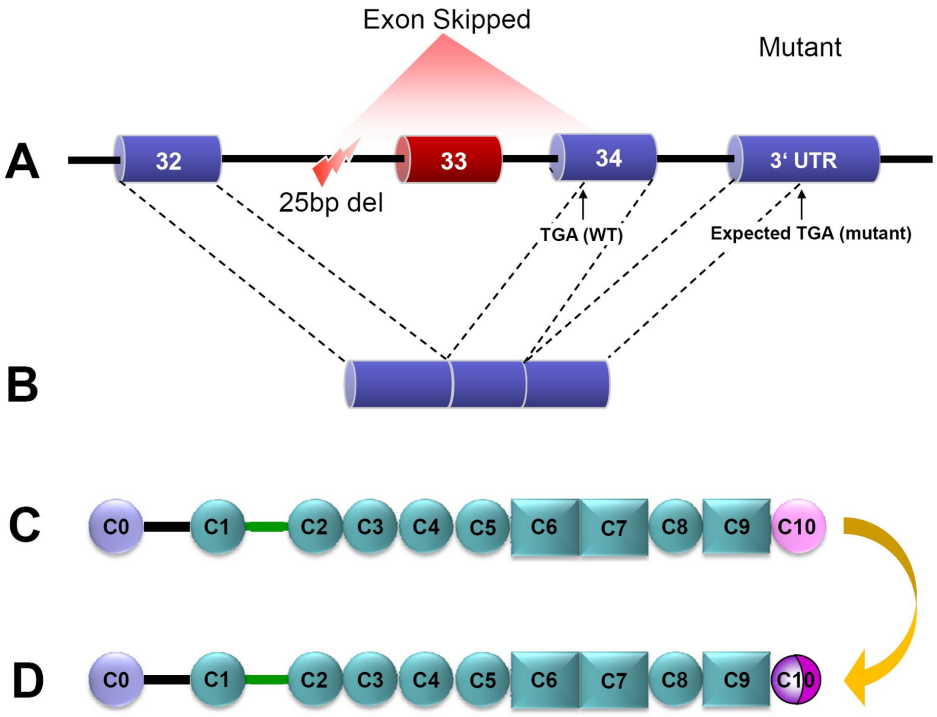
A common *MYBPC3* (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia

Perundurai S Dhandapany[1], Sakthivel Sadayappan[2], Yali Xue[3], Gareth T Powell[3], Deepa Selvi Rani[4], Prathiba Nallari[5], Taranjit Singh Rai[6], Madhu Khullar[6], Pedro Soares[7], Ajay Bahl[6], Jagan Mohan Tharkan[8], Pradeep Vaideeswar[9], Andiappan Rathinavel[10], Calambur Narasimhan[11], Dharma Rakshak Ayapati[12], Qasim Ayub[3,13], S Qasim Mehdi[13,18], Stephen Oppenheimer[14], Martin B Richards[7], Alkes L Price[15,], Nick Patterson[16], David Reich[16,17], Lalji Singh[4], Chris Tyler-Smith[3] & Kumarasamy Thangaraj[4]

[1]Department of Biochemistry, Madurai Kamaraj University, Madurai 625 021, India. [2]Molecular Cardiovascular Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. [3]Wellcome Trust Sanger Institute, Hinxton, Cambs, CB10 1SA, UK. [4]Centre for Cellular and Molecular Biology, Hyderabad 500 007, India. [5]Department of Genetics, Osmania University, Hyderabad 500 007, India. [6]Post Graduate Institute of Medical Education and Research, Chandigarh 160 012, India. [7]Faculty of Biological Sciences, Institute of Integrative and Comparative Biology, University of Leeds, Leeds LS2 9JT, UK. [8]Department of Cardiology, Sri Chitra Tirunal Institute of Medical Sciences and Technology, Thiruvananthapuram 695 011, India. [9]Department of Pathology (Cardiovascular & Thoracic Division), Seth GS Medical College and KEM Hospital, Mumbai 400 012, India. [10]Department of Cardio-Thoracic Surgery, Rajaji Government Hospital and Madurai Medical College, Madurai 625020, India. [11]Cardiology Unit, CARE Hospital, Hyderabad 500 001, India. [12]Department of Cardiac Surgery, Nizams Institute of Medical Sciences, Hyderabad 500 082, India. [13]Institute of Biotechnology and Genetic Engineering (KIBGE), University of Karachi, Karachi 75270, Pakistan. [14]Department of Anthropology, University of Oxford, Oxford, OX1 2JD, UK. [15]Harvard School of Public Health, Boston, Massachusetts 02115, USA. [16]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [17]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [18]Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi 74200, Pakistan. Correspondence should be addressed to K.T. (thangs@ccmb.res.in).

1. Supplementary Figures

Supplementary Figure 1



Supplementary Figure 1. Schematic representation of the consequences of the *MYBPC3* deletion

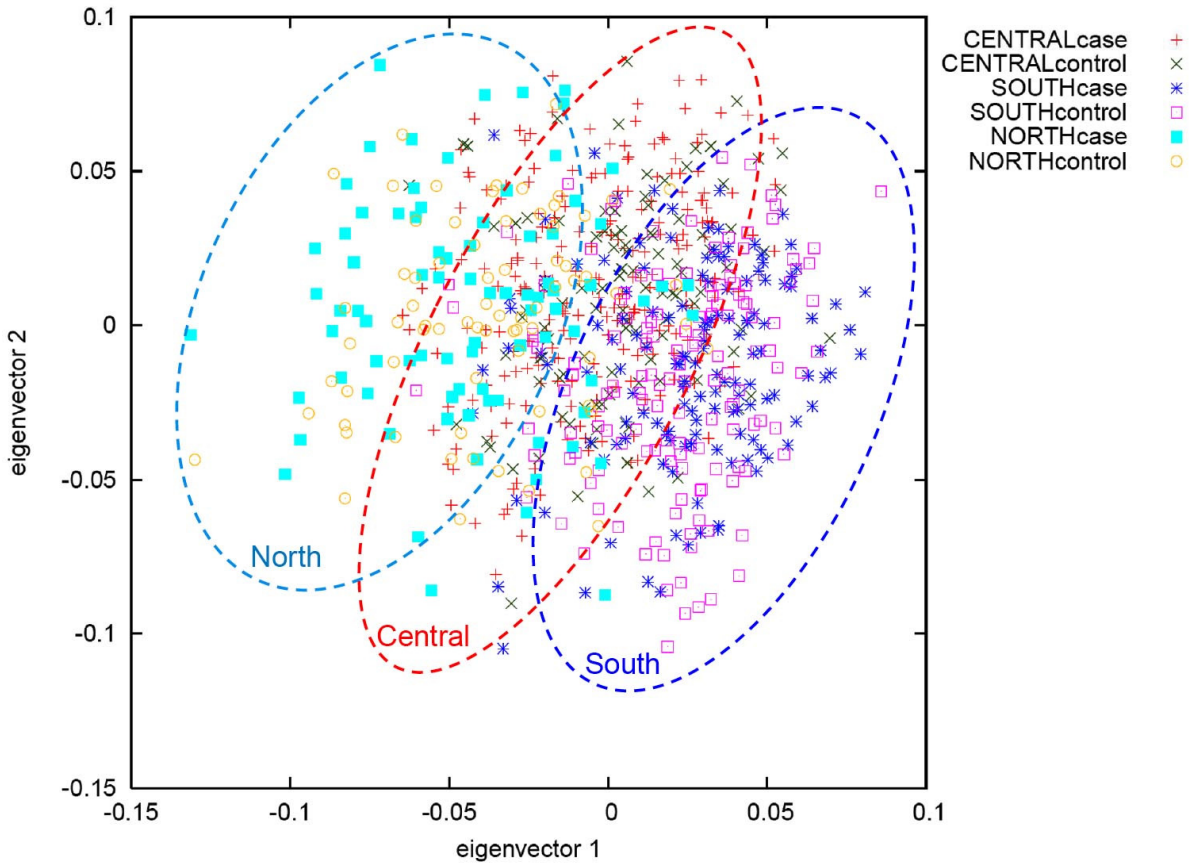
(A) Exon skipping. The deletion of 25bp in intron 32 at the branch point causes skipping of exon 33.

(B) The loss of exon 33 is associated with a shift in the open reading frame and an unscheduled translational stop in the 3'non-translated sequence

(C) Wild type MyBP-C protein

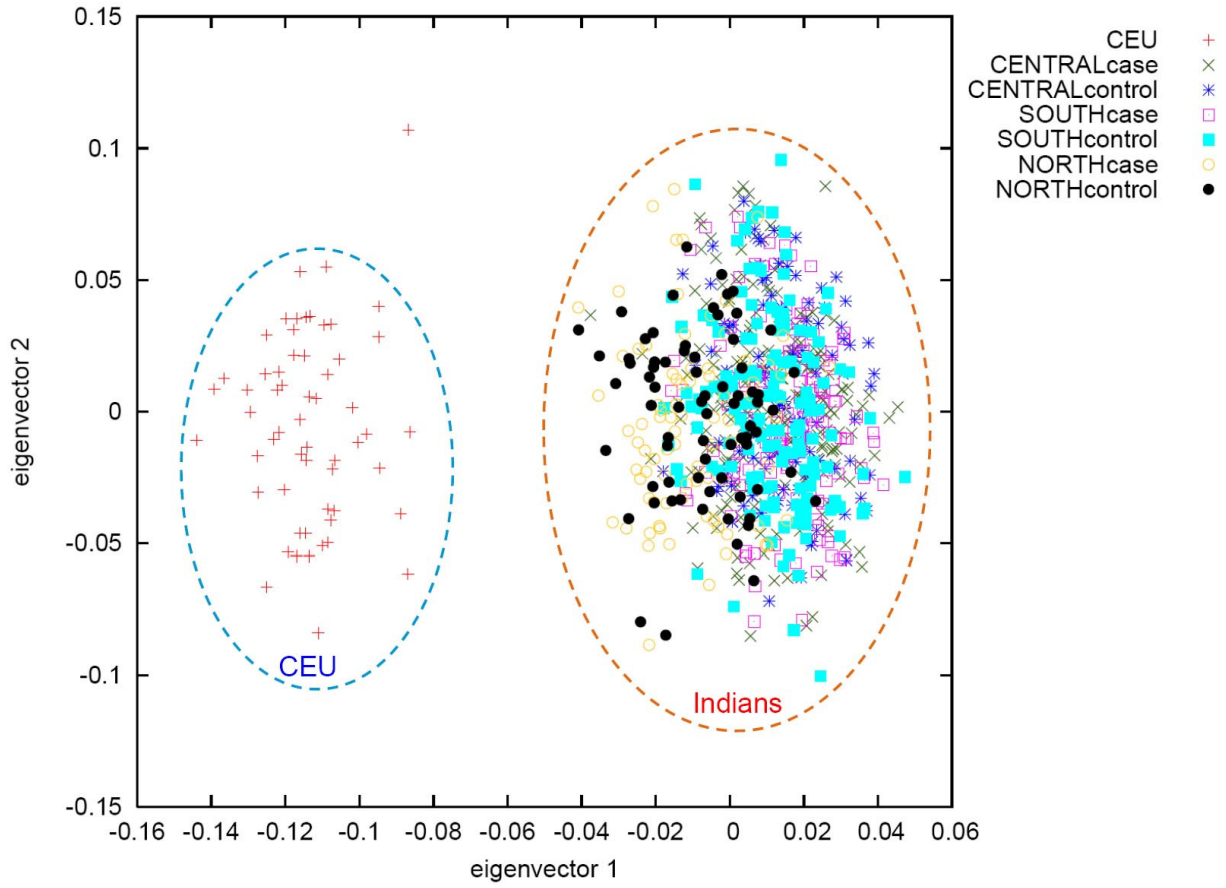
(D) The consequence of MyBP-C exon 33 skipping (due to the 25-bp deletion) is the substitution in domain C10 of 62 WT amino acids with a different 55 amino acids

Supplementary Figure 2



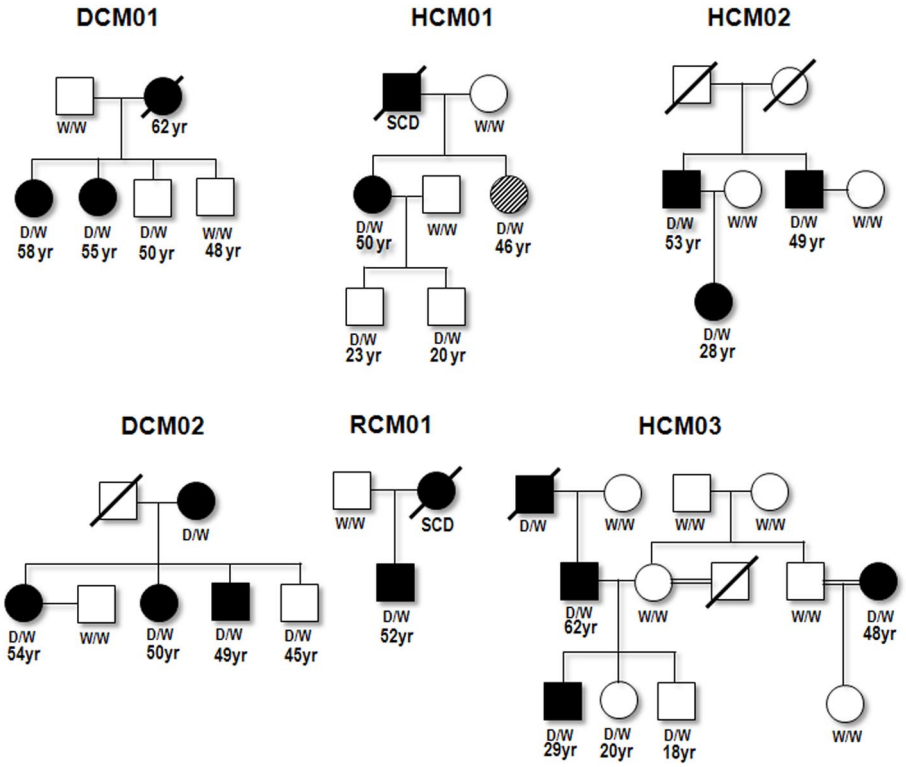
Supplementary Figure 2. Principal components analysis of South, Central and North Indian samples based on a panel of 50 ancestry-informative SNP markers (AIMs).

Supplementary Figure 3



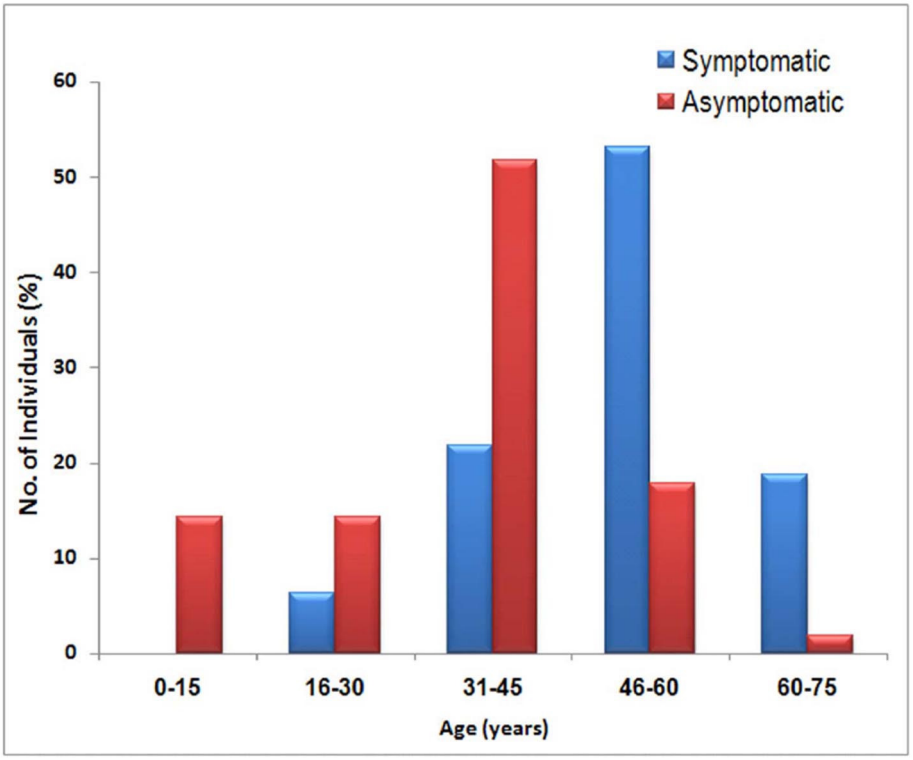
Supplementary Figure 3. Positions of South, Central and North Indian samples along the Europe-related genetic axis. We plot the top two eigenvectors of a principal components analysis of South, Central and North Indian samples together with CEU samples from HapMap. Only the first eigenvector, representing the Europe-related axis, is of interest. Note the North Indian samples were significantly closer to CEU on this axis than Central and South samples, confirming that the 50 AIMs are sufficient for detecting whether or not ancestry differences along this axis are present (Anova p-value < 1e-12)

Supplementary Figure 4A



Supplementary Figure 4A. Pedigrees of representative families with the 25 bp deletion. Squares indicate men, circles women, solid symbols affected family members, open symbols unaffected family members, and hatched symbols family members whose status is unknown; symbols with a slash indicate deceased individuals. The genetic status, when known, is shown below each individual (D indicates that the allele carries the 25 bp deletion, W that the allele does not carry 25 bp deletion, and D/W that the subject is heterozygous). SCD denotes sudden death from cardiac causes. HCM, DCM and RCM represent Hypertrophic, Dilated and Restrictive Cardiomyopathies, respectively.

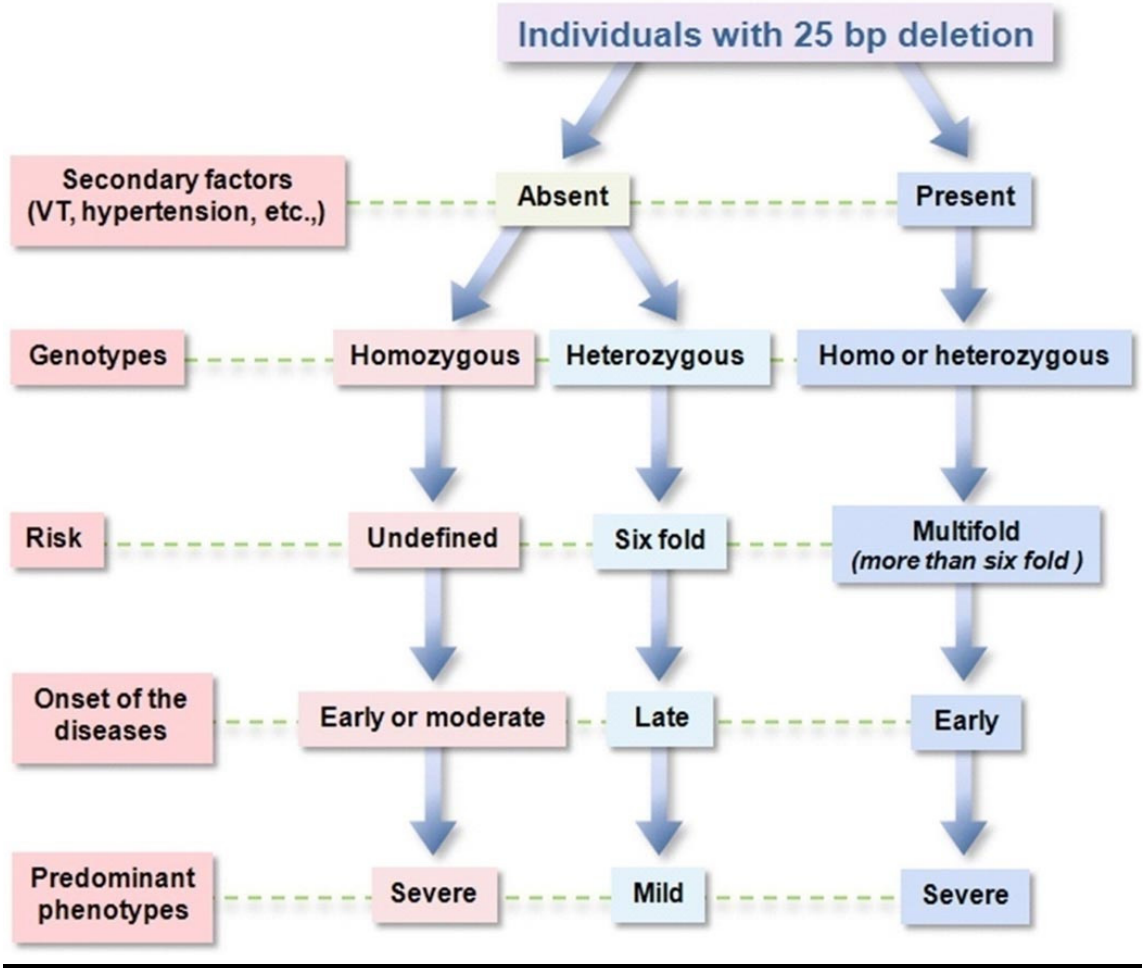
Supplementary Figure 4B



Supplementary Figure 4B. Age-dependent penetrance of the 25bp deletion among the family members of cardiomyopathy cases.

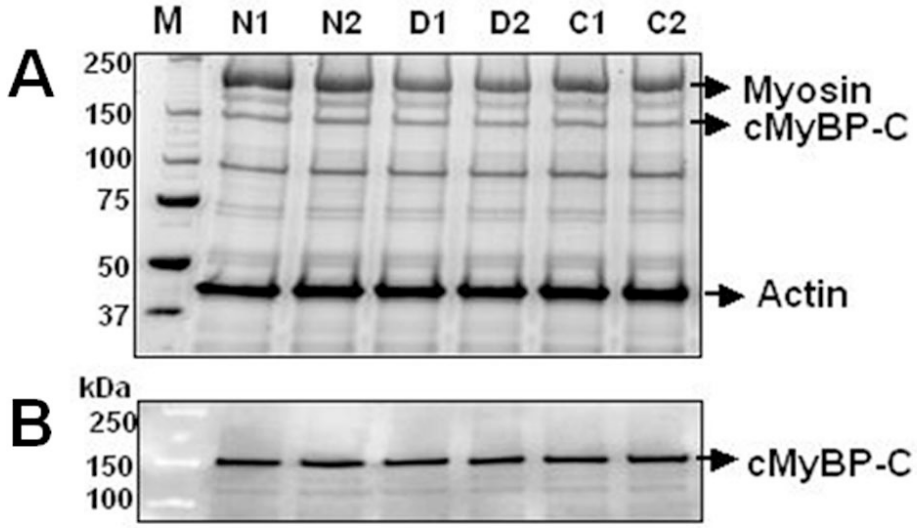
A total of eighty-eight deletion carriers (symptomatic and asymptomatic) among the 120 family members were compared among themselves for the five age classes (0-15, 16-30, 31-45, 46-60, 60-75). The selection criteria for symptomatic and asymptomatic carriers are outlined in Supplementary Table 3.

Supplementary Figure 4C



Supplementary Figure 4C. Schematic diagram showing the inferred impact of the deletion genotypes in Indian patients with cardiomyopathy

Supplementary Figure 5

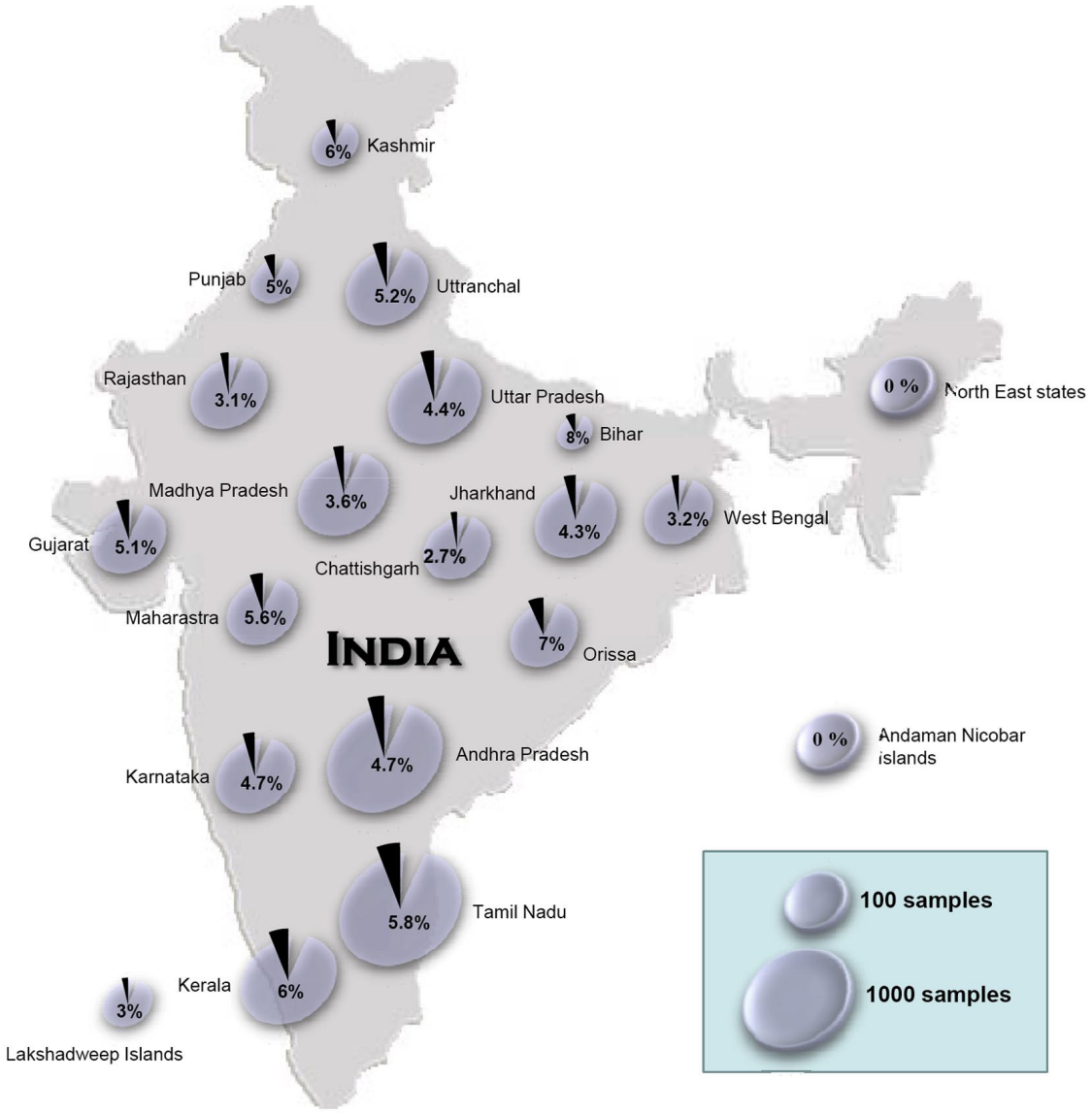


Supplementary Figure 5 . cMyBP-C protein in Myocardial biopsies

A, SDS-PAGE analysis of the total myofibrillar proteins from two different index patients with heterozygous deletions (D1 and D2). Reference samples of ventricular tissue from normal donor (N1 and N2) and patients with aortic stenosis (C1), and ventricular septum defect (C2) were used as controls.

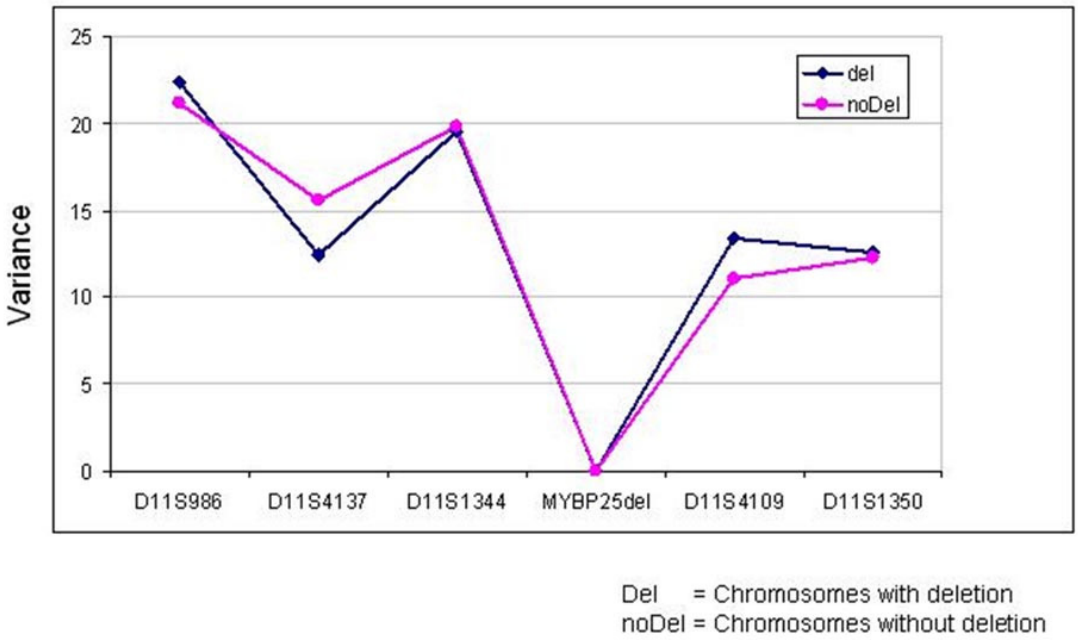
B, Western blot analyses of the same samples illustrating the absence of expected mutated protein (approx.140 kDa) in the patients with the 25bp deletion (D1 and D2). Cardiac MyBP-C was detected using a rabbit polyclonal antibody directed against C0-C1 of human cMyBP-C (S7). M, molecular weight markers.

Supplementary Figure 6



Supplementary Figure 6. Distribution of the 25 bp deletion of *MYBPC3* and its frequency in Indian States.

Supplementary Figure 7



Supplementary Figure 7. Variance of 25bp deleted and non-deleted chromosomes using five STR markers.

Supplementary Figure 8

	2428	3817	4009	4255	4555	6520	6638	7500	7927	8114	8262	8563	8668	9016	9287	9342	9665	9851	10253	10872	11138	11456	11579	11726	12723
BRU267_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
BRU310_del	C	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
BSK076_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	G	G
BSK078_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
PKH009_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	C	A	A	G	T	T	C	A	G
SDH121_del	T	A	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
SDH129_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
SDH155_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
SYD016_del	T	T	T	C	A	A	A	D	C	C	T	C	G	T	G	I	T	A	A	G	T	T	C	A	G
BRU267_undel	C	T	C	T	A	A	A	I	T	C	T	T	A	T	A	D	C	A	A	G	T	T	T	G	G
BRU310_undel	C	T	C	T	A	A	A	I	T	C	T	T	A	T	A	D	C	A	A	G	T	T	T	G	G
BSK076_undel	C	T	C	C	T	G	G	I	C	C	T	C	G	C	G	I	C	A	A	C	C	G	C	G	A
BSK078_undel	C	T	C	T	A	A	A	I	T	C	T	T	A	T	A	D	C	A	A	G	T	T	T	G	G
PKH009_undel	C	T	C	T	A	A	A	I	T	C	T	T	A	T	A	D	C	A	A	G	T	T	T	G	G
SDH121_undel	C	T	C	C	T	G	G	I	C	C	T	C	G	C	G	I	C	A	A	C	C	G	C	G	A
SDH129_undel	C	T	C	C	T	A	G	I	C	C	C	C	G	T	G	I	C	A	G	C	C	G	C	G	A
SDH155_undel	C	T	C	C	T	A	G	I	C	C	T	C	G	C	G	I	C	G	A	C	C	G	C	G	A
SYD016_undel	C	T	C	C	T	A	G	I	C	A	T	C	G	C	G	I	C	A	A	G	C	G	C	G	A

Supplementary Figure 8. Details of SNPs and insertion/deletion polymorphism

2. Supplementary Tables

Supplementary Table 1. Baseline characteristics of the patients studied

Characteristics of the Patients	**Group 1 (n= 321 cases)**	**Group-2 (n= 446 cases)**
Age (yrs)	48±8	49± 12
Male sex (%)	67	64
NYHA class III or IV (%)	35	40
Dyspnea (%)	70	66
Angina pectoris (%)	52	57
Syncope (%)	27	32
Hypertrophic Cardiomyopathy (n)	193	164
LVEDD, mm	35 ±7.41	37.00 ±5.69
LVESD, mm	17±8.15	18± 5.2
Septum, mm	22.4±5.32	23.12± 3.17
Abnormal ECG (%)	68	62
Dilated Cardiomyopathy (n)	118	277
LVEF at the time of enrollment	32 ± 12	29.57 ± 12.09
Restrictive Cardiomyopathy (n)	10	5
LA diameter, mm	49 ± 5.9	48.6 ± 2.9
LVEDD, mm	38.8± 13.3	35.8± 6
LVEF, (%)	58.20± 3.24	58.00± 3.24
Abnormal ECG (%)	85%	90%

Supplementary Table 2: Ancestry Informative Markers used in the present study

S.No	SNPs	Chromosome position	Genetic position	Physical position	Reference allele	Variant allele	CEU frequency	Indian Frequency
1	rs12913832	15	0.294016	26039213	A	G	0.216	0.935
2	rs1869829	2	1.512994	135711294	C	T	0.198	0.829
3	rs8050463	16	0.577824	30461660	A	G	0.181	0.785
4	rs6974649	7	1.385698	130270352	G	T	0.284	0.884
5	rs1257000	2	1.144331	97100193	A	G	0.871	0.279
6	rs1481356	1	2.378998	215297063	A	G	0.802	0.209
7	rs7127335	11	0.572031	38559782	A	G	0.267	0.855
8	rs1378942	15	0.891213	72864420	A	C	0.672	0.105
9	rs7034211	9	0.298405	14127320	C	T	0.259	0.831
10	rs4944957	11	0.782314	70845683	C	T	0.754	0.169
11	rs10800779	1	2.129964	198132682	C	T	0.19	0.767
12	rs1330303	9	0.355379	16705826	C	T	0.698	0.14
13	rs12416652	10	1.007332	77778952	C	T	0.198	0.765
14	rs2710994	7	0.42602	24756695	A	G	0.078	0.616
15	rs611555	6	0.183087	7020821	A	G	0.284	0.833
16	rs12945601	17	0.434115	17594136	C	T	0.733	0.18
17	rs12550668	8	0.242301	11617363	C	T	0.621	0.122
18	rs271762	1	2.55734	230976582	C	T	0.328	0.849
19	rs7999348	13	0.940914	98730923	C	T	0.207	0.738
20	rs1992906	5	1.874684	173115417	A	G	0.888	0.366
21	rs2158298	17	0.636443	36325760	A	G	0.655	0.141
22	rs756411	16	0.175355	7549913	A	G	0.121	0.657
23	rs10932521	2	2.174991	214694847	A	G	0.328	0.86
24	rs9655117	7	0.276422	14724599	A	G	0.069	0.558
25	rs998382	20	0.516865	30847797	C	T	0.293	0.819
26	rs1941411	11	1.280712	119635722	C	T	0.276	0.802
27	rs6890009	5	1.342041	131607932	G	T	0.371	0.866
28	rs10884942	10	1.3497	112006916	A	G	0.741	0.218
29	rs1981760	16	0.616971	49280575	C	T	0.224	0.75
30	rs594387	1	0.740561	46925406	C	T	0.155	0.69
31	rs7018694	9	0.409581	20242341	A	G	0.293	0.808
32	rs541805	9	0.492671	25924932	A	G	0.888	0.355
33	rs4759128	12	0.704256	53651487	A	G	0.905	0.401
34	rs7120706	11	1.232979	116335297	C	T	0.759	0.238
35	rs11240351	1	2.195556	201775996	C	T	0.448	0.912
36	rs5007171	3	0.664617	45124712	C	T	0.14	0.669
37	rs8106293	19	0.560374	38228216	C	T	0.767	0.262
38	rs1555741	1	0.680782	39162214	A	G	0.103	0.622
39	rs8081017	17	1.003372	67980475	C	T	0.19	0.726
40	rs16954273	16	0.709095	53694406	A	G	0.948	0.459
41	rs543759	18	0.233667	7498561	A	G	0.746	0.235
42	rs2232704	14	0.937742	93825872	C	T	0.828	0.292
43	rs9893521	17	0.127596	4946472	A	G	0.862	0.372
44	rs4685803	3	0.117294	4715592	A	G	0.336	0.847
45	rs11774017	8	1.089252	98297141	A	G	0.655	0.169
46	rs623345	6	1.294446	128148100	A	G	0.267	0.756
47	rs12595616	15	1.075976	89364517	C	T	0.307	0.791
48	rs3844335	8	0.800159	62356446	C	T	0.103	0.612
49	rs9376327	6	1.046437	100194102	C	T	0.974	0.541
50	rs12757113	1	0.517879	26281286	C	T	0.259	0.768

Supplementary Table 3. Genotypic-phenotypic correlation between symptomatic deletion carriers, asymptomatic carriers and non-carriers among the 120 members from 28 family affected with cardiomyopathy.

Patients' details	Symptomatic carriers (n=32)	Asymptomatic carriers (n=56)	Non carriers (n=32)
Age	55.7 ± 15.3	40 ± 14.1	40 ± 12.6
Sex, Males, %	45	55	35
Dyspnea, %	100	40	0
Angina Pectoris, %	42	22	0
Syncope, %	63	15	0
Hypertension, %	12	16	28
LVEDD, mm	38.2 ± 6.2	44 ± 4.3	44 ± 3.3
LVESD, mm	20.3 ± 3.8	26.3 ± 3.5	25.2 ± 3.4
Septum, mm	23.2 ± 4.2	11.6 ± 2.1	10.3 ± 2.4
LVEF, % (in DCM)	32 ± 8.2	53.6 ± 11.3	64.0 ± 6
Abnormal ECG, %	90	45	0
LVH, %	60	10	0
Aberrant T waves, %	40	12	0
Altered Q waves, %	10	5	0

Values are mean ±SD unless otherwise indicated. LVEDD, left ventricular end-diastolic diameter; LVESD, left ventricular end-systolic diameter

Supplementary Table 4. Details of the Indian populations screened for the 25bp deletion. Homozygous deleted individuals are shown in parenthesis.

Indian Populations	Total	Deleted	Frequency
South India			
TAMIL NADU			
Marvar	50	3	0.06
Kallar	60	7 (1)	0.11
Irulas	85	7	0.08
Kurumba	32	3	0.09
Agamudaiyar	48	3	0.06
Paniyan	50	3	0.06
Toda	100	7	0.07
Kota	74	4	0.05
Baduga	60	2	0.03
Chakkiliar	65	4	0.06
Khani	50	2	0.04
Malayali	90	5	0.05
Parayar	60	4	0.06
Paliyar	57	3	0.05
Malaikurvar	24	1	0.04
Kondareddy	16	0	-
Arundathi	40	2	0.05
Lingayath	85	3	0.03
Anglo-Indian	34	0	-
KERALA			
Oorali	47	2	0.04
Kattunayakkan	97	4 (1)	0.04
Kurcha	33	2	0.06
Kuruman	30	2	0.06
Muslims and others	200	15	0.07
Malayan	100	5	0.05
Ulladan	50	3	0.06
Karimbala	53	4	0.07
KARNATAKA			
Halakki	40	2	0.05
Kunabi	44	2	0.04
ANDHRA PRADESH			
Goud	100	7	0.07

Kolma	48	1	0.02
Yanadi	43	2	0.04
Chenchu	28	1	0.03
Siddi (African settlers)	20	0	-
Erkula	79	5 (1)	0.06
Thotti	54	4	0.07
Gowda	49	2	0.04
Muslims	35	1	0.02
Velamma	47	2	0.04
Padmashali	44	2	0.04
Sagara	60	3	0.05
Sugali	47	2	0.04
Madiga	67	3	0.04
Lambadi	46	2	0.04
Vysyas	35	1	0.02
West India			
MAHARASTRA			
Mahadekoli	89	4 (1)	0.04
Thakar	90	7	0.07
Parsi	50	2	0.04
Katkari	18	1	0.05
GUJARAT			
Rathwa	50	3	0.06
Koli	69	5	0.07
Tadvi	47	2 (1)	0.04
Girasia	49	2	0.04
Vatika Gond	50	2	0.04
Bhils	48	2	0.04
Central India			
MADHYA PRADESH			
Bhilala	53	1	0.01
Barela	20	7	0.35
KOL	96	3 (1)	0.03
Bhils	40	1	0.02
Gond	42	1	0.02
Mawasi	96	2	0.02
Sahariya	36	1	0.02
Keers	14	0	-
Korku	6	0	-
Baiga	30	1	0.03
North India			

KASHMIR			
Pandits	49	3	0.06
UTTRANCHAL			
Tharus	30	1	0.03
Buxas	48	2	0.04
UTTAR PRADESH			
Brahmin	35	2	0.05
Sidhi	22	1	0.04
Muslim	24	3 (1)	0.12
Haburra	48	1	0.02
Chipi & lakodiga	91	2	0.02
Rajput	86	7	0.08
Diwari	99	3	0.03
Yadava	64	2	0.03
Saharia	12	0	-
Rohidas	50	2	0.04
Kori	50	3	0.06
Lodhe	26	1	0.03
PUNJAB			
Ramgarhia	58	4	0.06
RAJASTHAN			
Bheels	50	2	0.04
Meena	50	1	0.02
Meghwal	26	1	0.03
CHATTISGARH			
Sathnami	64	2	0.03
Santhal	52	1	0.01
Gond	44	1	0.02
Kanwar	22	1	0.04
HIMALAYAS			
Bhist	112	5	0.04
Bhotia	58	0	-
Random samples	264	15	0.05
East India			
WEST BENGAL			
Random Samples	125	4	0.03
ORISSA			
Oraon	32	4	0.12
Kissan	14	1	0.07
Dhurva	33	2	0.06
Bonda	45	4	0.08

Gadaba	31	0	-
BIHAR			
Oraon	48	4	0.08
JHARKAND			
Ho	66	4	0.06
Munda	50	1	0.02
Oraon	89	4	0.04
North East India			
NAGALAND			
Chakhesang-Naga	55	0	-
Ao-Naga	80	0	-
Naga-Sema	53	0	-
ARUNACHAL PRADESH			
Nyshi	50	0	-
MIZORAM			
Mizo	45	0	-
North East (Random Samples)	160	0	-
ANDAMAN ISLANDS			
Great Andamanese	20	0	-
Nicobarese	48	0	-
Onges	46	0	-
LAKSHADWEEP ISLAND	80	4 (1)	0.05
	6273	287 (8)	0.04

Supplementary Table 5. Global distribution of the *MYBPC3* deletion in a panel of 8,358* individuals from around the world

Population Number	Population Names	Country	Total Number of Samples	Deleted	Frequency
1	Bantu-speakers	South Africa	8	0	0
2	San	Namibia	6	0	0
3	Northeastern Bantu	Kenya	12	0	0
4	Mbuti Pygmies	Democratic Republic of Congo	11	0	0
5	Biaka Pygmies	Central African Republic	34	0	0
6	Yoruba	Nigeria	23	0	0
7	Madenka	Senegal	18	0	0
8	Mozabite	Algeria (Mzab region)	26	0	0
9	Bedouin	Israel (Negev region)	39	0	0
10	Palestinian	Israel (Central)	47	0	0
11	Druze	Israel (Carmel region)	41	0	0
12	Tuscan	Italy	8	0	0
13	Sardinian	Italy	26	0	0
14	North Italian	[Italy (Bergamo region)	14	0	0
15	French	France	28	0	0
16	French Basque	France	24	0	0
17	Orcadian	Orkney Islands	15	0	0
18	Russian	Russia	25	0	0
19	Adygei	Russia (Caucasus region)	17	0	0

20	Makrani §	Pakistan	63	1	0.01
21	Brahui	Pakistan	95	7	0.07
22	Balochi	Pakistan	32	0	0
23	Hazara	Pakistan	31	0	0
24	Pathan	Pakistan	105	3	0.02
25	Kalash	Pakistan	56	0	0
26	Burusho	Pakistan	91	9	0.09
27	Kashmiri	Pakistan	20	0	0
28	Mohanna	Pakistan	56	3	0.05
29	Sindhi	Pakistan	143	9	0.05
30	Parsi	Pakistan	78	0	0
31	West Indians	India	560	30	0.05
32	North Indians	India	1534	67	0.04
33	North East Indians	India	443	0	0
34	East Indians	India	533	28	0.05
35	Central Indians	India	433	16	0.03
36	South Indians	India	2656	150	0.05
37	Sri Lankans	Sri Lanka	21	2	0.09
38	Andamanese	Andaman Islands	114	0	0
39	Senoi	Malaysia	54	0	0
40	Semang	Malaysia	110	0	0
41	Aboriginal Malays	Malaysia	108	4	0.03
42	Tengger	Indonesia	22	1	0.04
43	Bali	Indonesia	15	0	0
44	Toraja	Indonesia	24	0	0
45	Papuan	New Guinea	16	0	0
46	NAN Melanesian	Boungainville	21	0	0
47	Australian Aborigine	Australia	6	0	0
48	Cambodian	Cambodia	11	0	0
49	Miaozu	China	10	0	0
50	Dai	China	10	0	0

51	Lahu	China	10	0	0
52	Yizu	China	10	0	0
53	She	China	10	0	0
54	Naxi	China	10	0	0
55	Tujia	China	10	0	0
56	Han	China	44	0	0
57	Tu	China	10	0	0
58	Uygur	China	10	0	0
59	Mongola	China	10	0	0
60	Oroqen	China	10	0	0
61	Xibe	China	9	0	0
62	Daur	China	10	0	0
63	Hezhen	China	9	0	0
64	Japanese	Japan	30	0	0
65	Yakut	Russia (Siberia region)	22	0	0
66	Pima	Mexico	19	0	0
67	Maya	Mexico	24	0	0
68	Colombian	Colombia	13	0	0
69	Surui	Brazil	20	0	0
70	Karitiana	Brazil	23	0	0

*The study population also includes 120 samples from the HapMap project including JPT, CEU, YRI and HCB.

[§] Includes both Negroid (38) and Baloch (25) samples. The deletion is observed in Negroid (1/38).

Supplementary Table 6. Sequences of primers used for STR analysis

Primer	Forward	Reverse
1350	HEX: CATTCTGGGGTCTTTGAT	CGTGTGTGTGTATGTGGG
4109	TET: CTGGGAGTTAGGAGACCTGG	TTGAAGAGCCCTCACAGAC
1344	HEX: TGGCAGTTATGGTTGTCTTC	GCGCCTGGCTTGTACATATA
4137	FAM: ATTGCTTGGACCCAGG	TGATAAAGACATACCCGAGG
986	FAM: GAAGGACTCGGCTCCAG	GTAAGAGGATGGTAGGAGGG

Supplementary Table 7. *MYBPC3* allele-specific long PCR primers:

Primer	Forward	Reverse
LP_1	CTG GAT GGC TTC CCT CCC TCT C	ACA GTT CGA GGC GTC TTT GT
LP_2	TGT ACA AAG GGG TGG AGG AG	AGG GAG GGA AGC CAT CCA GGC
DLP_3	CCC TGC CAG GTC CCC TCT CTT T	ACA GTT CGA GGC GTC TTT GT
DLP_4	TTT GGG TGT GGA GAA TGA AT	CGG GCT ATA AAT AAG GTA AAG AGA G

Supplementary Table 8. Primers used for nested amplification and sequencing of the region upstream of the 25 bp deletion of the *MYBPC3* gene.

Primers	Sequence	Melting Temperature (°C)	Position on Chr 11 (bp)	Length (bp)	Amplicon size (bp)	Overlap Between Amplicons (bp)
MYUP1F	GGTGTGGAGAATGAATGAAACAT	60.11	47316281	23	614	
MYUP1R	ACCGGTAGCTCTTCTTCTTCTTG	60.415	47315668	23		
MYUP2F	AGAAAGATCTCTTTGGATCTGGG	60.082	47316054	23	650	340
MYUP2R	AAGGAGACTGTGGATGTGAGTGT	60.088	47315405	23		
MYUP3F	AGCGCAAGAAGAAGAAGAGCTAC	60.789	47315694	23	513	243
MYUP3R	CTGGTCATCCCATATCATCCTAA	60.048	47315182	23		
MYUP4F	CAGCCATGCTAGACACTCACAT	60.343	47315439	22	542	213
MYUP4R	ACAGTCTCAACACCTGTTCTGC	59.425	47314898	22		
MYUP5F	TGGGAATAATGACACTGCCTATC	60.218	47315238	23	535	294
MYUP5R	AGAACTGAACAAACCAGGAATCA	60.026	47314704	23		
MYUP6F	ATCTGTCTGTGGAAAGAGTCCTG	59.793	47315052	23	625	302
MYUP6R	AACATGGTGAAACTCCGTCTCTA	60.047	47314428	23		
MYUP7F	CTGATTCCTGGTTTGTTCAGTTC	60.026	47314727	23	538	254
MYUP7R	AGATCCATGCCCTAGACTCTGA	60.233	47314190	22		
MYUP8F	TTGTAGAGACGGAGTTTCACCAT	60.047	47314453	23	588	217
MYUP8R	CAGGACAGATATTTCTCTGGGTG	60.005	47313866	23		
MYUP9F	AGTTCAGAGTCTAGGGCATGGAT	60.511	47314372	23	666	302
MYUP9R	AGTTCAAGTGATTCTCCTGCCTC	61.144	47313707	23		
MYUP10F	TTGAGCAGATTCACCTGTAGTCA	59.928	47313946	23	540	351
MYUP10R	CCTCTGATAGGAATCTCCAGGAT	59.94	47313407	23		
MYUP11F	AGTGCGAAATCAGCAACTACATT	60.203	47313792	23	518	339
MYUP11R	GTCCTTCACCAGTATCGATGTGT	60.303	47313275	23		
MYUP12F	GAGGCAGGAGAATCACTTGAACT	61.144	47313571	23	624	250
MYUP12R	CTAAAGTGCTGGGATTACAGGCT	60.984	47312948	23		
MYUP13F	ACACATCGATACTGGTGAAGGAC	60.303	47313297	23	641	303
MYUP13R	AGGGACTGACATACACCAAGAGA	60.045	47312657	23		
MYUP14F	GATCACTTGAGGTCAGGAGTTTG	60.16	47312931	23	564	228
MYUP14R	CAACTTGACTGAAGTCACACAGC	60.006	47312368	23		
MYUP15F	TCAGTCCCTTAGTTCTGGAACAG	59.797	47312665	23	678	251
MYUP15R	GATGGGAACAACACACTATAGCC	59.781	47311988	23		
MYUP16F	TCAGTCAAGTTGCTTACCCTCTC	59.936	47312379	23	672	346
MYUP16R	CCTTGTCCTCCATGTTCTCAAT	60.361	47311708	22		
MYUP17F	GCTATAGTGTGTTGTTCCCATCC	59.781	47312009	23	615	256
MYUP17R	ACTCCAGAGCCACATTAAGACC	59.643	47311395	22		
MYUP18F	ATTGAGAACATGGAGGACAAGG	60.361	47311729	22	542	289
MYUP18R	CCCTACTATGGAGGGATTCAGAT	59.713	47311188	23		
MYUP19F	AGGATCTCCGGGTGACTGAC	61.48	47311442	20	655	211
MYUP19R	AGAAGTCCAGGGCCTTATAGTTG	60.031	47310788	23		

Supplementary Table 9. Primers used for amplification and sequencing of the region downstream of the 25 bp deletion of the *MYBPC3* gene.

Primers	Sequence	Melting Temperature (°C)	Position on Chr 11 (bp)	Length (bp)	Amplicon size (bp)	Overlap Between Amplicons (bp)
MYDOWN1F	ACTCTGGAGATTAGAAAGCCCTG	60.261	47310304	23	557	
MYDOWN1R	CAGGAGACACACTTGTCACACAT	60.127	47309748	23		
MYDOWN2F	AGATGGGCAATAGCTTCCAGA	61.095	47310069	21	556	277
MYDOWN2R	ACACCGAAATTGAGAAGAGTGAG	59.803	47309514	23		
MYDOWN3F	ATGTGTGACAAGTGTGTCTCCTG	60.127	47309770	23	627	210
MYDOWN3R	TATCTGCTCTCACTTTGCTGTGA	60.204	47309144	23		
MYDOWN4F	ACTCACTCTTCTCAATTTCGGTG	59.803	47309537	23	637	347
MYDOWN4R	ATCACAACACTGAGATTCTGCCT	60.189	47308901	23		
MYDOWN5F	TTGATCTTCACAGCAAAGTGAGA	60.045	47309173	23	596	226
MYDOWN5R	TGTCTTCCACTGCATAGAACTCC	60.675	47308578	23		
MYDOWN6F	ATCTCAGTGTTGTGATTGTGCTG	60.23	47308916	23	662	293
MYDOWN6R	GAAGGAAGGAAGGAGATGAGGT	60.075	47308255	22		
MYDOWN7F	TCTATGCAGTGGAAGACACTCCT	60.316	47308595	23	532	294
MYDOWN7R	GGTATGGGCAAGAAATGACTGTA	60.25	47308064	23		
MYDOWN8F	TAATCTCAGTTCAGGTGAGGACG	60.666	47308461	23	544	353
MYDOWN8R	AGTGGTCGAGGCTAGTGGAGT	60.324	47307918	21		
MYDOWN9F	AGGTGTCATTTAACACAAGGCAG	60.45	47308195	23	644	231
MYDOWN9R	TTTGTGTCCTTGAGACATTTGTG	60.06	47307552	23		
MYDOWN10F	ACTCCACTAGCCTCGACCACT	60.324	47307938	21	513	342
MYDOWN10R	ATGTGTCCCTCTGAGTGTGTCTT	60.088	47307426	23		
MYDOWN11F	ACTCCTCACAGAGTGAAACTGCT	59.606	47307691	23	580	219
MYDOWN11R	CTTCACATATGGAGATGGAGGAG	59.969	47307112	23		
MYDOWN12F	AAGACACACTCAGAGGGACACAT	60.088	47307448	23	641	290
MYDOWN12R	AGCCACAAGTACAAGACACCAAT	59.978	47306830	23		
MYDOWN13F	CACGATACTGAGACTCCTCCATC	60.146	47307147	23	627	293
MYDOWN13R	GTAGAATCACGGCCATCAAGTAA	60.377	47306521	23		
MYDOWN14F	ATTGGTGTCTTGTACTTGTGGCT	59.978	47306830	23	520	263
MYDOWN14R	ACGGGAACTGTGTTTAACATGAC	60.199	47306311	23		
MYDOWN15F	CAGCAGCTTAGGCTATGACAGG	61.423	47306658	22	584	303
MYDOWN15R	GGCTGTGGACATGTGTGTATTT	59.786	47306075	22		
MYDOWN16F	AGCTCTCTCTAGCTTGTGTCCAA	59.846	47306400	23	579	279
MYDOWN16R	GCTTCTTGCTCTACCTTTGATGA	60.035	47305822	23		
MYDOWN17F	AAATACACACATGTCCACAGCC	59.786	47306096	22	591	229
MYDOWN17R	GCTGTGTTTGTGCTTCATTCTTT	60.694	47305506	23		
MYDOWN18F	TCATCAAAGGTAGAGCAAGAAGC	60.035	47305844	23	688	292
MYDOWN18R	GTTTCGAGTTTAGAGGTGGTCCT	60.049	47305157	23		
MYDOWN19F	TAGACTCCTCTGTCCCTCAAGTG	59.924	47305424	23	685	221
MYDOWN19R	CTTCTGTCCTCTCTTTGGGAAAT	60.115	47304740	23		

3. Supplementary Methods

Clinical Evaluation of Patients and Control Subjects

A total of 800 registered cardiomyopathy patients (Group 1 and 2) from (1) Madurai Rajaji Hospital, Madurai; (2) Sri Chitra Tirunal Institute of Medical Sciences and Technology, Trivandrum; (3) Care Hospital, Hyderabad; (4) Nilofar; (5) Krishna Institute of Medical Science, Hyderabad; (6) Nizam Institute of Medical Sciences, Hyderabad; (7) Government Medical College Hospital, Kozhikode (representing South India 1, 2 and 3, South East and West); (8) Post Graduate Institute of Medical Education and Research, Chandigarh; and (9) Seth GS Medical College and KEM Hospital, Mumbai (representing North 1, 2, Central and West India) were recruited with informed written consent (Table 1 and 2). In addition, 33 postmortem samples with heart failure were also included in this study. The institutional review boards of the study centers approved the Protocol. The cardiomyopathy cases include all the three major forms of cardiomyopathies (hypertrophic, dilated and restrictive).

Diagnostic criteria of the index patients

***Hypertrophic cardiomyopathy*:** Individuals were diagnosed with HCM when echocardiography identified unexplained left ventricle hypertrophy greater than or equal to 13 mm, or they fulfilled proposed diagnostic criteria for HCM within the context of familial disease i.e., echocardiographic criteria and/or one of the following ECG abnormalities: left ventricular hypertrophy (Romhilt-Estes score ≥4); Q-waves (duration >0.04 sec and/or a depth >1/4 of ensuring R wave in at least two leads); and marked repolarization abnormalities (T-wave inversion in at least two leads).

***Restrictive cardiomyopathy*:** Individuals were diagnosed with RCM in accordance with previous established criteria [1]. when echocardiography revealed the following: Doppler measurements consistent with restrictive left ventricular filling pattern (increased ratio of early diastolic filling [E] to atrial filling [A] ≥ 2; decreased deceleration time [DT] < 150 ms; decreased isovolumic relaxation time [IVRT] < 70 ms); reduced or low-normal left ventricular end diastolic dimension (LVED); normal or near-normal left ventricular wall thickness and systolic function.

***Dilated cardiomyopathy*:** Individuals were diagnosed with DCM when echocardiographic demonstration of depressed systolic function of the left ventricle (LV) (LV ejection fraction (LVEF) <0.45 and/or fractional shortening <0.25) and a dilated LV (LV enddiastolic dimension >117% of the predicted value corrected for age and body surface area) in the absence of other cardiac or systemic causes) were noticed [2]

Details of the patients

The details of the patients are given in Supplementary Table 1. Details of the classification and phenotyping of the postmortem cases have been published previously [3]. Archived, paraffin-embedded myocardial tissue blocks were the source of DNA for genotyping.

Selection of controls

In group 1, the patients and controls were matched with respect to geographical region (controls were healthy volunteers from the hospitals mentioned above and were unrelated to the cardiomyopathy patients), ethnicity (self reported), sex (67% males vs 64% males), age (48±8 yrs vs 47±10 yrs), ECG parameters (85 % abnormal vs 100% normal) and Echocardiography (IVS; 22.4±5.32 mm vs 10.5±2 mm, LVEF; 32 ± 12 vs 58± 6.3). In group 2, the patients and controls were again matched for geographical region (as outlined in group 1), ethnicity (self reported), sex (64% males vs 68% males), age (49± 12yrs vs 48±8 yrs). The

controls were apparently healthy with no familial history or symptoms of cardiovascular diseases.

Population stratification analysis using 50 ancestry-informative markers (AIMs) and Mantel-Haenszel test

We ascertained 50 AIMs for inferring ancestry along a Europe-related axis using 86 samples from India, genotyped on the Affymetrix 6.0 chip (7 Bhil, 9 Tharu, 7 Meghawal, 10 Kurubma, 7 Hallaki, 5 Pandit, 4 Vaish, 2 Srivastava, 5 Vysya, 5 Naidu, 4 Velama, 5 Lodi, 4 Satnami, 4 Madiga, 3 Mala and 5 Kamsali) and 60 CEU samples from HapMap[4]. We chose 50 autosomal markers with the largest differences between India and CEU, requiring that markers be at least 2Mb apart. The average frequency difference between 86 India and 60 CEU samples at these markers was >50%. We genotyped a total of 456 cases and 338 controls selected from Table 1 and 2 (including the North, South and Central Indians samples) at these markers using the Sequenom iPLEX assay. 47 of the 50 markers passed quality control filters [Supplementary Table 2.]. We ran principal components analysis using the EIGENSOFT software[5,6]. Anova statistics for the difference between cases and controls from each population (North, Central and South) along the top eigenvector, representing the Europe-related axis, were also computed using the EIGENSOFT software [5,6]. Further data from the initial and six replicate association studies were combined in a meta-analysis using a Mantel-Haenszel test implemented in R and the result confirms our main association studies (M^2 = 47.8748, df = 2, p-value = 4.019e-11).

Selection criteria for family members

The 28 index patients and their family members were contacted and invited based on their interest to participate in the present genetic study. All the families were from South India and are unrelated. All the available family members were assessed with ECG and ECHO. Out of 28 families, 15 were HCM-related families, 12 were DCM-related families and one an RCM-related family. The details of the family members are given in Supplementary Table 2. Out of 28 families, six were selected as representative families (Supplementary Figure 4A) because they are high risk families with two or more cases of cardiomyopathies in first or second degree relatives. The remaining 22 families with moderate susceptibility displayed atleast one or two cases of cardiomyopathies or families associated with heart failure due to cardiomyopthies in first or second degree relatives. Out of 22 families, seven families were designated as cardiac death related families since the index patients (7) died in course of the study period (2000-2007). The familial study also shows that the 25 bp deletion is inheritable with incomplete penetrance. The same criteria used to screen the index patients were followed to distinguish symptomatic and asymptomatic carriers.

Population Screening

DNA samples of 6,273 individuals belonging to 107 ethnic groups were obtained from the DNA bank of the Centre for Cellular and Molecular Biology, Hyderabad, India (Collections of Dr. Lalji Singh and Dr. K. Thangaraj). The names of the populations and geographical locations are given in Supplementary Table **4**. A total of 2,085 samples belonging to different countries outside India were obtained from the HGDP-CEPH panel[7], The Wellcome Trust Sanger Institute, United Kingdom (Collection of Dr. Chris Tyler-Smith), Institute of Biotechnology and Genetic Engineering, Karachi, Pakistan (Collection of Prof. S. Qasim Medhi) and the University of Leeds, United Kingdom (Collection of Dr. Martin Richards). The samples were classified as North, South + Central, West, East and North East India as given in Supplementary Table 5 and the significance of differences in deletion carrier frequency between these regions of India was evaluated using a Fisher 2 by 5 test (http://www.quantitativeskills.com/sisa/statistics/fiveby2.htm).

Genotyping and DNA sequence analysis
Genomic DNA of individuals with cardiac failure was isolated from peripheral blood lymphocytes and genotyped using a pair of primers (Forward primer: 5'-GTT TCC AGC CTT GGG CAT AGT C-3' and Reverse primer: 5'-GAG GAC AAC GGA GCA AAG CCC-3') flanking the 25bp deletion region. PCR products (WT: 403bp and mutant 378bp) were analysed on 3.5% agarose gel, 6% Polyacrylamide gel, or using an ABI3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Samples with the 25bp deletion were further confirmed by direct sequencing of amplicons, using the BigDye terminator kit (Applied Biosystems, Foster City, CA, USA) and analysed on an ABI3730 DNA Analyser[8].

Protein analysis
Myocardial biopsies, wherever possible, were obtained from the individuals with cardiac failure with their informed written consent and stored at -80^{o}C. Tissue samples were transferred from -80^{o}C directly to SDS-sample buffer with 6 M urea, preheated at 90^{o}C, and homogenized. Insoluble material was pelleted and the supernatant analyzed on 14% SDS-polyacrylamide gels. For quantification, gels were stained with Coomassie brilliant blue and the actin band quantified densitometrically with a gel scanner. Samples were adjusted to equal protein concentrations by appropriate dilution in SDS-sample buffer. Equal volumes were then separated on 8–12% gradient gels and transferred to nitrocellulose membranes by electroblotting essentially as described[9]. Cardiac MyBP-C was detected using the rat polyclonal antibody directed against the recombinant NH_2-terminal module of C0-C1 of human cardiac MyBP-C[10]. Bound antibody was visualized with anti–rat or rabbit horseradish peroxidase conjugate and the ECL-system following the instructions of the supplier (Amersham Life Science). Blots were exposed on X-AR5 film (Eastman Kodak Co., Rochester, NY) for 30 s to 20 minutes.

Amplification of cDNA sequences
Total RNA was isolated from endomyocardial biopsies of the left ventricular myocardium according to the protocol described by RNeasy Minikit (Qiagen). A total of 100 ng RNA was reverse transcribed in a 20 µl volume as follows: 4 µl of reverse transcription buffer, 2 µl of 0.1 M DTT, 0.25 µl of RNA Guard, and 10 U H-Reverse Transcriptase Superscript from GIBCO BRL, Life Technologies (Eggenstein, Germany), 1 µl of 10 M pd (N_6)-primer, and 1 µl of 1 mM dNTP from Pharmacia LKB Biotechnology were added to preheated (68^{o}C, 5 minutes) RNA in a total volume of 20 µl. The reaction was incubated at 37^{o}C for 1.0 hour and terminated by heat inactivation (95^{o}C, 5 minutes). The cDNA products were then amplified in a 50 µl PCR reaction using the outer primer pair (Exon 32F: GCT ACA CTG CTA TGC TCT GCT GTG C and Exon 34R: CTG GCC ATC CCC AGG AGC CAG C). Blunt-ended PCR products were obtained using P*fu* Turbo DNA polymerase (Stratagene) and were directly cloned into pCR-Blunt II-Topo vectors (Invitrogen) and sequenced.

DNA Sequencing
Amplified PCR products, gel-isolated with a QiaxII Gel Extraction Kit (Qiagen). PCR-amplified cDNA, genomic DNA fragments, and plasmid DNA were sequenced on an ABI3730 DNA analyzer (Perkin-Elmer Corp., Applied Biosystems, Hitachi, Japan), using the ABI PRISM BigDye Terminator Cycle Sequencing Kit (Perkin-Elmer Corp., Foster City, CA), following the manufacturer's instructions.

Generation of recombinant adenovirus expressing wild type and truncated cMyBP-Cs
Two different cDNAs encoding WT and mutant mouse cMyBP-Cs were used in this study, since cMyBP-C is highly conserved between the human and mouse genomes. The WT cMyBP-C encodes the full-length protein and the mutant cMyBP-C, which results from 25 bp deletion in the intron 32, leads to the skipping of exon 33 (Accession number NM_008653). The procedure to construct the WT cDNAs was as described previously[10]. In the mutant

cMyBP-C, the exons 33 and 34 were completely removed, which altered the coding sequence for the region downstream of exon 32 and introduced a stop codon[10]. Replication-deficient recombinant adenoviruses were made by using the AdEasy system (Stratagene). The cells were normally transfected at a multiplicity of infection of ten. To distinguish the transgenic proteins from endogenous cMyBP-C protein, a myc epitope (EQKLISEEDL) was introduced after the initiation methionine codon of both the WT and mutant cDNAs.

Cardiomyocytes and Confocal analysis

Rat neonatal cardiomyocytes were isolated according to the manufacturer's protocol (Worthington Neonatal Cardiomyocyte Isolation System, Lakewood, NJ). Localization of cMyBP-C to the sarcomere was determined by immunohistochemical analyses using confocal microscopy. Alexa Fluor 488-conjugated goat anti-rabbit and Alexa Fluor 568-conjugated rabbit anti-mouse antibodies and TO-PRO-3 for nuclear staining were from Molecular Probes (Invitrogen). Polyclonal antibody against cMyBP-C (C0-C1), monoclonal antibody against α-tropomyosin (Sigma) and monoclonal antibody against the myc epitope (Roche) were utilized to immunostain neonatal cardiomyocytes. Cardiomyocytes were examined using confocal microscopy (Nikon PCM 2000) and software SimplePCI v.4 (C. Imaging Systems).

STR genotyping and calculation of variance

Five STR markers which are known closer to the 25 bp deletion (~3.4 MB up and downstream region), flanking the *MYBPC3* gene were amplified using a fluorescently labeled forward primer (Supplementary Table 5). Amplicons were analysed using an ABI3730 DNA Analyzer. Haplotypes were inferred from the genotypes obtained using PHASE [11]. The variance of each STR on the 25bp deleted chromosomes and the wild type chromosomes were then calculated**.**

Allele-specific amplification and resequencing of *MYBPC3* deleted and non-deleted alleles

We used allele-specific amplification to resequence ~10 kb haplotypes centered on the 25 bp deletion from nine heterozygous individuals (18 chromosomes). Long-range PCR was used to amplify the 25bp deleted and non-deleted alleles separately from heterozygous individuals, using a primer spanning or within the 25bp deleted region, respectively, paired with a second primer ~6.5 kb away (Supplementary Figure 7). A ~13 kb region of the *MYBPC3* gene surrounding the deletion was thus amplified in two fragments. The two amplicons from each chromosome were pooled and ~600 bp overlapping subfragments were amplified using 58 sets of primers (Supplementary Table 8 and 9). The amplicons were purified using SAP and ExoI (37^0C for 15 minutes and 80^0C for 15 minutes.) and sequenced using BigDye terminator cycle sequencing kit and ABI3730 DNA analyzer. SNPs found in a ~10 kb region of the 25 bp deleted and non-deleted chromosomes were flagged by Gap4 and checked manually in Mutation Surveyor. The two haplotypes from each individual were pooled and missing data (6.7%) were inferred by PHASE[11]. The haplotypes inferred by PHASE were checked for consistency with the directly-determined haplotypes. We found 24 variable sites, 23 SNPs and one insertion/deletion polymorphism (Supplementary Figure 7).

Estimation of the TMRCA of the deletion chromosomes

Five haplotypes were found among the nine deletion chromosomes. Two had structures that could be accounted for by single recombination events between the deletion consensus haplotype and observed non-deleted haplotypes, and were therefore considered likely to have arisen by recombination rather than mutation. The other three haplotypes could not be related to observed haplotypes by a simple recombination pathway and were therefore considered likely to have arisen by mutation. The deletion haplotypes differed from the consensus deletion haplotype by a mean of 0.22 substitutions per ~10.3 kb. Since the human and

chimpanzee sequences from this region differed by 83 positions in the 9.8 kb for which sequence data were available, and half of these were likely to have accumulated on the human lineage since its split from the chimpanzee lineage ~6.5 million years ago, a TMRCA of the human deletion haplotypes was estimated at ~33±23 thousand years using the ρ statistic[12].

4. Supplementary Note

Occurrence of the 25 bp deletion outside the Indian Subcontinent

The absence of the deletion from most populations tested in Indonesia and from the indigenous Semang and Senoi of the Malay Peninsula suggests a relatively recent intrusion in Southeast Asia from South Asia. The "Indianization" of Southeast Asia probably did not involve settlement until the fifth century AD and was focused around the Strait of Malacca and the Java Sea[13], so the arrival of the deletion (indicated by its presence in the Tenggerese of Java and Aboriginal Malays) is unlikely to date to before this time. Although there are few mtDNAs of Indian origin in Southeast Asia[14, 15], Y chromosomes of Indian ancestry have been found at a frequency of ~12% in Bali, where Indian influence had the greatest impact, suggesting Indian male settlement at least here[16]. Its presence in the Aboriginal Malays may seem surprising, given their isolation from the Indianized medieval cultures of the Malay Peninsula. However, it can readily be explained by introduction from an already Indianized island Southeast Asia after the seventh century (to which Aboriginal Malays trace part of their ancestry, although the time depth remains controversial)[14, 17] or even to more recent gene flow with Indian plantation labourers.

5. Supplementary References

1. Ammash, N.M., Seward, J.B., Bailey, K.R., Edwards, W.D.& Tajik, A.J. Clinical profile and outcome of idiopathic restrictive cardiomyopathy. *Circulation*. **101**, 2490-2496 (2000).
2. Mestroni, L. *et al*. Guidelines for the study of familial dilated cardiomyopathies. Collaborative Research Group of the European Human and Capital Mobility Project on Familial Dilated Cardiomyopathy. *Eur. Heart. J.* **20**, 93-102 (1999)
3. Phadke, R.S., Vaideeswar, P., Mittal, B. & Deshpande, J. Hypertrophic cardiomyopathy: an autopsy analysis of 14 cases. *J Postgrad. Med*. **47**,165-170 (2001).
4. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851-61 (2007).
5. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904-9 (2006).
6. Patterson, N. et al. Population structure and eigenanalysis. PLoS Genet 2, e190 (2006).
7. Cann, H.M. *et al.*, A human genome diversity cell line panel *Science.* **296**, 261-262 (2002).
8. Thangaraj, K. *et al*., Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol*. **13**, 86-93 (2003).
9. Towbin, H., Staehlin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci.* **76**, 4350-4354 (1979).
10. Sadayappan, S. *et al*., Cardiac myosin-binding protein-C phosphorylation and cardiac function *Circ. Res*. **25**, 1156-1163 (2005).
11. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978-989 (2001).
12. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37-48 (1999).
13. Bellwood, P. *Prehistory of the Indo-Malaysian archipelago* (Univ. of Hawaii Press, Honolulu, 1997).
14. Hill, C. *et al.,* Phylogeography and ethnogenesis of aboriginal Southeast Asians *Mol. Biol. Evol.* **23**, 2480-2491 (2006).
15. Hill, C. *et al.,* A mitochondrial stratigraphy for island southeast Asia. *Am. J. Hum. Genet.* ***80***, 29 (2007).
16. Karafet, T.M. *et al.,* Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Human Biology* **77**, 93-114 (2005).
17. Bellwood, P. *Asian Perspect*. **32**, 37 (1993).