homologous protein structure (template); **Prosa Z-score**: overall model quality according to the Prosa validation too; **IIE**: intron identity element

SIFT, FATHMM and RadialSVM predictions: **D**: Deleterious; **T**: Tolerated

Polyphen2 predictions: **D**: Probably damaging; **P**: possibly damaging; **B**: benign

Provean predictions: **D**: Deleterious; **N**: Neutral

LRT predictions: **D**: Deleterious; **N**: Neutral; **U**: Unknown

Mutation Taster predictions: **A**: disease_causing_automatic; **D**: disease_causing; **N**: polymorphism; **P**: polymorphism_automatic

Mutation Assessor predictions: **H**: high; **M**: medium; **L**: low; **N**: neutral. H/M means functional and L/N means non-functional

CADD_phred predictions: a cutoff was put at scaled C-score=15. A scaled C-score of greater of equal 15, indicates that the variant is predicted to be one of the ~3% most deleterious substitutions in the human genome

GERP++, PhyloP and SiPhy predictions: higher scores are more deleterious

2.3. Analysis of the missense variants

The effect of missense changes on the structure and function of a human protein was predicted by: (i) SIFT (Sorting Intolerant From Tolerant), (ii) PolyPhen-2 (Polymorphism Phenotyping v2) HDIV, that identifies human damaging mutations by assuming differences between human proteins and their closely related mammalian homologs as non-damaging; (iii) PolyPhen-2 HVAR, that identifies human disease-causing mutations by assuming common human nsSNPs as non-damaging; (iv) Provean (Protein Variation Effect Analyzer); (v) LRT (Likelihood Ratio Test) that identifies conserved amino acid positions and deleterious mutations using a comparative genomics data set of multiple vertebrate species; (vi) Mutation Taster; (vii) Mutation Assessor; (viii) FATHMM (Functional Analysis through Hidden Markov Models); (ix) RadialSVM (Radial Support Vector Machine); (x) LRT (Logistic Regression Test); (xi) CADD v1.3 (Combined Annotation–Dependent Depletion), a method for objectively integrating many diverse annotations into a single measure (C score) for each variant; and (xii) molecular modeling.

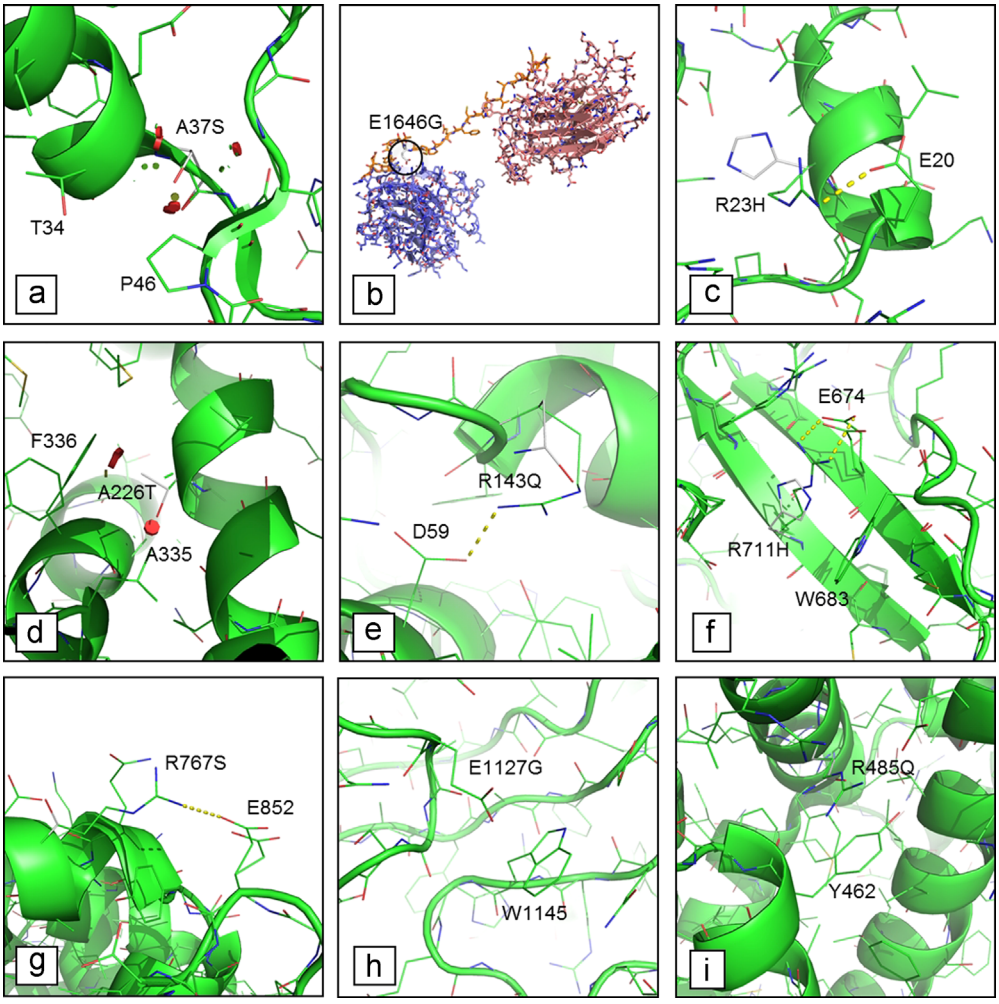

Fig. 1. Structural comparison of wild-type and mutant forms for (a) FLH2 A37S; (b) LAMA4 E1646G; (c) MYH6 R23H; (d) MYH7 A226T; (e) MYH7 R143Q; (f) MYOM1 R711H; (g) PKP2 R767S; (h) RYR2 E1127G; (i) RYR2 R485Q. The mutation is indicated in white. The predicted structural effects of mutations are: (a, d) steric hindrance (red circles); (b) local misfolding of linker domain (orange); (c, e, f, g) loss of important inter-residues contacts; (h) loss of a π-anion interaction; (i) loss of a π-cation interaction.

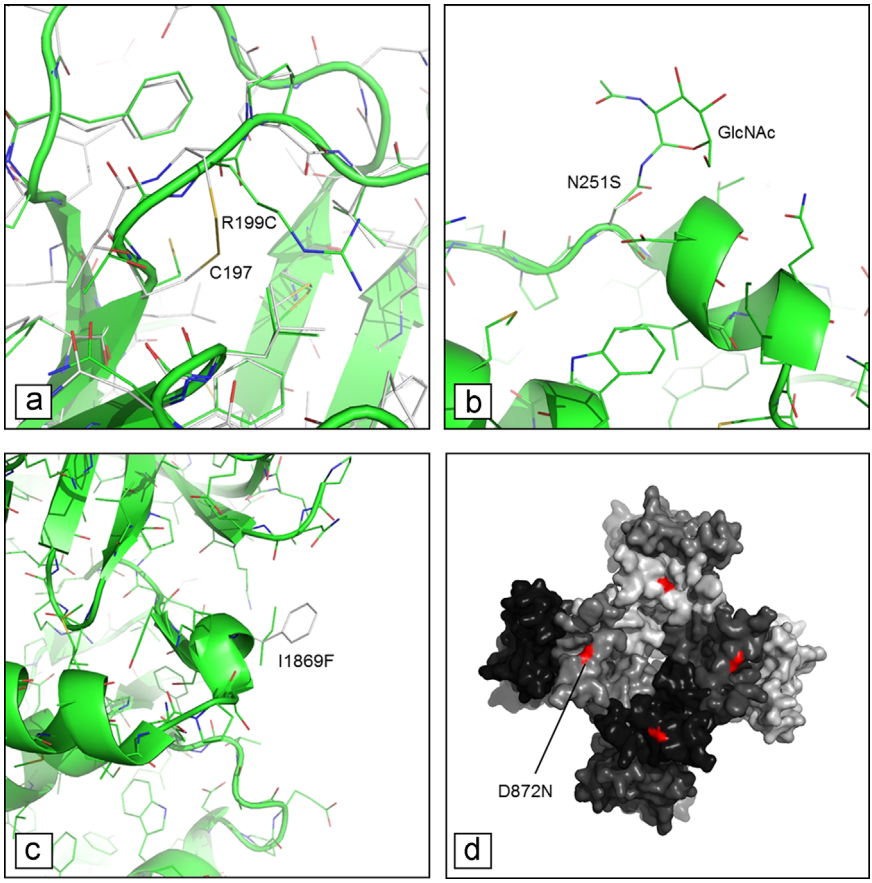

Fig. 2. Effects of nsSNVs for: (a) the cadherin domain of DSC2. The mutant R199C in the cadherin domain of DSC2 is predicted to introduce a disulfide bond with the near Cys197 residue (Cα-Cα distance ~6 Å), and possibly to result in local misfolding of the cadherin domain; (b) the melibiase domain of GLA. Mutant N215S of the melibiase domain of GLA results in the loss of a glycosylated site probably affecting the protein structure and/or function; (c) the FGF13 interaction domain of SCN5. Mutation I869F localizes on a solvent-exposed hydrophobic path of the domain of interaction with fibroblast growth factor 13 (FGF13). The I869F mutation could affect the recognition of the FGF13 protein; (d) the Na-Channel of SCN5. The mutant D872N results in the loss of a negative charge that is approximately located at the Na-channel domain of SCN5, probably affecting cations conductance of the channel. The approximate position of the negatively charged Asp872 residue is shown in red, in each of the four protein subunits forming the channel.

Regarding the molecular modeling, protein structure were experimentally determined by X-ray crystallography, or were inferred by homology modeling means (i.e., availability of a structural template with percentage of identity $>$ 20%). Protein models were built using the homology modeling approach implemented in modeler-9 package [2]. PSI-BLAST was used to find suitable structural templates for each sequence to model [3]. The sequences of each protein target to model and its structural template were then aligned by using the program CLUSTALW [4] and manually manipulated to optimize the matching of several characteristics, including the observed and predicted secondary structural elements, the hydrophobic regions in the three-dimensional structures, the structurally and functionally conserved residues, and *indel* regions in the structures. Then, ten different models were built for each target protein and evaluated using several criteria. The model displaying the lowest objective function [5], which measures the extent of violation of constraints from the structural templates, was taken as the representative model. Superimposition and root-mean-square deviation (RMSD) calculation of Cα traces of the 10 models were performed to detect the most variable and therefore less reliable modeled regions. These invariably corresponded to loop elements.

Procheck [6] was used to monitor the stereochemical quality of the representative models, whereas ProsaII [7] was used to measure the overall protein quality in packing and solvent exposure. Mutations on protein structures was carried out using the "Mutate model" script implemented in modeler-9 package [2]. The script takes as input a given three-dimensional structure of a protein (experimentally determined or predicted), and mutates a single residue. The residue sidechain's position is then optimized by energy minimization and refined by molecular dynamics simulations. Prediction of protein stability upon mutation was carried out using the DUET server [8]. Sequence identity between the modeled domain and its closest template ranged from 23% (Laminin G-like domain of LAMA4), to nearly 95% (N-terminal globular head domain of VCL). However, in spite of the low value of sequence identity measured in some cases, all of the models resulted in a good overall quality (Prosa Z-score < -2.00), except for CALR3 and SCN5. Given the short length of the predicted PB035848 domain of CALR3 (residues 294-347) and its sequence identity with its template (61%), the measured Prosa Z-score (-1.93) nonetheless indicated a model of quality comparable to a Nuclear Magnetic Resonance (NMR) structure [7] (Figs. 1 and 2).

Acknowledgments

This work funded by the Department of Molecular Medicine, Sapienza University of Rome. This work was also partially funded by the Department of Biochemical Sciences Sapienza University of Rome (prot. C26A149EC4).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.03.004.

References

[1] I. Bottillo, D. D'Angelantonio, V. Caputo, A. Paiardini, M. Lipari, C. De Bernardo, D. Giannarelli, A. Pizzuti, S. Majore, M. Castori, E. Zachara, F. Re, P. Grammatico, Molecular analysis of sarcomeric and non-sarcomeric genes in patients with hypertrophic cardiomyopathy, Gene 577 (2) (2016) 227–235.
[2] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using modeller, in: Andreas D. Baxevanis, et al., (Eds.), Current Protocols in Bioinformatics, 2006, Chapter 5 Unit 5 6.
[3] I. Friedberg, T. Kaplan, H. Margalit, Evaluation of PSI-BLAST alignment accuracy in comparison to structural alignments, protein science: a publication of the protein, Society 9 (11) (2000) 2278–2284.
[4] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic acids Res. 22 (22) (1994) 4673–4680.
[5] D.F. Burke, C.M. Deane, H.A. Nagarajaram, N. Campillo, M. Martin-Martinez, J. Mendes, F. Molina, J. Perry, B.V. Reddy, C. M. Soares, R.E. Steward, M. Williams, M.A. Carrondo, T.L. Blundell, K. Mizuguchi, An iterative structure-assisted approach to sequence alignment and comparative modeling, Proteins (Suppl 3) (1999) S55–S60.
[6] R.A. Laskowski, J.A. Rullmannn, M.W. MacArthur, R. Kaptein, J.M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR 8 (4) (1996) 477–486.
[7] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins 17 (4) (1993) 355–362.
[8] D.E. Pires, D.B. Ascher, T.L. Blundell, DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach, Nucleic Acids Res. 42 (2014) W314–W319.