UNIVERSITY OF NOVI SAD

FACULTY OF MEDICINE

DECIPHERING GENOTYPE-PHENOTYPE ASSOCIATIONS IN HYPERTROPHIC CARDIOMYOPATHY USING MACHINE LEARNING

DOCTORAL DISSERTATION

Mentors:
Prof. Dr. Lazar Velicki
Prof. Dr. Nataša Vučinić

Candidate:
Mila Glavaški

Novi Sad, 2023

УНИВЕРЗИТЕТ У НОВОМ САДУ

МЕДИЦИНСКИ ФАКУЛТЕТ

УТВРЂИВАЊЕ ПОВЕЗАНОСТИ ГЕНОТИПА И ФЕНОТИПА ХИПЕРТРОФИЧНЕ КАРДИОМИОПАТИЈЕ ПРИМЕНОМ МАШИНСКОГ УЧЕЊА

ДОКТОРСКА ДИСЕРТАЦИЈА

Ментор:
Проф. др Лазар Велицки
Проф. др Наташа Вучинић

Кандидат:
Мила Главашки

Нови Сад, 2023. године

УНИВЕРЗИТЕТ У НОВОМ САДУ **ОБРАЗАЦ – 5а**
МЕДИЦИНСКИ ФАКУЛТЕТ НОВИ САД

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА[1]

Врста рада:	Докторска дисертација
Име и презиме аутора:	Мила Главашки
Ментор (титула, име, презиме, звање, институција)	Проф. др сци. мед. Лазар Велицки, ванредни професор, Медицински факултет Нови Сад, Институт за кардиоваскуларне болести Војводине Проф. др Наташа Вучинић, ванредни професор, Медицински факултет Нови Сад
Наслов рада:	Утврђивање повезаности генотипа и фенотипа хипертрофичне кардиомиопатије применом машинског учења
Језик публикације (писмо):	Енглески језик (латиница) Српски језик (латиница)
Физички опис рада:	Унети број: Страница: 428 Поглавља: 7 Референци: 349 Табела: 131 Слика: 53 Графикона: 149 Прилога: 0
Научна област:	Медицина
Ужа научна област (научна дисциплина):	Кардиохирургија, кардиологија; генетика, лабораторијска медицина, биохемија; вештачка интелигенција, машинско учење
Кључне речи / предметна одредница:	хипертрофична кардиомиопатија; генотип; мутација; фенотип; машинско учење; алгоритми; вештачка интелигенција; ехокардиографија
Резиме на српском језику:	Hipertrofična kardiomiopatija (HCM) je najčešća nasledna kardiomiopatija. Dijagnoza HCM se postavlja na osnovu prisustva hipertrofije leve komore, uz isključivanje drugih uzroka hipertrofije. U pogledu genetičkih mutacija, HCM je heterogena bolest. Kliničke manifestacije i prognoza takođe mogu da budu veoma različite. Kod nekih pacijenata HCM je potpuno asimptomatska, dok kod drugih mogu da se razviju teška srčana insuficijencija i iznenadna srčana smrt. Povezanost genotipa i fenotipa HCM još uvek nije u potpunosti utvrđena. Mašinsko učenje je subdisciplina veštačke inteligencije u kojoj se kompjuterski algoritmi koriste za učenje kompleksnih šablona iz podataka. Cilj ovog istraživanja je bilo utvrđivanje povezanosti genotipa i fenotipa HCM primenom mašinskog učenja. Studija je bila multicentrična i retroprospektivna, obuhvatila je 143 odrasla pacijenta sa potvrđenom dijagnozom HCM. Anamnestički podaci, antropometrijska merenja, rezultati genetičkog testiranja, biohemijskih

[1] Аутор докторске дисертације потписао је и приложио следеће Обрасце:
5б – Изјава о ауторству;
5в – Изјава о истоветности штампане и електронске верзије и о личним подацима;
5г – Изјава о коришћењу.
Ове Изјаве се чувају на факултету у штампаном и електронском облику и не кориче се са тезом.

	analiza, nalazi transtorakalne ehokardiografije sa doplerom, kardiopulmonalnog testa fizičkim opterećenjem, elektrokardiograma (EKG) i EKG-holter-monitoringa su prikupljeni i korišćeni u daljoj analizi. HCM subfenotipi su identifikovani klasterizacijom. Povezanost genotipa i fenotipa je evaluirana korišćenjem Python modula Scikit-learn i SHapley Additive exPlanation (SHAP). Genotip-specifični nalazi ehokardiograma su identifikovani korišćenjem Python biblioteke za duboko učenje i računarski vid Fast AI, izradom modela za klasifikaciju ehokardiograma i naknadnom analizom regiona koji su najviše doprineli razlikovanju klasa. Četiri podtipa HCM su identifikovana na osnovu svih dostupnih podataka o fenotipu: klaster 0 ("AHOLD"), koji se razlikuje od ostalih na osnovu prečnika korena aorte (AO) i laktat dehidrogenaze (LDH), pri čemu su vrednosti AO > 30 mm i LDH > 300 U/L; klaster 1 ("RVSP ASCAOVS"), koji se razlikuje od ostalih na osnovu sistolnog pritiska desne komore (RVSP), dijametra ascedentne aorte (AscAO), i separacije aortnih kuspisa (AOvs), pri čemu su vrednosti AOvs > 27 m/s, AscAO < 31 mm i RVSP < 28 mmHg; klaster 2 ("weight"), koji se razlikuje od ostalih na osnovu telesne težine, sa vrednošću > 95 kg; i klaster 3 ("AV LVOT PG") koji se razlikuje od ostalih na osnovu srednjeg gradijenta pritisaka nad aortnom valvulom (AV meanPG), maksimalnog gradijenta pritisaka nad aortnom valvulom (AV maxPG), i maksimalnog gradijenta pritisaka nad izlaznim traktom leve komore (LVOT maxPG), pri čemu su vrednosti AV maxPG > 15 mmHg, AV meanPG > 6 mmHg, i LVOT maxPG > 15 mmHg. Algoritmi mašinskog učenja su potvrdili da utvrđivanje povezanosti genotipa i fenotipa HCM nije jednostavan zadatak. Predikcija ishoda fenotipa na osnovu informacije o mutiranim genima je moguća za prisustvo ili odsustvo sinusnog ritma i prisustvo ili odsustvo oštećenja miokarda. Modeli koji vrše predikciju prisustva ili odsustva sinusnog ritma su imali slične performanse kada su izrađeni samo na osnovu uzročnih gena za HCM i kada su izrađeni na osnovu svih analiziranih gena što sugeriše mogući značaj uzročnih gena za HCM i irelevantnost drugih analiziranih gena za ovaj ishod. Modeli koji vrše predikciju oštećenja miokarda su imali bolje performanse kada su korišćeni podaci o svim analiziranim genima (a ne samo o uzročnim genima za HCM), što sugeriše moguću važnu ulogu gena koji nisu uzročni, za ovaj ishod. Algoritmi mašinskog učenja su izvršili predikciju sledećih ishoda na osnovu podataka o genotipu i fenotipu: zamor, dispneja, bol u grudima, palpitacije, sinkopa, šum na srcu, pretibijalni edem, pokretanje mitralnog zalistka unapred (SAM), abnormalnost papilarnih mišića, hipokinezija, atrijalna fibrilacija, atrioventrikularni blok prvog stepena, blok leve grane (LBBB), blok desne grane (RBBB), prednji levi hemiblok, abnormalnosti ST segmenta, i negativni T talas. Prilikom predikcije zamora, najveći doprinos je imala kombinacija mutacije u *TNNT2* i maksimalnog odnosa disajne razmene (RER). Prilikom predikcije dispneje najveći doprinos imala je kombinacija mutacije u *MYBPC3* i vršne potrošnje kiseonika (peak VO2). Prilikom predikcije bola u grudima, najveći doprinos je imala kombinacija mutacije u *TNNI3* i koncentracije lipoproteina visoke gustine (eng. high-density lipoprotein, HDL). Prilikom predikcije šuma na srcu najveći doprinos imala je kombinacija mutacije u *MYH7* i podatka o implantiranju pejsmejkera/defibrilatora u porodičnoj istoriji, kao i kombinacija mutacije u *TNNT2* i zapremine leve pretkomore (LAV). Prilikom predikcije negativnog T talasa, najveći doprinos imala je kombinacija mutacije u *MYBPC3* i vrednosti transmitralnog maksimalnog gradijenta pritiska (MV maxPG). Identifikovani su genotip-specifični nalazi ehokardiograma: za mutaciju u *MYH7* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima), strukture koje najviše utiču na raspoznavanje

	su septum, izlazni trakt leve komore (LVOT), prednji zid, vrh srca, desna komora i mitralni aparat; za mutaciju u *TNNT2* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima) strukture koje najviše utiču na raspoznavanje su septum i desna komora; dok su za mutaciju u *MYBPC3* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima) ove strukture septum, leva komora i šupljina leve komore. Mašinsko učenje je na ovaj način doprinelo u određenoj meri izučavanju povezanosti genotipa i fenotipa HCM.
Датум прихватања теме од стране надлежног већа:	14.09.2021.
Датум одбране: (Попуњава одговарајућа служба)	
Чланови комисије: (титула, име, презиме, звање, институција)	Председник: Члан: Члан: Члан:
Напомена:	Нема

UNIVERSITY OF NOVI SAD
FACULTY OR CENTER

KEY WORD DOCUMENTATION[2]

Document type:	Doctoral dissertation
Author:	Mila Glavaški
Supervisor (title, first name, last name, position, institution)	Prof. dr. Lazar Velicki, associate professor, Faculty of Medicine Novi Sad, Institute of Cardiovascular Diseases Vojvodina Prof. dr. Nataša Vučinić, associate professor, Faculty of Medicine Novi Sad
Thesis title:	Deciphering Genotype-Phenotype Associations in Hypertrophic Cardiomyopathy Using Machine Learning
Language of text (script):	English language (latin script) Serbian language (latin script)
Physical description:	Number of: Pages: 428 Chapters: 7 References: 349 Tables: 131 Illustrations: 53 Graphs: 149 Appendices: 0
Scientific field:	Medicine
Scientific subfield (scientific discipline):	Cardiac surgery, cardiology; genetics, laboratory medicine, biochemistry; artificial intelligence, machine learning
Subject, Key words:	Cardiomyopathy, Hypertrophic; Genotype; Mutation; Phenotype; Machine Learning; Algorithms; Artificial Intelligence; Echocardiography
Abstract in English language:	Hypertrophic cardiomyopathy (HCM) is the most prevailing heritable cardiomyopathy. HCM is diagnosed by the existence of left ventricular hypertrophy despite the lack of abnormal loading conditions causing it. HCM is a heterogeneous disease regarding genetic mutations. Clinical manifestations and prognosis vary widely as well. Some patients are completely asymptomatic, in some others, severe heart failure and sudden cardiac death may arise. Definitive genotype-phenotype associations are still unknown. Machine learning (ML) is a subdiscipline of artificial intelligence, wherein computer algorithms are used for learning complex patterns from data. The aim of this research was to decipher genotype-phenotype associations in HCM using ML. The study was multi-centric and retroprospective, and involved 143 adult HCM patients. Medical and family history, anthropometric measurements, genetic testing, blood markers, transthoracic echocardiography with Doppler, cardiopulmonary exercise testing (CPET), ECG and ECG-holter-monitoring

[2] The author of doctoral dissertation has signed the following Statements:
5б – Statement on the authority,
5в – Statement that the printed and e-version of doctoral dissertation are identical and about personal data,
5г – Statement on copyright licenses.
The paper and e-versions of Statements are held at he faculty and are not included into the printed thesis.

	data were collected and further analysed. HCM subphenotypes were identified using clustering. Associations of genotype and phenotype were evaluated used Python modules Scikit-learn and SHapley Additive exPlanation (SHAP). Genotype-specific echocardiogram findings were identified using Python deep learning (DL) and computer vision library Fast AI, by generation of DL models for classification of ultrasonic images, and later analysis of the most decisive image regions. Four HCM subtypes were identified based on the overall phenotypic appearance: cluster 0 ("AHOLD"), distinguishable by aortic root diameter (AO) and lactate dehydrogenase (LDH), with values mostly AO > 30 mm, and LDH > 300 U/L; cluster 1 ("RVSP ASCAOVS"), distinguishable by right ventricle systolic pressure (RVSP), diameter of ascending aorta (AscAO), and aortic leaflet separation diameter (AOvs), with the values of RVSP < 28 mmHg, AscAO < 31 mm, and AOvs > 27 m/s; cluster 2 ("weight"), recognizable by weight, wherein values being mostly > 95 kg; and cluster 3 ("AV LVOT PG") distinguishable by aortic valve mean pressure gradient (AV meanPG), aortic valve peak pressure gradient (AV maxPG), and left ventricular outflow tract peak gradient (LVOT maxPG) wherein AV maxPG > 15 mmHg, AV meanPG > 6 mmHg, and LVOT maxPG > 15 mmHg. ML algorithms confirmed that the determination of genotype-phenotype associations in HCM is a cumbersome task. Two phenotypic outcomes that can be predicted from mutated genes are the absence or presence of sinus rhythm and the absence or presence of myocardial injury. Models predicting the absence or presence of sinus rhythm had similar performance when they were built using only causative genes and when using all analyzed genes, indicating potential importance of causative genes and irrelevance of non-causative genes for that outcome. On the other hand, models predicting myocardial injury — infarction had better performance when they were built using all analyzed genes (and not just causative ones), indicating a potentially significant role of non-causative genes in that outcome. The ML algorithms were able to predict phenotypic outcomes — fatigue, dyspnea, chest pain, palpitations, syncope, heart murmur, pretibial edema, systolic anterior motion, papillary muscle abnormalities, hypokinesia, atrial fibrillation (AF), first-degree atrioventricular (AV) block, left bundle branch block (LBBB), right bundle branch block (RBBB), left anterior hemiblock, ST segment abnormalities, and negative T wave — using genotypic and phenotypic data. The combination of a mutation in *TNNT2* and peak respiratory exchange ratio (RER) contributed the most in predicting fatigue. The combination of a mutation in *MYBPC3* and peak VO2 contributed the most in predicting dyspnea. The combination of a mutation in *TNNI3* and high-density lipoprotein (HDL) level contributed the most in predicting chest pain. The combination of a mutation in *MYH7* and pacemaker/defibrillator implants in family history, as well as the combination of a mutation in *TNNT2* and left atrial volume (LAV), contributed the most in predicting heart murmur. Lastly, the combination of a mutation in *MYBPC3* and transmitral maximal pressure gradient (MV maxPG) aided the most in predicting negative T wave. Genotype-specific echocardiogram findings were identified: for mutations in the *MYH7* gene (vs. mutation not detected), the most discriminative structures are the left ventricular outflow tract, septum, anterior wall, apex, right ventricle, and mitral apparatus; for mutations in the *TNNT2* gene (vs. mutation not detected), the most discriminative structures are septum and right ventricle; while for mutations in *MYBPC3* gene (vs. mutation not detected) these are septum, left ventricle, and left ventricle chamber. ML has thus been demonstrated to be useful in deciphering genotype-phenotype associations in HCM.

Accepted on Scientific Board on:	14.09.2021.
Defended: (Filled by the faculty service)	
Thesis Defend Board: (title, first name, last name, position, institution)	President: Member: Member: Member:
Note:	None

I am immensely grateful to my mentors, the world-class motivator and supervisor Professor Dr. Lazar Velicki, and remarkably devoted Professor Dr. Nataša Vučinić for all the support, knowledge, and care they provided me. Before I met you, I could not even imagine that such dedicated mentors existed.

I would like to thank my father Sava, mother Vera, and all of my friends and family immeasurably for all the love, support, and understanding throughout this process. You have all contributed crucially to who I am today, and I am blessed to share my journey with you.

Mila Glavaški

CONTENT

ABBREVIATIONS

2D — two-dimensional

ABC — ATP-binding cassette

ABCC9 — ATP-binding cassette subfamily C member 9 gene

ACC — American College of Cardiology

ACCF — American College of Cardiology Foundation

ACE — angiotensin-1 converting enzyme-1 gene

ACM — arrhythmogenic cardiomyopathy

ACTC1 — actin alpha cardiac muscle 1 gene

ACTN1 — actinin alpha 1 gene

ACTN2 — actinin alpha 2 gene

AF — atrial fibrillation

AGT — angiotensinogen gene

AGTR1 — angiotensin II receptor 1 gene

AHA — American Heart Association

ALP — actinin-associated LIM protein

ALT — alanine transaminase

AMPK — AMP-activated protein kinase

ANKRD1 — ankyrin repeat domain 1 gene

ANNs — artificial neural networks

AO — aortic root diameter

AOvs — aortic leaflet separation diameter

AP — average precision (precision-recall curve)

AR — aortic regurgitation

ASA — alcohol septal ablation

AscAO — diameter of ascending aorta

AST — aspartate transaminase

AT — anaerobic threshold

AUC — area under the receiver operating characteristic curve

AV — atrioventricular

AV Vmax — aortic valve peak velocity

AV maxPG — aortic valve peak pressure gradient

AV meanPG — aortic valve mean pressure gradient

AVA — aortic valve area

AVVTI — aortic valvular velocity time integral

AscAO — ascending aorta diameter

BDKRB2 — bradykinin B2 receptor gene

BMI — body mass index

CACNA1C — calcium voltage-gated channel subunit alpha1 C gene

CALR3 — calreticulin 3 gene

CARP1/Ankrd1 — cardiac ankyrin repeat protein/ankyrin repeat domain 1

CARP2/Ankrd2/Arpp — ankyrin repeat protein with PEST and proline rich region

CARP3/DARP/Ankrd23 — diabetes-related ankyrin repeat protein

CASQ2 — calsequestrin 2 gene

CAV3 — caveolin 3 gene

CHF — congestive heart failure

CK — creatine kinase

CK-MB — creatine kinase MB isoenzyme

CMA1 — chymase gene

CMRI — cardiac magnetic resonance imaging

COPD — chronic obstructive pulmonary disease

CPET — cardiopulmonary exercise testing

CRP — cysteine-rich protein

CRYAB — crystallin alpha B gene

CSRP3 — cysteine and glycine rich protein 3

CT — computed tomography

CTF1 — cardiotrophin 1 gene

CYP11B2 — aldosterone synthase gene

DCM — dilated cardiomyopathy

DCNNs — deep convolutional neural networks

DES — desmin gene

DL — deep learning

DMD — dystrophin gene

DSC2 — desmocollin 2 gene

DSG2 — desmoglein 2 gene

DSP — desmoplakin gene

DT — decision tree

E/E' — ratio of peak velocity of early diastolic transmitral flow to peak velocity of early diastolic mitral annular motion as determined by pulsed wave Doppler

ECG — electrocardiogram

ECM — extracellular matrix

EDN1 — endothelin 1 gene

EDVLV — end-diastolic volume of the left ventricle

EFLV — left ventricular ejection fraction

ELN — elastin gene

ESC — European Society of Cardiology

ESVLV — end-systolic volume of left ventricle

FABP — fatty acid-binding protein

FATZ — telethonin-binding protein at the Z-disc

FBN1 — fibrillin 1 gene

FGF — fibroblast growth factor

FHL1 — four-and-a-half LIM domains 1

FHL2 — four-and-a-half LIM domains 2

FHLs — four-and-a-half LIM domain proteins

FLNC — filamin C gene

FXN — frataxin gene

GAA — alpha glucosidase gene

GaussianNB — Gaussian Naive Bayes

GLA — galactosidase alpha gene

GSEA — gene set enrichment analysis

HCM — hypertrophic cardiomyopathy

HCN4 — hyperpolarization activated cyclic nucleotide gated potassium channel 4

HDAC2 — histone deacetylase 2

HDAC5 — histone deacetylase 5

HDL — high-density lipoprotein

HF — heart failure

ICD — implantable cardioverter-defibrillator

IGF — insulin-like growth factor

IGF2 — insulin-like growth factor 2 gene

IL6 — interleukin 6 gene

INDRA — "Integrated Network and Dynamical Reasoning Assembler"

IVSd — interventricular septum thickness

JNK — c-Jun N-terminal kinase

JPH2 — junctophilin 2 gene

LA — left atrium

LAMA4 — laminin subunit alpha 4

LAMP2 — lysosomal associated membrane protein 2

LAV — left atrial volume

LAVs — left atrial volume in systole

LBBB — left bundle branch block

LDB3 — LIM domain binding 3 gene

LDH — lactate dehydrogenase

LDL — low-density lipoprotein

Log Reg — logistic regression

LV — left ventricle

LVEDP — left ventricular end-diastolic pressure

LVH — left ventricle hypertrophy

LVIDd — left ventricular internal dimension at end-diastole

LVIDs — left ventricular internal dimension at end-systole

LVNC — left ventricular non-compaction cardiomyopathy

LVOT — LV outflow tract

LVOT Vmax — left ventricular outflow tract maximal velocity

LVOT maxPG — left ventricular outflow tract peak gradient

LVOTO — LV outflow tract obstruction

LVWT — LV wall thickness

MAAd — diastolic mitral annular area diameter

MADd — mitral annulus diameter at end-diastole

MARP — muscle ankyrin repeat protein

MHC — myosin heavy chain

MIAT — myocardial infarction associated transcript

ML — machine learning

MLP — muscle LIM protein

MR — mitral regurgitation

MR ERO — effective regurgitant orifice area for mitral regurgitation

MR RV — mitral regurgitation regurgitant volumes

MURF — muscle-RING-finger-proteins

MV maxPG — transmitral maximal pressure gradient

MV meanPG — transmitral mean pressure gradient

MVA — mitral valve area

MVVTI — mitral valve velocity-time integral

MYBPC3 — myosin binding protein C3

MYH6 — myosin heavy chain 6 gene

MYH7 — myosin heavy chain gene

MYL2 — myosin light chain 2 gene

MYL3 — myosin light chain 3 gene

MYLK2 — myosin light chain kinase 2 gene

MYOZ2 — myozenin 2 gene

MYPN — myopalladin gene

MZ — monozygotic

MeSH — Medical Subject Headings

MyBP-C — myosin-binding protein C

MyBP-H — myosin-binding protein H

NCX — sodium-calcium exchanger

ND — not detected

NEXN — nexilin F-actin binding protein

NFAT — nuclear factor of the activated T-cell

NKX2-5 — NK2 homeobox 5

NO — nitric oxide

NOTCH1 — Notch receptor 1 gene

NSVT — non-sustained ventricular tachycardia

NT-proBNP — N-terminal prohormone of brain natriuretic peptide

NYHA — New York Heart Association

OMGL — Oxford Medical Genetics Laboratory

PCR — polymerase chain reaction

PDLIM3 — PDZ and LIM domain 3 gene

PET — positron emission tomography

PI3K — phosphoinositide 3-kinase

PICP — C-terminal propeptide of type I procollagen

PKA — protein kinase A

PKC — protein kinase C

PKCß2 — protein kinase C beta 2

PKG-1 — cGMP-dependent protein kinase-1

PKP2 — plakophilin 2

PLA2G7 — platelet activating factor acetylhydrolase gene

PLN — phospholamban gene

PLWd — posterior left ventricle wall thickness at end-diastole

PRKAG2 — protein kinase AMP-activated non-catalytic subunit gamma 2

PSVT — paroxysmal supraventricular tachycardia

PTP — protein tyrosine phosphatase

PTPN11 — protein tyrosine phosphatase non-receptor type 11 gene

RAAS — renin–angiotensin–aldosterone system

RAF1 — Raf-1 Proto-Oncogene, Serine/Threonine Kinase gene

RAVs — right atrial volume at end systole

RBM20 — RNA binding motif protein 20 gene

RCM — restrictive cardiomyopathy

RER — respiratory exchange ratio

RF — random forest

ROS — reactive oxygen species

RV — right ventricle

RVSP — right ventricle systolic pressure

RYR1 — ryanodine receptor 1

SA — sinoatrial

SAM — systolic anterior motion

SBGN — Systems Biology Graphical Notation

SBML — Systems Biology Markup Language

SCD — sudden cardiac death

SDHA — succinate dehydrogenase complex flavoprotein subunit A

SERCA2a — ATP-dependent sarcoplasmic reticulum pump (SR Ca2+-ATPase)

SHAP — SHapley Additive exPlanation

SMAD2 — SMAD Family Member 2

SMOTE — Synthetic Minority Oversampling Technique

SVC — radial basis function kernel C-support vector classification

SVLV — stroke volume of the left ventricle

TAAd — tubular ascending aorta during end-diastole

TADd — transverse aortic arch during end-diastole

TAPSE — tricuspid annular plane systolic excursion

TCAP — titin-cap gene

TDI — tissue Doppler imaging

TGF — transforming growth factor

TGFB1 — transforming growth factor ß1 gene

TNF — tumor necrosis factor gene

TNNC1 — troponin C1, slow skeletal and cardiac type gene

TNNI3 — troponin I3, cardiac type gene

TNNT2 — troponin T2, cardiac type gene

TPM1 — tropomyosin 1 gene

TPM2 — tropomyosin 2 gene

TPM3 — tropomyosin 3 gene

TPM4 — tropomyosin 4 gene

TPOT — tree-based pipeline optimization

TR — tricuspid regurgitation

TRIM55 — tripartite motif containing 55 gene

TRIM63 — tripartite motif containing 63 gene

TRPM4 — transient receptor potential cation channel subfamily M member 4 gene

TTN — titin gene

TTR — transthyretin gene

TnC — troponin C

TnI — troponin I

TnT — troponin T

UPS — ubiquitin-proteasome system

VCL — vinculin gene

VEGF — vascular endothelial growth factor

VT — ventricular tachycardia

cGMP — cyclic guanosine monophosphate

cMyBP-C — cardiac isoform myosin-binding protein C

eGFR — estimated glomerular filtration rate

fMyBP-C — fast skeletal myosin-binding protein C

lncRNAs — long non-coding RNAs

mTOR — mechanistic target of rapamycin

rpm — revolutions per minute

sMyBP-C — slow skeletal myosin-binding protein C

1. INTRODUCTION

1.1. Hypertrophic cardiomyopathy (HCM)

HCM is the most prevailing heritable cardiomyopathy (1–5), having a recorded prevalence of 1 in 500 people among the general population worldwide (3,4,6–11), and recent research assessing a prevalence of up to 1 in 200 people (9,11–16). HCM affects 0.2-0.6% of population (13,17–20), causes disability and mortality across all ages (7), and represents a leading cause of sudden cardiac deaths among the young (21–23).

HCM is primarily considered to be inherited as an autosomal dominant trait (1,11,12,24), and to be caused by mutations in cardiac sarcomeric protein genes (1,12,25–27). However, recent research suggests that the genetic foundation of HCM is much more complex than originally postulated (12,28).

HCM is diagnosed by the existence of left ventricular hypertrophy (LVH) despite the lack of abnormal loading conditions causing it (29–31): in adults, left ventricular thickness found to be ≥ 15 mm (4,32), as quantified by any medical imaging technology — echocardiography (ECHO), cardiac magnetic resonance imaging (CMRI) or computed tomography (CT) (3,32). Secondary causes of LVH are systemic hypertension, subvalvular and valvular aortic stenosis, as well as infiltrative cardiomyopathies (3), and all of them must be excluded in order to diagnose HCM (3). Among first-degree family members of subjects with confirmed HCM, an unexplained left ventricular thickness found to be ≥ 13 mm is sufficient for HCM diagnosis (32).

Clinical manifestations and prognosis vary widely. Severe heart failure (HF) and sudden cardiac death (SCD) may arise in some HCM subjects (1,10,11). SCD can be the first manifestation of HCM (1). Current therapeutic strategies (pharmacological treatment and cardiac surgical procedures) provide symptomatic treatment for HCM, leaving the causes of the disease unaddressed (19). Treatment is usually individualized (3), and aims to alleviate symptoms and impede disease progression (32).

Research on HCM is complicated by the fact that animal models do not completely emulate human HCM (14).

1.2. Cardiomyopathies

Cardiomyopathies are heterogeneous myocardial diseases characterized by structural and functional defects. They can be subdivided into HCM, dilated cardiomyopathy (DCM), arrhythmogenic cardiomyopathy (ACM), restrictive cardiomyopathy (RCM), and unclassified cardiomyopathies as for example, left ventricular non-compaction cardiomyopathy (LVNC). The cardiomyopathies can be genetic or acquired (33,34), whereas HCM has a genetic basis (34).

1.3. Sarcomere

The myocardium is mainly composed of cardiac muscle cells (cardiomyocytes) (9). Cardiomyocytes contain myofibrils, which are positioned across the cell. Myofibrils are partitioned into contractile units known as sarcomeres (35,36). Sarcomere is a repeating building block and contractile unit of striated (cardiac and skeletal) muscles, and coordinates muscle contraction (33,37–40). HCM is sometimes depicted as a "disease of the sarcomere" (6,9,16), indicating that it is most often caused by mutations in genes that encode proteins associated with the cardiac sarcomere (29,36,41–43).

There are two principal components within the sarcomere: the thick filament and the thin filament. The thick filament is built by approximately 300 molecules of motor protein myosin. Each molecule of myosin is made of 2 protein units of α- or β-myosin heavy chain (MHC) and 4 myosin light chains (MLC). The thin filament contains repeating filamentous actin molecules, associated with the contraction-tuning troponin complex [troponin I (TnI), troponin T (TnT), and troponin C (TnC)], α-tropomyosin, and cardiac myosin-binding protein C (MyBP-C). MyBP-C links the thick and thin filaments and regulates actin-myosin interplay (Figure 1) (33,35,36,40). The striation pattern along a myofibril is the result of overlapping sets of actin and myosin filaments within the sarcomere (37).

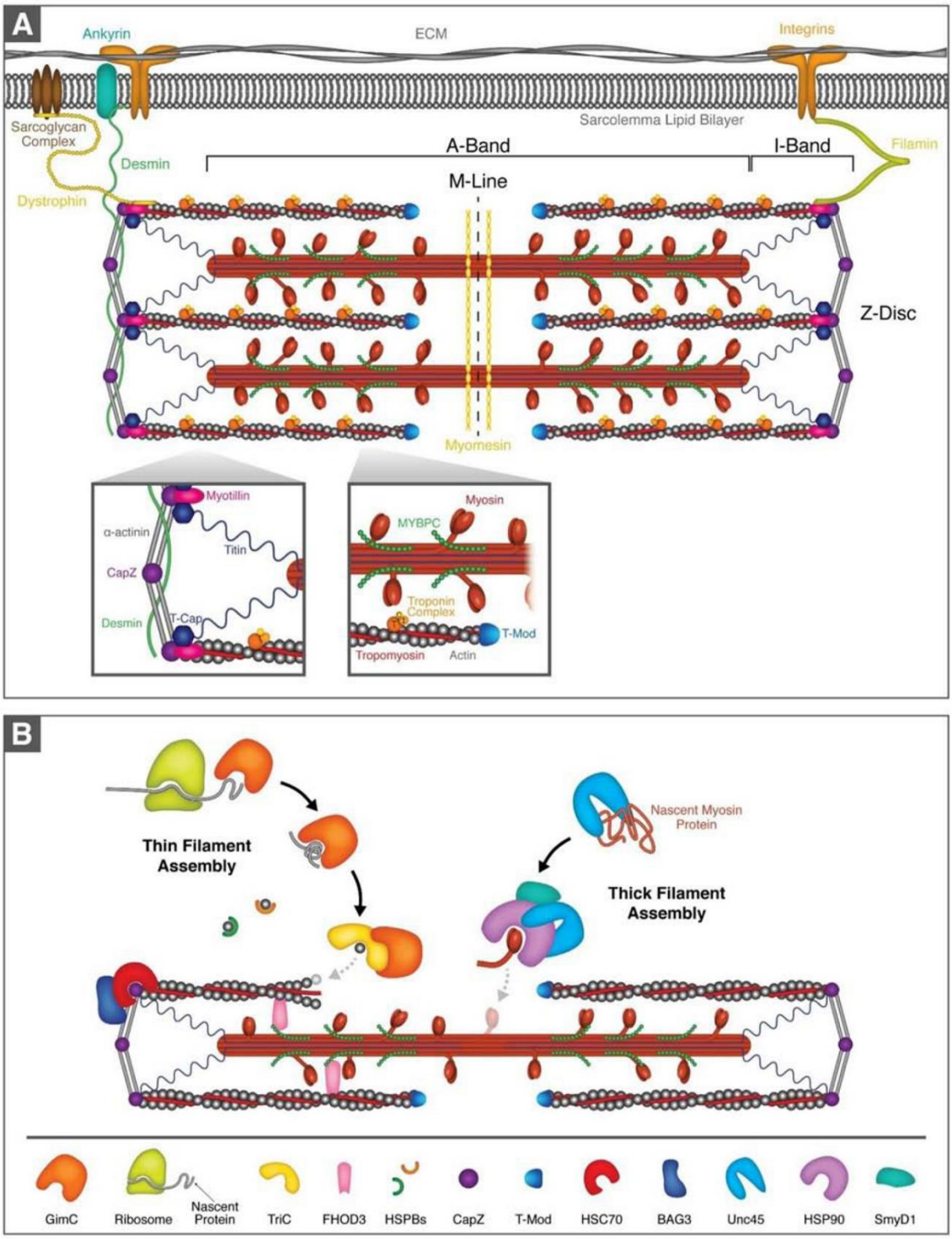

Figure 1. Cardiac sarcomere and assembly of myofilaments: A) sarcomere and sarcomere-associated proteins, B) thick and thin filament assembly (40).

Muscle contraction is a result of the actin-myosin interplay: the globular head of the myosin molecule folds towards and thereupon binds to actin, contracts, unbinds actin, and subsequently starts a new cycle. These connections between the myosin head and actin are termed cross-bridges, and MyBP-C regulates their kinetics (35). When a muscle contracts, actin and myosin filament arrays slide past each other, while the filaments shorten themselves very slightly (37).

Cardiomyocyte contraction and cardiac electrical activation are linked by the intracellular transport of calcium (35). Depolarization of the cardiomyocyte cell membrane activates the L-type voltage-dependent calcium (Ca^{2+}) channels in the T tubule of the cell membrane, leading to Ca^{2+} influx into the cell. This influx triggers the opening of ryanodine-receptor channels in the adjacent sarcoplasmic reticulum surrounding myofibrils ("calcium-induced calcium release"), with a rapid discharge of stored Ca^{2+} from the sarcoplasmic reticulum and a rise in cytosolic Ca^{2+}. Intracellular Ca^{2+} binds to TnC on the actin filaments, resulting in an allosteric conformational change in TnI and TnT, and removing tropomyosin from the myosin-binding sites of actin. Consequently, the myosin-binding sites of actin are exposed, allowing the cross-bridges to form (35–37,44). The globular head of the myosin then binds to actin and the contractile cycle can proceed (37).

After contraction, cytoplasmic Ca^{2+} levels must be decreased for myofilament relaxation to be initiated. Various mechanisms contribute to the removal of the cytoplasmic Ca^{2+} discharged from the myofilaments: in ventricular cardiomyocytes, the largest portion of cytoplasmic Ca^{2+} is returned to the sarcoplasmic reticulum via the adenosine triphosphate (ATP)-dependent sarcoplasmic reticulum pump (SR Ca^{2+}-ATPase (SERCA2a); ~70%), followed by the sodium-calcium exchanger (NCX; ~28%), and the sarcolemmal Ca^{2+}-ATPase (~1%). The contribution of mitochondrial Ca^{2+} uptake to the removal of the cytoplasmic Ca^{2+}, is small and plays a regulatory role — it enhances ATP production in case of increased workloads (9).

ATP is synthesized predominately in mitochondria via oxidative phosphorylation (44). ATP hydrolysis ($ATP + H_2O \rightarrow ADP + Pi$) releases energy, which drives myosin sliding along the actin filament, resulting in contraction (35,37,44). Uncoupling actin and myosin is also an energy consuming process (45). When a new molecule of ATP is bound

to myosin, it leads to the release of myosin from actin. TnI, TnT, and tropomyosin bind actin again, unexposing the myosin-binding sites. All this cumulatively results in muscle relaxation (36).

The sarcomere is symmetric, it has the M-disc in the center and borders at the Z-discs (9). In myofibrils, sarcomeres are linked at the protein-dense Z-discs (9,39) — the lateral boundaries of a sarcomere (39). Thin, actin-containing filaments are connected to the Z-disc. Thick, myosin-containing filaments are linked in the center via the M-band (33). "The force of muscle contraction occurs when the myosin motor protein attaches to the actin filament and pulls the Z-discs toward the M-band" (39). The elastic filament consists of titin, protein extending over half the sarcomere (33) — beginning at the Z-discs to the M-bands (9). The titin molecules provide a restoring force, when the sarcomere is stretched (37). Z-disc was for long considered a passive bond to titin and thin filaments. However, Z-discs also take part in various cellular mechanisms (39).

The sarcomere is dynamic. Apart from being a constant scaffold for structural and regulatory proteins, it responds to alterations in muscle load and injury (39). All the cytoskeletal elements must be coordinated to produce adequate contraction (33,39), and to enable continuous cardiac work and adaptations to the momentary requests of the body (33).

Sarcomeric function is hypersensitive: disruptions in the integrity of any component (single dysfunctional sarcomeric protein, changes in structure or dynamics of sarcomere, alterations in proteolytic degradation or expression, etc.) can result in cardiomyopathies (39). Compensatory mechanisms to overcome the defects are also sometimes activated, and some of them may, over time, become pathological too (33).

1.3.1. Proteins involved in sarcomeric function

1.3.1.1. α-Actinin

Human α-actinin is a highly conserved protein (46). There are 4 genes encoding α-actinin (39) and 4 main isoforms (46), which possess overlapping functions: ACTA1 and ACTA4 are non-muscle (and non-sarcomeric) isoforms that link actin filaments in cytoskeleton, ACTA2 and ACTA3 are skeletal muscle isoforms that are major components of the sarcomeric Z disc, and ACTA2 is cardiac muscle isoform (39,46).

Deficit of α-actinin-3 isoform is not pathogenic, most probably because individuals with this deficiency present high compensatory expression of α-actinin-2 (46,47). On the other hand, defects in α-actinin-2 may lead to cardiac and skeletal muscle diseases — including HCM and DCM (46,48).

α-actinin has many binding companions. Each of these interactions plays a different role in contraction production. Some of Z-disc proteins which interact with ACTA2 are: muscle LIM protein (MLP), N-terminus of titin, actinin-associated LIM protein (ALP), myotilin, myopalladin, myopodin, CapZ, cypher/oracle/ZASP, filamin, α-actinin, and telethonin-binding protein at the Z-disc (FATZ) (39).

1.3.1.2. Muscle LIM protein

Muscle LIM protein (MLP) is a protein with a LIM domain serving mechanosensing and structural roles (49), a member of the cysteine-rich protein (CRP) family (39). *CSRP1* is located in smooth muscle, *CSRP2* in fibroblasts and arteries, and *CSRP3* in striated (cardiac or skeletal) muscle. MLP is mainly considered a Z-disc protein; however, it also localizes to other regions in myocytes (e.g. M-band). MLP contains 2 LIM domains and they serve as protein-protein binding sites (39). MLP binds to α-actinin to stabilize the Z-disc (39), acts as a stretch sensor (39,50), signaling protein (39), and takes part in the remodeling of cardiac myofilaments (50), myogenesis and muscle differentiation (39).

1.3.1.3. Muscle ankyrin repeat proteins

The family of muscle ankyrin repeat proteins (MARPs) is consisted of: cardiac ankyrin repeat protein/ankyrin repeat domain 1 (CARP1/Ankrd1), ankyrin repeat protein with peptide sequence rich in proline (P), glutamic acid (E), serine (S), and threonine (T) and proline rich region (CARP2/Ankrd2/Arpp), and diabetes-related ankyrin repeat protein CARP3/DARP/Ankrd23). All of them contain 4 ankyrin repeat domains (39).

CARP1 (encoded by the gene *ANKRD1*) is mostly found in cardiomyocytes, and to a lower amount in skeletal muscles. CARP1 gene expression is part of the fetal gene program, and the expression increases after the induction of pathologic hypertrophy. CARP1 interplays with several proteins: calsequestrin, desmin, four-and-a-half LIM

domains 2 (FHL2), myopalladin, and talin; as well as numerous nuclear transcription factors implicated in cell differentiation, proliferation and apoptosis (39).

1.3.1.4. Titin

Titin (connectin) is the greatest protein in the human genome (39), a major source of passive stiffness in cardiomyocytes (51). It extends over half of the sarcomere (39). It is important in regulation of mechanical characteristics of the sarcomere during contraction and at rest (52), and in shaping viscoelastic features of myofilaments (53). The elastic features of titin change throughout heart development and progression of cardiac diseases, which strongly affects the diastolic function of the myocardium (53).

Titin has a major role in sarcomere formation and stability, as well as in signal transduction and mechanosensing (39,52). Titin is a key molecule in the passive mechanical properties of the myofilaments (39), because it acts as a molecular spring (39,53). Titin has several functionally diverse partners, which interact with it along its length (39). Titin is also known as the third filament (9).

Titin transcripts undergo alternative splicing, and as a result, they can theoretically produce more than one million titin isoforms (52). There are two titin isoforms in the human adult myocardium: N2B and N2BA (9). The presence of N2A and N2B elements is used to classify titin isoforms into major categories (52). N2BA isoform (containing N2B and N2A, the longer isoform) is different from N2B (containing only N2B, the shorter isoform) (9,54).

1.3.1.4.1. Z-disc titin

The N-terminal region of titin is fixed in the Z-disc (39) and binds α-actinin (39,46). At the I-band and Z-disc junction, titin binds tropomyosin, calpain, and obscurin, and simultaneously supports sarcomere stability and mechanosensing (39).

1.3.1.4.2. I-band titin

The N2B element is found only in heart. In the N2B isoform, part of the PEVK region is also missing. The PEVK region (termed so because it contains ~ 70% proline (P), glutamic acid (E), valine (V), and lysine (K) residues) enables titin's elastic features

(39). Increased PEVK phosphorylation is associated with increased cardiomyocyte passive stiffness in HF (54). Titin's PEVK region elongates during moderate stretching of striated muscle, and as a result, passive tension increases (55).

Titin's I-band region is involved in response to hypertrophy and mechanosensing. The N2A region interacts with MARPs to evoke a mechanosensory reaction. The N2B region interacts with four-and-a-half LIM domain proteins (FHLs) and in response to biomechanical stress activates hypertrophy pathways (39).

1.3.1.4.3. A-band titin

In the A-band, titin interreacts with MyBP-C and the myosin tails, and thus, links to the thick filaments (39).

1.3.1.4.4. M-band titin

M-band is the central part of the sarcomere, and the most dynamic part of the muscle (56). M-band titin plays a key role in structural support and signaling. The extreme C-terminus of titin binds obscurin, thus serving both signaling and structural functions. Muscle-RING-finger-proteins (MURF) 1 and 2 label titin for proteosomal degradation (39).

1.3.1.5. Actin

Actin is one of the most abundant (39,57) and preserved proteins (58), one of the most dynamic molecules (57) within cells. Actin is involved in protein and cell homeostasis (57). It is also the main constituent of the thin filament (39).

Six genes encode different actin isoforms: α-cardiac-, α-skeletal-, α-smooth-, β-cytoplasmic-, γ-smooth-, and γ-cytoplasmic-actin. Although they share high sequence homology (more than 87%), varied isoforms display tissue-specific expression. α-cardiac actin is encoded by *ACTA1* gene (9) and represents the main isoform, which forms thin filaments in cardiomyocytes (39).

Individual actin molecules are globular in solution (G-actin) but interreact to assembly filamentous polymers (F-actin). F-actin has the appearance of an arrowhead:

"barbed" ends are sticked into the Z-disc, and the "pointed" ends extend into the M-band (39).

1.3.1.6. Tropomyosin

Tropomyosin is a key element in the coordinated activation of the thin filament (39,59). It acts both as a regulatory protein and a structural support (39) and enables proper sarcomeric function by interplay with troponin T and actin (39,59). Tropomyosin regulates muscle contraction through a steric-mechanism, which governs myosin-crossbridge-actin interactions (60).

The tropomyosin family consists of 4 genes: *TPM1* (encoding α-TPM), *TPM2* (encoding β-TPM), *TPM3* (encoding γ-TPM), and *TPM4* (encoding δ-TPM) (39). Mutations in *TPM1* are usually studied and found in the context of HCM, but several mutations have been reported in DCM as well (61). α- and β-Tropomyosin are ~ 87% identical, but the ratio of their expression varies, based on the developmental stage and the fiber type (39). α-Tropomyosin is a highly conserved protein mostly found in cardiac and skeletal muscles (39,62), whilst β-tropomyosin is seen in slow-twitch muscle fibers (39).

Coiled-coils consist of two tropomyosin α-helical chains assembled in head-to-tail fashion, which interact with the groove of actin filaments (positively charged), and form dimers spanning 7 actin monomers (39,62). Each tropomyosin molecule is associated with single troponin complex and 7 actin monomers (39).

Ca^{2+} interacts with tropomyosin and troponin to switch the thin filament on and off; however, in myocardium, this regulation is the subject of a more precise control (60). At low calcium concentrations, tropomyosin caps the outer domain of actin and obstructs the myosin-binding site, causing muscle relaxation (60). This steric inhibition is reversed when intracellular calcium concentration is high: TnC binds calcium and intermediates the release of TnI, inducing a conformational change in tropomyosin, exposing the myosin-binding sites, and thereby enabling the globular myosin heads to interreact (39,59,60). This causes thin and thick filaments to move towards each other, causing sarcomere shortening and eventually muscle contraction (39,59).

1.3.1.7. Troponin complex

The troponin complex consists of TnT (that binds tropomyosin), TnC (that binds calcium), and TnI (that acts as an inhibitor, blocking myosin from binding to actin) (39). The troponin complex governs the position of tropomyosin over actin, and thereby their motor activity (39,60). It is a key regulator of cross-bridge cycling (39).

TnC is Ca^{2+}-binding subunit, and has 3 isoforms: cardiac, fast skeletal, and slow skeletal. The cardiac and slow skeletal isoforms are alternatively spliced from TNNC1 (39). When calcium is bound to the EF-motif, troponin C relieves the actin-inhibited contraction through conformational changes, which allow the N-terminal domain of TnC to interact with TnI (39,63).

TnI is the inhibitory subunit of the troponin complex. When TnC interacts with TnI, it detaches from actin, thus displacing tropomyosin to enable weak binding of myosin to actin. Three isoforms of TnI are encoded by different genes: cardiac, fast skeletal, and slow skeletal (39).

The function of TnT, tropomyosin-binding subunit, is to fix the troponin complex to tropomyosin. TnT has cardiac, slow, and fast skeletal isoforms. TnT organizes the troponin regulatory complex by attaching TnI and TnC to the thin filament. TnT further plays a role in muscle contraction through the adjustment of calcium sensitivity, actomyosin ATPase activity, and force generation in the sarcomere (39).

HCM is most often associated with troponin mutations, which cause an increase in Ca^{2+} activation (63).

1.3.1.8. Myosin

Two proteins are required for contraction: actin and myosin (64). Myosin is the molecular engine of muscle contraction, converting the chemical energy of ATP into movement (39,64). It is also the predominant constituent of the thick filament. Myosins are a group of proteins with structural and functional differences (there exist 11 classes in humans). Each myosin is made of several light chains, and 1 or 2 heavy chains and molecules. The head region of the MHC binds actin and generates muscle contraction via ATP hydrolysis. The neck region of the MHC transduces force, whereas the tail region of the MHC mediates interactions with myosin (24). Binding of myosin to actin is a multi-

phase sequence: the first step is weak binding, followed by the strong binding (9). MLCs are either “essential” or “regulatory”. The regulatory light chains are highly regulated by phosphorylation and regulate the functioning of MHC (39).

Multiple MLC and MHC isoforms exist, in order to adapt to the variation in energy demands during development and in different muscles — in the heart of large mammals, MHC-α (*MYH6*) is present during development and MHC-β (*MYH7*) is dominant in the adult. MHC can shift to the MHC-α (*MYH6*) in failing hearts. This fetal gene program is triggered during cardiac remodeling by stress signaling, and probably plays a role in disease progression (39).

Outcomes of mutations in myosin usually include myosin protein aggregates under sarcolemma, impaired myosin function, and/or muscle fiber degeneration (39).

Myosin modulators have been recently developed to treat cardiomyopathies. They are usually categorized as “myosin inhibitors” (aficamten, mavacamten) or “myosin activators” (danicamtiv, omecamtiv). However, this description is oversimplified (65). Randomized, double-blind, placebo-controlled EXPLORER-HCM trial has shown the benefits of disease-specific treatment using mavacamten (66).

1.3.1.9. Myosin-binding proteins

Thick filaments contain myosin-binding proteins, which link the thin and thick filaments. The myosin-binding protein family includes the myosin-binding protein H (MyBP-H, also known as H-protein) and MyBP-C (also known as C-protein). Both are located in the C zone of sarcomere (39).

MyBP-C is a rod-like protein (67), which regulates contraction force in striated muscles (68). It exists in 3 different isoforms: a slow skeletal (sMyBP-C, encoded by *MYBPC1* in humans), a fast skeletal (fMyBP-C, encoded by *MYBPC2* in humans), and a cardiac isoform, a fundamental structural protein of the heart muscle, (cMyBP-C, encoded by *MYBPC3* in humans) (39,68). While cardiac isoform is tissue-specific, slow skeletal isoform is expressed in the inter-atrial septum and the right atrium as well (39). These isoforms are highly conserved, with more than 90% homology (68).

Myosin-binding proteins are implicated in filament assembly and contraction regulation. They support sarcomeric integrity during interplay with thin and thick

filaments, as well as with titin. cMyBP-C might act as a regulator of actomyosin interaction and as a molecular "ruler," which coordinates the spacing between thin and thick filaments. The phosphorylation of myosin-binding proteins is crucial in their regulation (39,69).

Mutations in the *MYBPC3* gene are associated with over 45% of HCM cases, and 70% of genetic variants in *MYBPC3* are nonsense mutations (68).

1.3.1.10. Protein turnover

The lifetime of a cardiomyocyte spans many years, whereas the lifetime of the individual protein molecules is limited to days or weeks. New proteins must be continuously incorporated into the sarcomere to replace old ones, in order to preserve mechanical functioning. Three different systems are implicated in the removal of old proteins: ubiquitin-proteasome system (UPS), calpain system, and autophagy-lysosome pathway; the failure of these systems is common in the progression of heart disease. Molecular chaperones perform sarcomeric assembly maintenance, and their deficiency is linked to cardiomyopathy. Oxidative and mechanical stress predispose sarcomeric protein misfolding (40).

1.4. HCM genotype

Genotype is "the genetic makeup of an organism or of a specific characteristic", "the inherited genetic material coding for all processes in the organism's life" (70). Genotype relates to a particular gene or to a combination of the alleles of individual (70).

The first HCM-associated mutation was identified in 1989 in the *MYH7* gene, which encodes MHC, a sarcomeric protein of the thick filament (9,30,36). Thus far > 1400 HCM-associated mutations have been identified, of which approximately 90% have been found in the genes encoding proteins of the thin and thick filaments of the sarcomere (*MYH7*, *MYBPC3*, and *TNNT2*, which encode MHC, cMyBP-C, and cTnT) (9,14,31,71,72).

Mutation carriers are usually heterozygous and carry one disease (mutant) allele and one normal allele. However, in some rare cases, mutation carriers are homozygous, and present early (childhood) onset and severe disease. HCM in heterozygous mutation

carriers usually appears between the ages of 20-50 years. For these reasons, the first appearance of HCM seems to be predetermined by the dose of mutant protein (9).

Approximately 60% of HCM patients have an evident familial disease. Autosomal recessive and X-linked modes of inheritance have been described but are very uncommon. An X-linked inheritance implies the possibility that the case is actually a phenocopy condition, like Fabry disease. HCM phenocopies occur in syndromic conditions (e.g. Noonan syndrome) and in storage diseases (e.g. Anderson-Fabry disease) (29).

Pathogenic or likely pathogenic variants in sarcomeric protein genes cause 30-60% of HCM cases (depending on diagnostic criteria and approaches to variant classification) (73). Mutations in genes *MYH7* and *MYBPC3* are the 2 most frequent factors, being together accountable for more than 50% of HCM cases with pathogenic variants (15,25,29,36,69,73). Genetic and clinical studies have also supported the causal, but less common, pathogenic variants in other genes encoding sarcomere proteins including *TNNT2* (encoding cardiac troponin T), *TNNI3* (encoding cardiac troponin I), and *TPM1* (encoding α-tropomyosin) (30,36). Mutations in *TNNT2, TNNI3,* and *TPM1* are together accountable for < 10% of cases. Mutations in *ACTC1* (encoding cardiac α-actin), *MYL2* (encoding myosin regulatory light chain 2), and *MYL3* (encoding myosin essential light chain 3) are also identified as causes of HCM, though rare (29,30). Therefore, HCM is considered to be a disease of the sarcomere, however, it can be caused by other factors as well (12,69,71).

Mutations in other genes have also been described in patients with HCM: *TTN* (titin), *ACTN1* (encoding α-actinin), *MYH6* (myosin heavy chain or α-MHC), *TCAP* (telethonin), *CSRP3* (muscle LIM protein – a Z-disc protein), *FHL1* (four-and-a-half LIM domains 1), *PLN* (phospholamban), *ACTN2* (alpha-actinin-2, a Z-disc protein), *CRYAB* (crystallin alpha B), *FLNC* (filamin C), *MYOZ2* (myozenin 2, a Z-disc protein), *TNNC1* (cardiac troponin C), *TRIM55* (tripartite motif containing 55), and *TRIM63* (ubiquitin E3 ligase tripartite motif protein 63 or MuRF1) (25,29,35,36,73). While some of these genes are likely to be causal genes for HCM, others are only associated with HCM (Table 1-3) (29).

Table 1. Causal genes for HCM, adapted from (15) and (29)

Gene	Frequency in HCM (%)	Protein	Location in sarcomere	Function
MYH7	~40	β-Myosin heavy chain	Thick myofilament	ATPase activity, Force generation
MYBPC3	~40	Myosin binding protein-C	Thick myofilament	Cardiac contraction
TNNT2	~10	Cardiac troponin T	Thin myofilament	Regulator of acto-myosin interaction
TNNI3	< 5	Cardiac troponin I	Thin myofilament	Inhibitor of acto-myosin interaction
TPM1	< 1	α-tropomyosin	Thin myofilament	Places the troponin complex on cardiac actin
ACTC1	< 1	Cardiac α-actin	Thin myofilament	Acto-myosin interaction
MYL2	< 1	Regulatory myosin light chain	Thick myofilament	Myosin heavy chain 7 binding protein
MYL3	< 1	Essential myosin light chain	Thick myofilament	Myosin heavy chain 7 binding protein
CSRP3	< 1	Cysteine and glycine-rich protein 3	Z-disc	Muscle LIM protein (MLP), a Z disk protein

Table 2. Likely causal genes for HCM, adapted from (15) and (29)

Gene	Protein	Function
FHL1	Four-and-a-half LIM domains 1	Muscle development and hypertrophy
MYOZ2	Myozenin 2 (calsarcin 1)	Z disk protein
PLN	Phospholamban	Regulator of sarcoplasmic reticulum calcium
TCAP	Tcap (Telethonin)	Titin capping protein
TRIM63	Muscle ring finger protein 1	E3 ligase of proteasome ubiquitin system
TTN	Titin	Sarcomere function

Table 3. Genes associated with HCM, adapted from (15) and (29)

Gene	Protein	Function
ACTN2	Actinin, alpha 2	Z disk protein
ANKRD1	Ankyrin repeat domain 1	A negative regulator of cardiac genes
CASQ2	Calsequestrin 2	Calcium binding protein
CAV3	Caveolin 3	A caveolae protein
JPH2	Junctophilin2	Intracellular calcium signaling
LDB3	Lim domain binding 3	Z disk protein
MYH6	Myosin heavy chain alpha	Sarcomere protein expressed at low levels in the adult human heart
MYLK2	Myosin light chain kinase 2	Phosphorylate myosin light chain 2
NEXN	Nexilin	Z disc protein
TNNC1	Cardiac troponin C	Calcium sensitive regulator of myofilament function
VCL	Vinculin	Z disk protein

Mutations in several other genes: *PRKAG2* (γ2-subunit of AMP kinase), *GLA* (α-galactosidase A), and *LAMP2* (lysosome-associated membrane protein 2), are detected in ~ 2% of cases misdiagnosed as having HCM. Mutations in these genes implicate disparate mechanisms to induce hypertrophy, but also lead to expression of supplementary clinical phenotypes that do not appear in HCM (73). There is a mechanistic difference between phenocopy conditions and HCM: ventricular hypertrophy in phenocopy conditions results at least partly from storage of material (e.g. glycogen) and partly from functional defects in myocytes (e.g. impaired contraction) (29).

For majority of HCM genes, both *de novo* and familial pathogenic variants have been described (36).

The percentage of novel variant detection is 35-40%, with 56% of variants found in a single family and therefore considered "private" variants (74). Haplotype analyses of an identical variant in unrelated patients demonstrated that such genetic variants emerge independently (36). However, many unrelated *MYBPC3* variant carriers possess the same haplotype architecture in some homogeneous subpopulations in the Netherlands (75), Finland (76), Iceland (77), Japan (78), and India (79), which demonstrate clear founder effects in HCM. Due to late-onset and benign presentations, the presence of HCM-associated founding mutations in these populations demonstrates neutral or mild negative selection (36).

Causality cannot be definitely identified, especially in sporadic cases or small families. As a result of great genetic variety, the HCM-causing genes left undefined in ~ 40% of cases (29).

A subset of HCM cases (~ 5-7%) is attributable to digenic or oligogenic heterozygosity (36,72). Such conclusions not conform to the classic single gene disease definition, moreover suggest that a number of variants could jointly cause HCM phenotypes and each variant has a mild to moderate effect. Complex genotypes, inclusive of compound heterozygous or homozygous variants, have been described (36).

In general, genetic variants range from clinically negligible to causal. Genetic variants with very large effects are considered to be accountable for autosomal dominant single gene disorders. They have been detected by robust genetic methods, such as linkage and co-segregation analyses. Illustrations related to HCM involve genetic variants in

genes *MYH7* and *MYBPC3*. Genetic variants with clinically small or indistinguishable effects modestly impact the phenotype, still do not cause monogenic disorders. In the middle of the spectrum lies a subset of genetic variants exerting intermediary to large effect sizes and exhibiting incomplete penetrance — their effects are affected by other non-genetic and genetic factors. Such variants cause HCM with incomplete or low penetrance (29). Incomplete penetrance suggests that the majority of HCM mutation carriers have a low risk of pathological HCM phenotypes development over their lifetime (80).

The frequency of particular mutations in an HCM population is markedly low: the only two exceptions are the p.Arg502Trp mutation in *MYBPC3* (identified in ~1.5-3% of HCM cases) and the p.Val762Asp mutation in *MYBPC3* (occurring in 3.9% of the Japanese population) (29) — "p." stands as label used in standard nomenclature of mutations based on the amino acid sequence of the protein (81). Other mutations appear at a frequency of < 1% in the HCM population, and around half of them are detected in a single family or proband (29). Some other variants reported to be relatively common and "hot spots" are: Arg403Gln, Arg453Cys, and Arg663His in *MYH7*; Arg92Gln, Arg92Trp, and Arg104Val in *TNNT2*; as also Arg495Gln and c.1928-2A>G in *MYBPC3* (6) — the prefix "c." stands as label and indicates the use of standard mutation nomenclature based on reference coding DNA sequences (81). Therefore, significant "hot spots" for mutations in any of the known genes do not exist (29).

1.4.1. Digenic/oligogenic etiology of HCM

One mutation is generally sufficient to cause HCM, however, due to variable penetrance and expression, not always (23). Around 5% of HCM cases exhibit two (digenic) or more (oligogenic) causal mutations in the same or different genes (26,29).

1.4.2. Missense and nonsense mutations in HCM

The most pathogenic and likely pathogenic sarcomeric gene variants causing HCM encode missense residues, leading to a nonsynonymous amino acid substitution, and proposing a dominant negative effect. The notable exception is *MYBPC3*, wherein nonsense mutations are most often caused by frameshifts, insertions/deletions, or splice-

site variants. These create a premature stop codon and hence truncated protein is produced, which further leads to haploinsufficiency of the protein (29,35,36,73,82,83). There exist over 300 pathogenic HCM-associated mutations in the β-cardiac myosin (83).

In general, the missense mutations alter structure (e.g. protein misfolding, alteration of important kinase domains, amino acid substitution in a highly-conserved residue) and function (e.g. change in surface-exposed section of a molecule alters protein-protein interreaction), by altering the amino acid constitution of the encoded protein (29,35). Missense mutations usually cause the mutant protein to be integrated into sarcomere, however, its interactions with the regular proteins hamper normal sarcomeric functioning (poison polypeptide hypothesis) (35).

Missense mutations causing structural alterations in the encoded protein might decrease the effectiveness of the sarcomere formation. In the sarcomere of heterozygous individual, both alleles of the causal gene (normal and containing missense mutation) are transcribed and translated into the matching wild-type and mutant proteins. Effectiveness of transcription and translation of the wild-type and mutant alleles may differ (expected to be reduced for the mutant allele). Allelic imbalance is partly amended by the wild-type allele, but this allelic compensation is often incomplete. As a result, the mutation might reduce expression, and cause a moderate insufficiency of the corresponding protein. In HCM, there is a myocyte-to-myocyte variation in the expression of transcripts of mutant alleles. This heterogeneity may elucidate the varying phenotypic expression of HCM (29).

As a result of insertion/deletion mutations that induce a frame shift changes, altered proteins are produced. They are commonly targeted for degradation, leading to haploinsufficiency (29,35). Transcription and translation regulatory mechanisms prevent biosynthesis of the truncated proteins (29).

In line with the variety of HCM-associated mutations, initial defects in HCM are also diverse. Initial changes in HCM include altered translation efficiency and transcription rates, alterations in the affected protein structure, as well as in the functioning of the sarcomere (29).

1.4.3. Allelic imbalance and haploinsufficiency

Allelic imbalance and haploinsufficiency represent the main mechanisms of many Mendelian diseases (84).

Allelic imbalance represents disturbances in genetic homeostasis. It occurs when a higher protein expression is produced from one allele compared to another (deviating from the anticipated 1:1 expression proportion). Protein allelic imbalance occurs if the mutation cause production of altered protein that interferes with intrinsic protein folding, protein-protein interactions, or protein quality control, usually eventually decreasing mutant protein stability. Allelic imbalance is displayed by 25% of human genes. In case of HCM, allelic imbalance has been reported for several missense mutations in *MYBPC3*, *MYH7*, *TNNT2*, and *MYL2*. The ratio of wild-type to mutant protein differs regarding particular mutation (some specimens showed lower and some a higher proportion of mutant to the wild-type protein) (84).

Haploinsufficiency represents disturbances in protein homeostasis (84). Haploinsufficiency arises when a heterozygous mutation results in one allele of a gene being deleted or inactivated, and when a single functional copy of a gene is insufficient to maintain normal level of protein (69,84). The level of the functional protein does not have to be less than half of the normal, but it must be below the threshold required for proper functioning. This might also happen when a mutant protein is present still non-functional (as in the case of in-frame mutations) (84).

1.4.4. Genetic testing

The HCM prevalence has become higher with the usage of advanced genetic testing granting a molecular diagnosis ahead of clinical diagnosis (15). The positive result of genetic testing supports the diagnosis of HCM in a proband, but a negative result does not exclude it (29). If a proband with a positive genetic test is discovered, testing for the existence of the variant in relatives (cascade screening) needs to be executed (29,30). Detection of causative mutations in a proband with HCM enables presymptomatic diagnosis and clinical surveillance of relatives, as well as appropriate genetic counseling (6).

Relatives, who are not carriers of the causal mutation, are highly unlikely to develop HCM. Relatives of patients identified as HCM mutation carriers should be clinically examined. Phenotype-positive relatives might be diagnosed with HCM. Mutation-positive and phenotype-negative relatives should be clinically evaluated annually, or more often, if symptoms occur (29).

Many mutation-positive and phenotype-negative relatives will ultimately express the HCM phenotype. Nevertheless, because of incomplete penetrance, some mutation carriers never develop the HCM phenotype. Nonetheless, they need to be aware that they can pass the mutated gene to their offspring (29).

Genetic testing allows differentiation of HCM and the phenocopy conditions (29). Genes like *TTR*, *PRKAG2*, *LAMP2*, *GLA*, and *GAA* are associated with metabolic disorders that resemble HCM, but their clinical profiles, inheritance patterns, and treatments differ from one another and from HCM (85). Genetic testing for HCM detects phenocopy conditions in ~ 3% of individuals (29). Contemporary genetic testing for HCM fails to detect 50-60% of HCM patients (29).

Only 30-60% of individuals with clinical diagnosis of HCM carry sarcomere gene mutations (27,80,86). Up to 40% of the HCM cases are isolated and sporadic, wherein the proband does not have any known HCM mutations or a family history of HCM (80). Detection of pathogenic mutations is greater in individuals with a positive family history and in younger individuals, where it can outreach 50-60%. In other individuals, it is ~ 30-40% (12).

Many non-sarcomeric genes linked with HCM have been increasingly added in panels for HCM, despite limited evidence of a causal role in HCM. The choice of genes included in routinely used diagnostic panels, eventually lies on the spectrum between the extremes of adding only the validated genes on one side and the screening of all genes linked with HCM (irrespective of the strength of evidence) on the other (85).

Genetic screening of healthy individuals or patients with moderate hypertrophy is not advised. It can be deemed reasonable only in exceptional borderline cases (6).

1.4.5. Animal models and *in vitro* studies

Animal models do not completely reproduce human HCM (87). For instance, there exist major differences between humans and small rodents in the constitution of sarcomeric proteins isoforms, in particular, titin and MHC. On the other hand, phosphorylation sites are highly conserved (9). Knock-out, knock-in, and transgenic mice models have been used in deciphering pathogenesis mechanisms of HCM. However, functional research of HCM mutations in such models are complicated by the differences in sarcomere protein contents between them and humans. The MCH isoform encoded by *MYH7* is predominant in the human heart, whereas the myosin isoform encoded by *MYH6* is predominant in the murine heart. In order to study human mutation in *MYH7*, it has to be first transposed into *MYH6* in the mice models. Such transposition is not possible, since proteins encoded by *MYH6* and *MYH7* display major differences in their ATPase activities and acto-myosin kinetics (ATPase activity and the velocity of actin displacement of the *MYH6*-encoded protein are several folds higher than those of the *MYH7*-encoded protein). These differences hinder extrapolation of the discoveries in model organisms that predominantly express *MYH6* to human HCM. Research performed on hearts that predominantly express *MYH7* rather than *MYH6* (e.g. transgenic rabbits) propose discoveries that are more relevant to the functional outcomes of HCM mutations (29). In small rodents, the stiff (smaller) N2B titin isoform is predominant, whereas in pigs and humans, the compliant (long) N2BA titin isoform is predominant (9).

Single-cell studies stress the importance of considerable myocyte-to-myocyte variability of the gene expression and functioning as well as of Ca^{2+} sensitivity (29).

Animal and cell studies confirm altered calcium homeostasis to be a key component of pathophysiological processes leading to the development of LV hypertrophy (35).

1.5. HCM phenotype

Phenotype (from Greek words *phainein* — to show, and *typos* — type) is the composite of all the organism's characteristics or traits (70).

The clinical presentations of HCM are very variable. Some mutation carriers stay asymptomatic, while others develop severe HCM, terminal HF, or SCD (9,15,88).

HCM is characterized by asymmetric hypertrophy of the ventricle, most often of the septum, which sometimes obstructs blood flow to the aorta (obstructive HCM). Besides hypertrophy, HCM patients display myofibrillar and cellular disarray, fibrosis, and cardiac remodeling. Reduced myocardial efficiency and diastolic dysfunction appear even before the development of hypertrophy in asymptomatic mutation carriers. These initial changes in structure and functioning might be explained by defects in the sarcomere caused by HCM mutations (9).

1.5.1. Effects of mutations in sarcomeric protein-encoding genes

HCM mutations induce raised myofilament Ca^{2+} sensitivity. Increased tension caused by the HCM mutation initiates hypertrophic signaling via activation of MEK1–ERK and calcineurin signaling, thereby causing a rise in cardiomyocyte width and cardiac mass specific for concentric hypertrophy. Inhibition of MEK1–ERK signaling cause cardiomyocyte elongation, which is specific for eccentric hypertrophy. Therefore, it is suggested that the tension-based activation of diverse hypertrophic signaling underlies the hypertrophy (9).

Ca^{2+}-sensitizing effects of HCM mutations are potential substrates for ventricular arrhythmias. The degree of myofilament Ca^{2+} sensitization is mutation-specific (9).

1.5.1.1. Increased crossbridge kinetics and ATP utilization

It is suggested that HCM mutations lead to energy depletion by increasing ATP turnover during the crossbridge cycle. In HCM mutation carriers, reduced myocardial efficiency already exists at an early stage of HCM before the development of hypertrophy (9).

The HCM mutation-induced sarcomere alterations induce adverse remodeling and cause electrophysiological perturbations: action potential is extended, associated with CaMKII-mediated alterations in Ca^{2+} and late sodium (Na^{+}) currents. Cardiomyocytes of HCM patients show higher diastolic Ca^{2+} concentrations and increased occurrence of cellular arrhythmias, which is explained by disease-related phosphorylation changes of Ca^{2+}-handling proteins and sarcomere (9).

1.5.1.2. HCM: A disease of compensatory hypertrophy

At first, it was suggested that the incorporation of mutant proteins reduces contractility, which could lead to compensatory hypertrophy. Some authors find this suggestion inconsistent with laboratory and clinical findings. There are 3 arguments supporting the above perspective: 1. At first, experimentations with mutant proteins showed reduced functioning of mutant proteins, which had been earlier displayed as decreased motility of the proteins. However, more recent research has showed that some mutations actually cause increased motility in mutant proteins: mutation carriers, which do not develop hypertrophy, show increased motility of heart tissue that may be observed in an ECHO. Hence, the decrease of function of mutant proteins cannot be the only hypertrophy stimulus. 2. Hypertrophy in HCM is assymetric — it is clearly distinguishable from concentric hypertrophy observed in hearts with volume overload (e.g. in hypertension). 3. Patients usually develop HCM after puberty. This does not support hypertrophy as a compensatory mechanism, since mutation is present from the very beginning of heart development (89).

1.5.1.3. HCM: A disease of disrupted Ca^{2+} sensitivity

Several research studies have proposed that there is a disrupted Ca^{2+} sensitivity underlying HCM. Mutations in various genes encoding myofilament proteins increase the Ca^{2+} sensitivity. This in turn disturbs intracelular Ca^{2+} homeostasis, sarcoplasmatic reticulum Ca^{2+} uptake, and phosporylation of some proteins, which probably contribute to some aspects of HCM (89).

1.5.1.4. HCM: A disease of disturbed myocardial relaxation

Diastolic dysfunction is one of the prominent characteristics of HCM. It manifests in form of low Ca^{2+} concentrations during diastole, which precedes development of hypertrophy in HCM. Mutations could exacerbate the disease through the following: a) increase in mutant protein dosis, b) stimulation of mutant protein incorporation into sarcomere, c) favoring of fetal isoforms of other (non-mutant) sarcomeric proteins, and d) occurrence of post-translation modifications. Diastolic dysfunction could be also associated with energetic changes in myocardium, which drive a significant increase in

cellular ADP concentration. Rise in ADP concentration is a direct outcome of mutations that increase the myofilament's ATP-ase activity (89).

1.5.1.5. HCM: A disease of disrupted myocardial energy balance

There exists evidence of decreased phosphocreatin/ATP ratio in HCM, which indicates an insufficient energy regeneration in the myocardium. This supports the concept of energy deficient hypertrophic myocardium, and suggests that HCM mutations cause excessive cellular usage of ATP (which is also in the basis of the hypothesis that HF is a "fuel-less engine"). Reduced heart efficacy is present in both asymptomatic (mutation carriers) and symptomatic cases of HCM. Decrease of efficacy in asymptomatic mutation carriers suggests that energy deficits precede the development of myocardial hypertrophy in HCM (89).

1.5.1.6. HCM: A disease of changed concentrations of cellular ADP

It is well-known that ATP depletion rules the HCM development. However, this concept is questioned by reports showing that absolute cellular ATP concentrations never decrease to the extent of becoming a limiting factor in cardiomyocyte relaxation. There are 2 potential mechanisms that could explain how increased ADP concentrations could cause a rise in calcium concentration during diastole: 1. by direct buffering of Ca^{2+} in myofilaments (Ca^{2+} are captured by "sticky" myofillaments) and 2. by decrease in Ca^{2+} uptake by the sarcoplasmatic reticulum via Ca^{2+}-ATP-ase. Increased concentrations of ADP also influence membrane-membrane junctions via connexin 43 (89).

1.5.1.7. Myocardial ADP increase: Result of inefficient energy buffering

The swift regeneration of ATP from ADP in cardiomyocytes highly depends on the synchronization of mitochondrial ATP generation and its consumption in myofilaments. Mitochondrial structural and functional derangements in HCM patients may additionaly aggravate limited myocardial capacity for ATP regeneration. Reduced reserves of myocardial phosphocreatine and reduced creatine-kinase-dependent ATP regeneration are observed in HCM. Further, in HCM, increased myosin ATP-ase activity in

combination with cellular hypertrophy increases the strain on the ADP-buffering capacity of the phosphocreatine-creatine-kinase shuttle (89).

1.5.1.8. Mitochondrial ADP workload, creation of reactive oxygen species (ROS), and hypertrophic remodeling

In a healthy heart, excitation and contraction are closely linked with mitochondrial ATP production. During a rise in the cardiac workload, the elevated amplitude of Ca^{2+} concentration increases ATP consumption and thus speeds up ADP delivery to mitochondria. A hypertrophied myocardium requires more ATP than a healthy myocardium for any given calcium concentration. The resultant oxidative stress promotes hypertrophic cardiac remodeling via ERK signaling (89).

1.5.1.9. Inefficient myocardial coronary perfusion in HCM

The myocardial coronary perfusion is insufficient in case of HCM. In a healthy heart, coronary perfusion takes place in the time of diastole, while coronary flow is reduced during systole, as a consequence of the coronary vessels narrowing within the ventricular wall. Increased Ca^{2+} sensitivity of the myofilament leads to disrupted ventricular relaxation, which eventually restricts the coronary perfusion of the hypertrophied myocardium, especially during increased cardiac workload (e.g. during exercise). On the other hand, disrupted coronary perfusion results in decreased energy regeneration, which elevates ADP (and inorganic phosphate) concentration, decreases the phosphocreatine pool, and alters cellular pH. Although initial changes caused by mutations in cellular energetics might be reversible, ischemic injury causes irreversible myocardial changes (89).

1.5.2. Pathophysiologic features of HCM

Pathophysiological features of HCM are: cardiomyocyte hypertrophy (90–93) and disarray (91–94), myocardial remodeling (92,95,96) and fibrosis (3,23,92,97–99), coronary microvascular dysfunction (1,92,93,100,101), myocardial ischemia (1,18,92,101,102) and hypercontractility (10,92,103–105), impaired myocardial

relaxation (10,92,97,106), myocardial stiffness (92,94,97), and diastolic dysfunction (92,94,107).

1.5.2.1. Cardiomyocyte hypertrophy

Cardiac hypertrophy is a compensatory mechanism for managing biomechanical stresses, and for the maintenance of proper cardiac homeostasis and output (e.g. during exercise, pregnancy, and in hypertension, valvular disease, etc.) (108–110). While the fetal heart increases its mass primarily through cardiomyocyte proliferation, postnatal heart hypertrophy engages increased cardiomyocyte size (111) — adult cardiac hypertrophy involves an increase in the cardiomyocyte size, rather than an increase in the cardiomyocyte number. Several regulatory mechanisms are involved in cardiac hypertrophy: proliferation, translational regulation, epigenetic modifications, multiple signaling pathways, and immune responses (108–110).

It is not clear if cardiomyocyte hypertrophy is the cause for HCM or the result of the adaptive reactions of the heart to HCM (e.g. genetic disorders) (112).

1.5.2.2. Cardiomyocyte disarray

Extensive cardiomyocyte disarray is the hallmark of HCM, distinguishing it from other etiologies of LVH, for instance, those caused by pressure overload alone (80,113). The histopathology of endomyocardial biopsy specimens of patients with HCM displays enlarged cardiac myocytes in disarray with loss of normal parallel alignment, which gives the myocardium a whirl-like appearance (15,29). The cardiac myocytes contain pleiotropic nuclei, are enlarged, and have bizarre shapes. The myofibrillar disarray usually affects > 10% of the HCM myocardium; it is widely distributed, with a higher predisposition of the hypertrophied interventricular septum (29). The degree of myofibrillar disarray appears to be associated with the disease progression (113) (severe myocyte disarray has been observed in patients in whom SCD has occurred) (29).

Contractile imbalance — an unequal force generation by neighboring cardiomyocytes — might be a cause of myofibrillar disarray. Differences in force generation among cardiomyocytes from HCM patients are found to be much larger than those in the control group, at the same calcium concentration. It is proposed that this

heterogeneity in force generation among cardiomyocytes may disrupt the functional syncytium of the myocardium and lead to myocardial disarray (113,114).

1.5.2.3. Myocardial remodeling

Myocardial remodeling refers to alterations in heart architecture as a result of a variety of causes (115,116). It comprises metabolic, morphological, or electrical alteration (117). Myocardial remodeling is a multifactorial and complex maladaptive response associated with almost all cardiac diseases (118,119), and is also common across various types of cardiomyopathy (117). It involves physiological alterations in the mass, structure, form, and function of the cardiac cells as also the heart itself (119). In the context of cardiac hypertrophy, it represents a compensative response addressing stress or volume overload, by maintaining the cardiac output and lowering the increased wall tension. On the other hand, persistent overload accelerates interstitial fibrosis, cardiomyocytes loss, and cardiac failure (117).

Significant loss of cardiomyocytes and slower rate of cardiomyocyte regeneration represents the starting point for the myocardial remodeling. The main contributors to pathological myocardial remodeling are cardiomyocytes injury, cardiac hypertrophy and fibrosis, inflammatory response, and angiogenesis (119). Hyperdynamic contraction might be an essential mediator of the cardiac remodeling in HCM (120). Cardiac fibroblasts contribute to cardiac remodeling, specifically to hypertrophy and fibrosis. Their multifactorial contribution to myocardial remodeling makes them an attractive prospective therapeutic target (116).

Cardiac remodeling refers to the rearrangement of normal cardiac structures. It represents a chronic maladaptive progressive process characterized by apoptosis, fibrosis, necrosis, matrix components remodeling, myocardial hypertrophy, ventricular dilatation, and vascular dysfunction. Multiple pathogeneses are implicated in cardiac remodeling: genetic mutations, accelerated cell apoptosis, extracellular matrix anomalies, ischemia-related damage, increased hemodynamic overload, dysregulated neurohumoral stimulation, and immunological activation (117).

1.5.2.4. Myocardial fibrosis

Myocardial fibrosis is a common pathological process across several cardiac pathologies (121,122) and usually indicates poor prognosis (121). It is an increase in the quantity of collagen in the myocardium compared to cardiac myocytes (121–123). Myocardial fibrosis is a pathological process, wherein healthy myocardial tissue is replaced by fibrotic tissue and lose its normal function (124).

Myocardial fibrosis may be classified as replacement ('scarring') fibrosis and interstitial fibrosis. Replacement fibrosis is the result of both non-ischemic and ischemic myocyte loss. In such circumstances, collagen replaces injured myocytes. Interstitial fibrosis is a scattered distribution of excess collagen in the extracellular matrix, which may be reversible and treatable. It is not inevitably associated to myocyte injury. Interstitial and replacement fibrosis oftentimes exist together and represent varied components of pathologic cardiac remodeling (121–123). They are both major independent predictors of adverse cardiac outcomes (121,123).

Myocardial fibrosis is a result of the complex interplay of pro-fibrotic cell types, hormones, growth factors, and pro-inflammatory cytokines (121). It is associated with changes in the composition of collagen fibers type (i.e., overabundance of a particular collagen type) and their physical, chemical, and mechanical properties (i.e., excessive cross-linking) (123).

Myocardial fibrosis is found in up to 80% of HCM cases. In HCM patients who suffer SCD, many foci of myocardial scarring are found. Scarring frequently affects the midventricular septum and anterior wall, and is usually associated with severe hypertrophy. Besides the diffuse involvement of fibrosis in areas affected by hypertrophy, there are also right ventricular fibrosis insertion points, even in minimally symptomatic patients. In terminal HCM, over 1/3rd of the LV myocardium is replaced by fibrosis. The spatial interrelation between remodeled arterioles and fibrosis proposes ischemia as an underlying pathogenesis (123). Myocardial fibrosis is a suitable substrate for the maintenance and recurrence of arrhythmias (125).

1.5.2.5. Coronary microvascular dysfunction

HCM patients present coronary microvascular dysfunction (a decrease in coronary vasodilator reserve) in non-hypertrophied and hypertrophied LV sections, in the absence of epicardial coronary artery stenosis (1,126). Reduction in coronary flow reserve is a good predictor of future cardiovascular events (126).

Coronary microvascular dysfunction is caused by various factors: decreased capillary density, vascular remodeling, myocyte disarray, fibrosis, diastolic dysfunction, extravascular compression as a result of ventricular hypertrophy, and left ventricular outflow tract obstruction (LVOTO) (1,127).

Potential mechanisms of the coronary microvascular dysfunction are diverse: increased coronary vasoconstrictive reactivity at microvascular level [such as coronary microvascular spasm — replication of angina symptoms and ischemic electrocardiogram (ECG) changes, but without epicardial spasm during intracoronary acetylcholine provocation test], impaired endothelium-dependent and endothelium-independent coronary vasodilator capacities, and increased coronary microvascular resistance secondary to structural factors (such as vascular remodeling, luminal narrowing, vascular rarefaction, and external compression) (1,128). Besides the structural and functional alterations in small vessels, the hypertrophied myocardium requires more oxygen, and diastolic dysfunction contributes to imbalance between the oxygen demand and supply. All of this supports further development of myocardial ischemia (1,127).

Repetitive episodes of ischemia and chronic ischemia resulting from microvascular dysfunction progressively cause cardiomyocyte death and replacement fibrosis. These in turn lead to left ventricular remodeling with diastolic and, ultimately, systolic dysfunction (1,127).

HCM patients with coronary microvascular dysfunction and ischemia can remain asymptomatic, or can present symptoms of angina, atypical chest pain, or dyspnea. Atypical chest discomfort during effort or after meals is more often in HCM than in typical angina. Adults with HCM may simultaneously have atherosclerotic coronary artery disease, which worsens the prognosis (1).

An abrupt increase in oxygen consumption [like during exercise or atrial fibrillation (AF)] can produce ischemia when microvascular dysfunction is present. This could

provide explanation for ventricular arrhythmias and abnormal blood pressure during physical activity. Degree of microvascular dysfunction is suggested as being important to the HF development as well (1).

In cases of HCM, severe impairment of microvascular function is more frequent in individuals with sarcomeric mutations. Coronary vasodilator reserve is lower in individuals with obstructive HCM compared to those with non-obstructive HCM. Alcohol septal ablation (ASA) results in improvement of the coronary vasodilator reserve (126).

1.5.2.6. Myocardial ischemia

Several mechanisms contribute to myocardial ischemia in HCM (the same as for coronary microvascular dysfunction) and may act separately or simultaneously (127): systolic anterior motion (SAM) of the mitral valve, abnormal coronary prefusion, and microcirculatory resistance (101).

Normally, nutrients are delivered to tissues and metabolic waste is taken away from the tissues through circulating blood. Insufficient blood flow to a tissue (in this case myocardium) cause ischemia. Myocardial ischemia is most often caused by obstruction of coronary arteries, followed by a drop in oxygen tension within the myocardium. Myocardial ischemia markedly hampers the oxidative metabolism, and consequently induces energetic stress within cardiomyocytes. It slows down or blocks mitochondrial respiration and thereby oxidative phosphorylation. As a result, the production of ATP is reduced. In the lack of oxygen, glycolysis is intensified, which causes excessive production of lactic acid, and consequently intracellular acidosis (129).

In the lack of ATP, the Ca^{2+} pump within cardiomyocytes is not able to efficiently remove Ca^{2+} from the cytoplasm (129). Alterations in extracellular and intracellular Ca^{2+} control cause further Ca^{2+} overload (93). These abnormal Ca^{2+} levels numb ventricular contraction and promote cardiac dysfunction (129). Excessive contractions of energy-dependent muscle fibers cause electrophysiological disorders, myocardial structural damage, and reduced function (130).

Myocardial ischemia contributes to atrial and ventricular arrhythmias, sudden death, LV remodeling, and systolic dysfunction (1).

1.5.2.7. Myocardial hypercontractility

HCM is associated with hypercontractility (9). Impaired diastolic dysfunction with maintained or enhanced systolic function are the earliest presentations of HCM, which are caused by genetic mutations (9,83). These are apparent even before hypertrophy is detected (83). At the myofilament level, HCM-associated hypercontractility is frequently associated with increased Ca^{2+} sensitivity of the contractile apparatus and accelerated cross-bridge cycling (cross-bridges are hypercontractile, irrespective of Ca^{2+}) (131,132). Since hypercontractility occurs before hypertrophy, it is suggested that hypercontractility at the molecular scale is a direct cause of HCM (133). Increase in Ca^{2+} sensitivity causes hypercontractility, which further promotes hypertrophy (132).

1.5.2.8. Impaired myocardial relaxation

Myocardial relaxation reflects the activity of sarcomeres returning to precontractile condition (100). It is impaired in HCM due to altered calcium sensitivity (98). Myocardial relaxation is dependent on nitric oxide (NO) — local administration of NO improves LV relaxation. In HCM, a raise in the vascular endothelial growth factor (VEGF) is associated with fractional shortening of the LV and impaired ejection fraction (106).

1.5.2.9. Myocardial stiffness

Myocardial relaxation and myocardial stiffness have an essential role in diastolic LV function (134). Myocardial stiffness is affected by aging, because of cellular and extracellular matrix alterations and progressive physiological changes (135). Diastolic function mostly depends on ventricular compliance, which has a negative relationship with myocardial stiffness. Therefore, myocardial stiffness is markedly involved in disorders showing impaired cardiac compliance (136).

Myocardial stiffness mostly originates from cardiomyocyte- and extracellular matrix (ECM)-generated passive stiffness, but is also a variable property of the tissue (134,136). Diverse pathological alterations are implicated in myocardial stiffness, such as increased collagen in ECM, cardiac fibrosis, intrinsic cardiomyocyte stiffness as a result of cytoskeleton impairment, along with metabolic disorders (136).

As stated earlier, cardiac remodeling is a feature of HCM. Cardiac remodeling involves cardiomyocyte hypertrophy, the deposition of extracellular matrix proteins, and fibroblast proliferation. All these changes contribute to tissue stiffness and alter the extracellular mechanical stimulation of cardiomyocytes (which in turn affects their biomechanical properties). Collagens and titin enable cardiomyocytes to preserve their physiological stiffness. Proper physiological stiffness is essential to maintain the mechanical pressure of the beating heart. Pathological alterations in myocardial stiffness impair heart function (137).

Although actin and myosin play only minor roles in passive myocardial stiffness, both extracellular and intracellular components contribute to physiological and pathophysiological cardiac stiffness. On the extracellular side, cardiomyocytes are placed within an extracellular matrix. It consists of proteoglycans, glycosaminoglycans, collagens, laminins, and fibronectins. It is implicated in cell motility, adhesion, and signaling of cardiomyocytes. Molecular composition and cross-linking alter during development or disease, thereby changing the physiological stiffness of the myocardium. Cardiac stiffness changes during the transition from neonatal to adult hearts, due to changes in the quantity of collagen and in the collagen type I: collagen type III ratio. Collagen type I increases rigidity, whereas collagen type III enables elasticity. Further, upregulation of lysyl oxidase and increased collagen-lysyl oxidase crosslinking enhance myocardial stiffness, and result in diastolic dysfunction (137).

The chief regulator of intracellular stiffness in cardiomyocytes is titin. Titin plays the main role in the myofibrillar passive tension reaction to stretch in striated muscle cells. Mutations in *TTN* (the gene encoding titin) or its post-translational modifications triggered by disease, result in a change in the passive myocardial stiffness (137). The proportion of the titin isoforms — N2BA:N2B also influences diastolic dysfunction (136).

Desmin is the main intermediate filament in the cardiac muscle. It also contributes to the passive stiffness of cardiomyocytes. In cardiac diseases, desmin is phosphorylated, and its expression levels are modified (137).

1.5.2.10. Diastolic dysfunction

Some level of LV diastolic dysfunction is observed in almost all HCM patients. This dysfunction precedes cardiac hypertrophy (29,71,89,138).

Diastolic dysfunction is caused by the increased interstitial fibrosis, decreased relaxation, and increased stiffness of the thickened ventricular wall (29,138). Diastolic dysfunction causes a rise in the LV end diastolic pressure, which leads to elevation of left atrial, pulmonary venous, and pulmonary capillary pressures. The LV diastolic pressure rises strikingly during exertion, which causes exertional dyspnea, orthopnea, exercise intolerance, peripheral edema, and HF with preserved ejection fraction (29).

In the context of HCM, titin plays a critical role in diastolic dysfunction in HCM: post-translational modifications of titin (e.g. S-gluthationylation and phosphorylation) affect the elasticity of cardiomyocytes and diastolic properties of the LV (139).

Diastolic dysfunction is also characterized by high myocardial activation at low diastolic Ca^{2+} concentrations. High basal myofilament activation is usually sufficient to delay ventricular relaxation onset, and to restrict proper ventricular filling. This high basal myofilament activation is either a direct result of changes caused by an HCM mutation, or stems from secondary alterations caused by HCM mutations. HCM mutation-induced changes might involve: increased doses of mutant protein, increased incorporation of mutant protein into sarcomere, isoform switch to fetal isoforms, or post-translational modifications (89).

Additionally, it is proposed that high levels of ADP are the result of disturbed ATP regeneration and higher sarcomeric energetic requests, and represent a prominent cause of diastolic dysfunction (71,89). Diastolic dysfunction might first be caused by increased myofilament Ca^{2+} sensitivity and consequent great cross-bridge activity during diastole. Such dysfunction is likely to be exacerbated by ADP-mediated Ca^{2+} sensitization, decreased β-adrenergic receptor signaling, and oxidative stress (89,140). Severe diastolic dysfunction might lead to microvascular dysfunction (since coronary perfusion occur during diastole), and eventually to local ischemia as well as replacement fibrosis (140).

1.5.3. Patterns of left ventricular hypertrophy in HCM

The distribution of LV hypertrophy in HCM is diverse regarding both location and extent. HCM might affect any position within the LV (36), still the most common pattern of LV hypertrophy in HCM is asymmetric septal hypertrophy, followed by mid-ventricular, apical, concentric, and mass-like hypertrophy (Figure 2) (3,4,29,141–143). The right ventricle (RV) is rarely affected by hypertrophy (29). Differentiation of morphological subtypes of HCM does not contribute much to treatment strategies, except in the case of apical HCM. Apical HCM is "more sporadic, sarcomere mutations are detected less frequently, there is more AF, and SCD risk factors differ" (142). Hypertrophy of any structure adjoining the left ventricular outflow tract (LVOT) results in narrowing of the LVOT (144). Based on the extent and severity of the hypertrophy, individuals with HCM may develop LVOTO or mitral regurgitation (MR) (4). Other frequent pathologic findings are elongation of the mitral valve leaflet(s) and abnormal insertion of papillary muscle (29). Severity of LV hypertrophy has an essential role in prognosis and SCD risk assessment (3).

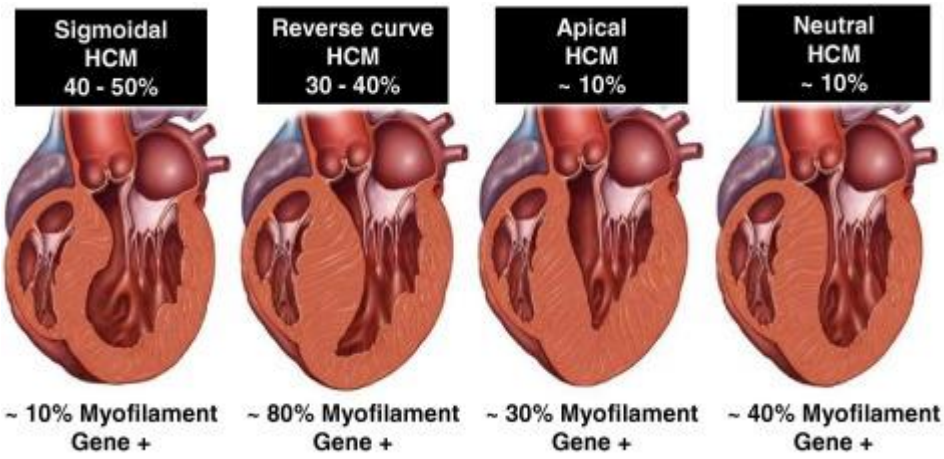

Figure 2. Most common septal morphologies in HCM (145)

In some patients, only particular segments of ventricular wall or septum are affected; in others, non-adjacent hypertrophied segments may be found. In others, hypertrophy might involve over half of the chamber (144).

1.5.4. Clinical presentations of HCM

The clinical presentations of HCM are strikingly variable (2,3,8,17,144), which sometimes makes HCM diagnosis challenging (3). Some patients are completely asymptomatic (3,5,6,93,144) and can be identified incidentally (3), while others manifest LVOTO (17,18,31,92,105,144), AF (3,11,17,92), SCD (3,5,8,92,144,146) or HF (3,6–8,34,92).

1.5.4.1. Symptoms

Classical symptoms in HCM are usually related to arrhythmias, diastolic dysfunction, LVOTO, MR, and myocardial ischemia (147). Most HCM cases remain asymptomatic or mildly symptomatic throughout their life (148), whereas others experience chest pain, fatigue, (exertional) dyspnea, palpitations, presyncope, and syncope (1,3,10,25,101,148–151), with dyspnea being the most common and syncope the least common (149). The frequency of dyspnea and palpitations increases with age (148). Chest pain may be associated with meals, exertion, or dehydration (152).

Dyspnea is the consequence of increased LV systolic pressures resulting from diastolic dysfunction, outflow tract obstruction, MR, and myocardial ischemia (3,153,154).

Chest pain in HCM is most commonly caused by microvascular coronary abnormalities and obstruction. In this regard, cardiac catheterization provides insights into the anatomy of the septal arteries and coronary artery bridging. Such insights may be useful in case of patients requiring ASA (147).

Pre-syncope and syncope develop in 25% of HCM cases (3,155). Despite comprehensive studies, the precise mechanisms in most of the cases are unknown (29,147). Potential mechanisms include LVOTO (3,29,147,155), inappropriate vasodilation, diastolic dysfunction-mediated hypotension (3,147,155), supraventricular arrhythmia, ventricular arrhythmia, sinus node dysfunction, complete heart block, ventricular arrhythmia, and volume depletion (3,147,155).

Unexplained, non-neurocardiogenic syncope, especially if more recent than 6 months ago, indicates an increased risk of SCD. Unexplained syncope is included in implantable cardioverter-defibrillator (ICD) decision-making (3).

Finally, HCM symptoms do not inevitably correlate with the degree of LVH or the severeness of LVOTO (36).

1.5.4.2. Left ventricular outflow tract obstruction

SAM is caused by elongation of the anterior and posterior mitral valve leaflets (independent of demographic variables, clinical variables, hypertrophy, LV wall thickness, and hemodynamic factors) along with anterior displacement of the papillary muscles (1,156). The anterior mitral valve leaflet moves into the LVOT, and in severe cases comes into contact with the hypertrophied basal septum during systole. This changes blood flow through the LVOT and causes hemodynamic instability. The extent of obstruction may differ. LVOTO is dynamic. It can exacerbate in cases of decreased preload (e.g. Valsalva), decreased afterload (vasodilators) and positive inotropes. Depending upon the obstruction degree, a reduction in cardiac output, exacerbation of MR, and diastolic dysfunction may appear inducing a rise in LV end-diastolic pressure that may influence coronary blood flow and cause arrhythmia (156).

LVOTO is found in 70% of HCM cases, either at rest or after physiological provocation (Valsalva maneuver) (1,157). LVOTO is a major therapeutic target in HCM, because it represents one of the major determinants of symptoms like dyspnea, chest pain, or syncope (157–159). LVOTO in HCM patients is associated with poor prognosis (1,160) and represents a risk factor for AF (29) and SCD (159).

Distortion of the mitral valve leaflets often causes secondary MR, a late systolic event that might represent a major cause of severe symptoms. The severity of MR is determined by the extent of outflow obstruction. The order of pathophysiologic events in HCM with obstruction and secondary MR has been defined as eject, obstruct, and leak (154).

1.5.4.3. Arrhythmias

Rhythm disorders in HCM usually consist of supraventricular and ventricular ectopic beats, and in rare cases, non-sustained or sustained ventricular tachycardia (VT). AF occurs in patients with severe LVOTO, abnormal size and function of LA, or

advanced clinical stage, and represents a major risk factor for thromboembolic stroke or HF.

Arrhythmias in HCM are results of mutation effects (e.g. altered calcium handling) or secondary processes, such as increased cardiomyocyte automaticity caused by hypertrophy, or re-entry caused by myocardial fibrosis. Arrhythmogenicity in HCM is also influenced by various pathophysiological factors: myocardial ischemia, altered hemodynamics, and maladaptive autonomic responses (15).

1.5.4.3.1. Atrial fibrillation

AF "manifests as regular and orderly loss of atrial electrical activity, which is replaced by rapid and disorderly fibrillation waves on the ECG" (125). It is the most frequent sustained arrhythmia in HCM and general population, with rates being 4-6-fold greater in HCM than in the general population (11,160–162). In HCM cases, the annual AF incidence is 2-3%, whereas lifetime prevalence is ~ 20-30%. AF is paroxysmal in 2/3rds of HCM patients and permanent or persistent in the remaining 1/3rd (160–162). It is more frequent in older individuals and in individuals with LVOTO (157).

AF is sometimes asymptomatic in patients with HCM (162,163), but can be also clinically presented as unspecific: dyspnea, hypotension, pulmonary edema, chest pain, palpitations, presyncope, syncope or stroke (160,163). Individuals with HCM are often more symptomatic of AF (161).

AF in HCM is caused by several factors: genetic component, anatomical, electrophysiological and hemodynamic abnormalities, abnormal calcium handling hypertrophy, as well as coronary microvascular dysfunction cause atrial ischaemia which are arrhythmogenic substrate for development of AF (29,157,160–162). However, pathophysiological and anatomical changes in HCM and AF are intertwined. AF is both the cause and effect of morphological and physiological changes in HCM. Thick left ventricular myocardium with reduced diastolic relaxation and filling raise LV pressure, causing elevated LA pressure. This process may induce atrial dilatation and remodeling, which ultimately leads to atrial enlargement. This in turn reduces the effective atrial refractory period, raises the dispersion of repolarization, and raises the capacity of ectopic triggers to preserve AF (29,161–163).

Atrial myofibril disarray and fibrosis are arrhythmogenic substrates for AF in HCM; they impair the conduction of sinus impulses and cause intra-atrial re-entry (157,160,161). Diastolic dysfunction in HCM can raise the risk of AF by increased atrial afterload, atrial myocyte stretching, and higher wall stress as a consequence of atrial dilation (162).

Regular atrial contraction is necessary for LV filling (160,163). In individuals with HCM, AF causes loss of organized atrial depolarization and contraction during diastole, and also reduces LV diastolic filling time. These factors, together with decreased LV compliance in a hypertrophic LV, may reduce cardiac output (160,161). The presence of atrial arrhythmia impairs myocardial ischemia, diastolic dysfunction, LVOTO, and MR (163). For these reasons, AF is usually poorly tolerated by individuals with HCM (29,160,161).

The most important independent predictors of AF in HCM are: LA diameter and volume (the strongest predictors), age, and New York Heart Association (NYHA) HF functional class (160,162). Varied cutoffs of LA size have been proposed for prediction of AF risk in HCM, however, the most accepted size is anteroposterior diameter > 45 mm (160). Age at diagnosis threshold, spaning from ≥ 40 years to > 50 years, is independently predictive of AF in HCM (160,162). NYHA class III/IV, left ventricular ejection fraction (EFLV) < 50%, and moderate to severe MR are associated with a higher risk of AF (160). A greater extent of septal hypertrophy is also associated with a higher risk of AF (160,161). Some authors also propose hypertension, vascular disease, obstructive sleep apnea, P-wave duration, ST-segment changes, premature ventricular contractions, LV size, LVOT gradient, LV fibrosis, and N-terminal pro-brain natriuretic peptide as clinical variables associated with the occurrence of AF in HCM (162).

The occurrence of AF is a critical moment in the progress of HCM (162) — the development of permanent or persistent AF in individuals with HCM frequently announces the beginning of a continual functional decline (161,162). AF is associated with high rates of thromboembolic complications, symptomatic HF, and mortality (29,157,160–162). AF is an independent predictor of overall mortality in HCM: there is up to 4-fold raise in the risk of death in case of AF, compared with sinus rhythm (160,161). Individuals who develop AF at a younger age carry poorer prognosis. Individuals with

AF have a higher rate of progression to NYHA Class III-IV HF symptoms, compared with non-AF HCM individuals (160,161).

The most important components of AF management involve rate and/or rhythm control, thromboembolic prophylaxis, strict risk factor and lifestyle control (160,161).

1.5.4.4. Sudden cardiac death

SCD is "an unexpected, non-traumatic death within 1 h of symptoms' onset in a patient known to have a potentially fatal cardiac condition, or when autopsy finds a cardiac or vascular anomaly as the probable cause, or when no extracardiac causes are found in the post-mortem examination, and therefore an arrhythmic event is the likely cause of death" (164).

SCD can be the first presentation of HCM (24,151,165). HCM is the main cause of SCD in patients < 35 years of age, but the precise incidence of SCD in this age is not clear (159). The risk of SCD in individuals with HCM is < 1% per year (24,36,159). In the U.S. National Registry of Sudden Death in Athletes, HCM was accountable for 36% of SCD cases, and, in a similar study in the U.K., HCM was diagnosed in 6% of the SCD cases (164).

However, a number of sudden deaths among athletes on sports-courts increased awareness about the association between physical activity and SCD in HCM (24,159). The overall risk of SCD for actively competing athletes with HCM is probably < 0.1% per year (159). The link between physical activity and SCD in HCM is still incompletely clear. Exercise involves a number of physiologic changes and some of them have been suggested as potential triggers for SCD: myocardial ischemia, sympathetic-vagal imbalance, hemodynamic changes, metabolic acidosis, electrolyte imbalance, and dehydration (159).

Myocardial fibrosis represents an arrhythmogenic substrate, and post-mortem SCD studies have reported the existence of myocardial scar in HCM patients (166). Individuals with HCM usually already present myocardial fibrosis and higher susceptibility towards myocardial ischemia, both of which may increase risk of arrhythmogenesis during exercise (159).

LVOTO provokes SCD, either through precipitation of ventricular arrhythmias caused by myocardial ischemia, or through increased left ventricular end-diastolic pressure (LVEDP) resulting in severe reduction in cardiac output and leading to electromechanical dissociation. Relief in LVOTO (e.g. through surgical myectomy) is associated with decreased rates of SCD (167).

Placement of an ICD is favored in individuals at a high risk for SCD (36,99). ICD implantation is an important preventive measure, since pharmacotherapy is generally not effective in preventing SCD in HCM (36,167). The 2011 American College of Cardiology Foundation (ACCF)/American Heart Association (AHA) guidelines recommend consideration of ICD implantation, established on the existence of ≥ 1 major risk factors (SCD family history especially at a young age, LV hypertrophy ≥ 30 mm, or unexplained syncope) or possible risk mediators (36,155,167,168). ACCF/AHA guidelines and European Society of Cardiology (ESC) guidelines also recommend ICD implantation. SCD events among first-degree relatives increase the risk of SCD (3,167).

Other potential indicators of an increased risk of SCD in HCM are: a history of ventricular fibrillation, sustained VT, or the presence of non-sustained VT; an abnormal reaction wherein blood pressure falls during physical activity; LVOTO; LV apical aneurysm; disease-causing gene mutations; younger age (especially under the age of 35 years) (3,21,24,36,155,167,169); and LA size (3,24,155,167,169).

Non-sustained ventricular tachycardia (NSVT) is a major risk factor for SCD, as its episodes might lead to ventricular fibrillation, which is the frequent cause of SCD (29). Finally, SCD is usually caused by cardiac arrhythmias, LVOTO, and ventricular diastolic dysfunction (36).

Periodic testing involving ambulatory cardiac monitoring, ECHO, stress testing, and CMRI enable identification of HCM patients who may be at high risk for SCD (151).

1.5.4.5. Heart failure

HF is a clinical condition wherein the heart is unable to meet cardiovascular requirements, producing symptoms like fatigue and dyspnea (153,170). Therapy for HF has developed significantly over the past decades, but mortality is still high (170).

Congestive heart failure (CHF) is a major public health issue, with 45 million people affected globally and an annual mortality rate of ~ 10% per year. Conventional CHF (heart pump failure) can be the consequence of systemic or pulmonary hypertension, ischemic heart disease, or valvular heart disease. CHF cause volume overload, reduced cardiac output/stroke volume, pulmonary venous congestion, lower extremity edema, and renal dysfunction (153).

HF in HCM and conventional CHF are very different conditions regarding pathophysiology and clinical presentations. HF in HCM represents a much more limited public health issue, and HCM mortality associated with HF is lower compared to that in CHF. This is a consequence of HCM characteristics, such as lower prevalence, a healthier and younger HCM patient population with less comorbidities, and treatments available capable of reverting the HF progression. Reversibility of HF is the most distinctive feature of HF in HCM — wherein mechanical impedance to LV outflow is relieved by surgical myectomy or ASA (153).

Several factors are involved in the pathophysiology of HF in HCM: diastolic dysfunction as a consequence of delayed LV relaxation, vascular remodeling, LVOTO, abnormal calcium homeostasis, decreased chamber compliance, myocardial ischemia, and abnormal vasomotor response (171).

In HF among HCM patients, peripheral and pulmonary edema and volume overload are not present. Most commonly, HCM patients demonstrate preserved systolic function (153).

ECHO plays an important role in predicting HCM progression to HF. Increased LV filling pressure is a negative prognostic factor in HCM patients. Right ventricular involvement (found in approximately 50% of HCM patients) increases the risk for developing HF symptoms — there is a direct correlation between RV wall thickness and HF symptoms (171).

Intraventricular obstruction raise the myocardial load and decreases the cardiac output and blood supply. The simultaneous existence of considerable MR also worsens the HF symptoms. LVOTO is a predictor of HF symptoms and HF progression (171).

1.6. Genotype-phenotype associations in HCM

HCM is a heterogeneous disease regarding both genetic mutations and clinical course (6,12,34,86,172). A high variety of involved mutations, the relative rareness of each of the individual mutations (private mutations are very common in families) (12,84), as well as incomplete penetrance (15,86,173,174) hamper the establishment of universal genotype-phenotype correlations. However, some trends have been identified (84).

The lack of clear genotype-phenotype associations restrict the use of genetic information in clinical management of HCM (85). Moreover, relatives with the same mutation often have different clinical presentations, progression and complications (some of which may remain completely phenotype negative) (6,15,80,84,168,175).

Childhood-onset of HCM is seen to occur more often, if there is a family history of early-onset HCM (175).

Similarly, just like the diversity in HCM genotypes, there is a diversity in the HCM mechanisms implicated in the pathogenesis of HCM. HCM mechanisms might be classified into groups; still, they are intertwined. The fundamental abnormality in HCM is the mutation. Proximal phenotypes are direct results of the mutations, specifically, the structure and function of the sarcomeric proteins. The secondary or intermediary phenotypes involve the molecular alterations, which appear as a response to proximal phenotypes (e.g. modified gene expression or activation of the signaling pathways, like the TGFB1 and MAPK). The tertiary effects involve resultant histological and pathological phenotypes, which are the outcome of intermediary (secondary) phenotypes. Tertiary effects eventually lead to the clinical phenotypes of HCM (29).

1.6.1. Genes

The presence of a mutation in some of genes encoding sarcomeric proteins is associated with earlier HCM onset (175,176), more severe LVH (175), and worse clinical outcomes (2,84).

More specifically, HCM patients with identified mutation in some of genes encoding sarcomeric proteins have worse microvascular dysfunction, compared to genotype-negative individuals (1). HCM patients having likely pathogenic or pathogenic mutations in any of the sarcomeric genes have a 2-fold increased risk of unfavorable

outcomes (such as death, HF, potentially fatal ventricular arrhythmias, etc.), compared to HCM cases with no mutation in sarcomeric genes. Individuals having variants of uncertain significance possess an intermediate risk. The lifetime risk of adverse outcomes in HCM is inversely associated with age at diagnosis (173).

1.6.1.1. *MYH7*

Mutations in the *MYH7* are usually associated with a poorer prognosis (e.g. they progress more often to end-stage HF), compared to mutations in *MYBPC3* or absent mutation (24,143,176) and an earlier disease onset (24,29).

Some critically important regions in the β myosin heavy chain are encoded by *MYH7*, such as the converter domain that have been linked with earlier onset and malignant arrhythmias (6).

HCM patients with mutation in *MYH7* present higher C-terminal propeptide of type I procollagen (PICP) levels (compared to carriers of mutation in *MYBPC3*), which suggests an increased myocardial collagen synthesis and profibrotic state (176).

The missense mutation Arg663His in the *MHY7* gene is associated with greater risk of AF (160,162).

1.6.1.2. *MYBPC3*

Mutations in *MYBPC3* are associated with a more moderate HCM phenotype (168,176), lower penetrance than mutations in *MYH7* (6), as well as with elderly onset (6,12,175). Because of its delayed onset, the reproductive age is not influenced and founder mutations (highly conserved within isolated populations) in *MYBPC3* are more frequent (6,175). Missense mutations in *MYBPC3* are predominant in children, whereas truncation mutations are more common in adults (6). No consensus has been identified regarding mutations in *MYBPC3* and disease severity, progression, and phenotype. The locations of truncating mutations in *MYBPC3* are not predictive of clinical outcomes either (84).

1.6.1.3. *TNNT2*

Mutations in *TNNT2* are associated with a moderate phenotype, yet pose a high risk of SCD (6,168,176). More recent research has shown that some mutations in *TNNT2* follow this rule, but there are also exceptions like p.Arg278 (because of its low frequency among the general population) (6). Hearts with mutations in *TNNT2* contain lesser fibrosis but more severe myocyte disarray. The increased rate of SCD in patients having mutation in *TNNT2* might be a consequence of myofilament Ca^{2+} sensitization or severe myocyte disarray (176).

1.6.1.4. Genetic negative HCM patients

A group of HCM patients, who have no mutation detected regardless of family members screening and exhaustive genetic testing, cannot help identifying an affected family member. This group has a distinct clinical course compared to sarcomere mutation-positive patients: they are diagnosed in older age, typically male, with more moderate LVH, and more probably already have a diagnosis of hypertension (88,174).

1.6.1.5. Gene dosage

Gene dosage also affects prognosis in HCM (6). Among HCM patients who exhibit 2 (digenic) or several (oligogenic) causal mutations in the same or different genes the severity of ventricular hypertrophy appears to be more pronounced. These patients also raise the assumption that the "absence" of causal genes can be elucidated by the digenic or oligogenic character of mutations in some HCM cases (29).

Homozygosity or compound heterozygosity is linked with early onset, severe clinical manifestations, and poor outcomes (higher incidence of SCD or HF events), but is uncommon (82,127,175). Patients with 2 or more sarcomeric HCM-causing mutations have an increased risk of lethal arrhythmias and poor outcomes (173).

1.6.1.6. Other

Thin-filament mutations are associated with an often atypical distribution of hypertrophy, prominent diastolic dysfunction, and a higher probability of restrictive

progression and SCD in the course of childhood, compared with more frequent thick-filament gene mutations (173).

An interesting aspect of sarcomeric mutations is pleiotropy — mutations in the very same gene could manifest as either HCM, DCM, RCM, or LVNC. Different mutations in genes like *MYH7* and *TNNT2* lead to varied phenotypes: HCM and DCM. The possible explanation might be the different locations of the mutations, which would affect various protein domains, causing different interreactions of mutant proteins with sarcomere components, and subsequent activation of various intermediary molecular events. Similarly, causal mutations might affect Ca^{2+} sensitivity of ATPase activity and force generation (HCM-causing mutations in thin filament proteins generally increase Ca^{2+} sensitivity of ATPase activity and myofibrillar force generation, whereas DCM-causing mutations decrease these) (29).

A tendency involving higher allele frequency of *MYBPC3* and *TNNI3* variants was identified, which could be a result of the reduced expressivity or penetrance of such variants (a milder disease progression compared with other sarcomeric disease gene variants). HCM cases with mutations in *MYH7* and *TNNT2* are generally younger at diagnosis (34 years), than HCM patients with mutations in *MYBPC3*, *TNNI3*, or those without mutations (176).

Early studies on large pedigrees with severe clinical presentations identified several "high-risk" mutations (e.g. *MYH7*-R403Q, *MYH7*-R453C25, and *TNNT2*-R92Q/W). However, studies on greater populations of unrelated individuals partially validated the devastating outcomes of those mutations, and were not able to find genotype-phenotype correlations for the most of variants (173).

1.6.2. Clinical courses

1.6.2.1. Patterns of left ventricular hypertrophy

Mutations in *MYH7* and *MYBPC3* are the most frequent in HCM involving basal septum (29). Mutations in *PRKAG2* and *LAMP2* usually lead to atypical distribution of hypertrophy, with an increased rate of arrhythmias and HF. Mutations in *TNNT2* and complex genotypes are associated with right atrial enlargement in HCM, leading to a poor prognosis. Mutations in *TNNI3* and *MYH7* are associated with restrictive phenotype

(6,175). Mutations in *ACTC1* cause apical hypertrophy and left ventricular non-compaction (6,175).

1.6.2.2. Ventricular arrhythmias and sudden cardiac death

HCM patients with mutations in *MYH7*, *TNNT2*, or *MYBPC3* present ventricular arrhythmias more frequently than individuals without sarcomeric mutations. Among these, the rate is highest in the *MYH7* group (176). The risk of SCD is small among patients carrying pathogenic/likely pathogenic variant without hypertrophy. Mutations in *TNNT2* might represent an exception (12).

1.6.3. Disease modifiers

HCM is a multifarious disease with diverse mutations, allelic imbalance, penetrance, heart contours, as well as disease modifiers, all being partly responsible for the definitive outcomes (83).

Since HCM phenotype involves the contribution and interplay of both genetic mutations and other factors (e.g. environmental factors and gene modifiers), definitive genotype-phenotype associations are still unknown (29,80,168,174). The etiology of HCM appears to be multifactorial. Sarcomeric dysfunction might be a mandatory though not necessarily starting point in HCM pathogenesis (80).

1.6.3.1. Molecular disease modifiers

Lack of clear genotype-phenotype associations in HCM underscores the importance of discovering supplementary elements that control the progression of HCM, and indicates that molecular mechanisms existing between genotype and clinical presentations may be crucial. Despite active research, molecular interactions in HCM are poorly understood (36,88,92,176).

Secondary molecular changes in pathways implicated in HCM pathogenesis, post-translational protein modifications, as well as epigenetic factors (e.g. microRNAs and small noncoding RNAs) affect histological and clinical HCM phenotypes (15,29). HCM phenotypes are directed by several factors, each of them undertaking a small effect (29).

1.6.3.1.1. Modifier genes

Genetic aspects beyond causal mutations affect the phenotypes of single gene disorders. This is especially the case with autosomal dominant diseases, characterized by age-dependent onset and variable expressivity. Mutations in genes encoding contractile sarcomeric proteins cause HCM; still, the contribution of the causal mutations to the definitive phenotype could be moderate, and other genes (together with the environment) could contribute significantly (177).

Genes involved in the pathophysiology of HCM are not all the same. HCM susceptibility genes (e.g. *MYBPC3*, *MYH7*, *TNNT2*, *TNNI3* etc.) are directly implicated in the pathophysiology of HCM, whereas modifier genes or variants are implicated in the adjustment of its phenotypic expression (177,178). Modifier genes are neither sufficient nor necessary to cause HCM. They constitute the genetic background of individuals, and the presence of DNA polymorphism makes genetic background relatively individual (29,177). The ultimate phenotype is the result of the causal mutations, environmental factors, and modifier genes (177).

Several histologic and morphologic HCM phenotypes are compensatory and regulated by numerous factors (177). Pathogenic variants of genes involved in regulation of cardiac hypertrophy and fibrosis may act as modifier genetic variants (29).

Gene modifier candidates for HCM are: angiotensin-1 converting enzyme-1 (*ACE*) (2,177,178), angiotensinogen (*AGT*) (2,177,178), angiotensin II receptor 1 (*AGTR1*) (2,177,178), chymase (*CMA1*) (177), bradykinin B2 receptor (*BDKRB2*) (177), aldosterone synthase (*CYP11B2*) (2,177,178), endothelin-1 (*EDN1*) (177), tumor necrosis factor α (*TNF*) (177), insulin-like growth factor 2 (*IGF2*) (177), transforming growth factor β1 (*TGFB1*) (177), interleukin 6 (*IL6*) (177), and platelet activating factor acetylhydrolase (*PLA2G7*) (177) (Table 4).

Table 4. Candidate gene modifiers for HCM, adapted from (177)

Symbol	Polymorphism	Reported results
ACE	I/D	DD is associated with higher risk of SCD DD is associated with severity of hypertrophy DD is more common in HCM patients Frequency of DD genotype unchanged No association with indices of hypertrophy
AGT	−6G/A	No association with indices of hypertrophy No association with indices of hypertrophy
AGT	M235T	235T allele more common in HCM
AGT	T174M	Frequency of T174M and M235T unchanged
AGTR1	1166A/C	No association with indices of hypertrophy C allele is associated with severity of hypertrophy
CMA1	1625A/G	No changes in frequency in HCM
BDKRB2	−412C/G T21M	T21M was found in HCM cases but not in controls
CYP11B2	−344T/C	No association with indices of hypertrophy
EDN1	8002G/A	A allele is associated with severity of hypertrophy
TNF	−308G/A	A allele is associated with severity of hypertrophy
IGF2	820G/A	No association with indices of hypertrophy
TGFB1	−509C/T	No association with indices of hypertrophy
IL6	−174G/C	No association with indices of hypertrophy
PLA2G7	994G/T (V279T)	T allele is more common in HCM and is associated with increased left ventricular dimension and decreased function

Several polymorphisms in genes encoding proteins involved in the renin-angiotensin-aldosterone system (RAAS), acting alone or together, might impact the phenotypic alterations found in HCM and are considered modifiers (178). This is because changes in "activation status" of the RAAS may lead to more prominent LVH and remodeling (36,178). RAAS contributes to ventricular hypertrophy through circulating angiotensin effects, and by local activation of RAAS in the myocardium. Polymorphisms in genes encoding proteins involved in the RAAS are associated with increased severity of LVOTO and progressive septal hypertrophy, while both of them are risk factors for adverse outcomes (178).

ACE DD genotype is linked with the degree of LVH. However, the role of RAAS polymorphism in HCM is still debatable (2). An insertion/deletion variant in the ACE, which is associated with variation in the plasma levels of ACE, modestly modifies cardiac hypertrophy and the risk of SCD in HCM (29).

1.6.3.1.2. Mitochondrial DNA variants

The heart demands a lot of energy. A single cardiomyocyte comprises hundreds of mitochondrias. Mitochondrial DNA mutations have been linked with childhood HCM. Furthermore, an association between several mitochondrial haplotypes and clinical progression of cardiomyopathies has been shown. The proposed mechanisms might involve polymorphisms in mitochondrial genes, causing decreased energy efficiency. For example, haplogroup H represents a susceptibility factor, whereas haplogroup J represents a protective factor for HCM. This is the result of haplogroup H having higher mitochondrial oxidative damage [this haplogroup has the highest maximal oxygen consumption (VO_2max)], in contrast with haplogroup J (the lowest VO_2max consumer) (2).

1.6.3.1.3. Epigenetics

Modifications of gene expression in the absence of alterations in the genetic code are referred to as epigenetics (179,180). Epigenetics refers to modifiers that have a role in switching genes “off” and “on” (179). These changes in gene expression are mediated by DNA methylation/demethylation, histone modification, nucleosome positioning, and non-coding RNA-mediated modifications (179,180). The effects of various environmental factors arise through alterations in the epigenome (88,180).

Along with genetic and environmental components, epigenetics are implicated in the pathophysiology of HCM (179). In individuals with HCM, phenotypic expression is shaped by epigenetic modifications. For instance, CpG islands methylation of the cardiac troponin T gene results in genetic instability, which leads to the deamination of this region, in turn leading to mutations that later predispose to HCM. CpG sites are spots where cytosine (C) is next to guanine (G), and CpG islands are domains with numerous CpG sites (179). There is evidence that end-stage cardiomyopathic hearts present differential methylation. Cardiac hypertrophy has also been associated with histone acetylation (2).

1.6.3.1.3.1. DNA methylation

Research suggests that an increased methylation of exonic CpG sites in the *MYBPC3* can lead to deamination of methylated CpGs, which consequently contributes

to mutation development. The frequentness of such mutations differ among particular genes. This mechanism of mutations can be accredited to highly methylated CpG sites deamination within genes. Furthermore, frequent HCM-causing mutation due to G-to-T transversion may be induced by the binding of carcinogens like acrolein and benzo(a)pyrene diol to methylated CpG sites (179).

1.6.3.1.3.2. Histone modification

Histones maintain chromatin in an active or silenced state by interreaction with DNA. Histone deacetylases 5 (HDAC5) inhibits histone deacetylases 2 (HDAC2) through deacetylation. Phosphorylation of HDAC5 occurs when the myocardium is stimulated by hypertrophic stress and activation of HDAC2 causes myocardial hypertrophy. Thus, histone deacetylases has a critical role in the cardiac hypertrophy development (179).

1.6.3.1.3.3. Micro RNAs (miRNA)

miRNA is a negative regulator, which operates through complementary mRNA silencing (179,180). Concentrations of miRNA are changing at various phases of HCM (2,179), so there is a potential for them to serve as severity markers in the course of HCM (179). In HCM, there are increased pro-fibrotic and pro-hypertrophic miRNAs and decreased miRNAs leading to opposite effects (Table 5). In cardiac tissue, miR-1 and miR-133 act as anti-hypertrophic factors, targeting several hypertrophic signaling molecules like transforming growth factor (TGF) β and fatty acid-binding protein (FABP). miR-29a is increased only in the plasma of patiens with obstructive HCM. This is in line with a correlation between the interventricular septum size and miR-29a (179).

Table 5. miRNAs in HCM [adapted from (2)]

miRNA	Effects
21, 23a, 24, 125, 129, 132, 195, 199, 208, 212, 222	Upregulated in hypertrophy
1, 133a, 29, 30b and 150	Downregulated in hypertrophy
27a, 29a, 199a-5p	Circulating miRNAs. Correlated with left ventricular mass
27a	Circulating miRNAs. Regulates HCM gene expression by targeting thyroid hormone receptor in cardiomyocyte
29a	Circulating miRNAs correlated with fibrosis

In this context, long non-coding RNAs (IncRNAs), which are also non-coding RNAs, play an important role. lncRNAs affect gene expression at the transcriptional and post-transcriptional levels. They are involved in development of HCM through the regulation of chromatin remodeling and interaction with the matching miRNAs. Levels of IncRNA myocardial infarction associated transcript (MIAT) are inversely correlated with miR-29a expression in HCM. Here, patients with no fibrosis exhibit increased lncRNA-MIAT and decreased miR-29a, compared to patients with fibrosis. The study has suggested that IncRNA-MIAT may serve as an endogenous miRNA buffer, which regulates the expression of miR-29a-3p (179).

Several miRNAs have been described as suitable biomarkers of HCM. There are some expectations that miRNAs may be useful as therapeutic agents for inherited cardiac diseases, as few miRNAs appear to be heart-tissue specific (2).

1.6.3.1.4. Signal pathways involved in cardiac hypertrophy

HCM mutations directly change the structure and function of the sarcomeric proteins and biophysical characteristics of the cardiomyocyte (15). The initial defects cause secondary (intermediary) molecular events such as activation of Ca^{2+}-sensitive, stress-responsive molecular pathways (29), and change cellular energy balance (15). These events produce the histological and morphological (tertiary) HCM phenotypes that are clinically presented as HCM (29). HCM mutations can trigger other signaling pathways: mitotic and trophic factors (e.g. calcineurin), mitogen-activated protein kinases, and TGF β pathways (15,29), and promote non-cardiac cells (e.g. fibroblasts). Genetic

variants in molecular pathways involved in cardiac hypertrophy and fibrosis can act as disease modifiers (15). The intermediate molecular events are activated in other cardiac hypertrophic responses as well (e.g. pressure overload-induced cardiac hypertrophy). Myocyte disarray, interstitial fibrosis, and cardiac hypertrophy are results of the intermediate molecules and pathways activation (29).

For instance, development of cardiac hypertrophy involves a complex set of pathways comprising dozens of receptors, ligands, transcriptional effectors, and cytoplasmic signal amplifiers. Some of the signal transduction molecules include MAPK, TGF, tyrosine kinase, fibroblast growth factor (FGF), insulin-like growth factor (IGF), protein kinase C (PKC), and c-Jun N-terminal kinase (JNK). These pathways additionally alter gene transcription, thereby causing alterations in protein synthesis, which then leads to cardiac hypertrophy. An example of extracellular signals transduction from the plasma membrane to the nucleus is the calcineurin/nuclear factor of the activated T-cell (NFAT) pathway. Cyclic guanosine monophosphate (cGMP)-dependent protein kinase-1 (PKG-1) is a significant downstream effector of cGMP-signaling in cardiomyocytes. A recent research has shown interaction between the phosphoinositide 3-kinase (PI3K) (Akt signaling pathway causing physiological cardiac hypertrophy) and the protein kinase C beta 2 (PKCβ2) (causing pathological cardiac hypertrophy) (179).

1.6.3.2. Non-molecular disease modifiers

Exercise (36,178), diet (36), alcohol consumption (178), microbial infection (88), cardiac loading conditions (36,178), environmental factors, and other diseases are non-molecular aspects that change the HCM phenotype. This proposes that the HCM phenotype is an outcome of the complex interplay between molecular and non-molecular factors (15,36).

1.6.3.2.1. Changes in loading conditions

Since cardiac hypertrophy is a result of the original defects derived by mutations, it is expected that changes in loading conditions (like systemic arterial hypertension) increase the severity of hypertrophy. Heavy physical activity is also expected to promote

cardiac hypertrophy in cases where mutations in genes encoding sarcomere proteins are already present (29,178).

1.6.3.2.2. Sex

Sex is suggested to be an important modifying factor in HCM (2), and could provide an explanation for some part of the clinical heterogeneity of HCM (174). Women have a tendency to develop HCM later in life, but when it occurs, affected women are more symptomatic (2,174). They are 6-13 years older at diagnosis than men (174).

A possible explanation for a delay in disease onset could be estrogen, which could have a role in cardiac hypertrophy inhibition through epigenetic modulations (174,176). On the other hand, the exposure of cardiomyocytes to androgens can produce hypertrophy. Studies have shown that males usually predominate over the life course up to 60 years of age when females become the more prevailing — pre-menopausal women might be more protected against developing HCM compared to men. However, another study has proposed that the high number of younger women (< 50 years) diagnosed with HF may indicate that menopause is unlikely to have a role in HCM (174).

Women are more likely to be sarcomere variant positive, while men are more likely to be sarcomere variant negative (174). Male patients predominate in the *MYBPC3*, *TNNT2*, and mutation negative groups. A well-balanced sex ratio in the *MYH7* group might be the result of higher penetrance, compared to the *MYBPC3* or sarcomeric mutation negative group (176).

Maximum LV wall thickness, LA diameter, and LV cavity size are often greater in men. However, the female heart is smaller than the male heart, and after correction for body surface area, they show relatively greater maximal LV wall thickness, LA diameter, and LV cavity size. Current HCM diagnostic criteria do not take this into account — women need to have relatively greater hypertrophy to come to at least 15 mm maximal LV wall thickness (current diagnostic criteria). Reduced heart size also explains the postponement in recognizing HCM, thus, more prominent symptoms are found in women. However, there is no difference between men and women in the severe hypertrophy prevalence (LV wall thickness $\geq$ 30 mm) (174).

Women have worse diastolic dysfunction, a higher prevalence of an obstructive phenotype, and more prominent HF symptoms (174). Regardless of similar extent of hypertrophy, women more often manifest dynamic LVOTO ($\geq$ 30 mmHg) than men. Women also have more MR, which is likely associated with the reduced cavity size. Women more frequently have HF symptoms, especially fatigue, exertional dyspnea, chest pain, and palpitations, and NYHA functional classes III to IV, compared to men (174).

Women more frequently undergo ASA and septal myectomy, which is probably related to the higher prevalence of obstruction and HF symptoms and to the increased age at diagnosis (174).

1.6.3.2.3. Exercise

Regular physical activity can improve symptoms of HCM patients, through improvement of LV diastolic dysfunction and better microcirculation via improvement of endothelial function (159).

The optimum exercise prescription in HCM remains unknown, but it appears that the benefits of regular physical activity can outweigh the risks of physical inactivity. Initiation of regular exercise training, in the early phases of HCM, can even produce genotype-positive relatives with a window of opportunity to prevent progress to the HCM phenotype, through favorable physiological adaptations (181).

However, HCM is the most frequent cause of SCD in young athletes. Hence, patients should avoid engaging in intense physical activities and participating in competitive sports (143). According to the current guidelines, the HCM diagnosis represents a cause for disqualification in competitive sports (2). Intensive physical activity is connected with an increase in LVH, and can serve as a trigger for malignant arrhythmias (2).

1.6.3.2.4. Comorbidities

1.6.3.2.4.1. Hypertension

The hypertensive heart is generally not easy to differentiate from HCM. Furthermore, the hypertension can have a negative effect on HCM progression by

promotion of additional LVH through diverse mechanisms, like neuroendocrine activation and increased afterload (143).

Hypertension enhances the extent of LVH via the induction of molecular pathways leading to cardiomyocyte hypertrophy — similar to intense physical exercise (2).

Numerous HCM patients develop hypertension during their life. Women usually have a prior diagnosis of hypertension more often than men (174).

1.6.3.2.4.2. Obstructive sleep apnea

Obstructive sleep apnea and HCM are a particularly frequent and destructive combination — obstructive sleep apnea is found in 32-71% of HCM patients (143,174). This wide range is probably the result of varied definitions of sleep-disordered breathing. HCM patients with obstructive sleep apnea are older, have greater limiting symptoms and exercise capacity, and more hypertension (174). Patients diagnosed with both HCM and obstructive sleep apnea show worse structural and functional heart impairment (143), increased prevalence of AF (143,174), and worse quality of life (143). Obstructive sleep apnea and HCM share some key pathophysiological mechanisms: myocardial hypertrophy, LA dilation, and overstimulation of the sympathetic nervous system (143).

1.6.3.2.4.3. Obesity

Obesity is linked with increased HCM penetrance, with a more severe phenotype, and more rapid and worse disease progression (140,174). Excess weight is linked with increased LV hypertrophy and mass (2,174). LVOTO is more frequent in overweight patients — it is observed in more than 50% HCM patients with body mass index > 30 (174).

Obesity affects the myocardium and amplifies the detrimental effects of sarcomere mutations via multiple mechanisms: LV remodeling, inflammation, perfusion defects, hemodynamic changes, neurohumoral activation, and metabolic perturbations (140). Obesity impacts the HCM phenotypic expression and progression by affecting cardiac function independently of mutation-induced effects. Further, it was also hypothesized that obesity intensifies mutation-induced pathogenic effects. In obese HCM patients, the heart appears to dilate exceedingly to increase stroke volume. That may propose that the

presence of a sarcomeric mutation decreases the myocardial capacity to handle increased obesity-related demands (140).

Obese patients with HCM more often present a significant LV outflow tract obstruction and are more symptomatic, as evaluated by NYHA functional class. They show decreased physical activity tolerance and capacity, compared to non-obese individuals. Obesity is linked with an increased LV mass index, LV cavity enlargement, greater posterior wall thickness and larger LA diameter (140).

There is a high prevalence of obesity in HCM patients. It is shown that while HCM patients are advised to regularly carry out non-strenuous physical activity, most of them do not follow these instructions due to fear of SCD, mental discomfort, or misinterpreted medical advice. This non-compliance contributes to the body weight increase, and the sedentary lifestyle further negatively impacts HCM progression (140).

Increased body weight predisposes individuals to develop HCM: high BMI in young adulthood is a predictor of HCM development later in life; each 1-unit increase in BMI increases in the risk of being diagnosed with HCM for 9% (140).

Obesity is closely related to diabetes. The clinical course is worse in diabetic patients with HCM. Compared to non-diabetic patients, diabetic HCM patients more often display LA enlargement, diastolic dysfunction, and MR. HCM patients with type 2 diabetes mellitus additionally show lower exercise capacity and worse NYHA functional class symptoms (140).

In obese individuals, cardiac workload and cardiac output are increased due to increased circulating blood volumes (and increased afterload, if hypertension coincides). HCM cardiomyocytes containing mutant protein compulsorily increase mitochondrial workload as a consequence of high ATP-usage by inefficient sarcomeres. Moreover, missense mutations in HCM impair length-dependent activation, which probably limits the cardiac contractile reserve when augmented preload appears. Persisting obesity-promoted preload elevation can consequently decrease the threshold for compensatory hypertrophy in HCM. There is an association between the quantity of truncal fat and septal thickness. The amount of epicardial fat is associated with the N-terminal prohormone of brain natriuretic peptide levels (140). Obesity represents an independent risk factor for HF in HCM cases (182).

Obesity may lead to cardiac adiposity, which in turn can induce endothelial dysfunction and local myocardial inflammation. Intramyocardial accumulation of fat is a source of cytokines and proinflammatory adipokines, which further contribute to vasodilation impairment, cardiac remodeling and stiffening. Cardiac adiposity is associated with lipotoxicity, which is shown to be detrimental to cardiomyocyte homeostasis. The epicardial fat volume is associated with the extent of cardiac hypertrophy, level of circulating biomarkers indicating myocyte injury, and diastolic dysfunction severity. Clinical manifestations observed in obese HCM patients may therefore be explained by myocardial adiposity (140).

Another factor related to obesity is hypertrophic stimuli driving LV remodeling, which affect molecular mechanisms. This prevents the incorporation and accumulation of mutant proteins (140).

In obese and diabetic patients, sympathetic nervous system overactivity is often seen. A high adrenergic drive occurs in symptomatic HCM with LVOTO. Chronic overstimulation of β-adrenergic receptor leads to its downregulation and pathway desensitization. Likewise, adrenergic receptor stimulation can append to the already high adrenergic drive in HCM, causing β-adrenergic signaling pathways impairment, and additional impairment of cardiomyocyte function. β-adrenergic overstimulation also induces oxidative stress, and thereby disrupt cardiomyocyte homeostasis (140).

Hyperlipidemia and hyperinsulinemia in obese and diabetic individuals cause a higher transfer of fatty acids to the myocardium. As time goes by, heart metabolism no longer has substrate flexibility and relies more on fatty acid oxidation. Glucose oxidation is more efficient than ATP production through fatty acid oxidation, regarding the amount of ATP molecules generated per oxygen atom used. In the HCM cardiomyocyte, this lack of balance in energy production can deteriorate the already existing mutation-related perturbations of the mitochondrial regeneration of ATP (140).

Also, in obese individuals, despite the increase in fatty acid oxidation, the uptake of fatty acids surpasses capacity of fatty acid oxidation, and leads to the intracellular lipid deposition, some of which may be transformed into toxic lipid species causing lipotoxicity. Lipotoxicity is connected with oxidative stress, mitochondrial dysfunction, endoplasmic reticulum stress, inflammation, and apoptosis. Lipid excess increases the

amount of acetyl-CoA precursors, which has been detected in failing hearts. Increased acetyl-CoA negatively regulates autophagy. Inhibited autophagy reduces mutant protein clearance, and therefore increasing the mutant protein dose (140).

Hyperglycemia promotes the generation and myocardial accumulation of glycation end-products, which induces diastolic dysfunction and inflammation (140).

1.6.3.2.4.4. Phenocopies

At first, term phenocopy was defined as phenotype variations induced by environmental, non-hereditary conditions, similar to or identic with the genotype-determined phenotype of some other individual. Phenocopy is now usually "defined as a phenotypic trait or disease that resembles the trait expressed by a particular genotype, but in an individual who is not a carrier of that genotype" (183).

Some HCM phenocopies, which can mimic HCM, are recognized and caused by mutations in *PRKAG2* (Wolff-Parkinson-White syndrome), *RAF1* and *PTPN11* (LEOPARD syndrome and Noonan syndrome), *TTR* (amyloidosis), *LAMP2* (Danon disease), *GLA* (Anderson-Fabry disease), *FXN* (Friederich's ataxia), and *GAA* (Pompe's disease) (6,168).

Fabry disease is particularly common (6) — approximately 1-3% of cases with clinically suspected HCM are deemed to have Fabry disease (168).

In amyloidosis, amyloid fibrils are accumulated extracellularly. Recently, advances in testing have discovered that transthyretin amyloidosis is more common than postulated in the past, and that some of the patients were formerly misdiagnosed as having HCM (168).

1.6.3.2.5. Environmental factors

A standard strategy towards assessing the genetic and environmental factors contribution to the varying progression of human diseases is to compare phenotypes in monozygotic (MZ) twins, who carry identical genetic sequences. It has been shown that a pair of twins carrying a pathogenic mutation in sarcomeric gene (*MYH7* G768R) had comparable LV wall thickness (LVWT) but varied fibrosis amounts. Interventricular septum thickness (IVSd) in 11 MZ twin pairs, demonstrated no significant heritability for

IVSd in HCM. On the contrary, another study showed consistent morphologic features and clinical development in identical twins diagnosed with HCM, proposing only a small influence of environment on HCM clinical presentations (88).

In a study of 11 MZ HCM twin pairs, followed up for over 5-14 years, LVWT discordance is demonstrated in all twin pairs. This suggests that environmental and epigenetic factors have a key role in hypertrophic remodeling. In the same study, LA sizes were consistent in 3 of 9 MZ twin pairs carrying HCM-causing mutations. An explanation of this result could be that LA dimension demonstrates defective ventricular relaxation, caused directly by sarcomeric dysfunction (88).

Another identical adult twin pair with HCM was reported as being extraordinarily similar, considering their clinical course and morphological features. However, there were no differing environmental influences between the twins — they reported no notable differences in diet, exercise, or lifestyle habits and showed no evidence of exposure to radiation or chemicals (184).

1.7. HCM in contemporary cardiology

The genetic basis of HCM has been traced back to 5-10 centuries ago. For this reason, HCM is not a "new" disease, however, its clinical identification is new. Contemporary HCM is to some extent a result of its often benign nature and reproductive fitness (185).

The prevalence of reporting HCM has increased with the usage of advanced genetic testing and precise imaging techniques. Genetic testing allows a molecular diagnosis before ventricular hypertrophy appears (15).

HCM should be diagnosed as the disease secondary to mutations in the sarcomere or associated structural cardiomyocyte proteins, but only after rigorous elimination of the secondary causes of LVH and of all systemic syndromes associated with LVH and mimicking HCM. Some of the secondary causes of LVH are arterial hypertension, anatomical heart disease, or medication toxicity. Other metabolic, endocrinological, neurological, neuromuscular, or autoimmune multisystem disorders and genetic syndromes can manifest the characteristic HCM features. They must be taken into account

(15), because the management of these disorders and syndromes is different from that in HCM (15,45).

Nowadays, numerous HCM patients are still not aware of their diagnosis due to lack of symptoms and discreet clinical presentation, lack of clinical experience with HCM, or subadequate diagnostics (185).

HCM identification and treatment have advanced notably over the previous decade — it has been converted from an odd disease with an unfavorable prognosis to a treatable disease with low mortality (< 1% per year) (11,185). With contemporary evidence-guided approach, life expectancy is mostly relatively high, and a greatly improved quality of life is seen. 2/3rds of HCM patients have a normal life span (15). However, these advances are fully available only in highly developed countries and regions (185).

HCM occurs worldwide, among different races and cultures (11,185). It is estimated that probably 20 million people are affected, but many of them are undiagnosed or undertreated. Identified HCM patients form only the "tip of the HCM patients iceberg" (11,185).

1.7.1. Clinical stages of HCM

The pathogenic HCM mutation starts up a lifelong myocardial remodeling. At the beginning, the "non-hypertrophic HCM" phase arises — the absence of LVH in patients with HCM-causing mutations. The prevalence of HCM in this phase has elevated since genetic testing and family screening has begun, and ~ 0.6% patients at this phase are "genotype-positive/phenotype-negative", and should be described as having "pre-clinical HCM", or "HCM without clinical signs" (15).

The first stage of HCM is characterized by the occurrence of LVH with or without LVOTO, hyperdynamic ventricular function, and mild symptoms, such as intermittent chest pain or decreased exercise tolerance. In ~ 50% of the HCM patients, the onset of this stage occurs by the 3rd decade, and in ~ 3/4ths of the HCM patients, by the 6th decade. Subtle abnormalities can be found upon performing ECG or ECHO during this phase: impaired LV relaxation, mild LA dilatation, or mitral valve abnormalities. Altered coronary microvascular function and higher amounts of type I collagen precursors have been reported in this stage. The risk of SCD is in general small, however, malignant

arrhythmias may arise, particularly in HCM patients carrying a mutation in the cardiac troponin T gene (15).

During the second stage of HCM, myocardial remodeling, expanding myocardial hypertrophy, and fibrosis with retained clinical state occur. The prevalence of this stage is about 15%. The characteristics of this phase are: a low to normal EFLV, atrial dilatation, moderate to severe diastolic dysfunction, thinning of LV walls, LV apical aneurysms, microvascular dysfunction, and onset of AF. Congestive symptoms begin to be detectable through higher levels of natriuretic peptides and by cardiopulmonary exercise testing impairment (CPET) (15).

The third stage represents the irreversible "end-stage" of HCM. In this phase, extreme LV fibrosis, progression to LV and atrial dilatation, systolic and diastolic dysfunction associated with hemodynamic decompensation, HF-related complications, or death are found. This stage has high morbidity and mortality. Less than 5% of HCM patients reach this end-stage phase (15).

In most of the patients, the disease is identified between 20-50 years of age. However, HCM may be also diagnosed during childhood, while some HCM patients remain asymptomatic and un-diagnosed until late in life (15).

Once the patient is suspected as having HCM, or once a relative has been diagnosed with HCM, rigorous workup is necessary including: genetic and laboratory testing, comprehensive medical and family history, physical examination, ECG, 24-hour ECG, CPET, transthoracic ECHO, and CMRI. The aim of this workup is exclusion of conditions that mimic HCM, and to determine the HCM stage. HCM is often diagnosed in asymptomatic patients by a relative being diagnosed with HCM or by the incidental detection of a heart murmur (15).

The most common initial tests for patients suspected of having HCM are ECG and ECHO. CMRI is used to supplement ECHO, if the images are inconclusive. When hypertrophy is found to be borderline on ECHO, CMRI can be useful for better visualization of all of the cardiac segments and identification of characteristic patterns of fibrosis. For those diagnosed with HCM, ambulatory ECG monitoring lasting at least 24 hours is recommended. Exercise testing can be used for assessment of the functional capacity and detection of latent outflow obstruction. Genetic testing might clarify an

underlying diagnosis — to exclude HCM phenocopies when suspicion has been based on phenotypic evaluation (45).

1.7.2. Genetic testing

Genetic testing has been recently introduced into the practice of cardiology (15,148). It is valuable in clarifying both HCM and phenocopy conditions diagnosis (15).

There are several benefits of pathogenic mutation identification: it facilitates cascade family screening and identifies first-degree family members at risk of HCM development. After detection of pathogenic mutation in the proband, predictive genetic testing should be offered to first-degree family members. Mutation positive family members should undergo a regular ECHO and ECG. Mutation negative family members can be discharged from follow-up. However, long-term follow-up of family members is required since late-onset cardiac expression is common in HCM. Physical activity is generally allowed, but must be individualized according to the character of sport activity, family history, genetic status, and full cardiac examination (86,148).

The additional motive for genetic testing comprises phenocopy conditions, since their management could be different: regarding different modes of inheritance (Fabry and Danon disease are X-linked), distinct progression (accelerated systolic dysfunction in Danon disease), distinct complications (conduction defects in PRKAG2 disease), and specific therapeutics (enzyme replacement treatment in Fabry or Pompe disease) (86,148).

In some situations, genetic testing can be considered for reproductive issues, like prenatal testing or pre-implantation diagnosis (148).

Genetic testing is used in individuals with HCM symptoms and signs to confirm HCM diagnosis, in patients who already have HCM diagnosis to guide genetic cascade screening in family members, as well as in first-degree family members estimated to be at risk of developing HCM after pre-test counseling (15).

Positive genetic testing supports the HCM diagnosis in a patient, but negative genetic testing cannot exclude it (15).

1.7.3. General laboratory and metabolic testing

Systematic laboratory and specific metabolic testing (like enzyme-assays to identify exact lysosomal or glycogen storage disorders) should be performed in suspected HCM cases (15).

Collagen synthesis and degradation cleavage products, increased levels of cytokines, cardiac troponin T, and other myocardial inflammation markers have been reported as interstitial fibrosis biomarkers. They may be used in HCM patient assessment. Several microRNAs (particularly miR29a) can be used as biomarkers for interstitial fibrosis and cardiac hypertrophy (15).

1.7.4. Electrocardiogram

An ECG is the most sensitive routine diagnostic test in the setting of HCM — only 5-10% of HCM patients have normal ECG at the time of presentation. Reported ECG abnormalities involve P wave abnormalities, as well as prominent septal Q waves (148), complete bundle branch block, evidence of LA enlargement (15), repolarization abnormalities, and left axis deviation (148). Sometimes, gigantic deeply-inverted T waves in the precordial leads may suggest apical or midventricular HCM (148). Delta-wave or extremely high-voltage QRS complexes should increase the likelihood of a phenocopy. Typical ECG patterns in athletes include sinus bradycardia or arrhythmia, first-degree atrioventricular (AV) block, incomplete right bundle branch block, and J-point elevation with ascending ST-segments (15).

Ambulatory (24-48 hours) ECG monitoring is used in the assessment of palpitation symptoms, when it can detect either a supraventricular or a ventricular arrhythmia including paroxysmal AF, or as a part of the risk assessment, when it can capture ventricular arrhythmias (148).

Although sensitive, the ECG is not specific; thus, a normal ECG must undergo more detailed diagnostic evaluation, usually imaging (148).

1.7.5. Cardiopulmonary exercise testing

Exercise testing is important in the clinical assessment of HCM patients. There are 4 important HCM aspects assessed by exercise testing. The first is the evaluation of stress-

induced LV obstruction (especially important in patients with symptomatic HCM). The second is the blood pressure response to exercise (abnormal one may indicate an increased risk of SCD). Third is the assessment for accompanying coronary artery disease and provocable ventricular arrhythmias (ventricular arrhythmias are rare in symptom-limited exercise testing, but when detected, they indicate an increased risk of SCD). The fourth important aspect of exercise testing is that, in asymptomatic or mildly symptomatic individuals, it enables risk stratification: low event rate is suggested in patients achieving >100% predicted metabolic equivalents (148,169). Peak oxygen consumption during CPET may assist in differentiate between HCM and the athlete's heart, being low in HCM and high in athletes (15).

It is worth mentioning that exercise testing is safe for HCM patients (148).

1.7.6. Echocardiography

ECHO is the primary imaging modality for establishing HCM diagnosis, for monitoring of the disease, as well as for basic hemodynamic status assessment (15,45,143,186).

In suspected HCM cases, the thickness of all LV segments should be measured. To prevent oblique orientation leading to overestimation of wall thickness, measurements need to be preferably taken from parasternal short axis views with orthogonal beam alignment. Only in case of apical hypertrophy are apical views important for the apical segments hypertrophy assessment. Otherwise, measurements from apical views should be avoided, to prevent possible overestimation of LV walls thickness (186).

As previously stated, pathological hypertrophy in adults refers to the presence of unexplained wall thickness ≥ 15 mm, in one or more LV myocardial segments. In patients with already established HCM diagnosis, lower cutoff values may be used in screening of first-degree family members, with an absolute wall thickness ≥ 13 mm (15,30,86,148). Two-dimensional (2D) ECHO is the first step in assessment of hypertrophy in HCM. Observation of chamber characteristics can be facilitated by the usage of intravenous contrast agents (148).

In adolescent and adult individuals active in competitive sports, physiological cardiac hypertrophy (also known as an athlete's heart) is presented with LV wall thickness

of 13-15 mm. In these individuals, there is a symmetrical enlargement of all heart chambers, and the LV cavity size is increased. This contrasts with HCM, wherein LV cavity is not enlarged. The asymmetrical septal hypertrophy distribution pattern greatly supports HCM. The morphological and functional features in athletes are considered reversible and benign (15).

SAM of the anterior mitral valve might present as well as provoke and aggravate LVOTO. This generates mechanical impedance in blood flow as it leaves the heart, which generates a pressure gradient between the LV cavity and the aorta. Some structural characteristics may contribute to the LVOT gradients development: excessive thickening of the interventricular septum (causes LVOT narrowing), elongated anterior or both leaflets of the mitral valve, and apical insertion of papillary muscles (15,143,187). Elongated mitral leaflets can protrude into the LVOT, while abnormalities of the subvalvular apparatus may contribute to obstruction by positioning the anterior leaflet anteriorly into the LVOT. Anterolateral papillary muscle displacement "pulls" the plane of leaflet coaptation anteriorly or up into the LVOT, predisposing the patient to SAM and LVOTO. A papillary muscle or abnormal, fibrotic secondary chord, which inserts itself directly onto the anterior leaflet, also tends to pull the anterior leaflet into the LVOT (187). LVOTO is defined as a "Doppler-derived peak instantaneous LVOT pressure gradient ≥ 30 mmHg, either at rest or during Valsalva maneuver or exercise" (186).

SAM can be identified with 2D and M-ECHO (143). SAM of the mitral apparatus produces a posteriorly directed mitral regurgitant jet. If a central or anteriorly directed jet is observed, the mitral apparatus abnormality is probably a co-contributor to the MR. Since resolution of the LV obstruction will not resolve MR in case of primary mitral valve disease, this will affect the choice of treatment (148).

MR is usually functional in HCM, and is a result of SAM of the anterior leaflet. In most cases, regurgitant jet as a consequence of SAM-related MR is directed posteriorly and is rarely severe. Difference in jet direction or a severe regurgitation should raise the suspicion that another contributing mechanism exists. In such cases, a transesophageal ECHO should be used in addition to transthoracic ECHO to confirm or refute the aforementioned possibility. In this case, the focus shoud be on examination intrinsic

valvular abnormalities: leaflet restriction secondary to “percussion injury”, leaflet fibrosis, a chordal rupture, or a myxomatous disease (143,186).

Management of obstructive and of non-obstructive forms of HCM differ; hence, it is important to clinically differentiate the two. Resting LVOTO is defined as peak gradient > 30 mmHg, and represents a predictor of HF and death (143).

Approximately 1/3rd of HCM patients have the non-obstructive HCM; another 1/3rd have physiologically provocable gradients (< 30 mmHg at rest and > 30 mmHg with physiologic provocation); and yet another 1/3rd have peak gradients > 30 mmHg at rest (15,143,172), usually due to SAM (143).

Gradients > 50 mmHg at rest or with provocation propose the need for percutaneous or surgical intervention, if symptoms cannot be controlled with medicaments. Exercise (stress) ECHO is the preferred approach for provoking the LVOT gradient, because it simulates well the conditions experienced by HCM patients in everyday life (143). The Valsalva maneuver (exhale with glottis closed) or the administration of nitrates are utilized for provoking LV obstruction in HCM (148,186).

Additional echocardiographic characteristics can be observed in LVOTO such as: fibrotic mitral valve changes, early systolic closure of the aortic valve, mid-systolic notching, and coarse fluttering of the aortic valve during systole (143).

Apart from classical HCM (with or without LVOTO), other anatomical varieties of HCM also exist: midcavity obstruction, midcavity obstruction with apical aneurism, apical HCM, or right ventricular obstruction. Some of these varieties may be managed differently (143).

Midcavity or midventricular obstruction with or without apical aneurism is a result of varied anatomic and hemodynamic elements. In HCM, midcavity obstruction is a result of the apposition of the lateral wall and septum or of the systolic apposition of the lateral wall and hypertrophied papillary muscle at the mid-LV level. In these cases, 2 distinct LV chambers are generated (proximal and distal), and they set up an “hour-glass” shaped LV (143).

Absence of SAM of the mitral valve and a narrowed mid-LV cavity, with or without a thin-walled, scarred LV apical aneurysm, sometimes containing thrombi, can be observed using ECHO in isolated midcavity HCM (143).

In apical HCM, the myocardial hypertrophy primarily affects the apex of the LV. This type can be overlooked during echocardiographic assessment. CMRI is most accurate in diagnosing apical HCM (143).

In rare cases of HCM patients, right ventricular outflow tract obstruction can be found (143). RV hypertrophy is rarely involved in HCM. Hence, the involvement of the RV or extreme ventricular hypertrophy increases the likelihood of a phenocopy mimicking HCM (15).

LA is an important parameter to be measured in HCM, since it has been shown that enlarged LA is associated with adverse outcomes. Enlargement > 48 mm (transverse linear dimensions) or volume > 118 mL has been correlated with an increased risk for AF, HF, and death (143).

Regarding left ventricular ejection function (LVEF), most of HCM patients have normal or supernormal LV systolic function. In HCM patients with sarcomeric mutations, myocardial mechanics abnormalities occur prior the clinical hypertrophy (143).

Increased stiffness and reduced chamber compliance are associated with decreased LV volume on account of excessive hypertrophy in HCM (143).

Tissue Doppler imaging (TDI) is a standard procedure for evaluation of HCM patients. Systolic myocardial velocity is reduced in HCM, even in non-hypertrophied sections. For sarcomeric mutation positive individuals with absence of LVH, a systolic TDI velocity in the lateral annulus < 13 cm/s is predictive for manifesting HCM (143).

Diastolic parameters, like the ratio of E to E', peak E wave-to-flow propagation velocity ratio, and color M mode flow propagation velocity, predict end-diastolic pressure of LV, exercise tolerance, and reduction in filling pressures after septal ablation or myomectomy (143).

The maximum instantaneous LV gradient is estimated by continuous wave Doppler. It usually takes place in late systole and has a specific late-peaking, dagger-shaped profile (148). Doppler profiles and timing intervals can help in the diastolic function assessment as well (148). Doppler ECHO of the LV inflow tract enables the diastolic function characterization in HCM (143).

1.7.7. Cardiac magnetic resonance imaging

Some morphological abnormalities (such as focal myocardial hypertrophy or left ventricular aneurysms) can be best detected by CMRI (15). Myocardial tissue abnormalities may be assessed most effectively using CMRI. Interstitial fibrosis is found in ~ 2/3rds of HCM patients (148). Focal fibrotic changes and interstitial fibrosis can also be detected using CMRI. Further, contrast CMRI can help in the identification of patients transitioning from clinical stage I to stage II of adverse remodeling. Moreover, CMRI is usually required to exclude phenocopies (15).

CMRI is superior to ECHO with contrast for LV apical assessment, particularly in the assessment of apical pouches across the HCM spectrum (148). CMRI provides information about chamber volumes and EFLV, similar to ECHO. Myocardial tagging utilising CMRI can also help in the evaluation of regional function (148).

1.7.8. Positron emission tomography (PET) and computed tomography (CT)

In HCM, myocardial ischemia can be most reliably and non-invasively assessed using PET. Myocardial ischemia in the absence of epicardial coronary artery disease is often found in HCM. Perfusion abnormalities in HCM have been linked with adverse outcomes, and obstruction relief results in normalization or improvement of perfusion. Perfusion abnormalities following pharmacological vasodilation in the absence of epicardial coronary artery disease can be observed using PET as well, and suggest microcirculatory abnormalities and adverse prognosis. Direct imaging of the coronary arteries is traditionally performed using invasive angiography; still, cardiac CT represents a new, more sensitive approach. Cardiac CT can detect abnormalities in the coronary arterial course (e.g. myocardial bridging) (148).

1.7.9. Pharmacological treatment

A part of contemporary HCM treatment is still directed toward symptomatic relief, and does not treat the cause of HCM. Such medications are beta blockers (e.g., metoprolol, propranolol), calcium channel blockers (e.g., verapamil, diltiazem), anti-arrhythmic drugs (e.g., amiodarone, disopyramide), and anticoagulants for thrombus formation prevention (188).

New disease-specific and cause-directed treatments have been developed, including aficamten (65,188), danicamtiv, mavacamten, and omecamtiv (65). EXPLORER-HCM trial showed that treatment with mavacamten improves LVOTO, NYHA class, exercise capacity, and health status of patients with obstructive HCM (66). In randomized trials, aficamten has been shown to reduce hypercontractility and LVOTO. Further, danicamtiv is seen to improve left atrial and left ventricular echocardiographic parameters. Also, omecamtiv mecarbil is found to be efficient in HF with decreased systolic function, thereby reducing HF-related outcomes or cardiovascular death (189).

1.7.10. Septal reduction therapy

Although pharmacological therapy is effective for managing symptoms in majority of HCM patients with a resting or provocable LV outflow gradient, some patients require septal mass reduction therapies (190).

Septal reduction therapy is indicated by both the European and American guidelines in cases of LVOTO gradient ≥ 50 mmHg (with or without provocative maneuver) and when moderate to severe symptoms (NYHA III or IV or recurrent syncope) persist despite optimized (maximum-tolerated) pharmacological therapy (NYHA class III-IV) (11,186,187,191).

Effective management of LVOTO consists of invasive septal reduction procedures (192). Septal reduction therapy consists of two chief modalities: surgical treatment (septal myectomy) and percutaneous treatment (ASA) (19,99,191,193,194). Both septal myectomy and ASA aim at the septal mass reduction, regression of obstruction, and symptoms improvement (186).

The outcomes of both treatments mentioned above are highly dependent on the administering center's experience. The best results in septal reduction therapies are attained at centers with a high volume of ASAs and septal myectomies (11,191,193). Further, treatment needs to be personalized, with respect to the functional conditions, the anatomical structure, the patient's wishes, and advantages and disadvantages of each technique (192). A team, including a cardiologist-expert in HCM, an interventional cardiologist, and a cardiac surgeon specialized in ASA and septal myectomy, should evaluate each patient during the decision-making process for the treatment (191,193).

Unlike other cardiac procedures, no randomized trials of septal myectomy and ASA have been performed (186,195).

1.7.10.1. Surgical approach

Surgical myectomy represents the elimination of a small quantity (5-10 g) of hypertrophied tissue from the interventricular septum (193). Although HCM was first described in 1958 by Teare, the first surgical HCM treatment had already been accomplished in 1961 (191).

Surgical HCM treatment is based on septal myectomy, which aims to reduce the LVOTO gradient (191) that is a cause of progressive limiting symptoms (192). It involves the removal of a relatively small amount of muscle from the ventricular septum (196). In 10-11% of HCM patients, realignment of papillary muscles and valve plication might be necessary as well (191).

Morrow's modified septal myectomy is the preferred technique. It is usually used in young patients or when other pathologies requiring accompanying treatment (e.g. coronary disease or mitral valve anomalies) are present (191).

In the obstructive HCM refractory to medical therapy, septal myectomy has been considered the gold standard septal reduction technique and provides the most effective symptomatic relief (15,187,192,193,196,197).

The surgical approach restores life quality (15,187,192) and exercise capacity (15), with LVOTO reduction in 90% of the cases and low mortality (septal myectomy mortality rates are 1-2%). It is one of the safest open-heart procedures, with a 30-day mortality of 0.4% at high-volume centers (11,15,187). Long-term survival of surgically-treated patients is comparable to that of the general population and risk of SCD is decreased (191–193). Literature supports superior long-term symptomatic relief in HCM patients who undergo surgical septal myectomy (15).

ACC and AHA guidelines propose surgical treatment as the first option, while it is stated that ASAs should be reserved for patients for whom surgery is contraindicated or the risk of surgical treatment is unacceptable (191,195). For younger patients, septal myectomy can be an appropriate alternative, given the low re-operation rate (156,186,187,193).

Further, surgical myectomy is recommended when there are other concomitant surgical indications like mitral valve replacement, coronary bypass, or realignment/partial excision of ectopic or hypertrophic papillary muscles (15,144,186,193). Heart transplantation is reserved for the most refractory cases (188).

However, ventricular septal defects are possible complication of the surgical approach, usually exhibited in patients with septum thickness < 20 mm or several concurrent surgical procedures (e.g. myectomy and coronary bypass grafting). Another possible complication is aortic valve regurgitation, and it can occur because of lack of septal support or dissection of septal coronary arteries. New regurgitation jets after surgical procedure must be cautiously assessed to rule-in or rule-out these complications (191).

1.7.10.2. Percutaneous approach

Three decades after the first septal myectomy for HCM, Sigwart executed the first non-surgical septal reduction ever, utilizing surgical alcohol (195). ASA is a catheter-based technique (15,144,191,195). It consists of the infusion of a small quantity of high-grade alcohol (1.5-2.5 mL) into a septal branch supplying the basal interventricular septum, to generate an iatrogenic infarction (causing myocardial necrosis). This, over the next few months, cause the myocardial remodeling, thereby reducing the LVOT gradient (144,156,191,193,195). Results are visible 3-6 months after ASA (this is the time it takes for the myocardium to reduce) (144,191). ASA is less invasive than surgical myectomy, and enables a shorter hospital stay and recovery time (156,187,191,195). Nevertheless, more arrhythmic episodes are observed after the procedure and in the long term, as a result of scar tissue generation (191). Given that ASA is associated with an increased risk of permanent pacemaker placement or repeated procedures, it should be reserved for individuals with high operative risk or a personal preference of avoiding surgery (186,187).

ASA has results comparable to surgical myectomy (regarding improvement in symptoms, functional status, exercise capacity, LVOT gradient reduction, long-term survival) (186,193,195). AV block is the most frequent complication of ASA (7-20%) (193).

In HCM patients with multiple comorbidities and/or advanced age, ASA can provide a more favorable risk-safety profile (20,156,186,187,196). In this context, the primary disadvantage of ASA is its effect on the conduction system (195). Septal myectomy tends to cause left bundle branch block, while ASA tends to cause right bundle branch block. In patients with pre-existing bundle branch blocks these need to be considered, in order to avoid the development of complete heart block requiring pacemaker implantation (187).

The 2011 Guideline stated that ASA should not be performed in subjects < 21 years of age (class III), and should be discouraged in subjects < 40 years of age or with septal thickness ≥ 30 mm (195).

With ASA, the distribution of the ablated myocardium is completely dependent on the perfusion pattern of the septal perforator branches. For that reason, adequate volume or correct location is less predictable to achieve, compared to septal myectomy that is performed under open visualization (144).

Alternative septal ablation methods have been introduced recently — endocardial radiofrequency ablation may produce modest reduction of basal septal hypertrophy LVOT gradients and functional class. It is administered from the LV endocardial surface, by bypassing the access from coronary arteries (186).

1.8. Machine learning (ML)

ML is a subdiscipline of artificial intelligence (198), wherein computer algorithms are used for learning complex patterns from data (199). ML algorithms easily detect trends and patterns, which could not be identified by the human or classical statistical approaches. ML algorithms handle multidimensional and multivariety data (200). Applications of ML have the potential to be useful across several medical domains (201), including cardiology (199).

In contrast to traditional rule-based algorithms, which apply rules explicitly programmed by humans to perform a sequence of operations, ML algorithms learn rules and patterns from the provided data itself. An ML algorithm is trained on provided data, and a trained ML model is then evaluated on test data it has never seen before (199).

There are different categories of ML algorithms (199). Within ML, tasks could be classified into supervised and unsupervised learning (198).

In supervised learning tasks, ML algorithms learn patterns from provided labeled examples. Relying on the labeled examples exclusively, the algorithm learns the most important features that drive its decisions, and formulates rules by itself to exploit those features. Formulated rules (ML model) can then be applied to data never seen before (199). More specifically, a training dataset with ground-truth labels is utilized to generate a model and optimize its performance. The learnt patterns can then be used for performing the same tasks in new datasets (198).

In unsupervised learning, algorithms learn patterns without labeled examples. Unsupervised learning uncovers patterns by analyzing unlabeled data. It can find similar examples (clustering), and select the most important features for differentiating the examples (199). Unsupervised learning enables the algorithm to create its own understanding of the data, unconstrained by human interpretation. The algorithm may find novel groupings in the data, which a human might not be aware of (199). Once clusters are recognized, they can be further analyzed or visualized (198).

1.8.1. Clustering

Clustering or cluster analysis refers to the division of data into groups (clusters) consisting of similar examples, based on their characteristics (199,202–204). Clustering categorizes the data based on the criterion that the points within one cluster should be maximally similar to each other and maximally different from the samples in other categories (203,205). Clustering has been successful in finding structures within medical datasets (204). For example, the clusters thus determined may lead to the identification of novel patient subgroups of diseases. Notably, identified clusters may suggest new groupings, but all insights must be clinically validated (199). Clustering is an unsupervised method, since it is carried out without prior knowledge of class-label patterns (204,205).

Clustering methods can be categorized into: hierarchical methods and partition methods (204,205). Partitional clustering produces a single, final clustering result. In contrast, hierarchical clustering provides a hierarchy of clusterings (205).

Hierarchical clustering typically constructs a tree structure (hierarchy of clusters), either out of separate clusters with single elements to a single cluster including all individuals or vice versa. There are 2 strategies for hierarchical clustering: an agglomerative (a "bottom-up") and a divisive (a "top-down") approach (202,203,205).

In the agglomerative approach, each observation is considered its own cluster; next, pairs of the most similar clusters are combined as one iteratively; and thus, as we move up the hierarchy, bigger clusters are created at each step. Finally, all the clusters are merged into a single large cluster containing all the objects. In the divisive approach, all observations are placed into one cluster at the beginning; splits are then performed recursively; and likewise, as we move down the hierarchy, an increasing number of smaller clusters are created at each step (202,203,205).

The output of hierarchical clustering is most often visualized as a dendrogram, a tree-based representation of the elements. It provides a good visualization of proximity among data objects and their clusters (202). Through a dendrogram, the processes of hierarchical clustering may be tracked (205). A dendrogram is more informative and intuitive than an unstructured collection of flat clusters (203,205). By cutting the dendrogram at the appropriate layer, a clustering with a desired resolution is gained (205).

1.8.2. Classification

Classification is another important task in ML (206) and is used in biomedicine (207). It is a supervised learning method (208). Classification in ML concerns the problem of identification (prediction) regarding which groups or categories an instance belongs to (206).

Classification can be binary, as in case of determining whether a pathology is present or not, or involve multiple classes, such as determining a particular pathology among several labels (206).

In single-label ML classifiers, each sample can have only one status, while multi-label classification enables each sample to have more than one status (209).

1.8.3. Artificial neural networks (ANNs)

The structure of ANNs is inspired by structures of the human brain (198,210,211). ANNs consist of an input layer (with the number of neurons usually equal to the number of input features), multiple hidden layers, and an output layer (with the number of neurons typically equal to the number of classes to be predicted). Each layer contains numerous artificial neurons usually connected to the neurons in the next layer via weights that reflect the relative importance of a given input (198,211). Each of the units "applies an activation function to the weighted sum of its inputs to calculate its output value". Such propagation of information continues until the final output values in the last layer are calculated (211). With increasing number of neurons in each hidden layer, and with an increasing number of hidden layers, the network becomes increasingly capable to solve more complex problems, but at the same time, it becomes more difficult to train (198).

1.8.4. Deep learning (DL)

DL is a representation learning method wherein input information is transformed into multiple levels of abstractions (212). Deep neural networks are ANNs with many hidden layers (198,211,213) and often non-linear activation functions (213). The composition of deep neural networks is able to model non-linear dependencies (211,213), and to integrate complex collections of data to produce reliable predictions (210,211,213,214). Contrary to other ML methods, DL does not require feature selection prior to training, since it learns higher-order representations directly from raw input data (211). Through layer-wise learning, it progressively builds up abstract representation and automatically extracts the relevant features needed for a given task (212,215).

For pattern recognition tasks, deep convolutional neural networks (DCNNs) are the most frequently used. With a sufficiently large training set, DCNNs can extract relevant features for a given task from the training samples (212).

DL has an important role in practical solutions and knowledge discovery from biomedical big data (with applications in bioinformatics, biomedicine, healthcare, and drug discovery) (216). DL is well-suited for these disciplines, wherein interdependent complex relationships between biological entities and processes exist in the data, and which often occur at multiple scales that are intrinsically noisy (213). DL is particularly

useful for large-scale image classification (210,214,215). Medical images exhibit a large variability due to the countless varied imaging modalities, sample preparation protocols, combinations of phenotypes of interest, and acquisition parameters (214).

Despite the high expectations of what artificial intelligence can bring to medicine, there are several challenges to be overcome before such tools can be integrated into clinical practice (212).

1.8.5. Automatic extraction of molecular mechanisms from databases and literature

Scientific literature may be perceived as a large repository of unstructured data (217). PubMed is a database of biomedical literature and enables the search and retrieval of such literature (218). Filters can be utilized to restrict the search using diverse criteria (publication date, species, etc.). Each publication in the PubMed database has a unique identifier (PMID). The Medical Subject Headings (MeSH) vocabulary thesaurus is utilized for indexing articles for PubMed (219). Other biomedical databases are ChEMBL (220), Pathway Commons (221), DrugBank (222), CTDbase (223), miRTarBase (224), and many more.

Molecular mechanisms underlining different diseases can be extracted from biomedical knowledge resources through automated or manual curation (225,226). The extracted information may be represented visually using visual pathway editors like CellDesigner (225). An automated approach example is the "Integrated Network and Dynamical Reasoning Assembler" (INDRA) (227). INDRA aggregates knowledge extracted from abstracts and full open-access articles, combines it with information from pathway databases using several different machine-reading systems, and subsequently assembles the same into executable models (228).

Knowledge in the form of graphs supports the research on complex systems as an intuitive visual representation and as a data structure readable both by humans and computers (229). Disease maps are comprehensive graph- and knowledge-based representations of molecular mechanisms of diseases (230). Interactive disease maps have been created so far for cancer (231), inflammation (232), Parkinson's disease (233),

Alzheimer's disease (234), asthma (235,236), rheumatoid arthritis (229), influenza A virus replication cycle (237), and others.

2. RESEARCH AIMS AND HYPOTHESES

Research aims:

1. Identification of HCM subphenotypes using clustering
2. Correlation determination between genotypes and subphenotypes in HCM
3. Determination of associations of genotype and phenotype outcomes in HCM using ML algorithms
4. Creation of ML models for HCM outcomes prediction based on genotypic and phenotypic data
5. Identification of genotype-specific echocardiogram findings in HCM using ML algorithms

Hypotheses:

1. HCM subphenotypes can be identified using clustering.
2. There is a correlation between genotypes and subphenotypes in HCM.
3. The associations of genotype and phenotype outcomes in HCM can be determined using ML algorithms.
4. ML models can predict HCM outcomes based on genotypic and phenotypic data.
5. Genotype-specific echocardiogram findings in HCM can be identified using ML algorithms.

3. MATERIALS AND METHODS

3.1. Study design

The study was multi-centric and retroprospective, wherein the retrospective period lasted 18 months and the prospective period was 12 months. Five institutions participated in the study: Institute of Cardiovascular Diseases Vojvodina (Republic of Serbia), University of Belgrade Clinical Centre (Republic of Serbia), Newcastle University Medical School and Newcastle upon Tyne Hospitals NHS Foundation Trust (United Kingdom), University Medical Centre Regensburg (Germany), and Careggi University Hospital Florence (Italy).

The study was approved by the Research Ethics Committees/Institutional Review Boards of each participating institution. The study was conducted in accordance with Good Clinical Practice and the Helsinki Declaration.

3.2. Subjects

The study involved 143 adult patients (≥ 18 years of age) with a confirmed diagnosis of HCM, who met the inclusion and exclusion criteria.

Inclusion criteria were as follows:

1. confirmed diagnosis of obstructive and/or non-obstructive HCM;
2. a history of unexplained LVH with either maximum wall thickness ≥ 15 mm or 13-14 mm (borderline hypertrophy) on ECHO and ≥ 1 fist-degree relative with HCM.

Exclusion criteria were as follows:

1. < 3 months post septal reduction (catheter-based intervention or surgery);
2. clinical decompensation in the previous 3 months, defined as NYHA class IV congestive HF symptoms;
3. resting blood pressure > 180/100 mm Hg;
4. systolic blood pressure < 100 mm Hg;
5. resting LVOT gradient > 50 mm Hg;
6. LVEF < 50% by ECHO;
7. implanted or scheduled pacemaker or cardio-defibrillator in the last 3 months;

8. renal insufficiency with a glomerular filtration rate < 30 mL/min per 1.73m^2;
9. pregnancy or planned pregnancy;
10. life expectancy < 12 months;
11. BMI > 40 kg/m^2;
12. a history of exercise induced ventricular arrhythmias or syncope;
13. presence or history of any other disease with a life expectancy of < 3 years;
14. history of treated or untreated malignancy of any organ system (except localized prostate cancer or localized basal or squamous cell carcinoma of the skin) within past 2 years, regardless of existence of local recurrence or metastases;
15. uncontrolled or life-threatening dysrhythmia, inclusive of symptomatic or sustained ventricular tachycardia and AF or atrial flutter with a resting ventricular rate > 110 beats per minute;
16. participation in organized or competitive sports activities (basketball, football, etc.), burst activity (racket sports, sprinting, etc.), heavy isometric exercise (bodybuilding, etc.) or opposition of refraining from the same for the duration of the study.

The patient eligibility was assessed through the evaluation of patients' electronic health records. Eligible patients were informed about the aims of the study, and signed informed consent forms for participation in the study. Patients' data were collected from hospital information system.

All participants in the study went through genetic counseling, wherein the importance of further cardiological screening of family members was unequivocaly underlined.

3.3. Protocol

3.3.1. Genotype-phenotype associations in HCM: Examination by machine learning algorithms using clinical data

3.3.1.1. Data

3.3.1.1.1. Demographics

Age and sex of patients were recorded.

3.3.1.1.2. Medical and family history

Current symptoms (chest pain, palpitations, fatigue, dyspnea, and syncope), signs (heart murmur, pulmonary crackles, pleural effusion, pretibial edema, and venous congestion), NYHA class, systolic and diastolic blood pressure, and also comorbidities [diabetes mellitus, thyroid disease, phaeochromocytoma, acromegaly, renal dysfunction, hepatic dysfunction, chronic obstructive pulmonary disease (COPD), anemia, neuromuscular disease, amyloidosis, and genetic disease] were registered.

Family history of HCM, DCM, SCD age < 40, SCD age between 40-59, SCD age ≥ 60, unexplained HF, pacemaker or defibrillator implants, cardiac transplantation, and evidence of systemic disease were recorded.

3.3.1.1.3. Anthropometric measurements

Height of the participants was measured using stadiometer, whilst weight was measured using beam scale. BMI was calculated.

3.3.1.1.4. Blood sampling

Blood samples were collected through the antecubital vein venipuncture in the morning after overnight fasting.

3.3.1.1.5. Genetic testing

After isolation of DNA from whole blood using QIAamp DNA Blood BioRobot MDx kit (QIAGEN GmbH, Hilden, Germany), a genetic panel of 41 genes was analyzed by polymerase chain reaction (PCR) and on the Next Generation Sequencing platform (Illumina, Inc., San Diego, USA). For this panel, the samples represent DNA isolated from the blood of patients. These genes are part of the commercially available panel, the TruSight Cardio Sequencing Kit (Illumina, Inc., San Diego, USA), for targeted sequencing for research purposes.

3.3.1.1.5.1. Gene panel

The core of the HCM gene panel analyzed consisted of: *ACTC1*, *ACTN2*, *ANKRD1*, *CSRP3*, *FHL1*, *GLA*, *LAMP2*, *MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *PLN*, *PRKAG2*, *TNNI3*,

TNNT2, and *TPM1*. These genes belong to the HCM gene panel used in the Oxford Medical Genetics Laboratory (OMGL) for blood samples. This core gene panel was designed to cover all HCM causative genes (*MYBPC3*, *MYH7*, *TNNI3*, *TNNT2*, *MYL2*, *MYL3*, *ACTC1*, *TPM1*) according to ACCF and AHA guidelines for HCM diagnosis and treatment (2011); as well as genes for differential diagnosis (*PRKAG2*, *GLA*, *FHL1*); and other validated genes associated with HCM (22). Additionally, the following genes were analysed: *ABCC9*, *CACNA1C*, *CTF1*, *DMD*, *DSC2*, *DSG2*, *DSP*, *ELN*, *FBN*, *HCN4*, *JPH2*, *LAMA4*, *MYH6*, *MYPN*, *NEXN*, *NKX2-5*, *NOTCH1*, *PDLIM3*, *PKP2*, *PTPN11*, *RBM2*, *RYR1*, *SDHA*, *TRPM4*, and *TTN*. All these genes are part of the Illumina TruSight Cardio Sequencing Panel, a commercially available panel for scientific purposes for targeted sequencing of genes implicated in inherited cardiac conditions. For all the genes analyzed, the sample represents DNA isolated from the blood of the patient.

This research study focuses on HCM causative genes. However, all the genes analyzed were used in the generation of models for predicting phenotypic outcomes using only genetic data, to attain more complete and prediction-capable genotypic information.

3.3.1.1.5.1.1. Causative genes

3.3.1.1.5.1.1.1. *MYH7*

The myosin heavy chain 7 (*MYH7*) gene is located on the long arm of chromosome 14 (14q11.2). *MYH7* is a protein-coding gene that encodes the beta (slow) heavy chain subunit of cardiac myosin. *MYH7* is expressed in the ventricle and skeletal muscles rich in slow-twitch type I muscle fibers. Mutations in *MYH7* are associated with HCM, myosin storage myopathy, DCM, and Laing early-onset distal myopathy (238).

3.3.1.1.5.1.1.2. *MYBPC3*

The myosin binding protein C3 (*MYBPC3*) gene is located on the short arm of chromosome 11 (11p11.2). *MYBPC3* is a protein-coding gene that encodes the cardiac isoform of the myosin-binding protein C. The cardiac isoform of *MYBPC3* is expressed exclusively in the myocardium. Mutations in this gene are found to cause familial HCM (239).

3.3.1.1.5.1.1.3. *TNNT2*

The troponin T2, cardiac type (*TNNT2*) gene is located on the long arm of chromosome 1 (1q32.1). *TNNT2* is a protein-coding gene that encodes the tropomyosin-binding subunit of the troponin complex. *TNNT2* is expressed predominantly in the heart. Mutations in this gene have been associated with familial HCM and DCM (240).

3.3.1.1.5.1.1.4. *TPM1*

The tropomyosin 1 (*TPM1*) gene is located on the long arm of chromosome 15 (15q22.2). *TPM1* is a protein-coding gene that encodes tropomyosin alpha-1 chain, which composes the main tropomyosin of striated muscle. *TPM1* is expressed mostly in the heart. Mutations in *TPM1* are associated with type 3 familial HCM (241).

3.3.1.1.5.1.1.5. *MYL2*

The myosin light chain 2 (*MYL2*) gene is located on the long arm of chromosome 12 (12q24.11). *MYL2* is a protein-coding gene that encodes the regulatory light chain associated with cardiac myosin beta (slow) heavy chain. *MYL2* is expressed predominantly in the heart. Mutations in *MYL2* are associated with mid-left ventricular chamber type HCM (242).

3.3.1.1.5.1.1.6. *MYL3*

The myosin light chain 3 (*MYL3*) gene is located on the short arm of chromosome 3 (3p21.31). *MYL3* is a protein-coding gene that encodes the myosin light chain 3. *MYL3* is expressed predominantly in the heart. Mutations in *MYL3* cause mid-left ventricular chamber type HCM (243).

3.3.1.1.5.1.1.7. *TNNI3*

The troponin I3, cardiac type (*TNNI3*) gene is located on the long arm of chromosome 19 (19q13.42). *TNNI3* is a protein-coding gene that encodes Troponin I (TnI)-cardiac protein. TnI together with troponin C (TnC) and troponin T (TnT) forms the troponin complex of the thin filament of striated muscle. TnI is the inhibitory subunit responsible for blocking actin-myosin interactions (enabling relaxation). *TNNI3* is

expressed only in the myocardium. Mutations in *TNNI3* cause familial HCM type 7 and familial RCM (244).

3.3.1.1.5.1.1.8. *ACTC1*

The actin alpha cardiac muscle 1 (*ACTC1*) gene is located on the long arm of chromosome 15 (15q14). *ACTC1* is a protein-coding gene that encodes cardiac muscle alpha actin (belonging to the actin family). The alpha actins are expressed in muscle tissues, and represent a predominant constituent of the contractile apparatus. Mutations in *ACTC1* have been associated with familial HCM and idiopathic DCM (245).

3.3.1.1.5.1.2. Other genes

3.3.1.1.5.1.2.1. *ACTN2*

The actinin alpha 2 (*ACTN2*) gene is located on the long arm of chromosome 1 (1q43). *ACTN2* is a protein-coding gene that encodes a muscle-specific alpha actinin isoform, which is expressed in both heart and skeletal muscles (246).

3.3.1.1.5.1.2.2. *ANKRD1*

The ankyrin repeat domain 1 (*ANKRD1*) gene is located on the long arm of chromosome 10 (10q23.31). *ANKRD1* is a protein-coding gene that encodes cardiac ankyrin repeat protein. Interactions between cardiac ankyrin repeat protein and the sarcomeric proteins titin and myopalladin propose that it can be implicated in the myofibrillar stretch-sensor system (247). *ANKRD1* is greatly expressed in cardiac and skeletal muscles (248).

3.3.1.1.5.1.2.3. *CSRP3*

The cysteine and glycine rich protein 3 (*CSRP3*) gene is located on the short arm of chromosome 11 (11p15.1). *CSRP3* is a protein-coding gene that encodes cysteine and glycine-rich protein 3 (also known as muscle LIM protein or cardiac LIM protein), a member of LIM-domain proteins, implicated in regulation of cellular differentiation and development. *CSRP3* is expressed in the cardiac and skeletal muscles. Mutations in *CSRP3* are presumed to cause heritable HCM and DCM (249,250).

3.3.1.1.5.1.2.4. *FHL1*

The four-and-a-half LIM domains 1 (*FHL1*) gene is located on the long arm of chromosome X (Xq26.3). *FHL1* is a protein-coding gene that encodes four-and-a-half LIM domains protein 1, a member of the four-and-a-half LIM only protein family. *FHL1* is expressed in the fat, heart, and several other tissues. Mutations in *FHL1* have been found in Emery-Dreifuss muscular dystrophy (251).

3.3.1.1.5.1.2.5. *GLA*

The galactosidase alpha (*GLA*) gene is located on the long arm of chromosome X (Xq22.1). *GLA* is a protein-coding gene that encodes enzyme alpha-galactosidase A. *GLA* has a ubiquitous expression in tissues. Mutations in *GLA* have an effect on the synthesis and stability of alpha-galactosidase A, and cause Fabry disease, a rare lysosomal storage disorder (252).

3.3.1.1.5.1.2.6. *LAMP2*

The lysosomal associated membrane protein 2 (*LAMP2*) gene is located on the long arm of chromosome X (Xq24). *LAMP2* is a protein-coding gene that encodes lysosomal associated membrane protein-2 (LAMP-2), a member of the family of membrane glycoproteins. It may play a role in the maintenance, protection, and adhesion of the lysosome. *LAMP2* has a ubiquitous expression in tissues (253).

3.3.1.1.5.1.2.7. *PLN*

The phospholamban (*PLN*) gene is located on the long arm of chromosome 6 (6q22.31). *PLN* is a protein-coding gene that encodes phospholamban, a main regulator of cardiac diastolic function. *PLN* is expressed in the heart as well as in other tissues in lower amounts. Mutations in *PLN* cause familial HCM and inherited human DCM with a refractory congestive HF (254).

3.3.1.1.5.1.2.8. *PRKAG2*

The protein kinase AMP-activated non-catalytic subunit gamma 2 (*PRKAG2*) gene is located on the long arm of chromosome 7 (7q36.1). *PRKAG2* is a protein-coding gene that encodes the γ2 regulatory subunit of AMP-activated protein kinase (AMPK). AMPK is an energy-sensing enzyme that monitors cellular energy status. *PRKAG2* has a ubiquitous expression in tissues. Mutations in *PRKAG2* have been associated with familial HCM, Wolff-Parkinson-White syndrome, as well as glycogen storage disease of the heart (255).

3.3.1.1.5.1.2.9. *JPH2*

Junctophilin 2 (*JPH2*) gene is located on the long arm of chromosome 20 (20q13.12). Junctional complexes between endoplasmic/sarcoplasmic reticulum and the plasma membrane are a common structural feature of all excitable cells. They enable communication between intracellular ion channels and cell surface. The protein encoded by *JPH2*, junctophilin 2, is a component of junctional complexes (256).

3.3.1.1.5.1.2.10. *DSG2*

Desmoglein 2 (*DSG2*) gene is located on the long arm of chromosome 18 (18q12.1). *DSG2* encodes desmoglein 2, a member of the desmoglein family. Desmogleins are Ca^{2+}-binding transmembrane glycoprotein constituents of desmosomes (which are cell-cell junctions between cell types that participate in intense mechanical stress, like in the myocardium). Mutations in *DSG2* have been associated with arrhythmogenic right ventricular dysplasia, familial, 10 (257).

3.3.1.1.5.1.2.11. *TRPM4*

Transient receptor potential cation channel subfamily M member 4 (*TRPM4*) gene is located on the long arm of chromosome 19 (19q13.33). The protein encoded by *TRPM4* is a calcium-activated nonselective ion channel. Although the activity of the encoded protein increases with rising intracellular Ca^{2+} levels, this channel does not transport calcium (258).

3.3.1.1.5.1.2.12. *TTN*

Titin (*TTN*) gene is located on the long arm of chromosome 2 (2q31.2). *TTN* encodes titin, an abundant protein in the striated muscle. One titin molecule stretches over half the length of a sarcomere. It guides the contractile machinery assembly in muscle cells. Mutations in *TTN* are associated with familial HCM 9 (259).

3.3.1.1.5.1.2.13. *RYR1*

Ryanodine receptor 1 (*RYR1*) gene is located on the long arm of chromosome 19 (19q13.2). Ryanodine receptor 1 is a calcium channel in the sarcoplasmic reticulum. Mutations in *RYR1* are associated with minicore myopathy along with external ophthalmoplegia, malignant hyperthermia susceptibility, and central core disease (260).

3.3.1.1.5.1.2.14. *NEXN*

Nexilin F-actin binding protein (*NEXN*) gene is located on the short arm of chromosome 1 (1p31.1). Nexilin F-actin binding protein is an actin-binding protein with the most probable function in cell migration and adhesion. Mutations in *NEXN* have been associated with DCM (261).

3.3.1.1.5.1.2.15. *DSC2*

Desmocollin 2 (*DSC2*) gene is located on the long arm of chromosome 18 (18q12.1). Desmocollin 2 is a member of the desmocollin protein family. Desmocollins are one of main constituents of the desmosome. Desmosomes are cell-cell junctions found mostly in cells experiencing mechanical stress. Mutations in *DSC2* are associated with arrhythmogenic right ventricular dysplasia-11 (262).

3.3.1.1.5.1.2.16. *ABCC9*

ATP binding cassette subfamily C member 9 (*ABCC9*) gene is located on the short arm of chromosome 12 (12p12.1). ATP binding cassette subfamily C member 9 is a member of the superfamily of ATP-binding cassette (ABC) transporters. They transport diverse molecules across intra- and extra-cellular membranes. Mutations in *ABCC9* are associated with DCM type 10 (263).

3.3.1.1.5.1.2.17. *DSP*

Desmoplakin (*DSP*) gene is located on the short arm of chromosome 6 (6p24.3). Desmoplakin anchors intermediate filaments to desmosomal plaques and establishes a necessary element of functional desmosomes. Mutations in *DSP* cause several cardiomyopathies and keratodermas (264).

3.3.1.1.5.1.2.18. *FBN1*

Fibrillin 1 (*FBN1*) gene is located on the long arm of chromosome 15 (15q21.1). Fibrillin 1 is a member of the fibrillin protein family. Fibrillin-1 is an extracellular matrix glycoprotein, a constituent of calcium-binding microfibrils, which enable force-bearing structural support. Mutations in *FBN* are associated with Marfan syndrome, Shprintzen-Goldberg syndrome, Weill-Marchesani syndrome, neonatal progeroid syndrome, and ectopia lentis syndrome (265).

3.3.1.1.5.1.2.19. *CTF1*

Cardiotrophin 1 (*CTF1*) gene is located on the short arm of chromosome 16 (16p11.2). Cardiotrophin 1 is a cytokine that leads to cardiac myocyte hypertrophy *in vitro* (266).

3.3.1.1.5.1.2.20. *CACNA1C*

Calcium voltage-gated channel subunit alpha1 C (*CACNA1C*) gene is located on the short arm of chromosome 12 (12p13.33). Calcium voltage-gated channel subunit alpha1 C is an alpha-1 subunit of a voltage-dependent calcium channel. Calcium channels perform the Ca^{2+} ions influx into the cell upon membrane polarization. The alpha-1 subunit constitutes the pore of the channel (267).

3.3.1.1.5.1.2.21. *ELN*

Elastin (*ELN*) gene is located on the long arm of chromosome 7 (7p11.23). Elastin is a constituent of elastic fibers. Mutations and deletions in *ELN* are associated with autosomal dominant cutis laxa and supravalvular aortic stenosis (268).

3.3.1.1.5.1.2.22. *NOTCH1*

Notch receptor 1 (*NOTCH1*) gene is located on the long arm of chromosome 9 (9q34.3). Notch receptor 1 is a member of the NOTCH protein family. Notch signaling controls interreactions between adjoining cells. Notch receptor 1 has a role in the development of different tissues and cells. Mutations in *NOTCH1* are associated with aortic valve disease, T-cell acute lymphoblastic leukemia, chronic lymphocytic leukemia, head and neck squamous cell carcinoma, and Adams-Oliver syndrome (269).

3.3.1.1.5.1.2.23. *PTPN11*

Protein tyrosine phosphatase non-receptor type 11 (*PTPN11*) gene is located on the long arm of chromosome 12 (12q24.13). Protein tyrosine phosphatase non-receptor type 11 is a member of the protein tyrosine phosphatase (PTP) family. PTP signaling controls several different cellular processes such as cell growth and differentiation, as well as mitotic cycle and oncogenic transformation. Mutations in *PTPN1* are a cause of Noonan syndrome and acute myeloid leukemia (270).

3.3.1.1.5.1.2.24. *MYH6*

Myosin heavy chain 6 (*MYH6*) gene is located on the long arm of chromosome 14 (14q11.2). Cardiac muscle myosin contains 2 light chain subunits, 2 heavy chain subunits, and 2 regulatory subunits. *MYH6* encodes the alpha heavy chain subunit of cardiac myosin. Mutations in *MYH6* cause familial HCM and atrial septal defect 3 (271).

3.3.1.1.5.1.2.25. *RBM20*

RNA binding motif protein 20 (*RBM20*) gene is located on the long arm of chromosome 10 (10q25.2). RNA binding motif protein 20 binds RNA and controls splicing. Mutations in *RBM20* have been associated with familial DCM (272).

3.3.1.1.5.1.2.26. *DMD*

Dystrophin (*DMD*) gene is located on the short arm of chromosome X (Xp21.2-p21.1) (273). Its protein product dystrophin connects the sarcomere to the extracellular matrix shielding the sarcolemma from contraction-produced injury (274). *DMD* gene is

the greatest in human genome (274,275). There are at least 17 *DMD* transcript variants (275). The mutation are frequent in *DMD* gene (274). Deletion, duplication, and point mutation at this gene locus can cause Duchenne muscular dystrophy, Becker muscular dystrophy, or cardiomyopathy (273).

3.3.1.1.5.1.2.27. *LAMA4*

Laminin subunit alpha 4 (*LAMA4*) gene is located on the long arm of chromosome 6 (6q21). Laminins are extracellular matrix glycoproteins, the principal non-collagenous components of basement membranes. They play a role in signaling, migration, adhesion, differentiation, and metastasis. Laminins contain 3 non-identical chains. Each of the chains is encoded by a distinct gene. *LAMA4* encodes the alpha chain isoform laminin, alpha 4. Its precise role is unknown (276).

3.3.1.1.5.1.2.28. *SDHA*

Succinate dehydrogenase complex flavoprotein subunit A (*SDHA*) gene is located on the short arm of chromosome 5 (5p15.33). Succinate dehydrogenase complex flavoprotein subunit A is the principal catalytic subunit of succinate-ubiquinone oxidoreductase, a part of the mitochondrial respiratory chain. Mutations in *SDHA* gene have been associated with Leigh syndrome (277).

3.3.1.1.5.1.2.29. *HCN4*

Hyperpolarization activated cyclic nucleotide gated potassium channel 4 (*HCN4*) gene is located on the long arm of chromosome 15 (15q24.1). Hyperpolarization activated cyclic nucleotide gated potassium channel 4 is essential for the cardiac-pacemaking process. Mutations in *HCN4* gene have been linked to AF with bradyarrhythmia (familial sinus bradycardia) (278).

3.3.1.1.5.1.2.30. *PKP2*

Plakophilin 2 (*PKP2*) gene is located on the short arm of chromosome 12 (12p11.21). Plakophilin proteins have a role in connecting cadherins to the cytoskeleton's intermediate filaments. *PKP2* gene is necessary to maintain transcription of genes

controlling intracellular Ca^{2+} cycling. Mutations in this gene are associated with Brugada syndrome, ACM, and idiopathic ventricular fibrillation (279).

3.3.1.1.5.1.2.31. *PDLIM3*

PDZ and LIM domain 3 (*PDLIM3*) gene is located on the long arm of chromosome 4 (4q35.1). PDZ and LIM domain 3 protein contains a PDZ and a LIM domains, which indicates that it may be implicated in cytoskeletal assembly. It colocalizes with skeletal muscle alpha-actinin-2 at the Z lines. Aberrant alternative splicing of *PDLIM3* may be involved in myotonic dystrophy (280).

3.3.1.1.5.1.2.32. *NKX2-5*

NK2 homeobox 5 (*NKX2-5*) gene is located on the long arm of chromosome 5 (5q35.1). NK2 homeobox 5 is a homeobox-containing transcription factor involved in heart development. Mutations in *NKX2-5* gene cause atrial septal defect with AV conduction defect, tetralogy of Fallot, and congenital hypothyroidism non-goitrous type 5 (281).

3.3.1.1.5.1.2.33. *MYPN*

Myopalladin (*MYPN*) gene is located on the long arm of chromosome 10 (10q21.3). *MYPN* encodes a protein, which interacts with nebulette in the cardiac muscle. Myopalladin has both structural and regulatory functions (282).

3.3.1.1.6. Blood markers

A complete blood count was done from a whole blood sample using a hematology analyzer.

Serum and plasma samples were used for blood-markers testing. Serum was prepared by allowing the whole blood samples to clot after collection, by leaving it undisturbed at room temperature, followed by centrifugation performed at 3500 revolutions per minute (rpm) for 15 minutes. The plasma was prepared by centrifugation of whole blood samples, performed at 3500 rpm for 15 minutes.

Blood-markers testing included: glucose, urea, creatinine, uric acid, sodium, potassium, calcium, alanine transaminase (ALT), aspartate transaminase (AST), creatine kinase (CK), creatine kinase MB isoenzyme (CK-MB), lactate dehydrogenase (LDH), troponin, N-terminal prohormone of brain natriuretic peptide (NT-proBNP), total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, triglycerides, total protein, and albumin. Estimated glomerular filtration rate (eGFR) was calculated using Cockcroft-Gault equation.

3.3.1.1.7. Transthoracic echocardiography with Doppler

A transthoracic ECHO with Doppler was performed. ECHOs were recorded in real time during 3 cardiac cycles in the standard parasternal (long-axis) and apical views (apical 4, apical 2, and apical long). Parasternal short-axis views were acquired at 3 levels: basal, midpapillary, and apical. Apical 4-chamber view was used for right ventricular evaluation. Peak velocity of the LVOT was recorded from the apical 5-chamber view using pulse Doppler and used to calculate the pressure gradient.

The following parameters were registered: left atrial (LA) size, left atrial volume (LAV), left atrial volume in systole (LAVs), transmitral maximal pressure gradient (MV maxPG), transmitral mean pressure gradient (MV meanPG), mitral valve velocity-time integral (MVVTI), mitral annulus diameter at end-diastole (MADd), diastolic mitral annular area diameter (MAAd), mitral valve area (MVA), MR, vena contracta, effective regurgitant orifice area for mitral regurgitation (MR ERO), mitral regurgitation regurgitant volumes (MR RV), SAM, papillary muscle abnormalities, mitral leaflet abnormalities, calcification of mitral annulus, interventricular septal thickness at end-diastole (IVSd), posterior LV wall thickness at end-diastole (PLWd), left ventricular internal dimension at end-systole (LVIDs), left ventricular internal dimension at end-diastole (LVIDd), end-diastolic volume of the left ventricle (EDVLV), end-systolic volume of left ventricle (ESVLV), stroke volume of the left ventricle (SVLV), EFLV, LVOT maximal velocity (LVOT Vmax), LVOT peak gradient (LVOT maxPG), LVOT peak gradient (LVOT maxPG) after Valsalva maneuver, myocardial fibrosis, hypokinesia, akinesia, dyskinesia, hyperkinesia, ratio of peak velocity of early diastolic transmitral flow to peak velocity of early diastolic mitral annular motion as determined by pulsed

wave Doppler (E/E'), diastolic dysfunction grade, aortic valve peak pressure gradient (AV maxPG), aortic valve mean pressure gradient (AV meanPG), aortic valvular velocity time integral (AVVTI), aortic valve peak velocity (AV Vmax), aortic valve area (AVA) by planimetry, aortic valve area (AVA) by Doppler, aortic regurgitation (AR), aortic valve (unicuspid/bicuspid/tricuspid), aortic root diameter (AO), aortic leaflet separation diameter (AOvs), ascending aorta diameter (AscAO), right atrial volume at end systole (RAVs), transverse aortic arch during end-diastole (TADd), tubular ascending aorta during end-diastole (TAAd), tricuspid regurgitation (TR), tricuspid annular plane systolic excursion (TAPSE), and right ventricle systolic pressure (RVSP).

3.3.1.1.8. Cardiopulmonary exercise testing

A progressive CPET with a cycle ergometer was conducted to evaluate participants' cardiac, pulmonary, and circulatory functions. ECG was continuously monitored at rest, during the exercise, and in the recovery phase. Heart rate and blood pressure were regularly measured. At the beginning, resting baseline values were recorded for 3 minutes. Next, the participants pedaled for 3 minutes without any resistance at a pedal frequency of 60-70 rpm, and subsequently resistance has been constantly increased at the predetermined ramp rate of 10 W/minute. The CPET was ended, if participants voluntary stopped the exercise due to severe symptoms (like exhaustion or breathlessness), or were not able to keep pedal-frequency > 60 rpm, or if complications (such as significant ST-segment deviation) arose.

The following CPET parameters were registered: peak oxygen consumption (peak VO_2), anaerobic threshold (AT), peak heart rate, peak respiratory exchange ratio (RER), and ventilatory efficiency slope (VE/VCO_2 slope).

3.3.1.1.9. Electrocardiography and ECG-holter-monitoring

The electrocardiography was conducted in supine position using a standard 12-lead-electrocardiograph. For identification of sporadic arrhythmia, participants wore an ECG-holter-monitor for 24 hours and kept a diary of symptoms and activities.

Heart rate and heart axis were determined. Based on ECG, heart rhythm was evaluated: presence of sinus rhythm, atrial flutter, AF, paroxysmal supraventricular

tachycardia (PSVT), and NSVT. Conduction parameters were registered as well: AV block I, AV block II (Mobitz 1), AV block II (Mobitz 2), AV block III, left bundle branch block (LBBB), right bundle branch block (RBBB), left anterior hemiblock, and right anterior hemiblock. Further, PR interval, QRS duration, Sokolow index, presence of significant Q wave, ST segment abnormalities, and negative T wave in ECG were also evaluated.

3.3.1.2. Data analysis

3.3.1.2.1. Data wrangling

Most of the preparation, transformation, and structuring of raw genotypic and phenotypic data into format appropriate for further analysis, as well as data exploration were done using Python's library Pandas v. 1.4.3 (283).

3.3.1.2.2. Identification of HCM subphenotypes using clustering

Since clustering is division of data into groups (clusters) of similar examples, and clustering here was done based on patients' phenotypic data, groups of similar patients were found based on their phenotypic data. Therefore, identified clusters represent HCM subphenotypes, and terms clusters and subphenotypes were used interchangeably.

Appropriately prepared phenotypic data were analyzed by clustering algorithms (unsupervised ML). Hierarchical clustering (agglomerative clustering, affinity="euclidean", linkage="ward") was accomplished employing Scikit-learn module v. 1.1.1. and SciPy library v. 1.8.1 and dendrogram was visualized using MatPlotLib v. 3.5.2 (283).

Only first visits of the patients were included in the cluster analysis. The data from the second visit would interfere with cluster analysis, so they were omitted. No combinations of features were made — features were analyzed in their original form. Features containing > 30% missing values were eliminated from the dataset, and all other missing values were imputed by KNNImputer of Scikit-learn (n_neighbors = 12, weights = "uniform"). Numerical features were standardized employing Scikit-learn StandardScaler (283).

Since both numerical and categorical data were analyzed, to minimize further data manipulation, K-Prototypes algorithm from Kmodes package was used for clustering (appropriate for direct use, when dataset contain mixed categorical and numerical values). For determining the optimal number of clusters, elbow method was used and the result was validated using Kneelocator (https://pypi.org/project/kneed/). Cluster characteristics were visualized employing Seaborn library v. 0.11.2 (283).

Currently there is no technique to interpret K-Prototypes clustering directly. For this reason, two indirect methods to interpret K-Prototypes clustering were used — first, decision tree classification (classes were set up as determined in the clustering, and predicted using on all the data used for clustering) with dendrogram generation (using sklearn.tree.plot_tree). The second method involved feature importance computation for random forest classification (using the fitted attribute of the random forest algorithm, feature_importances_), which was performed after the clustering (classes were set up as determined in the clustering). Since decision trees make locally optimal choices, random forest classification was performed to gain more stable and general feature importance (283).

3.3.1.2.3. Genotype-phenotypic outcomes associations' identification and generation of models for outcome prediction based on genotypic and phenotypic data

Associations of genotype and phenotypic outcomes were evaluated used Python modules Scikit-learn and SHapley Additive exPlanation (SHAP). Models for outcome prediction based on genotypic and phenotypic data were built using Python modules Scikit-learn and SHAP. In order to generate synthetic samples for the minority class in classification tasks, Synthetic Minority Oversampling Technique (SMOTE) from imbalanced-learn library was used. Feature selection was performed using Scikit-learn SelectKBest (score_func = f_classif), SelectPercentile, Scikit-learn VarianceThreshold (threshold = 0.02), Recursive Feature Elimination, and domain knowledge. Feature contribution to model predictions was assessed using SHAP module v. 0.41.0 (283).

The presence of diverse phenotypic outcomes was predicted in classification tasks. Altogether, 268 visits were covered by the analysis. For majority of participants, data

were collected for 2 visits. For a minority of patients, data were obtained for the first visit only, due to loss of follow-up (283).

Feature engineering was performed: numerical features were combined by division, categorical by multiplication, and numerical and categorical with each other by their multiplication. Custom features were also generated as a multiplication or sum of features that conjointly represent a rational clinical entity (283).

Features with > 30% missing values were excluded from the analysis, and all other missing values were imputed employing KNNImputer of Scikit-learn (n_neighbors = 12, weights = "uniform"). Standardization of numerical features was performed using Scikit-learn StandardScaler. Imputation and standardization were implemented as a pipeline, and the pipeline was applied to training and test data separately. Training sets contained 188 (75.80%) visits (283). In cases where data were available for two visits, both visits were assigned to either training or test set exclusively (284).

For Scikit-learn, estimators' parameters default values were applied, and additionally, for logistic regression, class_weight = "balanced" was applied (283). For each classification task, tree-based pipeline optimization (TPOT) was also employed (285).

Five-fold cross-validation was applied. Accuracy, precision, recall, F1-score, AUC, and average precision (AP, under the PR curve) were used as evaluation metrics (283).

Global feature importance was assessed as mean absolute Shapley values per feature across the data, using SHAP v. 0.41.0 (283).

3.3.1.2.4. Identification of genotype-specific echocardiogram findings

Ultrasonic records in DICOM format were converted to JPG images using RadiAnt DICOM Viewer v. 2021.2.2. Images were grouped based on the views they represented. Ultrasonic images showing beginning of P-wave and T-wave on ECG (as a representative image of ventricular diastole and ventricular systole phase) were detached, each to a separate image batch. In order to remove background noise (patient details, date, heart rate, ECG, etc.), custom masker was used. The masker was built using Python OpenCV library. Masked ultrasonic images, showing only the region of interest, were used for further analysis (Figure 3).

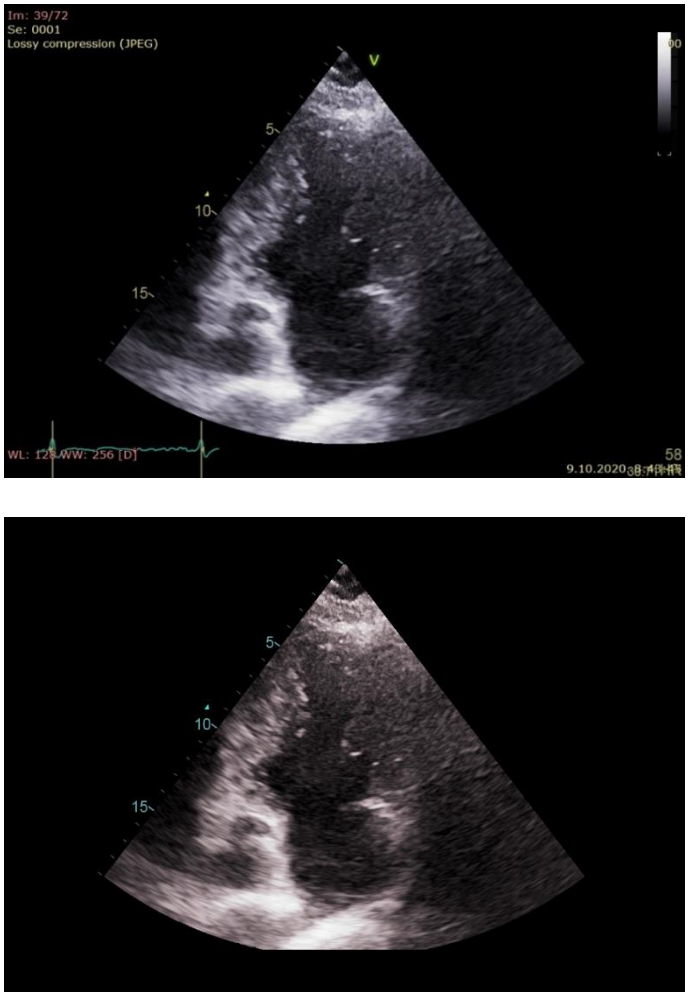

Figure 3. Same ultrasonic image before (up) and after (down) custom masker application

Genotype-specific ECHO findings were identified using Python DL and computer vision library Fast AI. This was accomplished by generation of DL models for classification of ultrasonic images, based on the underlying genotype and later analysis of the most decisive image regions (which most directed the prediction of the model towards one class or another). For the creation of classification model Autokeras was consulted. Models were built using the following setting: image_size = 224, batch_size = 4, validation_percentage = 0.15, without data augmentation, using pretrained ResNet18.

3.3.2. Molecular mechanisms of genotype-phenotype associations in HCM: Examination by machines using literature data

3.3.2.1. HCM Map: Interactive knowledge resource about molecular mechanisms of HCM

In order to generate a disease-specific knowledge base containing a formalized visual review of all principal molecular pathways implicated in HCM, HCM Map was generated (287).

The HCM Map was built in Systems Biology Markup Language (SBML) using CellDesigner software v. 4.4.2. The HCM Map was visualized as a Systems Biology Graphical Notation (SBGN) diagram, and was made interactive and publicly available using Molecular Interaction NEtwoRks VisuAlization (MINERVA) platform (287).

The HCM Map was constructed manually through the incorporation of molecular interactions extracted from PubMed database articles about human HCM molecular mechanisms. The construction begun with a comprehensive manual article search in PubMed, using relevant key phrases such as “gene hypertrophic cardiomyopathy,” “signaling hypertrophic cardiomyopathy,” “micro RNA hypertrophic cardiomyopathy,” “noncoding RNA hypertrophic cardiomyopathy,” and more. The filter “10 years” (for including the period 2010-2020) was applied for selection of the articles. A well-demonstrated “consensus” information was collected from major reviews, incorporated into the HCM Map in the first phase, and specific aspects from recent original research articles were integrated afterward (287).

The following plugins were added: Gene set enrichment analysis (GSEA) plugin (286), Adverse drug reactions plugin (286), Disease-variant associations plugin (286), and Map exploration plugin (286).

3.3.2.2. Molecular mechanisms of genotype-phenotype associations in HCM collected and represented by machines

The study aimed to gather knowledge related to molecular aspects of HCM and its clinical presentations, which are dispersed in scientific papers. For this purpose, by the automated extraction of molecular mechanisms, 4 models were generated by machines and subsequently analysed (287).

3.3.2.2.1. Models' generation

The INDRA-assembled PubMed HCM model was automatically assembled, applying INDRA's PubMed literature client with the search term "hypertrophic cardiomyopathy" (major_topic = True), and filtering out retrieved results published before January 1, 2010 (287). The literature was automatically read by REading and Assembling Contextual and Holistic mechanisms from text (REACH) reading system (288). The extracted statements were then grounded, mapped, preassembled (de-duplicated and organized in a hierarchy), and assembled employing Cytoscape networks assembler (287).

INDRA-assembled PubMed+PathwayCommons HCM model was assembled automatically using Pathway Commons database via BioPAX API of INDRA and PubMed literature client of INDRA (287). The first set of statements was gathered from the Pathway Commons database (221) via BioPAX API of INDRA, with "neighborhood" query, for genes: *ACTC1*, *ACTN2*, *ANKRD1*, *CALR3*, *CASQ2*, *CAV3*, *CRYAB*, *CSRP3*, *DES*, *FHL1*, *FLNC*, *GAA*, *GLA*, *JPH2*, *LAMP2*, *LDB3*, *MYBPC3*, *MYH6*, *MYH7*, *MYL2*, *MYL3*, *MYLK2*, *MYOZ2*, *MYPN*, *NEXN*, *PLN*, *PRKAG2*, *TCAP*, *TNNC1*, *TNNI3*, *TNNT2*, *TPM1*, *TTR*, and *VCL* (287). The second set of statements for this model was composed by PubMed literature client of INDRA with the search term "hypertrophic cardiomyopathy" (major_topic = True), and by filtering out retrieved results published before January 1, 2010 (287). The literature was automatically read using the REACH reading system (288). All the statements from both sets were gathered, grounded, mapped, preassembled (de-duplicated and organized in a hierarchy), and later assembled employing networks assembler of Cytoscape (287).

Statements for Truncated INDRA DB model were drawn from the INDRA Database with the MeSH query constraint "Cardiomyopathy, Hypertrophic, Familial." Only completely correctly extracted statements were included into the model. The criterion for correctness was that all aspects of the statement (labels, subject, object, interaction direction, and type) had to be extracted the identically as careful manual curation would. The statements then were manually transcribed as interactions and nodes in a network table in XLSX format (287).

Statements for INDRA DB model were drawn from the INDRA Database with the MeSH query constraint “Cardiomyopathy, Hypertrophic, Familial.” All statements found were included into the model. The statements were transcribed as interactions and nodes in a network table in XLSX format (287).

3.3.2.2.2. Network analysis

For further network analysis, all models were imported to Cytoscape v. 3.8.2 (289) and later uploaded to NDEx v. 2.5.0 (290–292). Top 10% of elements for each of the 12 centrality measures [closeness, eccentricity, betweenness, degree, clustering coefficient, radiality, bottleneck, stress, edge percolated component (EPC), density of maximum neighborhood component (DMNC), maximum neighborhood component (MNC), maximal clique centrality (MCC)] in each network were visualized (287) employing the Cytoscape CytoHubba app v. 0.1 (293) and uploaded to NDEx (287). Partition of nodes into shells, based on the rank of their importance in the network, was performed employing the Cytoscape Wk-shell-decomposition app v. 1.1.0 (287,294). Models with a lower level of noise were created employing the Cytoscape PE-measure app v. 1.0 (287,295).

3.3.2.3. Molecular mechanisms of different clinical presentations of HCM collected and represented by machines

3.3.2.3.1. Generation of models

Molecular mechanisms of HCM and its various clinical presentations were compiled using the INDRA database (227). From all PubMed articles published after 1 January 2010, molecular mechanisms were collected in the form of INDRA statements (296), separately for HCM and 19 of its clinical presentations. INDRA statements were found by PubMed identifiers (PMIDs) in the INDRA database, employing REST Client API. PMIDs were congregated through the INDRA PubMed client (228) using the search terms: hypertrophic cardiomyopathy, myofibrillar disarray, cardiomyocyte disarray, cardiomyocyte hypertrophy, myocardial fibrosis, myocardial hypercontractility, myocardial stiffness, impaired myocardial relaxation, impaired cardiac relaxation, myocardial remodeling, cardiac remodeling, diastolic dysfunction, coronary

microvascular dysfunction, myocardial ischemia, atrial fibrillation, sudden cardiac death, left ventricular outflow tract obstruction, heart failure, major adverse cardiovascular events, and rehospitalization (use_text_word = True, major_topic = True) (92).

Intersections of the sets containing INDRA statements for HCM and its clinical presentations were generated; each of them was transcribed to a network table, visualized, analysed using Cytoscape v. 3.8.2 (92,289), and uploaded to NDEx v 2.5.0 (290–292).

3.3.2.3.2. Network analysis

The most important nodes of intersection networks were identified and partitioned into shells employing Cytoscape Wk shell decomposition v. 1.1.0 (92,294). The reliability of interactions was calculated using Cytoscape PE-measure application v. 1.0 (92,295), and models with a lower level of noise were created. Cooperatively working elements were found using NCMine Cytoscape plugin version 1.3.0 (92,297), whereas the nodes' centrality scores were calculated using Cytoscape CytoHubba app version 0.1 (92,293).

3.3.2.4. HCM clinical course dynamics on smaller scale

The HCM Clinical interactive knowledge resource was built using CellDesigner v. 4.4.2, following literature and databases exploration and compaction, as well as gathering of the elements found. The elements incorporated are in line with the HCM clinical practice guidelines. Components were annotated using the Minimal Information Requested In the Annotations of Models (MIRIAM). The HCM Clinical was converted into an interactive HCM knowledge resource and made publicly available using the Molecular Interaction NEtwoRks VisuAlization (MINERVA) platform v. 15.1.2.

3.4. Definitions of terms

Outcome is defined as any phenotypic expression that can be determined qualitatively, quantitatively, or by standard medical practices.

(Normal) Sinus rhythm is the (normal) rhythm of the heart where electrical stimuli are initiated in the sinoatrial (SA) node.

Cardiac cycle refers to all of the events that occur from the beginning of one heartbeat to the beginning of the next.

3.5. Statistical analyses

Statistical analyses were conducted using licensed SPSS software v. 28.0.1.1. (IBM SPSS Statistics, Armonk, New York). Mean values of continuous variables were compared using ANOVA, whereas categorical variables were compared through the chi-square test. The statistical significance was set at the p value of < 0.05. Phi coefficient was used as a measure of relationship between presence/absence of mutations in genes and presence/absence of phenotypic outcomes. Comparison of agreement between real and determined clusters were estimated using Cohen's kappa.

4. RESULTS

4.1. Genotype-phenotype associations in HCM: Examination by machine learning algorithms using clinical data

4.1.1. Genetic variants

Detected variants are shown in Table 6.

Table 6. Detected variants

Mutated gene	Genetic variant Coding DNA reference sequence	Genetic variant Protein reference sequence	Reference SNP ID	Consequence	Clinical significance by ClinVar database
MYBPC3	c.3192dup	p.Lys1065Glnfs*12	rs397516007	Frameshift Variant	Pathogenic
MYBPC3	c.1624G>C	p.Glu542Gln	rs121909374	Stop Gained	Pathogenic
MYBPC3	c.1615A>G	p.Ile539Val	-	-	-
MYBPC3	c.772C>A	p.Glu258Lys	-	-	-
MYBPC3	c.772G>A	p.Glu258Lys	rs397516074	Missense Variant	Pathogenic/Likely pathogenic
MYBPC3	c.2398G>A	p.Gly800Arg	rs727504574	Missense Variant	Uncertain significance
MYBPC3	c.2846dup	p.Met949Ilefs*102	-	-	-
MYBPC3	c.1458-1G>A	p.?	rs397515903	Splice Acceptor Variant	Pathogenic
MYBPC3	c.2689_2698del	p.Gly897Alafs*24	-	-	-
MYBPC3	c.407-1G>A	p.?	-	-	-
MYBPC3	c.1591G>C	p.Gly531Arg	rs397515912	Missense Variant	Pathogenic/Likely pathogenic
MYBPC3	c.913_914del	p.Phe305Profs*27	rs397516080	Frameshift Variant	Pathogenic/Likely pathogenic
MYBPC3	c.1020C>G	p.Tyr340*	-	-	-
MYBPC3	c.1505G>A	p.Arg502Gln	rs397515907	Missense Variant	Pathogenic/Likely pathogenic
MYBPC3	c.2728C>A	p.Pro910Thr	rs397515985	Missense Variant	Conflicting interpretations of pathogenicity Likely benign(1);Uncertain significance(9)
MYBPC3	c.1112C>G	p.Pro371Arg	rs397515887	Missense Variant	Uncertain significance
MYBPC3	c.1504C>T	p.Arg502Trp	rs375882485	Missense Variant	Conflicting interpretations of pathogenicity Likely pathogenic(2);Pathogenic(15);Uncertain significance(2)
MYBPC3	c.1928-2A>G	p.?	rs397515937	Splice Acceptor Variant	Pathogenic

Table 6 (continued).

MYBPC3	c.2441_2443del	p.Lys814del	rs727504288	Inframe Deletion	Conflicting interpretations of pathogenicity Likely pathogenic(2);Uncertain significance(10)
MYBPC3	c.3226_3227insT	p.Asp1076fs	rs397516008	Frameshift Variant	Pathogenic
MYBPC3	c.517_519del	p.?	-	-	-
MYBPC3	c.655G>C	p.Val219Leu	rs397516068	Missense Variant	Pathogenic/Likely pathogenic
MYBPC3	c.1174del	p.Ala392Leufs*14	rs1565628486	Frameshift Variant	Pathogenic
MYBPC3	c.1789C>T	p.Arg597Trp	rs201596087	Missense Variant	Uncertain significance
MYBPC3	c.1090G>A	p.Ala364Thr	rs794727046	Missense Variant	Pathogenic/Likely pathogenic
MYBPC3	c.2429G>T	p.Arg810Leu	rs375675796	Missense Variant	Conflicting interpretations of pathogenicity Likely pathogenic(3);Uncertain significance(1)
MYBPC3	c.1224-52G>A	p.?	rs786204336	Intron variant	Pathogenic/Likely pathogenic
MYBPC3	c.2373dupG	p.Trp792Valfs	rs397515963	Frameshift Variant	Pathogenic
MYBPC3	c.3065G>C	p.Arg1022Pro	rs397516000	Missense Variant	Conflicting interpretations of pathogenicity Likely pathogenic(3);Pathogenic(1) ;Uncertain significance(4)
MYBPC3	c*26+2T>C	p.?	-	-	-
MYH7	c.4130C>T	p.Thr1377Met	rs397516201	Missense Variant	Pathogenic
MYH7	c.2167C>T	p.Arg723Cys	rs121913630	Missense Variant	Pathogenic
MYH7	c.2346C>A	p.Ser782Arg	rs730880736	Missense Variant	Likely pathogenic
MYH7	c.2302G>A	p.Gly768Arg	rs727503260	Missense Variant	Pathogenic/Likely pathogenic
MYH7	c.2207T>C	p.Ile736Thr	rs727503261	Missense Variant	Pathogenic
MYH7	c.2389G>A	p.Ala797Thr	rs3218716	Missense Variant	Pathogenic/Likely pathogenic
MYH7	c.2722C>G	p.Leu908Val	rs121913631	Missense Variant	Pathogenic
MYH7	c.715G>A	p.Asp239Asn	rs397516264	Missense Variant	Likely pathogenic
TNNT2	c.274C>T	p.Arg92Trp	-	-	-
TNNT2	c.517_519del	p.Glu173del	-	-	-

Table 6 (continued).

TNNI3	c.557G>A	p.Arg186Gln	rs397516357	Missense Variant	Pathogenic/Likely pathogenic
TNNI3	c.511G>A	p.Ala171Thr	rs121917761	Missense Variant	Uncertain significance
TNNI3	c.625G>A	p.Glu209Lys	rs727504268	Missense Variant	Uncertain significance
MYH6	c.5519A>G	p.Lys1840Arg	rs373629059	Missense Variant	Uncertain significance
MYH6	c.4193G>A	p.Arg1398Gln	rs150815925	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(7); Likely benign(1)
TPM1	c.375-3C>T	p.?	rs202228866	Intron Variant	Conflicting interpretations of pathogenicity Benign(5);Likely benign(2);Uncertain significance(2)
TPM1	c.4871C>T	p.?	-	-	-
MYL2	c.58A>C	p.Met20Leu	-	-	-
MYL3	c.170C>G	p.Ala57Gly	rs139794067	Missense Variant	Conflicting interpretations of pathogenicity Pathogenic(3); Likely pathogenic(2); Uncertain significance(6)
PKP2	c.1114G>C	p.Ala372Pro	rs200586695	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Likely benign(8)
MYPN	c.3335C>T	p.Pro1112Leu	rs71534278	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(2); Benign(4); Likely benign(5)
ABCC9	c.3394_3395dupAC	p.Pro1133LeufsTer70	-	-	-
JPH2	c.692G>A	p.Arg231Gln	rs761591158	Missense Variant	Uncertain significance
CSRP3	c.10T>C	p.Trp4Arg	rs45550635	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(3); Benign(3); Likely benign(6)
CACNA1C	c.3830T>A	p.Leu1277His	-	-	-
PDLIM3	c.331-537G>T	p.?	-	-	-
PTPN11	c.824A>C	p.Asn275Thr	rs397507528	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(3); Benign(2)
NEXN	c.1419_1421delAAG	p.Arg475del	rs794729091	-	Conflicting interpretations of pathogenicity Uncertain significance(4); Likely benign(1)

Table 6 (continued).

NEXN	c.1949_1951delGAG	p.Gly650del	rs397517853; rs760927219,	-	Conflicting interpretations of pathogenicity Uncertain significance(5); Likely benign(1)
HCN4	c.3417-3425 (cCTCCCCCAGga/cga)	p.2423-2431	-	-	-
ANKRD1	c.197G>A	p.Arg66Gln	rs150797476	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(3); Benign(1); Likely benign(5)
DSC2	c.2393G>A	p.Arg798Gln	rs61731921	Missense Variant	Benign
DSC2	c.1787C>T	p.Ala596Val	rs148185335	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Benign(1); Likely benign(10)
PKP2	c.419C>T	p.Ser140Phe	rs150821281	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(4); Benign(1); Likely benign(11)
DSG2	c.877A>G	p.Ile293Val	rs2230234	Missense Variant	Benign
TRPM4	c.536G>A	p.Gly179Glu	-	-	-
HCN4	c.3317C>T	p.Pro1106Leu	-	-	-
RYR1	c.14833C>T	p.Arg4945Ter	-	-	-
DSP	c.528T>A	p.Ser176Arg	-	-	-
DSP	c.913A>T	p.Ile305Phe	rs17604693	Missense Variant	Benign/Likely benign
FBN1	c.2956G>A	p.Ala986Thr	rs112287730	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Benign(10); Likely benign(10)
CTF1	c.277G>T	p.Ala93Ser	-	-	-
NOTCH1	c.1060G>T	p.Val354Leu	-	-	-
RBMD20	c.3584C>A	p.Ser1195Tyr	rs753102653	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(4); Likely benign(2)
DMD	c.8762A>G	p.His2921Arg	rs1800279	Missense Variant	Benign/Likely benign
TRPM4	c.3611C>T	p.Pro1204Leu	rs150391806	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Benign(4); Likely benign(3)
TRPM4	c.913G>A	p.Gly305Arg	-	-	-
TRPM4	c.2158G>A	p.Glu720Lys	-	-	-

Table 6 (continued).

SDHA	c.1367C>T	p.Ser456Leu	rs76896145	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Benign(1)
ELN	c.1828G>C	p.Gly610Arg	-	-	-
ELN	c.1363G>A	p.Ala455Thr	-	-	-
NKX2-5	c.73C>T	p.Arg25Cys	rs28936670	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Benign(4); Likely benign(2)
LAMA4	c.1436A>G	p.Asp479Gly	-	-	-
FBN1	c.1081G>T	p.Asp361Tyr	-	-	-
TTN	c.16546G>T	p.Asp5516Tyr	rs72648940	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(8); Benign(2); Likely benign(8)
TTN	c.17721G>T	p.Lys5907Asn	-	-	-
TTN	c.30274C>T	p.His10092Tyr	rs72650011	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(1); Benign(7); Likely benign(5)
TTN	c.43690T>A	p.Ser14564Thr	rs181189778	Missense Variant	Conflicting interpretations of pathogenicity Uncertain significance(6); Benign(2); Likely benign(5)

4.1.2. Identification of HCM subphenotypes using clustering

Clusters of HCM patients were determined based on phenotypic data.

4.1.2.1. Hierarchical clustering

As a result of hierarchical cluster analysis, a dendrogram has been generated (a tree diagram, wherein the most similar HCM patients are placed on branches that are close together) (Figure 4).

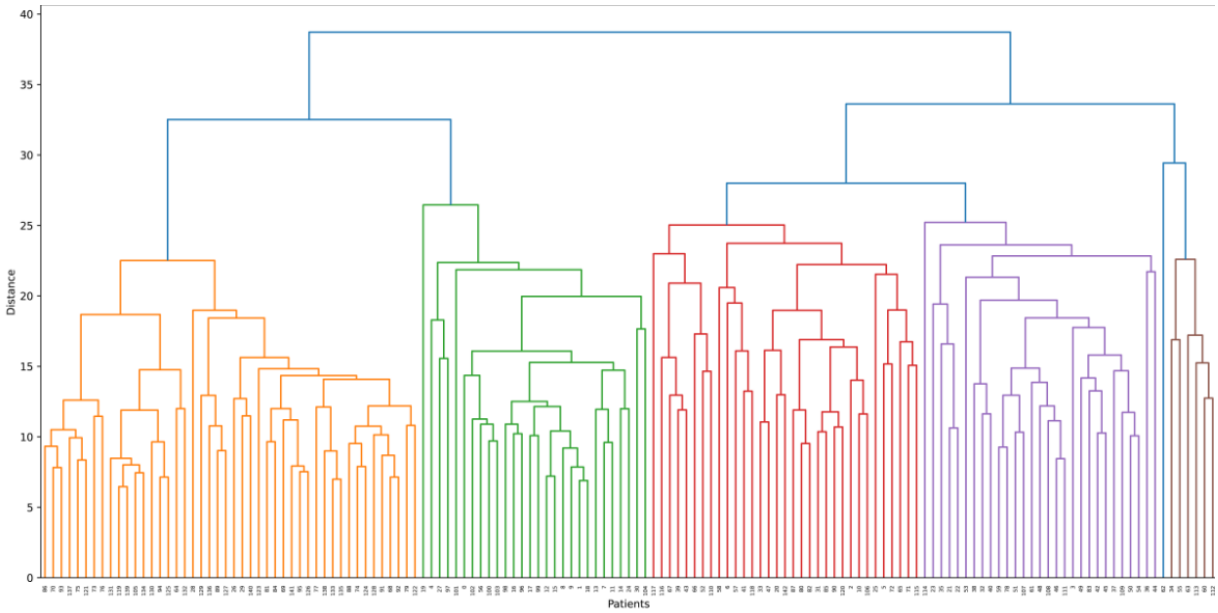

Figure 4. Dendrogram — result of hierarchical cluster analysis

A set of clusters is derived from a dendrogram, by cutting at varied horizontal levels. In this case, the most prominent aggregations happen, if we divide all patients into 2, 4, or 6 clusters.

An approximate interpretation of clustering-logic is presented in the form of a visualized decision tree (Figures 5-7). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by hierarchical clustering (Tables 7-9).

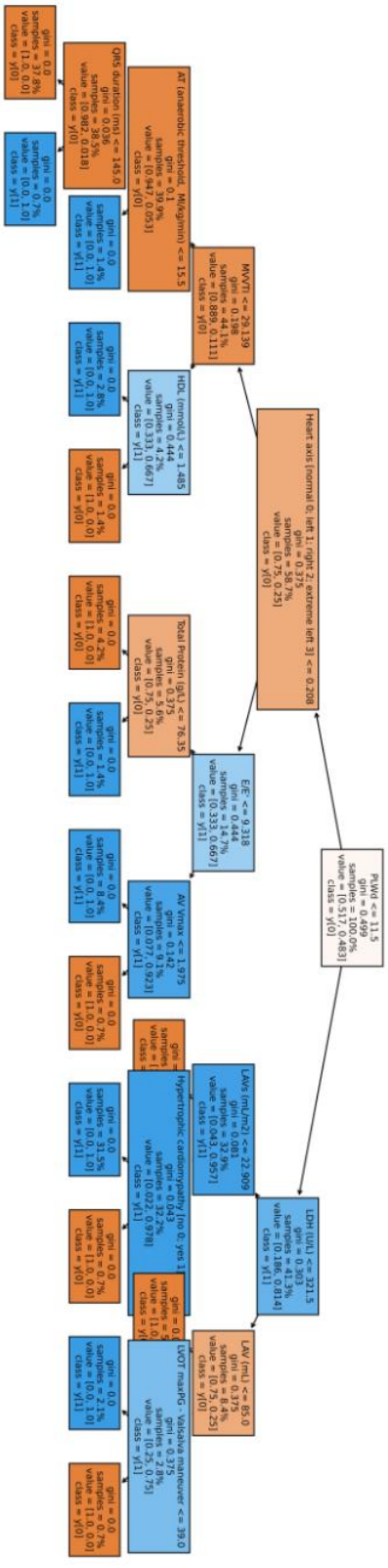

Figure 5. An approximate interpretation of clustering-logic for 2 clusters

Table 7. Feature importance — Top 35 features for distinguishing 2 clusters

Feature	Estimated importance
LVOT Vmax	0.059688
MVVTI	0.055190
PLWd	0.053490
Albumin	0.041790
LDH	0.040685
E/E'	0.039324
MV maxPG	0.034698
Heart axis [normal/left/right/extreme left]	0.028349
LAV	0.023633
AO	0.019106
Random glucose	0.018822
QRS duration	0.018626
LAVs	0.018459
Weight	0.017638
RAVs	0.017498
LA	0.016575
EFLV	0.016143
ESVLV	0.014526
AV Vmax	0.014466
AT (anaerobic threshold)	0.014386
AV meanPG	0.013739
AV maxPG	0.012468
Sokolow index	0.012265
Peak heart rate	0.011912
Triglycerides	0.011650
HDL	0.011483
Peak VE/VCO_2	0.011092
Age	0.010751
Calcium	0.010320
TAPSE	0.010200
MV meanPG	0.009797
Creatine-kinase	0.009785
AOvs	0.009746
AscAO (mm)	0.009344
Diastolic	0.009286

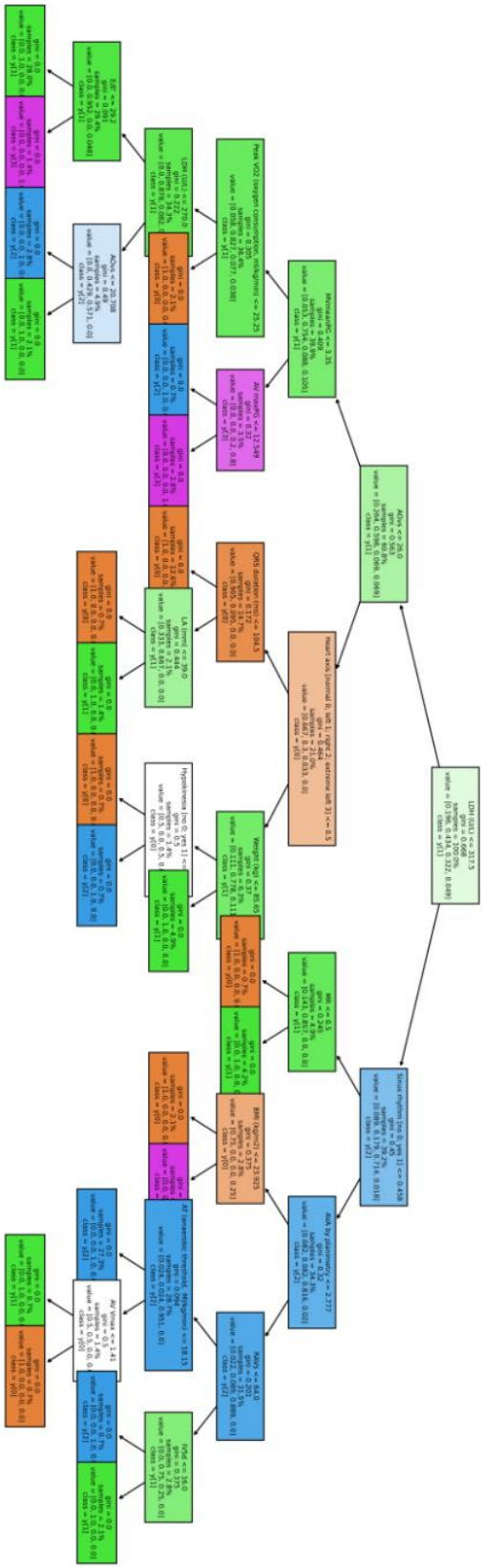

Figure 6. An approximate interpretation of clustering-logic for 4 clusters

Table 8. Feature importance — Top 40 features for distinguishing 4 clusters

Feature	Estimated importance
LDH	0.061884
AO (mm)	0.055299
MV meanPG	0.046228
AOvs	0.033707
PLWd	0.029291
E/E'	0.028763
MVVTI	0.027145
MV maxPG	0.026808
AscAO	0.024202
Peak VE/VCO_2	0.022777
RAVs	0.022454
LVOT Vmax	0.020590
AVVTI	0.019889
HDL	0.019392
Albumin	0.019389
LAVs	0.016331
Calcium	0.015438
QRS duration	0.014732
Weight	0.014536
Heart murmur [no/yes]	0.014079
LAV	0.013832
NYHA class	0.012940
AV Vmax	0.012782
ESVLV	0.012625
Random glucose	0.011767
AVA by planimetry	0.011489
LA	0.010609
AV meanPG	0.010592
EFLV	0.010500
HCM in family history [no/yes]	0.010493
Heart axis [normal/left/right/extreme left]	0.010487
Diastolic dysfunction grade	0.010210
AV maxPG	0.010105
Peak heart rate	0.009650
LVOT maxPG - Valsalva maneuver	0.009561
Sokolow index	0.009249
NT-BNP	0.009109
RVSP	0.009008
Heart rate	0.008841
LVIDd	0.008485

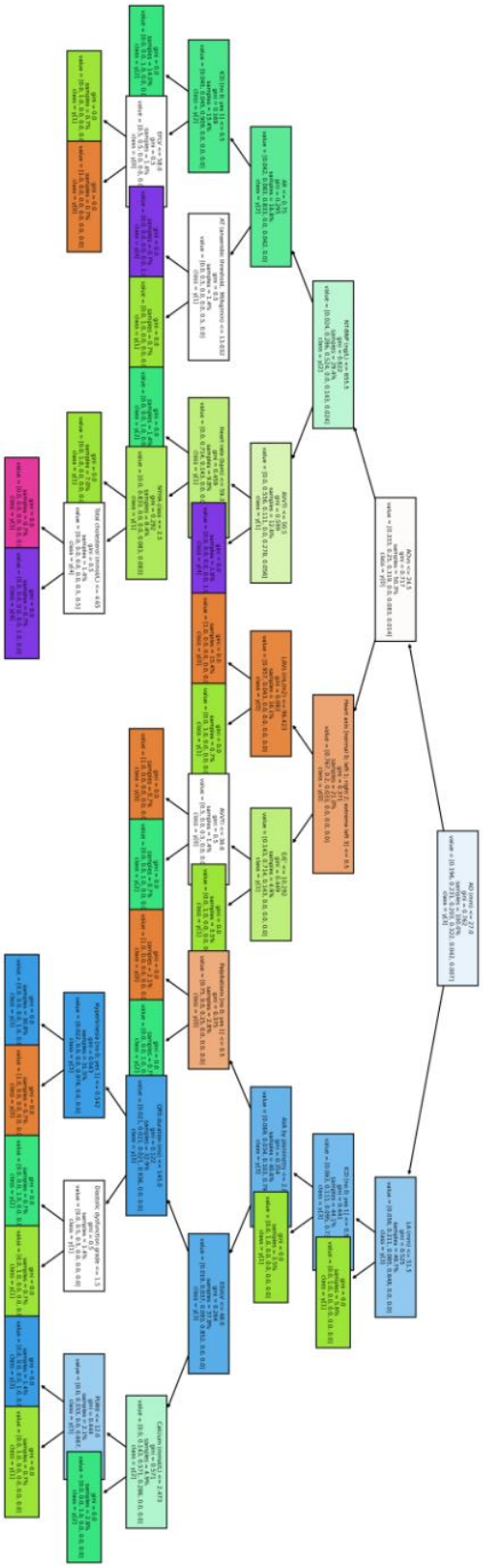

Figure 7. An approximate interpretation of clustering-logic for 6 clusters

Table 9. Feature importance — Top 45 features for distinguishing 6 clusters

Feature	Estimated importance
LDH	0.056124
AO	0.054735
MV meanPG	0.037623
AOvs	0.032800
Peak VE/VCO_2	0.029921
MVVTI	0.027920
PLWd	0.026490
LVOT Vmax	0.022202
MV maxPG	0.021949
E/E'	0.021736
NT-BNP	0.019012
LAV	0.018839
RAVs	0.018373
LA	0.017998
Albumin	0.016098
Heart murmur [yes/no]	0.015882
AscAO	0.015036
AVVTI	0.014664
Heart axis [normal/left/right/extreme left]	0.014625
Diastolic dysfunction grade	0.013922
LAVs	0.013861
Hypertrophic cardiomypathy	0.013206
AVA by planimetry	0.012182
Troponin	0.012135
HDL	0.011520
Heart rate	0.011465
Height	0.011373
EFLV	0.010996
NYHA class	0.010971
Sinus rhythm [yes/no]	0.010752
Calcium	0.010746
Weight	0.010435
LDL	0.010257
PR interval	0.009927
Sokolow index	0.009848
AV maxPG	0.009406
Random glucose	0.009334
ICD [no/yes]	0.009256
IVSd	0.009047
Total cholesterol	0.008985
Atrial fibrillation [no/yes]	0.008854
AV meanPG	0.008706
Peak heart rate	0.008595
Triglycerides	0.008520
ESVLV	0.008393

4.1.2.2. K-Prototype clustering

The optimal number of clusters in this dataset was found to be 4, which was determined using Elbow method (283). However, to systematically inspect the dataset, sets of 2, 3, 5, 6, and 7 clusters were also investigated in greater detail. None of the patients presented pleural effusion, venous congestion, phaeochromocytoma, acromegaly, amyloidosis, dyskinesia, and AV block III.

An approximate interpretation of the clustering-logic is presented in the form of a visualized decision tree. Feature importance of random forest trained on the same dataset was determined, with labels as assigned by K-Prototype clustering.

4.1.2.2.1. Two clusters

Characteristics of the 2 clusters determined are shown in Figures 8-25.

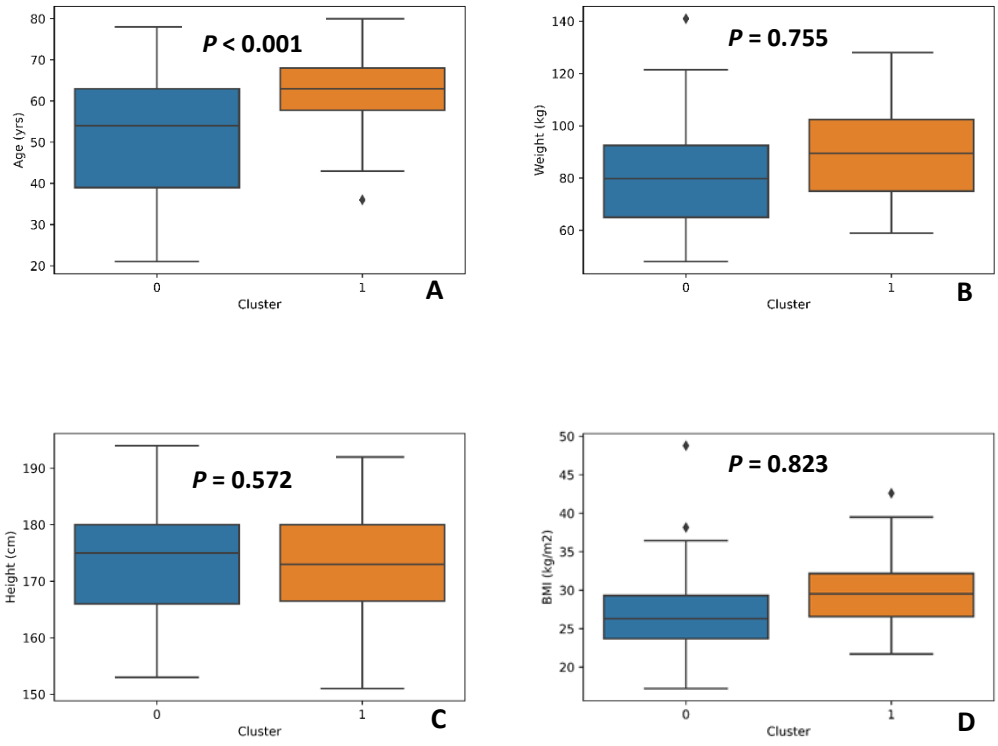

Figure 8. Two clusters setting, features: A — Age, B — Weight, C — Height, D — BMI

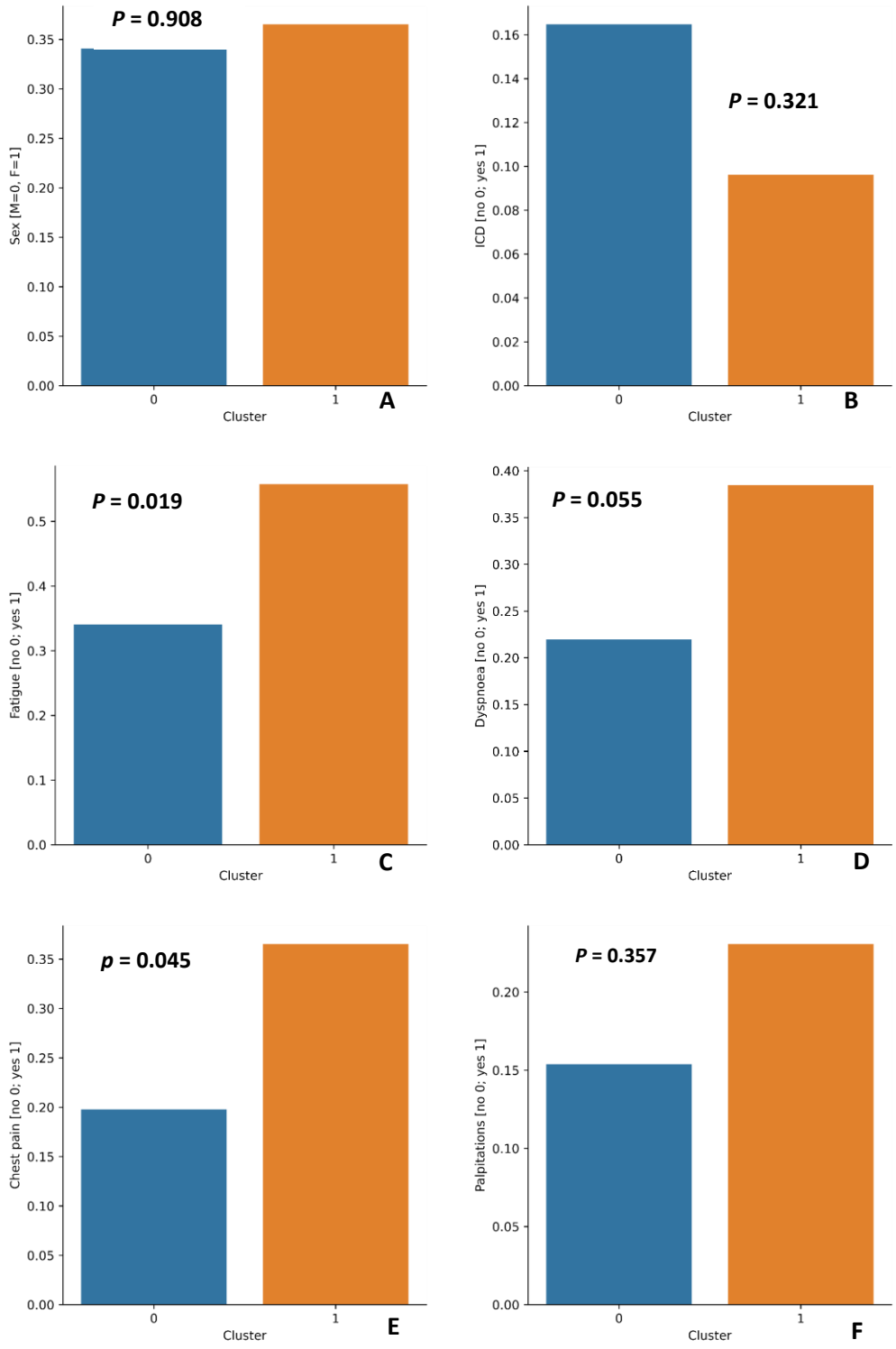

Figure 9. Two clusters setting, features: A — Sex, B — ICD, C — Fatigue, D — Dyspnea, E — Chest pain, F — Palpitations

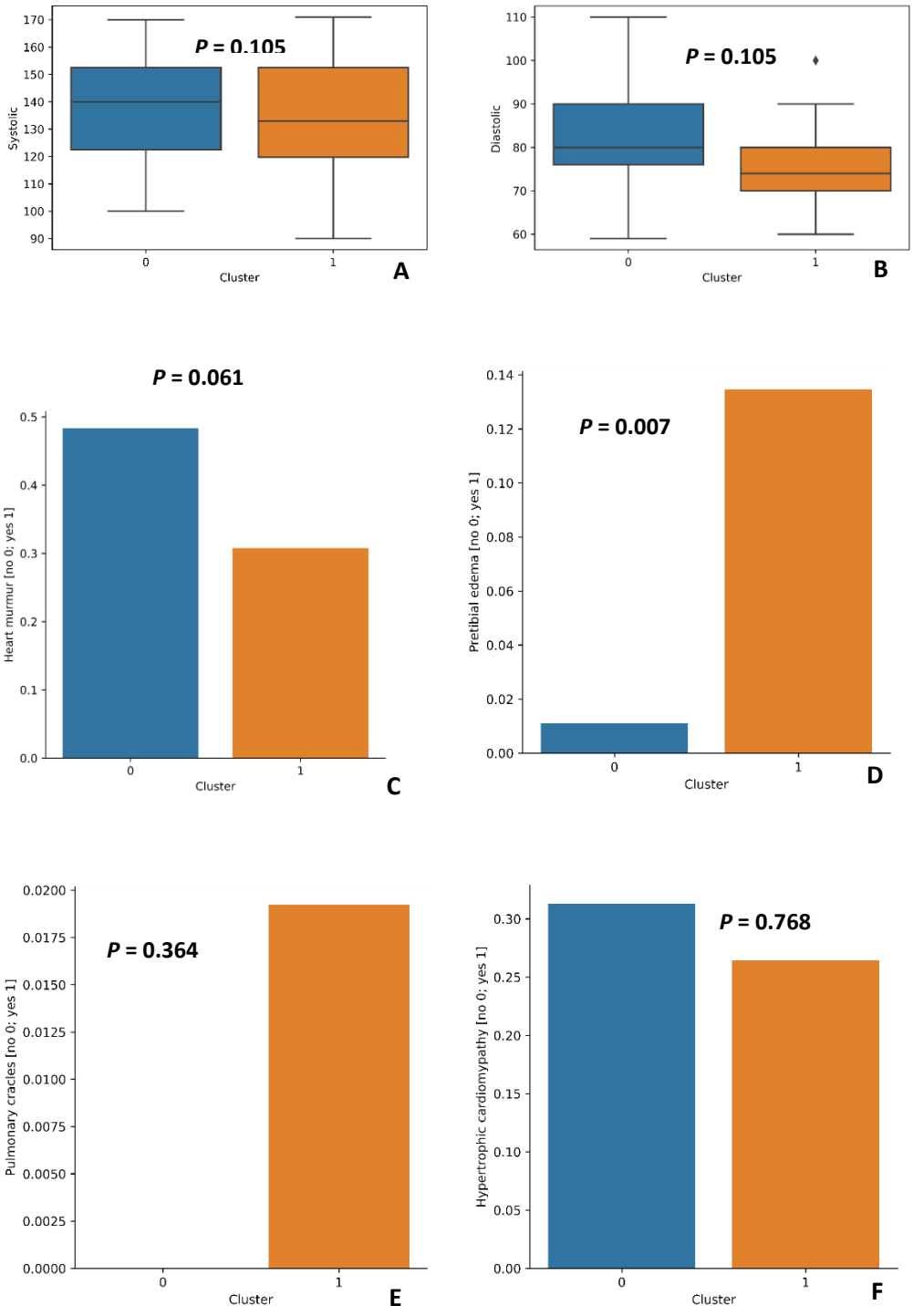

Figure 10. Two clusters setting, features: A — Systolic blood pressure, B — Diastolic blood pressure, C — Heart murmur, D — Pretibial edema, E — Pulmonary crackles, F — HCM in family history

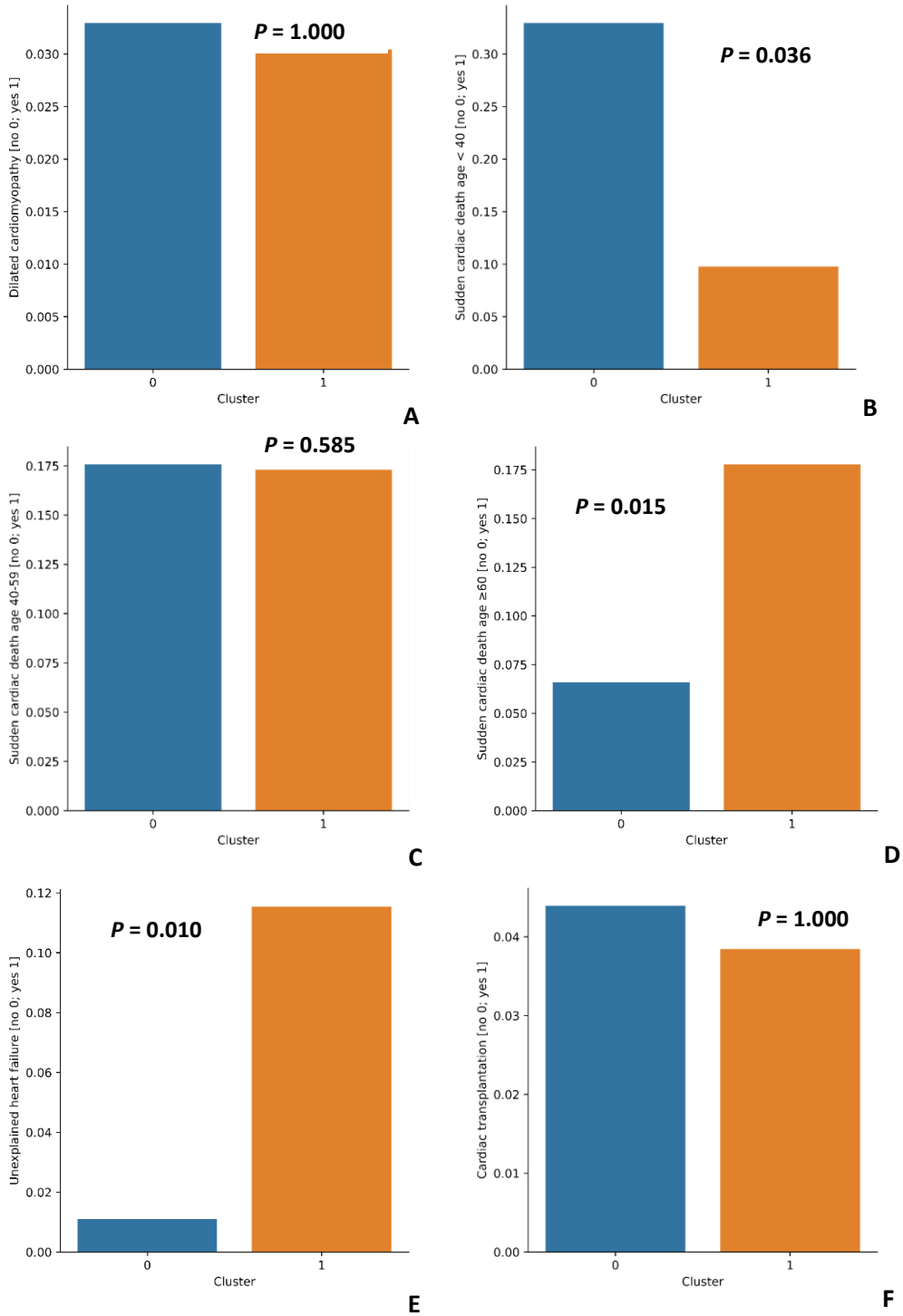

Figure 11. Two clusters setting, features: A — DCM in family history, B — SCD in age < 40 in family history, C — SCD in age 40-59 in family history, D — SCD in age ≥ 60 in family history, E — Unexplained HF in family history, F — Cardiac transplantation in family history

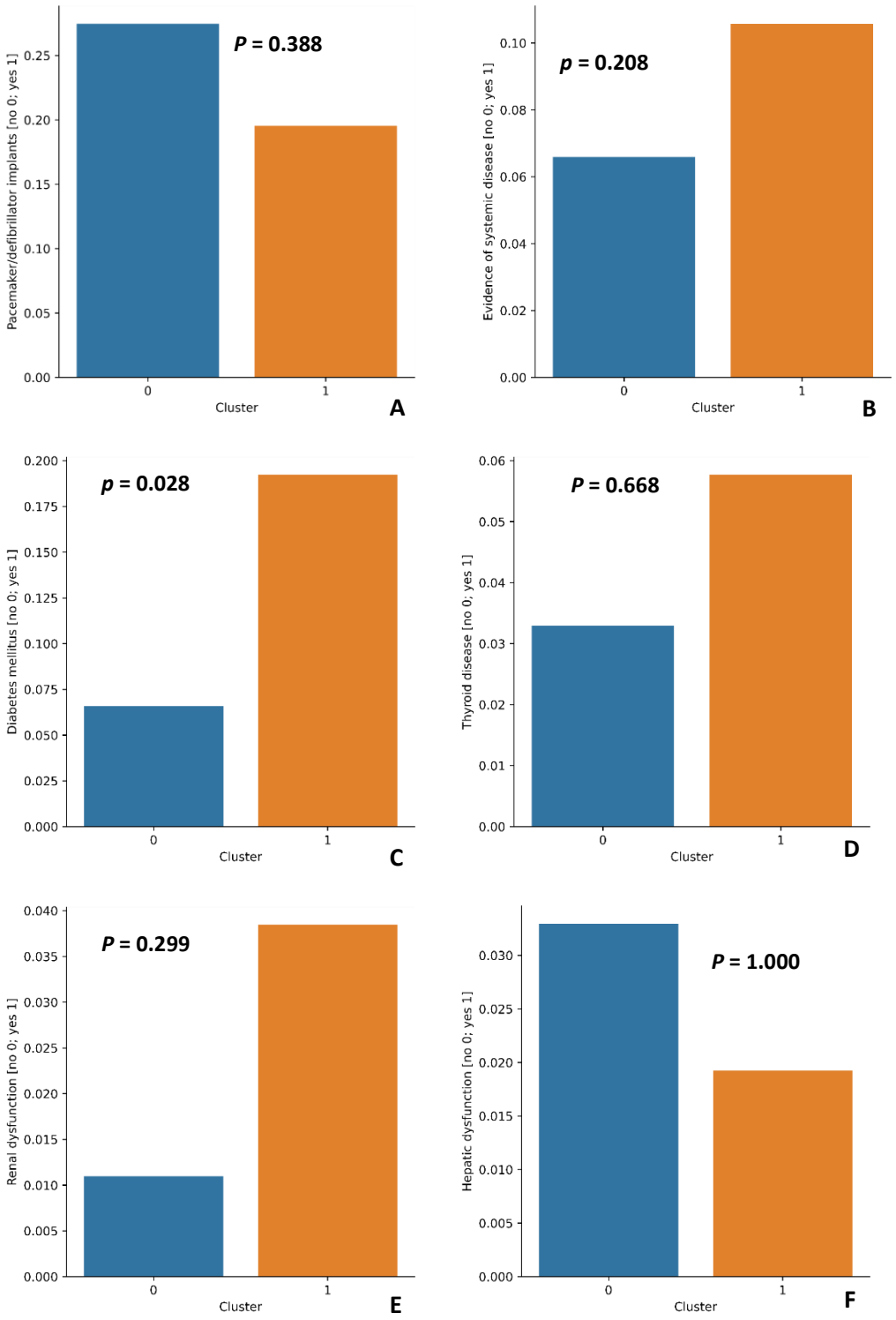

Figure 12. Two clusters setting, features: A — Pacemaker/defibrillator implants in family history, B — Evidence of systemic disease in family history, C — Diabetes mellitus, D — Thyroid disease, E — Renal dysfunction, F — Hepatic dysfunction

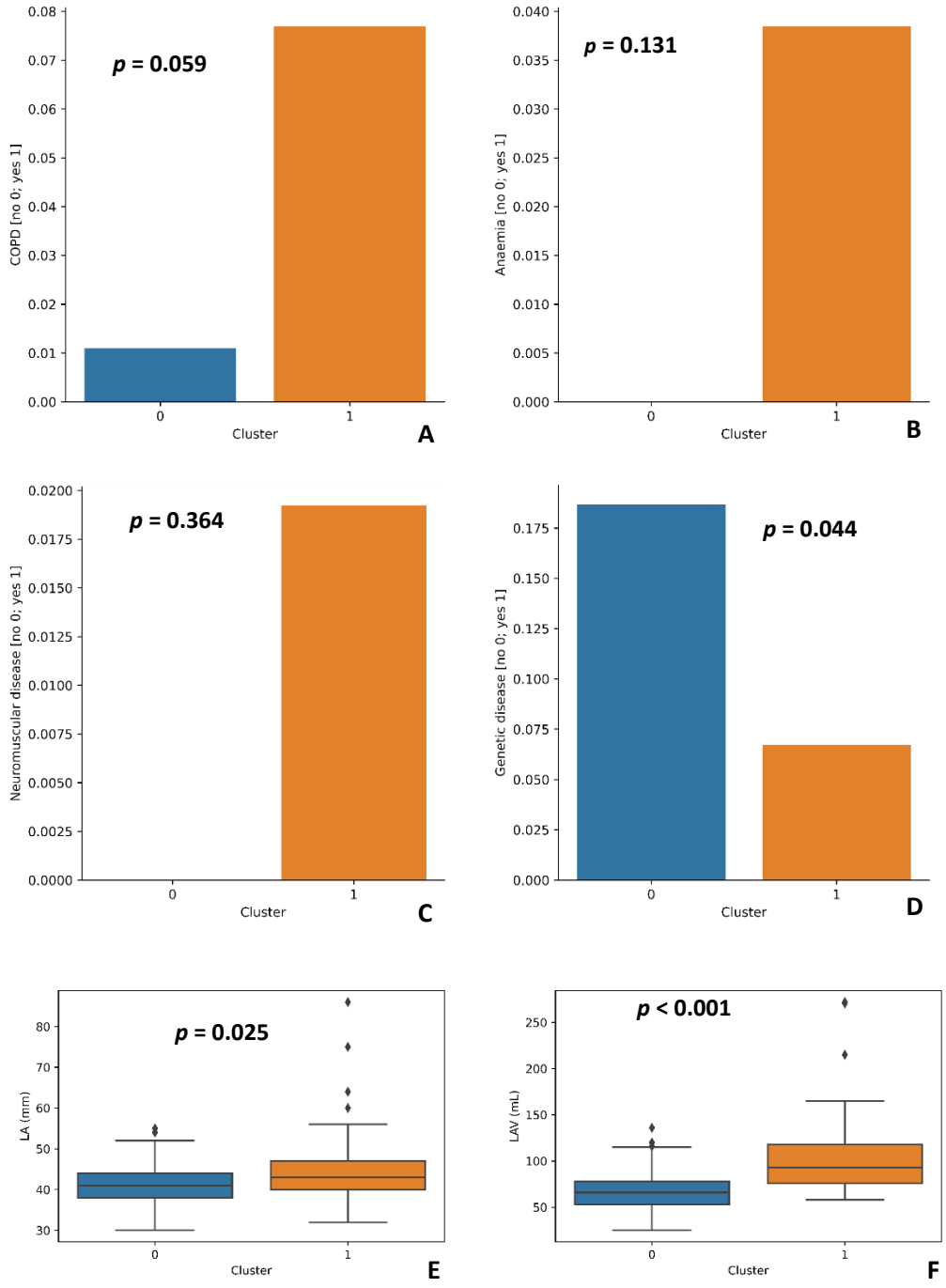

Figure 13. Two clusters setting, features: A — COPD, B — Anemia, C — Neuromuscular disease, D — Genetic disease, E — LA, F — LAV

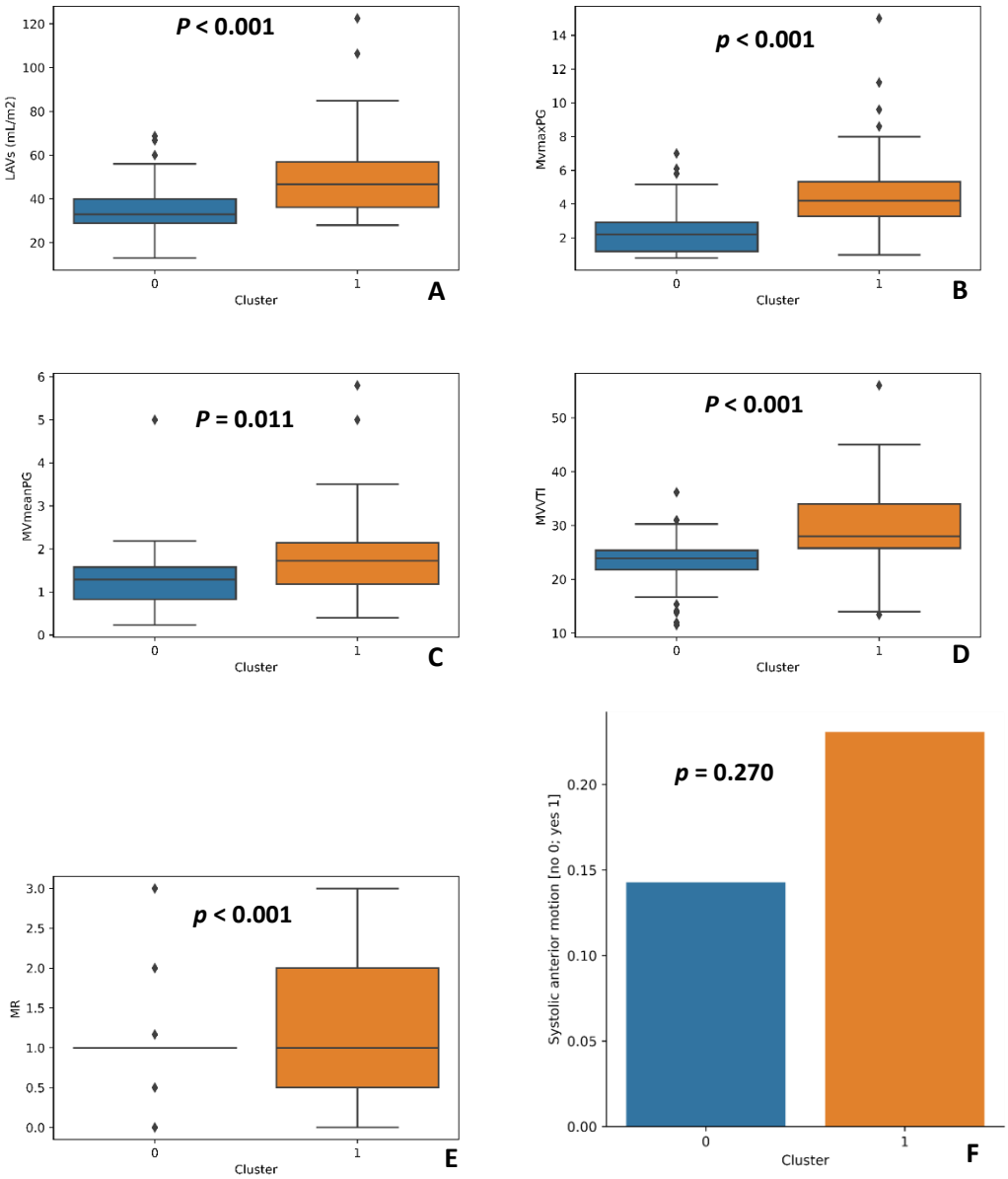

Figure 14. Two clusters setting, features: A — LAVs, B — MV maxPG, C — MV meanPG, D — MVVTI, E — MR, F — SAM

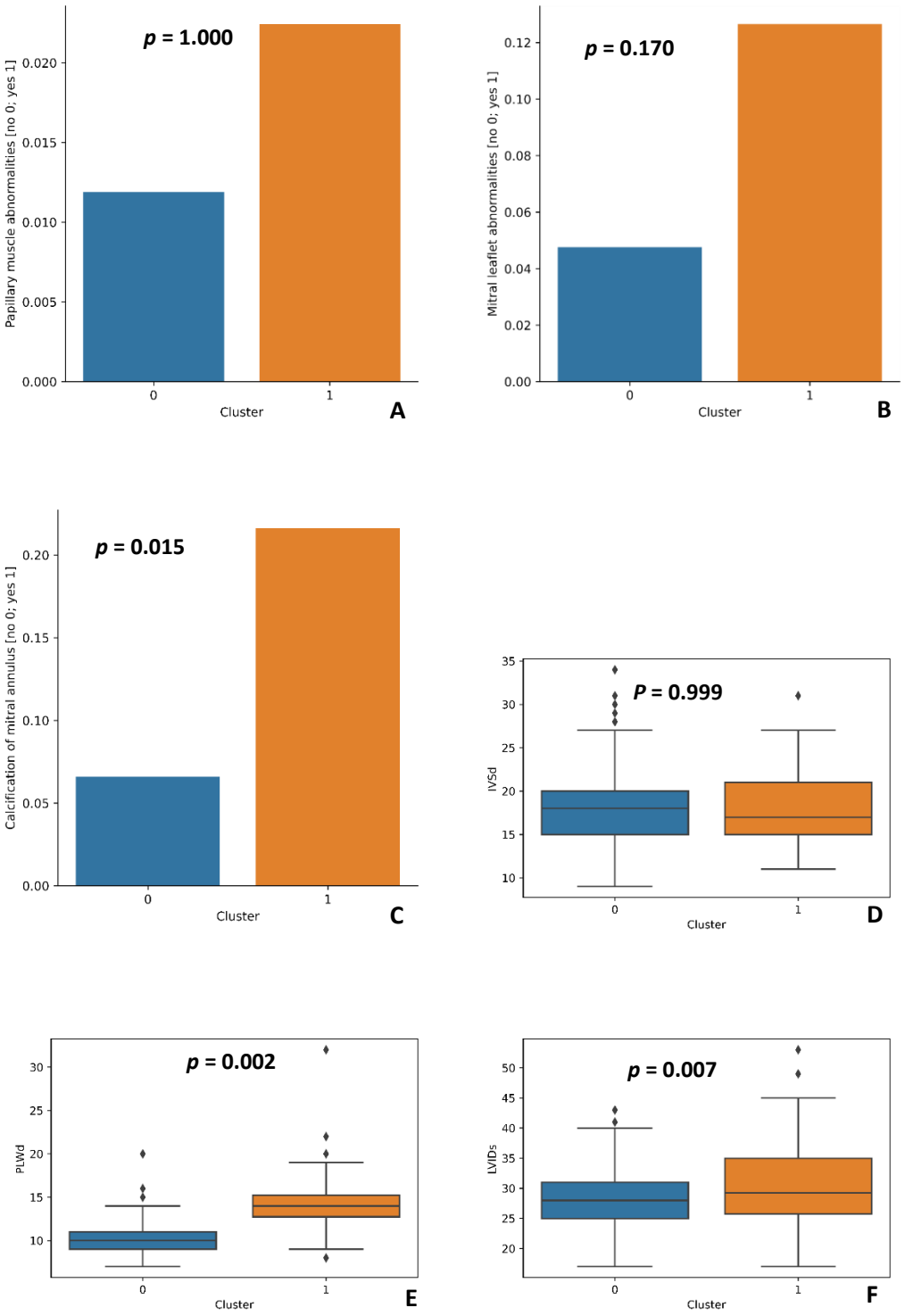

Figure 15. Two clusters setting, features: A — Papillary muscle abnormalities, B — Mitral leaflet abnormalities, C — Calcification of mitral annulus, D — IVSd, E — PLWd, F — LVIDs

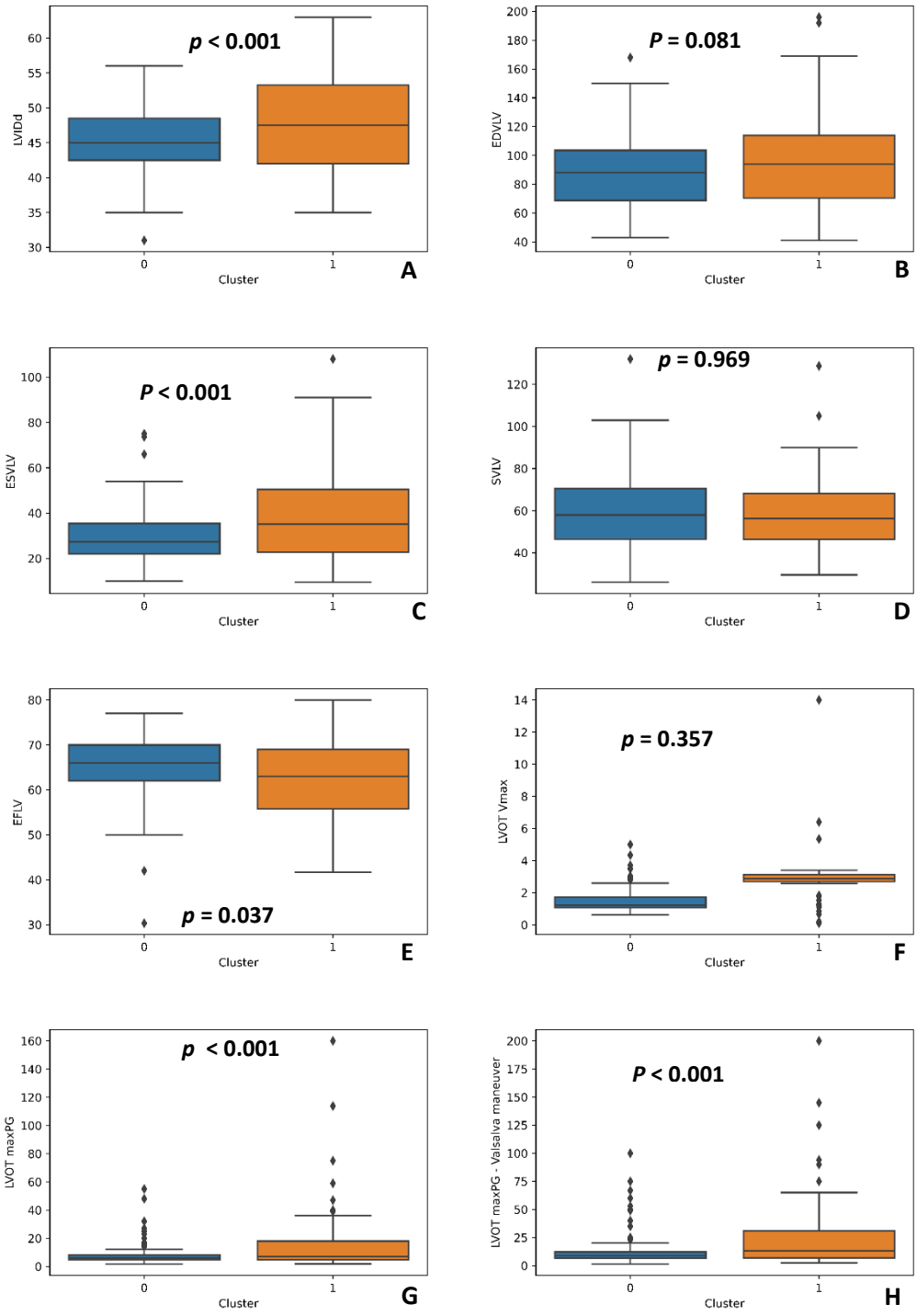

Figure 16. Two clusters setting, features: A — LVIDd, B — EDVLV, C — ESVLV, D — SVLV, E — EFLV, F — LVOT Vmax, G — LVOT maxPG, H — LVOT maxPG — Valsalva maneuver

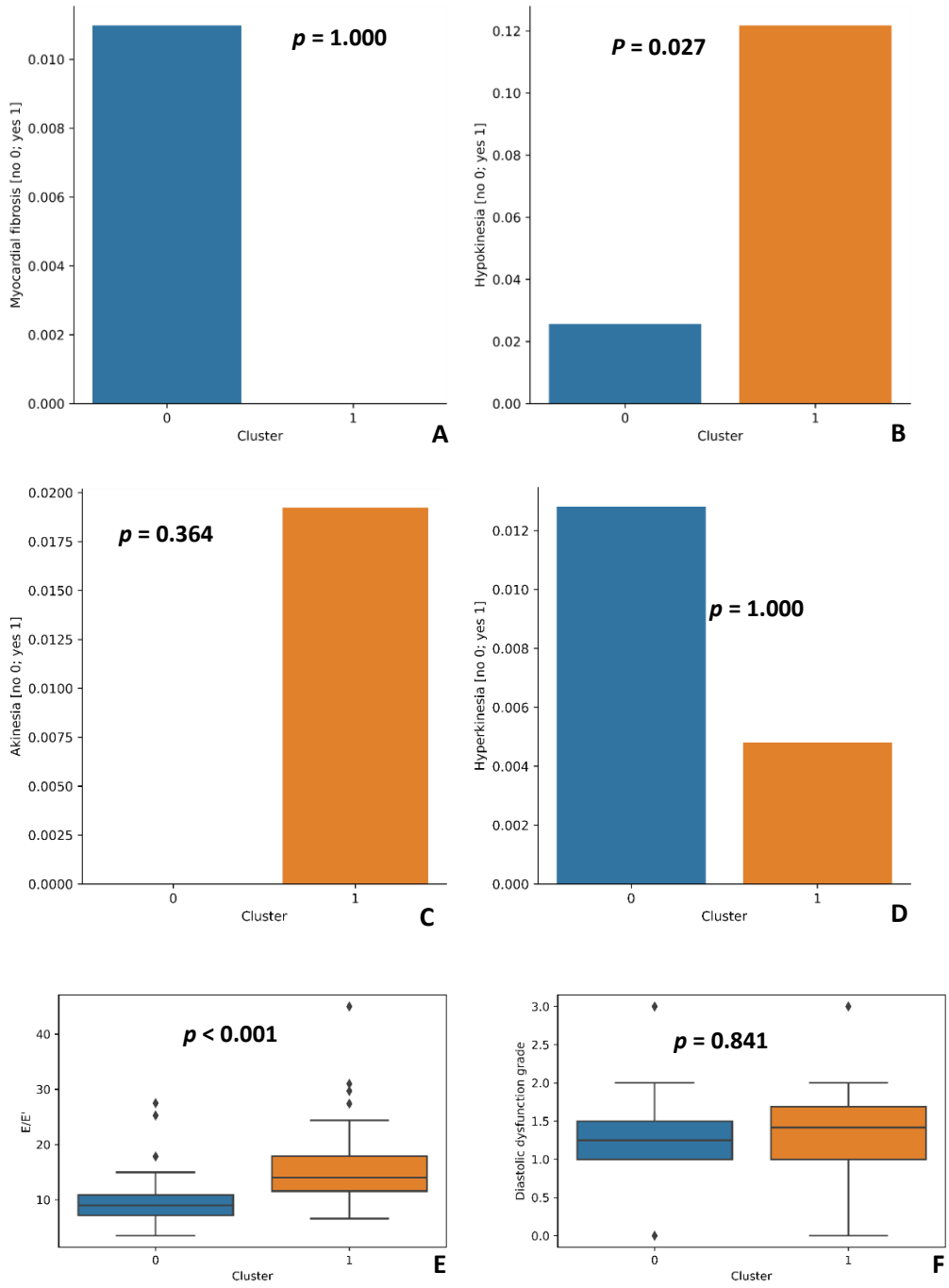

Figure 17. Two clusters setting, features: A — Myocardial fibrosis, B — Hypokinesia, C — Akinesia, D — Hyperkinesia, E — E/E', F — Diastolic dysfunction grade

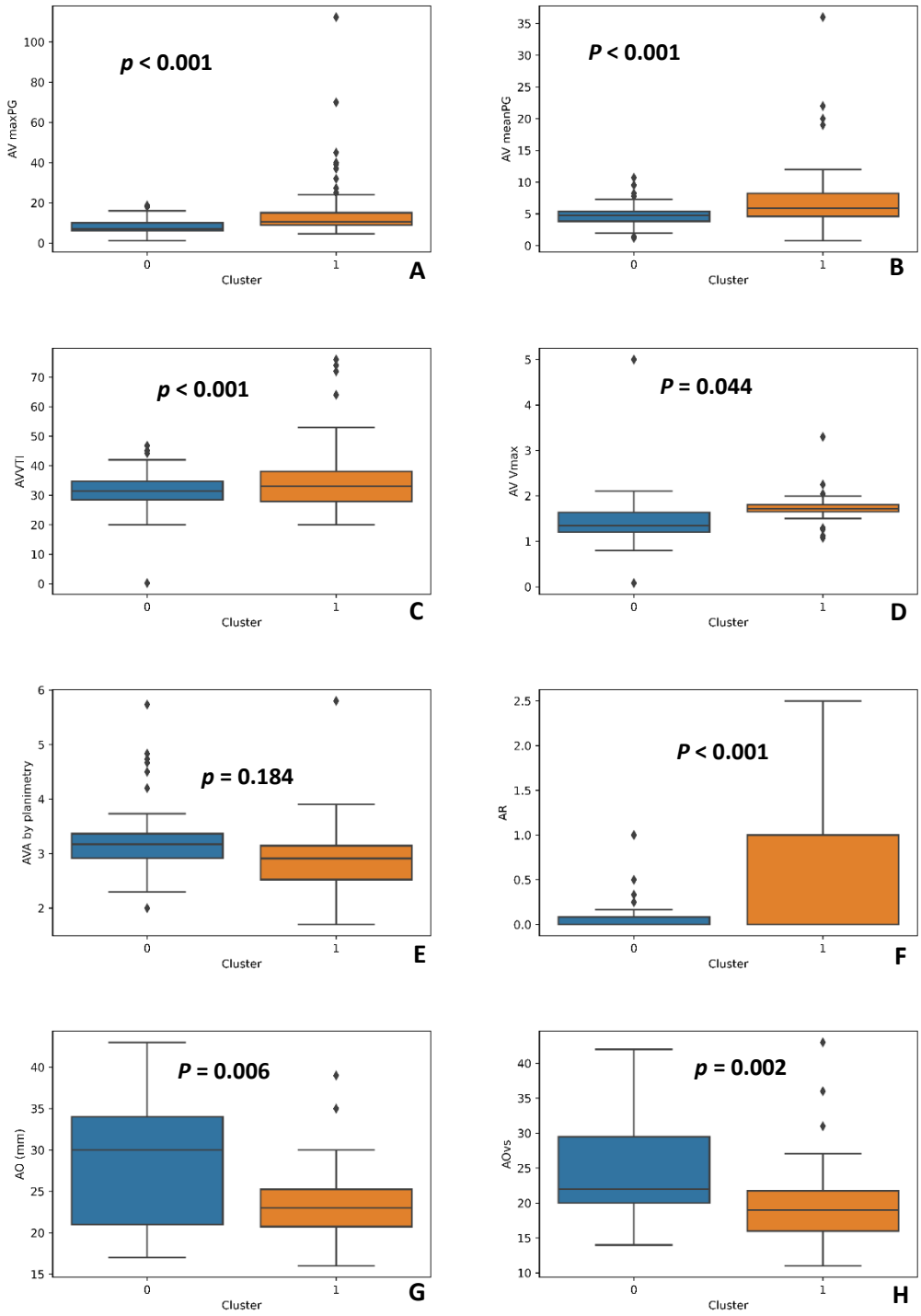

Figure 18. Two clusters setting, features: A — AV maxPG, B — AV meanPG, C — AVVTI, D — AV Vmax, E — AVA by planimetry, F — AR, G — AO, H — AOvs

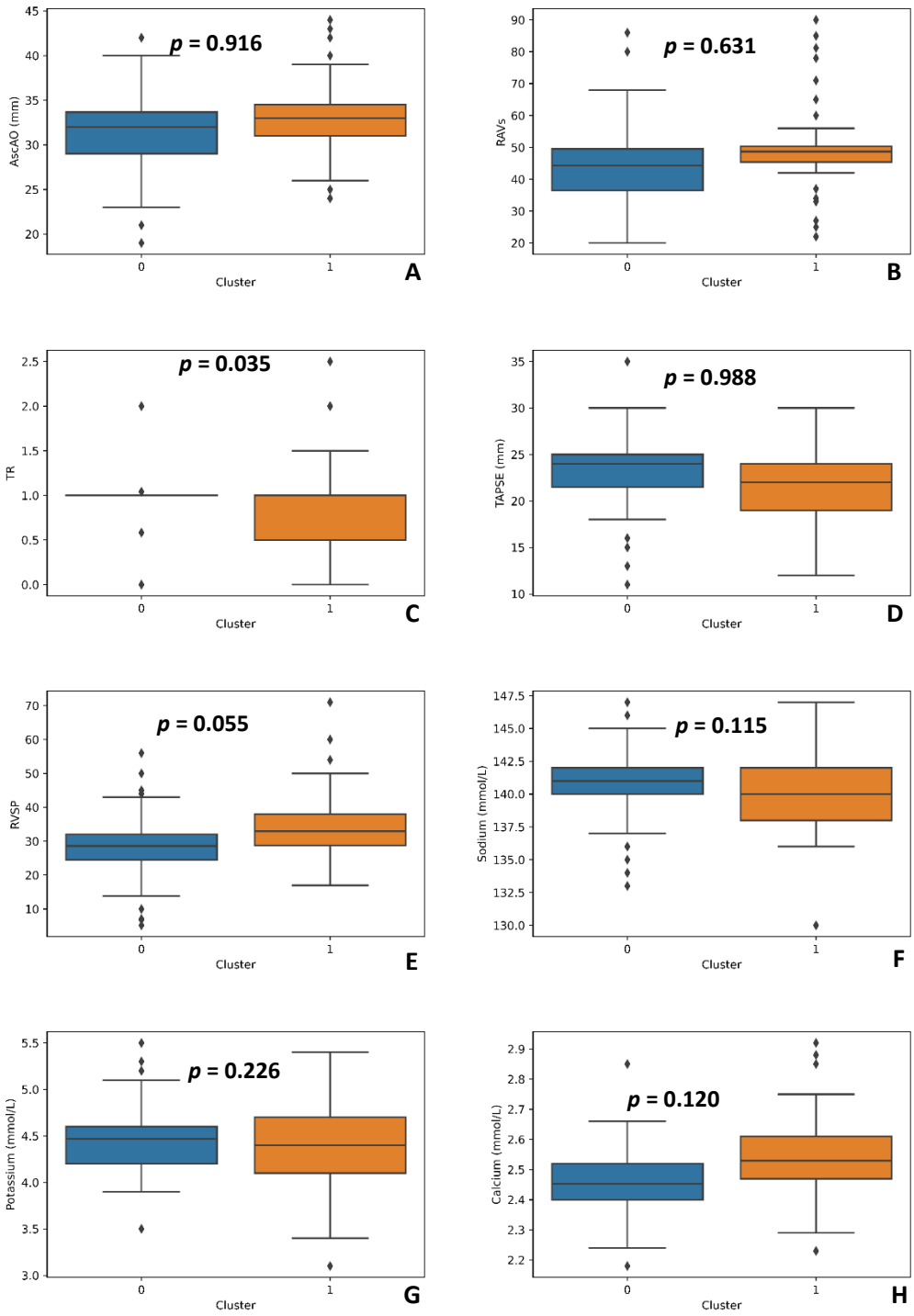

Figure 19. Two clusters setting, features: A — AscAO, B — RAVs, C — TR, D — TAPSE, E — RVSP, F — Sodium, G — Potassium, H — Calcium

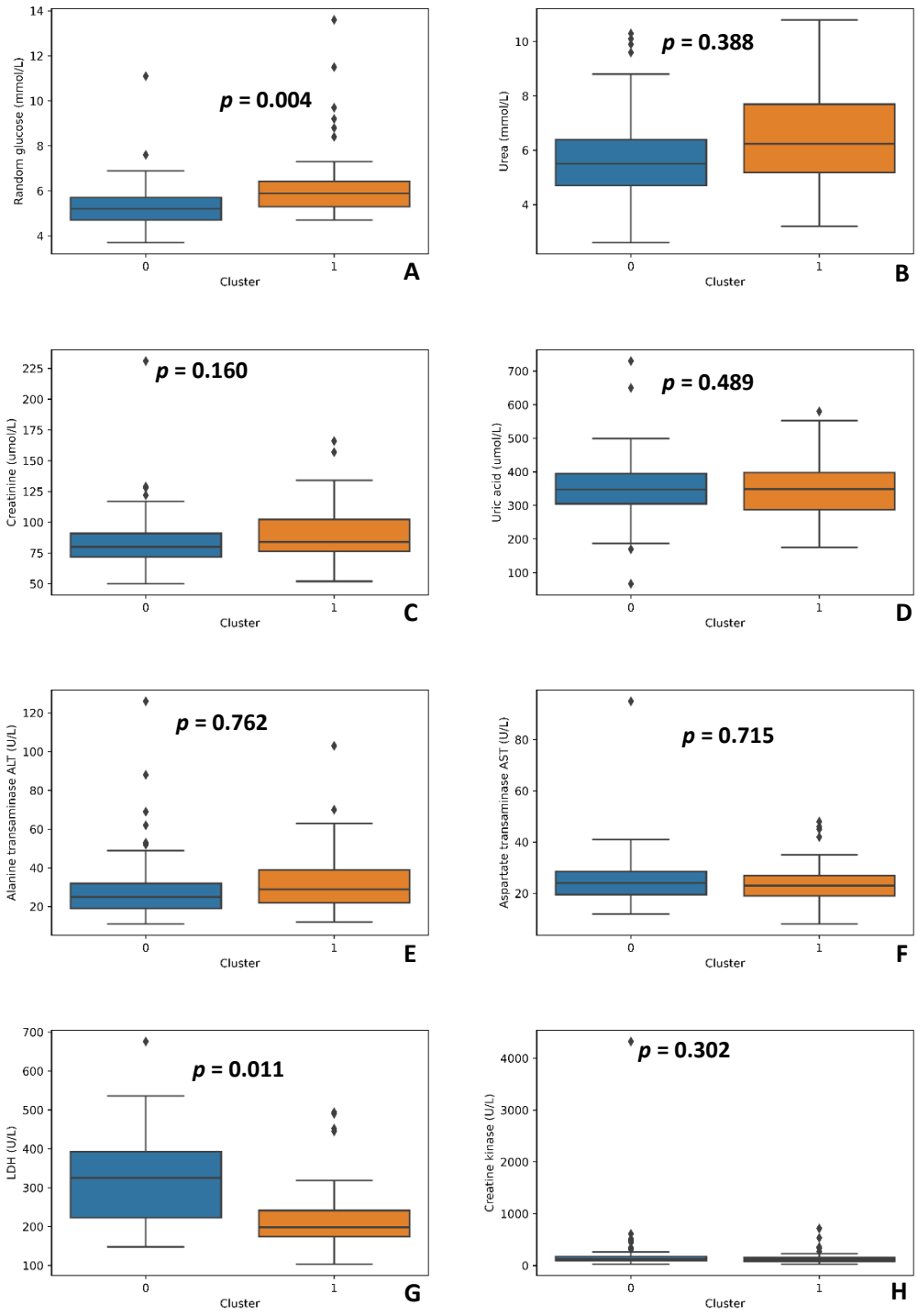

Figure 20. Two clusters setting, features: A — Random glucose, B — Urea, C — Creatinine, D — Uric acid, E — ALT, F — AST, G — LDH, H — Creatine-kinase

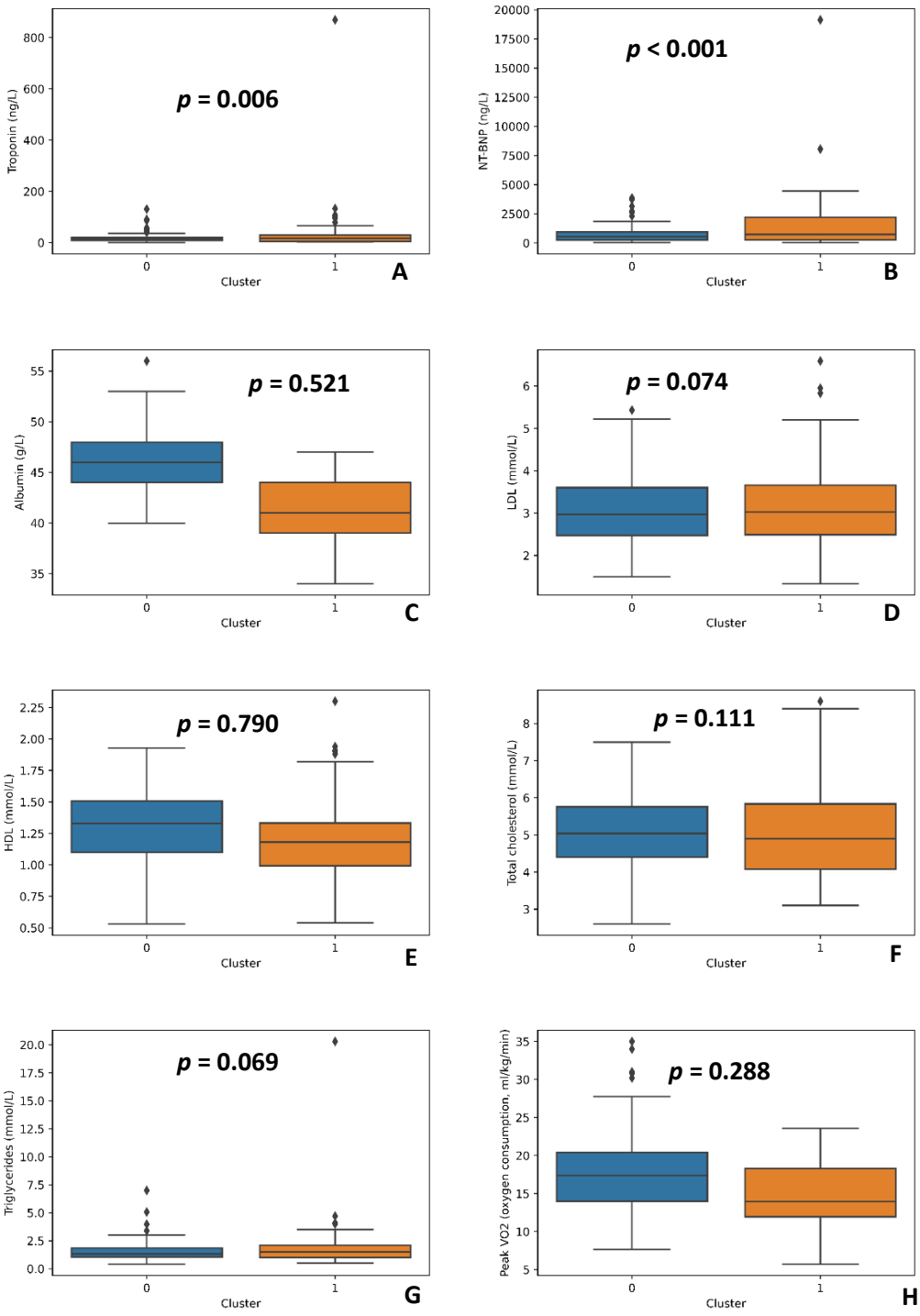

Figure 21. Two clusters setting, features: A — Troponin, B — NT-BNP, C — Albumin, D — LDL, E — HDL, F — Total cholesterol, G — Triglyceride, H — Peak VO_2

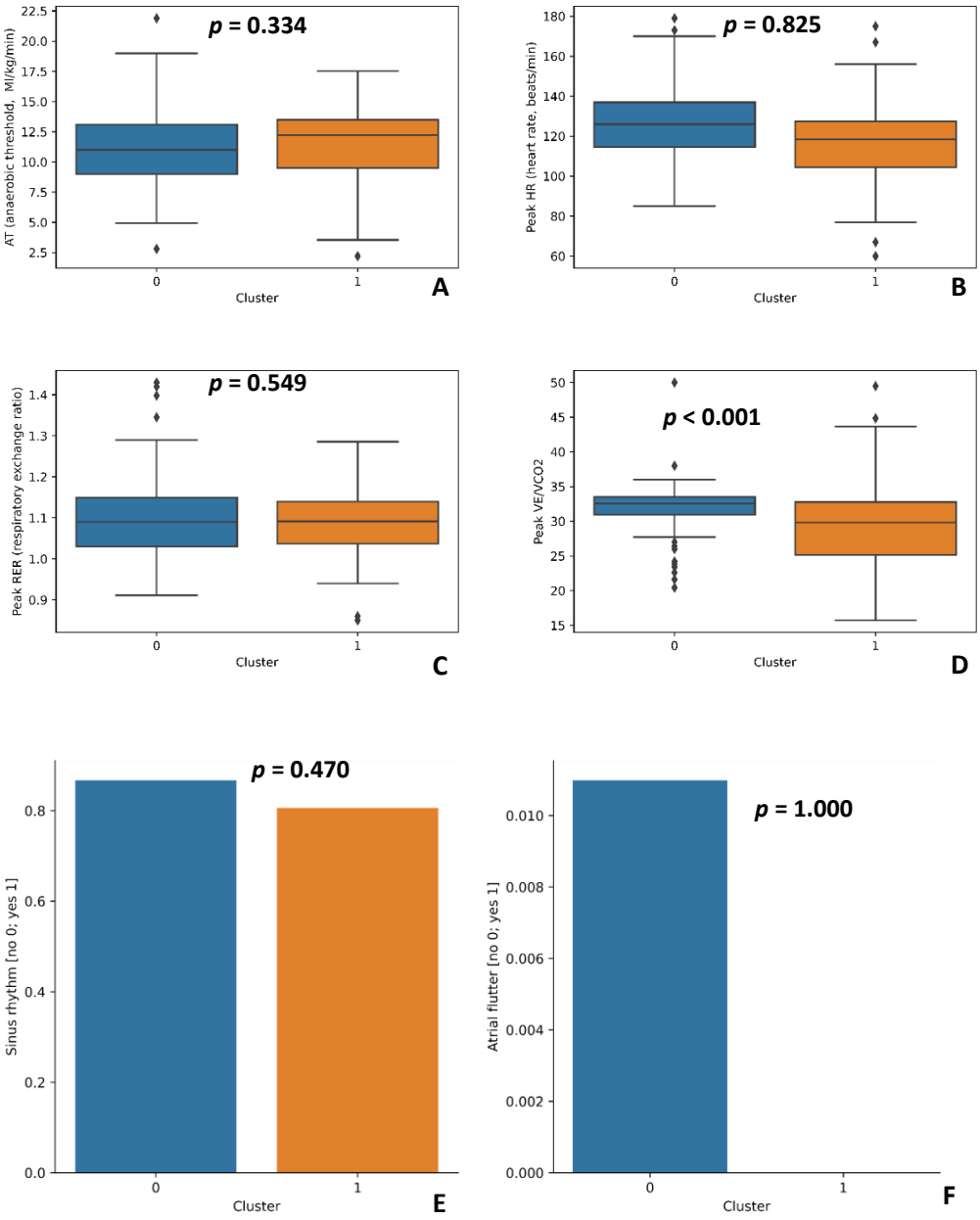

Figure 22. Two clusters setting, features: A — AT, B — Peak HR, C — Peak RER, D — Peak VE/VCO_2, E — Sinus rhythm, F — Atrial flutter

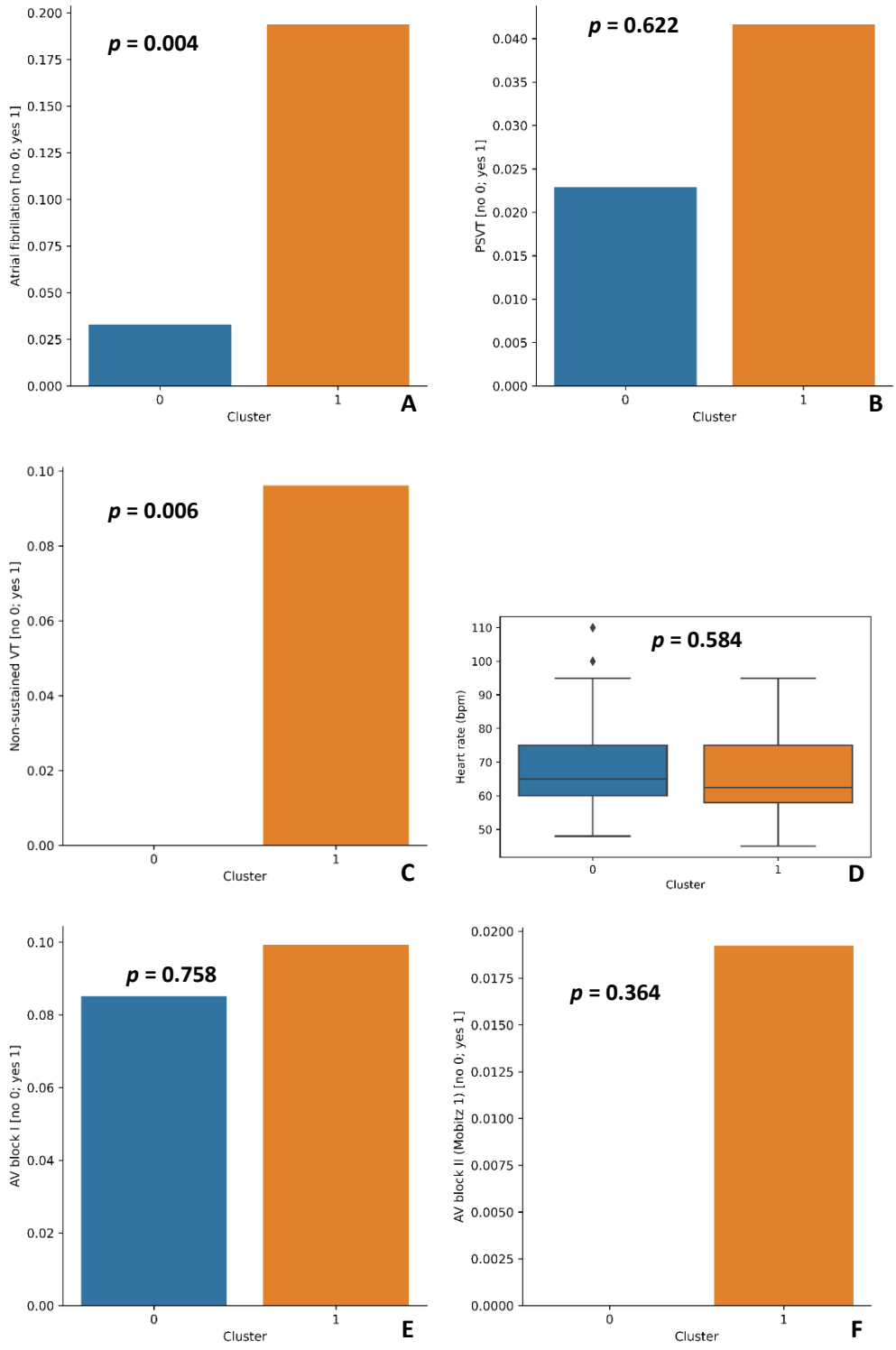

Figure 23. Two clusters setting, features: A — AF, B — PSVT, C — Non-sustained VT, D — Heart rate, E — AV block I, F — AV block II (Mobitz 1)

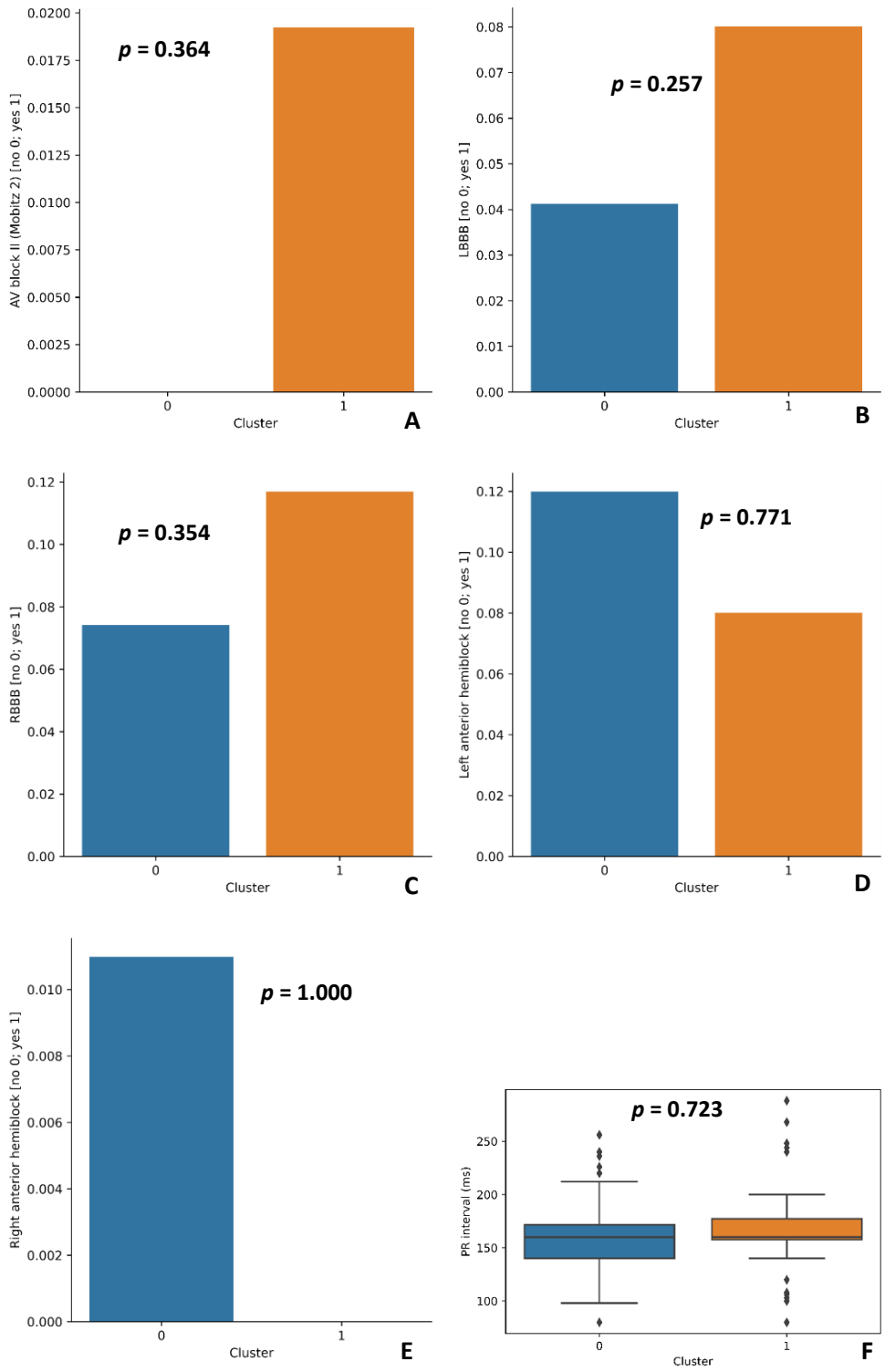

Figure 24. Two clusters setting, features: A — AV block II (Mobitz 2), B — LBBB, C — RBBB, D — Left anterior hemiblock, E — Right anterior hemiblock, F — PR interval

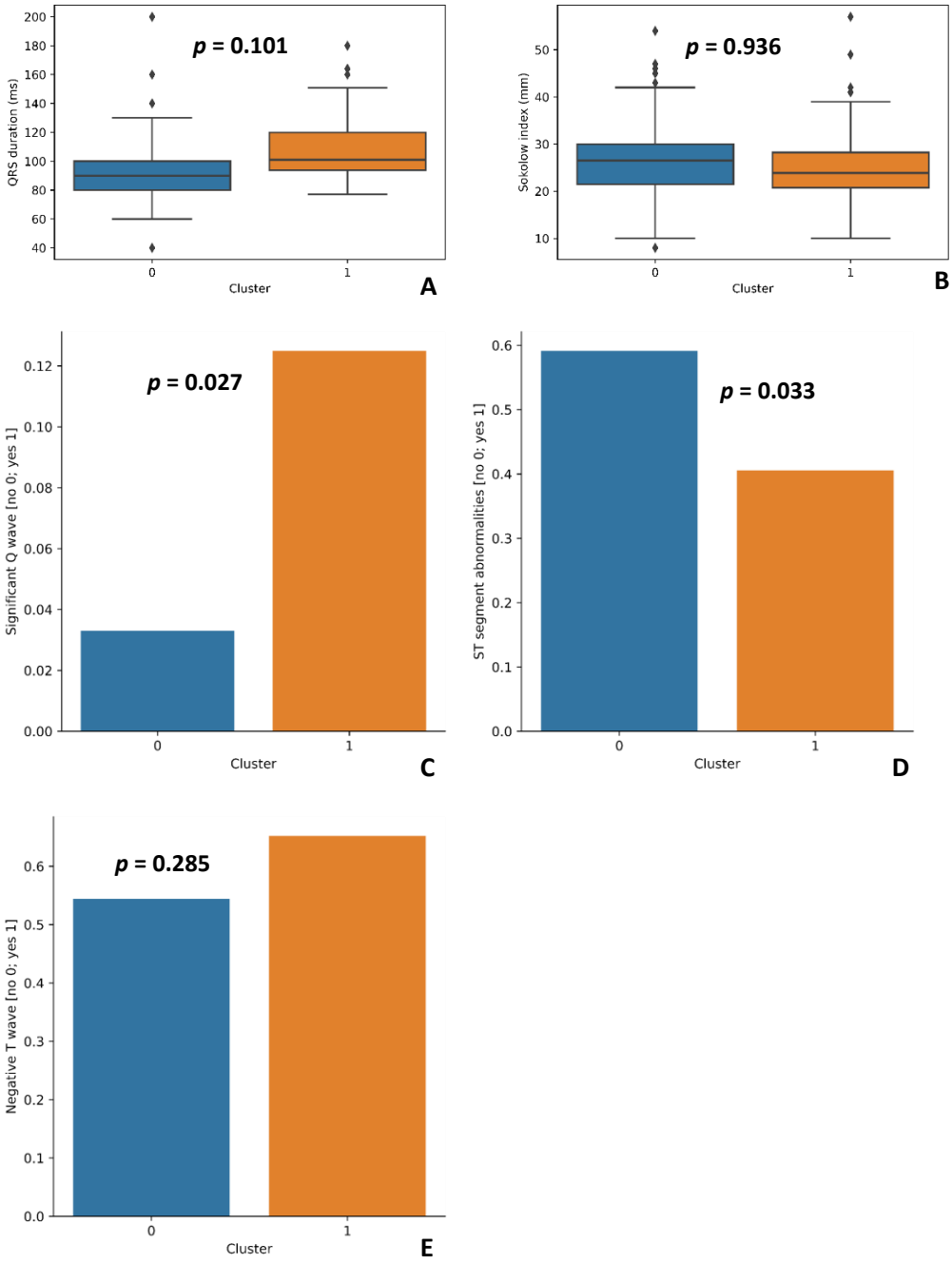

Figure 25. Two clusters setting, features: A — QRS duration, B — Sokolow index, C — Significant Q wave, D — ST segment abnormalities, E — Negative T wave

An approximate interpretation of clustering-logic for 2 clusters is presented in the form of visualized decision tree (Figure 26). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by clustering (Table 10).

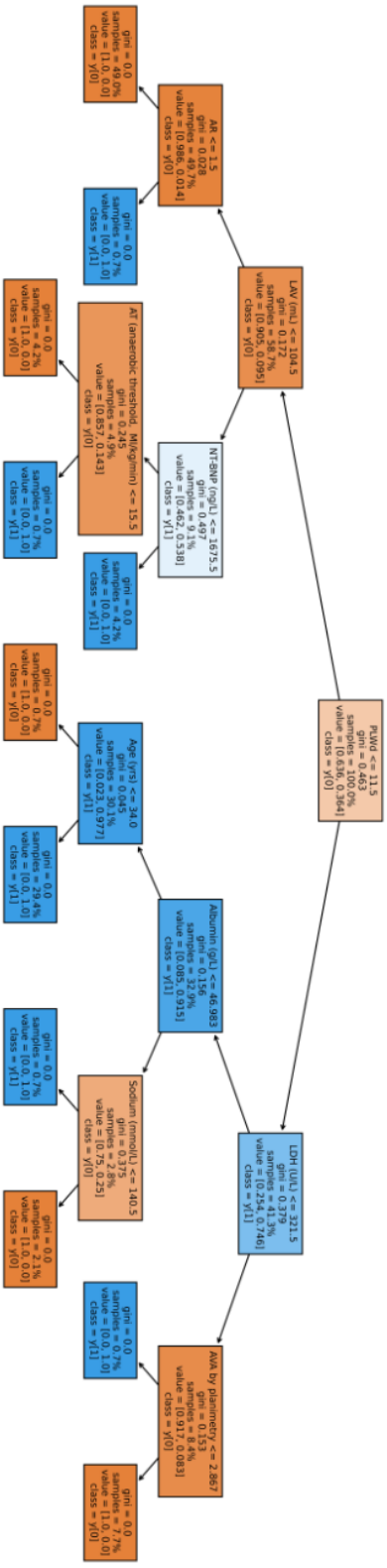

Figure 26. An approximate interpretation of clustering-logic for 2 clusters

Table 10. Feature importance — Top 35 features for distinguishing 2 clusters

Feature	Estimated importance
PLWd	0.080593
LVOT Vmax	0.074177
Serum albumin	0.066732
E/E'	0.060873
LAV	0.035834
MVVTI	0.035781
LDH	0.035777
LAVs	0.033546
MV maxPG	0.032848
AO	0.028702
AV Vmax	0.023693
AV maxPG	0.023128
AOvs	0.020798
Random glucose	0.016864
MV meanPG	0.013688
QRS duration	0.013257
Age	0.013242
Serum calcium	0.012902
AV meanPG	0.012264
NT-BNP	0.011428
AVA by planimetry	0.011422
RVSP	0.010845
Peak VO_2	0.010418
Diastolic	0.010385
LA	0.010270
Peak HR	0.010065
EFLV	0.009574
RAVs	0.009526
TAPSE	0.009351
AT	0.008802
PR interval	0.008787
Sodium	0.008229
Weight	0.008145
AscAO	0.008133
Troponin	0.008042

4.1.2.2.2. Three clusters

Characteristics of 3 clusters determined are shown in Figures 27-44.

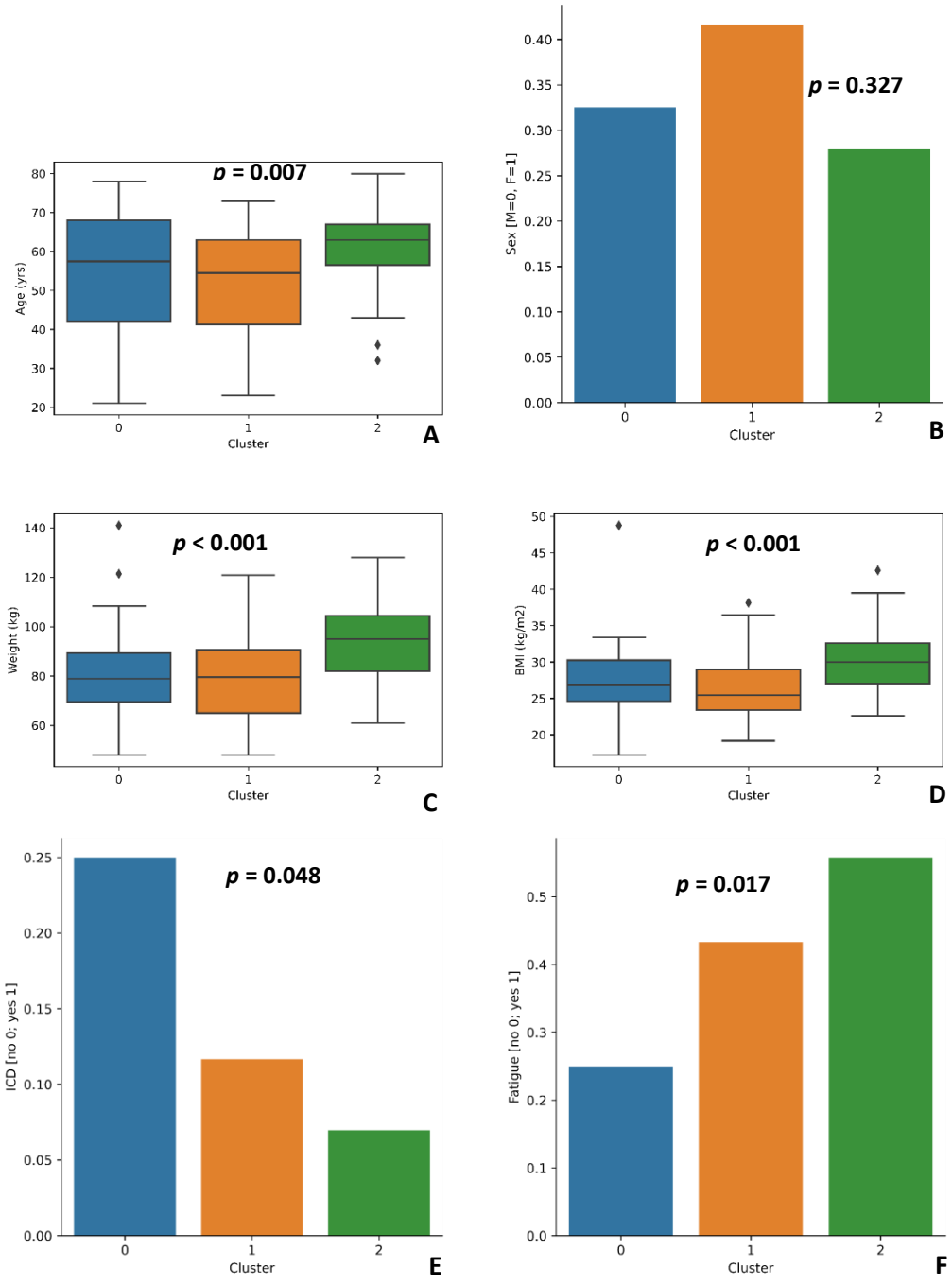

Figure 27. Three clusters setting, features: A — Age, B — Sex, C — Weight, D — BMI, E — ICD, F — Fatigue

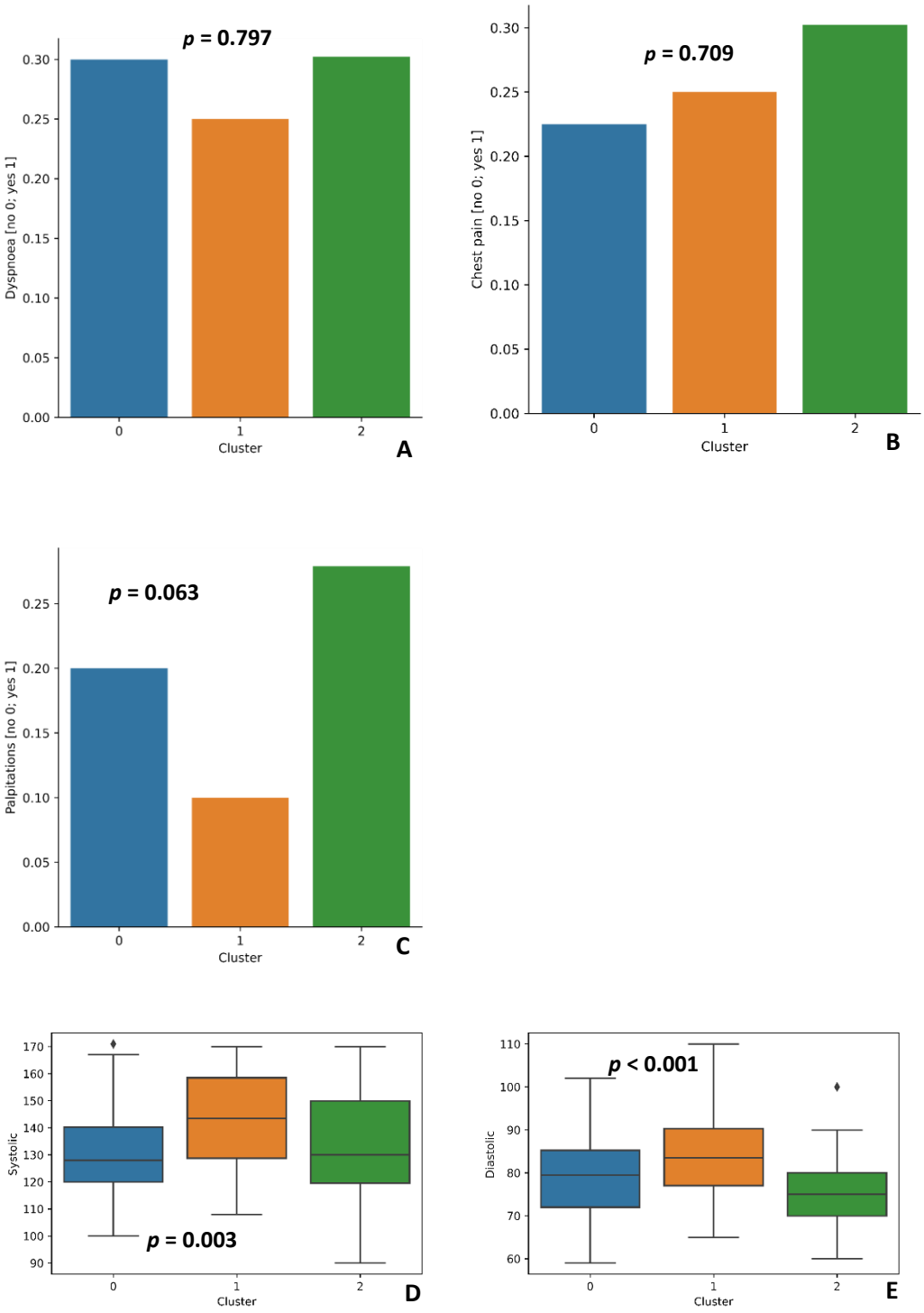

Figure 28. Three clusters setting, features: A — Dyspnea, B — Chest pain, C — Palpitations, D — Systolic blood pressure, E — Diastolic blood pressure

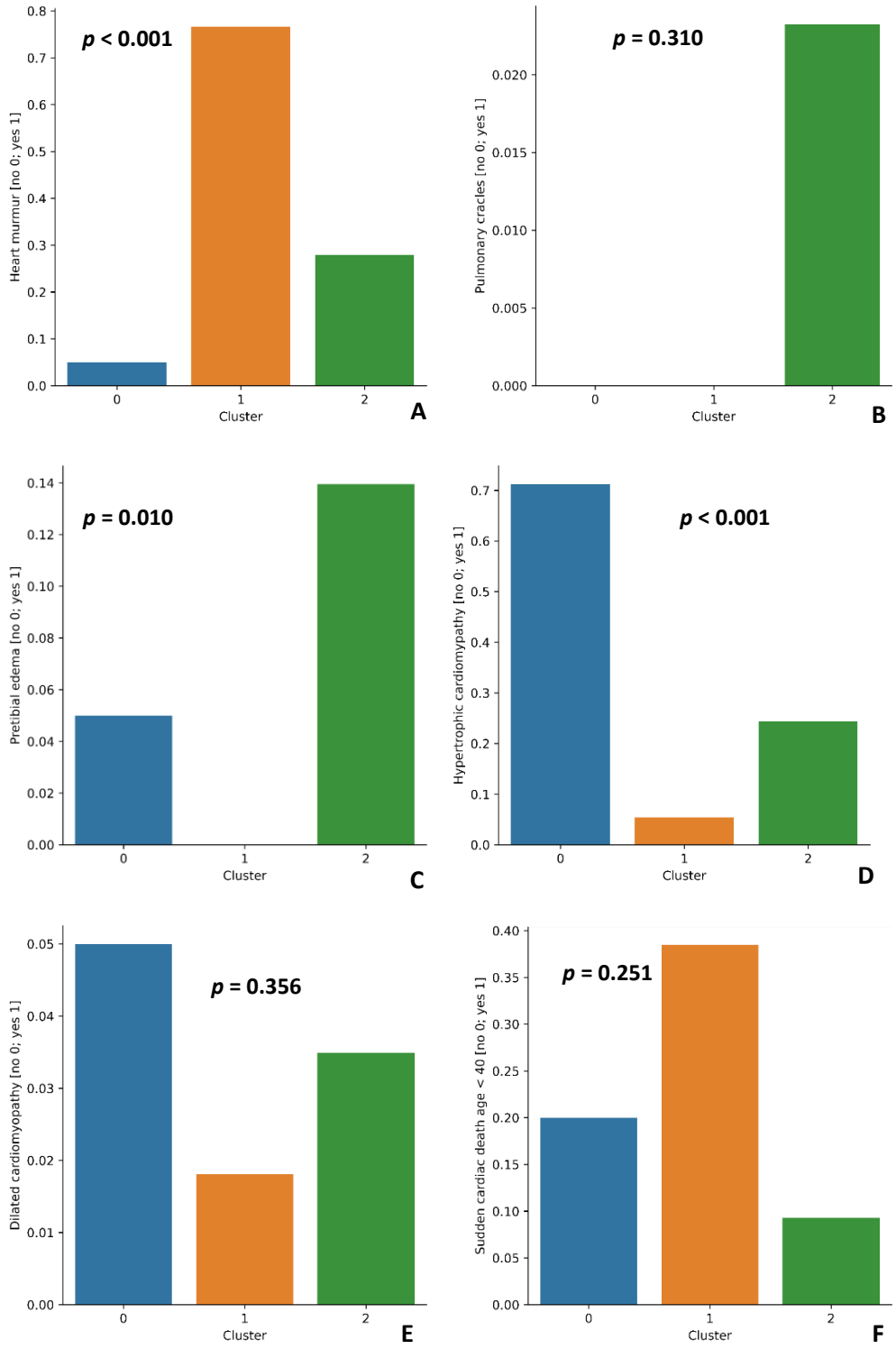

Figure 29. Three clusters setting, features: A — Heart murmur, B — Pulmonary crackles, C — Pretibial edema, D — HCM in family history, E — DCM in family history, F — SCD in age < 40 in family history

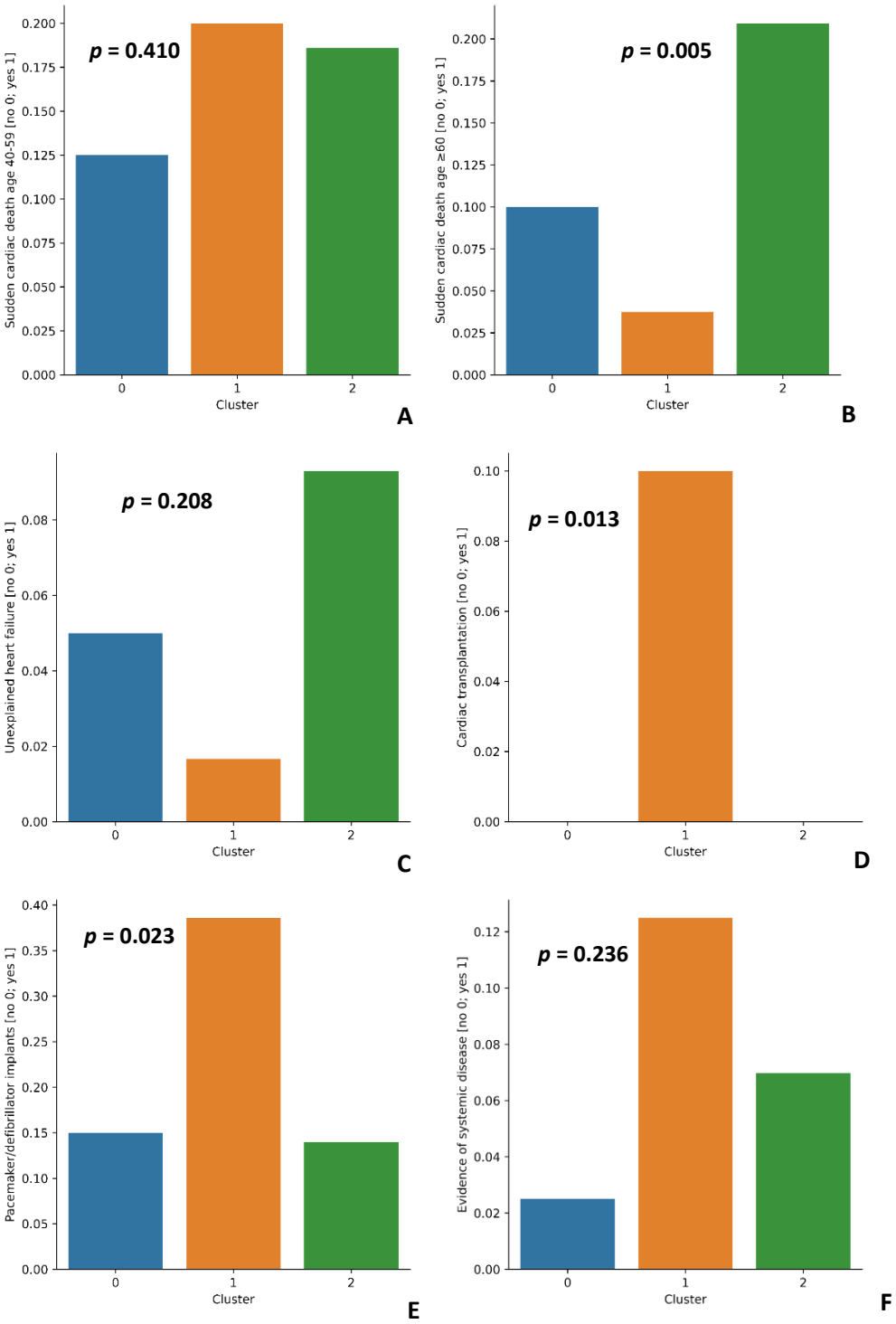

Figure 30. Three clusters setting, features: A — SCD in age 40-59 in family history, B — SCD in age ≥ 60 in family history, C — Unexplained HF in family history, D — Cardiac transplantation in family history, E — Pacemaker/defibrillator implants in family history, F — Evidence of systemic disease in family history

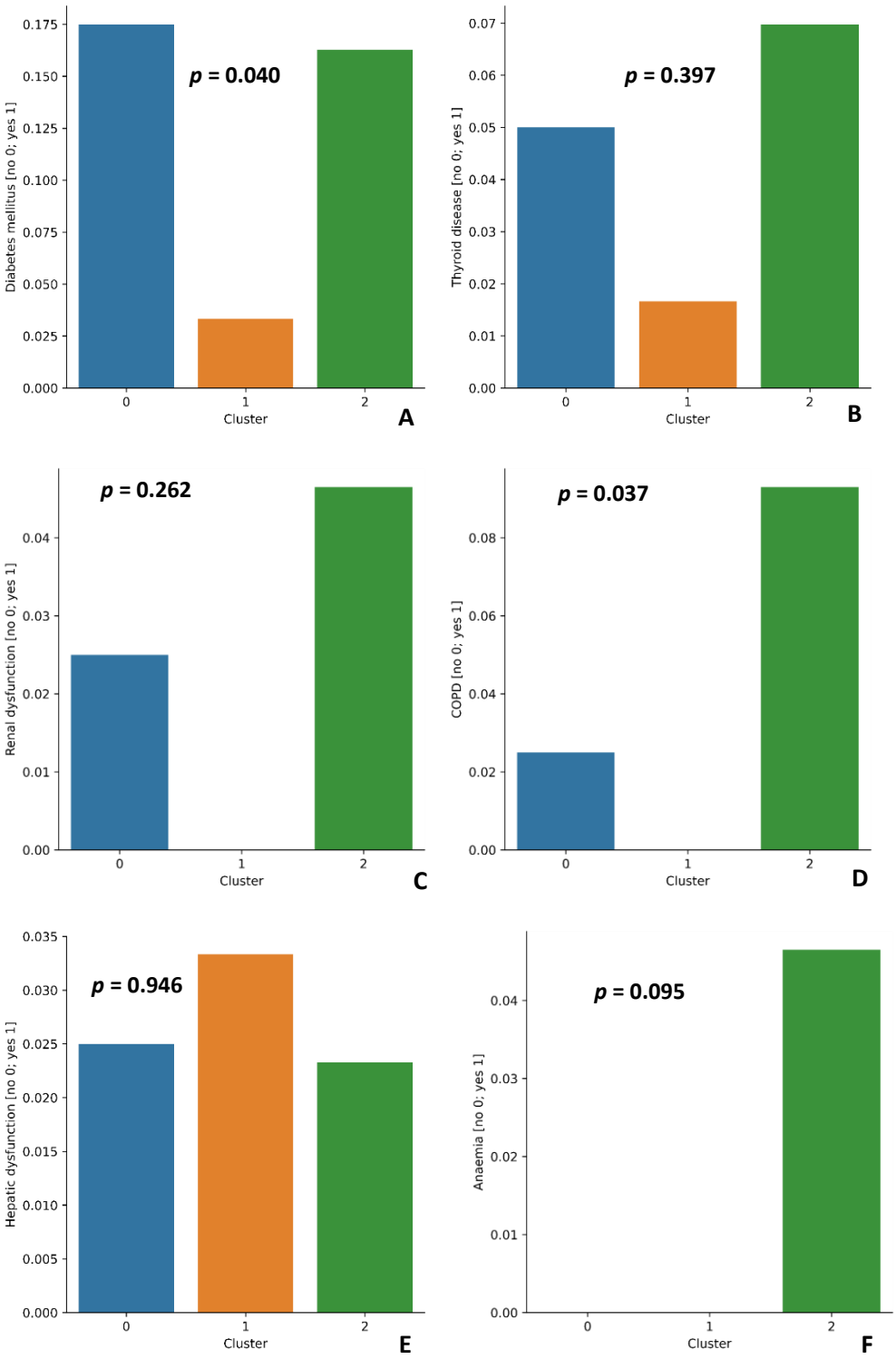

Figure 31. Three clusters setting, features: A — Diabetes mellitus, B — Thyroid disease, C — Renal dysfunction, D — COPD, E — Hepatic dysfunction, F — Anemia

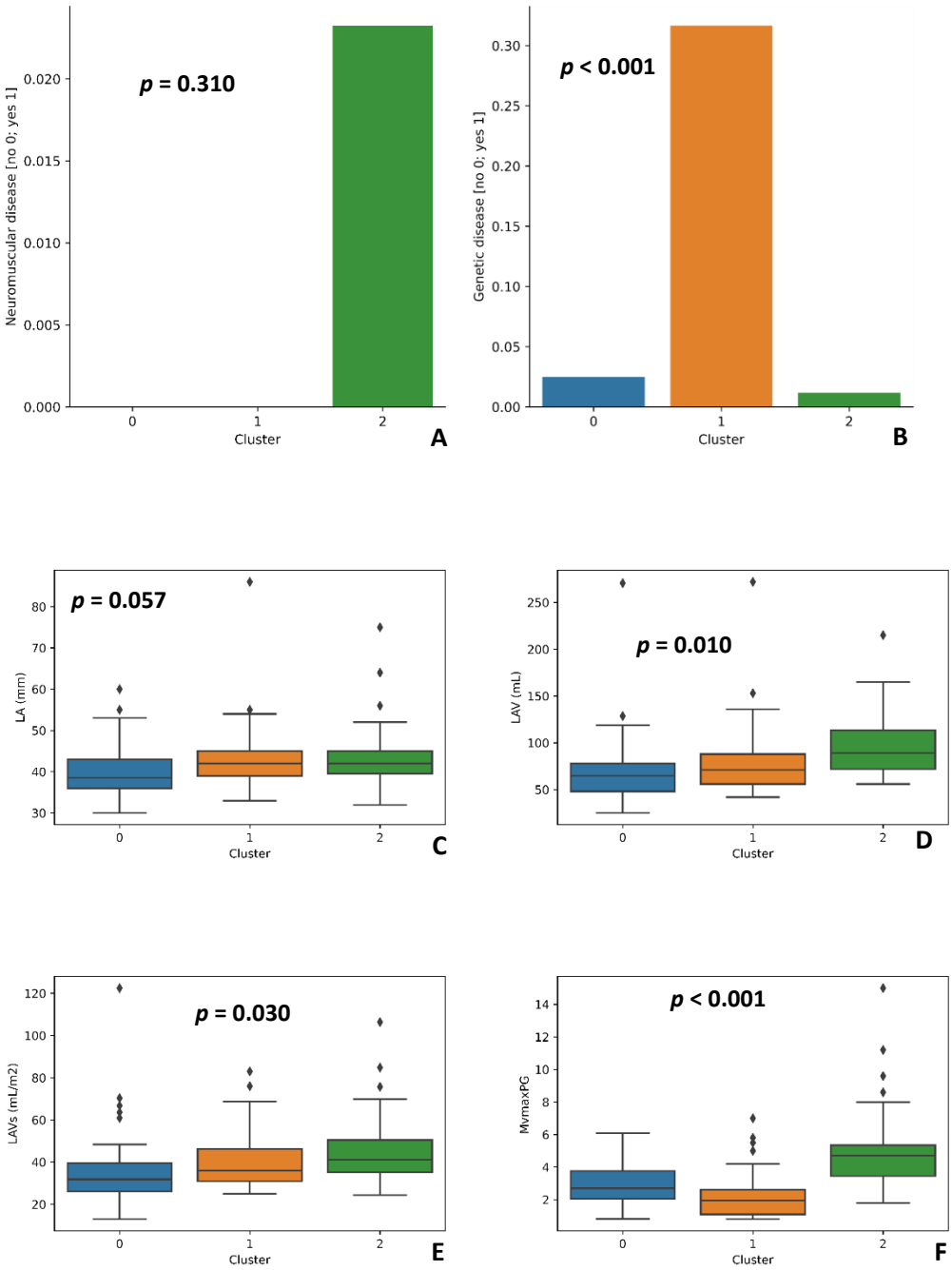

Figure 32. Three clusters setting, features: A — Neuromuscular disease, B — Genetic disease, C — LA, D — LAV, E — LAVs, F — MV maxPG

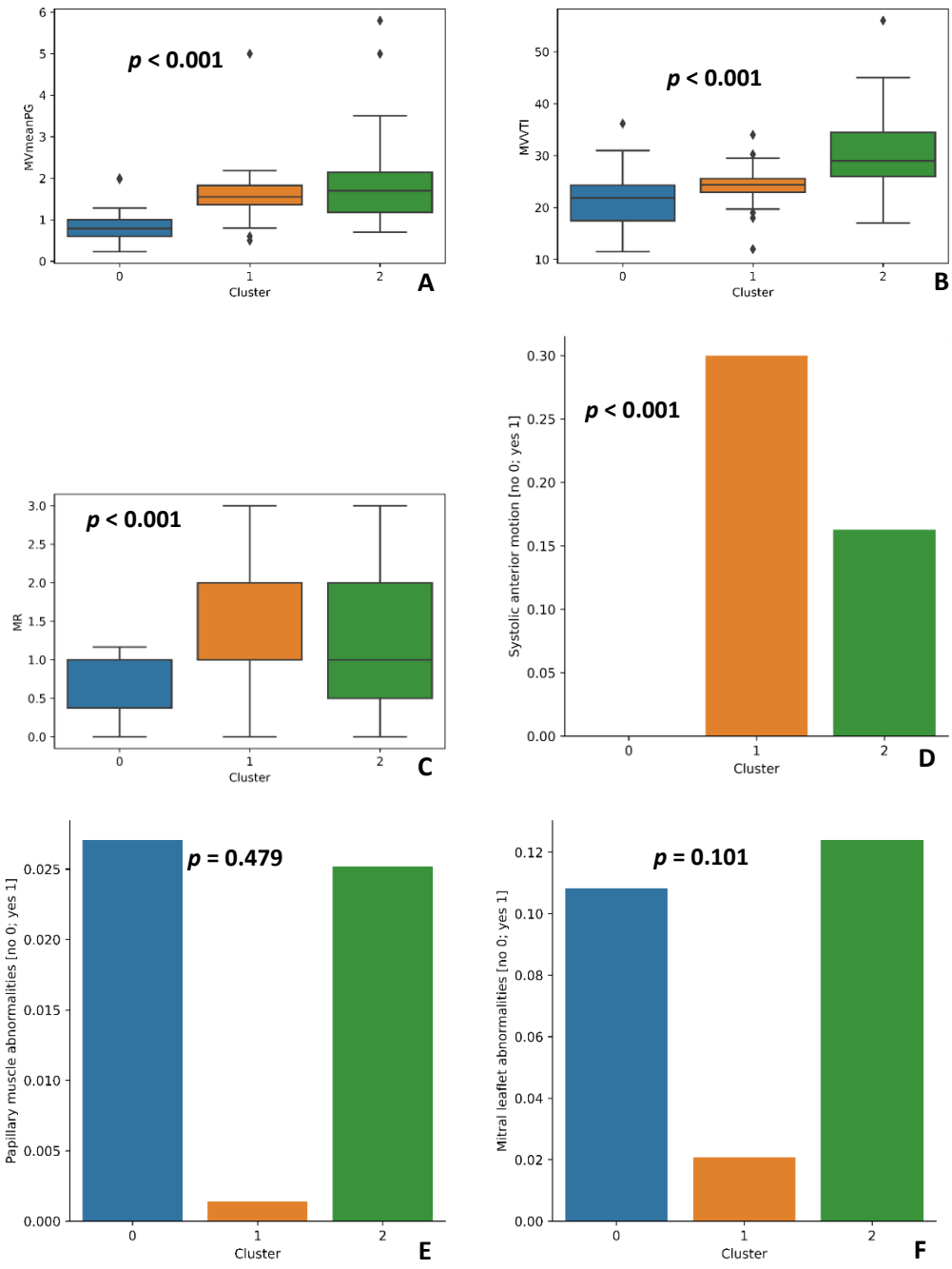

Figure 33. Three clusters setting, features: A — MV meanPG, B — MVVTI, C — MR, D — SAM, E — Papillary muscle abnormalities, F — Mitral leaflet abnormalities

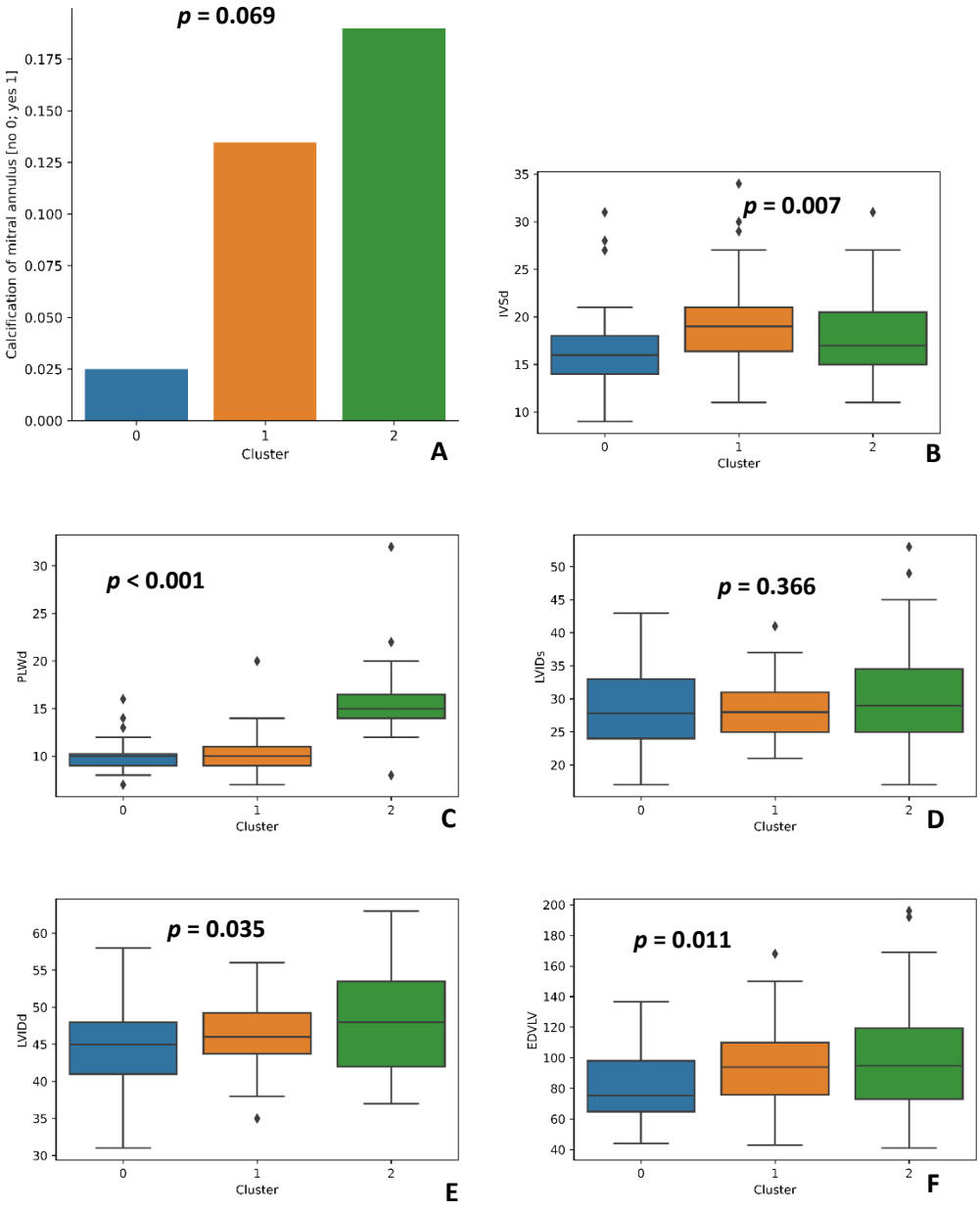

Figure 34. Three clusters setting, features: A — Calcification of mitral annulus, B — IVSd, C — PLWd, D — LVIDs, E — LVIDd, F — EDVLV

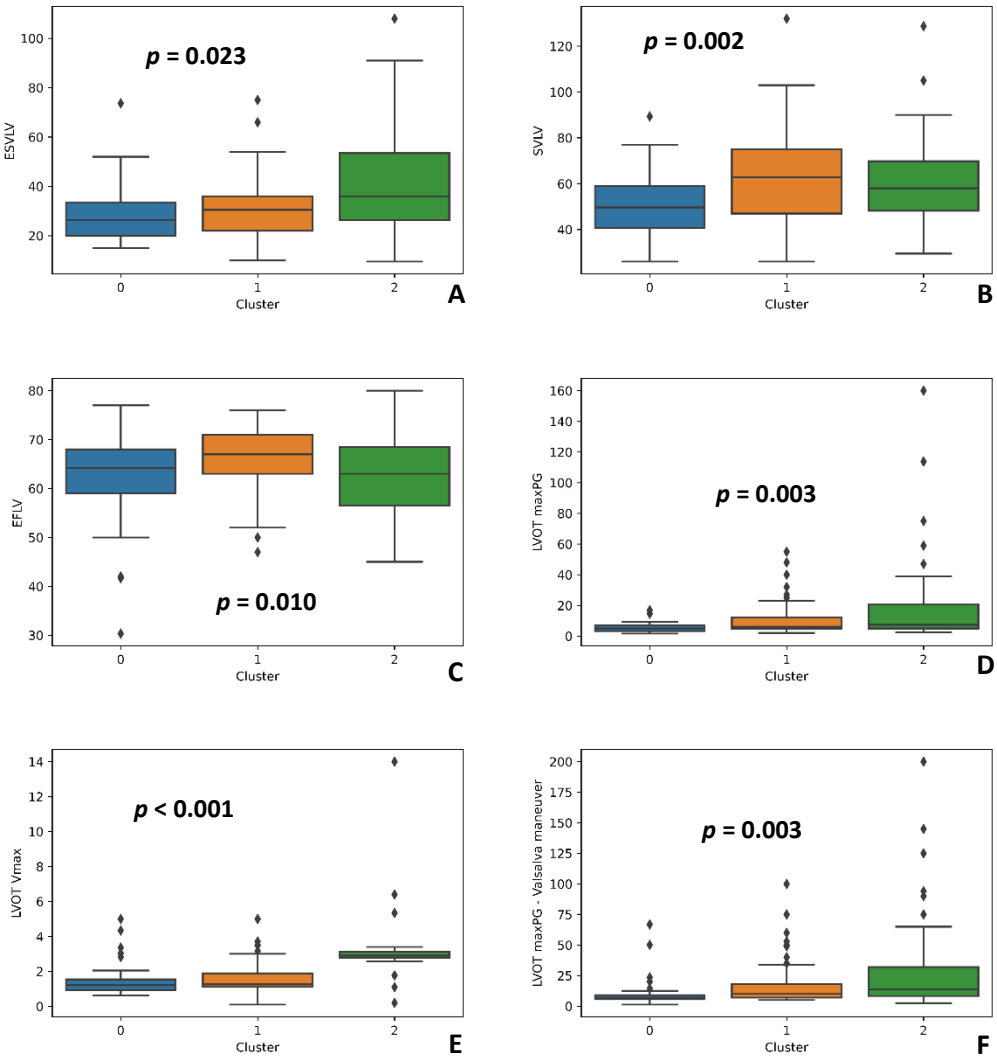

Figure 35. Three clusters setting, features: A — ESVLV, B — SVLV, C — EFLV, D — LVOT maxPG, E — LVOT Vmax, F — LVOT maxPG - Valsalva maneuver

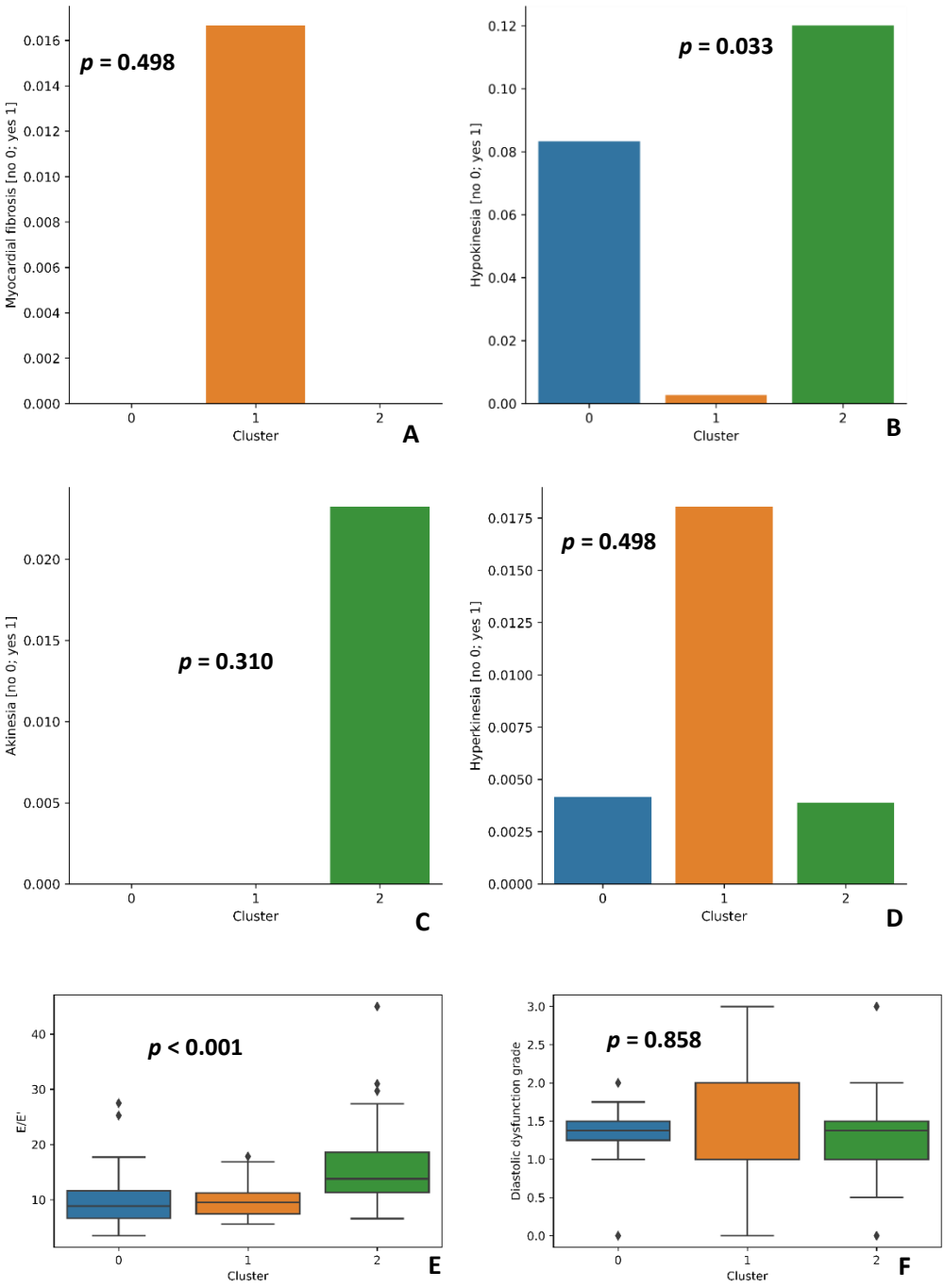

Figure 36. Three clusters setting, features: A — Myocardial fibrosis, B — Hypokinesia, C — Akinesia, D — Hyperkinesia, E — E/E', F — Diastolic dysfunction grade

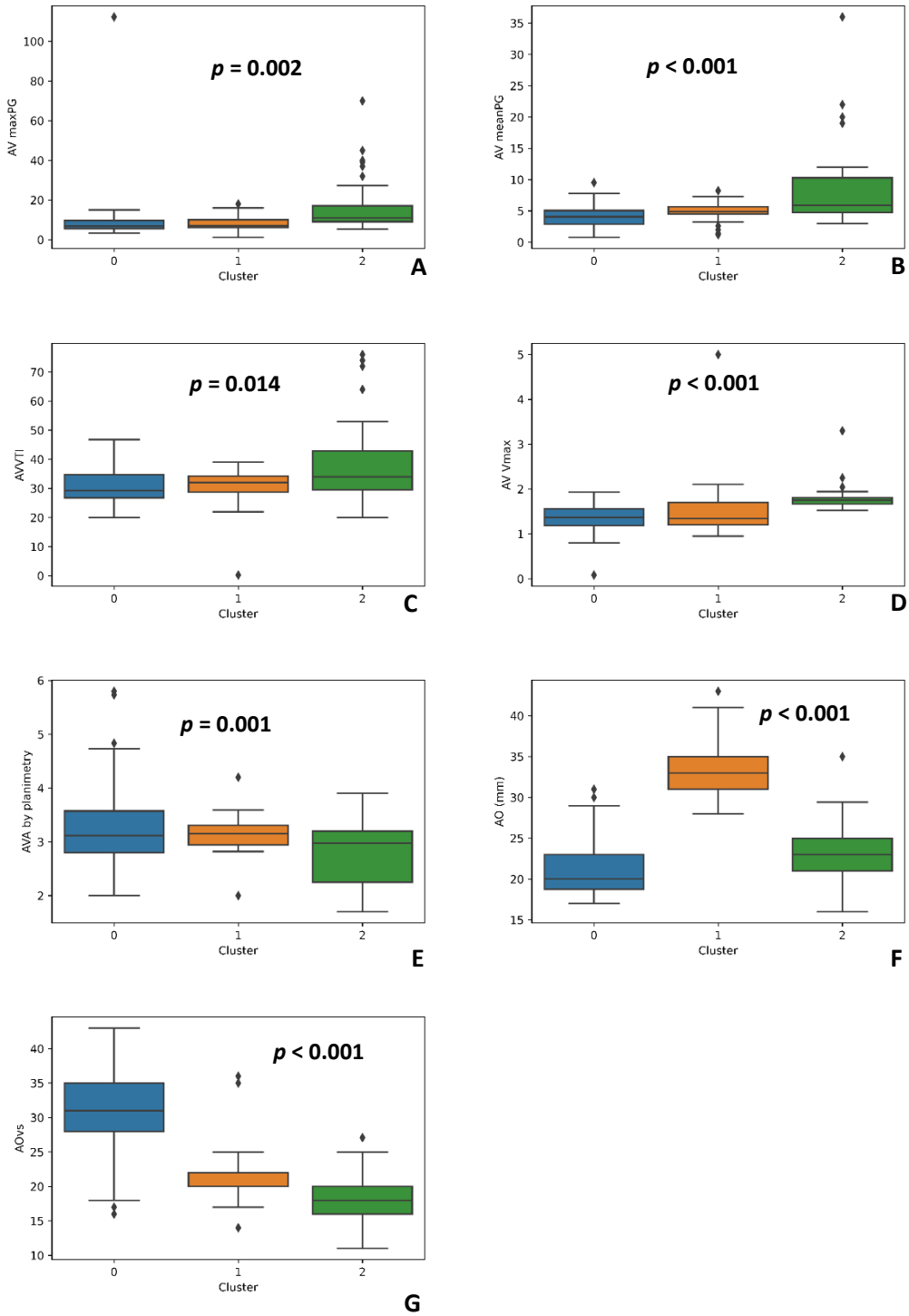

Figure 37. Three clusters setting, features: A — AV maxPG, B — AV meanPG, C — AVVTI, D — AV Vmax, E — AVA by planimetry, F — AO, H — AOvs

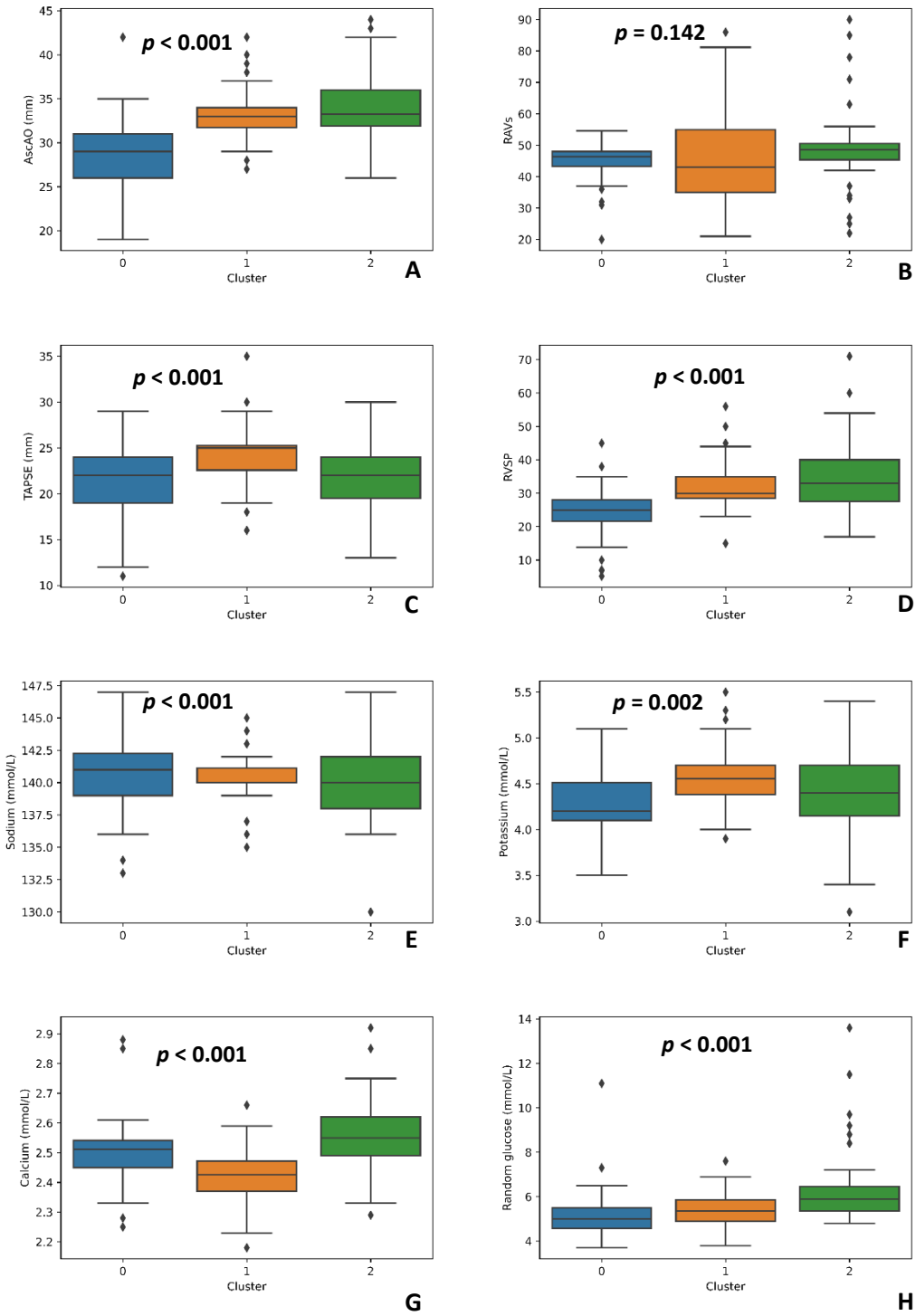

Figure 38. Three clusters setting, features: A — AscAO, B — RAVs, C — TAPSE, D — RVSP, E — Sodium, F — Potassium, G — Calcium, H — Random glucose

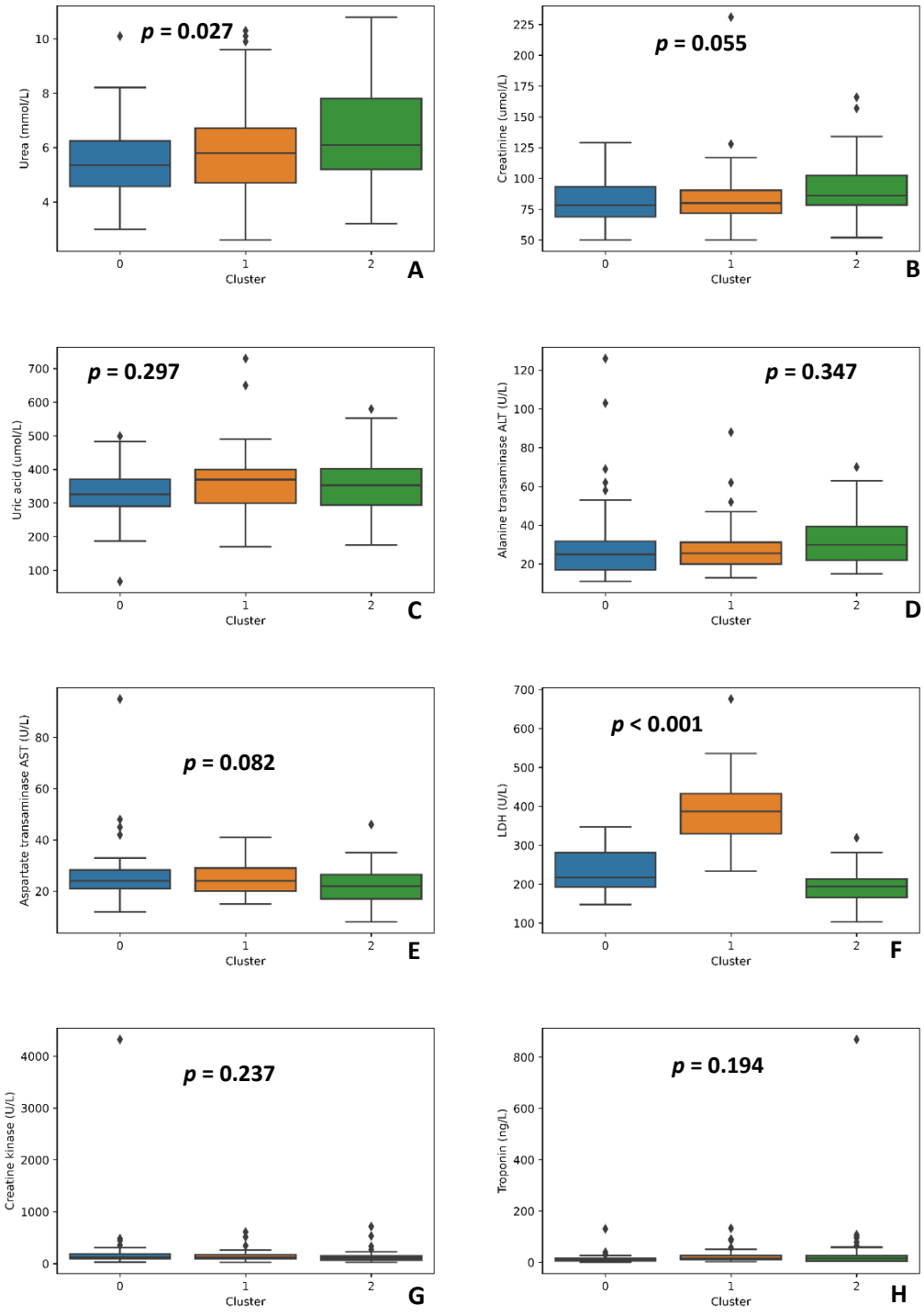

Figure 39. Three clusters setting, features: A — Urea, B — Creatinine, C — Uric acid, D — ALT, E — AST, F — LDH, G — Creatine-kinase, H — Troponin

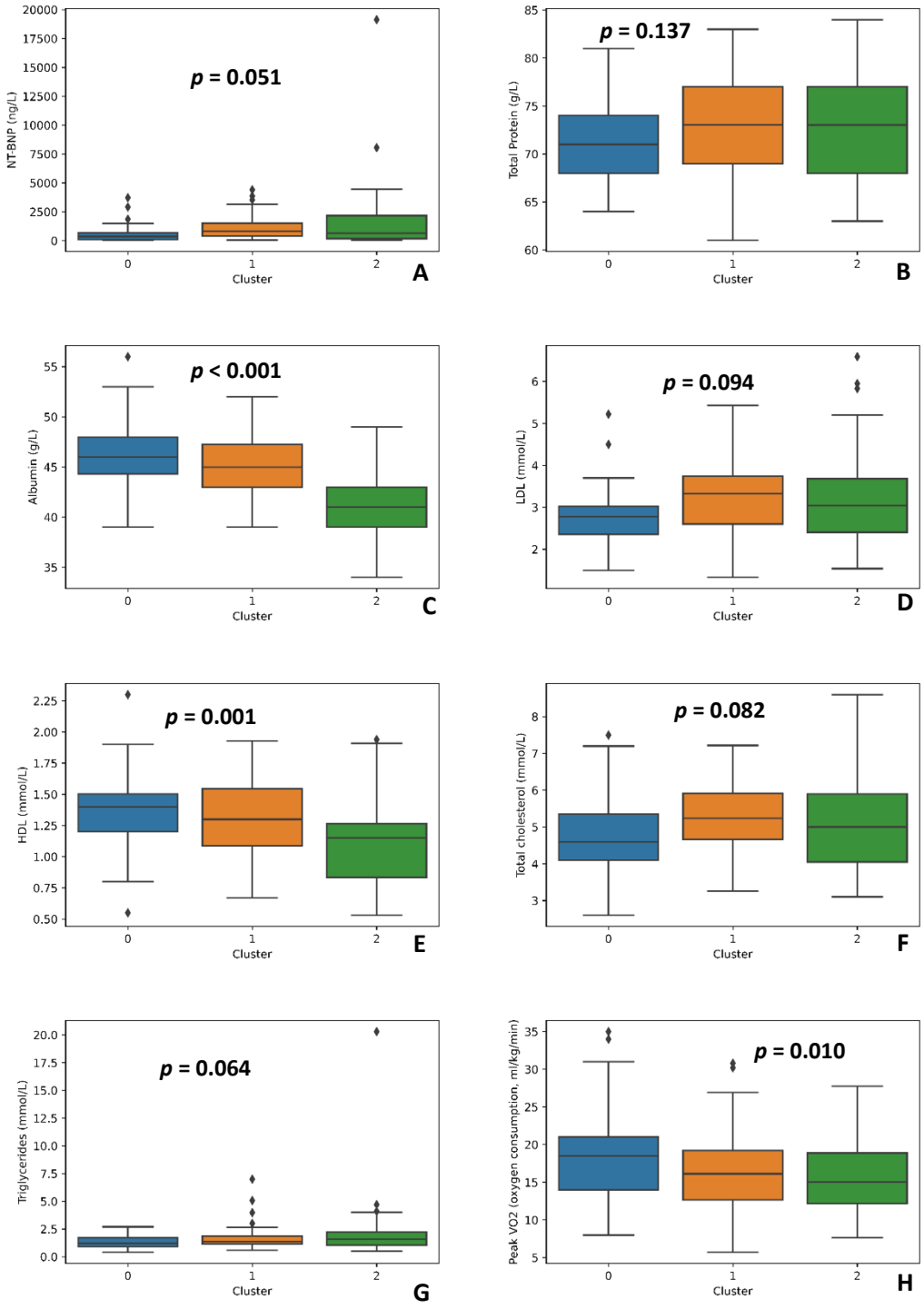

Figure 40. Three clusters setting, features: A — NT-BNP, B — Total protein, C — Albumin, D — LDL, E — HDL, F — Total cholesterol, G — Triglyceride, H — Peak VO_2

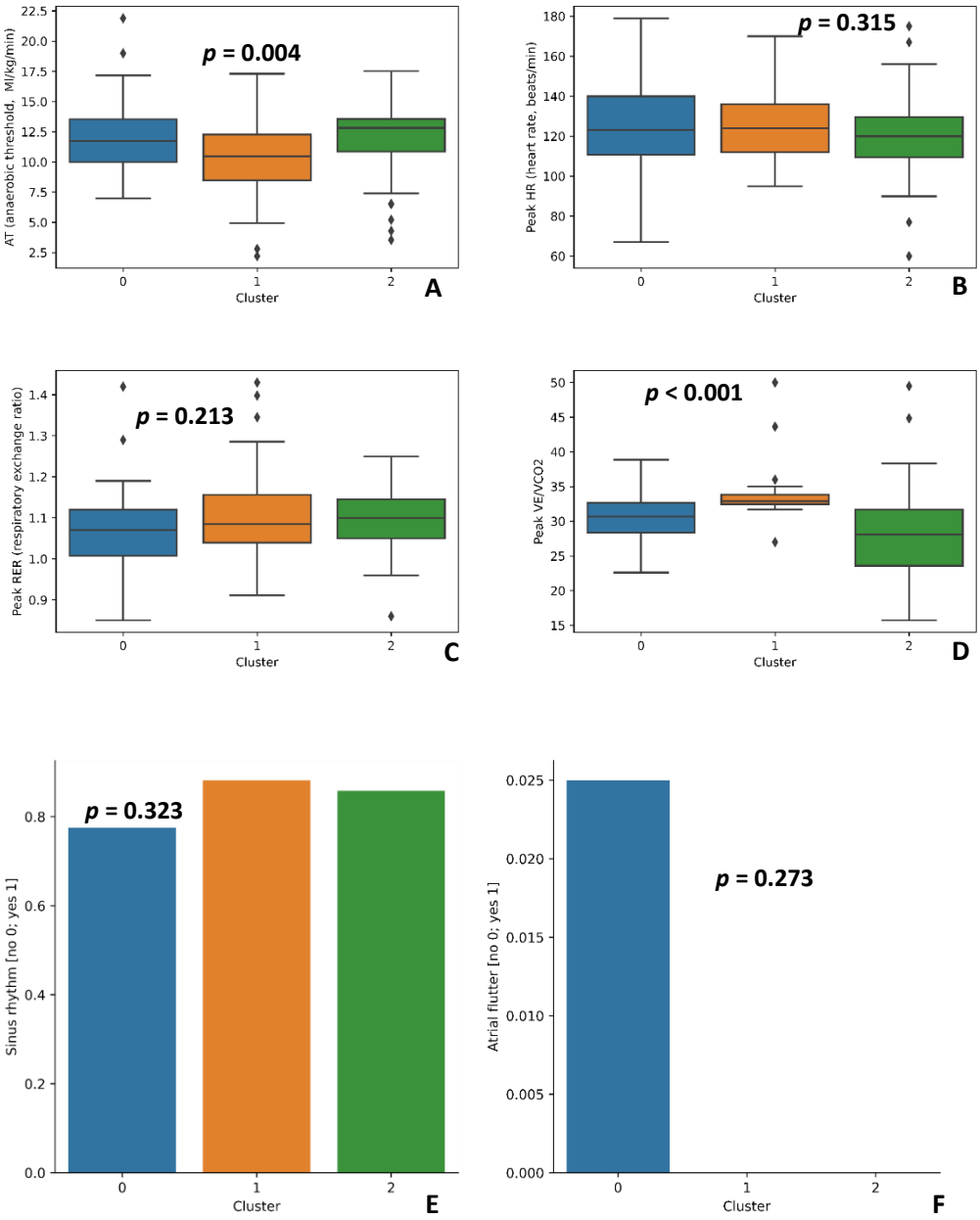

Figure 41. Three clusters setting, features: A — AT, B — Peak HR, C — Peak RER, D — Peak VE/VCO_2, E — Sinus rhythm, F — Atrial flutter

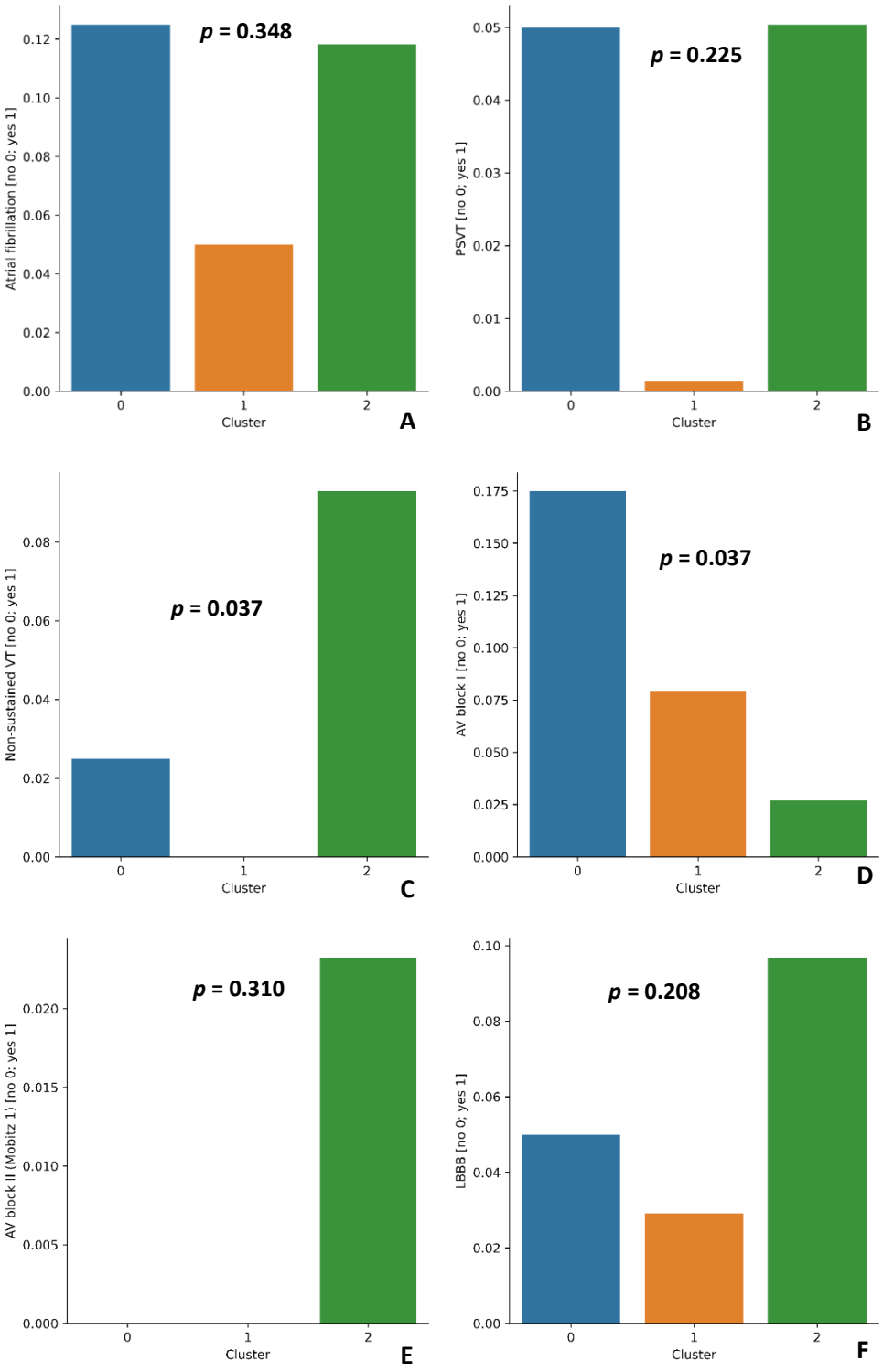

Figure 42. Three clusters setting, features: A — AF, B — PSVT, C — Non-sustained VT, D — AV block I, E — AV block II (Mobitz 1), F — LBBB

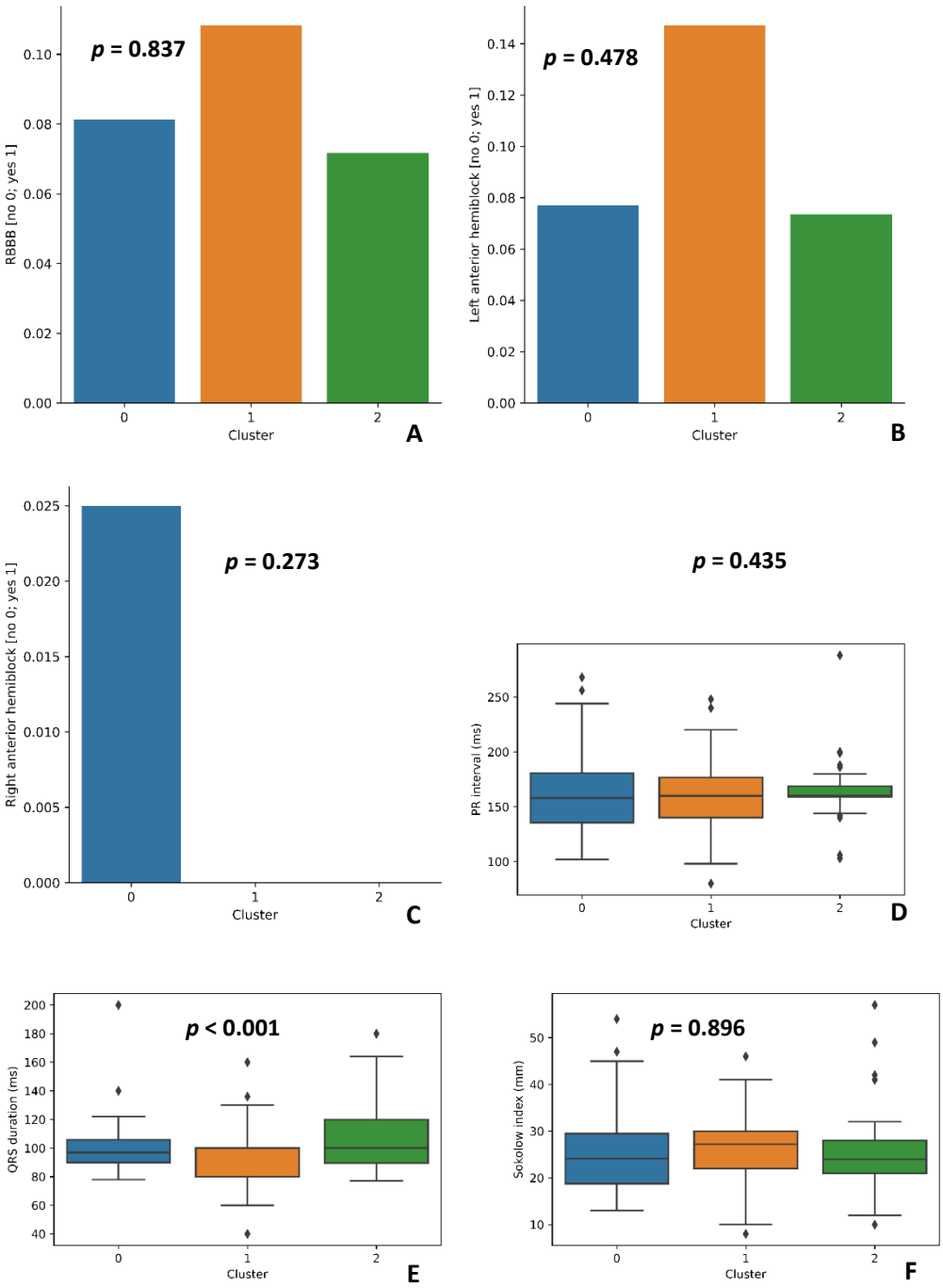

Figure 43. Three clusters setting, features: A — RBBB, B — Left anterior hemiblock, C — Right anterior hemiblock, D — PR interval, E — QRS duration, F — Sokolow index

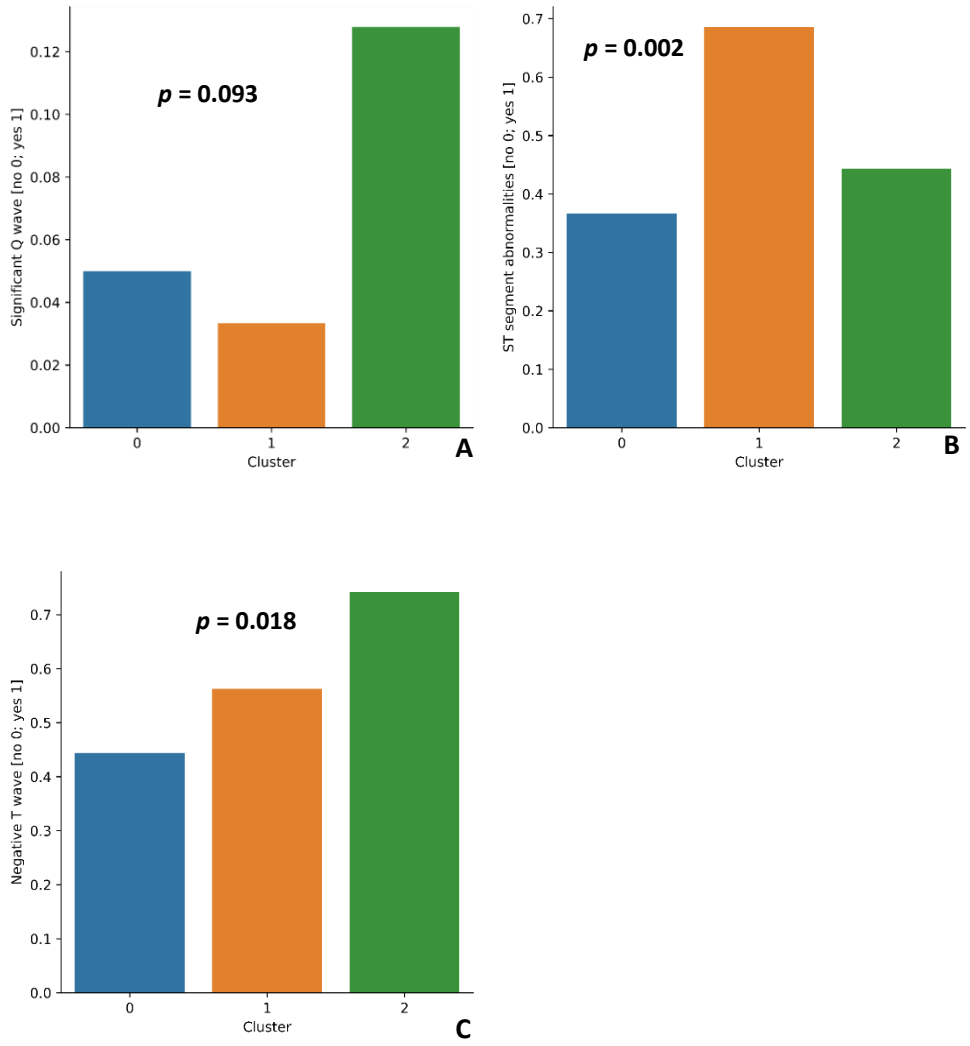

Figure 44. Three clusters setting, features: A — Significant Q wave, B — ST segment abnormalities, C — Negative T wave

An approximate interpretation of clustering-logic for 3 clusters is presented in the form of visualized decision tree (Figure 45). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by clustering (Table 11).

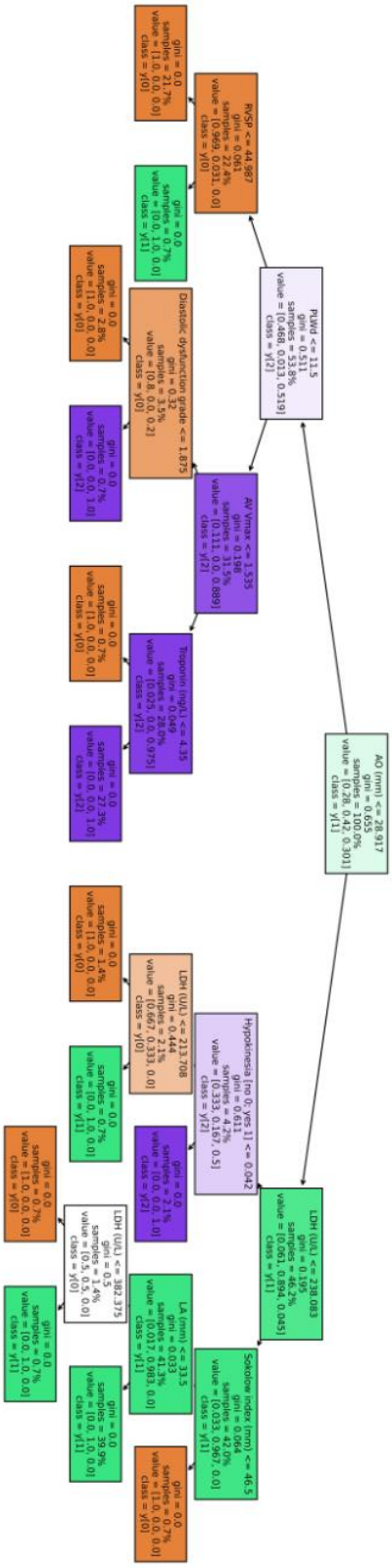

Figure 45. An approximate interpretation of clustering-logic for 3 clusters

Table 11. Feature importance — Top 40 features for distinguishing 3 clusters

Feature	Estimated importance
AO	0.101542
LDH	0.077974
AOvs	0.075634
PLWd	0.069079
LVOT Vmax	0.049774
MV meanPG	0.047133
Peak VE/VCO_2	0.043810
Albumin	0.028354
MVVTI	0.027188
HCM in family history [yes/no]	0.025341
AscAO (mm)	0.024483
MV maxPG	0.021908
AV Vmax	0.020156
Heart murmur [yes/no]	0.017667
E/E'	0.015789
AV maxPG	0.015206
QRS duration	0.015160
Serum calcium	0.012818
RVSP	0.012465
BMI	0.011648
Weight	0.010973
AV meanPG	0.010871
MR	0.010865
AVA by planimetry	0.010404
NYHA class	0.010226
LVOT maxPG - Valsalva maneuver	0.009823
Height	0.007761
NT-BNP	0.007440
RAVs	0.006483
EDVLV	0.006385
PR interval	0.006343
Heart rate	0.006067
ESVLV	0.006051
LAV	0.005837
TR	0.005767
ALT	0.005680
Anaerobic threshold	0.005592
LVOT maxPG	0.005365
LAVs	0.005342
Total cholesterol	0.004976

4.1.2.2.3. Four clusters

Characteristics of 4 clusters determined are shown in Figures 46-62.

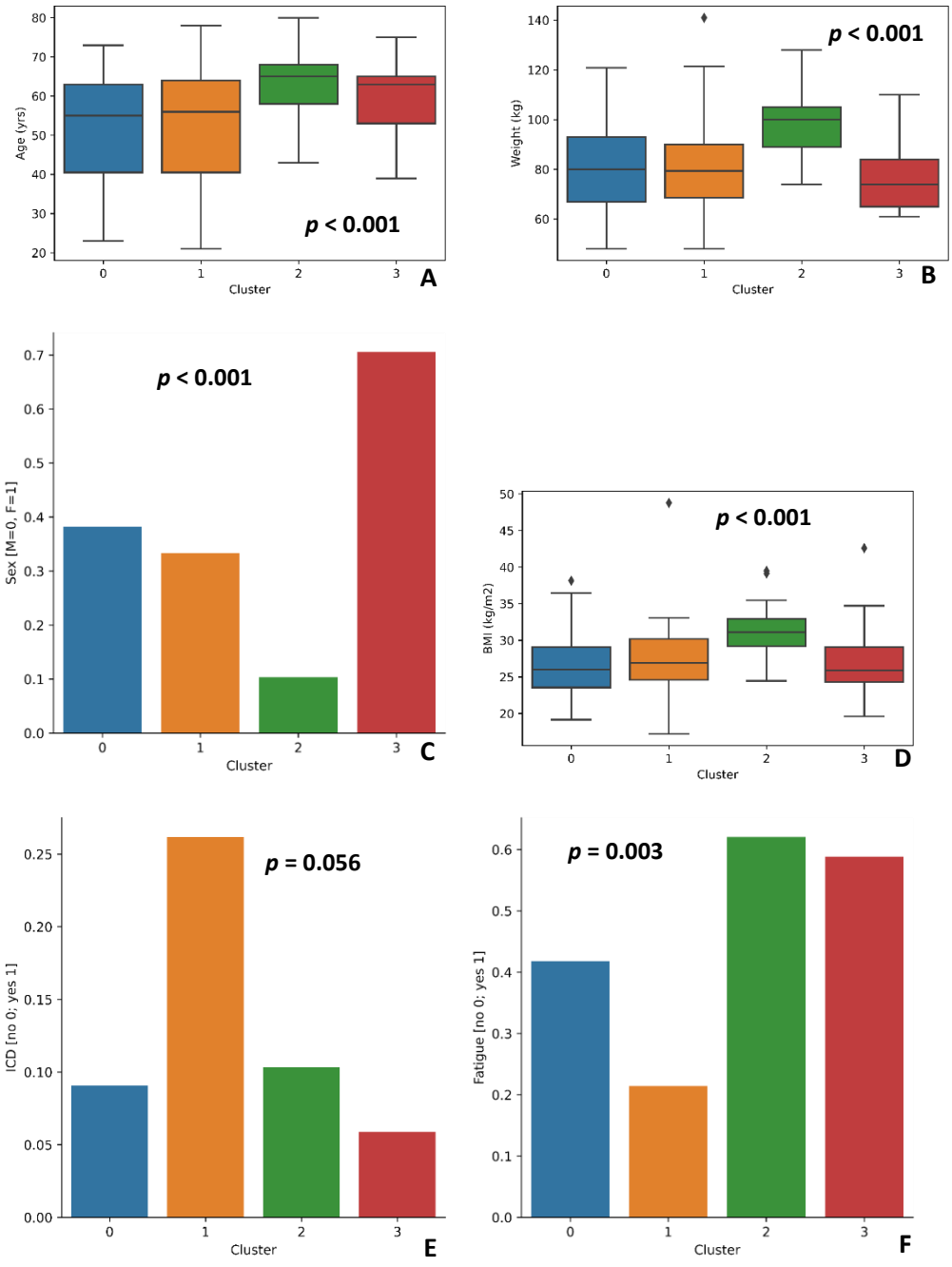

Figure 46. Four clusters setting, features: A — Age, B — Weight, C — Sex, D — BMI, E — ICD, F — Fatigue (283)

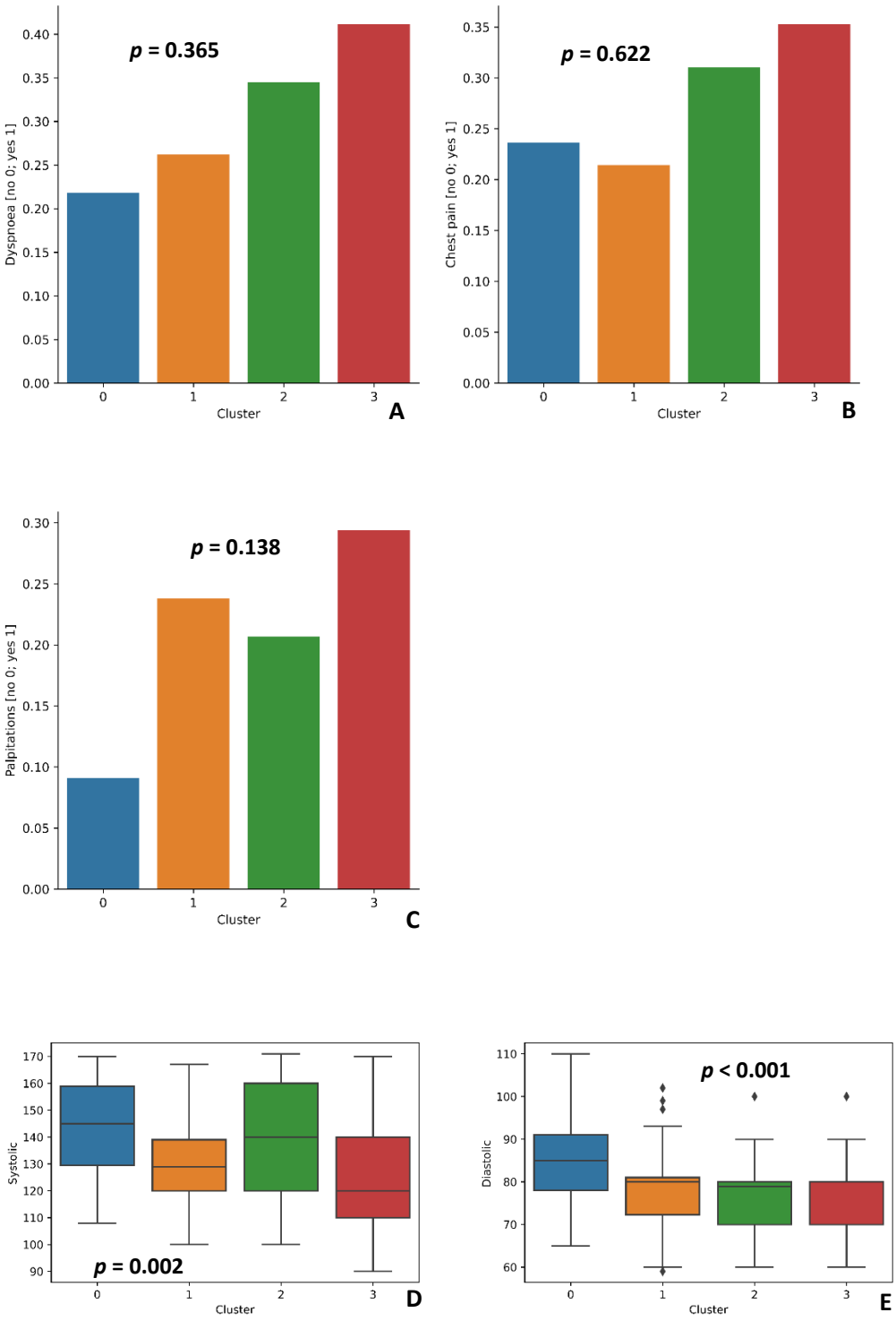

Figure 47. Four clusters setting, features: A — Dyspnea, B — Chest pain, C — Palpitations, D — Systolic blood pressure, E — Diastolic blood pressure (283)

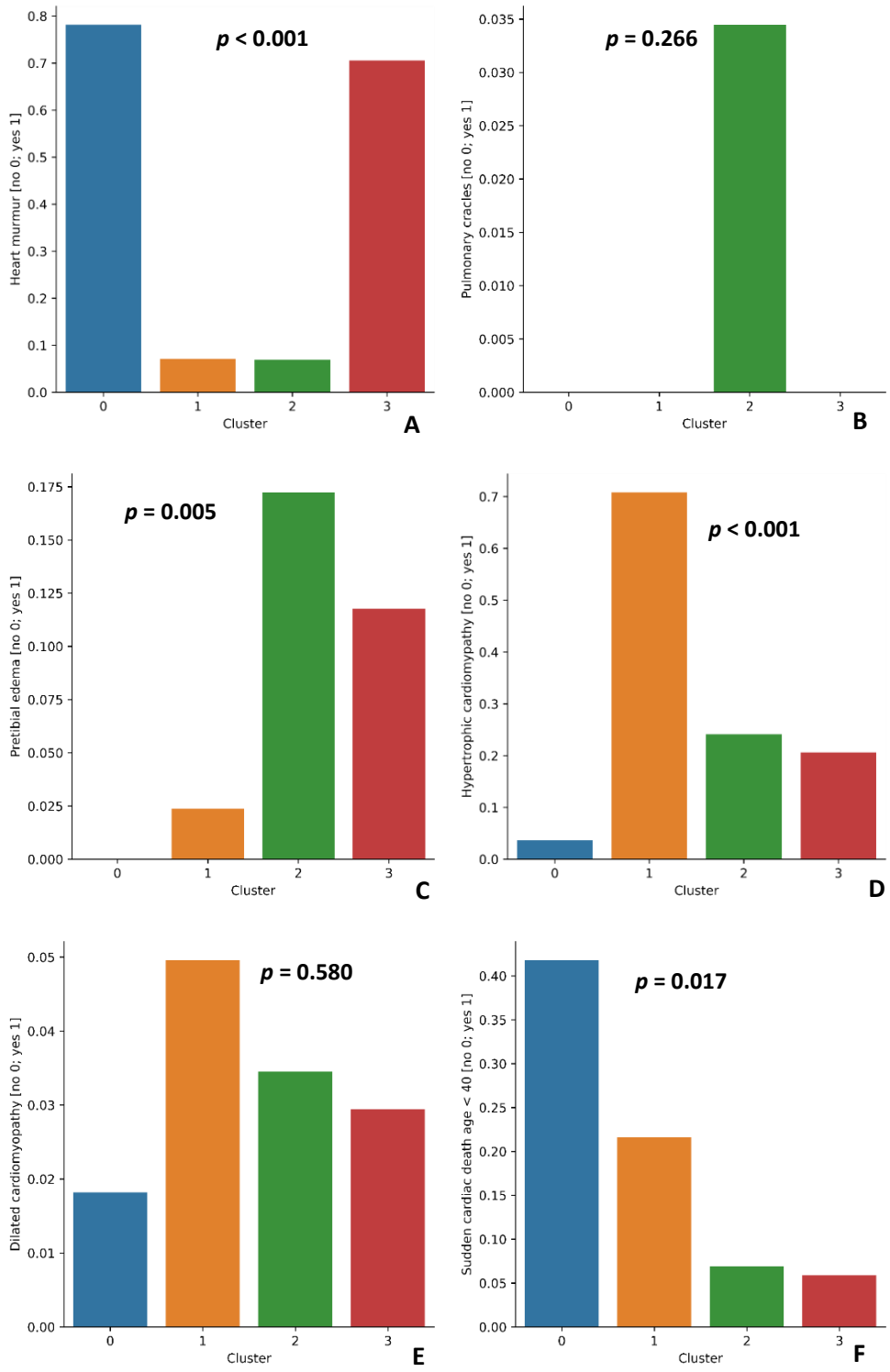

Figure 48. Four clusters setting, features: A — Heart murmur, B — Pulmonary crackles, C — Pretibial edema, D — HCM in family history, E — DCM in family history, F — SCD in age < 40 in family history (283)

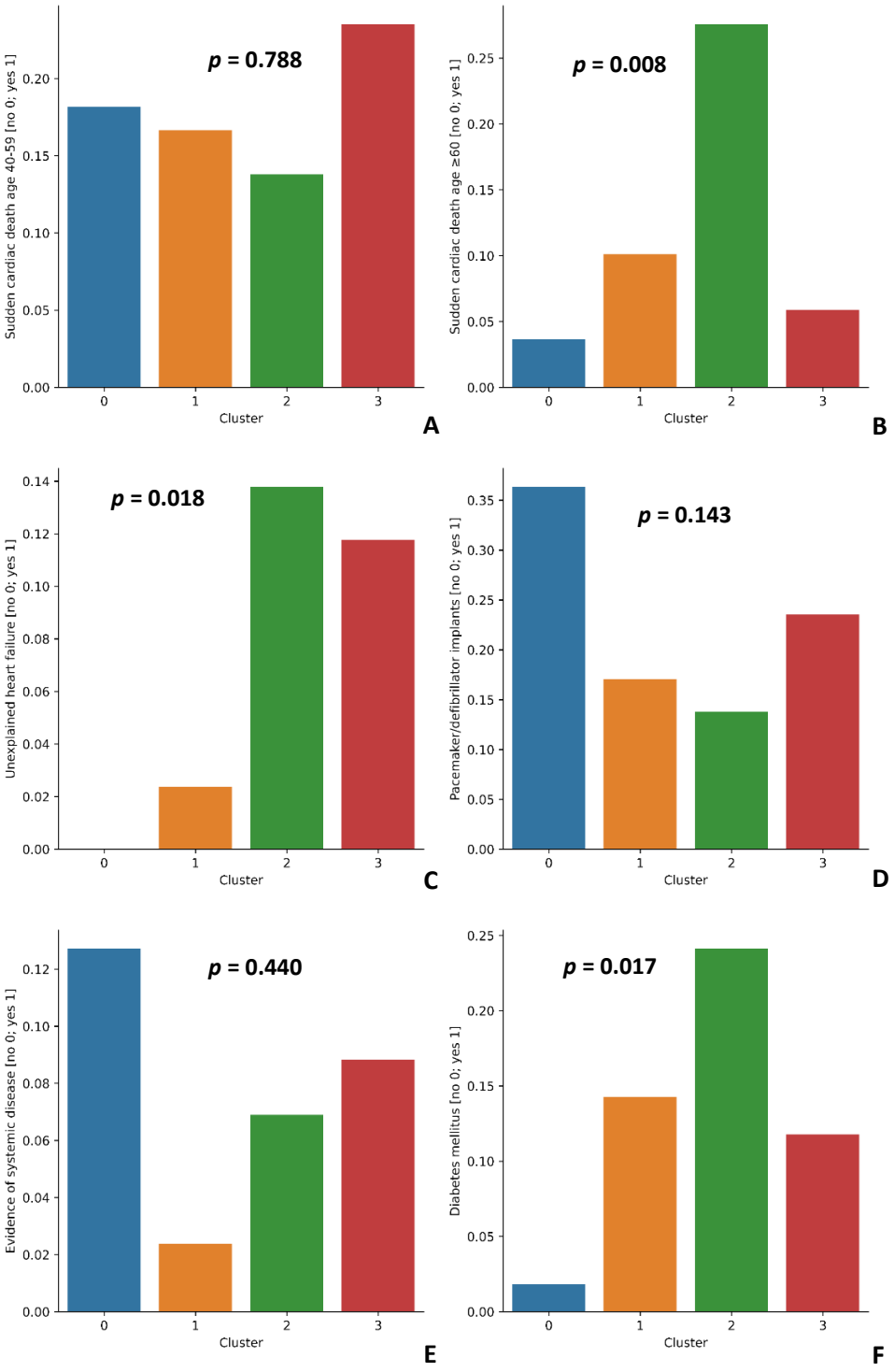

Figure 49. Four clusters setting, features: A — SCD in age 40-59 in family history, B — SCD in age ≥ 60 in family history, C — Unexplained HF in family history, D — Pacemaker/defibrillator implants in family history, E — Evidence of systemic disease in family history, F — Diabetes mellitus (283)

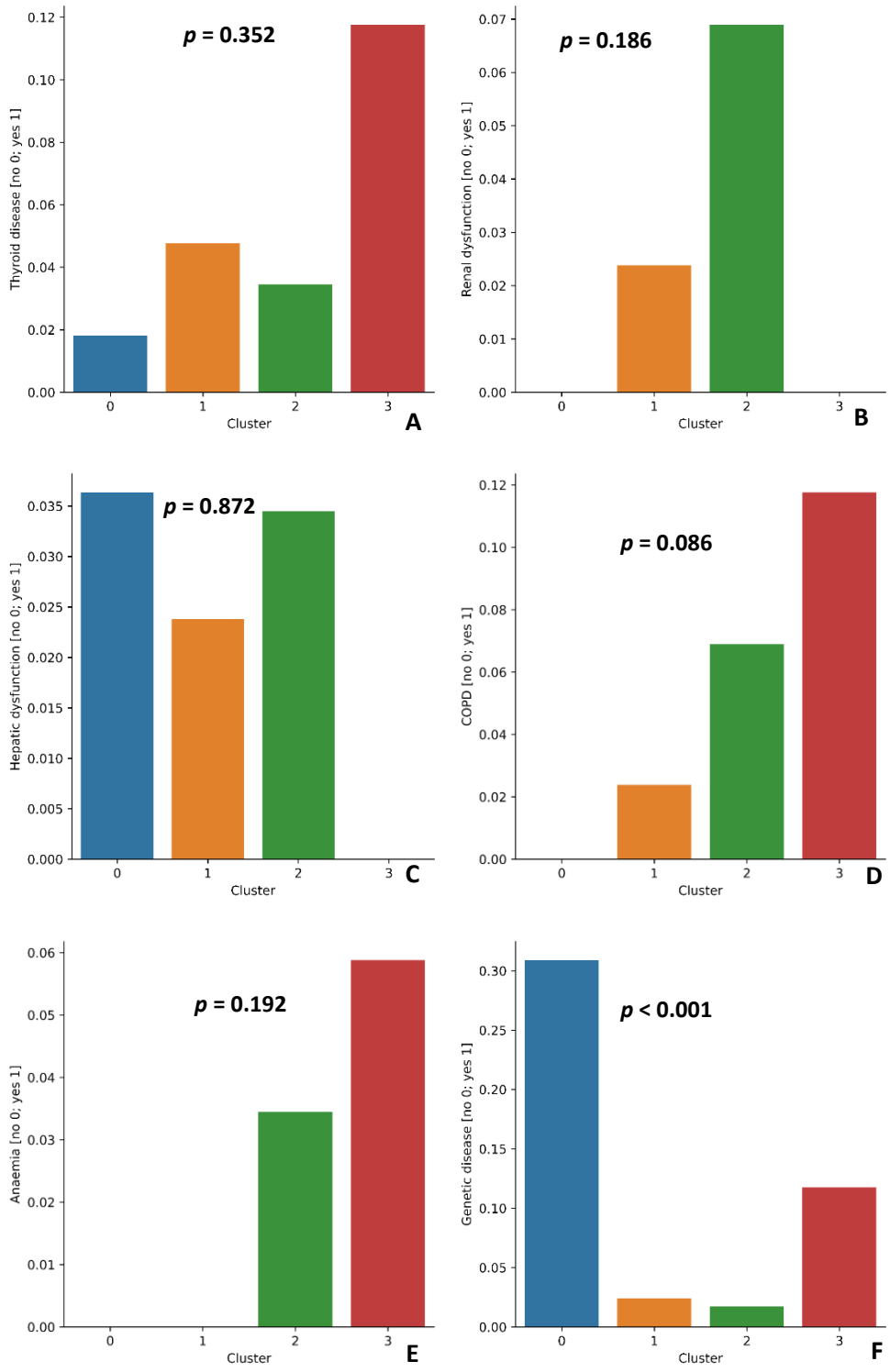

Figure 50. Four clusters setting, features: A — Thyroid disease, B — Renal dysfunction, C — Hepatic dysfunction, D — COPD, E — Anemia, F — Genetic disease (283)

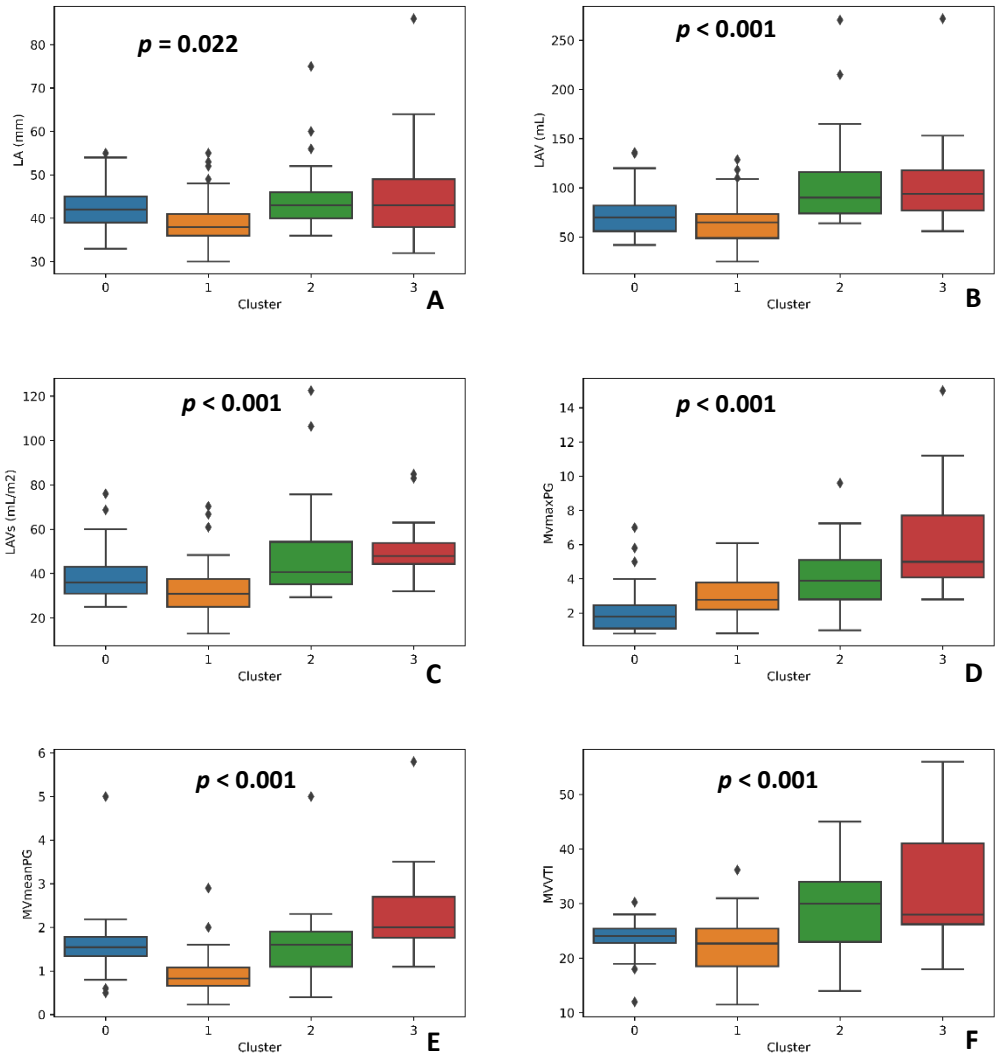

Figure 51. Four clusters setting, features: A — LA, B — LAV, C — LAVs, D — MV maxPG, E — MV meanPG, F — MVVTI (283)

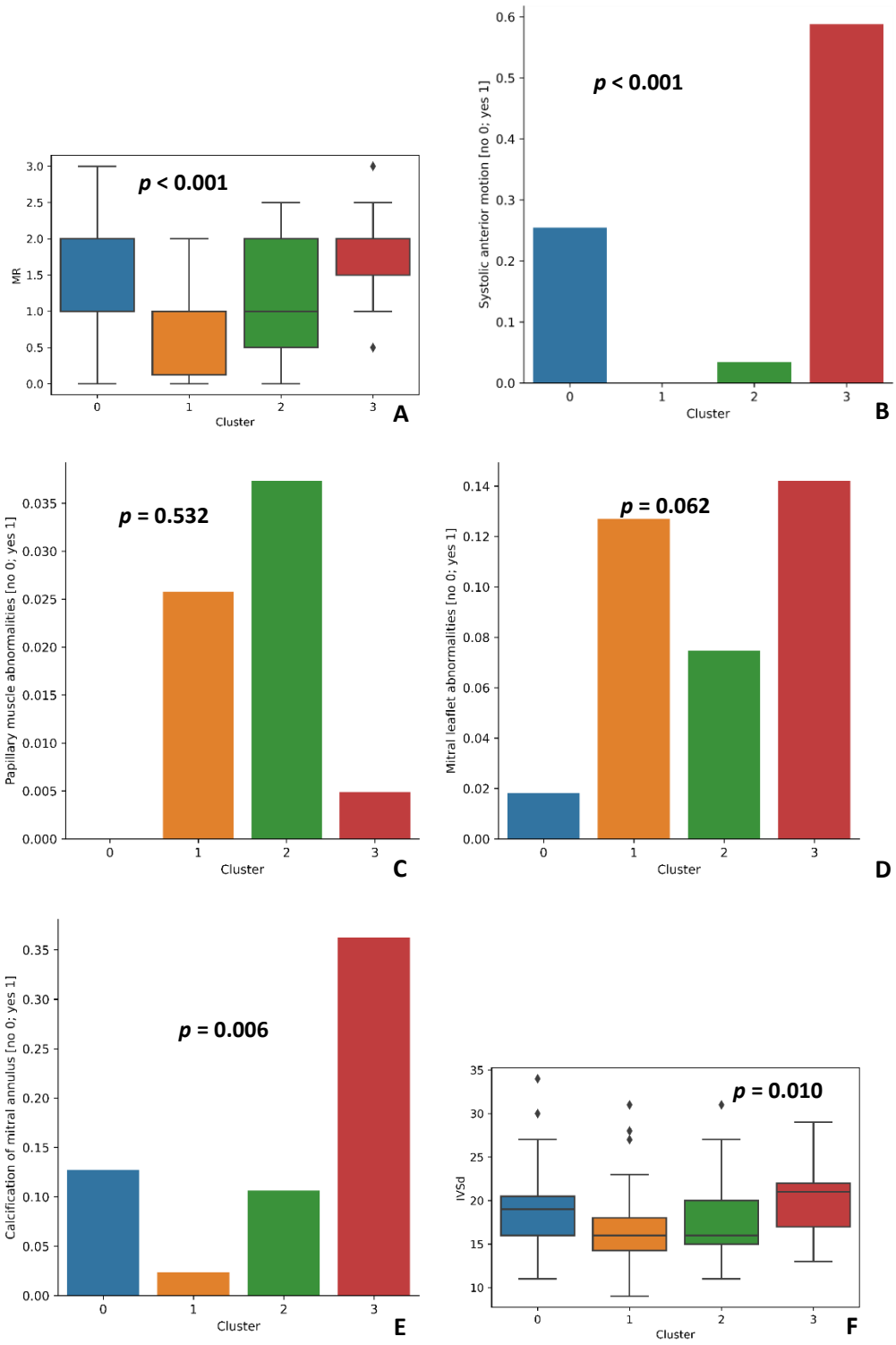

Figure 52. Four clusters setting, features: A — MR, B — SAM, C — Papillary muscle abnormalities, D — Mitral leaflet abnormalities, E — Calcification of mitral annulus, F — IVSd (283)

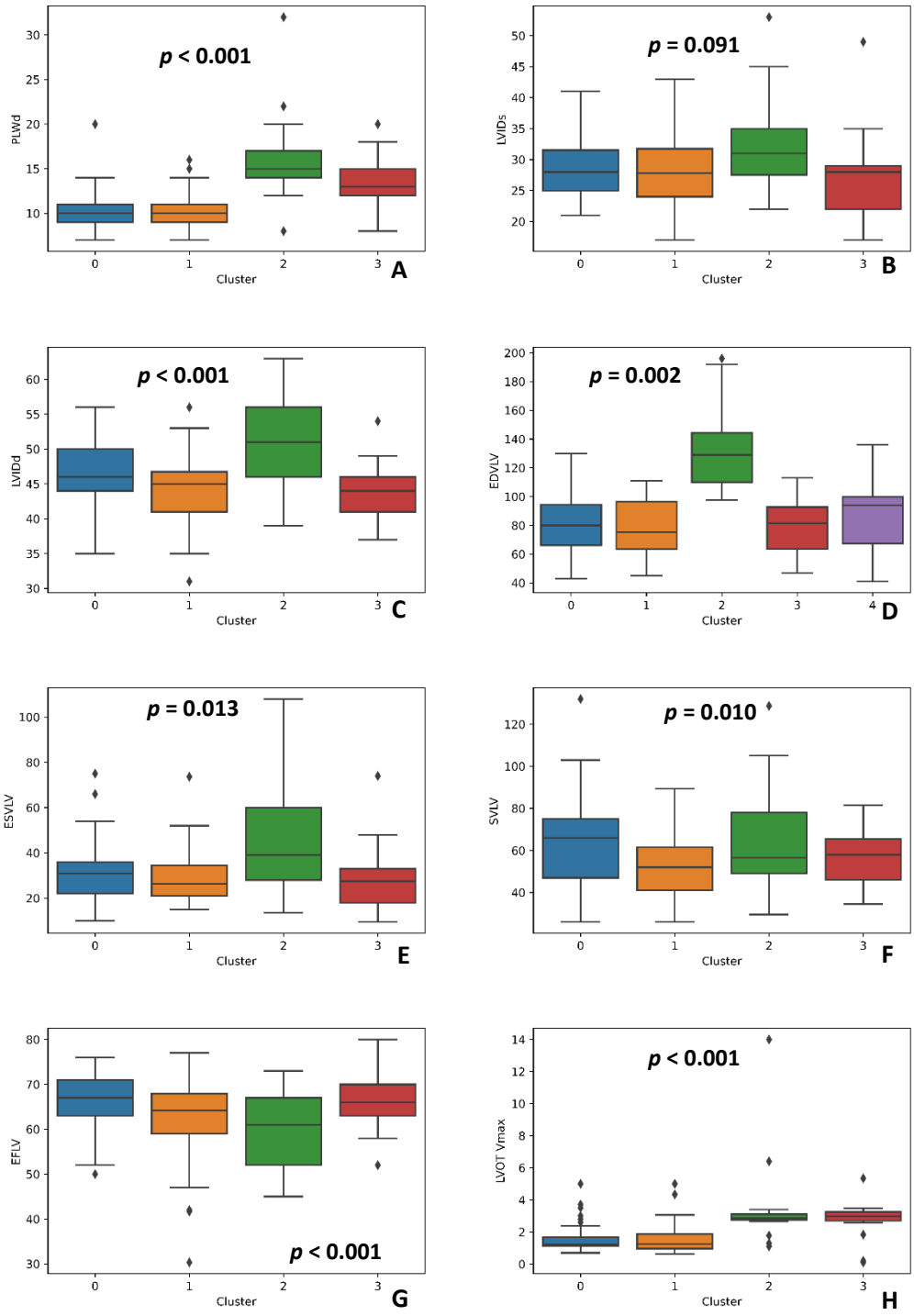

Figure 53. Four clusters setting, features: A — PLWd, B — LVIDs, C — LVIDd, D — EDVLV, E — ESVLV, F — SVLV, G — EFLV, H — LVOT Vmax (283)

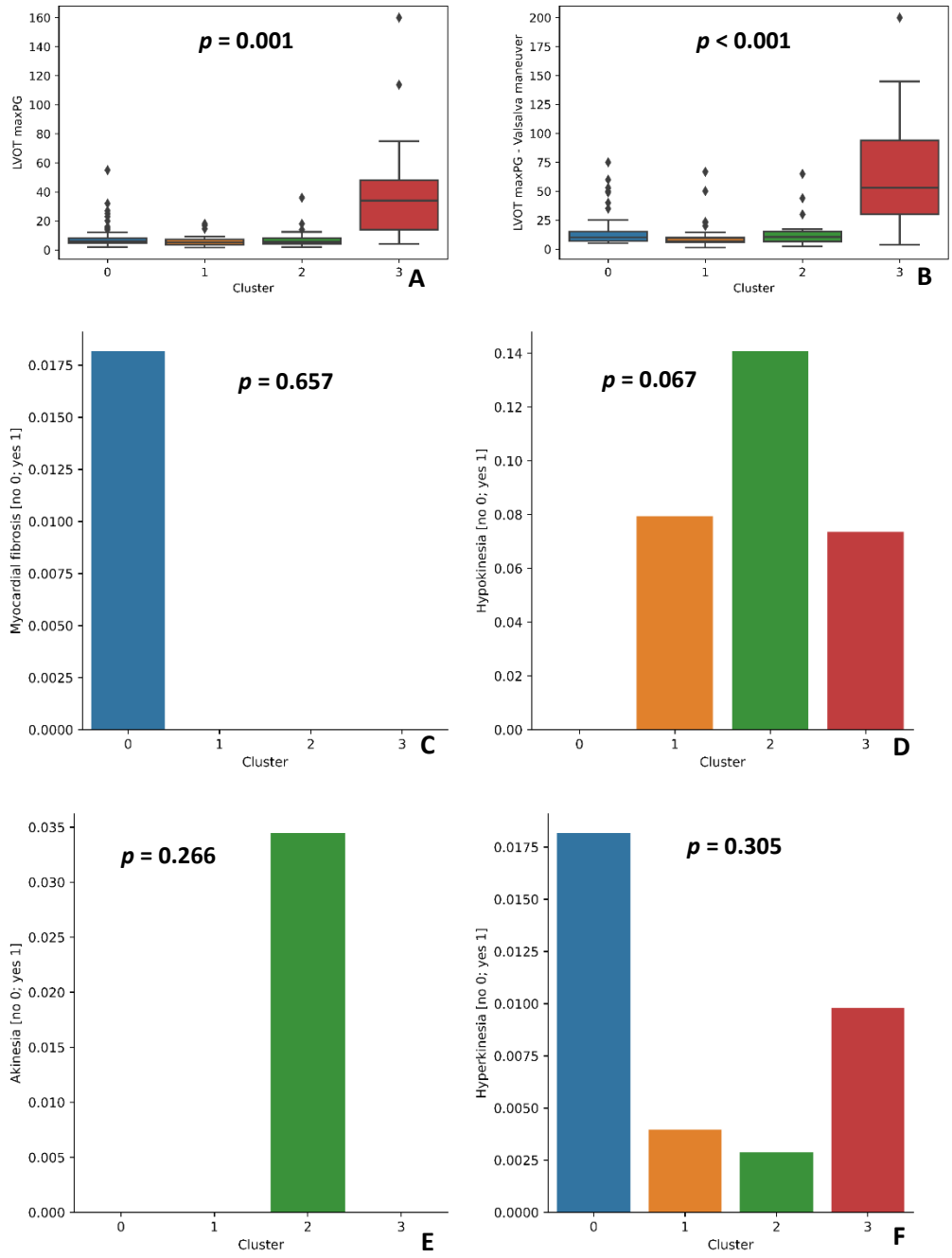

Figure 54. Four clusters setting, features: A — LVOT maxPG, B — LVOT maxPG — Valsalva maneuver, C — Myocardial fibrosis, D — Hypokinesia, E — Akinesia, F — Hyperkinesia (283)

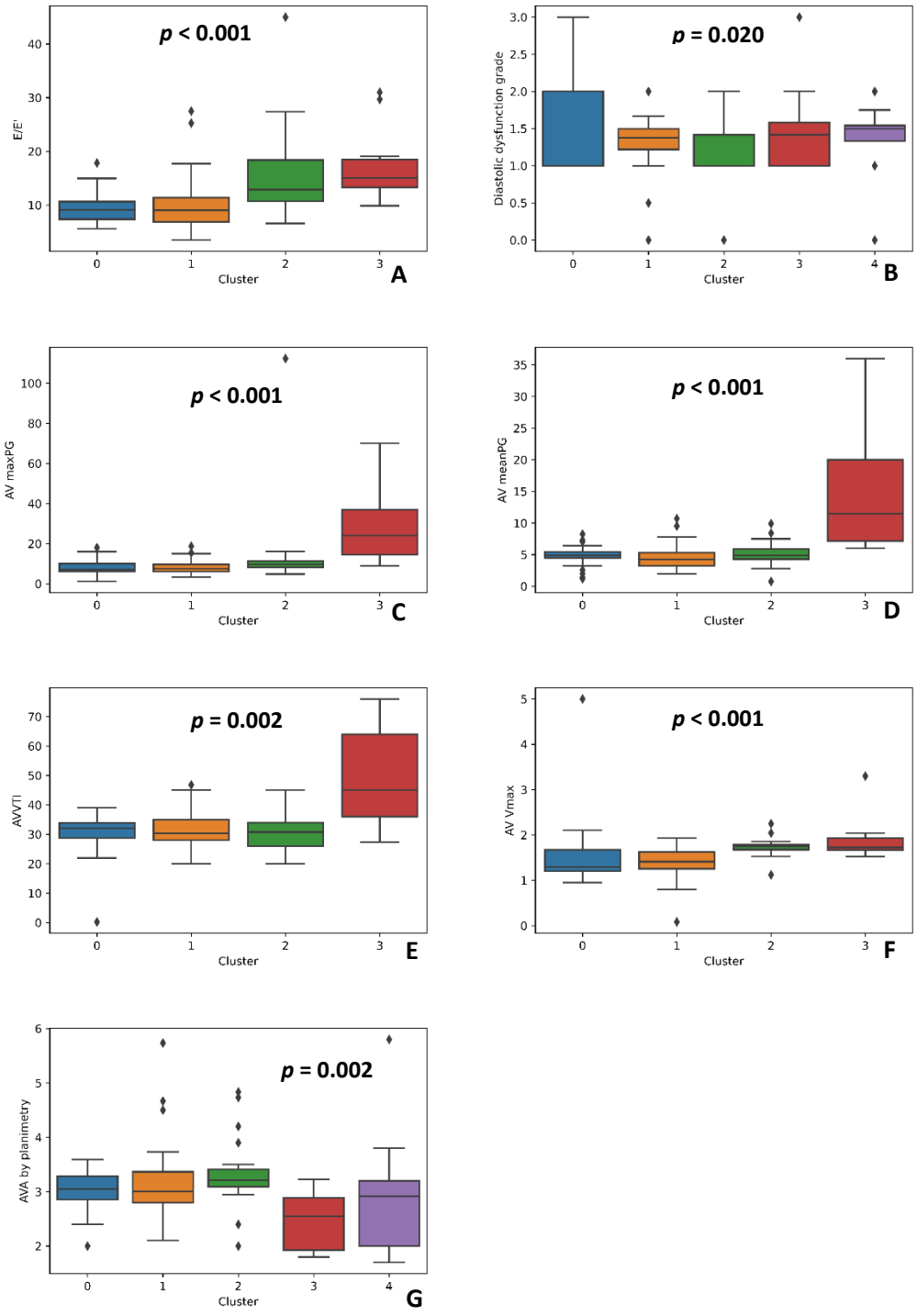

Figure 55. Four clusters setting, features: A — E/E', B — Diastolic dysfunction grade, C — AV maxPG, D — AV meanPG, E — AVVTI, F — AV Vmax, G — AVA by planimetry (283)

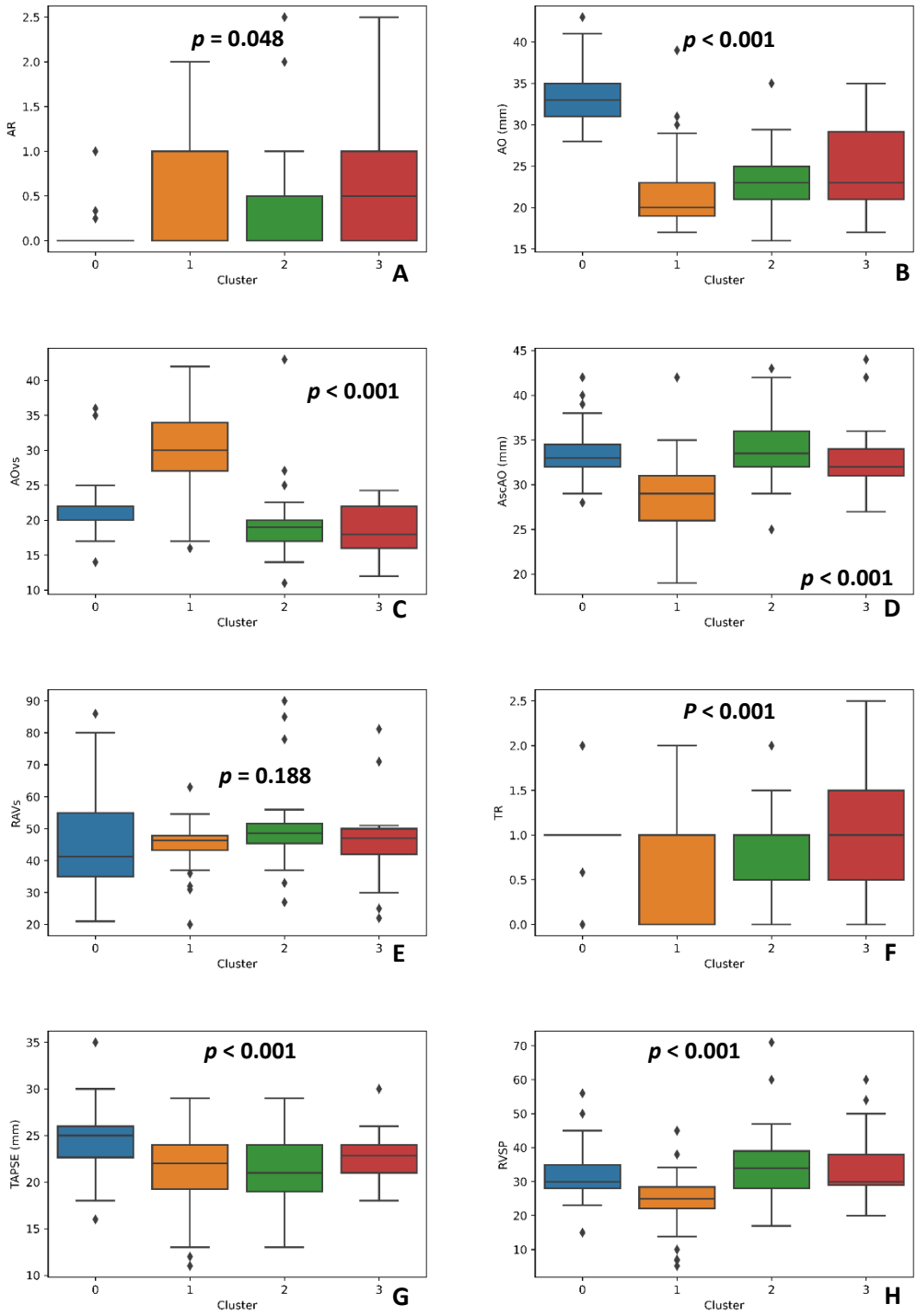

Figure 56. Four clusters setting, features: A — AR, B — AO, C — AOvs, D — AscAO, E — RAVs, F — TR, G — TAPSE, H — RVSP (283)

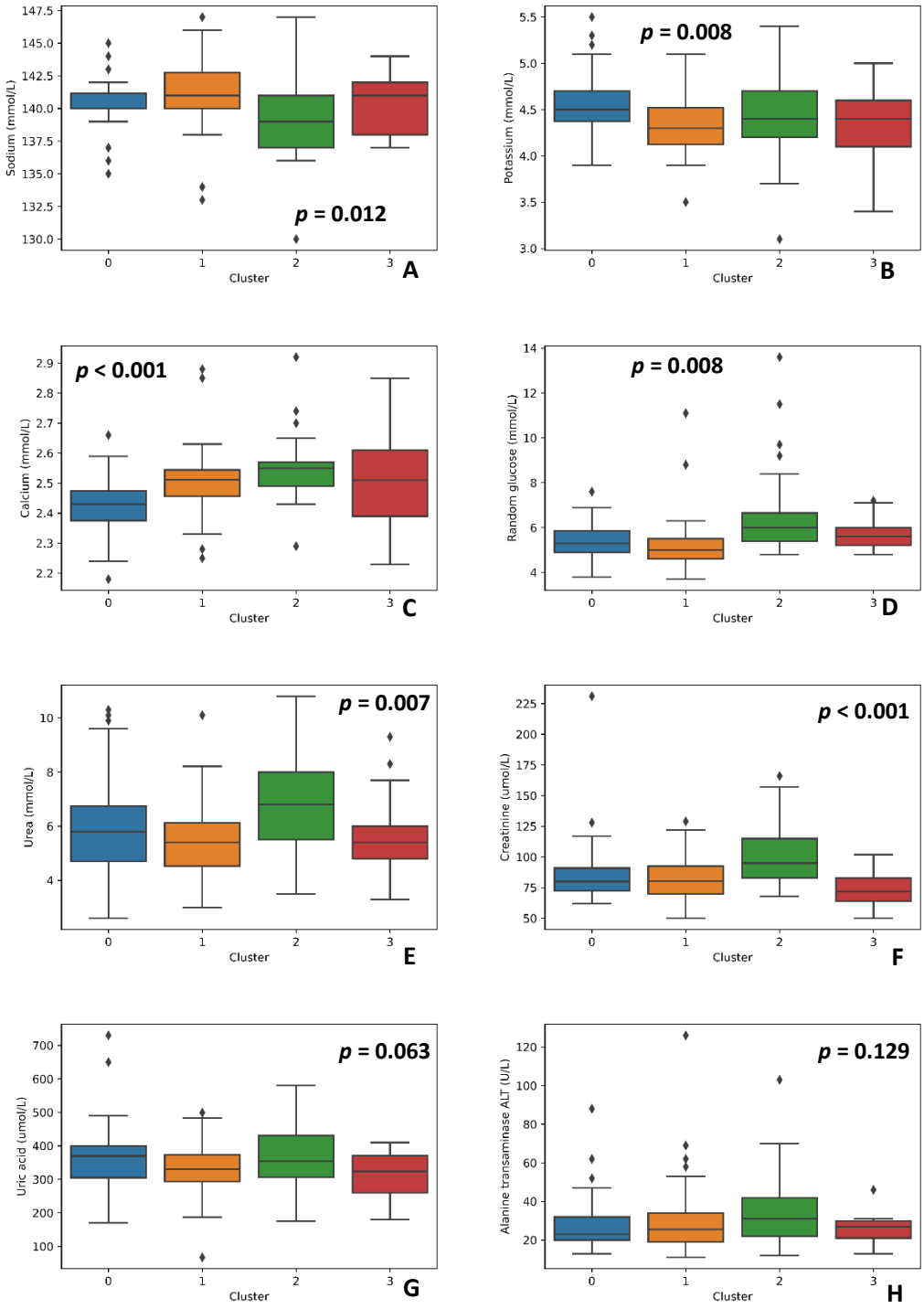

Figure 57. Four clusters setting, features: A — Sodium, B — Potassium, C — Calcium, D — Random glucose, E — Urea, F — Creatinine, G — Uric acid, H — ALT (283)

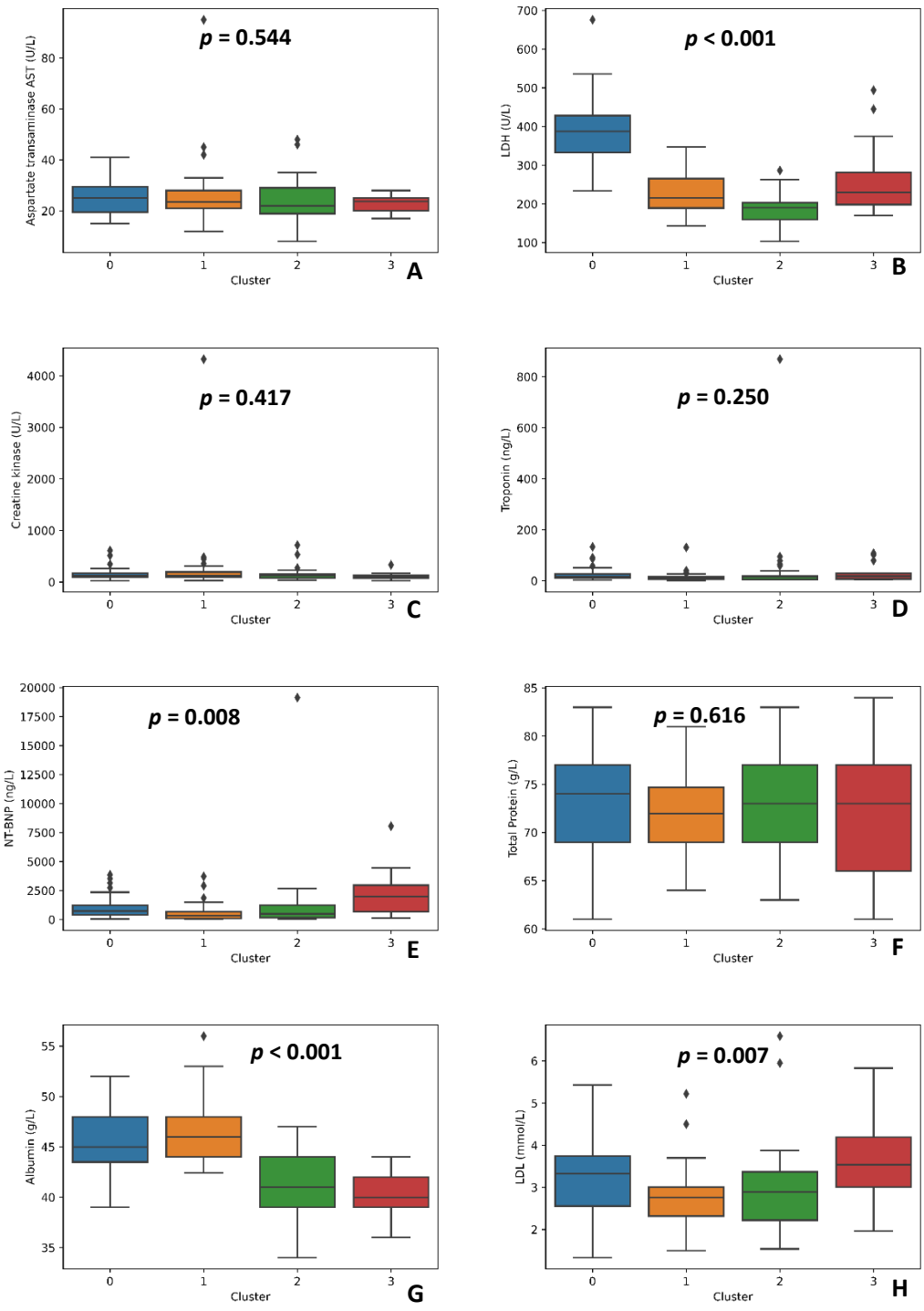

Figure 58. Four clusters setting, features: A — AST, B — LDH, C — Creatine-kinase, D — Troponin, E — NT-BNP, F — Total protein, G — Albumin, H — LDL (283)

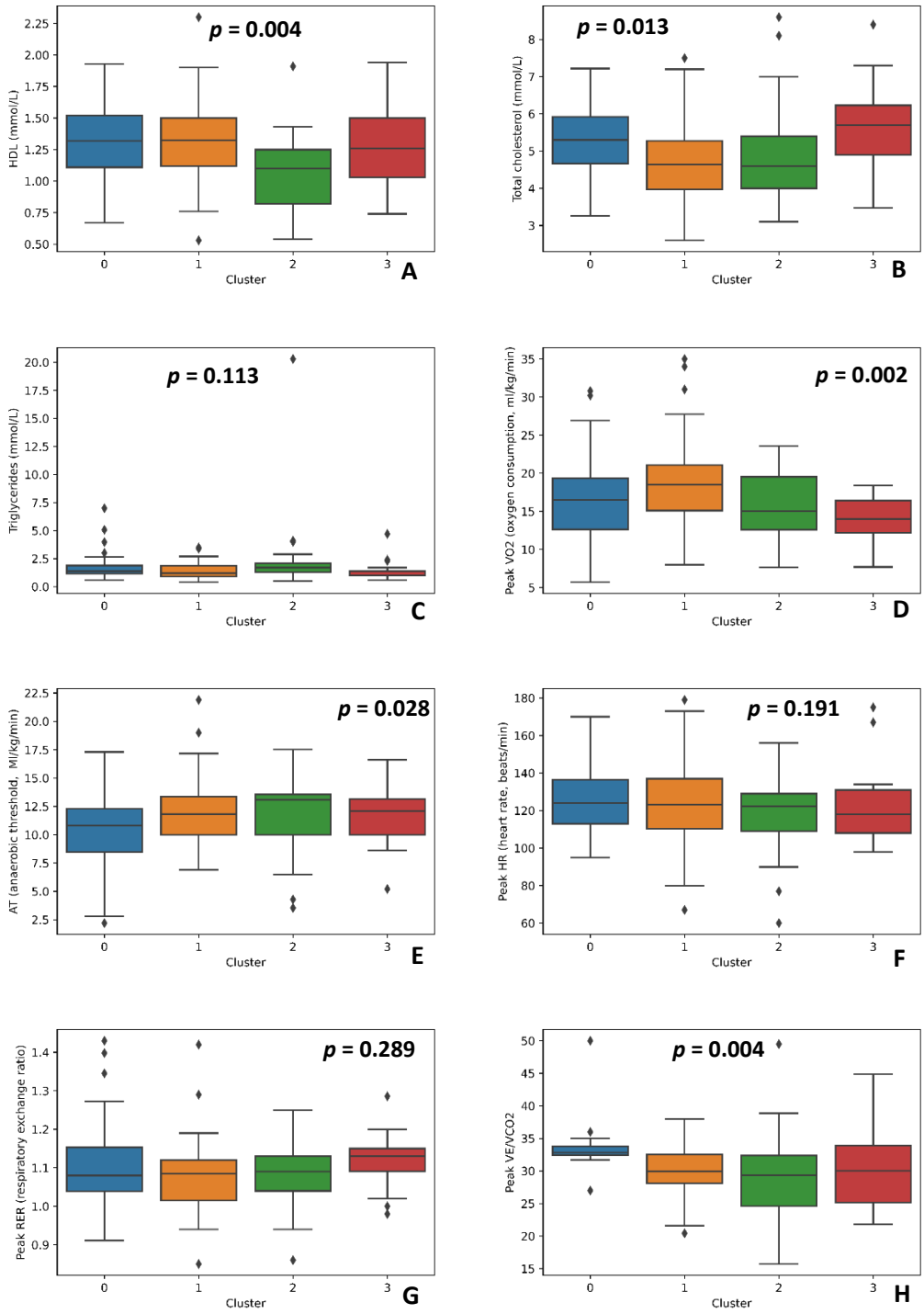

Figure 59. Four clusters setting, features: A — HDL, B — Total cholesterol, C — Triglycerides, D — Peak VO_2, E — AT, F — Peak HR, G — Peak RER, H — Peak VE/VCO_2 (283)

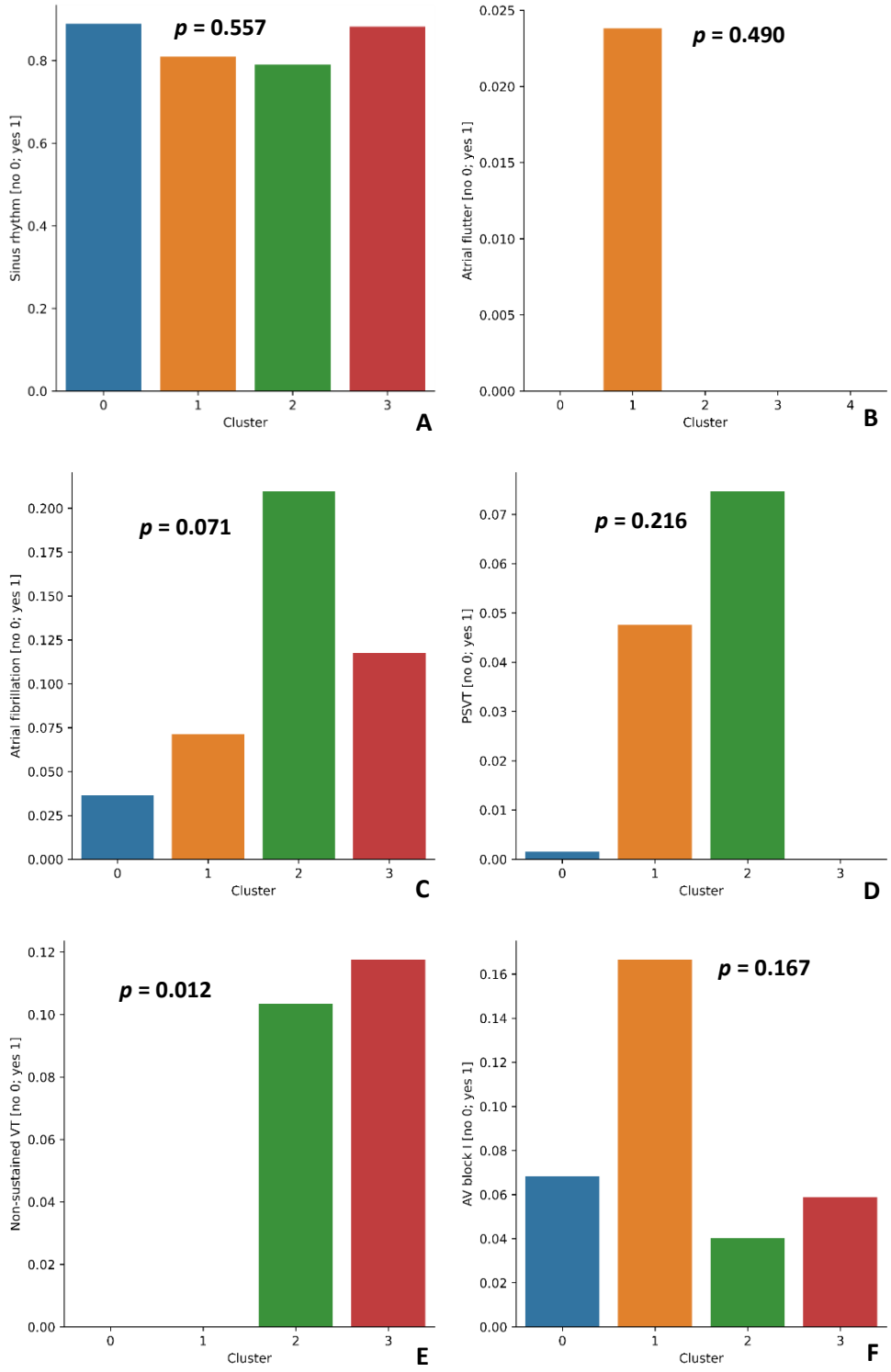

Figure 60. Four clusters setting, features: A — Sinus rhythm, B — Atrial flutter, C — AF, D — PSVT, E — Non-sustained VT, F — AV block I (283)

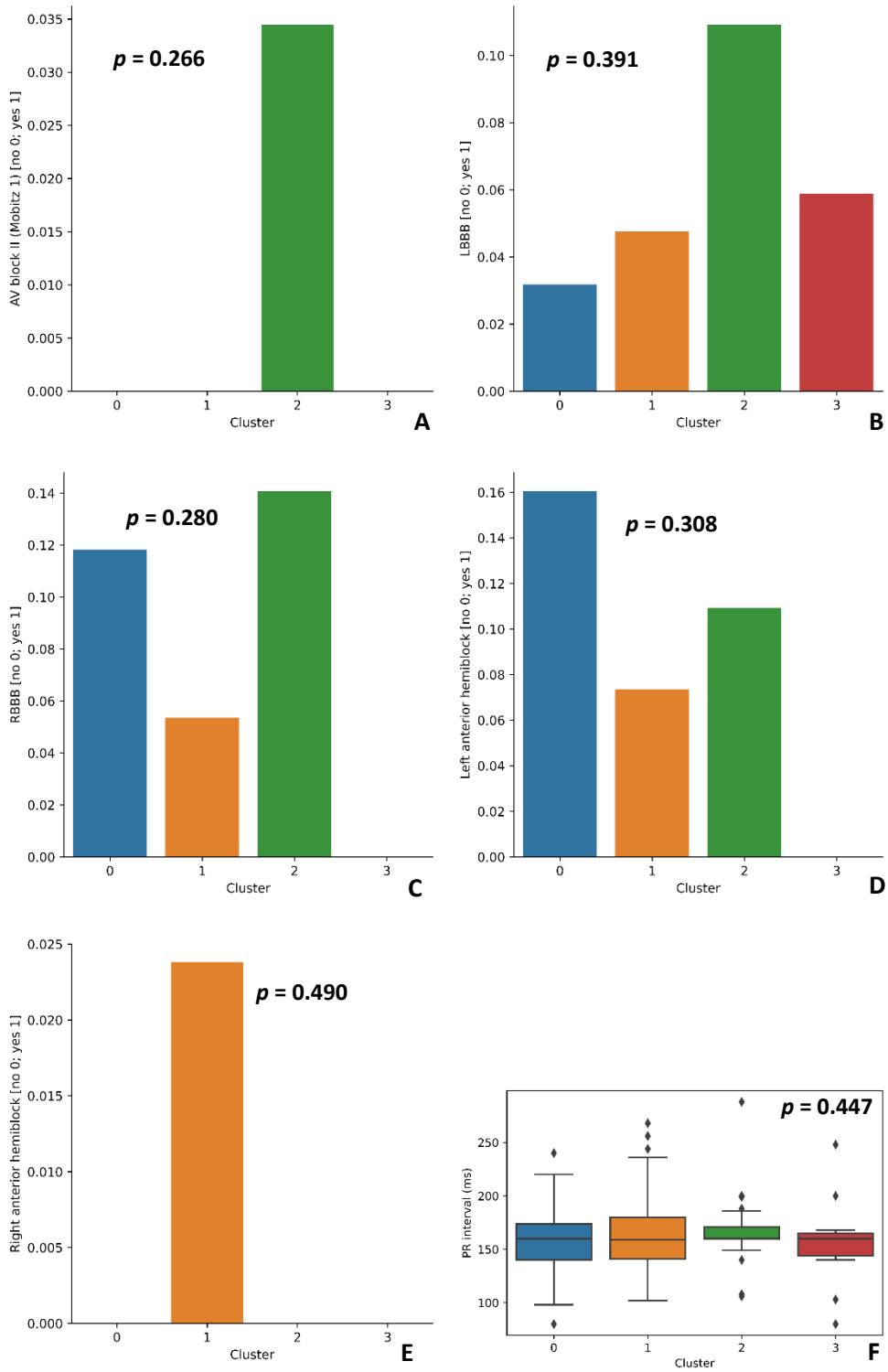

Figure 61. Four clusters setting, features: A — AV block II (Mobitz 1), B — LBBB, C — RBBB, D — Left anterior hemiblock, E — Right anterior hemiblock, F — PR interval (283)

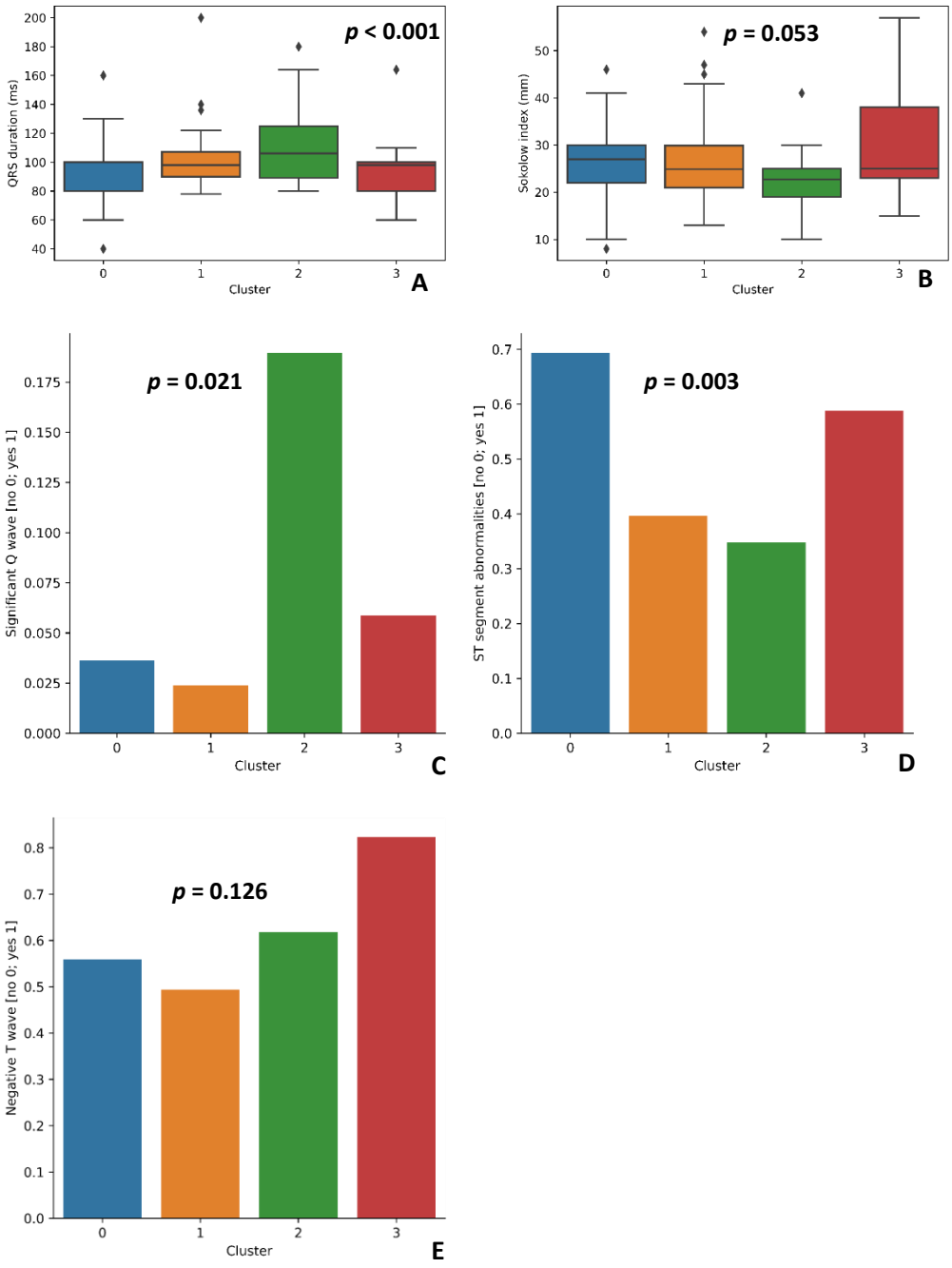

Figure 62. Four clusters setting, features: A — QRS duration, B — Sokolow index, C — Significant Q wave, D — ST segment abnormalities, E — Negative T wave (283)

An approximate interpretation of clustering-logic for 4 clusters is presented in the form of visualized decision tree (Figure 63). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by clustering (Table 12).

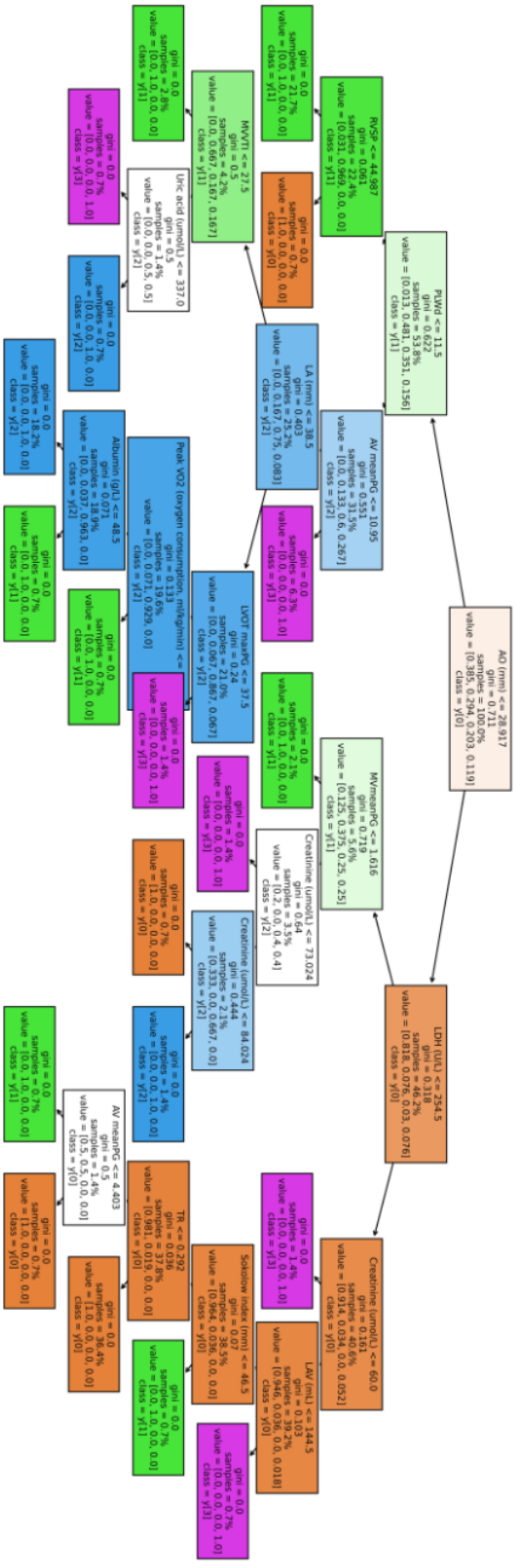

Figure 63. An approximate interpretation of clustering-logic for 4 clusters

Table 12. Feature importance — Top 45 features for distinguishing 4 clusters

Feature	Estimated importance
LDH	0.072933
AO	0.063513
AOvs	0.053407
PLWd	0.042644
LVOT Vmax	0.039411
MV meanPG	0.034072
MV maxPG	0.032063
Peak VE/VCO_2	0.031522
Heart murmur [yes/no]	0.027068
AV maxPG	0.026379
AscAO	0.024348
HCM in family history [yes/no]	0.024280
Serum albumin	0.023722
Weight	0.017189
LVOT maxPG	0.017161
MVVTI	0.015820
AV meanPG	0.014308
RVSP	0.014291
NYHA class	0.013766
AV Vmax	0.013508
LA	0.013420
AVVTI	0.013285
BMI	0.013265
LAV	0.013225
Diastolic	0.013080
LAVs	0.012232
QRS duration	0.011427
LVIDd	0.011076
Serum calcium	0.010857
NT-BNP	0.009222
AVA by planimetry	0.009159
E/E'	0.009069
RAVs	0.008986
EFLV	0.008804
Serum creatinine	0.008517
LVOT maxPG - Valsalva maneuver	0.008415
Heart rate	0.008279
LDL	0.007798
Age	0.007743
IVSd	0.007711
Sokolow index	0.007636
MR	0.007281
Peak HR	0.006946
Anaerobic threshold	0.006793
ESVLV	0.006754

4.1.2.2.4. Five clusters

Characteristics of 5 clusters determined are shown in Figures 64-81.

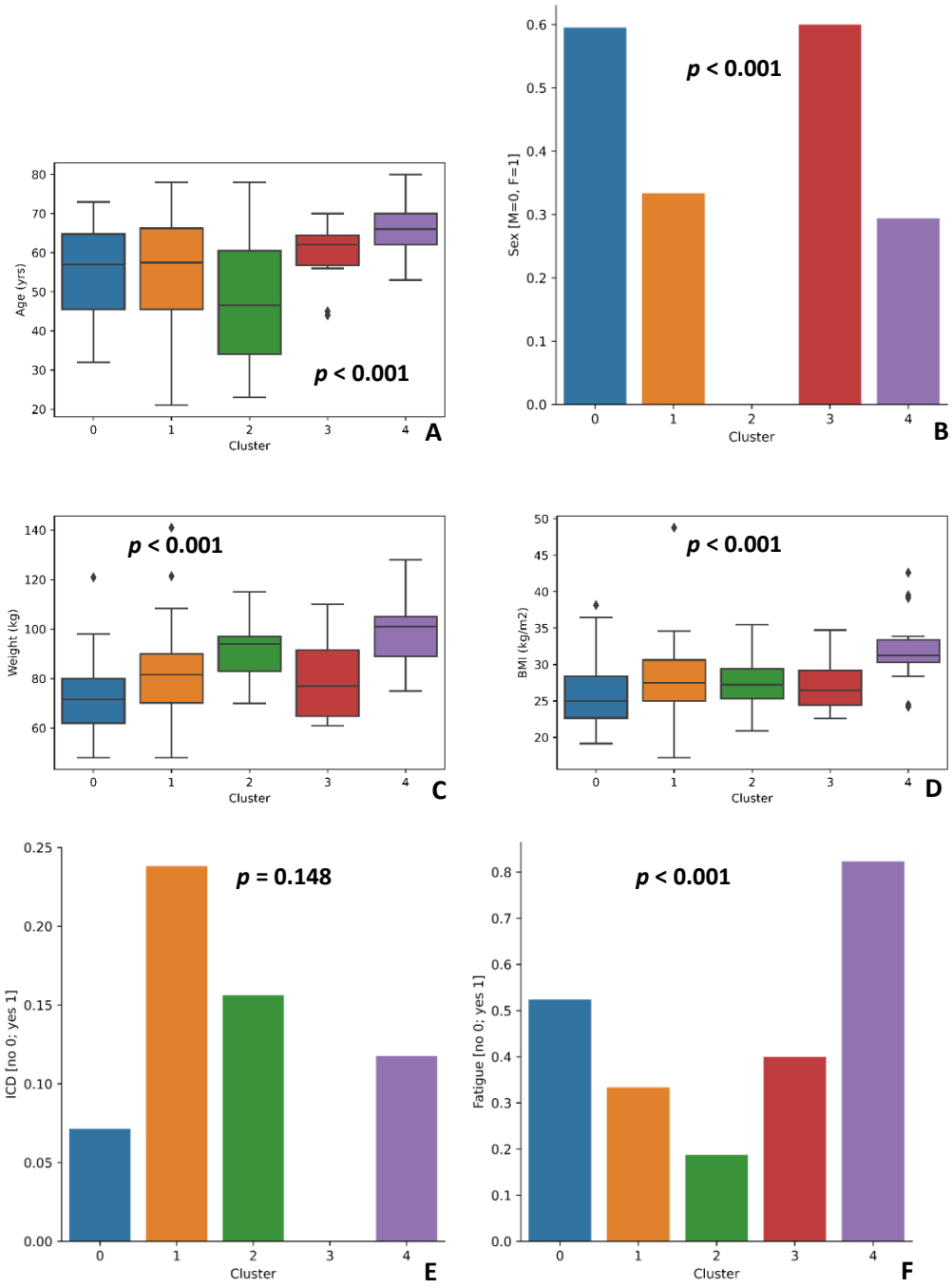

Figure 64. Five clusters setting, features: A — Age, B — Sex, C — Weight, D — BMI, E — ICD, F — Fatigue

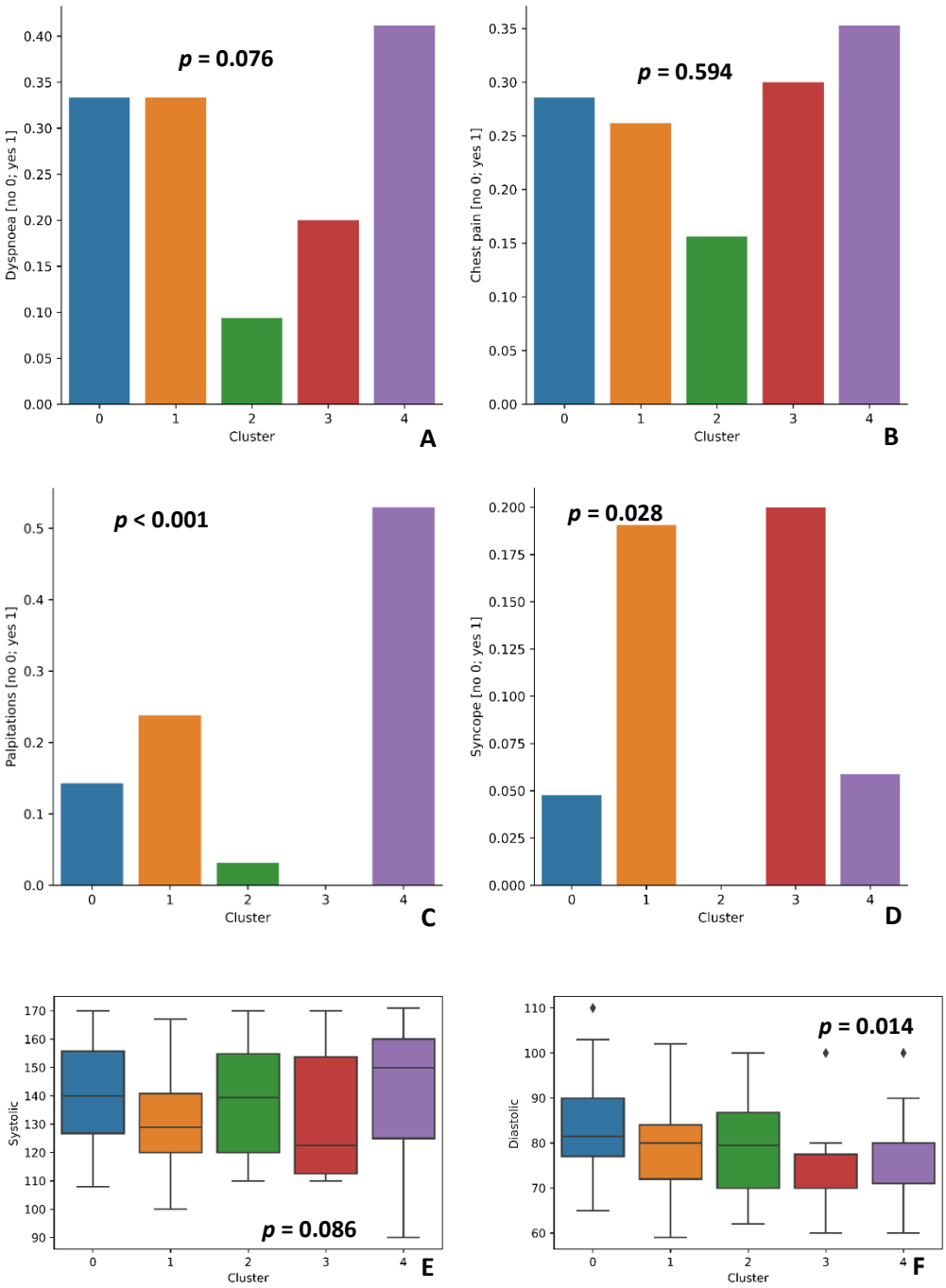

Figure 65. Five clusters setting, features: A — Dyspnea, B — Chest pain, C — Palpitations, D — Syncope, E — Systolic blood pressure, F — Diastolic blood pressure

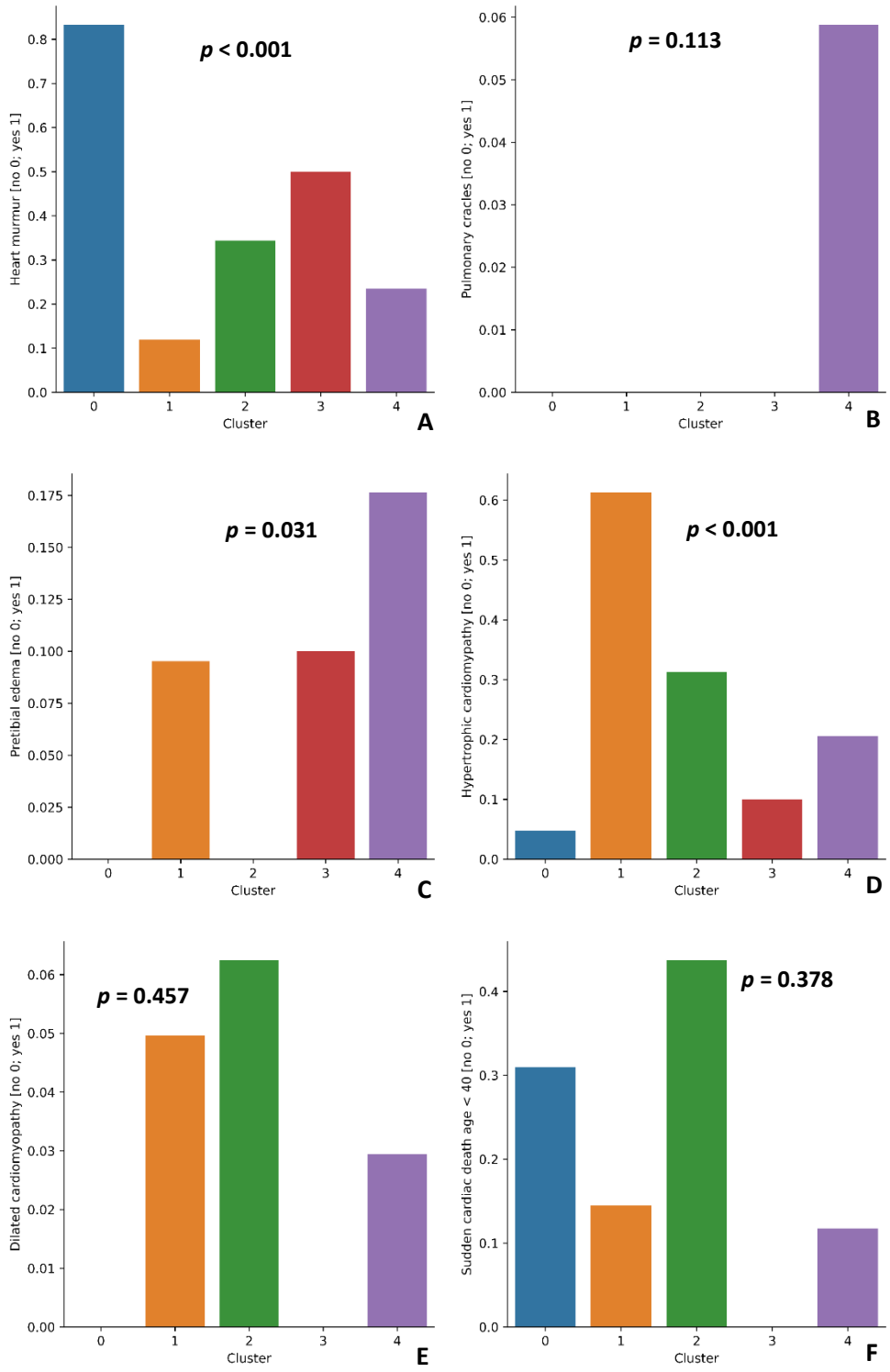

Figure 66. Five clusters setting, features: A — Heart murmur, B — Pulmonary crackles, C — Pretibial edema, D — HCM in family history, E — DCM in family history, F — SCD in age < 40 in family history

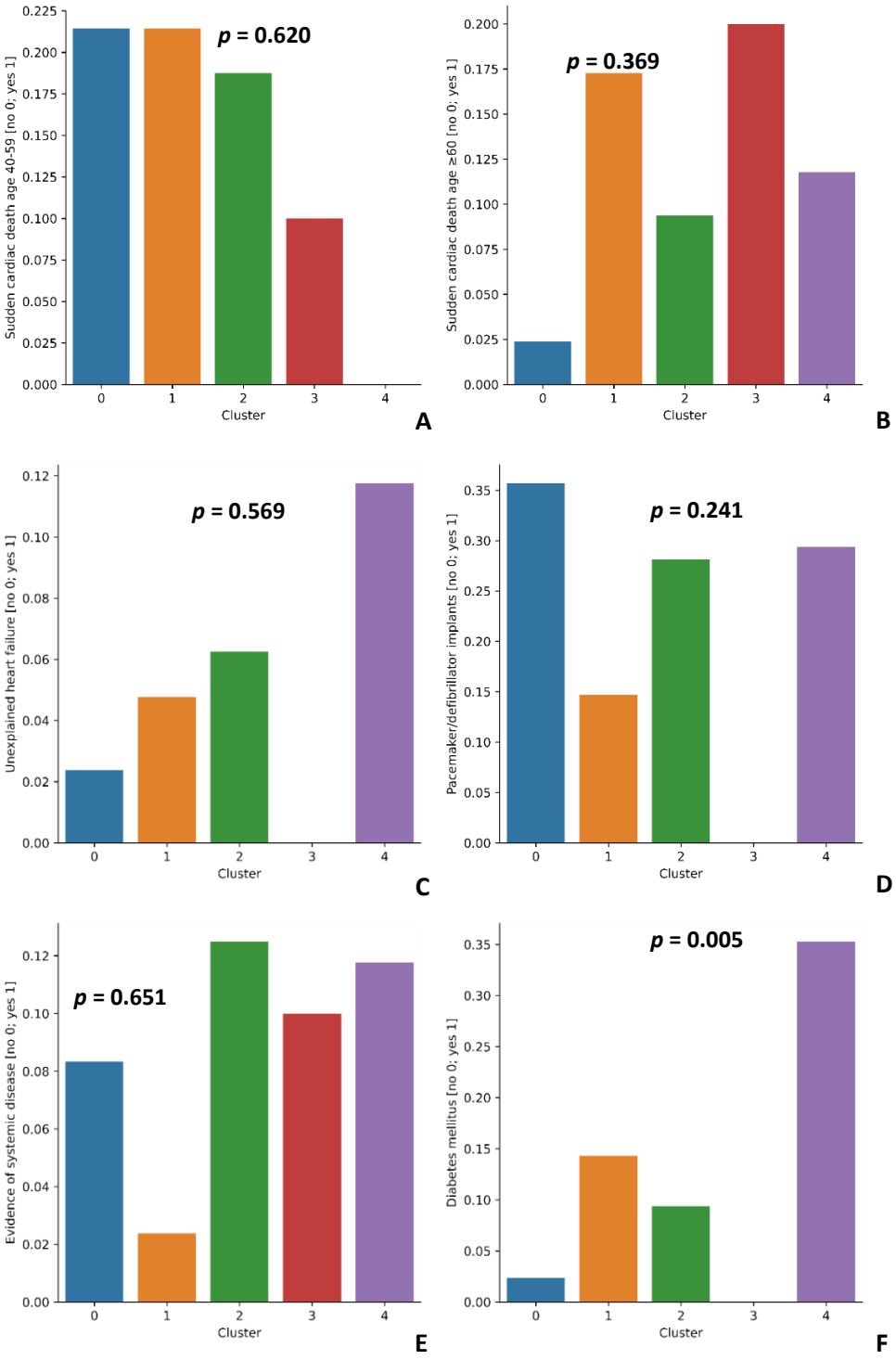

Figure 67. Five clusters setting, features: A — SCD in age 40-59 in family history, B — SCD in age ≥ 60 in family history, C — Unexplained HF in family history, D — Pacemaker/defibrillator implants in family history, E — Evidence of systemic disease in family history, F — Diabetes mellitus

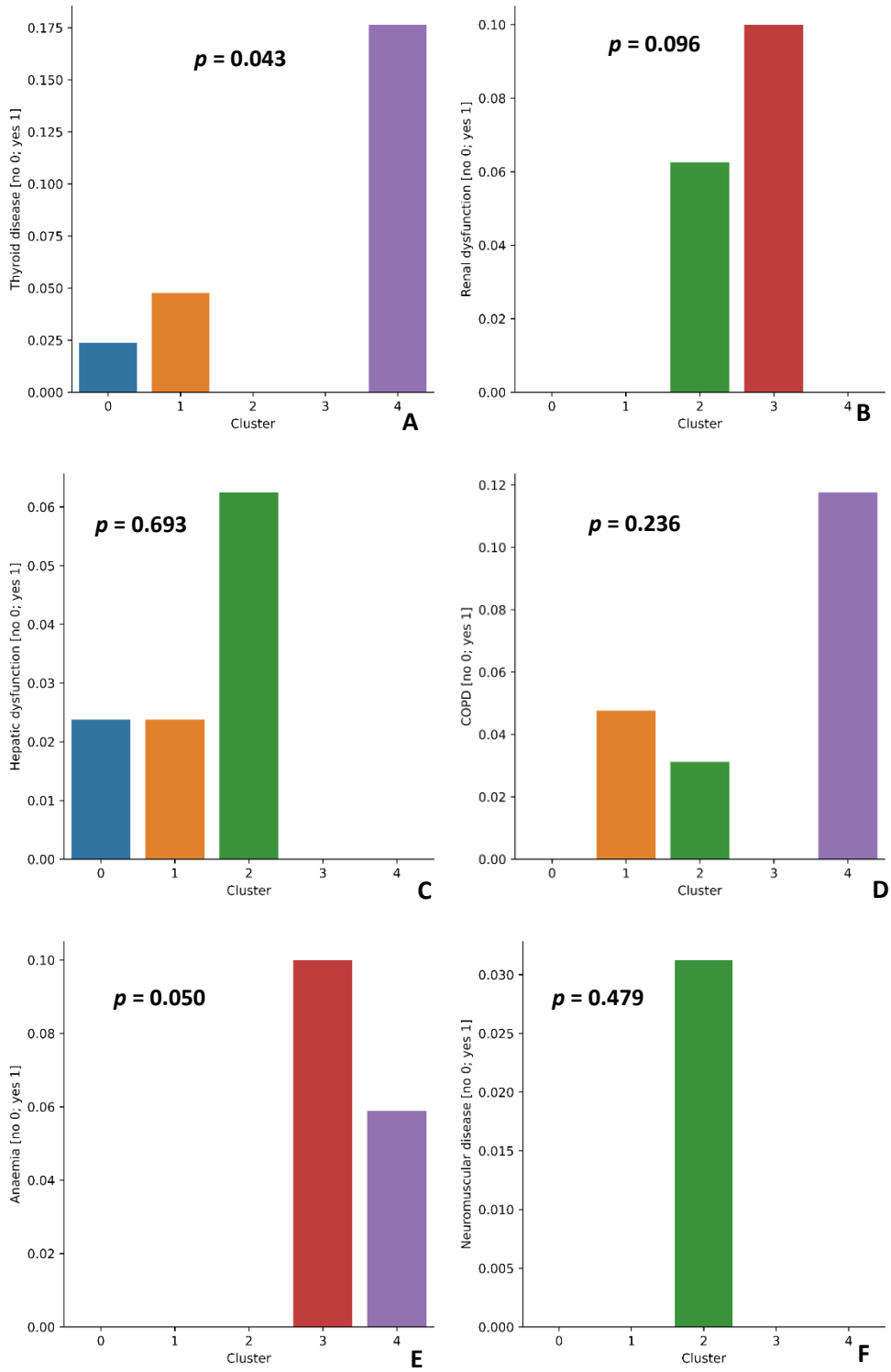

Figure 68. Five clusters setting, features: A — Thyroid disease, B — Renal dysfunction, C — Hepatic dysfunction, D — COPD, E — Anemia, F — Neuromuscular disease

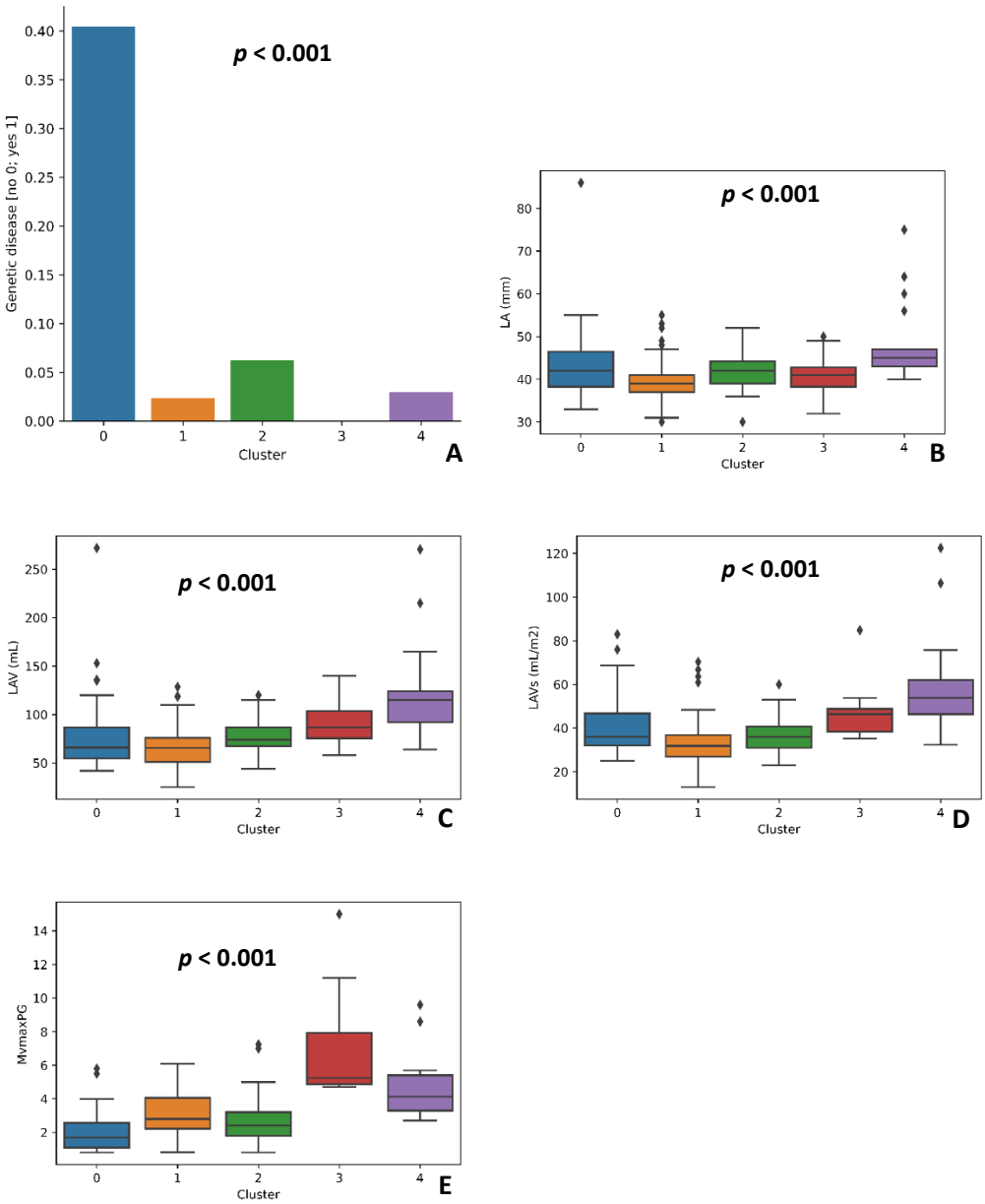

Figure 69. Five clusters setting, features: A — Genetic disease, B — LA, C — LAV, D — LAVs, E — MV maxPG

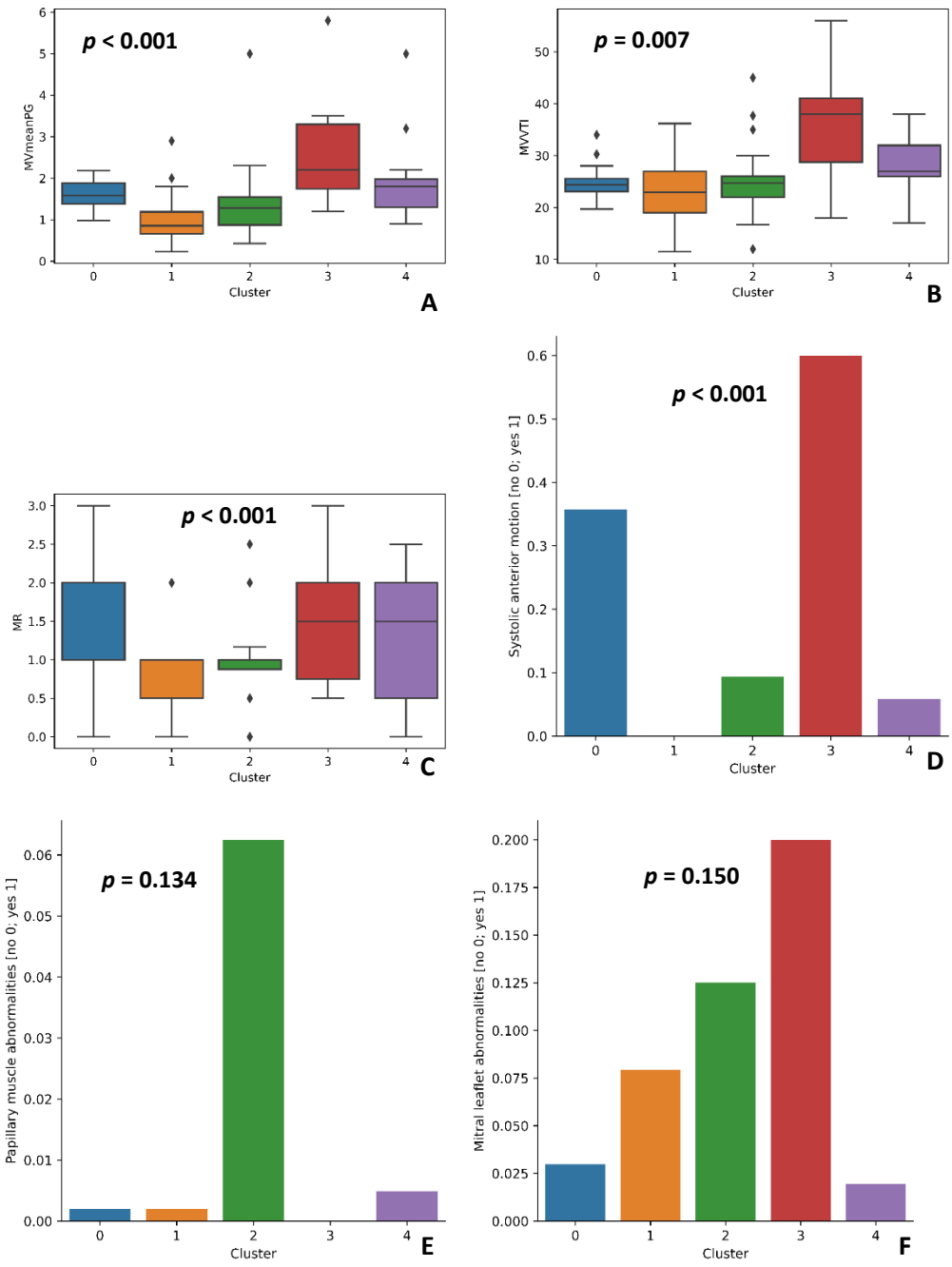

Figure 70. Five clusters setting, features: A — MV meanPG, B — MVVTI, C — MR, D — SAM, E — Papillary muscle abnormalities, F — Mitral leaflet abnormalities

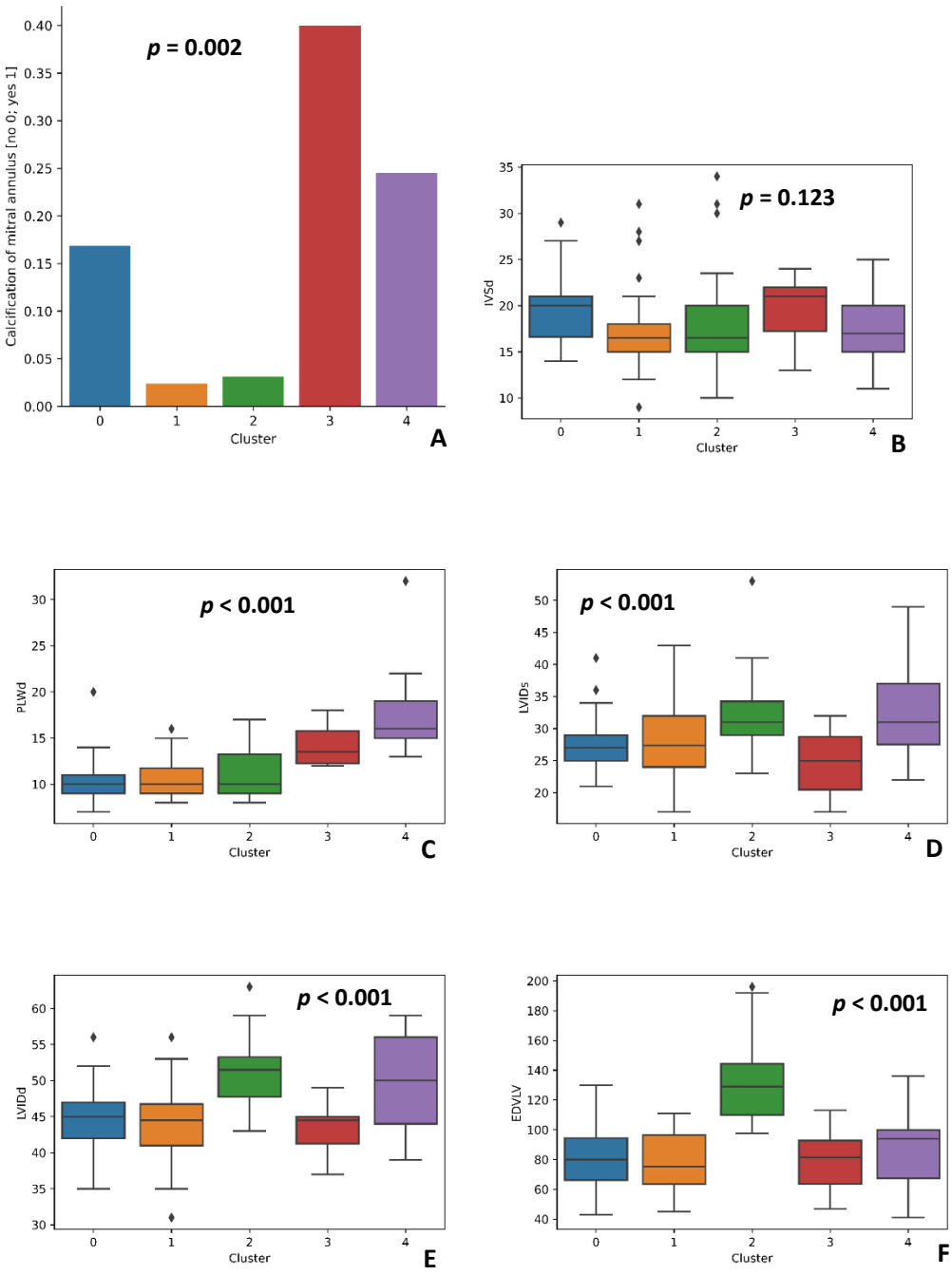

Figure 71. Five clusters setting, features: A — Calcification of mitral annulus, B — IVSd, C — PLWd, D — LVIDs, E — LVIDd, F — EDVLV

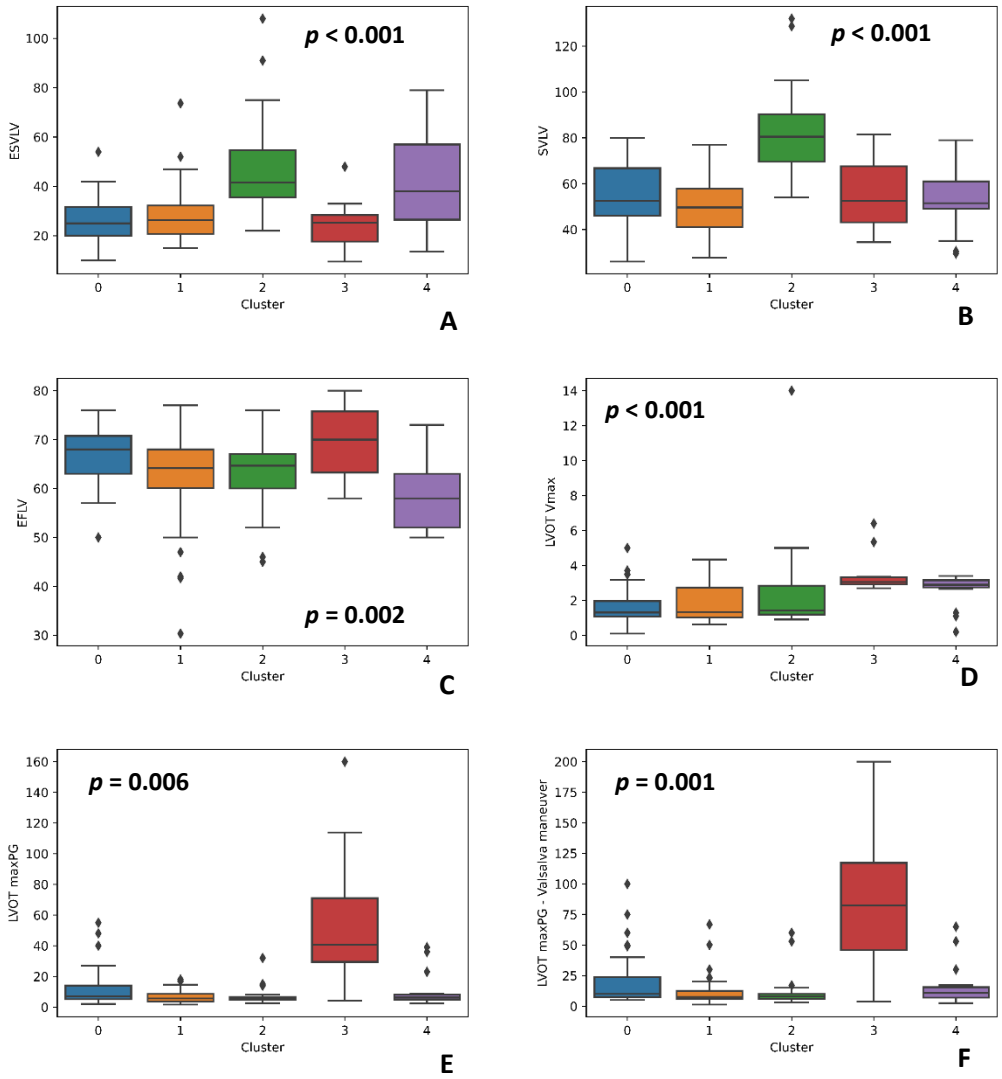

Figure 72. Five clusters setting, features: A — ESVLV, B — SVLV, C — EFLV, D — LVOT Vmax, E — LVOT maxPG, F — LVOT maxPG - Valsalva maneuver

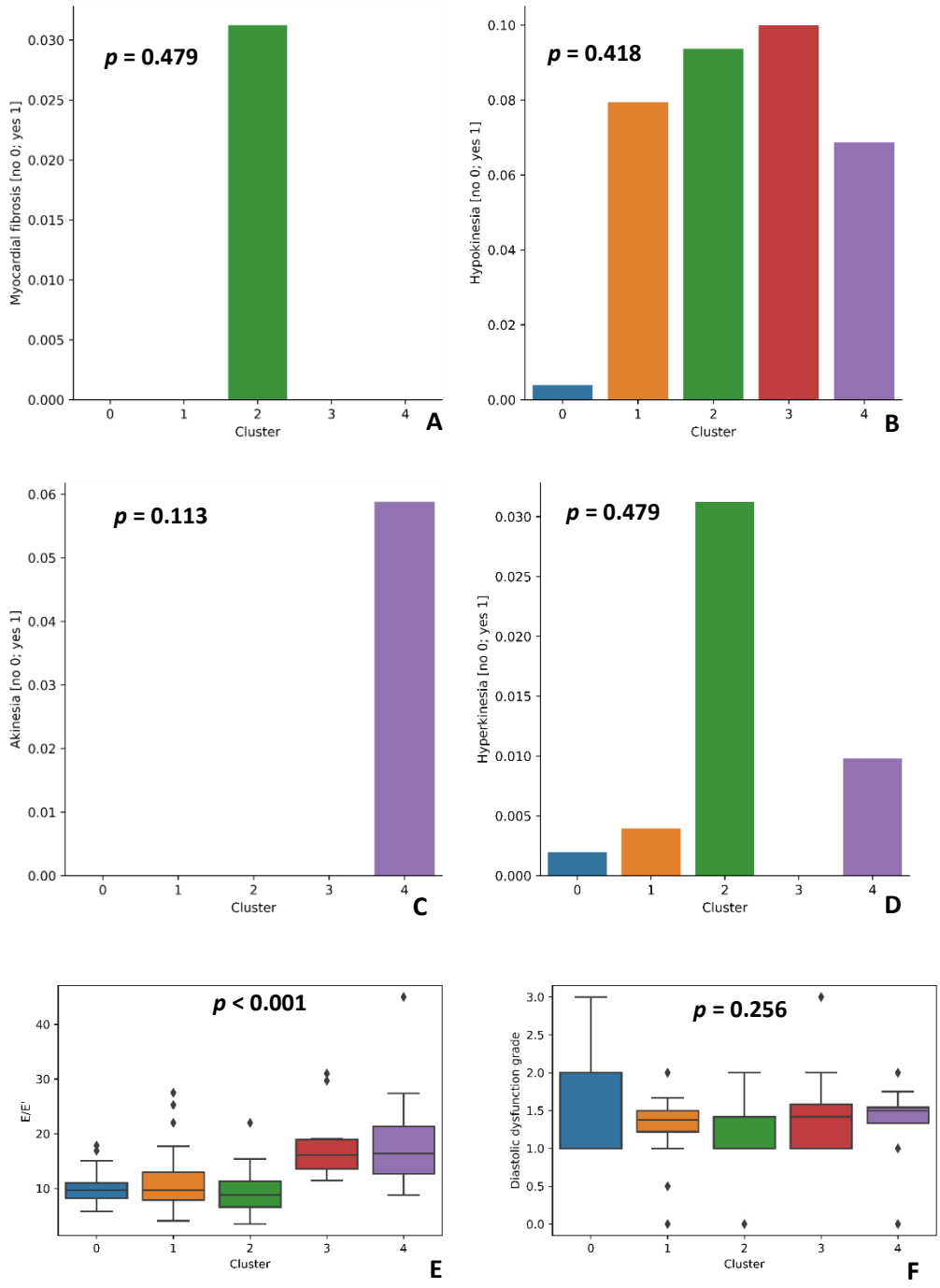

Figure 73. Five clusters setting, features: A — Myocardial fibrosis, B — Hypokinesia, C — Akinesia, D — Hyperkinesia, E — E/E', F — Diastolic dysfunction grade

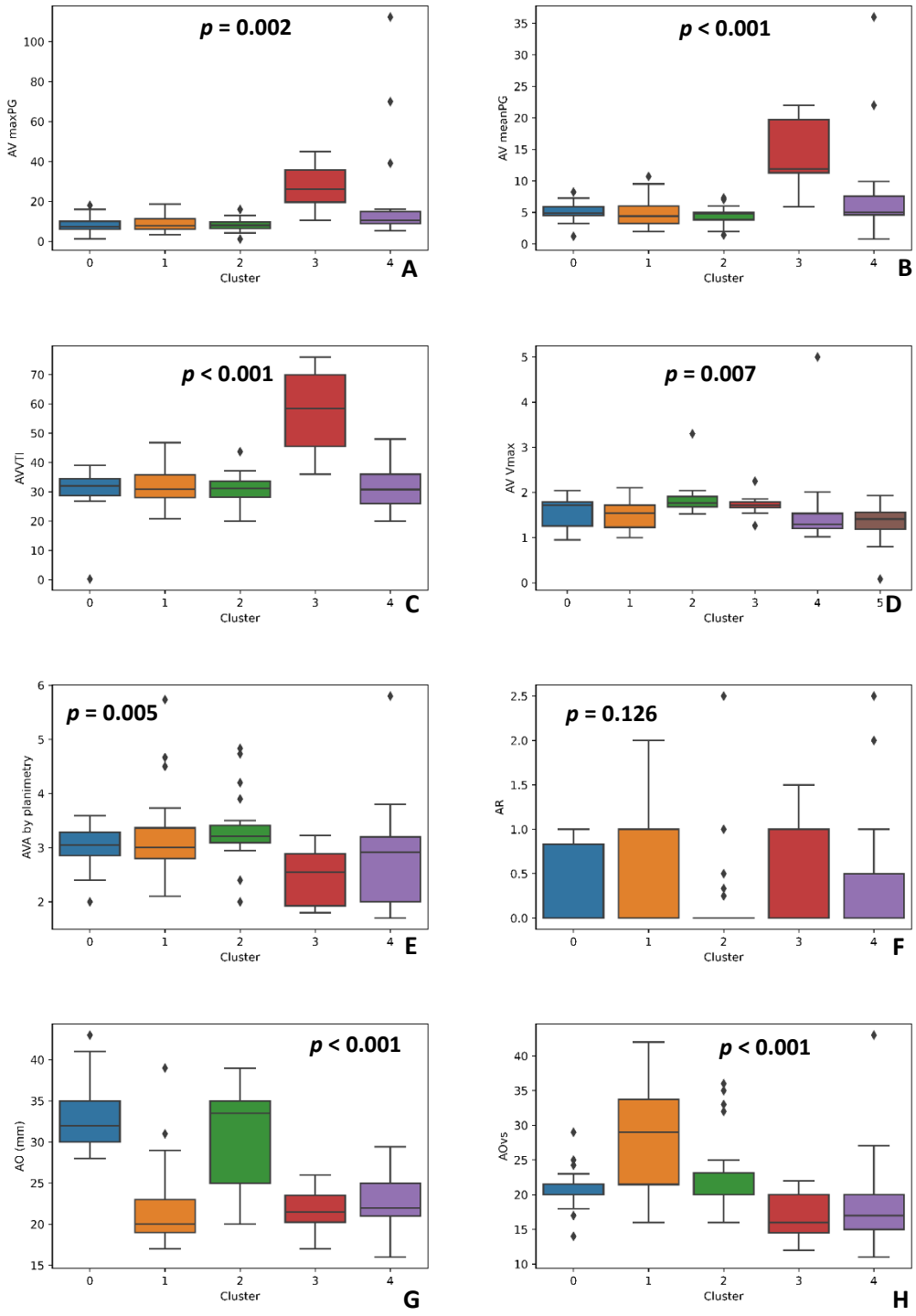

Figure 74. Five clusters setting, features: A — AV maxPG, B — AV meanPG, C — AVVTI, D — AV Vmax, E — AVA by planimetry, F — AR, G — AO, H — AOvs

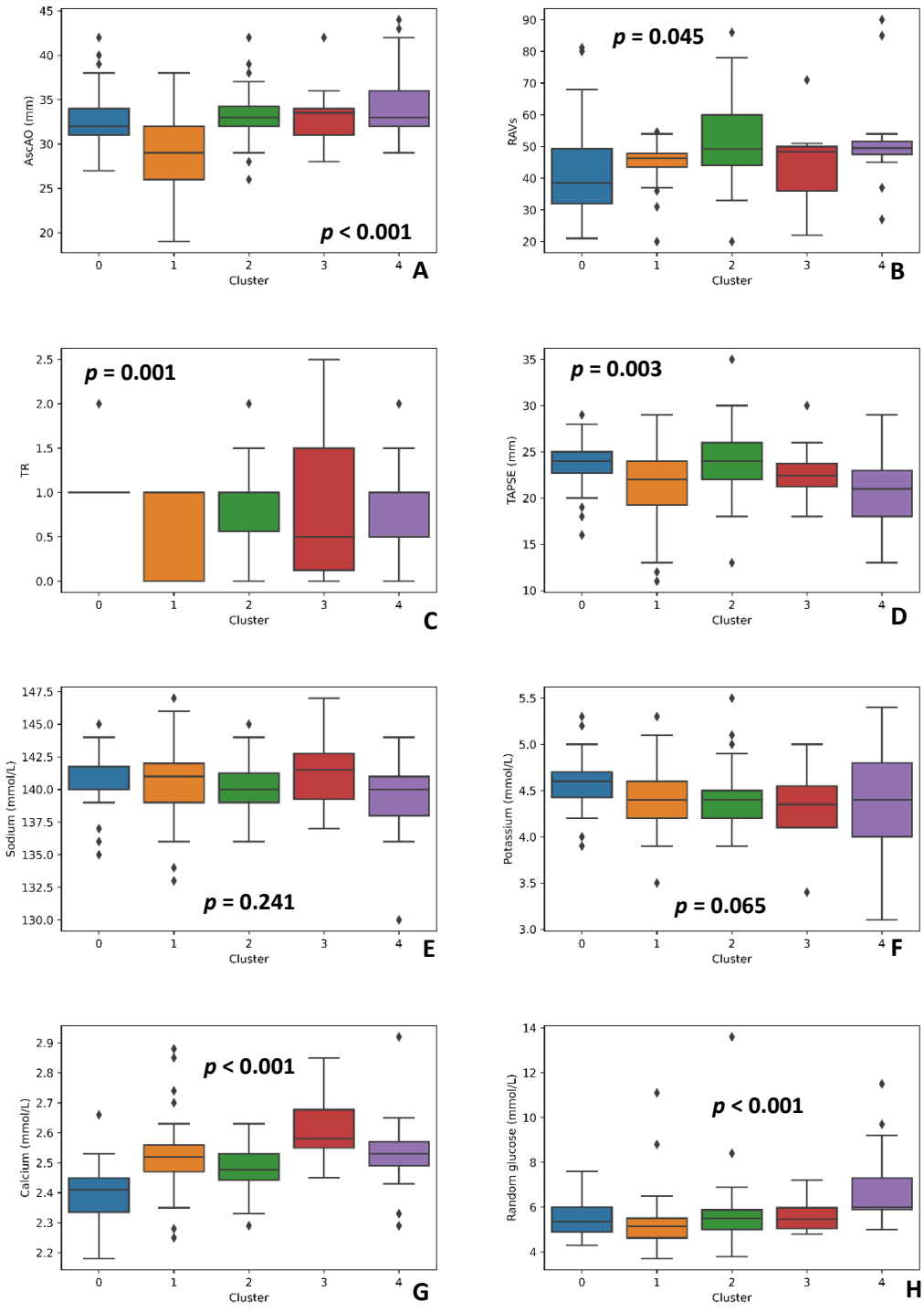

Figure 75. Five clusters setting, features: A — AscAO, B — RAVs, C — TR, D — TAPSE, E — Sodium, F — Potassium, G — Calcium, H — Random glucose

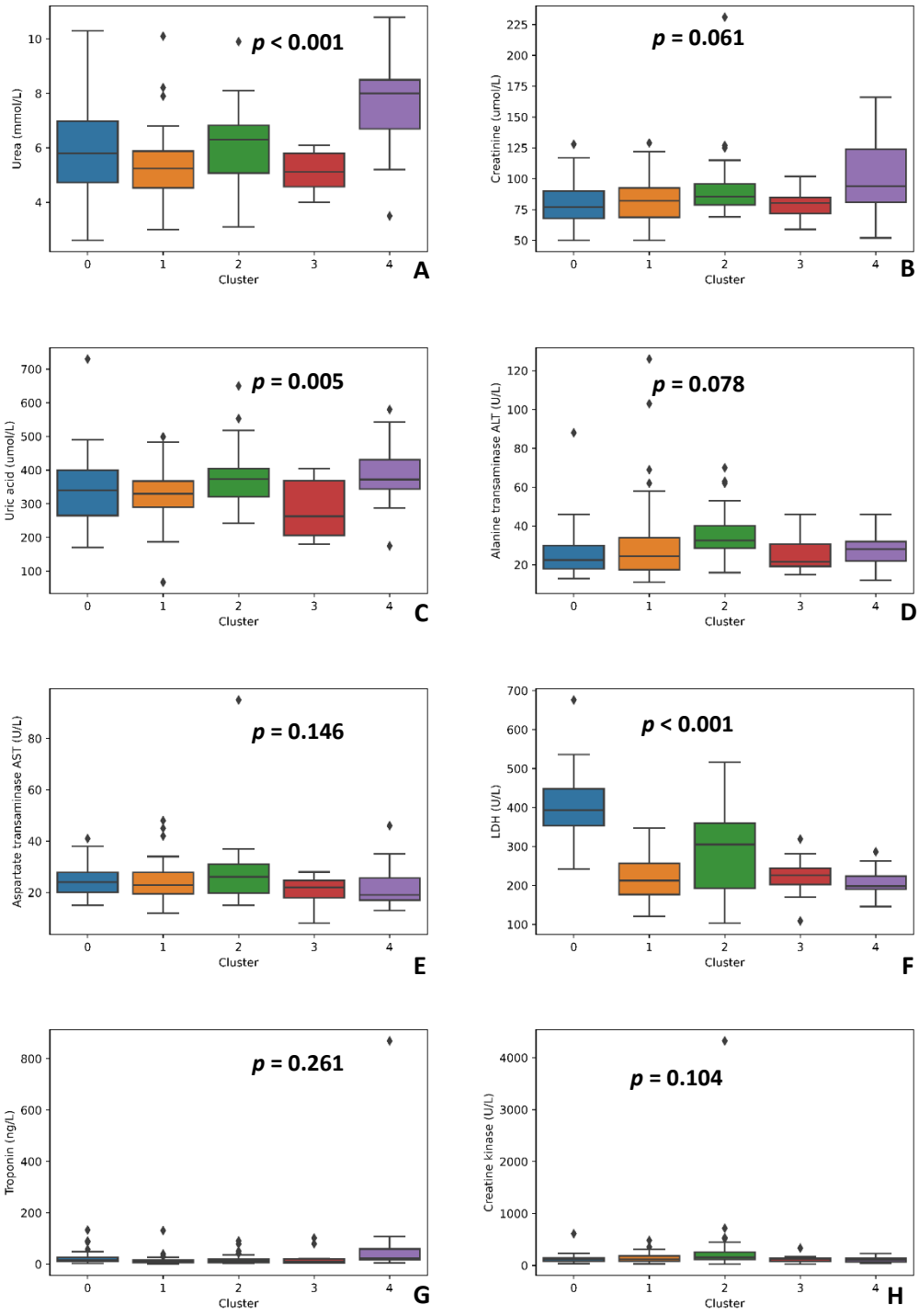

Figure 76. Five clusters setting, features: A — Urea, B — Creatinine, C — Uric acid, D — ALT, E — AST, F — LDH, G — Troponin, H — Creatine-kinase

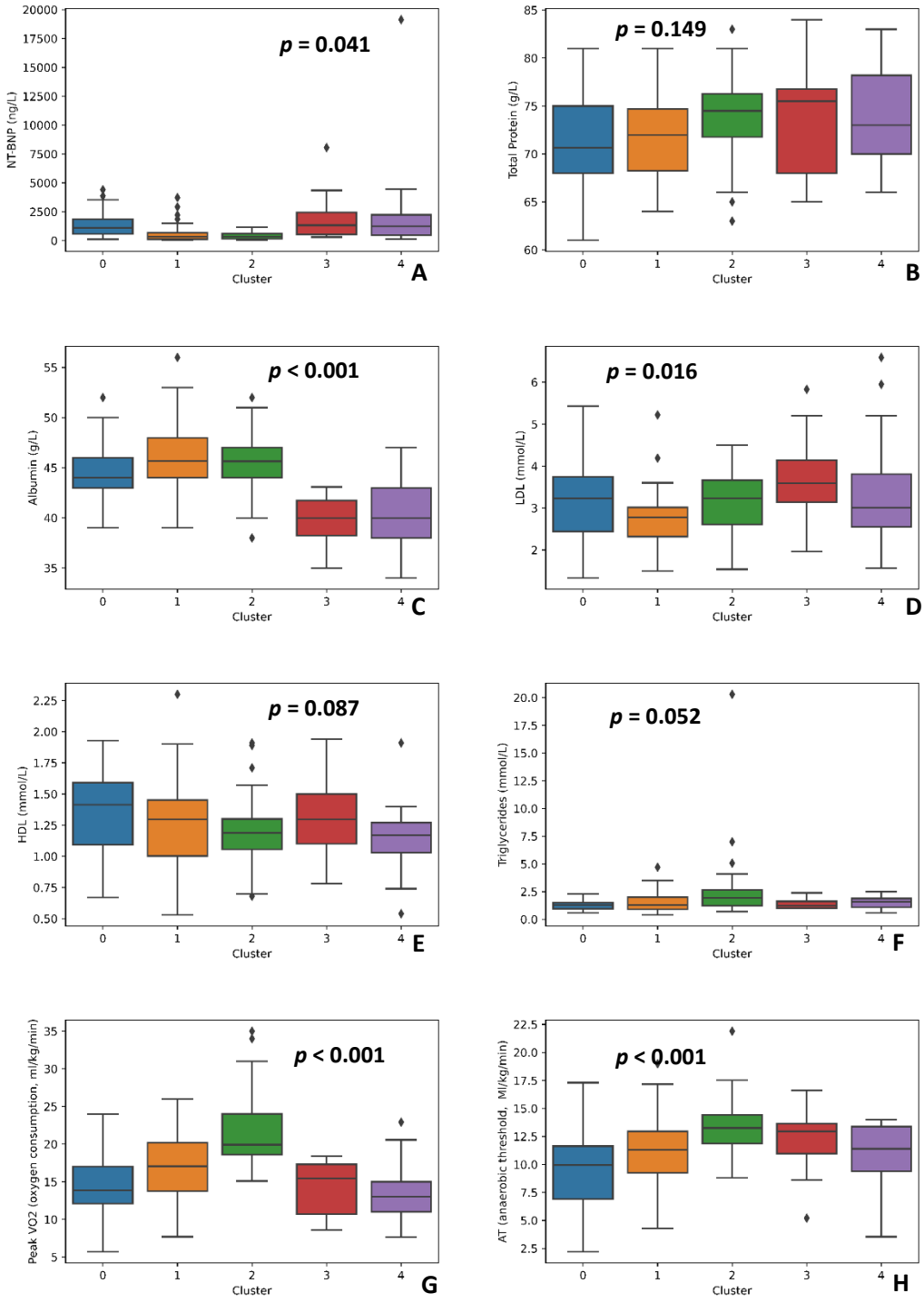

Figure 77. Five clusters setting, features: A — NT-BNP, B — Total protein, C — Albumin, D — LDL, E — HDL, F — Triglycerides, G — Peak VO_2, H — AT

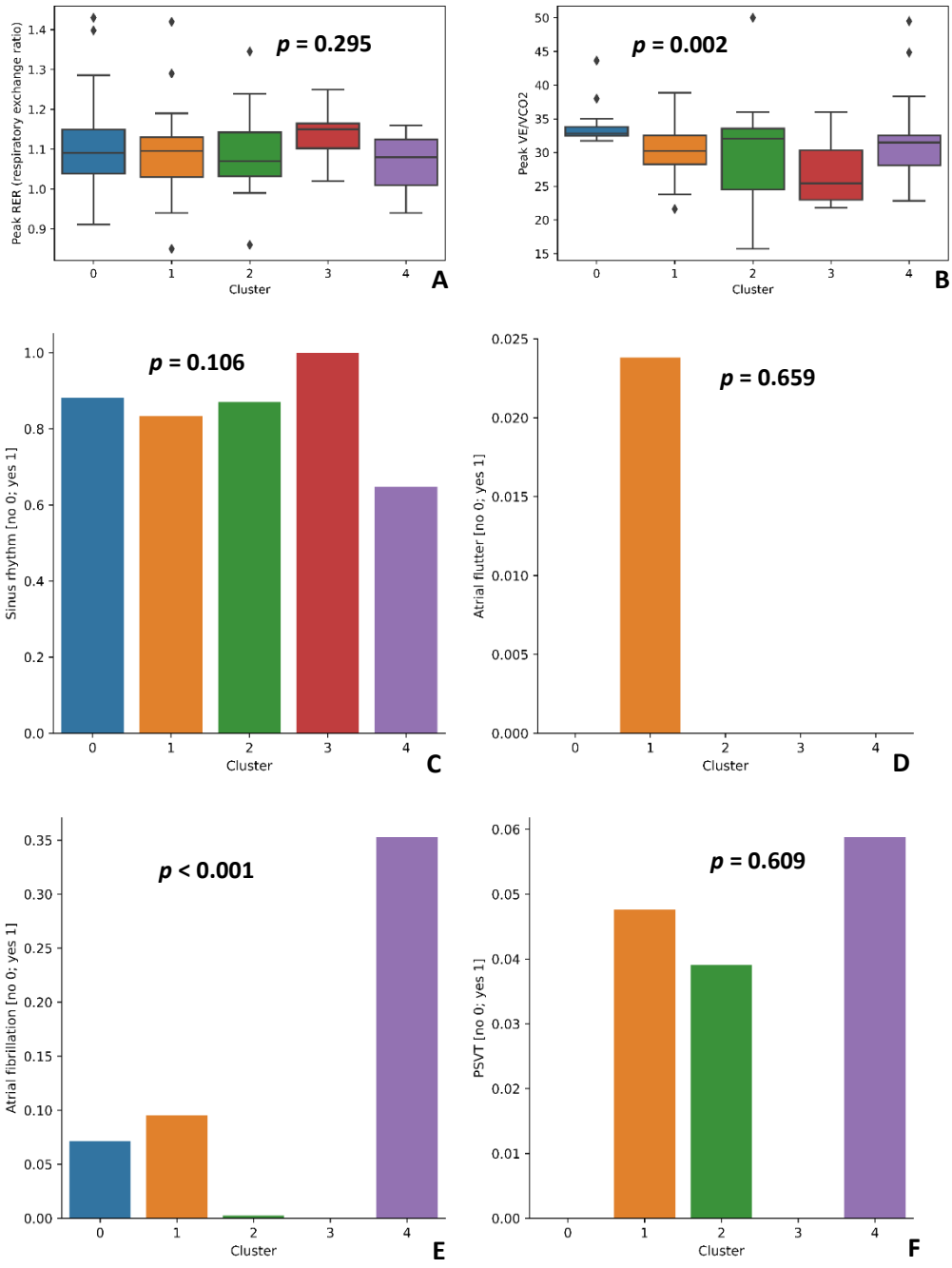

Figure 78. Five clusters setting, features: A — Peak RER, B — Peak VE/VCO_2, C — Sinus rhythm, D — Atrial flutter, E — AF, F — PSVT

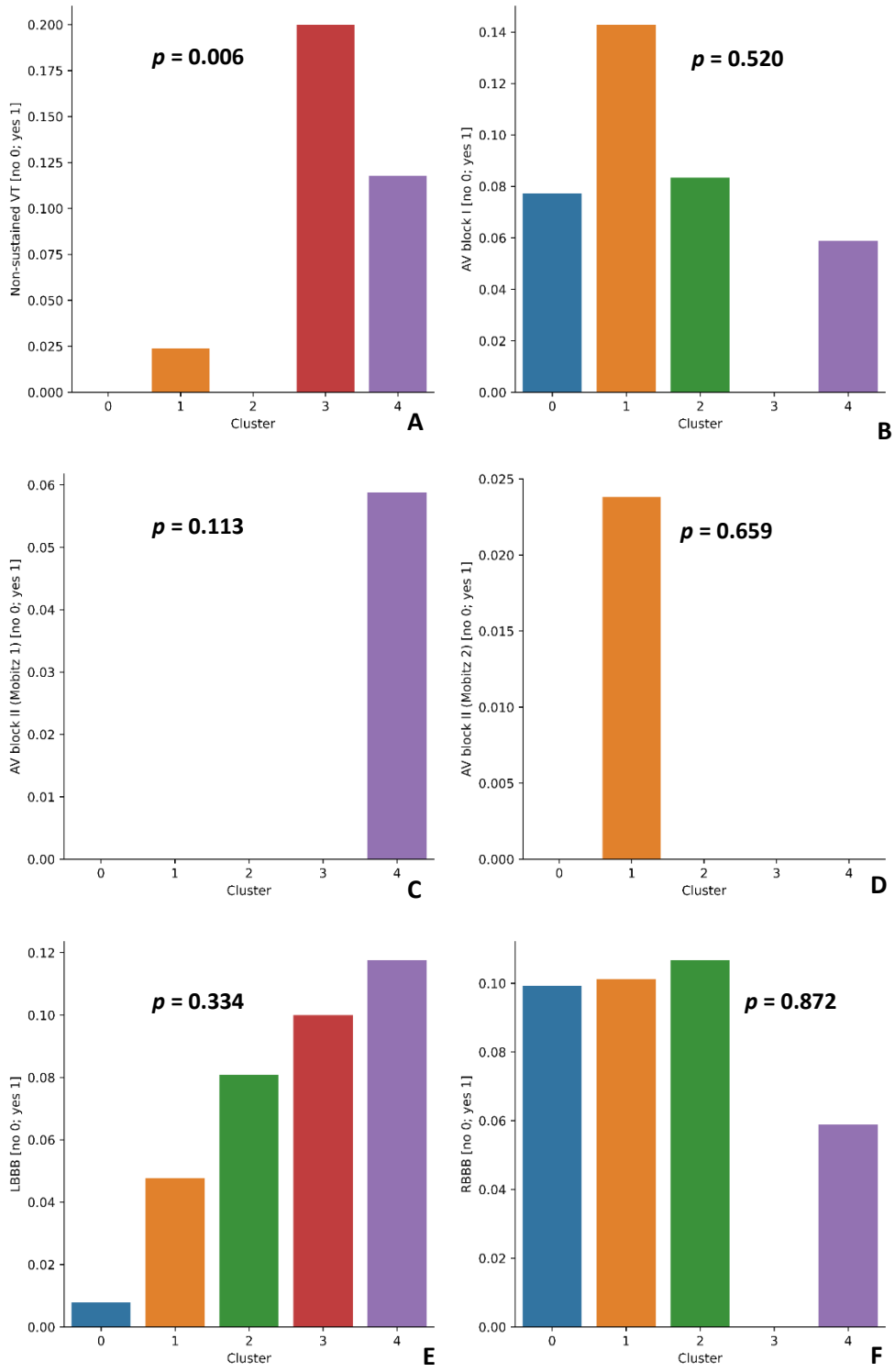

Figure 79. Five clusters setting, features: A — Non-sustained VT, B — AV block I, C — AV block II (Mobitz 1), D — AV block II (Mobitz 2), E — LBBB, F — RBBB

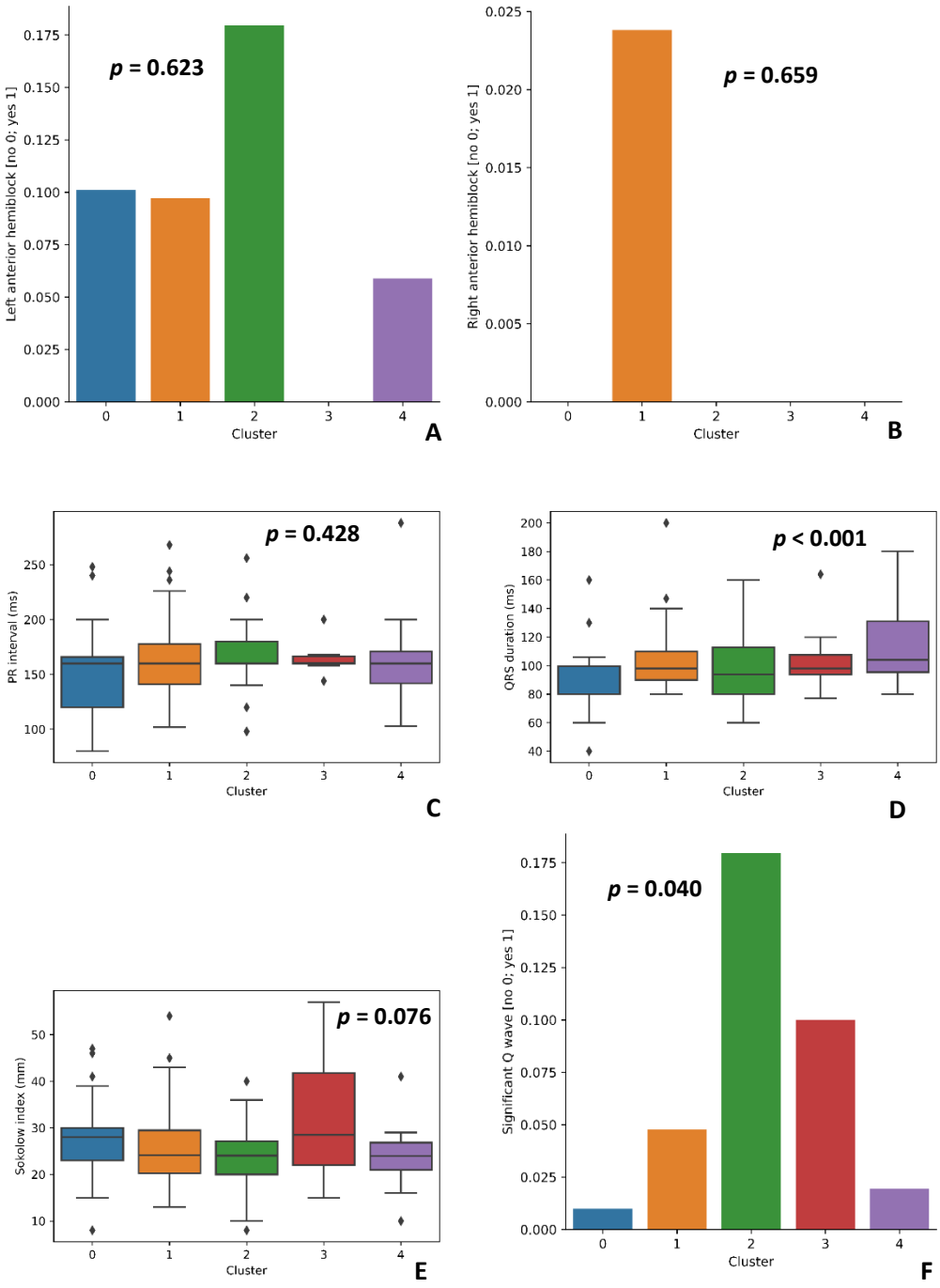

Figure 80. Five clusters setting, features: A — Left anterior hemiblock, B — Right anterior hemiblock, C — PR interval, D — QRS duration, E — Sokolow index, F — Significant Q wave

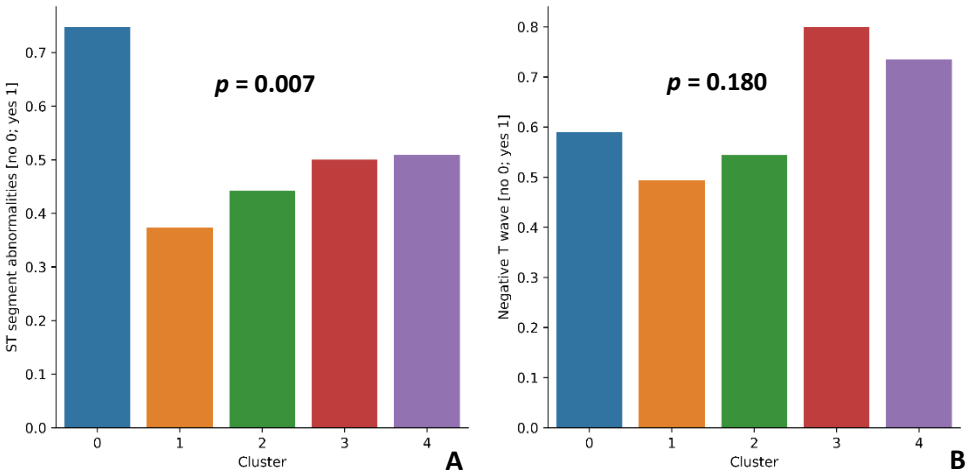

Figure 81. Five clusters setting, features: A — ST segment abnormalities, B — Negative T wave

An approximate interpretation of clustering-logic for 5 clusters is presented in the form of visualized decision tree (Figure 82). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by clustering (Table 13).

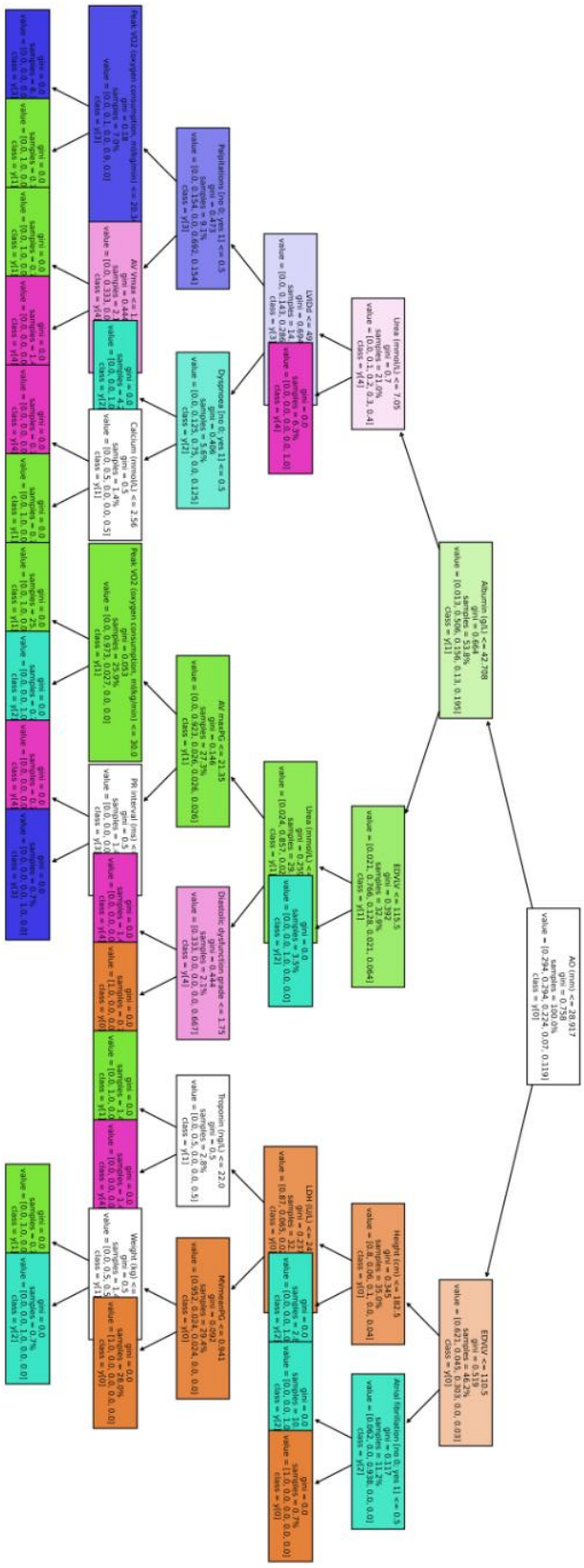

Figure 82. An approximate interpretation of clustering-logic for 5 clusters

Table 13. Feature importance — Top 50 features for distinguishing 5 clusters

Feature	Estimated performance
AO	0.064219
LDH	0.055462
EDVLV	0.050539
MV meanPG	0.038548
SVLV	0.036455
ESVLV	0.034233
AOvs	0.032217
Peak VO_2	0.025251
Weight	0.024397
Height	0.021852
Peak VE/VCO_2	0.020364
RAVs	0.018605
MV maxPG	0.017974
Calcium	0.017628
Albumin	0.016222
PLWd	0.016192
RVSP	0.015337
LAVs	0.015174
LVIDd	0.014638
E/E'	0.014228
NT-BNP	0.013645
AVVTI	0.013381
LAV	0.013117
HCM in family history [yes/no]	0.012657
Anaerobic threshold	0.012502
Serum urea	0.012461
AV meanPG	0.012326
LVOT maxPG	0.011570
Age	0.011425
LVOT Vmax	0.011268
AV maxPG	0.011199
NYHA class	0.010263
BMI	0.010210
MVVTI	0.010070
AscAO	0.009943
QRS duration	0.009362
TR	0.009313
LA	0.009019
LDL	0.008725
Heart rate	0.008618
Creatine-kinase	0.008565
Troponin	0.008330
EFLV	0.008095
Triglycerides	0.007970
MR	0.007632
ALT	0.007523
Random glucose	0.007375
Heart murmur [yes/no]	0.007115
AVA by planimetry	0.007060
Uric acid	0.006565

4.1.2.2.5. Six clusters

Characteristics of the 6 clusters determined are shown in Figures 83-100.

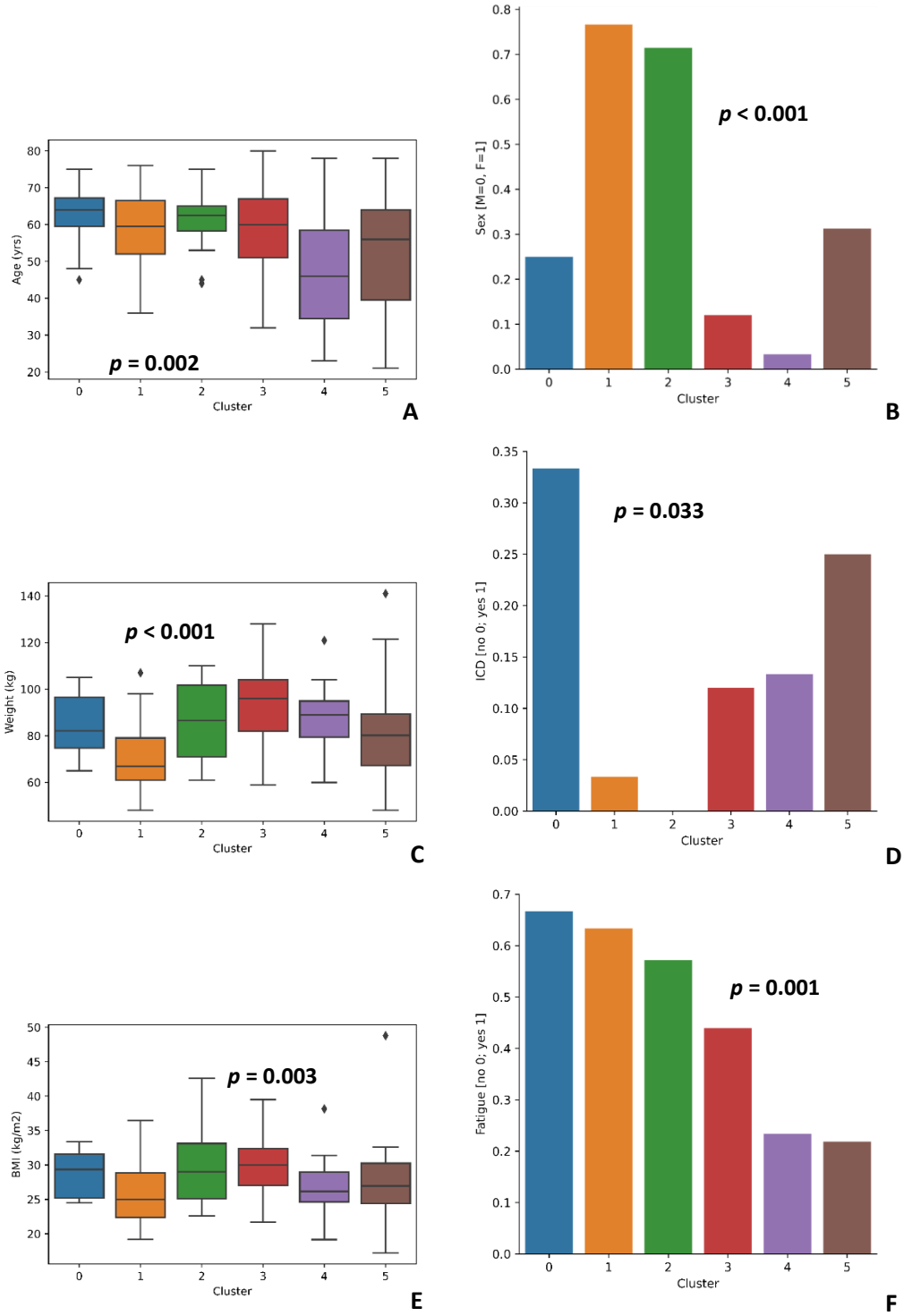

Figure 83. Six clusters setting, features: A — Age, B — Sex, C — Weight, D — ICD, E — BMI, F — Fatigue

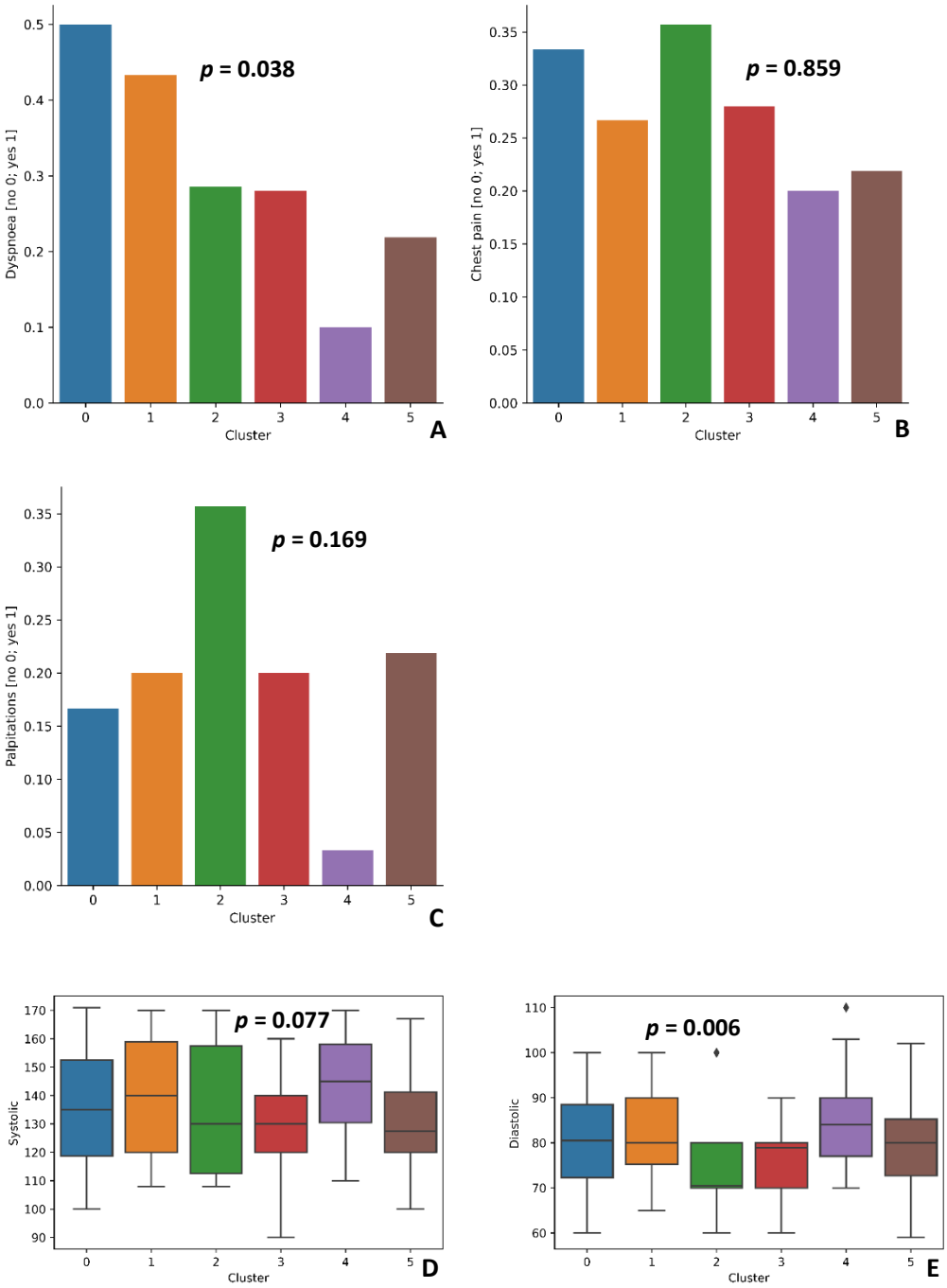

Figure 84. Six clusters setting, features: A — Dyspnea, B — Chest pain, C — Palpitations, D — Systolic blood pressure, E — Diastolic blood pressure

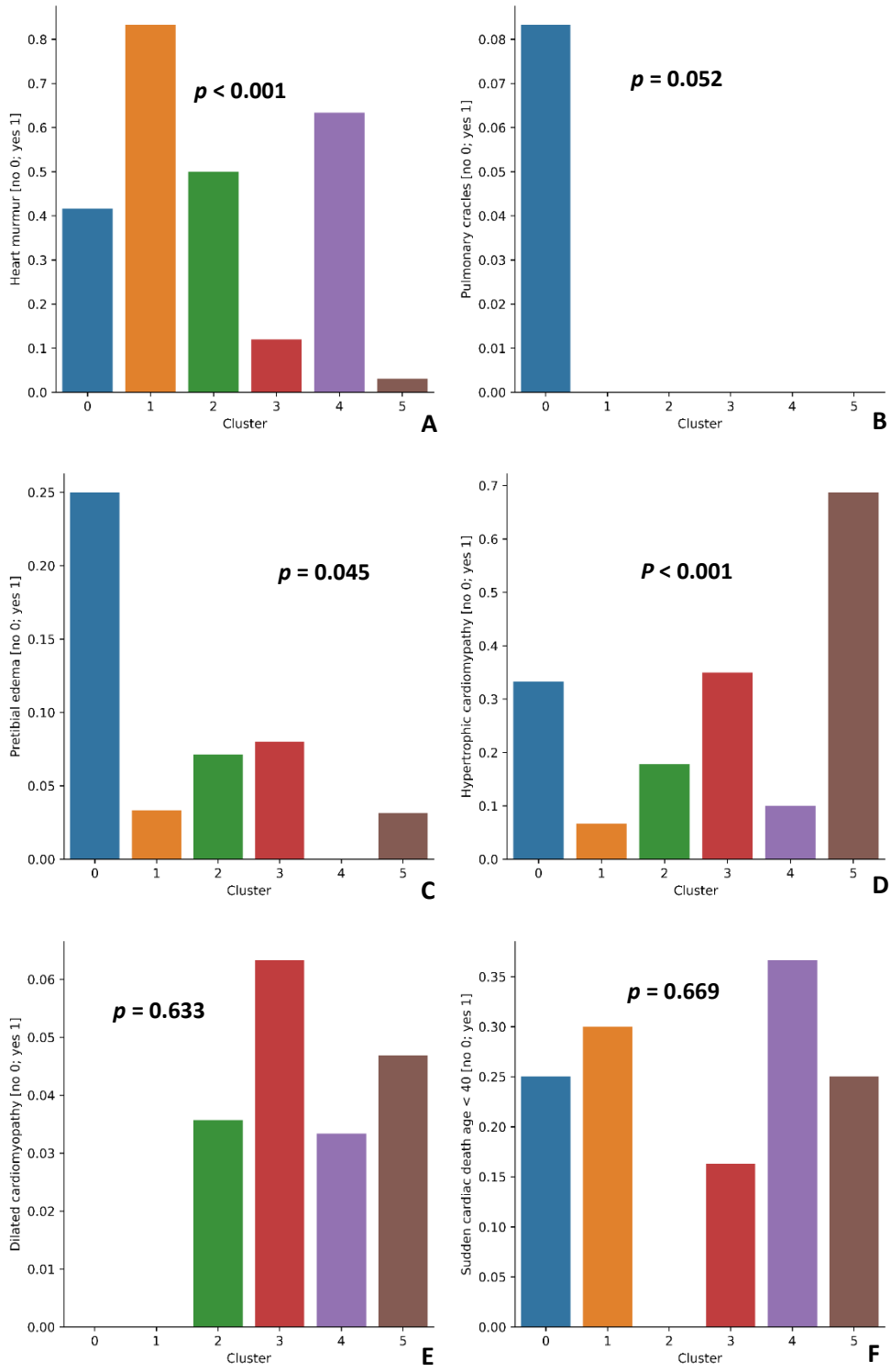

Figure 85. Six clusters setting, features: A — Heart murmur, B — Pulmonary crackles, C — Pretibial edema, D — HCM in family history, E — DCM in family history, F — SCD in age < 40 in family history

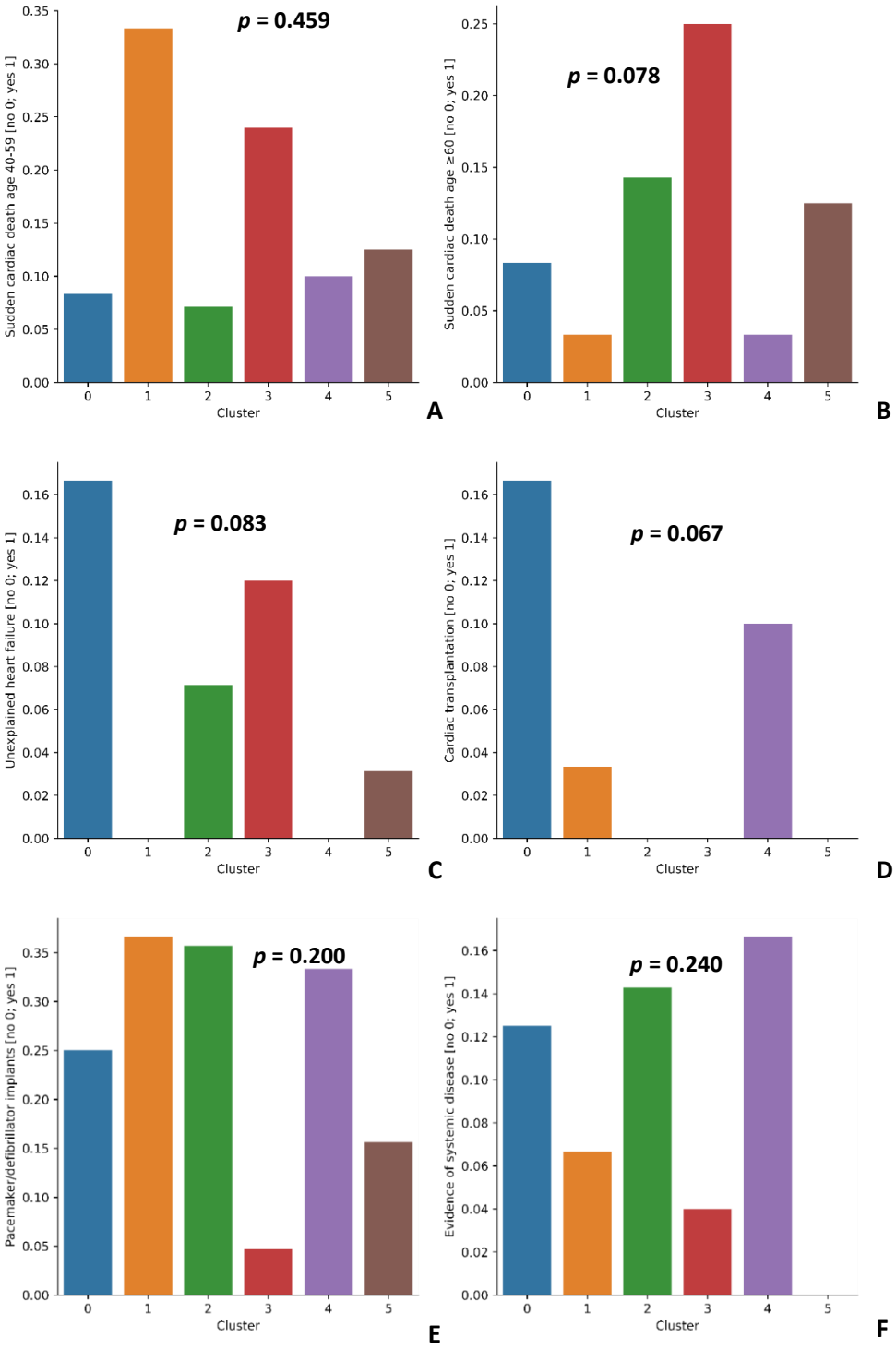

Figure 86. Six clusters setting, features: A — SCD in age 40-59 in family history, B — SCD in age ≥ 60 in family history, C — Unexplained HF in family history, D — Cardiac transplantation in family history, E — Pacemaker/defibrillator implants in family history, F — Evidence of systemic disease in family history

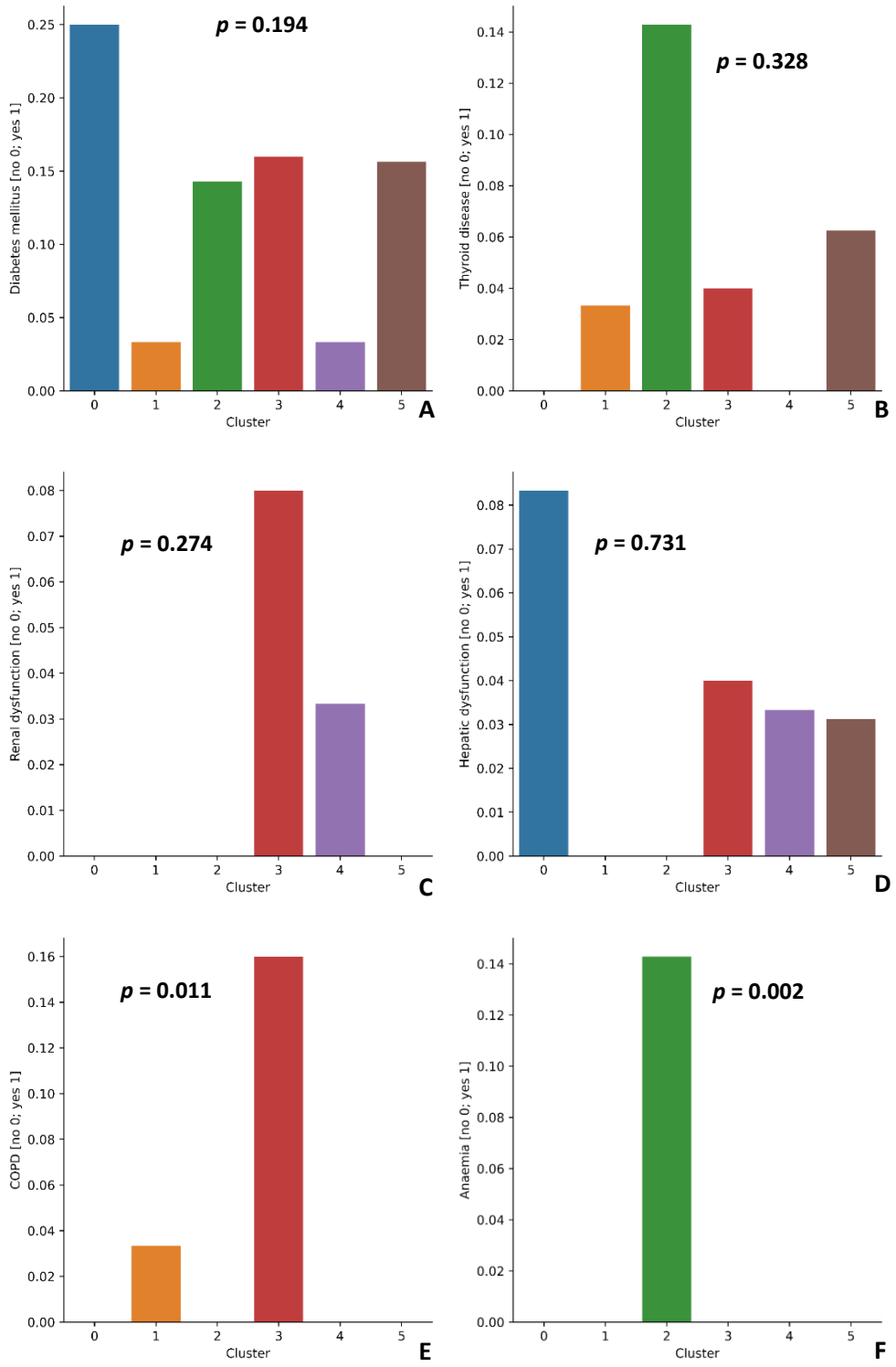

Figure 87. Six clusters setting, features: A — Diabetes mellitus, B — Thyroid disease, C — Renal dysfunction, D — Hepatic dysfunction, E — COPD, F — Anemia

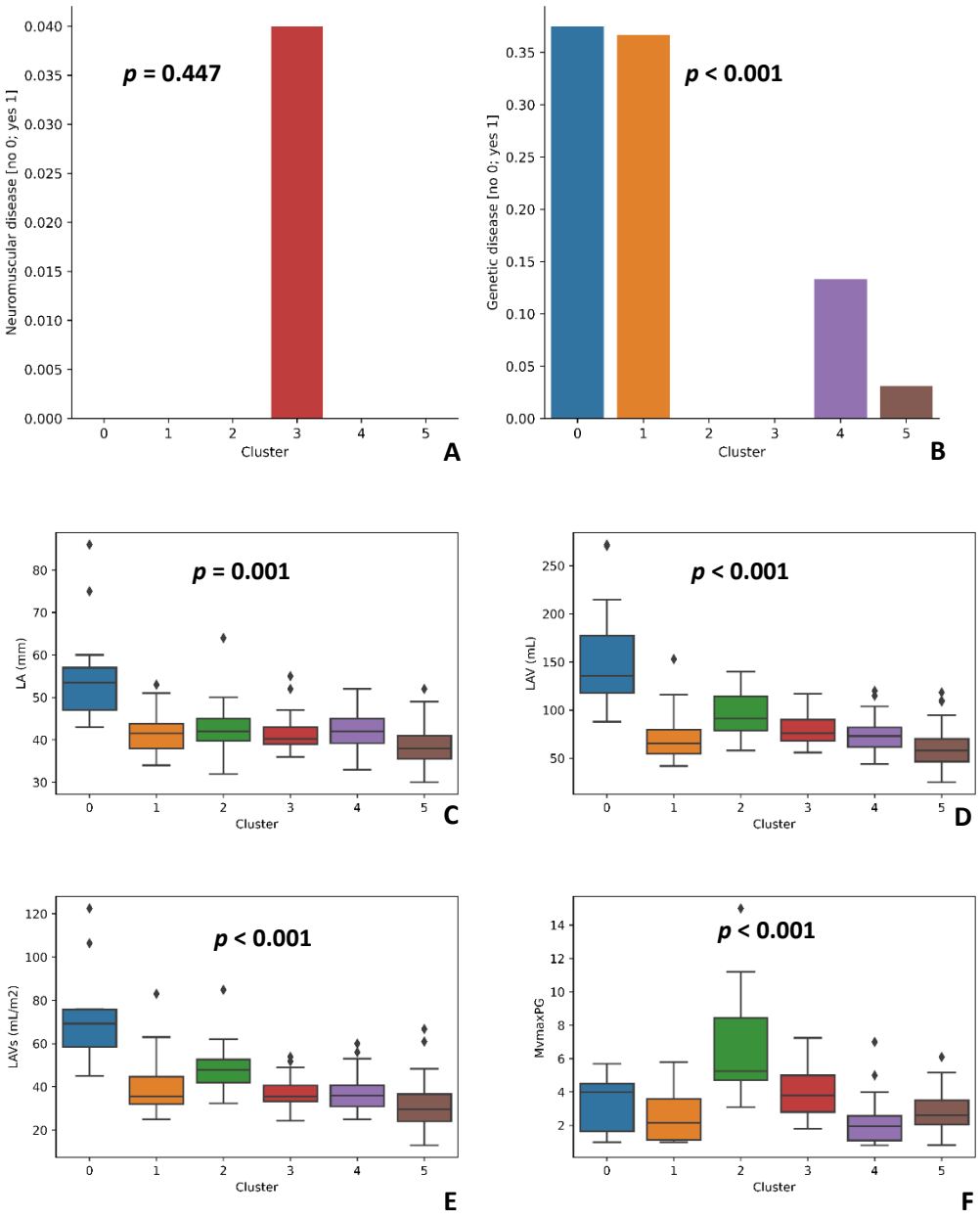

Figure 88. Six clusters setting, features: A — Neuromuscular disease, B — Genetic disease, C — LA, D — LAV, E — LAVs, F — MV maxPG

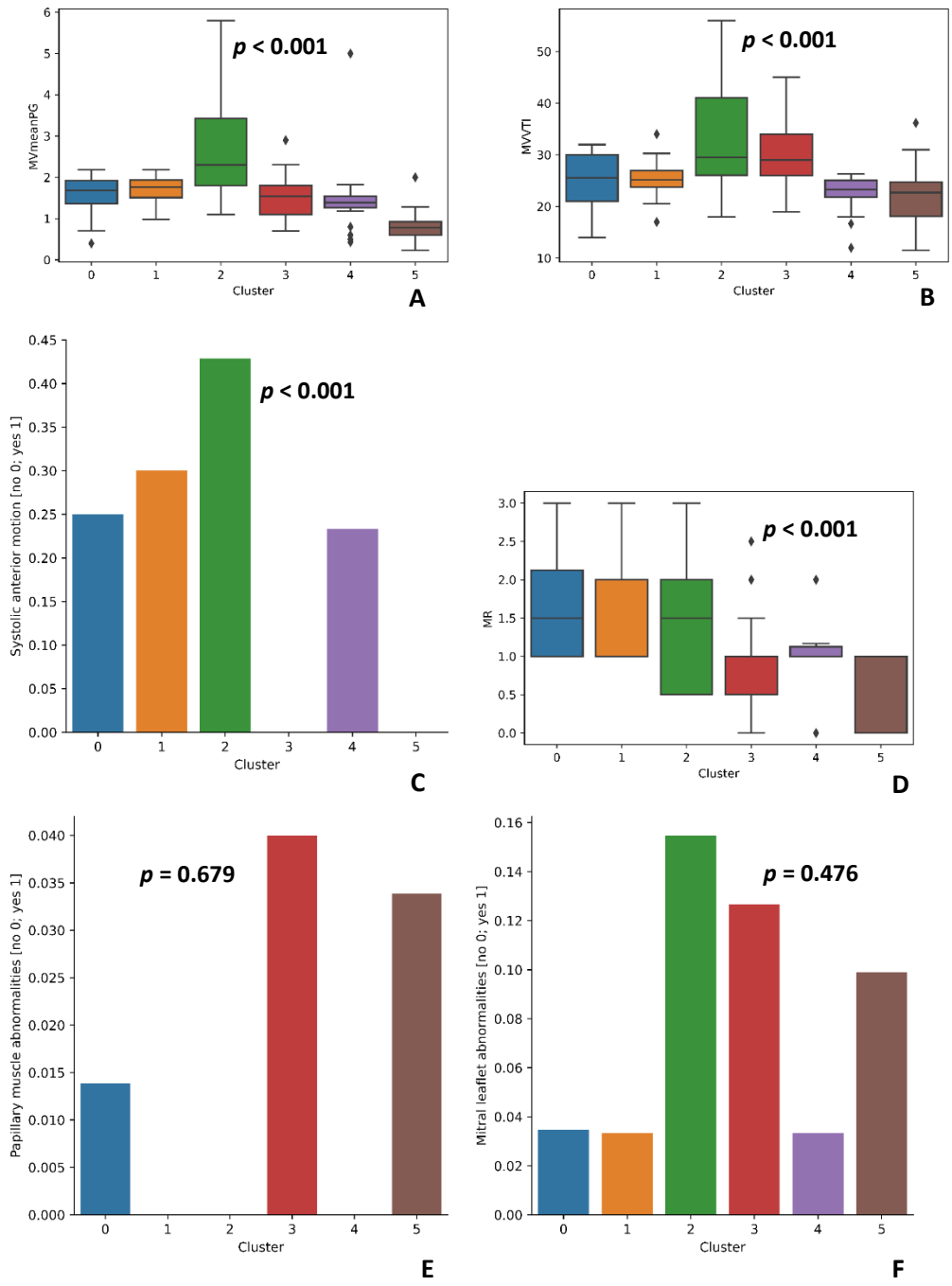

Figure 89. Six clusters setting, features: A — MV meanPG, B — MVVTI, C — SAM, D — MR, E — Papillary muscle abnormalities, F — Mitral leaflet abnormalities

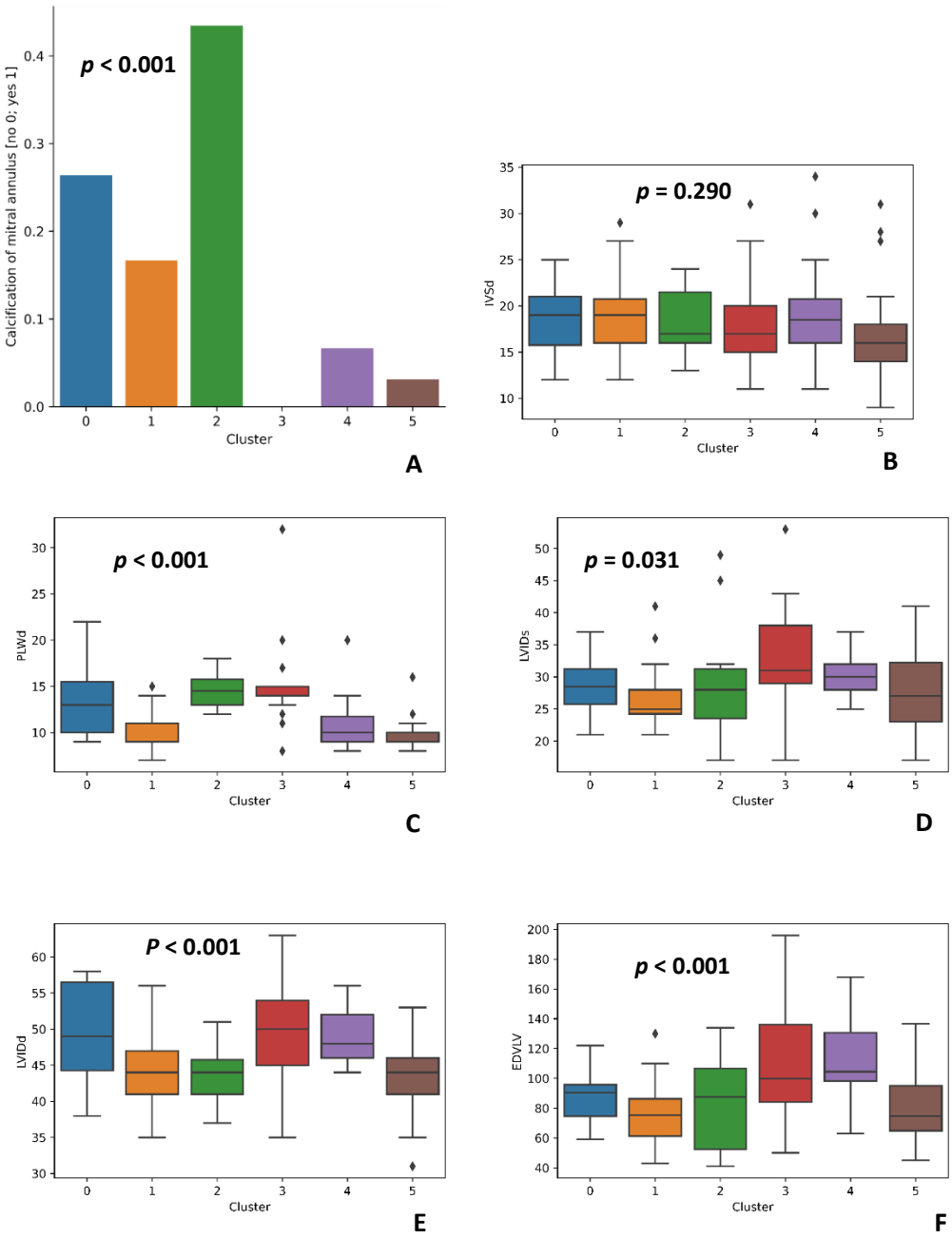

Figure 90. Six clusters setting, features: A — Calcification of mitral annulus, B — IVSd, C — PLWd, D — LVIDs, E — LVIDd, F — EDVLV

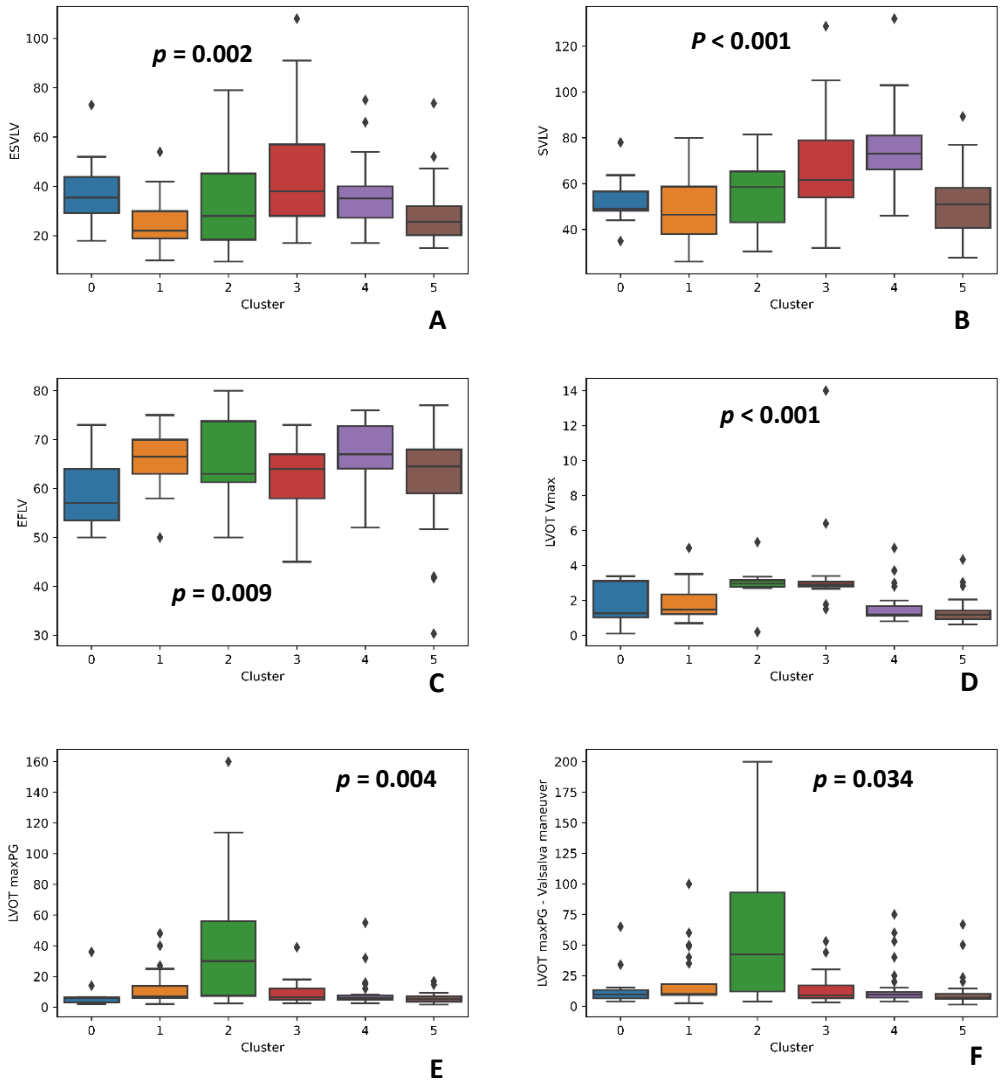

Figure 91. Six clusters setting, features: A — ESVLV, B — SVLV, C — EFLV, D — LVOT Vmax, E — LVOT maxPG, F — LVOT maxPG - Valsalva maneuver

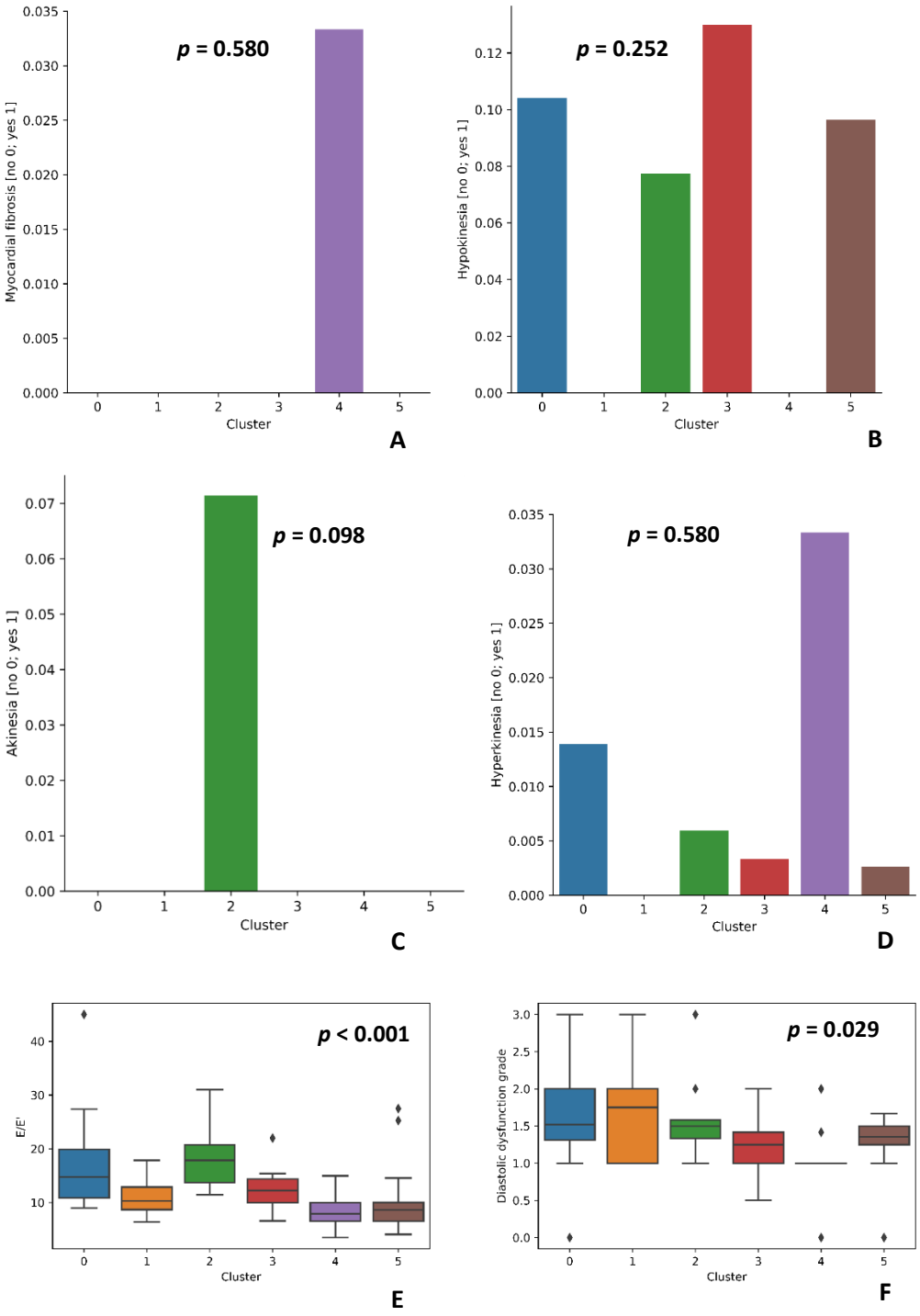

Figure 92. Six clusters setting, features: A — Myocardial fibrosis, B — Hypokinesia, C — Akinesia, D — Hyperkinesia, E — E/E', F — Diastolic dysfunction grade

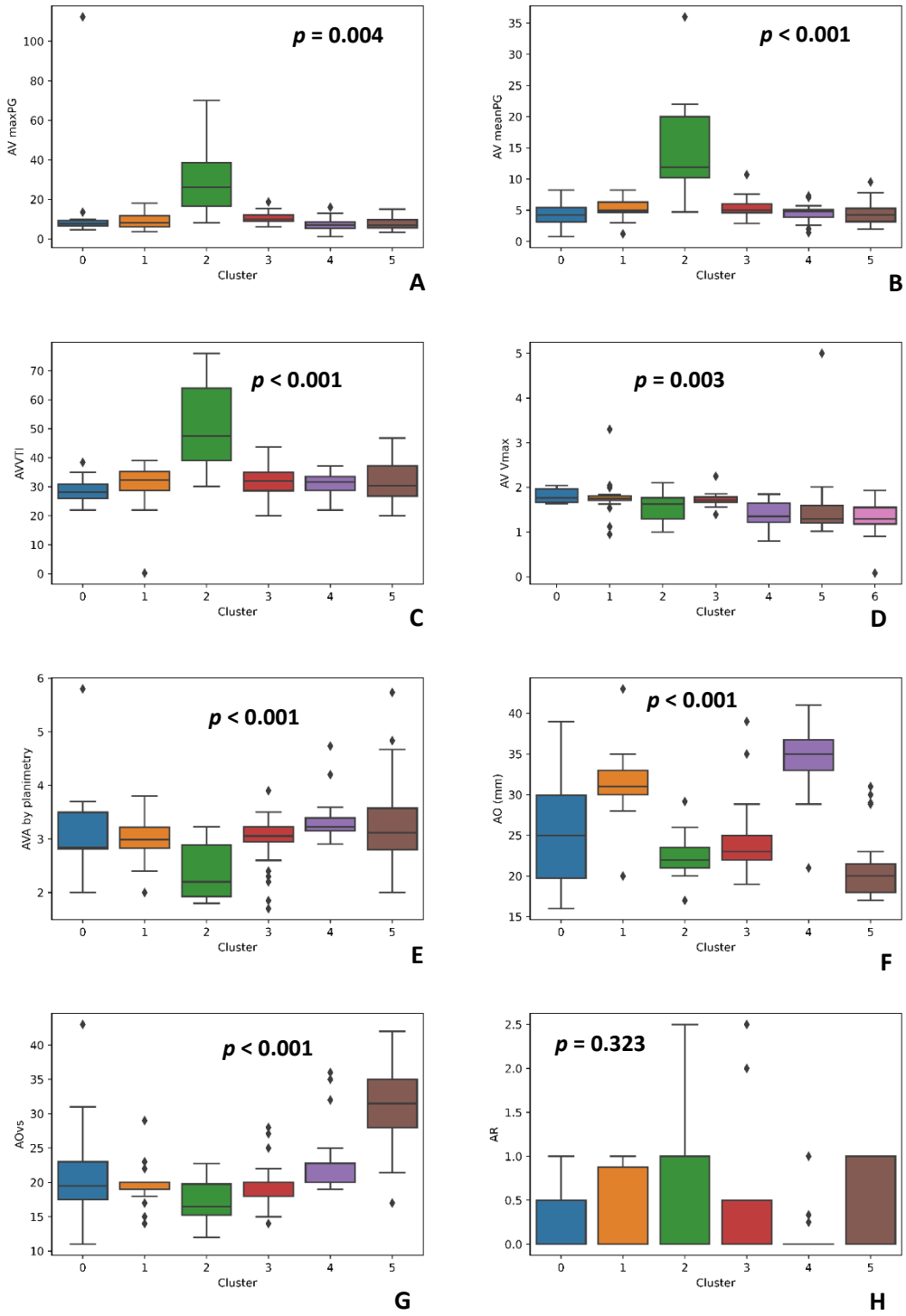

Figure 93. Six clusters setting, features: A — AV maxPG, B — AV meanPG, C — AVVTI, D — AV Vmax, E — AVA by planimetry, F — AO, G — AOvs, H — AR

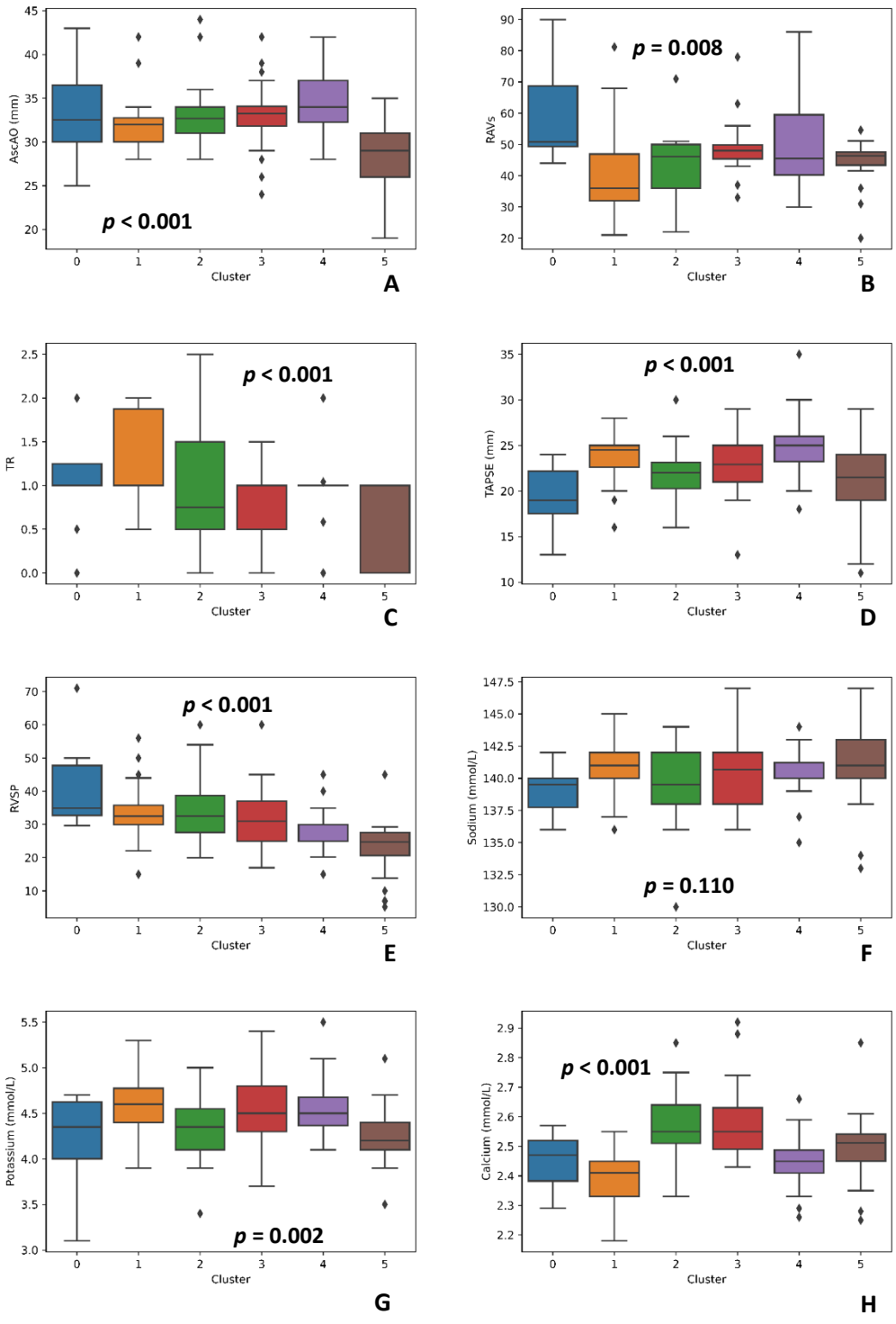

Figure 94. Six clusters setting, features: A — AscAO, B — RAVs, C — TR, D — TAPSE, E — RVSP, F — Sodium, G — Potassium, H — Calcium

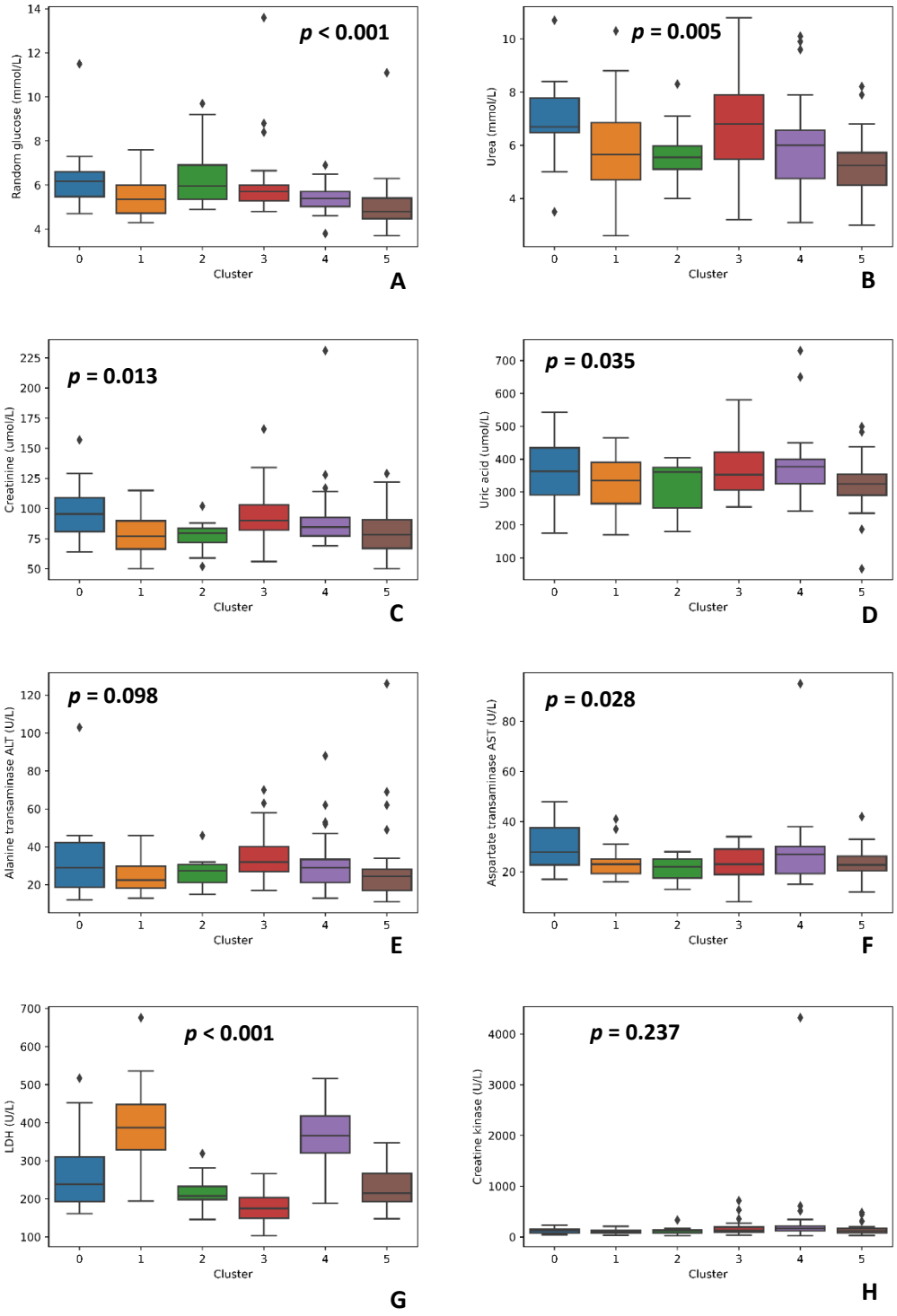

Figure 95. Six clusters setting, features: A — Random glucose, B — Urea, C — Creatinine, D — Uric acid, E — ALT, F — AST, G — LDH, H — Creatine-kinase

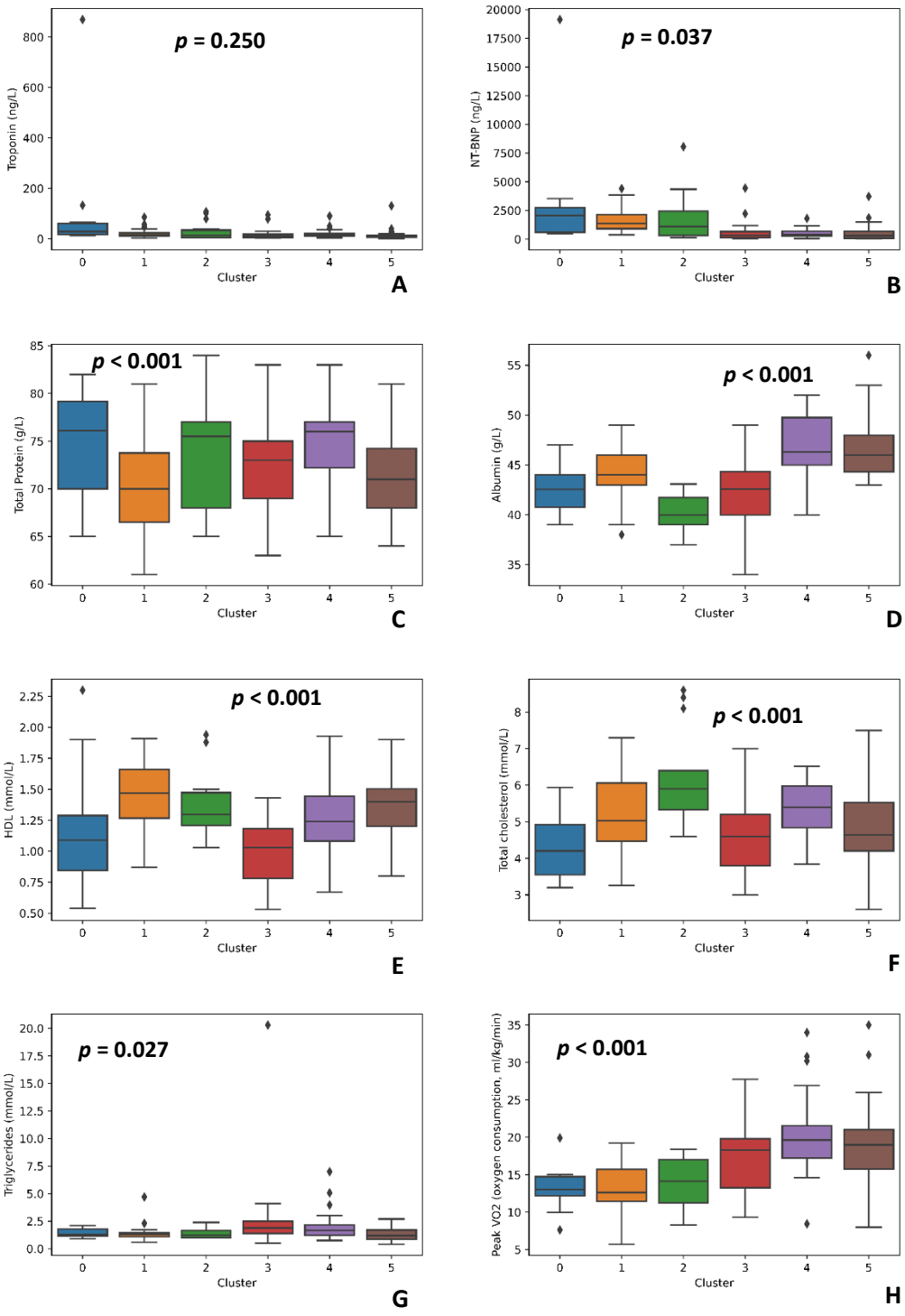

Figure 96. Six clusters setting, features: A — Troponin, B — NT-BNP, C — Total protein, D — Albumin, E — HDL, F — Total cholesterol, G — Triglycerides, H — Peak VO_2

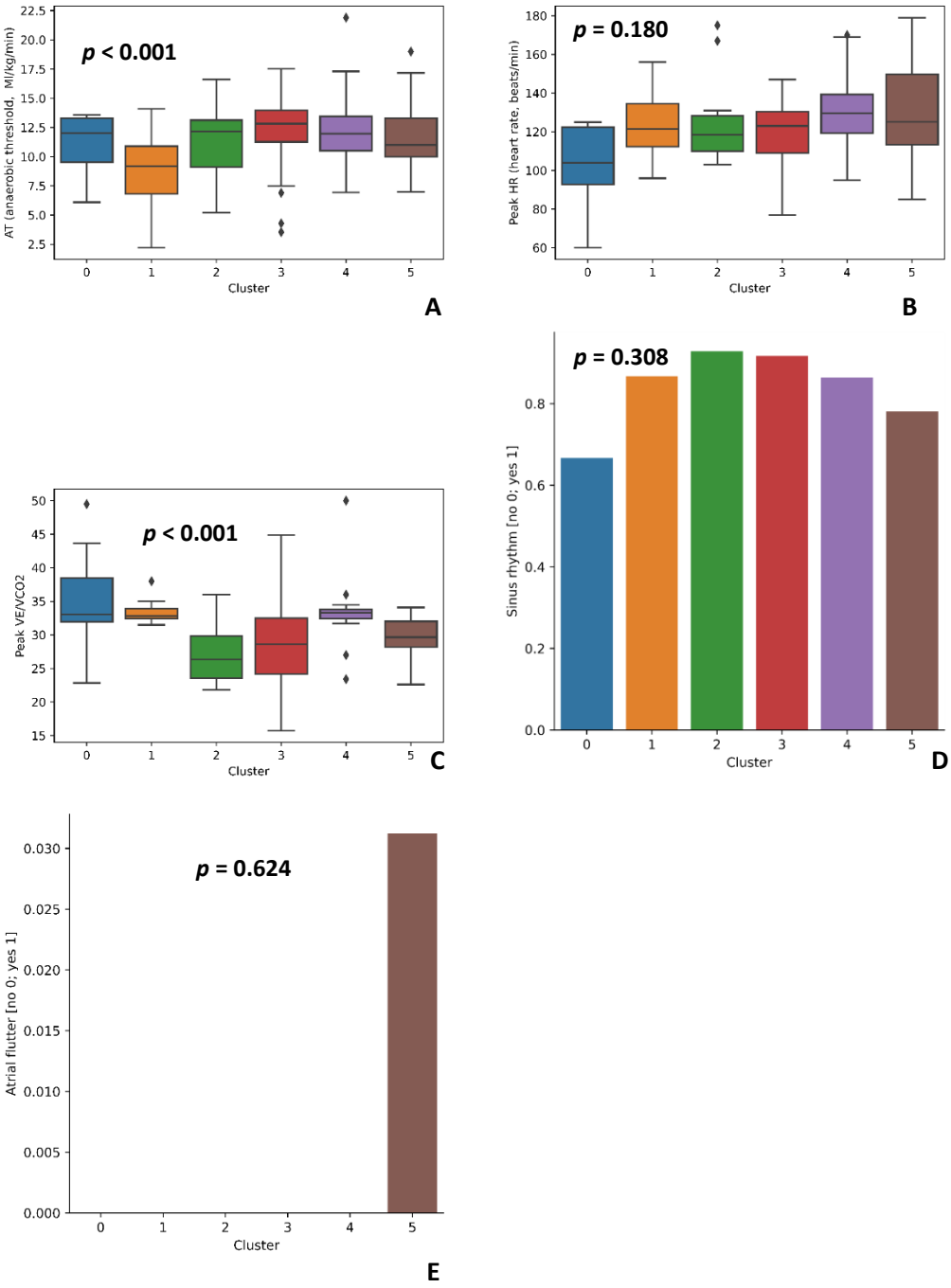

Figure 97. Six clusters setting, features: A — AT, B — Peak HR, C — Peak VE/VCO2, D — Sinus rhythm, E — Atrial flutter

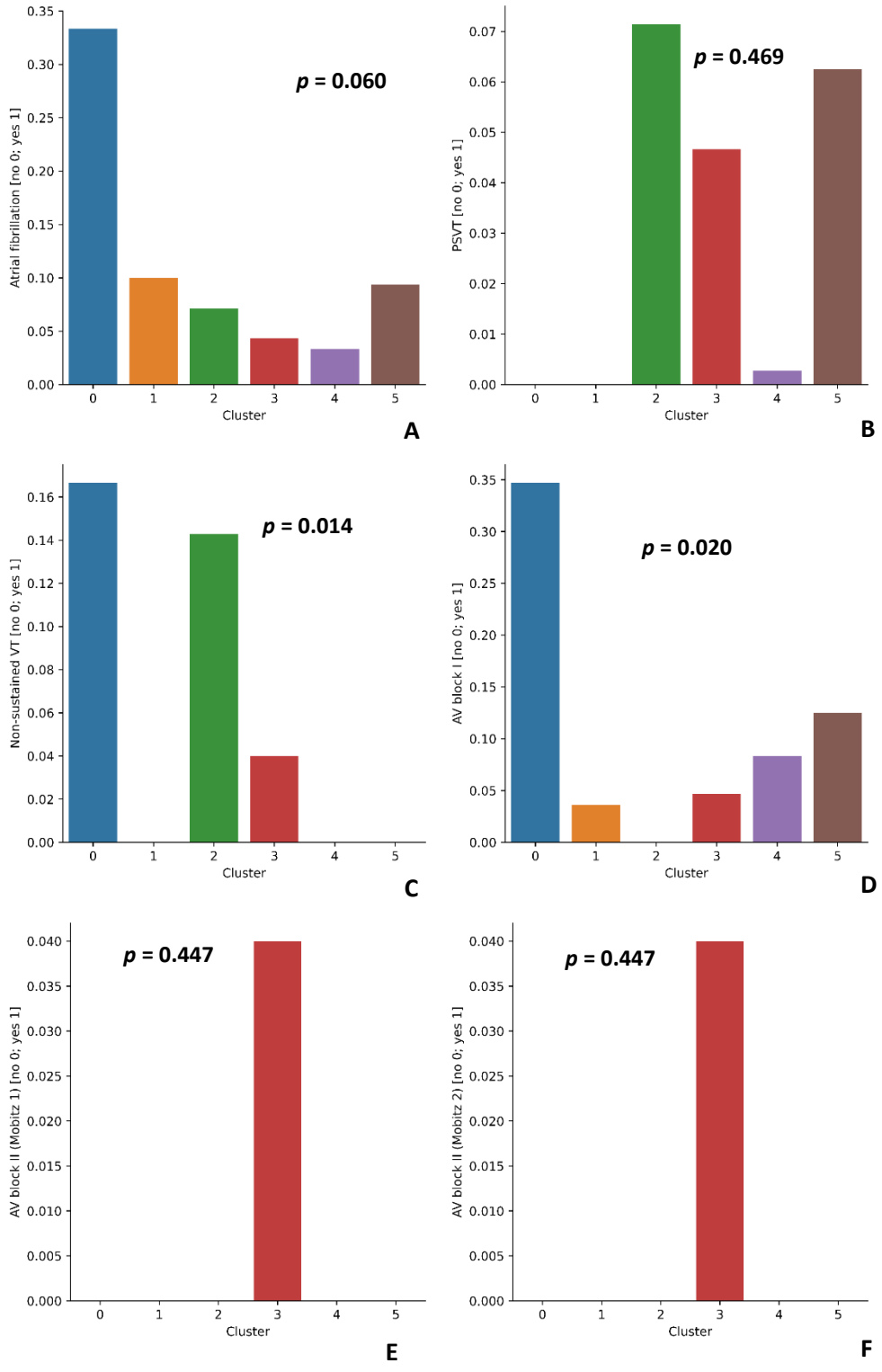

Figure 98. Six clusters setting, features: A — AF, B — PSVT, C — Non-sustained VT, D — AV block I, E — AV block II (Mobitz 1), F — AV block II (Mobitz 2)

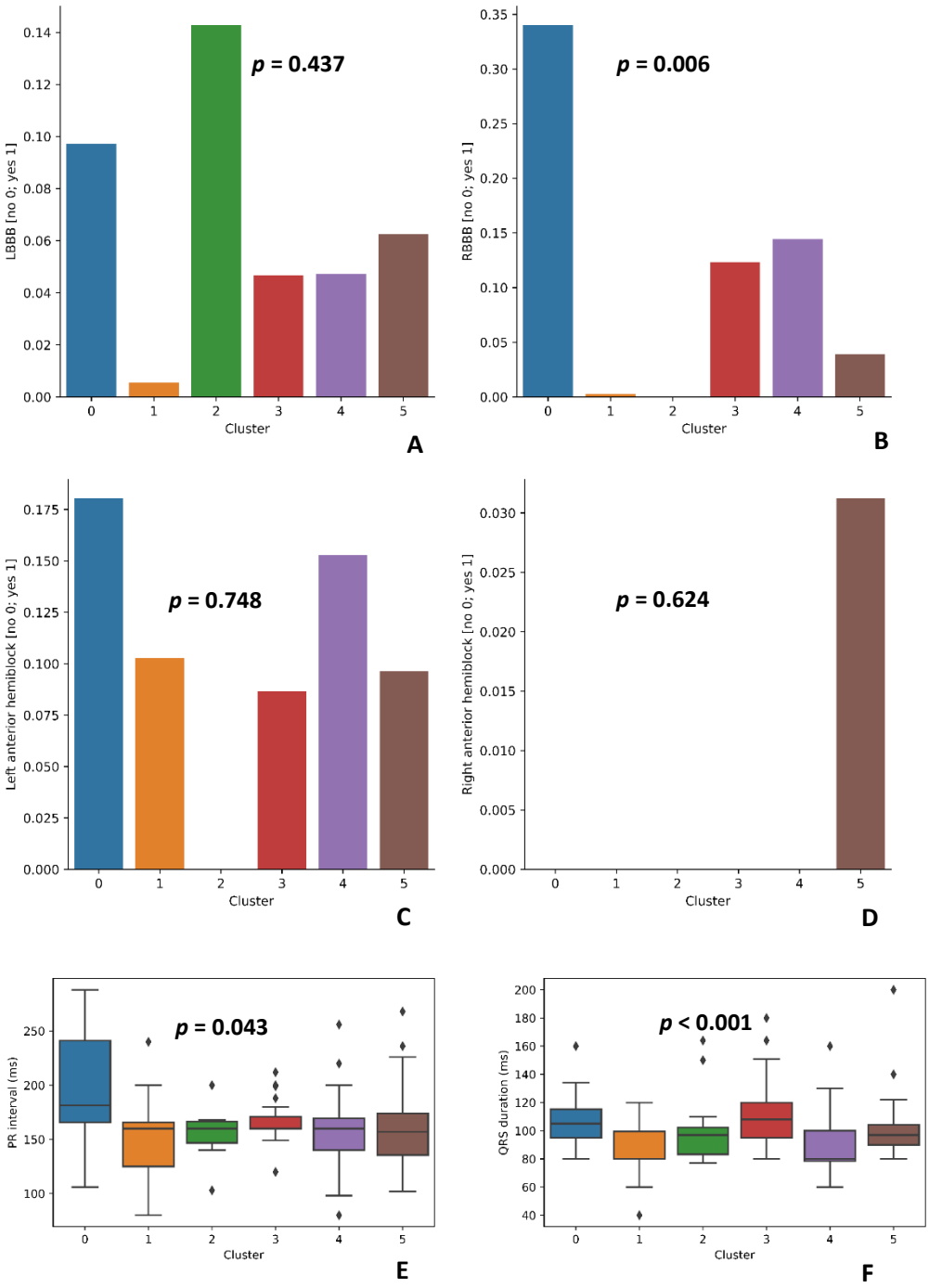

Figure 99. Six clusters setting, features: A — LBBB, B — RBBB, C — Left anterior hemiblock, D — Right anterior hemiblock, E — PR interval, F — QRS duration

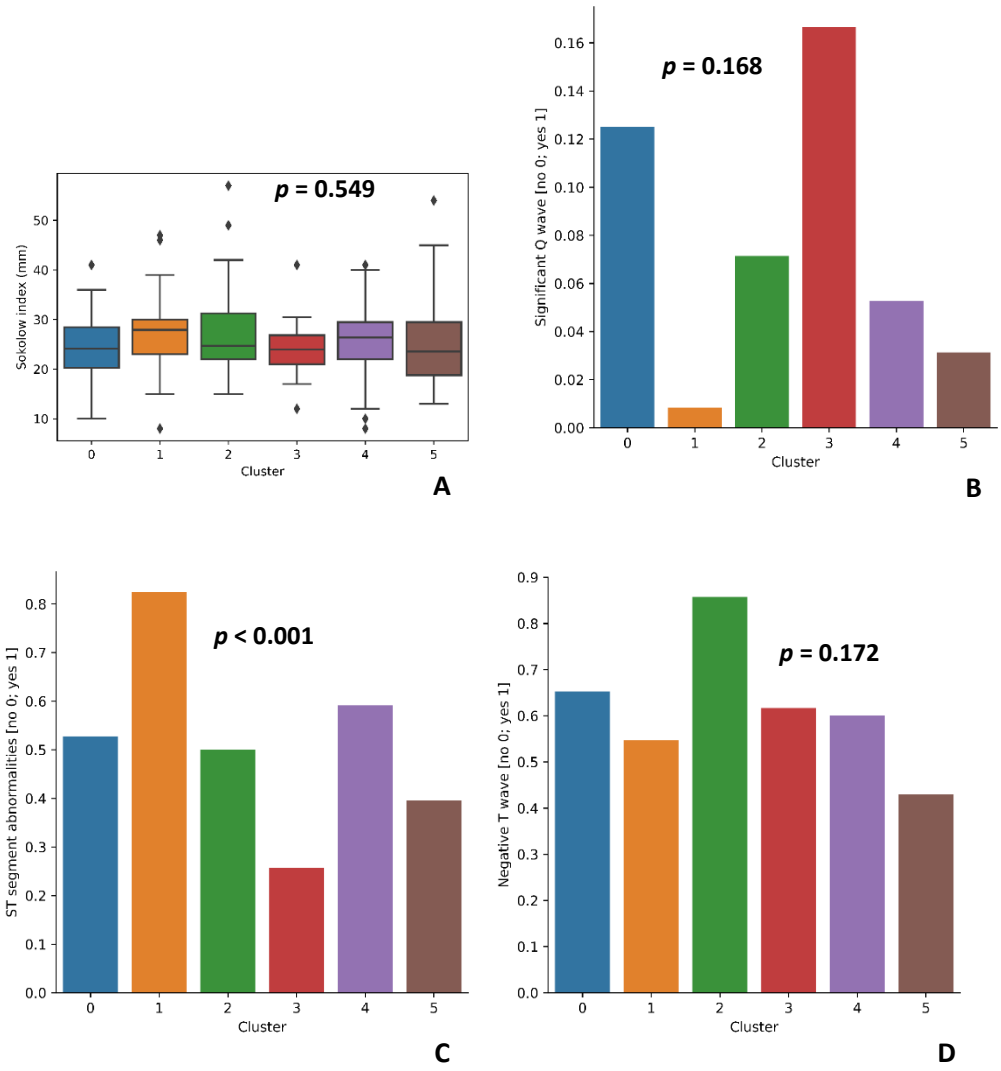

Figure 100. Six clusters setting, features: A — Sokolow index, B — Significant Q wave, C — ST segment abnormalities, D — Negative T wave

An approximate interpretation of clustering-logic for 6 clusters is presented in the form of a visualized decision tree (Figure 101). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by clustering (Table 14).

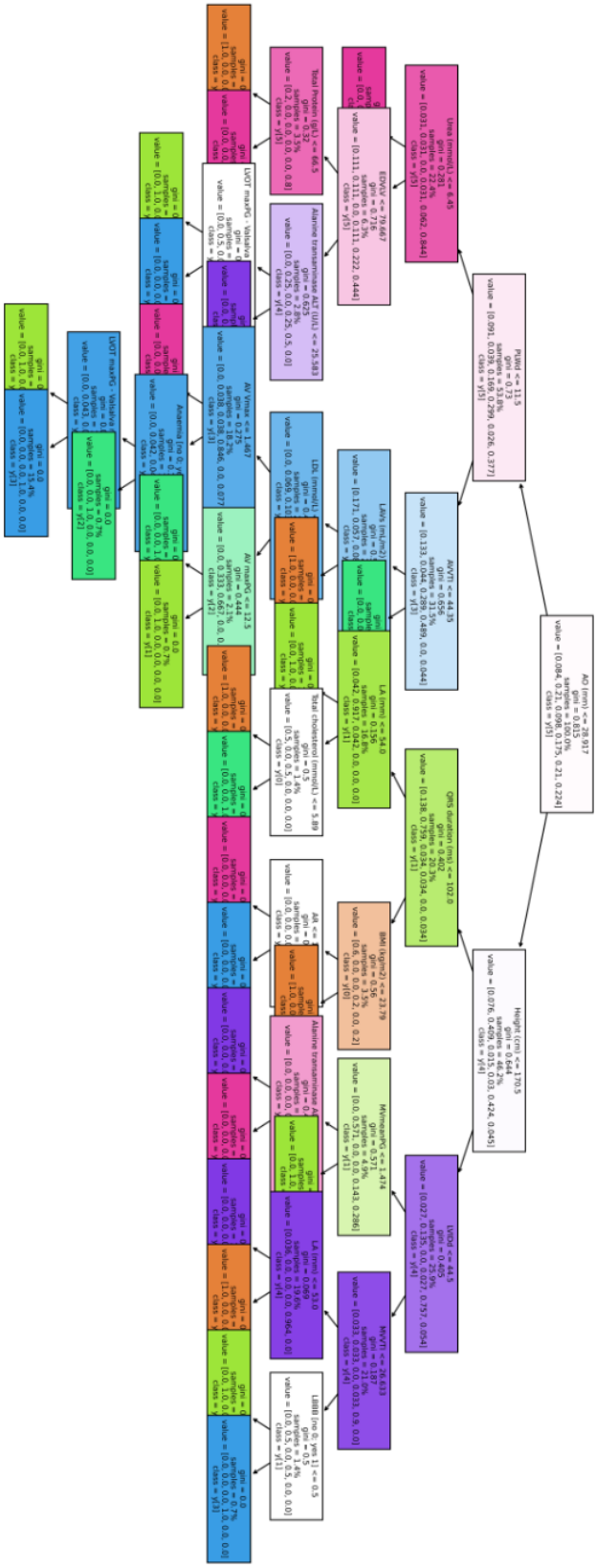

Figure 101. An approximate interpretation of clustering-logic for 6 clusters

Table 14. Feature importance — Top 35 features for distinguishing 6 clusters

Feature	Estimated importance
AO	0.053284
LDH	0.046597
AOvs	0.040991
MV meanPG	0.033790
LVOT Vmax	0.026312
PLWd	0.026020
Peak VE/VCO_2	0.024505
Height	0.024011
EDVLV	0.022810
AV meanPG	0.022360
LAV	0.021264
LAVs	0.020985
Peak VO_2	0.020242
AV maxPG	0.018780
NT-BNP	0.018620
RVSP	0.018227
NYHA class	0.017358
Albumin	0.016636
Heart murmur [yes/no]	0.016513
Serum calcium	0.015658
MVVTI	0.015298
AscAO	0.015032
E/E'	0.014929
SVLV	0.014906
HDL	0.014718
LA	0.014164
Weight	0.014051
AVVTI	0.013854
ESVLV	0.013028
RAVs	0.012670
LDL	0.012563
MV maxPG	0.012536
Random glucose	0.012080
Sex	0.012015
LVIDd	0.011660

4.1.2.2.6. Seven clusters

Characteristics of the 7 clusters determined are shown in Figures 102-118.

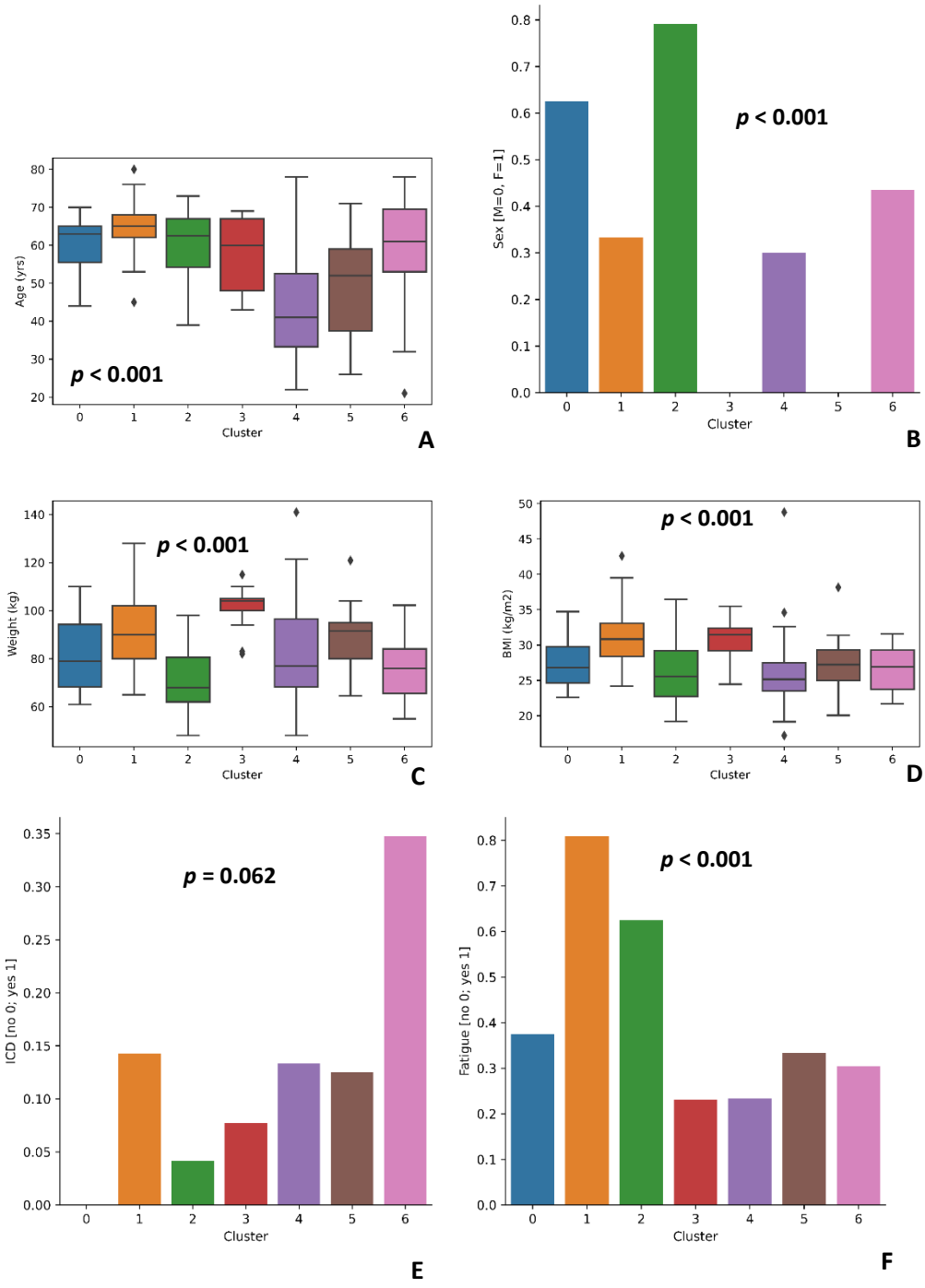

Figure 102. Seven clusters setting, features: A — Age, B — Sex, C — Weight, D — BMI, E — ICD, F — Fatigue

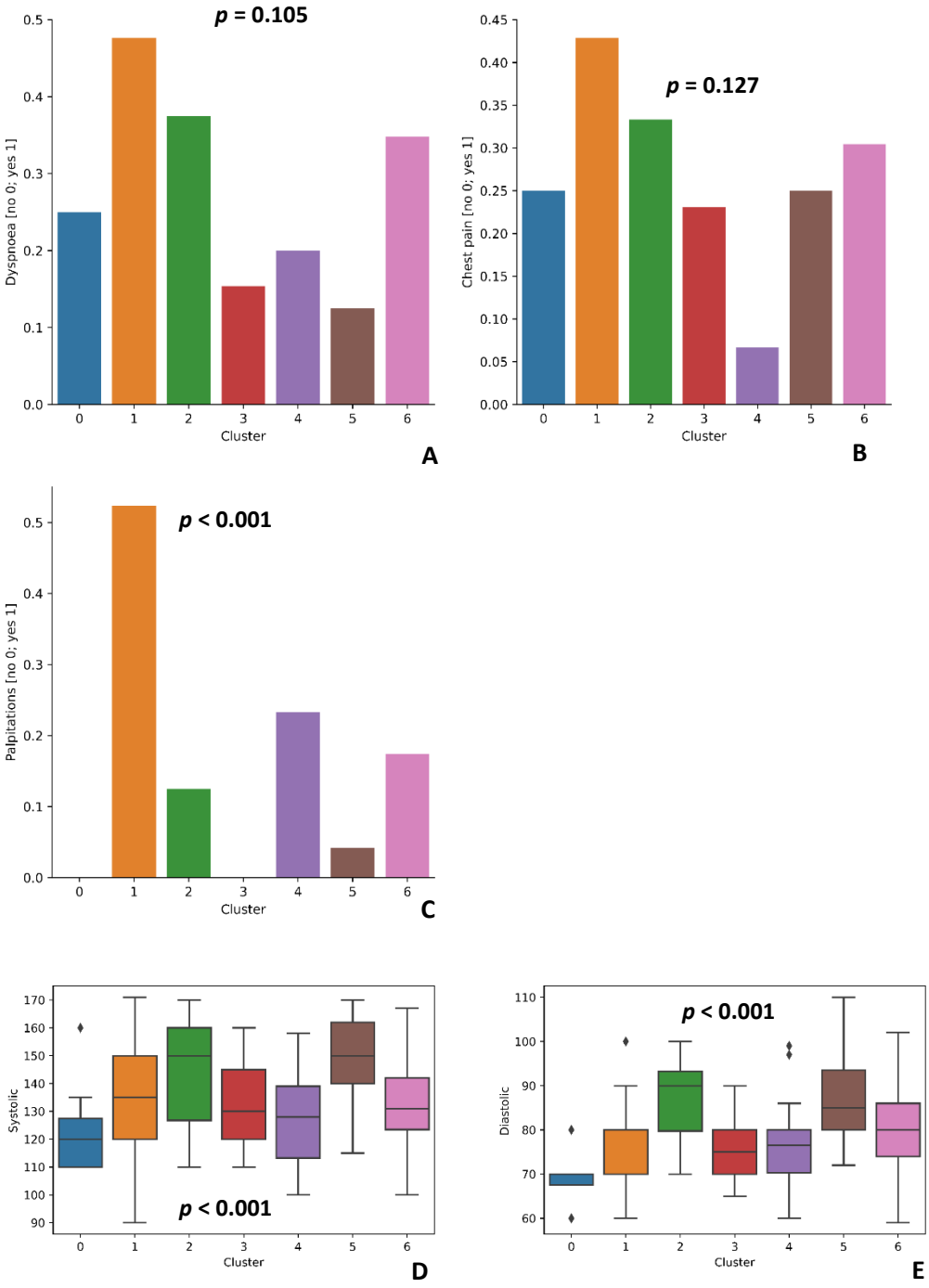

Figure 103. Seven clusters setting, features: A — Dyspnea, B — Chest pain, C — Palpitations, D — Systolic blood pressure, E — Diastolic blood pressure

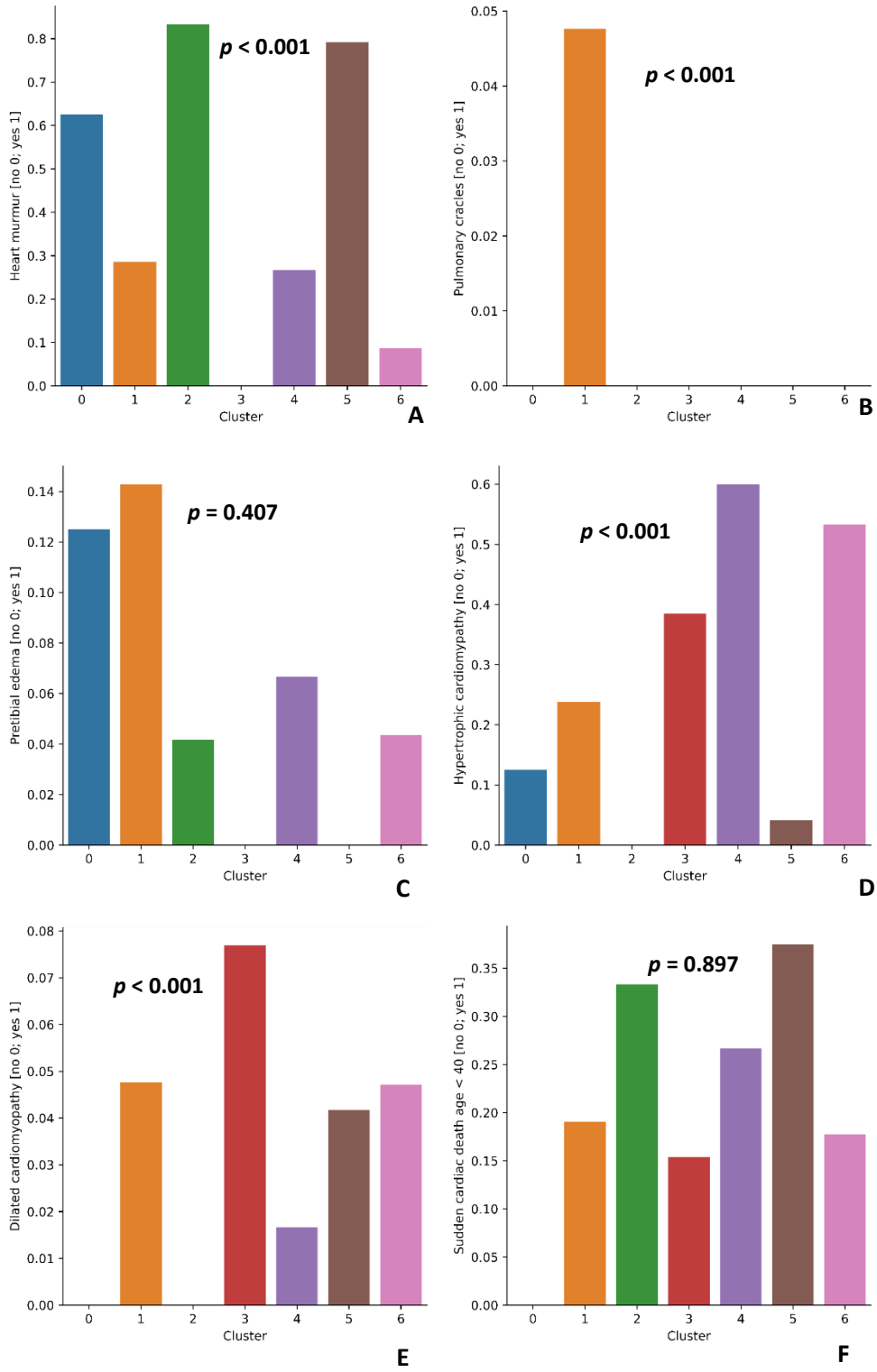

Figure 104. Seven clusters setting, features: A — Heart murmur, B — Pulmonary crackles, C — Pretibial edema, D — HCM in family history, E — DCM in family history, F — SCD in age < 40 in family history

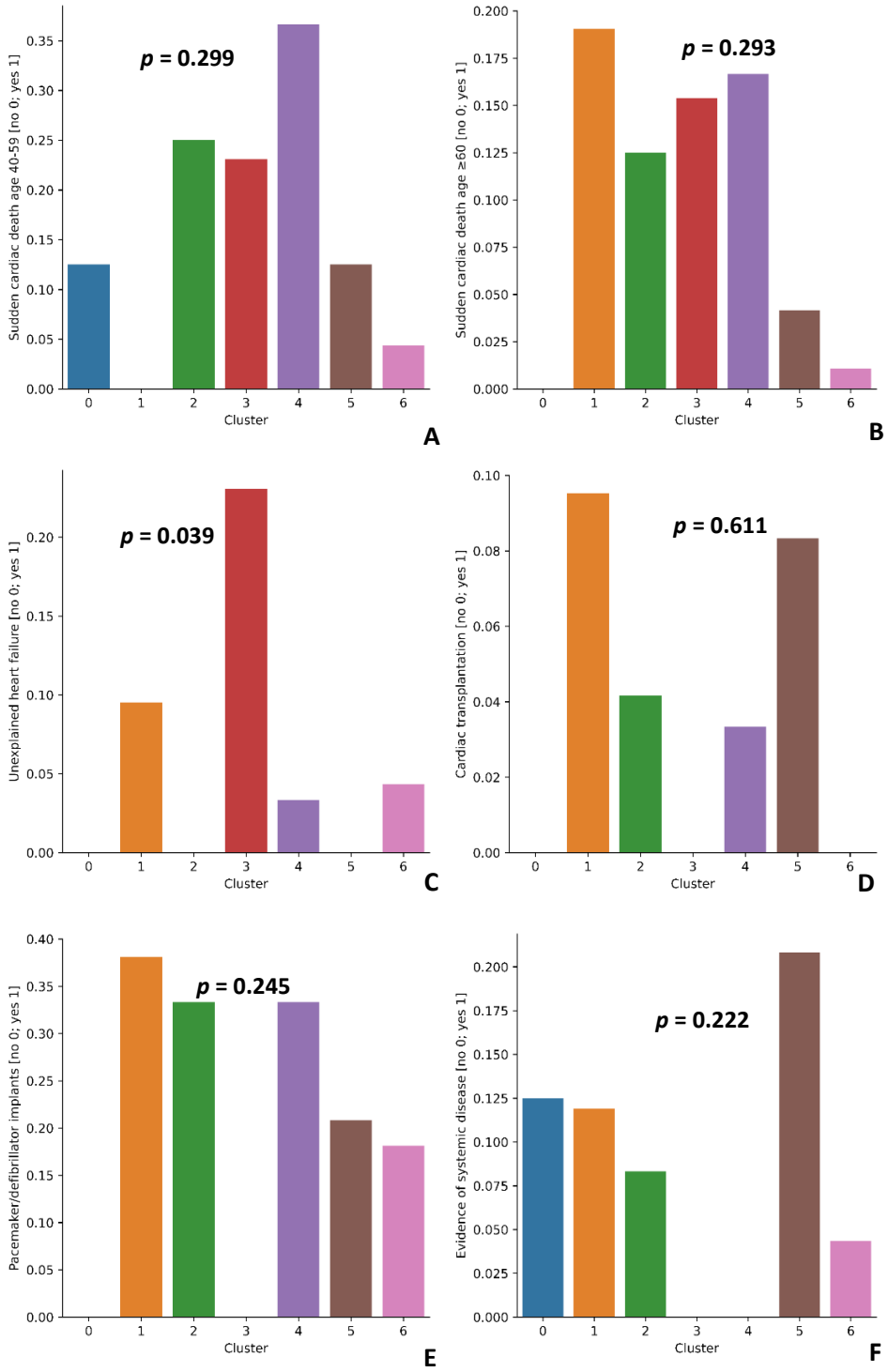

Figure 105. Seven clusters setting, features: A — SCD in age 40-59 in family history, B — SCD in age ≥ 60 in family history, C — Unexplained HF in family history, D — Cardiac transplantation in family history, E — Pacemaker/defibrillator implants in family history, F — Evidence of systemic disease in family history

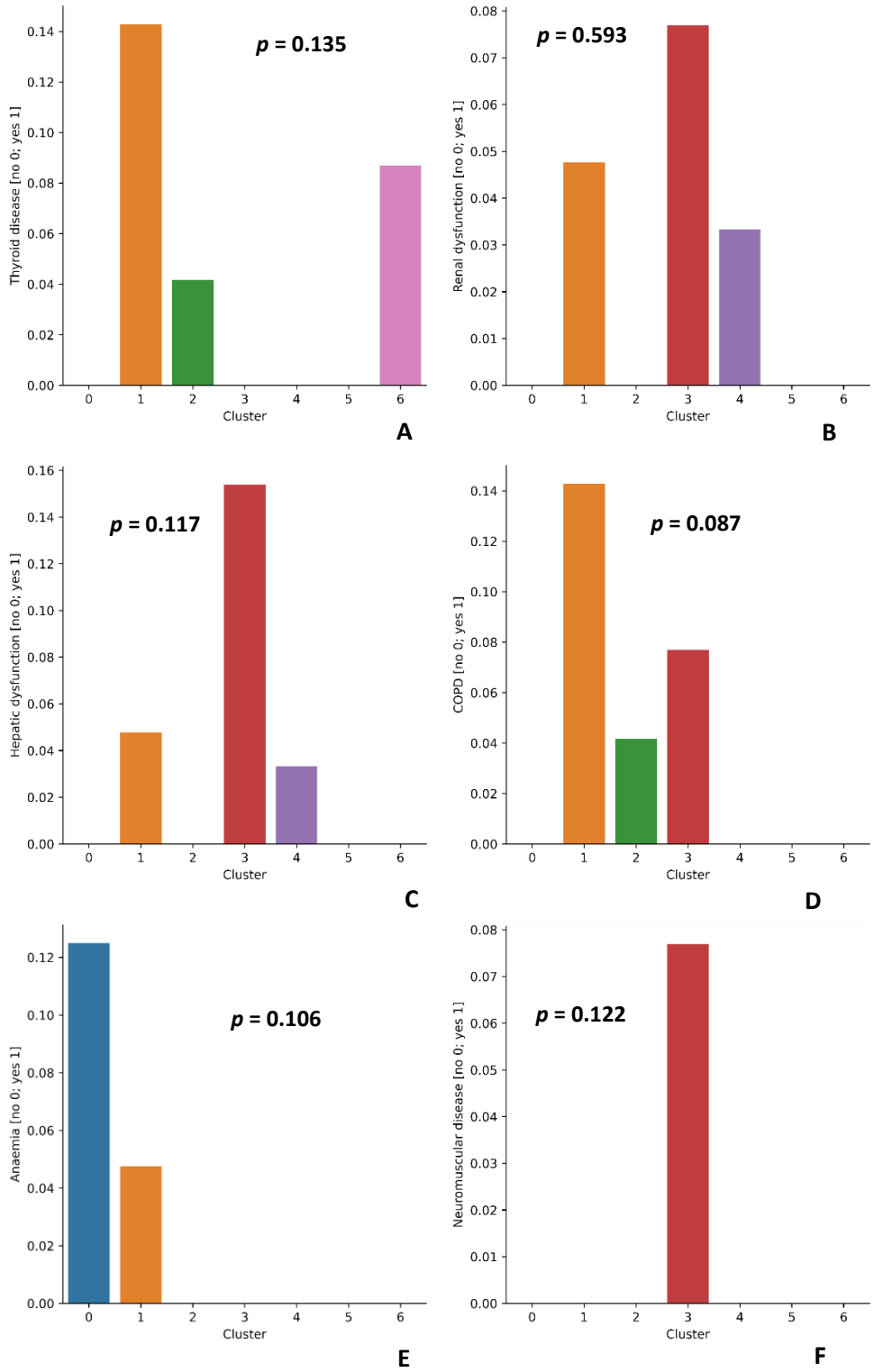

Figure 106. Seven clusters setting, features: A — Thyroid disease, B — Renal dysfunction, C — Hepatic dysfunction, D — COPD, E — Anemia, F — Neuromuscular disease

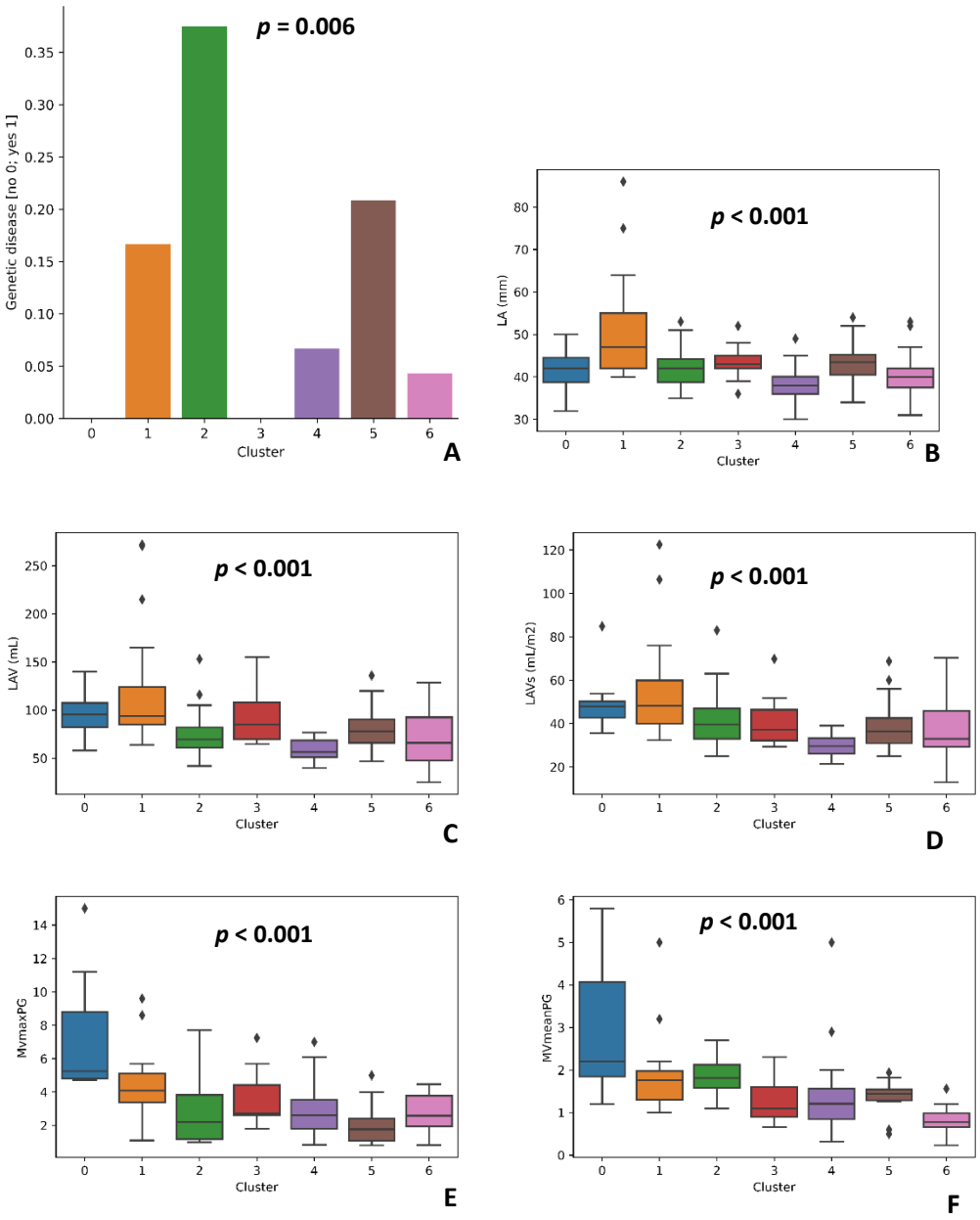

Figure 107. Seven clusters setting, features: A — Genetic disease, B — LA, C — LAV, D — LAVs, E — MV maxPG, F — MV meanPG

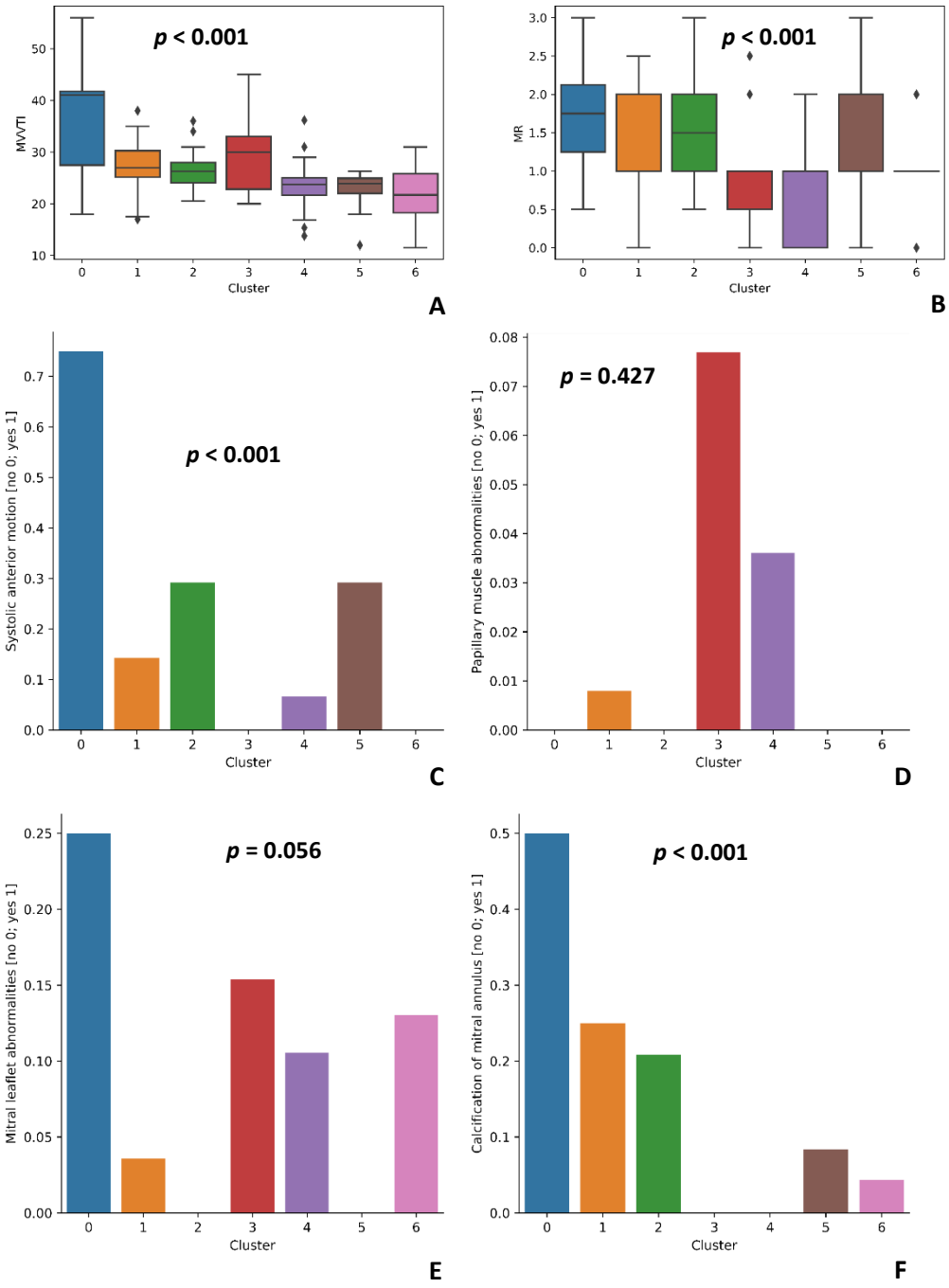

Figure 108. Seven clusters setting, features: A — MVVTI, B — MR, C — SAM, D — Papillary muscle abnormalities, E — Mitral leaflet abnormalities, F — Calcification of mitral annulus

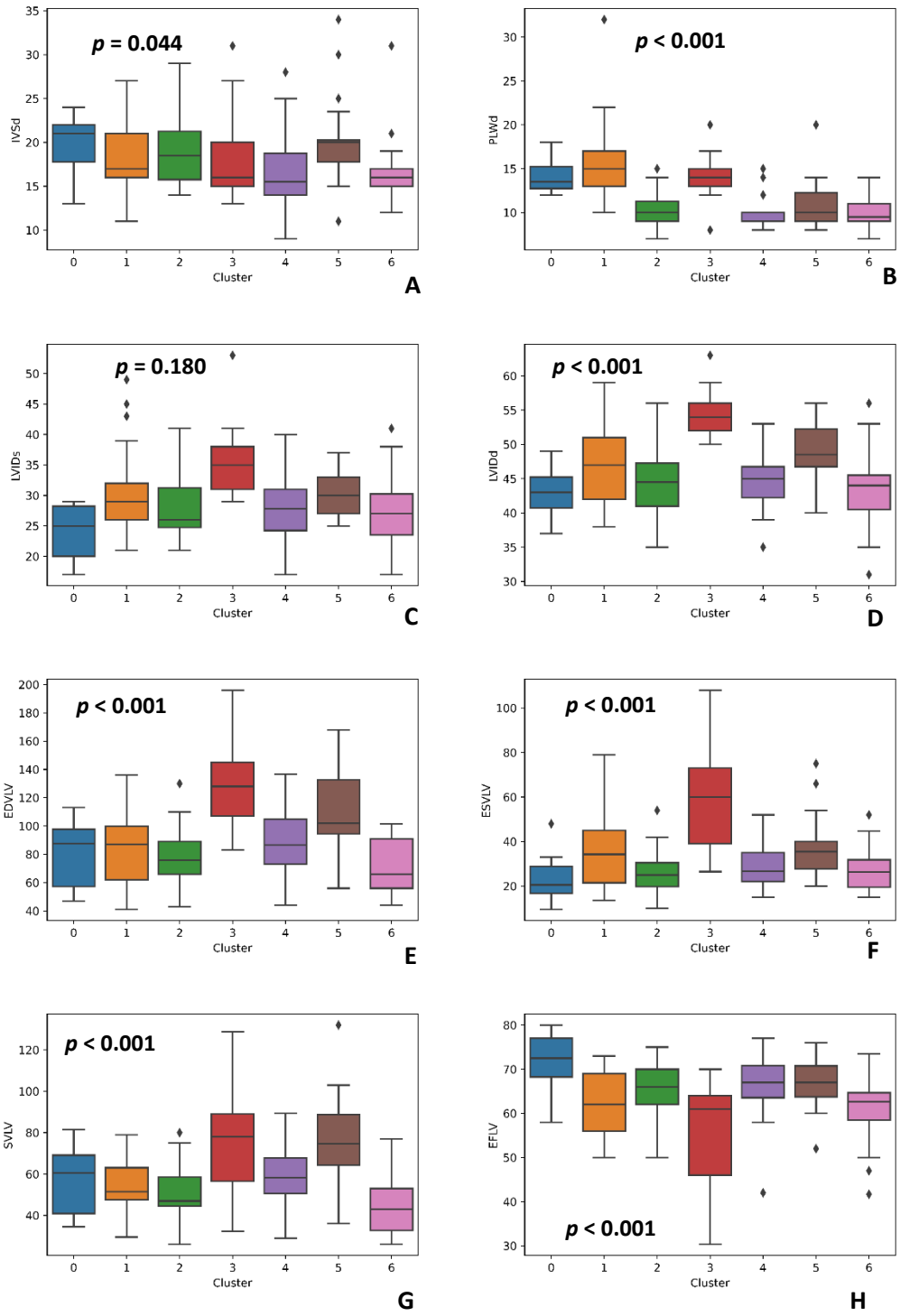

Figure 109. Seven clusters setting, features: A — IVSd, B — PLWd, C — LVIDs, D — LVIDd, E — EDVLV, F — ESVLV, G — SVLV, F — EFLV

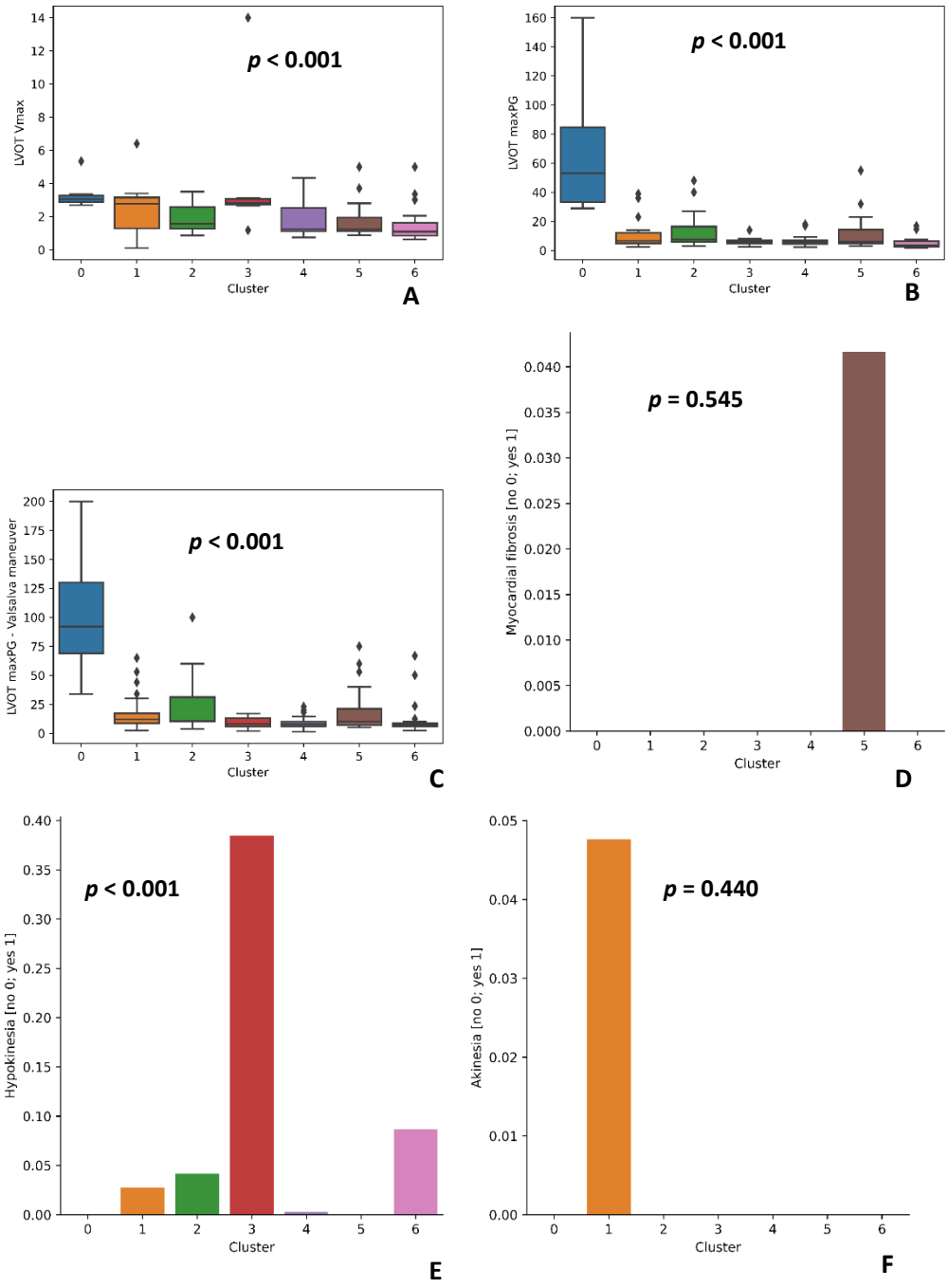

Figure 110. Seven clusters setting, features: A — LVOT Vmax, B — LVOT maxPG, C — LVOT maxPG - Valsalva maneuver, D — Myocardial fibrosis, E — Hypokinesia, F — Akinesia

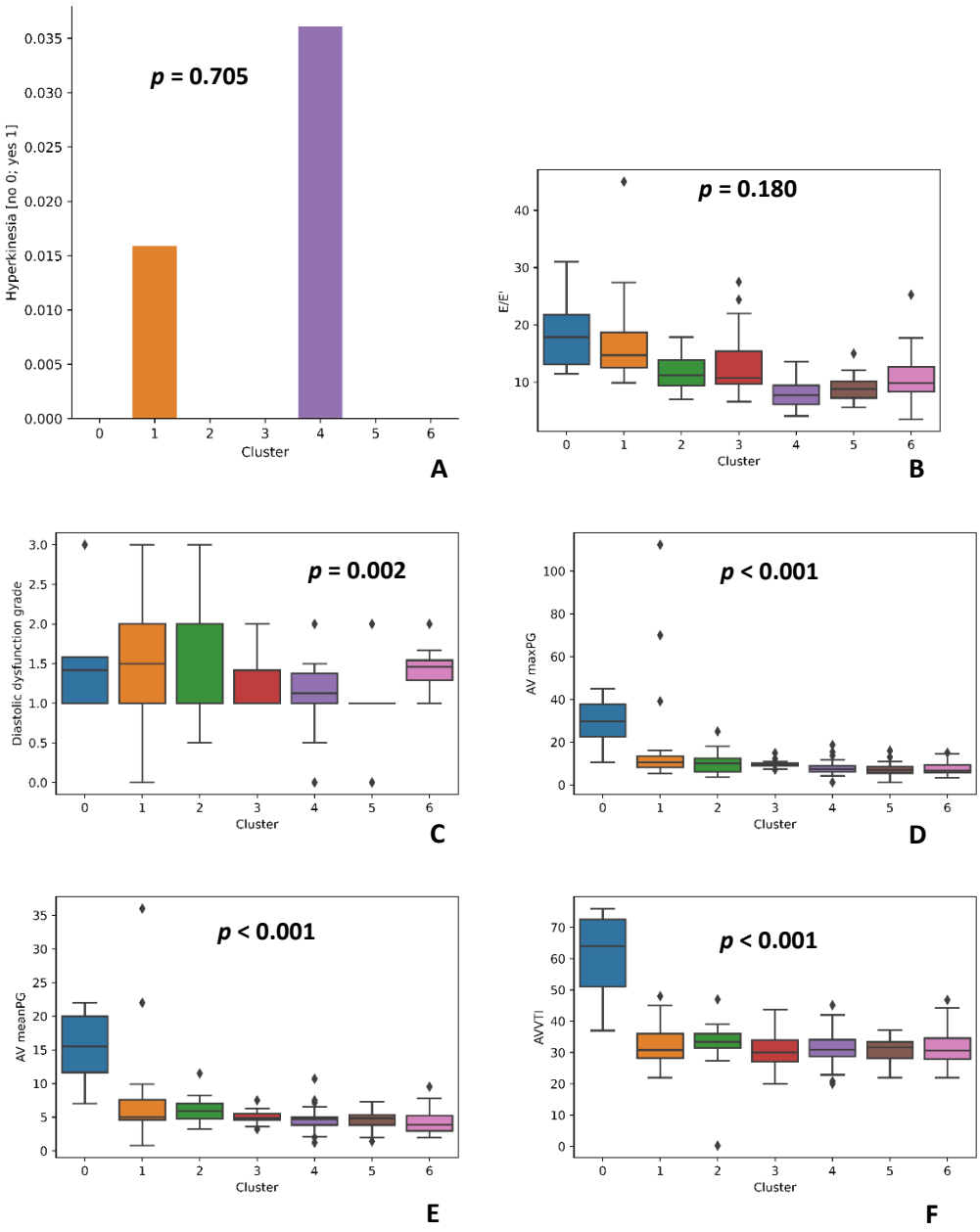

Figure 111. Seven clusters setting, features: A — Hyperkinesia, B — E/E', C — Diastolic dysfunction grade, D — AV maxPG, E — AV meanPG, F — AVVTI

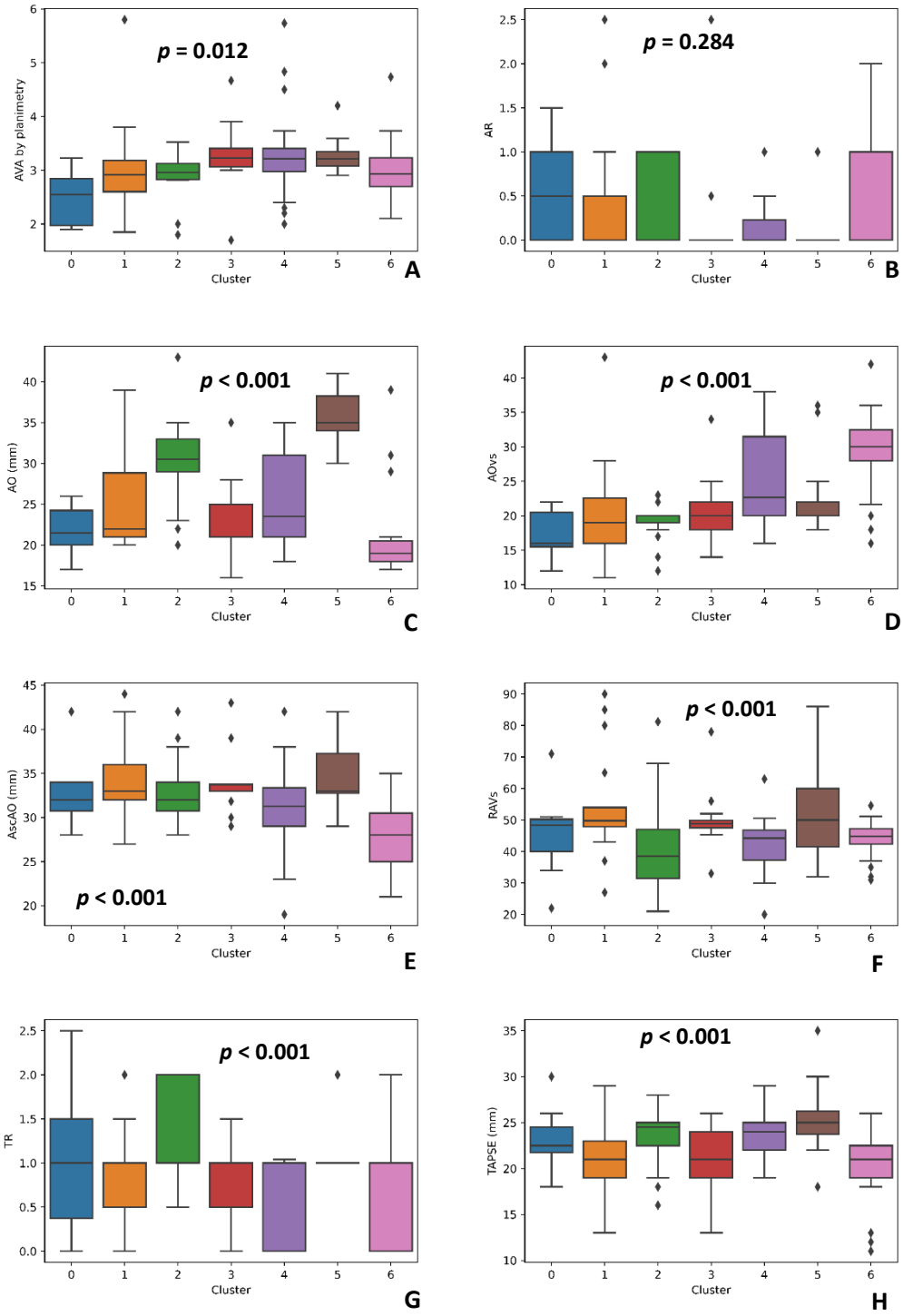

Figure 112. Seven clusters setting, features: A — AVA by planimetry, B — AR, C — AO, D — AOvs, E — AscAO, F — RAVs, G — TR, H — TAPSE

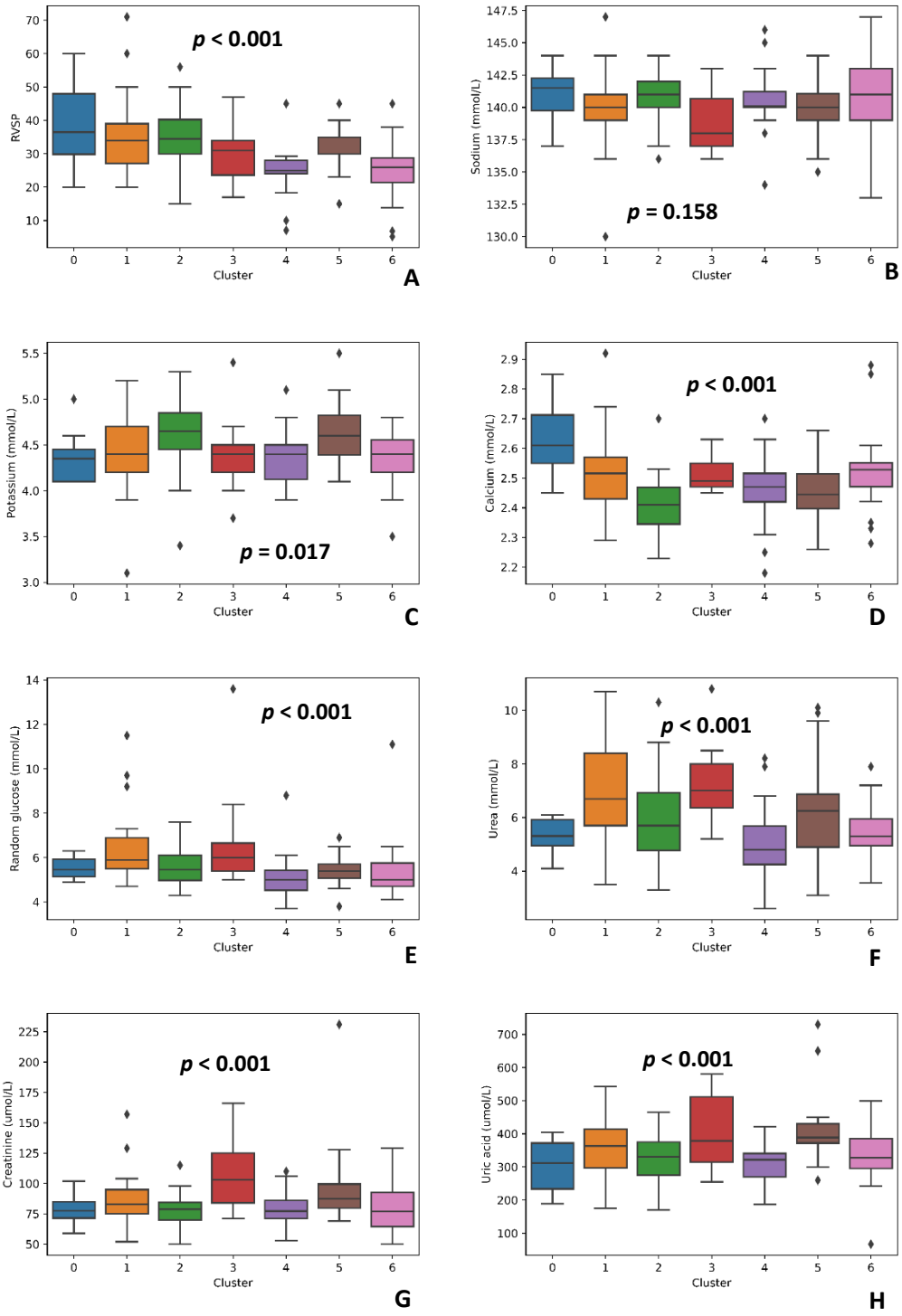

Figure 113. Seven clusters setting, features: A — RVSP, B — Sodium, C — Potassium, D — Calcium, E — Random glucose, F — Urea, G — Creatinine, H — Uric acid

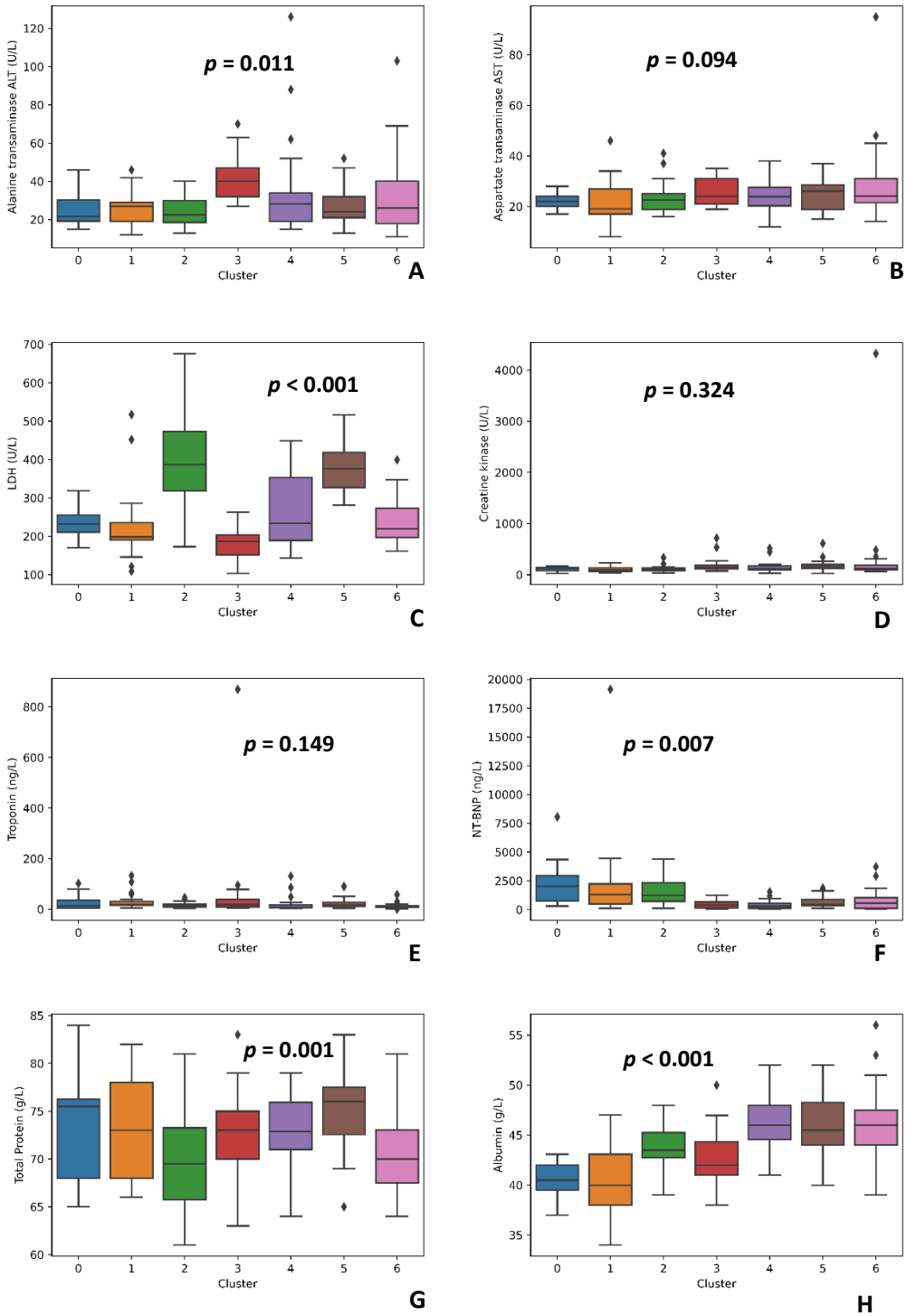

Figure 114. Seven clusters setting, features: A — ALT, B — AST, C — LDH, D — Creatine-kinase, E — Troponin, F — NT-BNP, G — Total protein, H — Albumin

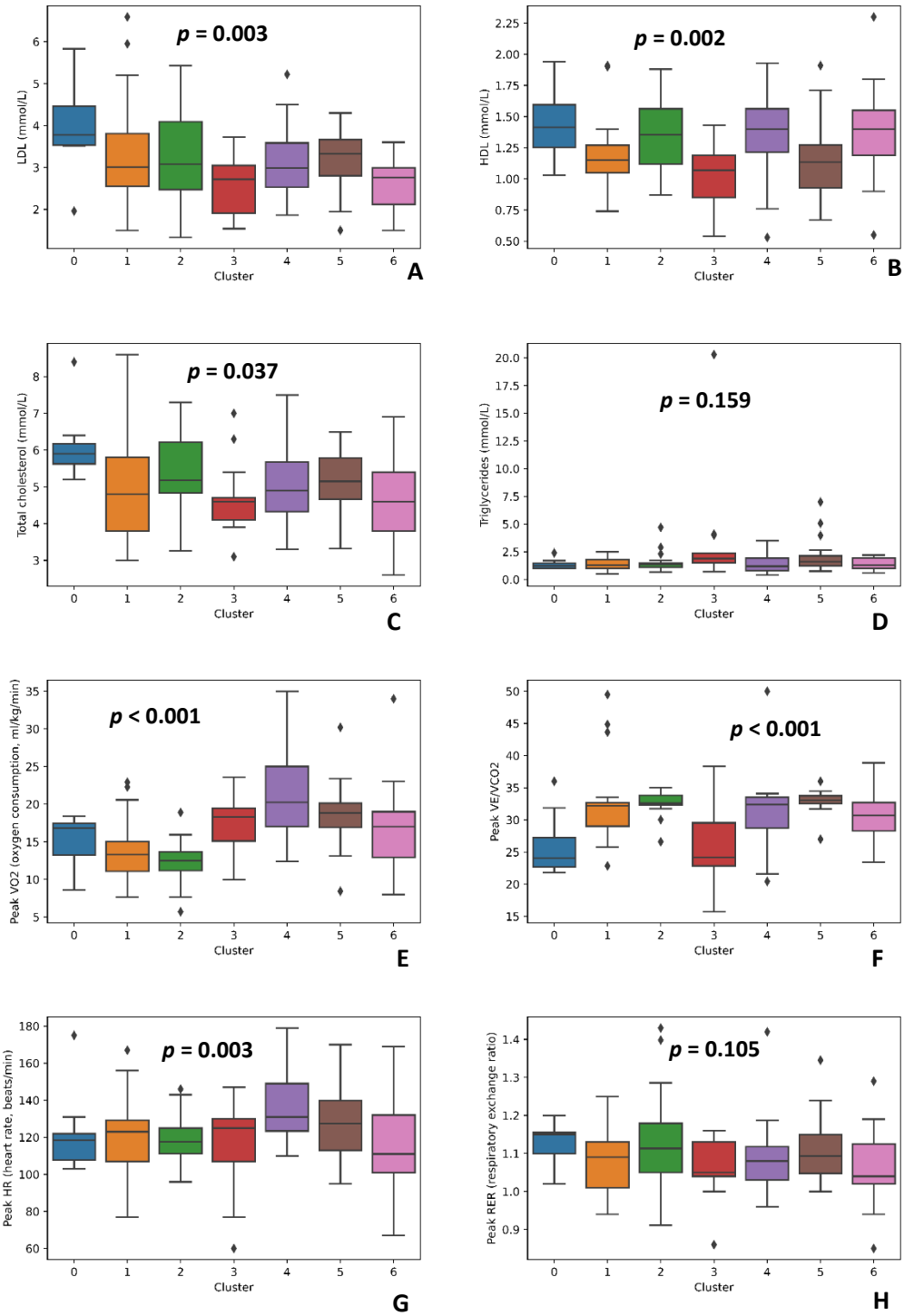

Figure 115. Seven clusters setting, features: A — LDL, B — HDL, C — Total cholesterol, D — Triglycerides, E — A Peak VO_2, F — Peak VE/VCO_2, G — Peak HR, H — Peak RER

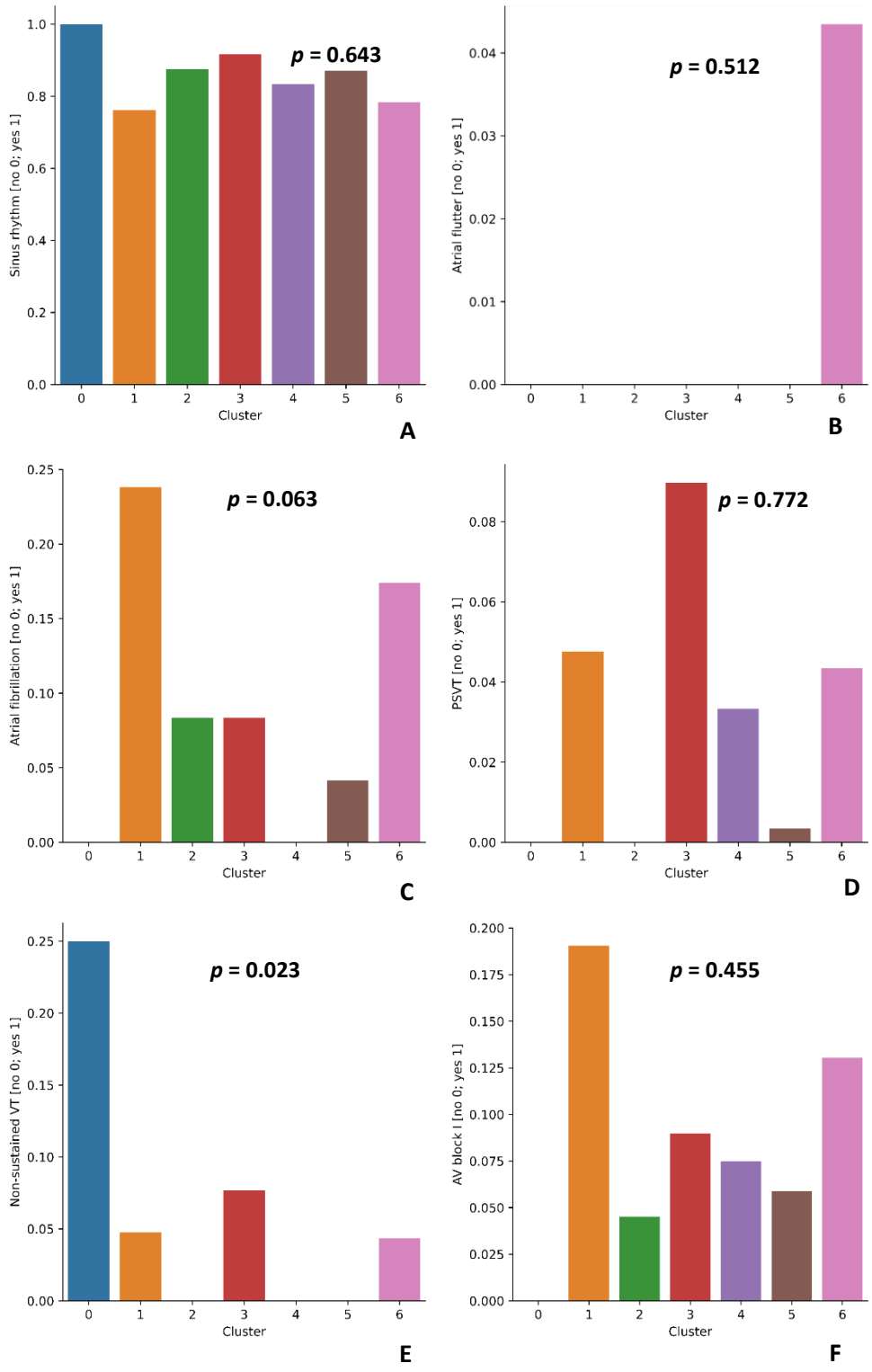

Figure 116. Seven clusters setting, features: A — Sinus rhythm, B — Atrial flutter, C — AF, D — PSVT, E — Non-sustained VT, F — AV block I

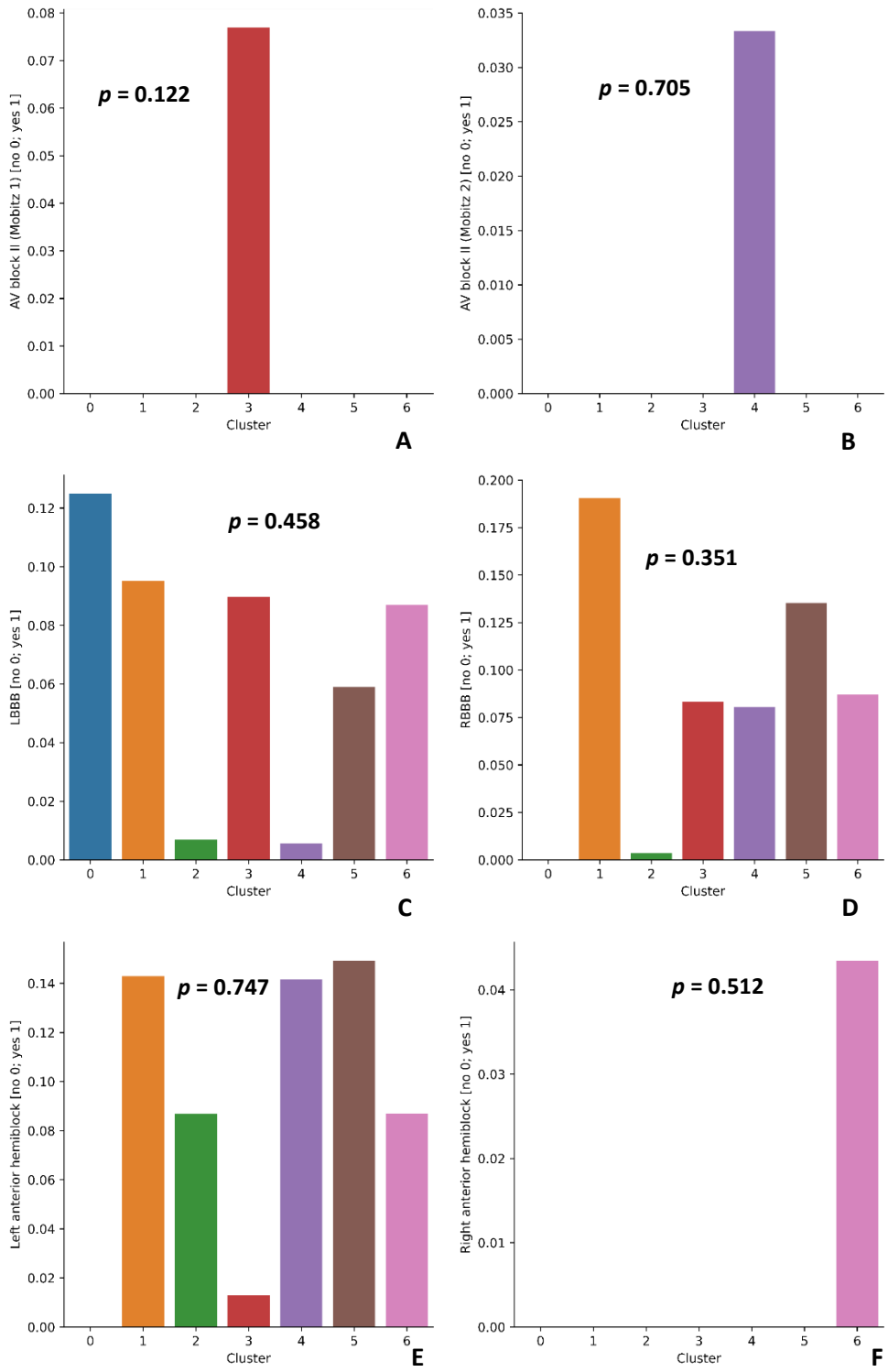

Figure 117. Seven clusters setting, features: A — AV block II (Mobitz 1), B — AV block II (Mobitz 2), C — LBBB, D — RBBB, E — Left anterior hemiblock, F — Right anterior hemiblock

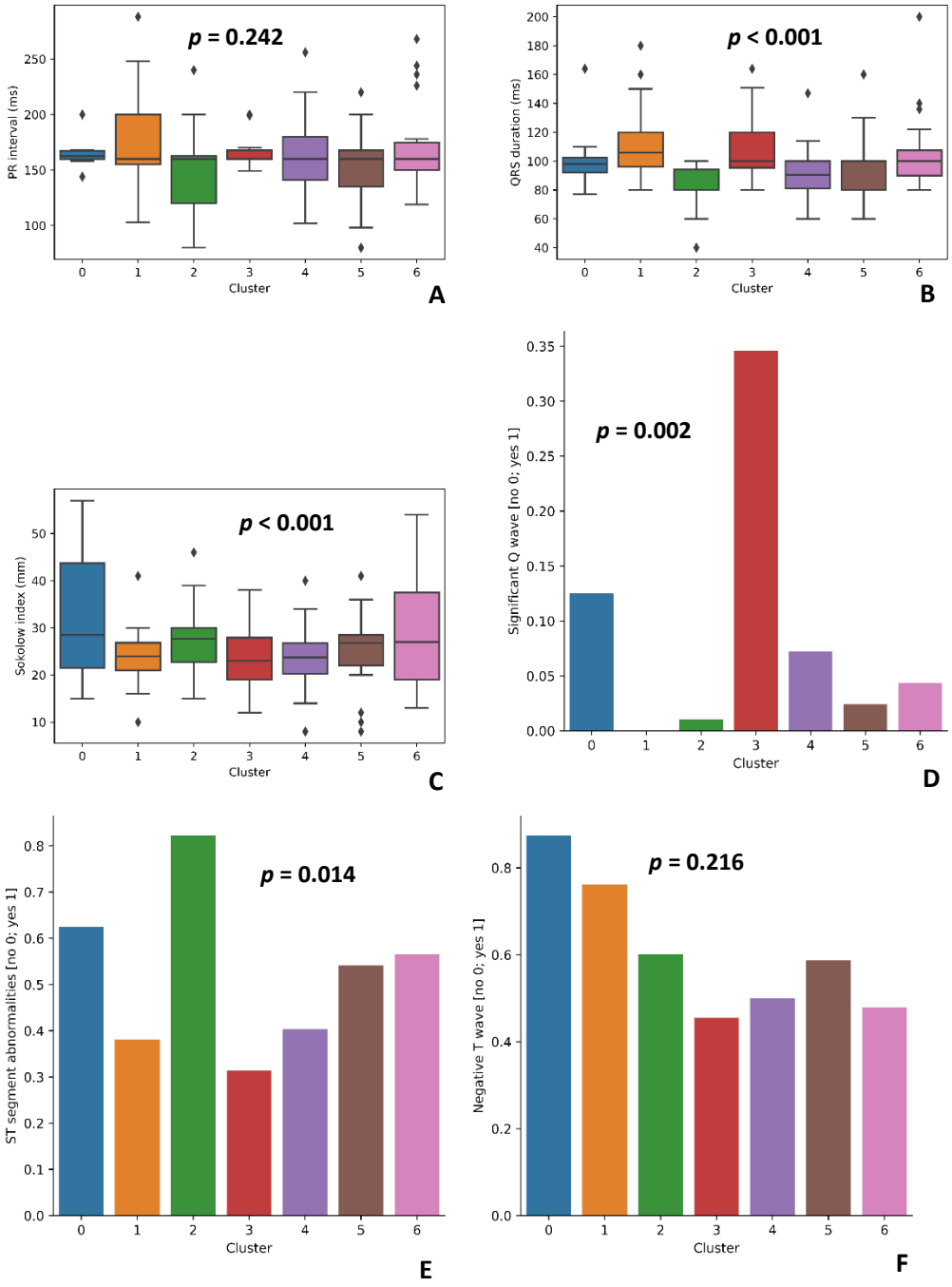

Figure 118. Seven clusters setting, features: A — PR interval, B — QRS duration, C — Sokolow index, D — Significant Q wave, E — ST segment abnormalities, F — Negative T wave

An approximate interpretation of clustering-logic for 7 clusters is presented in the form of visualized decision tree (Figure 119). Feature importance of random forest trained on the same dataset was determined, with labels as assigned by clustering (Table 15).

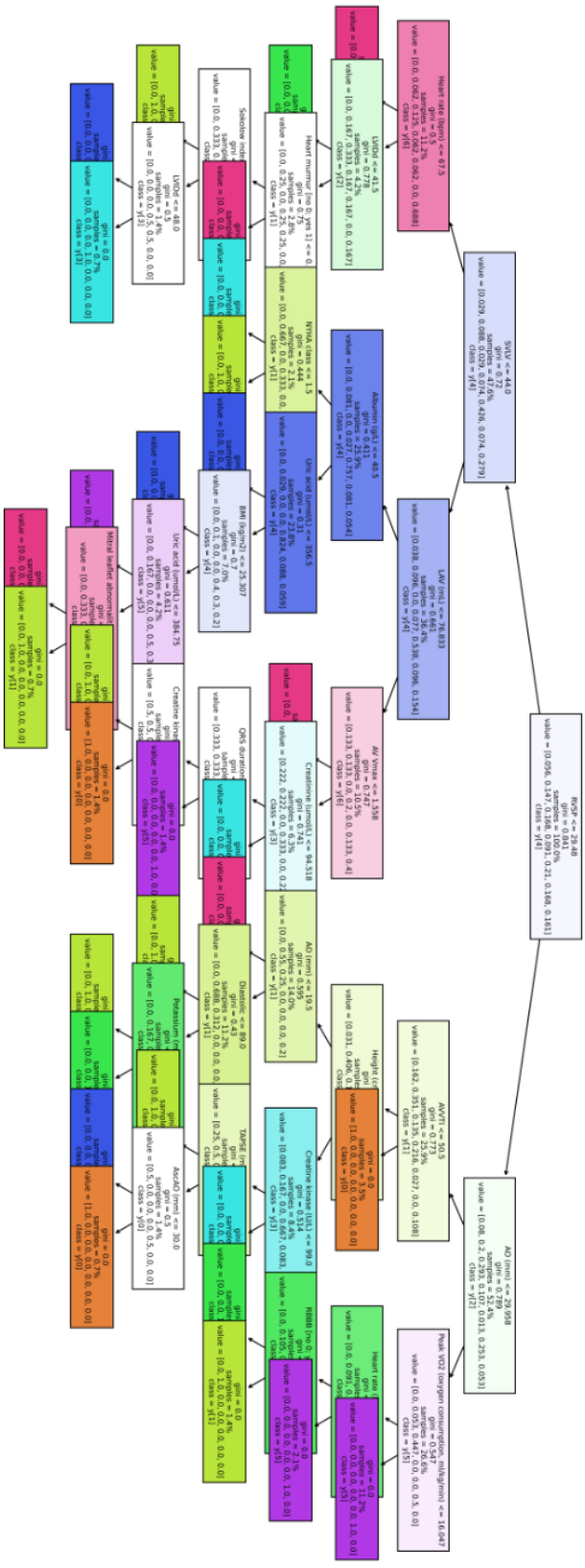

Figure 119. An approximate interpretation of clustering-logic for 7 clusters

Table 15. Feature importance — Top 35 features for distinguishing 7 clusters

Feature	Estimated importance
SVLV	0.015489
ESVLV	0.015328
AOvs	0.015201
MVVTI	0.015017
LVIDd	0.014685
NT-BNP	0.014286
Peak VE/VCO_2	0.012190
LVOT maxPG - Valsalva maneuver	0.012095
AscAO	0.011887
Albumin	0.011760
MV maxPG	0.011565
QRS duration	0.011490
Heart murmur [yes/no]	0.011335
ALT	0.010735
Sex	0.010719
Serum calcium	0.010673
Peak HR	0.010617
RAVs	0.010432
Serum urea	0.010237
Serum creatinine	0.010163
LVOT Vmax	0.010121
Sokolow index	0.010029
EFLV	0.010012
HDL	0.009956
BMI	0.009584
Creatine-kinase	0.009566
Systolic	0.009432
Diastolic dysfunction grade	0.009342
AV meanPG	0.009158
Troponin	0.008994
LDL	0.008901
Total serum protein	0.008707
Random glucose	0.008429
Heart rate	0.008322
TAPSE	0.008313

4.1.3. Genotype-cluster associations

4.1.3.1. Hierarchical clustering

According to hierarchical clustering, there is no association between clusters and genotype (Figure 120).

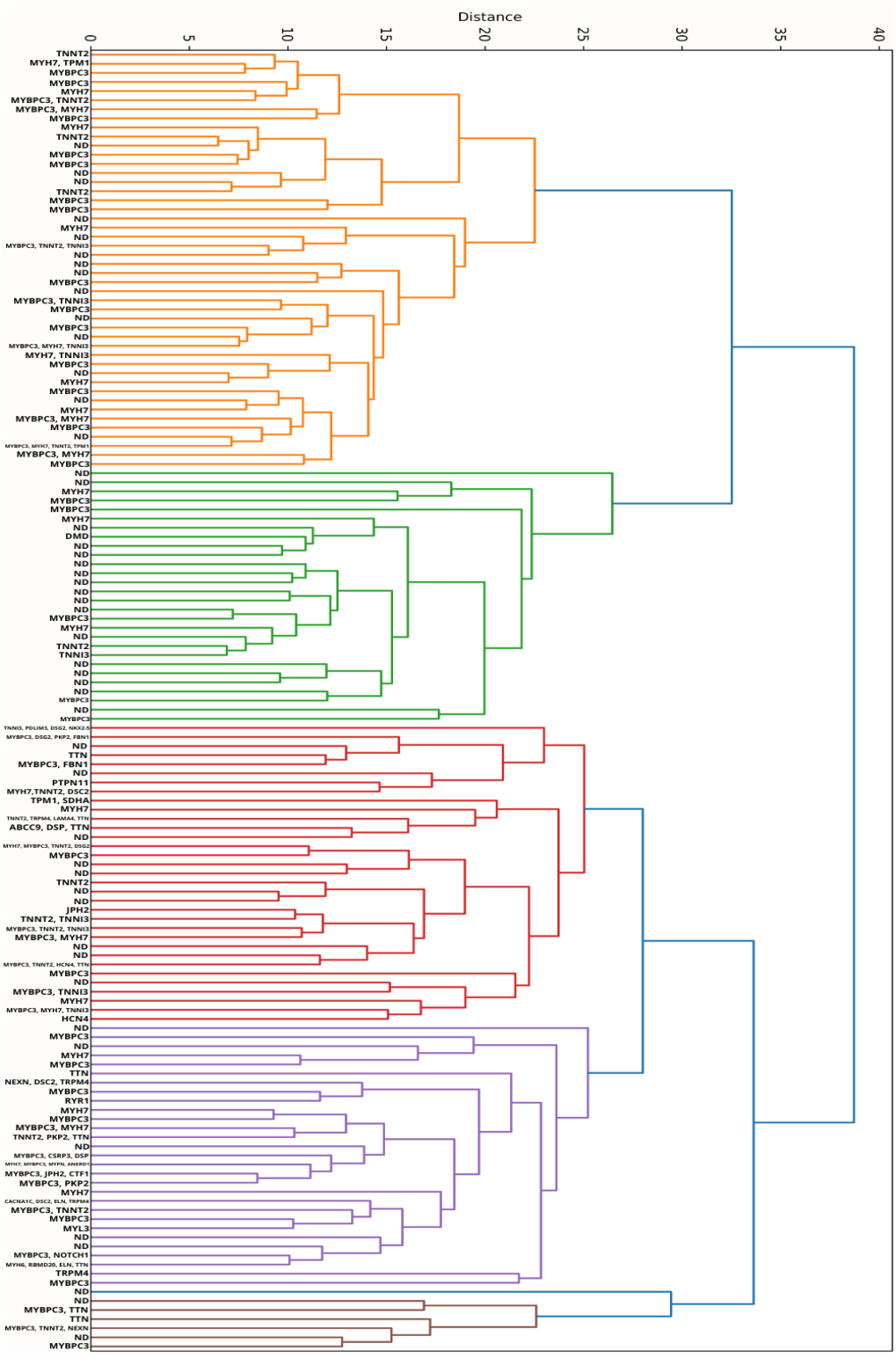

Figure 120. Hierarchical clustering — dendrogram

4.1.3.2. K-Prototype clustering

Some statistically significant correlations were found between clusters determined using K-Prototype clustering and mutated genes.

4.1.3.2.1. Two clusters

No statistically significant correlation between genotypes and subphenotypes (belonging to determined clusters) was found in the 2-subphenotypes (2-clusters) setting (Figure 121).

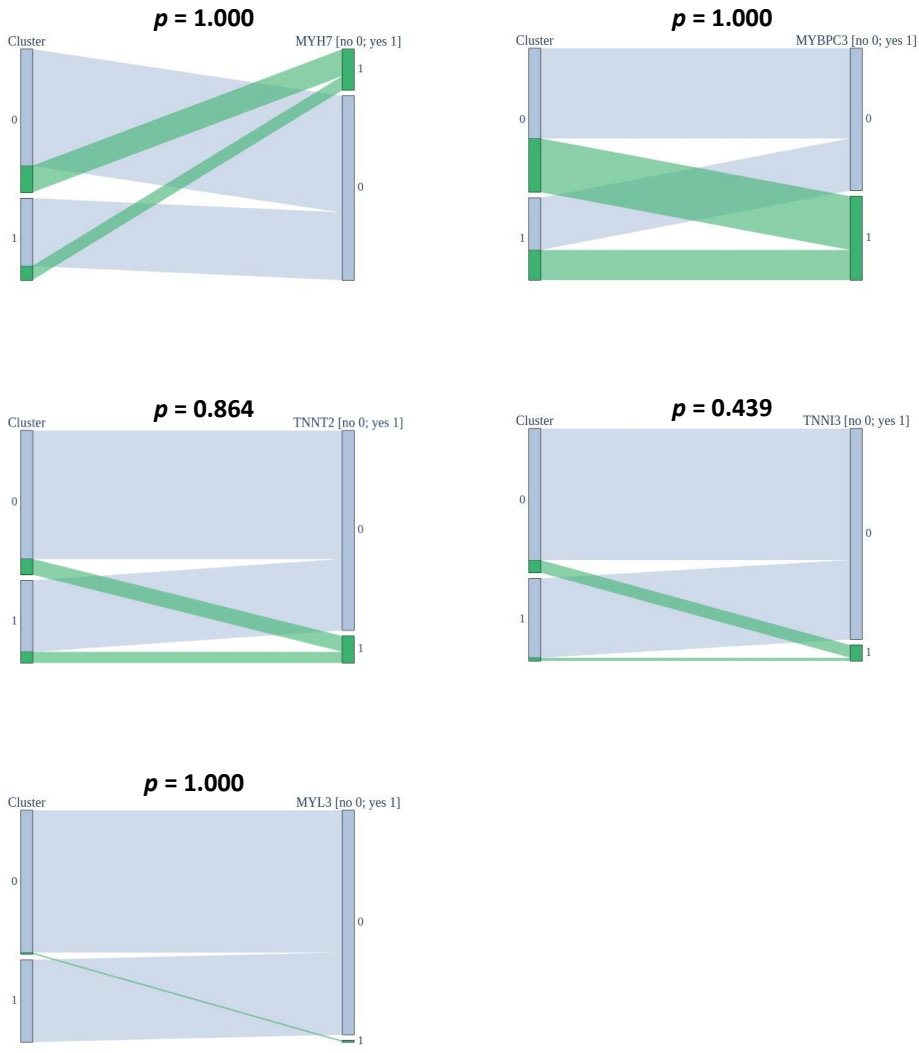

Figure 121. Correlations between mutated genes and subphenotypes in 2-subphenotypes (2-clusters) setting

4.1.3.2.2. Three clusters

A statistically significant correlation between mutated *TNNI3* and subphenotypes (p = 0.041) was found in the 3-subphenotypes (3-clusters) setting (Figure 122).

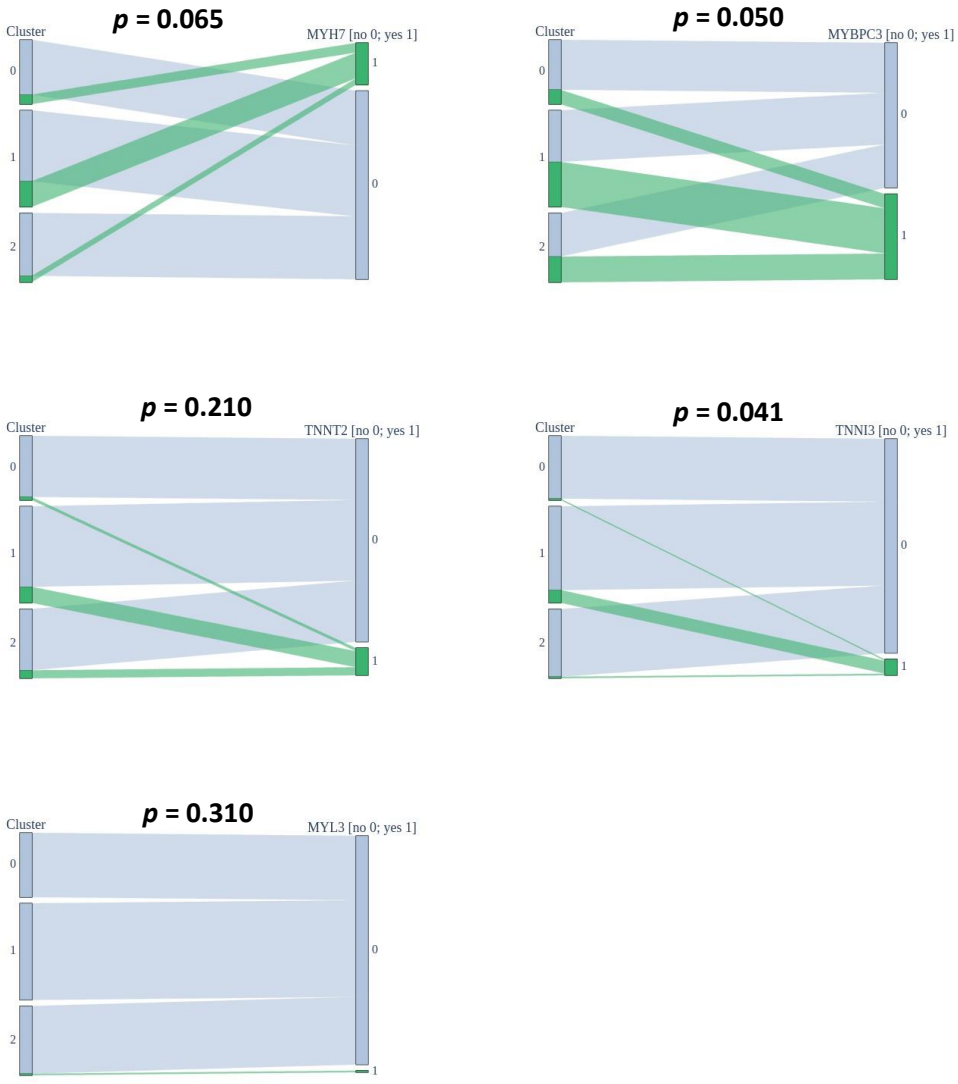

Figure 122. Correlations between mutated genes and subphenotypes in 3-subphenotypes (3-clusters) setting

4.1.3.2.3. Four clusters

Statistically significant correlations between mutated *MYBPC3* and subphenotypes (p = 0.038), and mutated *TNNI3* and subphenotypes (p = 0.045) were found in the 4-subphenotypes (4-clusters) setting (Figure 123).

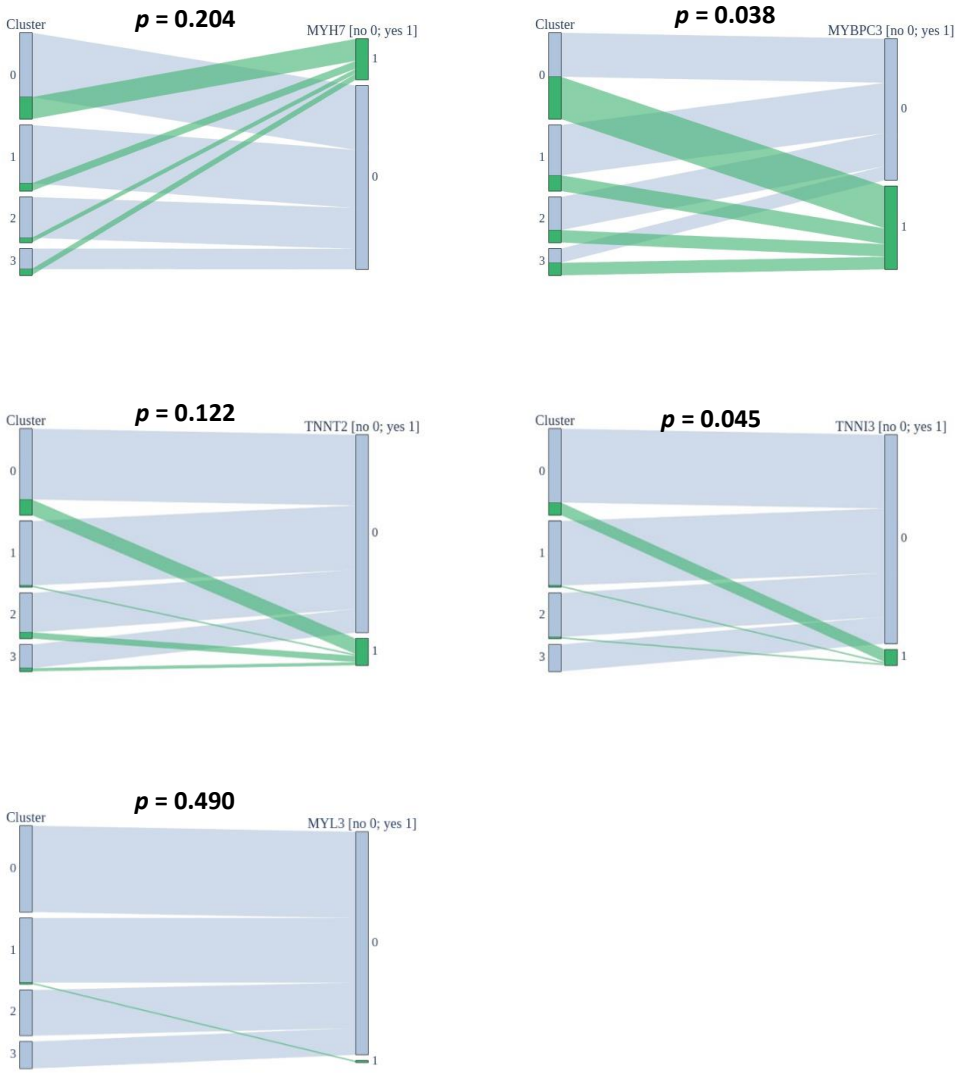

Figure 123. Correlations between mutated genes and subphenotypes in 4-subphenotypes (4-clusters) setting (283)

4.1.3.2.4. Five clusters

A statistically significant correlation between mutated *MYH7* and subphenotypes ($p = 0.045$) was found in 5-subphenotypes (5-clusters) setting (Figure 124).

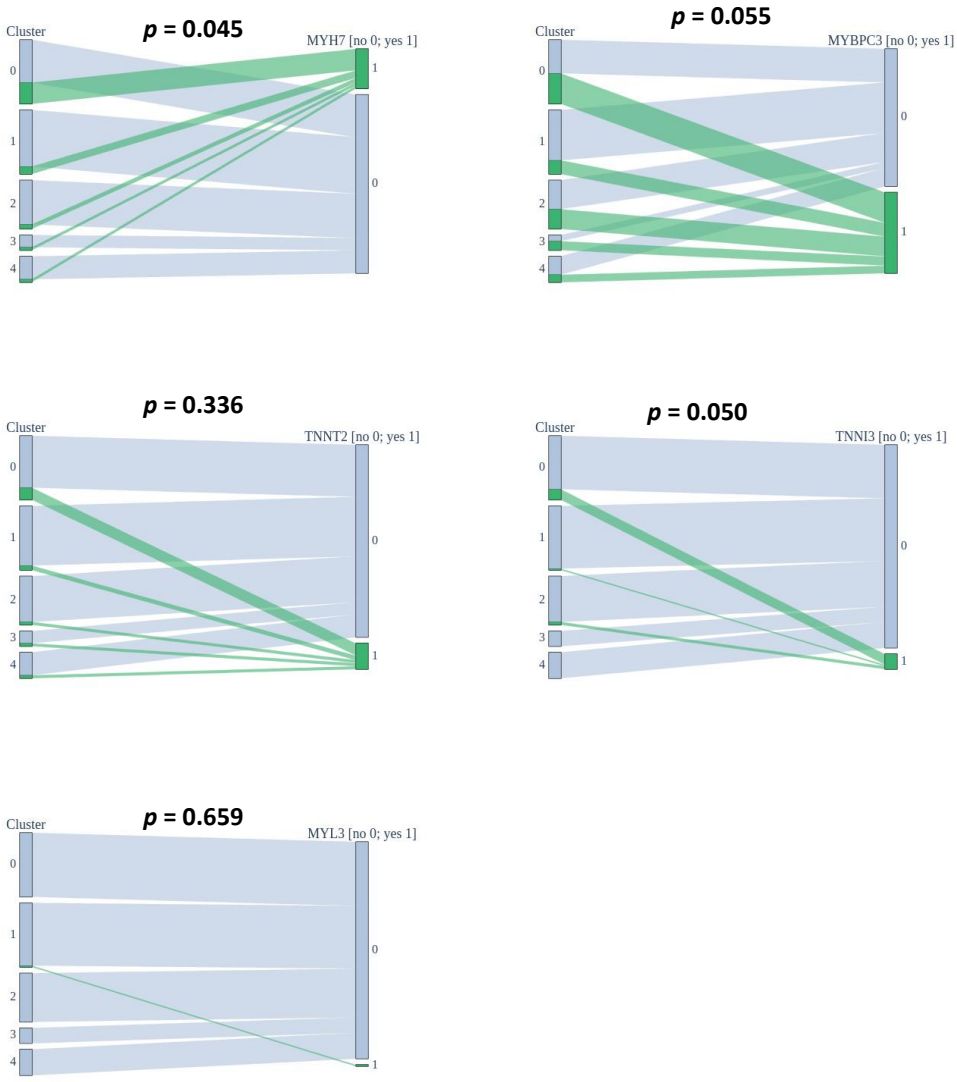

Figure 124. Correlations between mutated genes and subphenotypes in 5-subphenotypes (5-clusters) setting

4.1.3.2.5. Six clusters

A statistically significant correlation between mutated *MYBPC3* and subphenotypes (p = 0.044) was found in 6-subphenotypes (6-clusters) setting (Figure 125).

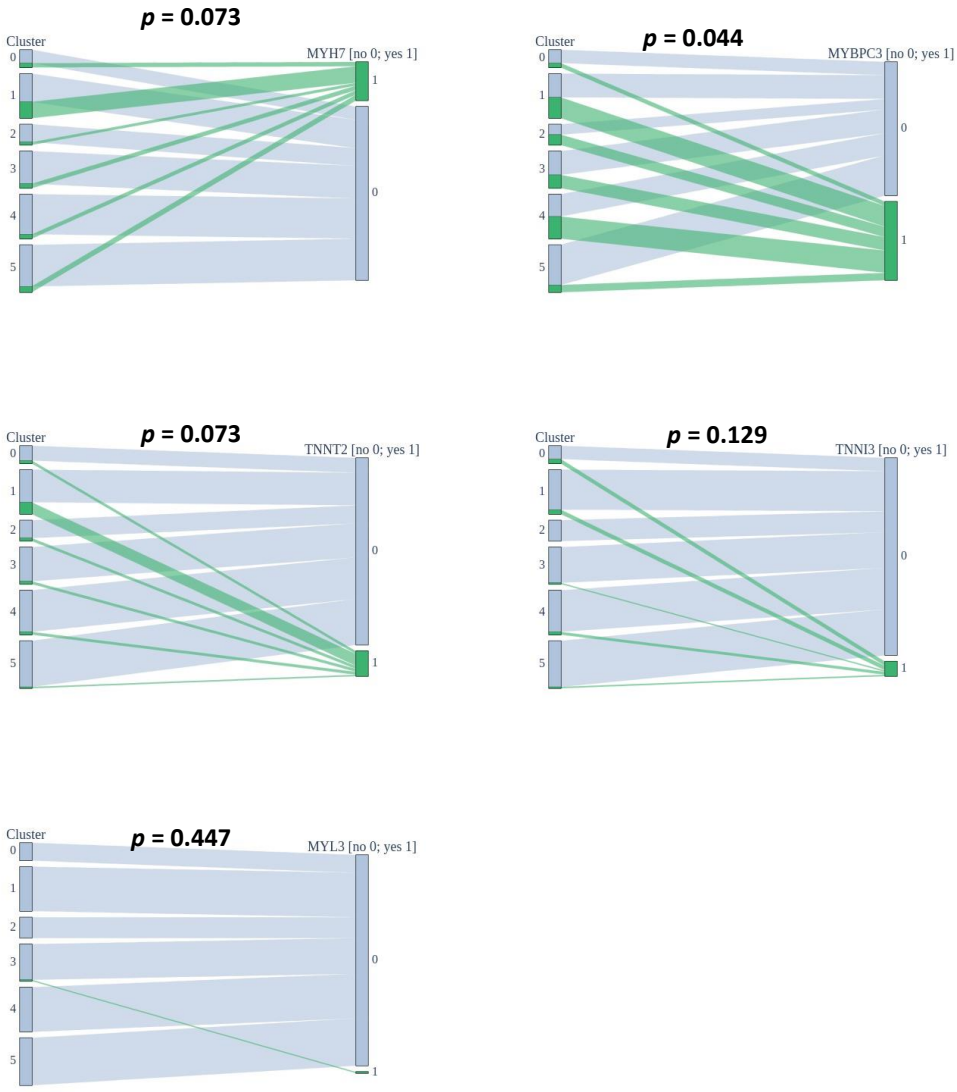

Figure 125. Correlations between mutated genes and subphenotypes in 6-subphenotypes (6-clusters) setting

4.1.3.2.6. Seven clusters

No statistically significant correlation between genotypes and subphenotypes (belonging to determined clusters) was found in 7-subphenotypes (7-clusters) setting (Figure 126).

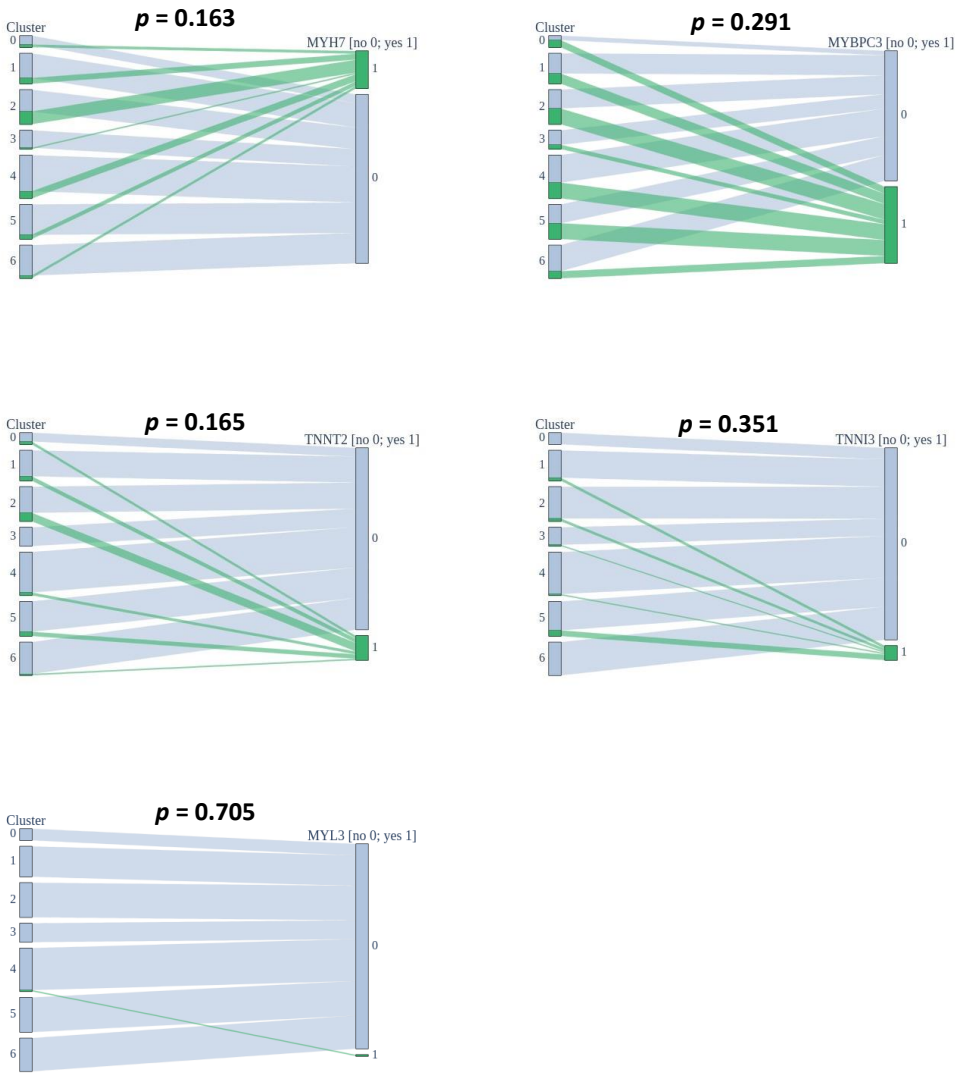

Figure 126. Correlations between mutated genes and subphenotypes in 7-subphenotypes (7-clusters) setting

4.1.4. Genotype-phenotype associations

4.1.4.1. Predicting phenotypic outcomes using only genetic data

Performance of generated models and relationships between mutated genes and phenotypic outcomes have been shown in Tables 16-105. Some composite outcomes were predicted as well: signs (consisting of heart murmur, pulmonary crackles, pleural effusion, pretibial edema, and venous congestion), symptoms (consisting of fatigue, dyspnea, chest pain, palpitations, and syncope), mitral valve abnormalities (consisting of papillary muscle abnormalities, mitral leaflet abnormalities, calcification of mitral annulus, and SAM), ventricular conduction disorders (consisting of LBBB, RBBB, left anterior hemiblock, and right anterior hemiblock), left ventricular kinetics disorders (consisting of hypokinesia, akinesia, dyskinesia, hyperkinesia, and myocardial fibrosis), atrial conduction disorders (consisting of AV block I, AV block II [Mobitz 1], AV block II [Mobitz 2], and AV block III), myocardial injury — infarction (consisting of significant Q wave, ST segment abnormalities, negative T wave).

Table 16. Predicting palpitations from mutated causative genes [Support: Yes (n = 14); No (n = 54)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.31	0.74	0.37	0.85	0.23	0.00	0.48	0.50	0.38	0.46	0.20	0.21	0.18	0.21
RF	0.74	0.74	0.85	0.85	0.00	0.00	0.48	0.50	0.46	0.46	0.20	0.21	0.21	0.21
Log Reg	0.44	0.79	0.59	0.89	0.14	0.00	0.48	0.50	0.36	0.50	0.20	0.21	0.18	0.21
GaussianNB	0.26	0.26	0.26	0.26	0.26	0.26	0.48	0.50	0.40	0.40	0.20	0.21	0.18	0.18
Ridge	0.74	0.74	0.85	0.85	0.00	0.00	0.48	0.50	0.46	0.46	0.20	0.21	0.21	0.21
SVC	0.74	0.74	0.85	0.85	0.00	0.00	0.49	0.50	0.46	0.46	0.20	0.21	0.21	0.21
Linear SVC	0.74	0.74	0.85	0.85	0.00	0.00	0.48	0.50	0.46	0.46	0.20	0.21	0.21	0.21

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 17. Relationships between mutated causative genes and palpitations (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
TNNT2	0.097529
TNNI3	-0.058930
MYL3	0.069403
TPM1	0.069403

Between mutated causative genes and palpitations negligible relationships were found.

Table 18. Predicting palpitations from mutated genes [Support: Yes (n = 14); No (n = 54)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.62	0.76	0.75	0.86	0.24	0.27	0.62	0.46	0.49	0.56	0.26	0.21	0.20	0.24
RF	0.76	0.76	0.86	0.86	0.27	0.27	0.62	0.46	0.56	0.56	0.26	0.21	0.24	0.24
Log Reg	0.59	0.79	0.71	0.88	0.26	0.12	0.64	0.65	0.50	0.53	0.31	0.31	0.21	0.23
GaussianNB	0.28	0.28	0.27	0.27	0.29	0.29	0.64	0.49	0.44	0.44	0.31	0.21	0.19	0.19
Ridge	0.78	0.78	0.87	0.87	0.29	0.29	0.63	0.65	0.57	0.57	0.30	0.30	0.25	0.25
SVC	0.79	0.79	0.88	0.88	0.30	0.30	0.57	0.60	0.58	0.58	0.25	0.25	0.27	0.27
Linear SVC	0.78	0.78	0.87	0.87	0.29	0.29	0.63	0.65	0.57	0.57	0.29	0.30	0.25	0.25

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 19. Relationships between mutated genes (and their combinations) and palpitations (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TTN	0.173773
PTPN11	0.186814
HCN4	0.186814
ND	-0.161280

Between mutated genes (and their combinations) and palpitations negligible relationships were found.

Table 20. Predicting fatigue from mutated causative genes [Support: Yes (n = 28); No (n = 40)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.65	0.65	0.75	0.76	0.40	0.33	0.68	0.66	0.59	0.58	0.60	0.54	0.48	0.48
RF	0.65	0.65	0.76	0.76	0.37	0.33	0.68	0.66	0.59	0.58	0.60	0.54	0.48	0.48
Log Reg	0.53	0.66	0.60	0.77	0.43	0.34	0.68	0.66	0.51	0.59	0.60	0.54	0.42	0.51
GaussianNB	0.62	0.62	0.70	0.70	0.48	0.48	0.68	0.68	0.59	0.59	0.60	0.60	0.47	0.47
Ridge	0.65	0.65	0.76	0.76	0.33	0.33	0.68	0.66	0.58	0.58	0.60	0.54	0.48	0.48
SVC	0.63	0.63	0.75	0.75	0.32	0.32	0.65	0.65	0.57	0.57	0.51	0.51	0.47	0.47
Linear SVC	0.65	0.65	0.76	0.76	0.33	0.33	0.68	0.66	0.58	0.58	0.60	0.54	0.48	0.48

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 21. Relationships between mutated causative genes and fatigue (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.080645
TNNT2	0.162619
MYBPC3	-0.070188
TNNI3	0.095797
TPM1	0.114920

Between mutated causative genes and fatigue negligible relationships were found.

Table 22. Predicting fatigue from mutated genes [Support: Yes (n = 28); No (n = 40)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.71	0.71	0.77	0.78	0.58	0.55	0.75	0.74	0.67	0.66	0.70	0.70	0.56	0.56
RF	0.71	0.71	0.78	0.78	0.57	0.57	0.70	0.74	0.67	0.67	0.67	0.70	0.56	0.56
Log Reg	0.75	0.71	0.80	0.78	0.68	0.55	0.74	0.74	0.73	0.66	0.70	0.70	0.61	0.56
GaussianNB	0.71	0.71	0.80	0.80	0.47	0.47	0.77	0.77	0.65	0.65	0.75	0.75	0.57	0.57
Ridge	0.71	0.71	0.78	0.78	0.57	0.57	0.75	0.75	0.67	0.67	0.70	0.70	0.56	0.56
SVC	0.71	0.71	0.78	0.78	0.57	0.57	0.76	0.70	0.67	0.67	0.67	0.56	0.56	0.56
Linear SVC	0.71	0.71	0.78	0.78	0.55	0.55	0.75	0.75	0.66	0.66	0.70	0.70	0.56	0.56

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 23. Relationships between mutated genes (and their combinations) and fatigue (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TNNT2	0.162619
TRPM4	0.135189
TTN	0.324315
NEXN	0.114920
ABCC9	0.114920
FBN1	0.114920
PTPN11	0.114920
LAMA4	0.114920
TPM1	0.114920
SDHA	0.114920
HCN4	0.114920
PKP2	0.182939
ND	-0.202304
TNNT2_TRPM4	0.114920
TNNT2_LAMA4	0.114920
TNNT2_PKP2	0.114920
MYBPC3_FBN1	0.114920
MYBPC3_PKP2	0.141064
TRPM4_LAMA4	0.114920
TTN_ABCC9	0.114920
TTN_DSP	0.114920
TTN_LAMA4	0.114920
TTN_PKP2	0.114920
NEXN_DSC2	0.114920
ABCC9_DSP	0.114920
TPM1_SDHA	0.114920

Between mutated genes (and their combinations) and fatigue negligible relationships were mostly found, with exception of mutation in *TTN* wherein relationship is moderate positive.

Table 24. Predicting dyspnea from mutated causative genes [Support: Yes (n = 18); No (n = 50)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.65	0.69	0.73	0.81	0.50	0.16	0.66	0.46	0.65	0.51	0.37	0.28	0.35	0.27
RF	0.69	0.69	0.81	0.81	0.22	0.16	0.66	0.46	0.52	0.51	0.37	0.28	0.28	0.27
Log Reg	0.62	0.74	0.71	0.85	0.43	0.00	0.53	0.46	0.60	0.50	0.32	0.28	0.32	0.26
GaussianNB	0.41	0.41	0.43	0.43	0.39	0.39	0.53	0.46	0.51	0.51	0.32	0.28	0.27	0.27
Ridge	0.69	0.69	0.81	0.81	0.16	0.16	0.53	0.46	0.51	0.51	0.32	0.28	0.27	0.27
SVC	0.69	0.69	0.81	0.81	0.16	0.16	0.36	0.47	0.51	0.51	0.22	0.30	0.27	0.27
Linear SVC	0.69	0.69	0.81	0.81	0.16	0.16	0.53	0.46	0.51	0.51	0.32	0.28	0.27	0.27

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 25. Relationships between mutated causative genes and dyspnea (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.132383
TNNT2	0.055589
TNNI3	-0.069304
MYL3	-0.056459
TPM1	0.051196

Between mutated causative genes and dyspnea negligible relationships were found.

Table 26. Predicting dyspnea from mutated genes [Support: Yes (n = 18); No (n = 50)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.57	0.66	0.67	0.79	0.41	0.15	0.68	0.68	0.57	0.49	0.38	0.38	0.30	0.26
RF	0.66	0.68	0.79	0.80	0.15	0.15	0.67	0.68	0.49	0.50	0.39	0.39	0.26	0.26
Log Reg	0.56	0.71	0.64	0.83	0.42	0.00	0.71	0.69	0.58	0.48	0.39	0.39	0.30	0.26
GaussianNB	0.28	0.28	0.08	0.08	0.41	0.41	0.64	0.63	0.49	0.49	0.35	0.35	0.26	0.26
Ridge	0.68	0.68	0.80	0.80	0.15	0.15	0.71	0.66	0.50	0.50	0.40	0.37	0.26	0.26
SVC	0.69	0.69	0.81	0.81	0.16	0.16	0.44	0.67	0.51	0.51	0.27	0.40	0.27	0.27
Linear SVC	0.68	0.68	0.80	0.68	0.15	0.15	0.69	0.66	0.50	0.50	0.38	0.37	0.26	0.26

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 27. Relationships between mutated genes (and their combinations) and dyspnea (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	0.132383
TTN	0.205244
PTPN11	0.158851
MYH6	0.112074
RBMD20	0.112074
PKP2	0.115773
ND	-0.156712
MYBPC3_DSG2	0.112074
MYBPC3_PKP2	0.194990
DSG2_FBN1	0.112074
DSG2_PKP2	0.112074
TTN_ELN	0.112074
TTN_MYH6	0.112074
TTN_RBMD20	0.112074
FBN1_PKP2	0.112074
ELN_MYH6	0.112074
ELN_RBMD20	0.112074
MYH6_RBMD20	0.112074

Between mutated genes (and their combinations) and dyspnea negligible relationships were mostly found. Most prominent relationship found was between mutation in *TTN* and dyspnea wherein relationship is weak positive.

Table 28. Predicting chest pain from mutated causative genes [Support: Yes (n = 15); No (n = 53)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.49	0.81	0.52	0.89	0.44	0.24	0.72	0.72	0.65	0.57	0.42	0.42	0.29	0.32
RF	0.81	0.81	0.89	0.89	0.24	0.24	0.72	0.72	0.57	0.57	0.42	0.42	0.32	0.32
Log Reg	0.49	0.78	0.52	0.88	0.44	0.00	0.71	0.71	0.65	0.50	0.32	0.32	0.29	0.22
GaussianNB	0.49	0.49	0.51	0.51	0.46	0.46	0.72	0.72	0.67	0.67	0.42	0.42	0.30	0.30
Ridge	0.81	0.81	0.89	0.89	0.24	0.24	0.72	0.71	0.57	0.57	0.42	0.32	0.32	0.32
SVC	0.81	0.81	0.89	0.89	0.24	0.24	0.72	0.72	0.57	0.57	0.42	0.40	0.32	0.32
Linear SVC	0.81	0.81	0.89	0.89	0.24	0.24	0.72	0.72	0.57	0.57	0.42	0.42	0.32	0.32

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 29. Relationships between mutated causative genes and chest pain (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.137235
MYBPC3	-0.190476
TNNI3	0.217479
MYL3	0.063276
TPM1	-0.050621

Between mutated causative genes and chest pain negligible relationships were mostly found. Most prominent relationship found was between mutation in *TNNI3* and chest pain wherein relationship is weak positive.

Table 30. Predicting chest pain from mutated genes [Support: Yes (n = 15); No (n = 53)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.50	0.81	0.56	0.89	0.41	0.24	0.68	0.71	0.61	0.57	0.40	0.43	0.27	0.32
RF	0.81	0.81	0.89	0.89	0.24	0.24	0.68	0.71	0.57	0.57	0.40	0.43	0.32	0.32
Logistic Regression	0.54	0.78	0.62	0.88	0.44	0.00	0.69	0.69	0.64	0.50	0.34	0.32	0.28	0.22
GaussianNB	0.50	0.50	0.54	0.54	0.45	0.45	0.70	0.71	0.66	0.66	0.34	0.41	0.29	0.29
Ridge	0.81	0.81	0.89	0.89	0.24	0.24	0.71	0.72	0.57	0.57	0.41	0.44	0.32	0.32
SVC	0.78	0.78	0.88	0.88	0.00	0.00	0.66	0.47	0.50	0.50	0.40	0.26	0.22	0.22
Linear SVC	0.81	0.81	0.89	0.89	0.24	0.24	0.71	0.72	0.57	0.57	0.42	0.44	0.32	0.32

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 31. Relationships between mutated genes (and their combinations) and chest pain (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	0.137235
MYBPC3	-0.190476
TNNI3	0.217479
PTPN11	0.177173
PDLIM3	0.125000
NKX2-5	0.125000
TNNI3_DSG2	0.125000
TNNI3_PDLIM3	0.125000
TNNI3_NKX2_5	0.125000
DSG2_PDLIM3	0.125000
DSG2_NKX2_5	0.125000
PDLIM3_NKX2_5	0.125000

Between mutated genes (and their combinations) and chest pain negligible relationships were mostly found. Most prominent relationship found was between mutation in *TNNI3* and chest pain wherein relationship is weak positive.

Table 32. Predicting syncope from mutated causative genes [Support: Yes (n = 6); No (n = 62)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.38	0.91	0.53	0.95	0.09	0.00	0..54	0.56	0.36	0.50	0.09	0.10	0.08	0.09
RF	0.91	0.85	0.95	0.92	0.00	0.00	0.55	0.56	0.50	0.47	0.11	0.10	0.09	0.09
Log Reg	0.47	0.91	0.63	0.95	0.05	0.00	0.53	0.55	0.33	0.50	0.09	0.10	0.08	0.09
GaussianNB	0.18	0.18	0.22	0.22	0.12	0.12	0.55	0.56	0.40	0.40	0.11	0.10	0.08	0.08
Ridge	0.91	0.91	0.95	0.95	0.00	0.00	0.53	0.55	0.50	0.50	0.09	0.10	0.09	0.09
SVC	0.91	0.91	0.95	0.95	0.00	0.00	0.44	0.57	0.50	0.50	0.08	0.10	0.09	0.09
Linear SVC	0.91	0.91	0.95	0.95	0.00	0.00	0.53	0.56	0.50	0.50	0.09	0.10	0.09	0.09

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

No relationships between mutated causative genes and syncope with phi coefficients having values > 0.05 or < -0.05 were found — no relationships between mutated causative genes and syncope were found.

Table 33. Predicting syncope from mutated genes [Support: Yes (n = 15); No (n = 53)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.41	0.91	0.56	0.95	0.13	0.00	0.68	0.47	0.45	0.50	0.19	0.09	0.08	0.09
RF	0.90	0.91	0.95	0.95	0.00	0.00	0.68	0.46	0.49	0.50	0.19	0.09	0.09	0.09
Log Reg	0.49	0.91	0.64	0.95	0.10	0.00	0.57	0.50	0.42	0.50	0.10	0.09	0.08	0.09
GaussianNB	0.22	0.22	0.31	0.31	0.10	0.10	0.56	0.45	0.35	0.35	0.10	0.08	0.07	0.07
Ridge	0.87	0.87	0.93	0.93	0.00	0.00	0.60	0.51	0.48	0.48	0.11	0.10	0.09	0.09
SVC	0.91	0.91	0.95	0.95	0.00	0.00	0.54	0.50	0.50	0.50	0.17	0.09	0.09	0.09
Linear SVC	0.87	0.87	0.93	0.93	0.00	0.00	0.63	0.47	0.48	0.48	0.14	0.09	0.09	0.09

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 34. Relationships between mutated genes (and their combinations) and syncope (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
DSG2	0.132391
FBN1	0.132391
MYBPC3_DSG2	0.208248
MYBPC3_FBN1	0.132391
DSG2_FBN1	0.208248
DSG2_PKP2	0.208248
FBN1_PKP2	0.208248

Between mutated genes (and their combinations) and syncope negligible relationships were mostly found. Some weak positive relationships were found between combinations of genes and syncope.

Table 35. Predicting pretibial edema from mutated causative genes [Support: Yes (n = 6); No (n = 62)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.44	0.94	0.61	0.97	0.00	0.00	0.52	0.52	0.23	0.50	0.06	0.06	0.06	0.06
RF	0.94	0.94	0.97	0.97	0.00	0.00	0.52	0.53	0.50	0.50	0.06	0.06	0.06	0.06
Logistic Regression	0.51	0.94	0.68	0.97	0.00	0.00	0.53	0.55	0.27	0.50	0.06	0.07	0.06	0.06
GaussianNB	0.24	0.24	0.35	0.35	0.07	0.07	0.52	0.52	0.36	0.36	0.06	0.06	0.05	0.05
Ridge	0.94	0.94	0.97	0.97	0.00	0.00	0.53	0.55	0.50	0.50	0.06	0.07	0.06	0.06
SVC	0.94	0.94	0.97	0.97	0.00	0.00	0.52	0.50	0.50	0.50	0.06	0.06	0.06	0.06
Linear SVC	0.94	0.94	0.97	0.97	0.00	0.00	0.54	0.59	0.50	0.50	0.06	0.07	0.06	0.06

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 36. Relationships between mutated causative genes and pretibial edema (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.086918
MYBPC3	-0.078678

Between mutated causative genes and pretibial edema negligible relationships were found.

Table 37. Predicting pretibial edema from mutated genes [Support: Yes (n = 4); No (n = 64)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.93	0.93	0.96	0.96	0.29	0.00	0.58	0.46	0.61	0.49	0.11	0.05	0.13	0.06
RF	0.93	0.93	0.96	0.96	0.00	0.00	0.81	0.67	0.49	0.49	0.20	0.11	0.06	0.06
Log Reg	0.94	0.94	0.97	0.97	0.60	0.00	0.82	0.34	0.85	0.50	0.26	0.05	0.39	0.06
GaussianNB	0.32	0.32	0.48	0.48	0.04	0.04	0.50	0.49	0.29	0.29	0.06	0.06	0.05	0.05
Ridge	0.94	0.94	0.97	0.97	0.00	0.00	0.91	0.42	0.50	0.50	0.32	0.08	0.06	0.06
SVC	0.94	0.94	0.97	0.97	0.00	0.00	0.29	0.53	0.50	0.50	0.05	0.07	0.06	0.06
Linear SVC	0.93	0.93	0.96	0.96	0.00	0.00	0.81	0.33	0.49	0.49	0.20	0.05	0.06	0.06

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 38. Relationships between mutated genes (and their combinations) and pretibial edema (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
DSG2	0.171630
TRPM4	0.149258
TTN	0.299958
ABCC9	0.171630
DSP	0.104582
FBN1	0.171630
LAMA4	0.171630
TNNT2_TRPM4	0.171630
TNNT2_LAMA4	0.171630
MYBPC3_DSG2	0.259390
MYBPC3_FBN1	0.171630
MYBPC3_PKP2	0.130469
DSG2_FBN1	0.259390
DSG2_PKP2	0.259390
TRPM4_LAMA4	0.171630
TTN_ABCC9	0.171630
TTN_DSP	0.171630
TTN_LAMA4	0.171630
ABCC9_DSP	0.171630
FBN1_PKP2	0.259390

Between mutated genes (and their combinations) and pretibial edema negligible relationships were mostly found. Some weak positive relationships were found between combinations of genes and pretibial edema. Most prominent relationship found was between mutation in *TTN* and pretibial edema wherein relationship is low-moderate positive.

Table 39. Predicting SAM from mutated causative genes [Support: Yes (n = 11); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.50	0.83	0.61	0.91	0.30	0.00	0.50	0.55	0.55	0.50	0.19	0.18	0.18	0.17
RF	0.83	0.83	0.91	0.91	0.00	0.00	0.50	0.60	0.50	0.50	0.19	0.20	0.17	0.17
Log Reg	0.65	0.83	0.77	0.91	0.26	0.00	0.50	0.60	0.54	0.50	0.19	0.20	0.18	0.17
GaussianNB	0.33	0.33	0.33	0.33	0.33	0.33	0.60	0.60	0.60	0.60	0.20	0.20	0.20	0.20
Ridge	0.83	0.83	0.91	0.91	0.00	0.00	0.50	0.60	0.50	0.50	0.19	0.20	0.17	0.17
SVC	0.83	0.83	0.91	0.91	0.00	0.00	0.49	0.62	0.50	0.50	0.17	0.22	0.17	0.17
Linear SVC	0.83	0.83	0.91	0.91	0.00	0.00	0.50	0.60	0.50	0.50	0.19	0.20	0.17	0.17

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 40. Relationships between mutated causative genes and SAM (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.064275
MYBPC3	0.098106
TNNI3	-0.051922

Between mutated causative genes and SAM negligible relationships were found.

Table 41. Predicting SAM from mutated genes [Support: Yes (n = 11); No (n = 55)]

	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.52	0.82	0.64	0.90	0.27	0.00	0.53	0.65	0.53	0.49	0.19	0.26	0.17	0.17
RF	0.80	0.82	0.89	0.90	0.00	0.00	0.54	0.64	0.48	0.49	0.20	0.23	0.17	0.17
Log Reg	0.48	0.83	0.60	0.91	0.26	0.00	0.53	0.64	0.51	0.50	0.20	0.23	0.17	0.17
GaussianNB	0.38	0.38	0.42	0.42	0.33	0.33	0.63	0.63	0.59	0.59	0.23	0.23	0.20	0.20
Ridge	0.83	0.83	0.91	0.91	0.00	0.00	0.57	0.64	0.50	0.50	0.21	0.23	0.17	0.17
SVC	0.83	0.83	0.91	0.91	0.00	0.00	0.64	0.59	0.50	0.50	0.24	0.27	0.17	0.17
Linear SVC	0.83	0.83	0.91	0.91	0.00	0.00	0.57	0.64	0.50	0.50	0.21	0.23	0.17	0.17

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 42. Relationships between mutated genes (and their combinations) and SAM (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
HCN4	0.219937

Between mutated genes (and their combinations) and SAM mostly negligible relationships were found. Most prominent relationship found was between mutation in *HCN4* and SAM wherein relationship is weak positive.

Table 43. Predicting mitral leaflet abnormalities from mutated causative genes [Support: Yes (n = 6); No (n = 57)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.49	0.90	0.62	0.95	0.24	0.00	0.53	0.53	0.64	0.50	0.11	0.11	0.13	0.10
RF	0.90	0.90	0.95	0.95	0.00	0.00	0.53	0.53	0.50	0.50	0.11	0.11	0.10	0.10
Log Reg	0.49	0.90	0.62	0.95	0.24	0.00	0.53	0.53	0.64	0.50	0.11	0.11	0.13	0.10
GaussianNB	0.32	0.32	0.41	0.41	0.19	0.19	0.53	0.53	0.55	0.55	0.11	0.11	0.10	0.10
Ridge	0.90	0.90	0.95	0.95	0.00	0.00	0.53	0.53	0.50	0.50	0.11	0.11	0.10	0.10
SVC	0.90	0.90	0.95	0.95	0.00	0.00	0.50	0.53	0.50	0.50	0.10	0.11	0.10	0.10
Linear SVC	0.90	0.90	0.95	0.95	0.00	0.00	0.53	0.53	0.50	0.50	0.11	0.11	0.10	0.10

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 44. Relationships between mutated causative genes and mitral leaflet abnormalities (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.106952
MYBPC3	-0.088080
TNNI3	0.054644

Between mutated causative genes and mitral leaflet abnormalities negligible relationships were found.

Table 45. Predicting mitral leaflet abnormalities from mutated genes [Support: Yes (n = 6); No (n = 57)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.63	0.87	0.77	0.93	0.15	0.00	0.51	0.49	0.50	0.48	0.14	0.11	0.10	0.10
RF	0.87	0.87	0.93	0.93	0.00	0.00	0.53	0.52	0.48	0.48	0.14	0.11	0.10	0.10
Log Reg	0.68	0.90	0.80	0.95	0.17	0.00	0.54	0.61	0.53	0.50	0.13	0.12	0.10	0.10
GaussianNB	0.30	0.30	0.41	0.41	0.15	0.15	0.46	0.46	0.46	0.46	0.09	0.09	0.09	0.09
Ridge	0.87	0.87	0.93	0.93	0.00	0.00	0.53	0.64	0.48	0.48	0.13	0.14	0.10	0.10
SVC	0.87	0.87	0.93	0.93	0.00	0.00	0.52	0.48	0.48	0.48	0.10	0.09	0.10	0.10
Linear SVC	0.87	0.87	0.93	0.93	0.00	0.00	0.52	0.48	0.48	0.48	0.14	0.09	0.10	0.10

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 46. Relationships between mutated genes (and their combinations) and mitral leaflet abnormalities (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	-0.106952
TRPM4	0.104094
RYR1	0.134862
DSC2	0.191661
CACNA1C	0.302026
ELN	0.370810
MYH6	0.213045
RBMD20	0.213045
TRPM4_DSC2	0.191661
TRPM4_CACNA1C	0.302026
TRPM4_ELN	0.302026
TTN_ELN	0.213045
TTN_MYH6	0.213045
TTN_RBMD20	0.213045
DSC2_CACNA1C	0.302026
DSC2_ELN	0.302026
CACNA1C_ELN	0.302026
ELN_MYH6	0.213045
ELN_RBMD20	0.213045
MYH6_RBMD20	0.213045

Between mutated genes (and their combinations) and mitral leaflet abnormalities negligible and weak positive relationships were mostly found. Most prominent single-gene relationships found were between mutations in *CACNA1C* and mitral leaflet abnormalities, and *ELN* and mitral leaflet abnormalities wherein relationships are moderate positive.

Table 47. Predicting calcification of mitral annulus from mutated genes [Support: Yes (n = 8); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.62	0.86	0.74	0.92	0.25	0.00	0.45	0.40	0.57	0.49	0.12	0.12	0.15	0.13
RF	0.86	0.86	0.92	0.92	0.00	0.00	0.42	0.39	0.49	0.49	0.12	0.11	0.13	0.13
Log Reg	0.65	0.87	0.77	0.93	0.27	0.00	0.34	0.37	0.59	0.50	0.10	0.11	0.15	0.13
GaussianNB	0.22	0.22	0.25	0.25	0.20	0.20	0.54	0.50	0.45	0.45	0.14	0.13	0.12	0.12
Ridge	0.87	0.87	0.93	0.93	0.00	0.00	0.36	0.35	0.50	0.50	0.11	0.11	0.13	0.13
SVC	0.87	0.87	0.93	0.93	0.00	0.00	0.53	0.39	0.50	0.50	0.14	0.11	0.13	0.13
Linear SVC	0.86	0.86	0.92	0.92	0.00	0.00	0.38	0.39	0.49	0.49	0.11	0.11	0.13	0.13

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 48. Relationships between mutated genes (and their combinations) and calcification of mitral annulus (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
CSRP3	0.108388
PTPN11	0.255031
HCN4	0.255031
MYBPC3_DSP	0.108388
MYBPC3_CSRP3	0.108388
DSP_CSRP3	0.108388

Between mutated genes (and their combinations) and calcification of mitral annulus negligible relationships were mostly found. Two weak positive relationships found were between mutation in *PTPN11* and calcification of mitral annulus, and between mutation in *HCN4* and calcification of mitral annulus.

Table 49. Predicting sinus rhythm from mutated causative genes [Support: Yes (n = 50); No (n = 10)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.40	0.80	0.28	0.00	0.49	0.89	0.59	0.56	0.52	0.48	0.86	0.85	0.84	0.83
RF	0.83	0.83	0.00	0.00	0.91	0.91	0.59	0.56	0.50	0.50	0.86	0.85	0.83	0.83
Log Reg	0.40	0.83	0.28	0.00	0.49	0.91	0.61	0.59	0.52	0.50	0.88	0.86	0.84	0.83
GaussianNB	0.40	0.40	0.14	0.14	0.54	0.54	0.56	0.56	0.36	0.36	0.85	0.84	0.80	0.80
Ridge	0.83	0.83	0.00	0.00	0.91	0.91	0.59	0.56	0.50	0.50	0.86	0.85	0.83	0.83
SVC	0.83	0.83	0.00	0.00	0.91	0.91	0.57	0.55	0.50	0.50	0.85	0.84	0.83	0.83
Linear SVC	0.83	0.83	0.00	0.00	0.91	0.91	0.59	0.56	0.50	0.50	0.86	0.85	0.83	0.83

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 50. Relationships between mutated causative genes and sinus rhythm (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.062113
MYBPC3	0.141487
TPM1	-0.230073

Between mutated causative genes and sinus rhythm negligible relationships were mostly found. Most prominent relationship found was between mutation in *TPM1* and sinus rhythm wherein relationship is weak negative.

Table 51. Predicting sinus rhythm from mutated genes [Support: Yes (n = 50); No (n = 10)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.52	0.80	0.33	0.00	0.62	0.89	0.53	0.53	0.59	0.48	0.84	0.83	0.86	0.83
RF	0.83	0.83	0.00	0.00	0.91	0.91	0.54	0.53	0.50	0.50	0.86	0.84	0.83	0.83
Log Reg	0.52	0.83	0.33	0.00	0.62	0.91	0.54	0.52	0.59	0.50	0.86	0.83	0.86	0.83
GaussianNB	0.32	0.32	0.20	0.20	0.41	0.41	0.52	0.57	0.39	0.39	0.83	0.84	0.81	0.81
Ridge	0.83	0.83	0.00	0.00	0.91	0.91	0.50	0.52	0.50	0.50	0.84	0.83	0.83	0.83
SVC	0.83	0.83	0.00	0.00	0.91	0.91	0.44	0.53	0.50	0.50	0.82	0.84	0.83	0.83
Linear SVC	0.82	0.82	0.00	0.00	0.90	0.90	0.50	0.52	0.49	0.49	0.84	0.83	0.83	0.83

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 52. Relationships between mutated genes (and their combinations) and sinus rhythm (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYBPC3	0.141487
DSG2	-0.162273
ABCC9	-0.230073
DSP	-0.132000
PTPN11	-0.230073
TPM1	-0.230073
SDHA	-0.230073
MYBPC3_DSG2	-0.162273
DSG2_FBN1	-0.162273
DSG2_PKP2	-0.162273
TTN_ABCC9	-0.230073
TTN_DSP	-0.230073
ABCC9_DSP	-0.230073
FBN1_PKP2	-0.162273
TPM1_SDHA	-0.230073

Between mutated genes (and their combinations) and sinus rhythm negligible and weak negative relationships were found.

Table 53. Predicting left anterior hemiblock from mutated causative genes [Support: Yes (n = 5); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.50	0.92	0.63	0.96	0.21	0.00	0.47	0.51	0.64	0.50	0.09	0.09	0.11	0.08
RF	0.92	0.92	0.96	0.96	0.00	0.00	0.47	0.51	0.50	0.50	0.09	0.09	0.08	0.08
Log Reg	0.50	0.92	0.63	0.96	0.21	0.00	0.47	0.51	0.64	0.50	0.09	0.09	0.11	0.08
GaussianNB	0.27	0.27	0.35	0.35	0.15	0.15	0.46	0.51	0.51	0.51	0.08	0.09	0.08	0.08
Ridge	0.92	0.92	0.96	0.96	0.00	0.00	0.47	0.51	0.50	0.50	0.09	0.09	0.08	0.08
SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.47	0.65	0.50	0.50	0.08	0.12	0.08	0.08
Linear SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.47	0.51	0.50	0.50	0.09	0.09	0.08	0.08

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

No phi coefficients having values > 0.05 or < -0.05 were found — no relationships between mutated causative genes and left anterior hemiblock were found.

Table 54. Predicting left anterior hemiblock from mutated genes [Support: Yes (n = 5); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.45	0.92	0.58	0.96	0.20	0.00	0.53	0.54	0.61	0.50	0.10	0.10	0.11	0.08
RF	0.92	0.92	0.96	0.96	0.00	0.00	0.51	0.54	0.50	0.50	0.10	0.10	0.08	0.08
Log Reg	0.60	0.92	0.73	0.96	0.25	0.00	0.50	0.49	0.69	0.50	0.09	0.09	0.14	0.08
GaussianNB	0.40	0.40	0.53	0.53	0.18	0.18	0.47	0.58	0.58	0.58	0.09	0.11	0.10	0.10
Ridge	0.92	0.92	0.96	0.96	0.00	0.00	0.49	0.49	0.50	0.50	0.09	0.09	0.08	0.08
SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.73	0.37	0.50	0.50	0.14	0.08	0.08	0.08
Linear SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.49	0.51	0.50	0.50	0.09	0.10	0.08	0.08

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 55. Relationships between mutated genes (and their combinations) and left anterior hemiblock (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TTN_PKP2	0.143625

Negligible relationships between mutated genes (and their combinations) and left anterior hemiblock were found.

Table 56. Predicting RBBB from mutated causative genes [Support: Yes (n = 5); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.73	0.92	0.84	0.96	0.20	0.00	0.65	0.38	0.58	0.50	0.19	0.07	0.10	0.08
RF	0.92	0.88	0.96	0.94	0.00	0.00	0.65	0.38	0.50	0.48	0.19	0.07	0.08	0.08
Log Reg	0.88	0.92	0.94	0.96	0.36	0.00	0.65	0.38	0.66	0.50	0.19	0.07	0.18	0.08
GaussianNB	0.20	0.20	0.25	0.25	0.14	0.14	0.62	0.38	0.47	0.47	0.18	0.07	0.08	0.08
Ridge	0.92	0.92	0.96	0.96	0.00	0.00	0.65	0.38	0.50	0.50	0.19	0.07	0.08	0.08
SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.55	0.41	0.50	0.50	0.10	0.07	0.08	0.08
Linear SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.65	0.38	0.50	0.50	0.19	0.07	0.08	0.08

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 57. Relationships between mutated causative genes and RBBB (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.118811
TNNT2	0.213873

Negligible relationships between mutated causative genes and RBBB were mostly found. Most prominent relationship found was between mutation in *TNNT2* and RBBB wherein relationship is weak positive.

Table 58. Predicting RBBB from mutated genes [Support: Yes (n = 5); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.55	0.95	0.70	0.97	0.13	0.57	0.35	0.35	0.48	0.70	0.07	0.07	0.08	0.45
RF	0.95	0.95	0.97	0.97	0.57	0.57	0.68	0.60	0.70	0.70	0.46	0.12	0.45	0.45
Log Reg	0.68	0.92	0.80	0.96	0.17	0.00	0.45	0.34	0.55	0.50	0.09	0.07	0.09	0.08
GaussianNB	0.20	0.20	0.25	0.25	0.14	0.14	0.22	0.22	0.47	0.47	0.06	0.06	0.08	0.08
Ridge	0.92	0.92	0.96	0.96	0.00	0.00	0.50	0.18	0.50	0.50	0.12	0.06	0.08	0.08
SVC	0.95	0.95	0.97	0.97	0.57	0.57	0.81	0.54	0.70	0.70	0.49	0.12	0.45	0.45
Linear SVC	0.93	0.93	0.96	0.96	0.33	0.33	0.45	0.18	0.60	0.60	0.09	0.06	0.27	0.27

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 59. Relationships between mutated genes (and their combinations) and RBBB (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	0.118811
TNNT2	0.213873
TRPM4	0.113266
TTN	0.186557
LAMA4	0.319365
HCN4	0.319365
PKP2	0.172852
ND	-0.127865
TNNT2_TRPM4	0.319365
TNNT2_LAMA4	0.319365
TNNT2_PKP2	0.319365
TRPM4_LAMA4	0.319365
TTN_LAMA4	0.319365
TTN_PKP2	0.319365

Between mutated genes (and their combinations) and RBBB negligible and weak relationships were mostly found. Most prominent single-gene relationships found were between mutations in *LAMA4* and RBBB, and *HCN4* and RBBB wherein relationship is moderate positive.

Table 60. Predicting LBBB from mutated causative genes [Support: Yes (n = 4); No (n = 56)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.52	0.93	0.67	0.97	0.12	0.00	0.67	0.67	0.51	0.50	0.10	0.10	0.07	0.07
RF	0.93	0.93	0.97	0.97	0.00	0.00	0.67	0.67	0.50	0.50	0.10	0.10	0.07	0.07
Log Reg	0.73	0.93	0.84	0.97	0.20	0.00	0.67	0.67	0.62	0.50	0.10	0.10	0.10	0.07
GaussianNB	0.22	0.22	0.28	0.28	0.15	0.15	0.67	0.67	0.58	0.58	0.10	0.10	0.08	0.08
Ridge	0.93	0.93	0.97	0.97	0.00	0.00	0.67	0.67	0.50	0.50	0.10	0.10	0.07	0.07
SVC	0.93	0.93	0.97	0.97	0.00	0.00	0.54	0.67	0.50	0.50	0.07	0.10	0.07	0.07
Linear SVC	0.93	0.93	0.97	0.97	0.00	0.00	0.67	0.67	0.50	0.50	0.10	0.10	0.07	0.07

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 61. Relationships between mutated causative genes and LBBB (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.091816
TNNT2	-0.067697
MYBPC3	0.161491

Negligible relationships between mutated causative genes and LBBB were found.

Table 62. Predicting LBBB from mutated genes [Support: Yes (n = 4); No (n = 56)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.48	0.93	0.65	0.97	0.00	0.00	0.77	0.70	0.26	0.50	0.34	0.32	0.07	0.07
RF	0.93	0.93	0.97	0.97	0.00	0.00	0.77	0.68	0.50	0.50	0.34	0.11	0.07	0.07
Log Reg	0.43	0.93	0.60	0.97	0.00	0.00	0.74	0.45	0.23	0.50	0.33	0.07	0.07	0.07
GaussianNB	0.25	0.25	0.35	0.35	0.12	0.12	0.62	0.54	0.48	0.48	0.09	0.07	0.06	0.06
Ridge	0.93	0.93	0.97	0.97	0.00	0.00	0.75	0.45	0.50	0.50	0.34	0.07	0.07	0.07
SVC	0.93	0.93	0.97	0.97	0.00	0.00	0.76	0.54	0.50	0.50	0.34	0.08	0.07	0.07
Linear SVC	0.93	0.93	0.97	0.97	0.00	0.00	0.75	0.51	0.50	0.50	0.34	0.07	0.07	0.07

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 63. Relationships between mutated genes (and their combinations) and LBBB (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYBPC3	0.161491

Negligible relationships between mutated genes (and their combinations) and LBBB were found.

Table 64. Predicting AF from mutated genes [Support: Yes (n = 7); No (n = 53)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.62	0.88	0.74	0.94	0.26	0.00	0.41	0.43	0.60	0.50	0.11	0.12	0.15	0.12
RF	0.88	0.87	0.94	0.93	0.00	0.00	0.40	0.41	0.50	0.49	0.11	0.11	0.12	0.12
Log Reg	0.62	0.88	0.74	0.94	0.26	0.00	0.46	0.50	0.60	0.50	0.12	0.12	0.15	0.12
GaussianNB	0.37	0.37	0.46	0.46	0.24	0.24	0.44	0.43	0.58	0.58	0.12	0.12	0.14	0.14
Ridge	0.88	0.88	0.94	0.94	0.00	0.00	0.49	0.52	0.50	0.50	0.12	0.13	0.12	0.12
SVC	0.88	0.88	0.94	0.94	0.00	0.00	0.49	0.55	0.50	0.50	0.12	0.13	0.12	0.12
Linear SVC	0.87	0.87	0.93	0.93	0.00	0.00	0.46	0.48	0.49	0.49	0.12	0.12	0.12	0.12

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 65. Relationships between mutated genes (and their combinations) and AF (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYBPC3	-0.152631
DSG2	0.191774
ABCC9	0.271904
DSP	0.166448
FBN1	0.117091
PTPN11	0.271904
LAMA4	0.117091
TPM1	0.271904
SDHA	0.271904
HCN4	0.117091
TNNT2_TRPM4	0.117091
TNNT2_LAMA4	0.117091
MYBPC3_DSG2	0.191774
MYBPC3_FBN1	0.117091
DSG2_FBN1	0.191774
DSG2_PKP2	0.191774
TRPM4_LAMA4	0.117091
TTN_ABCC9	0.271904
TTN_DSP	0.271904
TTN_LAMA4	0.117091
ABCC9_DSP	0.271904
FBN1_PKP2	0.191774
TPM1_SDHA	0.271904

Negligible and weak relationships between mutated genes (and their combinations) and AF were found.

Table 66. Predicting significant Q wave from mutated causative genes [Support: Yes (n = 4); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.47	0.93	0.64	0.96	0.00	0.00	0.64	0.49	0.25	0.50	0.11	0.07	0.07	0.07
RF	0.93	0.93	0.96	0.96	0.00	0.00	0.63	0.49	0.50	0.50	0.11	0.07	0.07	0.07
Log Reg	0.47	0.93	0.64	0.96	0.00	0.00	0.64	0.49	0.25	0.50	0.11	0.07	0.07	0.07
GaussianNB	0.19	0.19	0.27	0.27	0.08	0.08	0.64	0.49	0.33	0.33	0.11	0.07	0.05	0.05
Ridge	0.93	0.93	0.96	0.96	0.00	0.00	0.64	0.49	0.50	0.50	0.11	0.07	0.07	0.07
SVC	0.93	0.93	0.96	0.96	0.00	0.00	0.44	0.49	0.50	0.50	0.06	0.07	0.07	0.07
Linear SVC	0.93	0.93	0.96	0.96	0.00	0.00	0.63	0.49	0.50	0.50	0.11	0.07	0.07	0.07

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 67. Relationships between mutated causative genes and significant Q wave (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.096618
TNNT2	0.094110

Negligible relationships between mutated causative genes and significant Q wave were found.

Table 68. Predicting significant Q wave from mutated genes [Support: Yes (n = 4); No (n = 55)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.49	0.93	0.66	0.96	0.00	0.00	0.77	0.72	0.26	0.50	0.34	0.33	0.07	0.07
RF	0.93	0.93	0.96	0.96	0.00	0.00	0.77	0.60	0.50	0.50	0.34	0.09	0.07	0.07
Log Reg	0.53	0.93	0.69	0.96	0.00	0.00	0.77	0.45	0.28	0.50	0.34	0.07	0.07	0.07
GaussianNB	0.22	0.22	0.32	0.32	0.08	0.08	0.59	0.54	0.35	0.35	0.08	0.08	0.06	0.06
Ridge	0.93	0.93	0.96	0.96	0.00	0.00	0.74	0.45	0.50	0.50	0.33	0.07	0.07	0.07
SVC	0.93	0.93	0.96	0.96	0.00	0.00	0.63	0.53	0.50	0.50	0.31	0.08	0.07	0.07
Linear SVC	0.93	0.93	0.96	0.96	0.00	0.00	0.74	0.45	0.50	0.50	0.33	0.07	0.07	0.07

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 69. Relationships between mutated genes (and their combinations) and significant Q wave (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TRPM4	0.142783
NOTCH1	0.366026
LAMA4	0.366026
TNNT2_TRPM4	0.366026
TNNT2_LAMA4	0.366026
MYBPC3_NOTCH1	0.366026
TRPM4_LAMA4	0.366026
TTN_LAMA4	0.366026

Between mutated genes and significant Q wave negligible relationships were mostly found. Single-gene exceptions are relationships found between mutation in *LAMA4* and significant Q wave, and mutation in *NOTCH1* and significant Q wave wherein relationship is moderate positive.

Table 70. Predicting ST segment abnormalities from mutated causative genes [Support: Yes (n = 30); No (n = 29)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.58	0.58	0.55	0.55	0.60	0.60	0.60	0.61	0.58	0.58	0.58	0.58	0.55	0.55
RF	0.54	0.58	0.49	0.55	0.58	0.60	0.60	0.61	0.54	0.58	0.58	0.58	0.53	0.55
Log Reg	0.54	0.54	0.49	0.49	0.58	0.58	0.60	0.60	0.54	0.54	0.58	0.58	0.53	0.53
GaussianNB	0.51	0.51	0.38	0.38	0.59	0.59	0.60	0.62	0.51	0.51	0.58	0.58	0.51	0.51
Ridge	0.54	0.54	0.49	0.49	0.58	0.58	0.60	0.60	0.54	0.54	0.58	0.58	0.53	0.53
SVC	0.58	0.58	0.55	0.55	0.60	0.60	0.60	0.60	0.58	0.58	0.58	0.58	0.55	0.55
Linear SVC	0.54	0.54	0.49	0.49	0.58	0.58	0.60	0.61	0.54	0.54	0.58	0.58	0.53	0.53

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 71. Relationships between mutated causative genes and ST segment abnormalities (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.058556
MYBPC3	-0.119448
TNNI3	0.069978
MYL3	0.099220

Negligible relationships between mutated causative genes and ST segment abnormalities were found.

Table 72. Predicting ST segment abnormalities from mutated genes [Support: Yes (n = 30); No (n = 29)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.53	0.53	0.50	0.50	0.55	0.55	0.63	0.63	0.52	0.52	0.60	0.60	0.52	0.52
RF	0.56	0.54	0.50	0.51	0.61	0.57	0.63	0.63	0.56	0.54	0.60	0.60	0.54	0.53
Log Reg	0.56	0.56	0.52	0.52	0.59	0.59	0.62	0.62	0.56	0.56	0.58	0.58	0.54	0.54
GaussianNB	0.58	0.58	0.24	0.24	0.71	0.71	0.55	0.55	0.57	0.57	0.54	0.54	0.55	0.55
Ridge	0.56	0.56	0.52	0.52	0.59	0.59	0.62	0.62	0.56	0.56	0.58	0.58	0.54	0.54
SVC	0.53	0.53	0.48	0.48	0.56	0.56	0.61	0.62	0.52	0.52	0.57	0.57	0.52	0.52
Linear SVC	0.56	0.56	0.52	0.52	0.59	0.59	0.62	0.62	0.56	0.56	0.58	0.58	0.54	0.54

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 73. Relationships between mutated genes (and their combinations) and ST segment abnormalities (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYBPC3	-0.119448
JPH2	-0.128247
TRPM4	-0.212208
RYR1	-0.104442
NEXN	-0.104442
DSC2	-0.148474
CTF1	-0.104442
CSRP3	-0.104442
CACNA1C	-0.104442
ELN	-0.128247
NOTCH1	-0.104442
PTPN11	-0.104442
LAMA4	-0.104442
PKP2	-0.166436
ND	0.195798
TNNT2_TRPM4	-0.104442
TNNT2_LAMA4	-0.104442
TNNT2_PKP2	-0.104442
MYBPC3_JPH2	-0.104442
MYBPC3_DSP	-0.104442
MYBPC3_CTF1	-0.104442
MYBPC3_CSRP3	-0.104442
MYBPC3_NOTCH1	-0.104442
MYBPC3_PKP2	-0.128247
JPH2_CTF1	-0.104442
TRPM4_DSC2	-0.148474
TRPM4_CACNA1C	-0.104442
TRPM4_ELN	-0.104442
TRPM4_LAMA4	-0.104442
TTN_LAMA4	-0.104442
TTN_PKP2	-0.104442
NEXN_DSC2	-0.104442
DSC2_CACNA1C	-0.104442
DSC2_ELN	-0.104442
DSP_CSRP3	-0.104442
CACNA1C_ELN	-0.104442

Negligible relationships between mutated genes (and their combinations) and ST segment abnormalities were mostly found. The only exception is relationship between mutation in *TRPM4* and ST segment abnormalities wherein relationship is weak negative.

Table 74. Predicting negative T wave from mutated causative genes [Support: Yes (n = 34); No (n = 24)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.63	0.59	0.56	0.50	0.68	0.66	0.62	0.51	0.62	0.58	0.65	0.59	0.64	0.62
RF	0.58	0.58	0.42	0.47	0.67	0.65	0.62	0.46	0.55	0.56	0.65	0.56	0.60	0.61
Log Reg	0.63	0.58	0.56	0.42	0.68	0.67	0.62	0.46	0.62	0.55	0.65	0.56	0.64	0.60
GaussianNB	0.61	0.61	0.41	0.41	0.71	0.71	0.62	0.51	0.57	0.57	0.65	0.59	0.61	0.61
Ridge	0.58	0.58	0.42	0.42	0.67	0.67	0.62	0.46	0.55	0.55	0.65	0.56	0.60	0.60
SVC	0.59	0.59	0.50	0.50	0.66	0.66	0.53	0.46	0.58	0.58	0.61	0.56	0.62	0.62
Linear SVC	0.58	0.58	0.42	0.42	0.67	0.67	0.62	0.46	0.55	0.55	0.65	0.56	0.60	0.60

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 75. Relationships between mutated causative genes and negative T wave (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
TNNT2	-0.081670
MYBPC3	-0.056908
TNNI3	0.061806
MYL3	0.087633

Negligible relationships between mutated causative genes and negative T wave were found.

Table 76. Predicting negative T wave from mutated genes [Support: Yes (n = 34); No (n = 25)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.68	0.63	0.63	0.56	0.72	0.68	0.51	0.54	0.67	0.62	0.66	0.67	0.68	0.64
RF	0.63	0.63	0.52	0.50	0.69	0.70	0.51	0.58	0.61	0.60	0.66	0.69	0.64	0.69
Log Reg	0.64	0.63	0.60	0.50	0.68	0.70	0.69	0.65	0.64	0.60	0.73	0.71	0.66	0.63
GaussianNB	0.64	0.64	0.51	0.51	0.72	0.72	0.70	0.61	0.62	0.62	0.73	0.66	0.64	0.64
Ridge	0.63	0.63	0.50	0.50	0.70	0.70	0.63	0.64	0.60	0.60	0.70	0.69	0.63	0.63
SVC	0.61	0.61	0.49	0.49	0.68	0.68	0.55	0.51	0.59	0.59	0.65	0.63	0.62	0.62
Linear SVC	0.61	0.61	0.49	0.49	0.68	0.68	0.57	0.62	0.59	0.59	0.68	0.69	0.62	0.62

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 77. Relationships between mutated genes (and their combinations) and negative T wave (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TRPM4	-0.135687
TTN	0.159062
RYR1	-0.118251
CACNA1C	-0.118251
NOTCH1	-0.118251
LAMA4	-0.118251
TNNT2_TRPM4	-0.118251
TNNT2_LAMA4	-0.118251
MYBPC3_NOTCH1	-0.118251
MYBPC3_PKP2	-0.145204
TRPM4_CACNA1C	-0.118251
TRPM4_ELN	-0.118251
TRPM4_LAMA4	-0.118251
TTN_LAMA4	-0.118251
DSC2_CACNA1C	-0.118251
DSC2_ELN	-0.118251
CACNA1C_ELN	-0.118251

Negligible relationships between mutated genes (and their combinations) and negative T wave were found.

Table 78. Predicting signs from mutated causal genes [Support: Yes (n = 28); No (n = 40)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.47	0.56	0.58	0.72	0.28	0.00	0.55	0.55	0.44	0.47	0.44	0.44	0.39	0.41
RF	0.53	0.53	0.69	0.69	0.06	0.00	0.55	0.55	0.46	0.45	0.44	0.44	0.40	0.41
Log Reg	0.47	0.56	0.58	0.72	0.28	0.00	0.55	0.55	0.44	0.47	0.44	0.44	0.39	0.41
GaussianNB	0.46	0.46	0.57	0.57	0.24	0.24	0.55	0.55	0.42	0.42	0.44	0.44	0.38	0.38
Ridge	0.56	0.56	0.72	0.72	0.00	0.00	0.55	0.55	0.47	0.47	0.44	0.44	0.41	0.41
SVC	0.56	0.56	0.72	0.72	0.00	0.00	0.48	0.55	0.47	0.47	0.41	0.44	0.41	0.41
Linear SVC	0.56	0.56	0.72	0.72	0.00	0.00	0.55	0.55	0.47	0.47	0.44	0.44	0.41	0.41

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 79. Relationships between mutated causative genes and signs (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
TNNT2	0.154861
MYBPC3	0.053789
TNNI3	-0.098472
MYL3	0.111798

Negligible relationships between mutated causative genes and signs were found.

Table 80. Predicting signs from mutated genes [Support: Yes (n = 28); No (n = 40)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.38	0.54	0.34	0.70	0.42	0.00	0.57	0.55	0.41	0.46	0.45	0.45	0.37	0.41
RF	0.53	0.51	0.69	0.67	0.06	0.06	0.57	0.55	0.46	0.44	0.45	0.45	0.40	0.40
Log Reg	0.38	0.54	0.40	0.70	0.36	0.00	0.53	0.54	0.39	0.46	0.44	0.44	0.37	0.41
GaussianNB	0.40	0.40	0.44	0.44	0.35	0.35	0.52	0.54	0.40	0.40	0.43	0.44	0.37	0.37
Ridge	0.54	0.54	0.70	0.70	0.00	0.00	0.55	0.54	0.46	0.46	0.45	0.44	0.41	0.41
SVC	0.54	0.54	0.70	0.70	0.00	0.00	0.46	0.52	0.46	0.46	0.41	0.44	0.41	0.41
Linear SVC	0.54	0.54	0.70	0.70	0.00	0.00	0.55	0.54	0.46	0.46	0.45	0.44	0.41	0.41

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 81. Relationships between mutated genes (and their combinations) and signs (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TNNT2	0.154861
DSC2	-0.113963
ABCC9	0.111798
FBN1	0.111798
MYL3	0.111798
PTPN11	0.111798
LAMA4	0.111798
TNNT2_TRPM4	0.111798
TNNT2_LAMA4	0.111798
MYBPC3_FBN1	0.111798
TRPM4_DSC2	-0.113963
TRPM4_LAMA4	0.111798
TTN_ABCC9	0.111798
TTN_DSP	0.111798
TTN_LAMA4	0.111798
ABCC9_DSP	0.111798

Negligible relationships between mutated genes (and their combinations) and signs were found.

Table 82. Predicting symptoms from mutated causative genes [Support: Yes (n = 44); No (n = 24)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.57	0.63	0.26	0.07	0.70	0.77	0.52	0.52	0.49	0.50	0.65	0.65	0.64	0.65
RF	0.63	0.63	0.07	0.00	0.77	0.77	0.52	0.52	0.50	0.49	0.65	0.65	0.65	0.64
Log Reg	0.57	0.65	0.26	0.00	0.70	0.79	0.52	0.52	0.49	0.50	0.65	0.65	0.64	0.65
GaussianNB	0.54	0.54	0.28	0.28	0.67	0.67	0.52	0.52	0.48	0.48	0.65	0.65	0.64	0.64
Ridge	0.63	0.63	0.00	0.00	0.77	0.77	0.52	0.52	0.49	0.49	0.65	0.65	0.64	0.64
SVC	0.63	0.63	0.00	0.00	0.77	0.77	0.47	0.53	0.49	0.49	0.65	0.67	0.64	0.64
Linear SVC	0.63	0.63	0.00	0.00	0.77	0.77	0.52	0.52	0.49	0.49	0.65	0.65	0.64	0.64

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 83. Relationships between mutated causative genes and symptoms (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.100252
MYBPC3	-0.068465
TNNI3	0.087186
MYL3	0.071027
TPM1	0.071027

Negligible relationships between mutated causative genes and symptoms were found.

Table 84. Predicting symptoms from mutated genes [Support: Yes (n = 44); No (n = 24)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.44	0.60	0.42	0.07	0.46	0.75	0.49	0.48	0.47	0.48	0.68	0.66	0.64	0.64
RF	0.63	0.62	0.00	0.00	0.77	0.76	0.52	0.48	0.49	0.48	0.69	0.66	0.64	0.64
Logistic Regression	0.46	0.65	0.37	0.00	0.52	0.79	0.53	0.53	0.46	0.50	0.66	0.68	0.63	0.65
GaussianNB	0.38	0.38	0.52	0.52	0.12	0.12	0.48	0.54	0.51	0.51	0.66	0.68	0.65	0.65
Ridge	0.63	0.63	0.00	0.00	0.77	0.77	0.48	0.48	0.49	0.49	0.66	0.67	0.64	0.64
SVC	0.63	0.63	0.00	0.00	0.77	0.77	0.45	0.54	0.49	0.49	0.63	0.69	0.64	0.64
Linear SVC	0.63	0.63	0.00	0.00	0.77	0.77	0.48	0.48	0.49	0.49	0.66	0.66	0.64	0.64

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 85. Relationships between mutated genes (and their combinations) and symptoms (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	0.100252
TTN	0.200446
RYR1	-0.126270
CSRP3	-0.126270
CACNA1C	-0.126270
PKP2	0.113067
ND	-0.115648
MYBPC3_DSP	-0.126270
MYBPC3_CSRP3	-0.126270
TRPM4_CACNA1C	-0.126270
TRPM4_ELN	-0.126270
DSC2_CACNA1C	-0.126270
DSC2_ELN	-0.126270
DSP_CSRP3	-0.126270
CACNA1C_ELN	-0.126270

Negligible relationships between mutated genes (and their combinations) and symptoms were mostly found. The only exception is relationship between mutation in *TTN* and symptoms wherein relationship is weak positive.

Table 86. Predicting mitral valve abnormalities from mutated causative genes [Support: Yes (n = 18); No (n = 45)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.57	0.71	0.70	0.83	0.27	0.00	0.52	0.45	0.48	0.50	0.30	0.26	0.28	0.29
RF	0.71	0.68	0.83	0.81	0.00	0.00	0.45	0.45	0.50	0.48	0.26	0.26	0.29	0.29
Log Reg	0.57	0.71	0.70	0.83	0.27	0.00	0.52	0.45	0.48	0.50	0.30	0.26	0.28	0.29
GaussianNB	0.49	0.49	0.64	0.64	0.16	0.16	0.52	0.52	0.39	0.39	0.30	0.30	0.26	0.26
Ridge	0.71	0.71	0.83	0.83	0.00	0.00	0.52	0.45	0.50	0.50	0.30	0.26	0.29	0.29
SVC	0.71	0.71	0.83	0.83	0.00	0.00	0.45	0.52	0.50	0.50	0.27	0.30	0.29	0.29
Linear SVC	0.71	0.71	0.83	0.83	0.00	0.00	0.52	0.45	0.50	0.50	0.30	0.26	0.29	0.29

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 87. Relationships between mutated causative genes and mitral valve abnormalities (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.136852
MYBPC3	0.082912
TNNI3	-0.077475
MYL3	-0.063104
TPM1	-0.063104

Negligible relationships between mutated causative genes and mitral valve abnormalities were found.

Table 88. Predicting mitral valve abnormalities from mutated genes [Support: Yes (n = 22); No (n = 53)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.47	0.65	0.52	0.79	0.39	0.00	0.56	0.58	0.50	0.46	0.34	0.37	0.29	0.29
RF	0.61	0.64	0.75	0.78	0.12	0.07	0.60	0.58	0.46	0.47	0.36	0.37	0.28	0.29
Log Reg	0.47	0.71	0.53	0.83	0.37	0.00	0.56	0.57	0.49	0.50	0.35	0.35	0.29	0.29
GaussianNB	0.31	0.31	0.16	0.16	0.41	0.41	0.56	0.59	0.46	0.46	0.36	0.37	0.28	0.28
Ridge	0.65	0.65	0.79	0.79	0.00	0.00	0.56	0.58	0.46	0.46	0.36	0.36	0.29	0.29
SVC	0.68	0.68	0.81	0.81	0.00	0.00	0.45	0.49	0.48	0.48	0.30	0.31	0.29	0.29
Linear SVC	0.64	0.64	0.78	0.78	0.07	0.07	0.56	0.58	0.47	0.47	0.35	0.37	0.29	0.29

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 89. Relationships between mutated genes (and their combinations) and mitral valve abnormalities (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
CACNA1C	0.140974
ELN	0.173009
PTPN11	0.140974
TPM1	-0.100711
MYH7_DSG2	0.140974
TNNT2_TNNI3	-0.100711
TNNT2_DSG2	0.140974
TRPM4_CACNA1C	0.140974
TRPM4_ELN	0.140974
DSC2_CACNA1C	0.140974
DSC2_ELN	0.140974
CACNA1C_ELN	0.140974

Negligible relationships between mutated genes (and their combinations) and mitral valve abnormalities were found.

Table 90. Predicting ventricular conduction disorders from mutated causative genes [Support: Yes (n = 13); No (n = 47)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.55	0.78	0.67	0.88	0.31	0.00	0.59	0.56	0.52	0.50	0.26	0.25	0.22	0.22
RF	0.78	0.72	0.88	0.83	0.00	0.00	0.59	0.56	0.50	0.46	0.25	0.25	0.22	0.22
Log Reg	0.55	0.78	0.67	0.88	0.31	0.00	0.59	0.56	0.52	0.50	0.26	0.25	0.22	0.22
GaussianNB	0.47	0.47	0.60	0.60	0.20	0.20	0.59	0.56	0.41	0.41	0.25	0.25	0.20	0.20
Ridge	0.78	0.78	0.88	0.88	0.00	0.00	0.59	0.56	0.50	0.50	0.26	0.25	0.22	0.22
SVC	0.78	0.78	0.88	0.88	0.00	0.00	0.60	0.56	0.50	0.50	0.27	0.25	0.22	0.22
Linear SVC	0.78	0.78	0.88	0.88	0.00	0.00	0.59	0.56	0.50	0.50	0.26	0.25	0.22	0.22

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 91. Relationships between mutated causative genes and ventricular conduction disorders (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
TNNT2	0.070941
MYBPC3	0.067379
MYL3	-0.053515
TPM1	-0.053515

Negligible relationships between mutated causative genes and ventricular conduction disorders were found.

Table 92. Predicting ventricular conduction disorders from mutated genes [Support: Yes (n = 13); No (n = 47)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.47	0.77	0.58	0.87	0.27	0.00	0.64	0.59	0.46	0.49	0.40	0.36	0.21	0.22
RF	0.78	0.78	0.88	0.88	0.00	0.00	0.64	0.59	0.50	0.50	0.39	0.36	0.22	0.22
Log Reg	0.47	0.78	0.59	0.88	0.24	0.00	0.67	0.61	0.44	0.50	0.40	0.36	0.20	0.22
GaussianNB	0.17	0.17	0.00	0.00	0.29	0.29	0.52	0.52	0.38	0.38	0.22	0.22	0.18	0.18
Ridge	0.77	0.77	0.87	0.87	0.00	0.00	0.67	0.62	0.49	0.49	0.40	0.36	0.22	0.22
SVC	0.78	0.78	0.88	0.88	0.00	0.00	0.47	0.55	0.50	0.50	0.33	0.36	0.22	0.22
Linear SVC	0.77	0.77	0.87	0.87	0.00	0.00	0.68	0.62	0.49	0.49	0.40	0.36	0.22	0.22

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 93. Relationships between mutated genes (and their combinations) and ventricular conduction disorders (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
TTN	0.188641
LAMA4	0.191657
HCN4	0.191657
TNNT2_TRPM4	0.191657
TNNT2_LAMA4	0.191657
TNNT2_PKP2	0.191657
TRPM4_LAMA4	0.191657
TTN_LAMA4	0.191657
TTN_PKP2	0.191657

Negligible relationships between mutated genes (and their combinations) and ventricular conduction disorders were found.

Table 94. Predicting left ventricular kinetics disorders from mutated causative genes [Support: Yes (n = 5); No (n = 57)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.31	0.92	0.41	0.96	0.16	0.00	0.63	0.63	0.53	0.50	0.11	0.11	0.09	0.08
RF	0.92	0.92	0.96	0.96	0.00	0.00	0.63	0.63	0.50	0.50	0.11	0.11	0.08	0.08
Log Reg	0.31	0.92	0.41	0.96	0.16	0.00	0.53	0.63	0.53	0.50	0.09	0.11	0.09	0.08
GaussianNB	0.26	0.26	0.34	0.34	0.15	0.15	0.53	0.63	0.51	0.51	0.09	0.11	0.08	0.08
Ridge	0.92	0.92	0.96	0.96	0.00	0.00	0.63	0.63	0.50	0.50	0.11	0.11	0.08	0.08
SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.47	0.63	0.50	0.50	0.08	0.11	0.08	0.08
Linear SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.63	0.63	0.50	0.50	0.11	0.11	0.08	0.08

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 95. Relationships between mutated causative genes and left ventricular kinetics disorders (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.102562
TNNI3	0.105570
TPM1	0.319860

Negligible relationships between mutated causative genes and left ventricular kinetics disorders were mostly found. The only exception is relationship between mutation in *TPM1* and left ventricular kinetics disorders wherein relationship is moderate positive.

Table 96. Predicting left ventricular kinetics disorders from mutated genes

[Support: Yes (n = 5); No (n = 57)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.34	0.92	0.47	0.96	0.13	0.00	0.61	0.62	0.46	0.50	0.11	0.13	0.08	0.08
RF	0.92	0.92	0.96	0.96	0.00	0.00	0.59	0.62	0.50	0.50	0.11	0.13	0.08	0.08
Logistic Regression	0.47	0.92	0.61	0.96	0.15	0.00	0.61	0.62	0.53	0.50	0.11	0.13	0.09	0.08
GaussianNB	0.34	0.34	0.45	0.45	0.16	0.16	0.61	0.61	0.55	0.55	0.10	0.10	0.09	0.09
Ridge	0.92	0.92	0.96	0.96	0.00	0.00	0.58	0.62	0.50	0.50	0.10	0.13	0.08	0.08
SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.55	0.48	0.50	0.50	0.10	0.10	0.08	0.08
Linear SVC	0.92	0.92	0.96	0.96	0.00	0.00	0.59	0.62	0.50	0.50	0.11	0.13	0.08	0.08

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 97. Relationships between mutated genes (and their combinations) and left ventricular kinetics disorders (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	-0.102562
TNNI3	0.105570
DSG2	0.144453
TRPM4	0.115223
RYR1	0.144453
CSRP3	0.144453
TPM1	0.319860
SDHA	0.319860
PDLIM3	0.225617
NKX2-5	0.225617
MYBPC3_DSP	0.144453
MYBPC3_CSRP3	0.144453
TNNI3_DSG2	0.225617
TNNI3_PDLIM3	0.225617
TNNI3_NKX2_5	0.225617
DSG2_PDLIM3	0.225617
DSG2_NKX2_5	0.225617
DSP_CSRP3	0.144453
TPM1_SDHA	0.319860
PDLIM3_NKX2_5	0.225617

Negligible relationships between mutated genes and left ventricular kinetics disorders were mostly found. Single-gene exceptions are relationships between mutation in *PDLIM3* and left ventricular kinetics disorders, and between mutation in *NKX2-5* and left ventricular kinetics disorders, wherein relationships are weak positive; as well as between mutation in *TPM1* and left ventricular kinetics disorders, and mutation in *SDHA* and left ventricular kinetics disorders, wherein relationships are moderate positive.

Table 98. Predicting atrial conduction disorders from mutated causative genes

[Support: Yes (n = 6); No (n = 53)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.54	0.90	0.69	0.95	0.13	0.00	0.39	0.39	0.45	0.50	0.09	0.09	0.09	0.10
RF	0.90	0.90	0.95	0.95	0.00	0.00	0.39	0.39	0.50	0.50	0.09	0.09	0.10	0.10
Log Reg	0.54	0.90	0.69	0.95	0.13	0.00	0.39	0.39	0.45	0.50	0.09	0.09	0.09	0.10
GaussianNB	0.27	0.27	0.32	0.32	0.22	0.22	0.39	0.39	0.59	0.59	0.09	0.09	0.12	0.12
Ridge	0.90	0.90	0.95	0.95	0.00	0.00	0.39	0.39	0.50	0.50	0.09	0.09	0.10	0.10
SVC	0.90	0.90	0.95	0.95	0.00	0.00	0.62	0.48	0.50	0.50	0.15	0.10	0.10	0.10
Linear SVC	0.90	0.90	0.95	0.95	0.00	0.00	0.39	0.39	0.50	0.50	0.09	0.09	0.10	0.10

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 99. Relationships between mutated causative genes and atrial conduction disorders (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	0.101205

Negligible relationships between mutated causative genes and atrial conduction disorders were found.

Table 100. Predicting atrial conduction disorders from mutated genes [Support: Yes (n = 5); No (n = 69)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.49	0.90	0.64	0.95	0.12	0.00	0.40	0.52	0.42	0.50	0.09	0.12	0.09	0.10
RF	0.90	0.90	0.95	0.95	0.00	0.00	0.39	0.52	0.50	0.50	0.09	0.12	0.10	0.10
Logistic Regression	0.51	0.90	0.66	0.95	0.12	0.00	0.44	0.46	0.43	0.50	0.10	0.10	0.09	0.10
GaussianNB	0.39	0.39	0.49	0.49	0.25	0.25	0.57	0.59	0.66	0.66	0.12	0.12	0.14	0.14
Ridge	0.90	0.90	0.95	0.95	0.00	0.00	0.41	0.46	0.50	0.50	0.09	0.10	0.10	0.10
SVC	0.90	0.90	0.95	0.95	0.00	0.00	0.47	0.53	0.50	0.50	0.10	0.12	0.10	0.10
Linear SVC	0.90	0.90	0.95	0.95	0.00	0.00	0.41	0.52	0.50	0.50	0.09	0.12	0.10	0.10

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 101. Relationships between mutated genes (and their combinations) and atrial conduction disorders (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
MYH7	0.101205
HCN4	0.133487

Negligible relationships between mutated genes (and their combinations) and atrial conduction disorders were found.

Table 102. Predicting myocardial injury from mutated causative genes [Support: Yes (n = 43); No (n = 16)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.44	0.73	0.20	0.00	0.57	0.84	0.49	0.44	0.38	0.50	0.72	0.71	0.69	0.73
RF	0.73	0.73	0.00	0.00	0.84	0.84	0.53	0.44	0.50	0.50	0.74	0.71	0.73	0.73
Log Reg	0.46	0.73	0.20	0.00	0.59	0.84	0.43	0.44	0.39	0.50	0.70	0.71	0.69	0.73
GaussianNB	0.44	0.44	0.23	0.23	0.56	0.56	0.44	0.44	0.40	0.40	0.71	0.71	0.69	0.69
Ridge	0.73	0.73	0.00	0.00	0.84	0.84	0.43	0.44	0.50	0.50	0.70	0.71	0.73	0.73
SVC	0.73	0.73	0.00	0.00	0.84	0.84	0.54	0.44	0.50	0.50	0.75	0.71	0.73	0.73
Linear SVC	0.73	0.73	0.00	0.00	0.84	0.84	0.43	0.44	0.50	0.50	0.70	0.71	0.73	0.73

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 103. Relationships between mutated causative genes and myocardial injury (only coefficients > 0.05 or < -0.05 are shown)

Gene	Phi coefficient
MYH7	-0.052373
MYL3	0.061386

Negligible relationships between mutated causative genes and myocardial injury were found.

Table 104. Predicting myocardial injury from mutated genes [Support: Yes (n = 43); No (n = 16)]

Estimator	Accuracy		F1-score				AUC from estimator		AUC from predictions		AP (PR) from estimator		AP (PR) from predictions	
			No		Yes									
Oversampling	-	+	-	+	-	+	-	+	-	+	-	+	-	+
DT	0.36	0.59	0.17	0.08	0.47	0.74	0.52	0.52	0.32	0.43	0.75	0.75	0.67	0.70
RF	0.63	0.69	0.00	0.00	0.77	0.82	0.51	0.52	0.43	0.48	0.76	0.77	0.70	0.72
Log Reg	0.42	0.73	0.15	0.00	0.56	0.84	0.61	0.55	0.35	0.50	0.80	0.78	0.68	0.73
GaussianNB	0.34	0.34	0.38	0.38	0.29	0.29	0.61	0.57	0.47	0.47	0.80	0.79	0.72	0.72
Ridge	0.69	0.69	0.00	0.00	0.82	0.82	0.58	0.57	0.48	0.48	0.79	0.78	0.72	0.72
SVC	0.73	0.73	0.00	0.00	0.84	0.84	0.50	0.56	0.50	0.50	0.72	0.76	0.73	0.73
Linear SVC	0.69	0.69	0.00	0.00	0.82	0.82	0.51	0.52	0.48	0.48	0.76	0.77	0.72	0.72

DT — decision tree; RF — random forest; Log Reg — logistic regression; GaussianNB — Gaussian Naive Bayes; SVC — radial basis function kernel C-support vector classification; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

Table 105. Relationships between mutated genes (and their combinations) and myocardial injury (only coefficients > 0.10 or < -0.10 are shown)

Gene	Phi coefficient
DSG2	-0.119060
TRPM4	-0.109135
TTN	0.167710
RYR1	-0.168812
DSC2	-0.158170
CACNA1C	-0.168812
ELN	-0.113067
PKP2	-0.122279
MYBPC3_DSG2	-0.119060
MYBPC3_PKP2	-0.207289
DSG2_FBN1	-0.119060
DSG2_PKP2	-0.119060
TRPM4_DSC2	-0.158170
TRPM4_CACNA1C	-0.168812
TRPM4_ELN	-0.168812
DSC2_CACNA1C	-0.168812
DSC2_ELN	-0.168812
FBN1_PKP2	-0.119060
CACNA1C_ELN	-0.168812

Negligible relationships between mutated genes and myocardial injury were found.

4.1.4.2. Predicting phenotypic outcomes using both genetic and other phenotypic data

ML algorithms were able to predict phenotypic outcomes — fatigue, dyspnea, chest pain, palpitations, syncope, heart murmur, pretibial edema, SAM, papillary muscle abnormalities, hypokinesia, AF, AV block I, LBBB, RBBB, left anterior hemiblock, ST segment abnormalities, and negative T wave — using genotypic and phenotypic data. Performance of generated models has been shown in Table 106 (283).

Table 106. Performance of generated models for predicting phenotypic outcomes using both genetic and other phenotypic data (283)

Predicted feature	Estimator	Accuracy	Precision	Recall	F1-Score	AUC from estimator	AUC from predictions	AP (PR) from estimator	AP (PR) from predictions
Fatigue	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dyspnea	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Chest pain	RF	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Palpitations	SVC	0.97	0.97	0.83	0.91	0.92	0.92	0.86	0.86
Syncope	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Heart murmur	SVC	0.89	0.82	0.76	0.87	0.88	0.88	0.88	0.88
Pretibial edema	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Systolic anterior motion	SVC	0.99	0.98	0.93	0.97	0.97	0.97	0.95	0.95
Papillary muscle abnormalities	LogReg	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Hypokinesia	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Atrial fibrillation	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AV block I	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LBBB	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RBBB	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Left anterior hemiblock	SVC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ST segment abnormalities	RF	0.93	0.81	0.89	0.90	0.97	0.94	0.98	0.96
Negative T wave	SVC	0.93	0.88	0.85	0.92	0.93	0.93	0.92	0.92

SVC — radial basis function (RBF) kernel C-support vector classification; RF — random forest; LogReg — logistic regression; AUC — area under the receiver operating characteristic curve; AP (PR) — average precision (precision-recall curve)

4.1.4.2.1. Symptoms

HCM symptoms were predicted by subgroups of other genotypic and phenotypic features (Figures 127-131) (283).

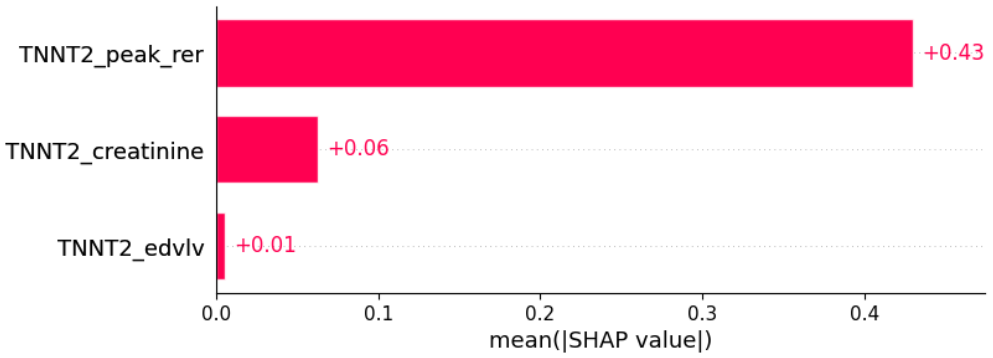

Figure 127. Fatigue was predicted by the shown subset of features. Their relative importance is indicated (283).

*TNNT2*_peak_rer = mutation in *TNNT2* x peak respiratory exchange ratio; *TNNT2*_creatinine = mutation in *TNNT2* x serum creatinine; *TNNT2*_edvlv = mutation in *TNNT2* x EDVLV

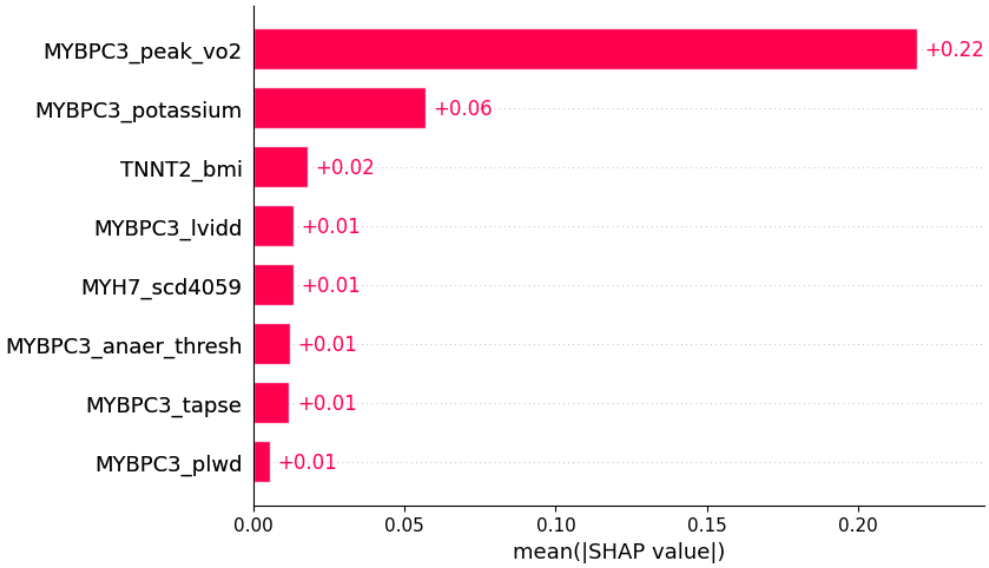

Figure 128. Dyspnea was predicted by the shown subset of features. Their relative importance is indicated (283).

*MYBPC3*_peak_vo2 = mutation in *MYBPC3* x peak VO_2; *MYBPC3*_potassium = mutation in *MYBPC3* x serum potassium; *TNNT2*_bmi = mutation in *TNNT2* x body mass index; *MYBPC3*_lvidd = mutation in *MYBPC3* x LVIDd; *MYH7*_scd4059 = mutation in *MYH7* x SCD in age 40–59 in family history; *MYBPC3*_anaer_thresh = mutation in *MYBPC3* x anaerobic threshold; *MYBPC3*_tapse = mutation in *MYBPC3* x TAPSE; *MYBPC3*_plwd = mutation in *MYBPC3* x PLWD

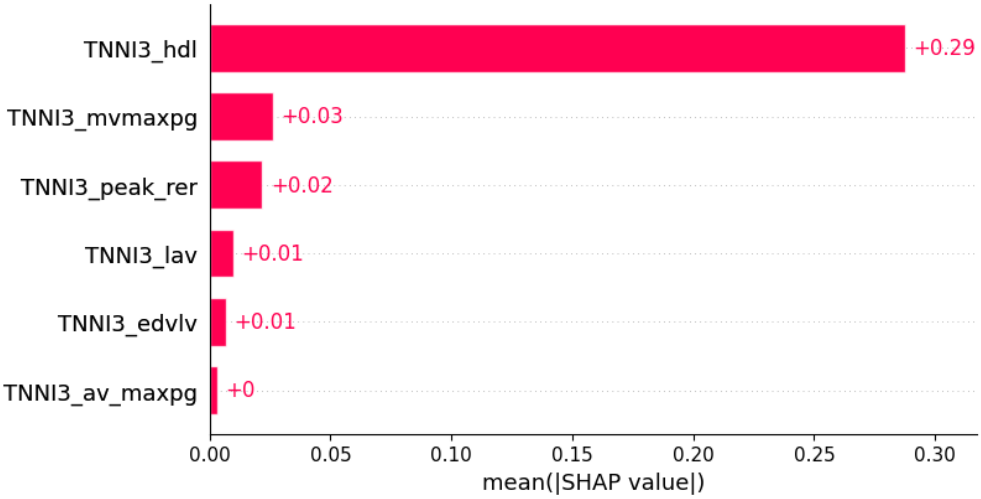

Figure 129. Chest pain was predicted by the shown subset of features. Their relative importance is indicated (283).

*TNNI3*_hdl = mutation in *TNNI3* x serum HDL; *TNNI3*_mvmaxpg = mutation in *TNNI3* x MV maxPG; *TNNI3*_peak_rer = mutation in *TNNI3* x peak respiratory exchange ratio; *TNNI3*_lav = mutation in *TNNI3* x LAV; *TNNI3*_edvlv = mutation in *TNNI3* x EDVLV; *TNNI3*_av_maxpg = mutation in *TNNI3* x AV maxPG

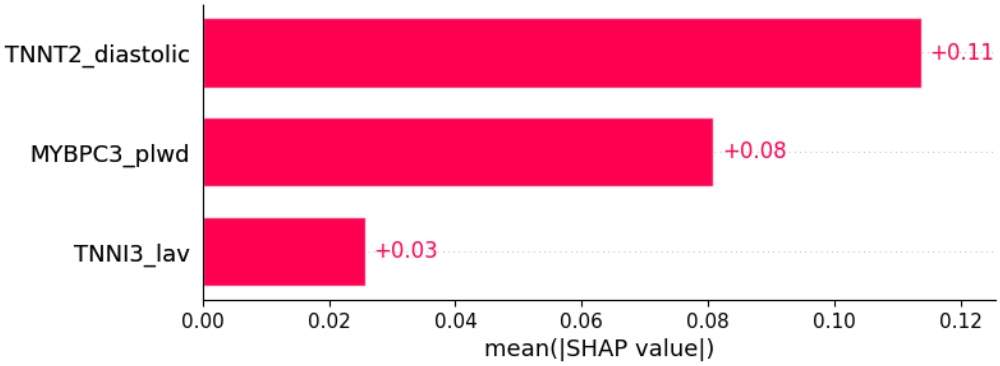

Figure 130. Palpitations were predicted by the shown subset of features. Their relative importance is indicated (283).

*TNNT2*_diastolic = mutation in *TNNT2* x diastolic blood pressure; *MYBPC3*_plwd = mutation in *MYBPC3* x PLWD; *TNNI3*_lav = mutation in *TNNI3* x LAV

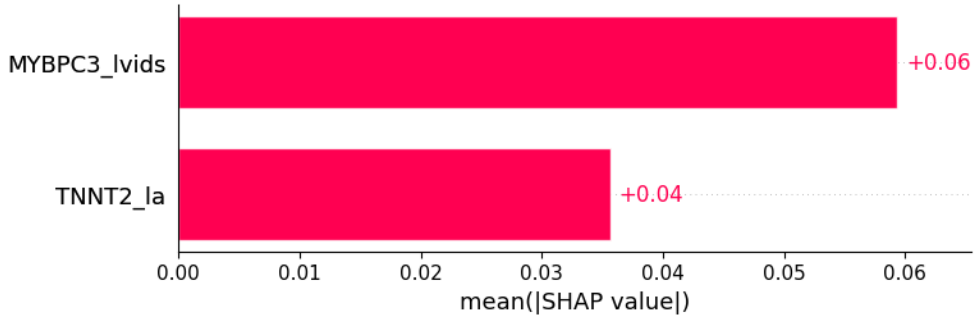

Figure 131. Syncope was predicted by the shown subset of features. Their relative importance is indicated (283).

*MYBPC3*_lvids = mutation in *MYBPC3* x LVIDs; *TNNT2*_la = mutation in *TNNT2* x LA

4.1.4.2.2. Signs

HCM signs were predicted by subgroups of other genotypic and phenotypic features (Figures 132 and Figure 133) (283).

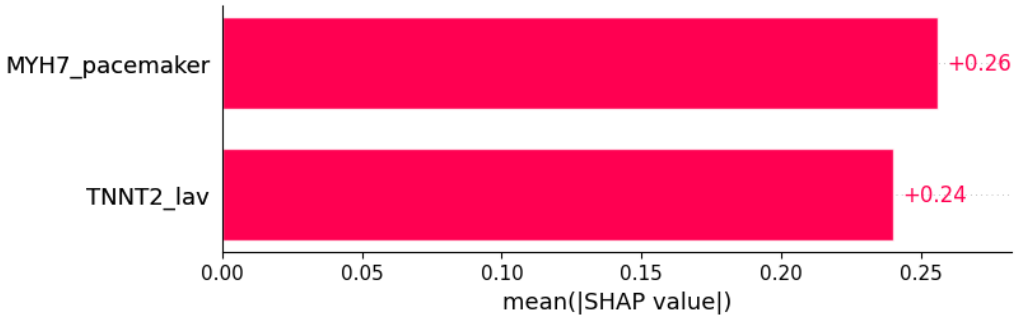

Figure 132. Heart murmur was predicted by the shown subset of features. Their relative importance is indicated (283).

*MYH7*_pacemaker = mutation in *MYH7* x pacemaker/defibrillator implants in family history; *TNNT2*_lav = mutation in *TNNT2* x LAV

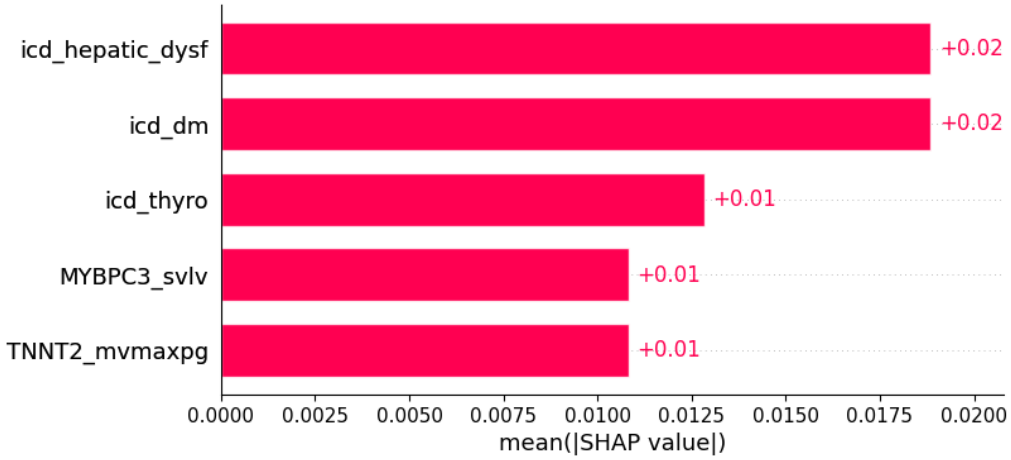

Figure 133. Pretibial edema was predicted by the shown subset of features. Their relative importance is indicated (283).

icd_hepatic_dysf = ICD x hepatic dysfunction; icd_dm = ICD x diabetes mellitus; icd_thyro = ICD x thyroid disease; *MYBPC3*_svlv = mutation in *MYBPC3* x SVLV; *TNNT2*_mvmaxpg = mutation in *TNNT2* x MV maxPG

4.1.4.2.3. Echocardiography

Certain ultrasonic findings in HCM were predicted by subgroups of other genotypic and phenotypic features (Figures 134-136) (283).

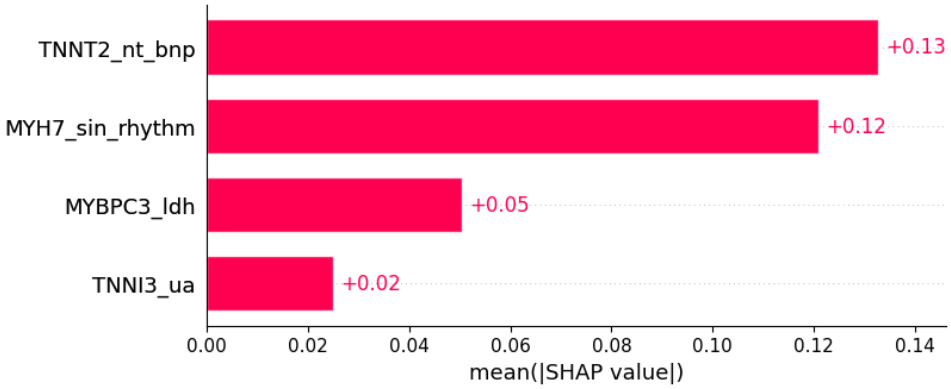

Figure 134. SAM was predicted by the shown subset of features. Their relative importance is indicated (283).

*TNNT2*_nt_bnp = mutation in *TNNT2* x NT BNP; *MYH7*_sin_rhythm = mutation in *MYH7* x sinus rhythm; *MYBPC3*_ldh = mutation in *MYBPC3* x LDH; *TNNI3*_ua = mutation in *TNNI3* x serum uric acid

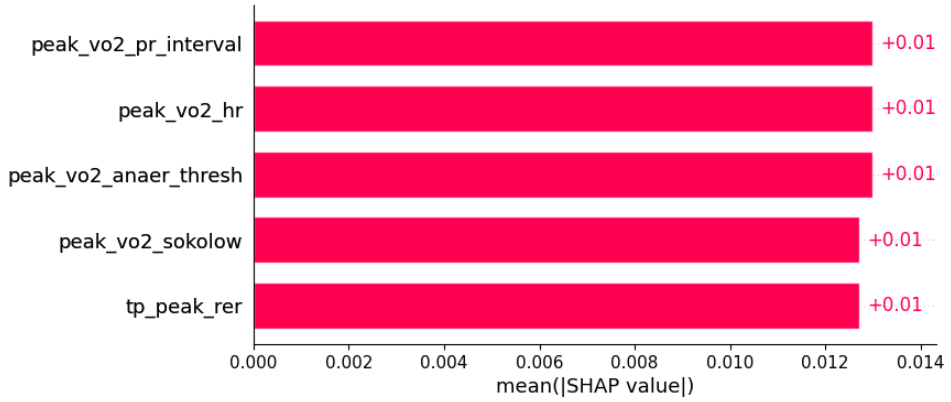

Figure 135. Papillary muscle abnormalities were predicted by the shown subset of features. Their relative importance is indicated (283).

peak_vo2_pr_interval = peak VO_2/PR interval; peak_vo2_hr = peak VO_2/heart rate; peak_vo2_anaer_thresh = peak VO_2/anaerobic threshold; peak_vo2_sokolow = peak VO_2/Sokolow index; tp_peak_rer = total protein in serum/peak respiratory exchange ratio

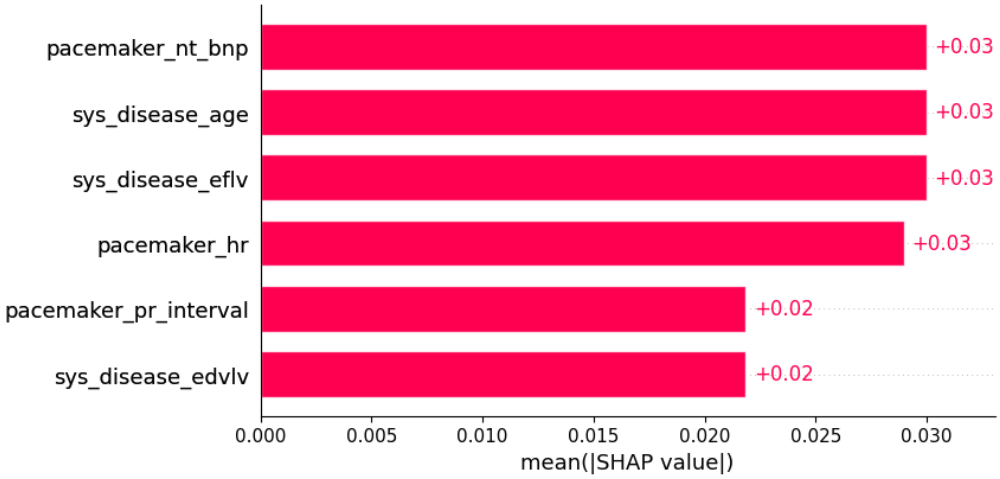

Figure 136. Hypokinesia was predicted by the shown subset of features. Their relative importance is indicated (283).

pacemaker_nt_bnp = pacemaker/defibrillator implants in family history x NT BNP; sys_disease_age = evidence of systemic disease in family history x age; sys_disease_eflv = evidence of systemic disease in family history x EFLV; pacemaker_hr = pacemaker/defibrillator implants in family history x heart rate; pacemaker_pr_interval = pacemaker/defibrillator implants in family history x PR interval; sys_disease_edvlv = evidence of systemic disease in family history x EDVLV

4.1.4.2.4. Conduction and rhythm disorders

Certain conduction and rhythm disorders in HCM were predicted by subgroups of other genotypic and phenotypic features (Figures 137-141) (283).

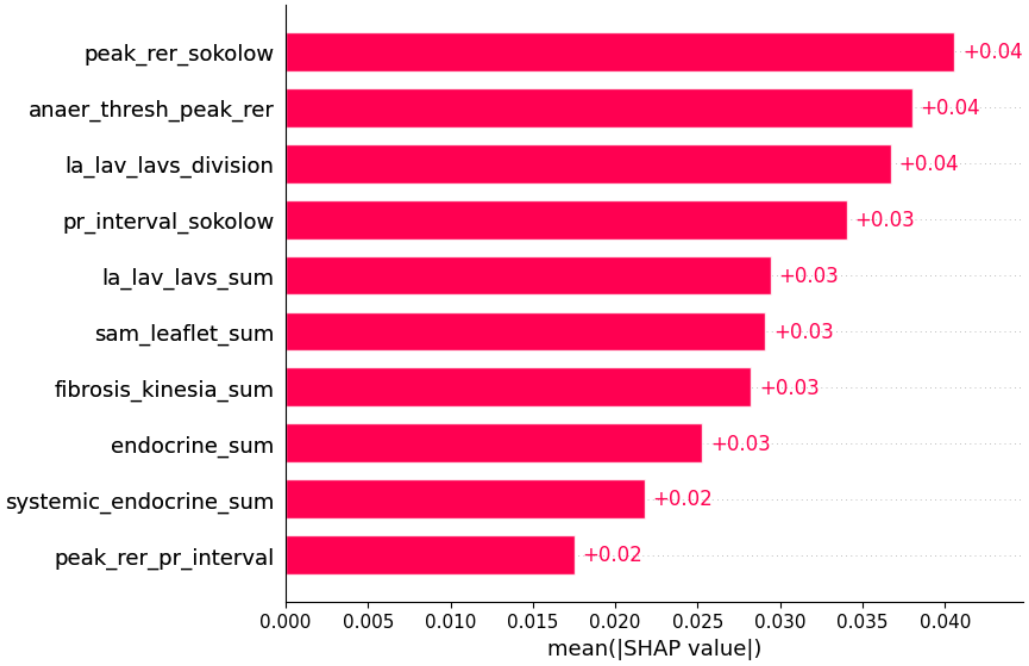

Figure 137. AF was predicted by the shown subset of features. Their relative importance is indicated (283).

peak_rer_sokolow = peak respiratory exchange ratio x Sokolow index; anaer_thresh_peak_rer = anaerobic threshold/peak respiratory exchange ratio; la_lav_lavs_division = LA/LAV/LAVs; pr_interval_sokolow = PR interval x Sokolow index; la_lav_lavs_sum = LA+LAV+LAVs; sam_leaflet_sum = systolic anterior motion + mitral regurgitation + papillary muscle abnormalities + mitral leaflet abnormalities + calcification of mitral annulus; fibrosis_kinesia_sum = myocardial fibrosis + hypokinesia + akinesia + dyskinesia + hyperkinesia; endocrine_sum = diabetes mellitus + thyroid disease + phaeochromocytoma + acromegaly; systemic_endocrine_sum = evidence of systemic disease in family history + diabetes mellitus + thyroid disease + phaeochromocytoma + acromegaly + neuromuscular disease + amyloidosis + genetic disease as a comorbidity; peak_rer_pr_interval = peak respiratory exchange ratio/PR interval

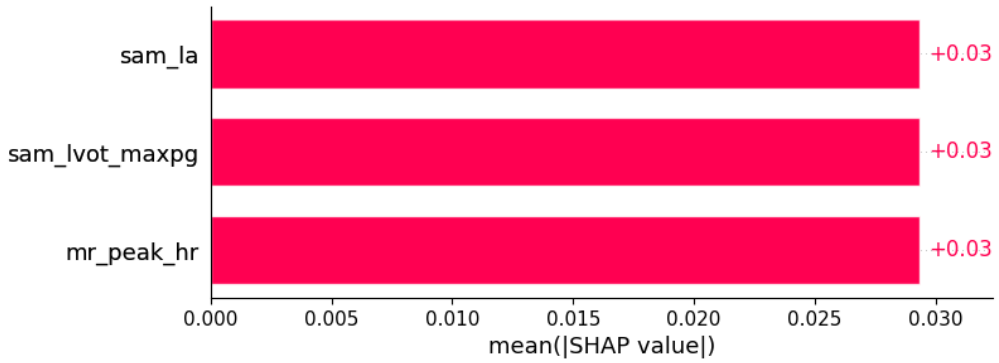

Figure 138. AV block I was predicted by the shown subset of features. Their relative importance is indicated (283).

sam_la = systolic anterior motion x LA; sam_lvot_maxpg = systolic anterior motion x LVOT maxPG; mr_peak_hr = mitral regurgitation x peak HR in cardiopulmonary exercise test

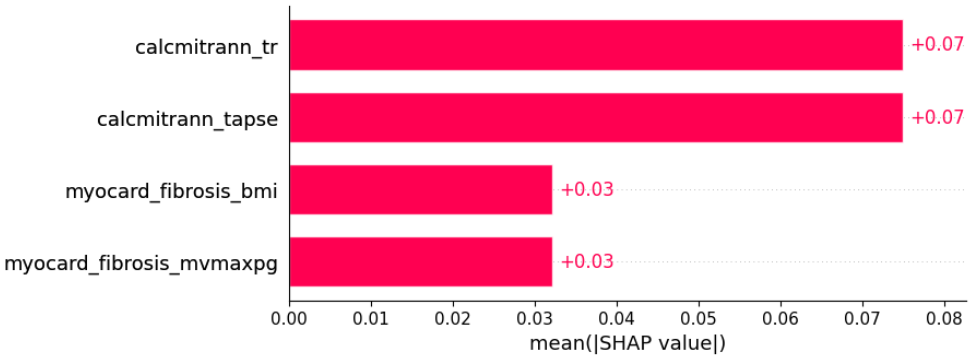

Figure 139. LBBB was predicted by the shown subset of features. Their relative importance is indicated (283).

calcmitrann_tr = calcification of mitral annulus x tricuspid regurgitation; calcmitrann_tapse = calcification of mitral annulus x TAPSE; myocard_fibrosis_bmi = myocardial fibrosis x body mass index; myocardial_fibrosis_mvmaxpg = myocardial fibrosis x MV maxPG

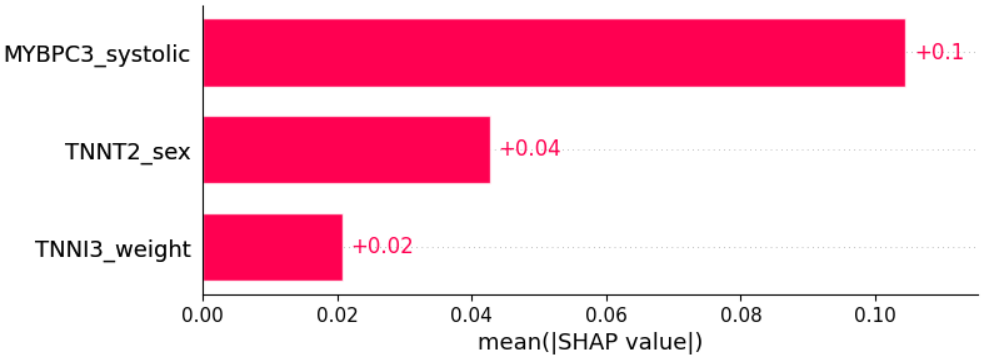

Figure 140. RBBB was predicted by the shown subset of features. Their relative importance is indicated (283).

*MYBPC3*_systolic = mutation in *MYBPC3* x systolic blood pressure; *TNNT2*_sex = mutation in *TNNT2* x sex; *TNNI3*_weight = mutation in *TNNI3* x weight

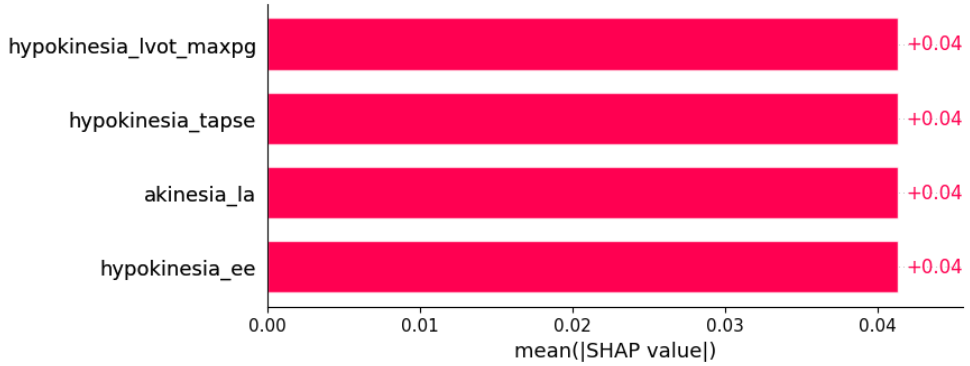

Figure 141. Left anterior hemiblock was predicted by the shown subset of features. Their relative importance is indicated (283).

hypokinesia_lvot_max_pg = hypokinesia x LVOT maxPG; hypokinesia_tapse = hypokinesia x TAPSE; akinesia_la = akinesia x LA; hypokinesia_ee = hypokinesia x E/E'

4.1.4.2.5. Ischemia

Certain ECG findings indicating myocardial ischemia in HCM were predicted by subgroups of other genotypic and phenotypic features (Figures 142 and 143) (283).

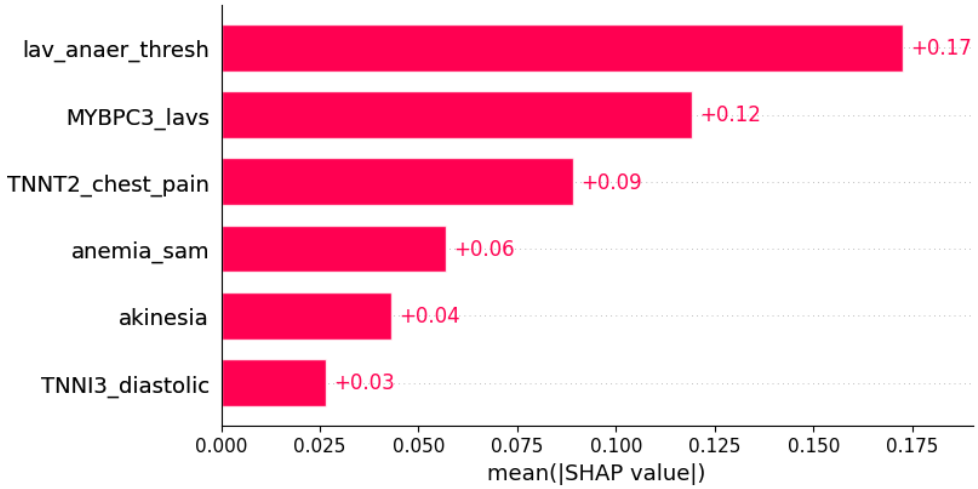

Figure 142. ST segment abnormalities were predicted by the shown subset of features. Their relative importance is indicated (283).

lav_anaer_thesh = LAV/anaerobic threshold; *MYBPC3*_lavs = mutation in *MYBPC3* x LAVs; *TNNT2*_chest_pain = mutation in *TNNT2* x chest pain; anemia_sam = anemia x systolic anterior motion; *TNNI3*_diastolic = mutation in *TNNI3* x diastolic blood pressure

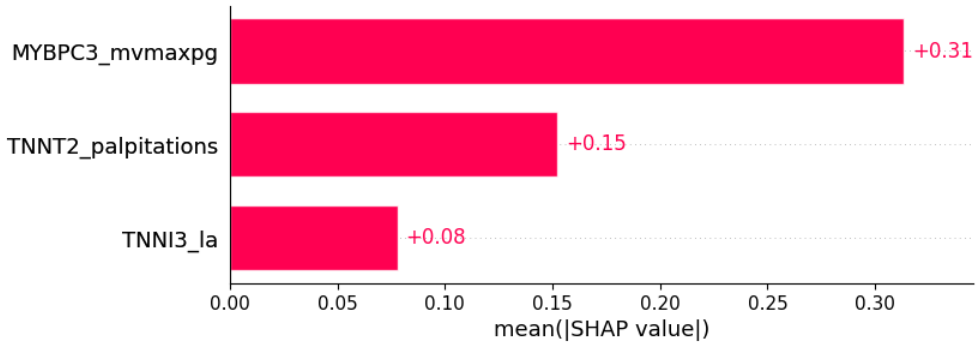

Figure 143. Negative T wave was predicted by the shown subset of features. Their relative importance is indicated (283).

*MYBPC3*_mvmaxpg = mutation in *MYBPC3* x MV maxPG; *TNNT2*_palpitations = mutation in *TNNT2* x palpitations; *TNNI3*_la = mutation in *TNNI3* x LA

4.1.5. Genotype-specific echocardiogram findings

Genotype-specific echocardiogram findings were identified.

4.1.5.1. *MYH7*-specific echocardiogram findings

4.1.5.1.1. Parasternal long axis view, ventricular diastole

A model for echocardiographic images classification (parasternal long axis view during ventricular diastole) of patients with mutation in *MYH7* gene and without detected mutation was created, with performance shown in Table 107 and Figure 144. The discriminative parts for classification of a given image (areas that contributed the most to the decision for the classification) are shown in Figure 145 (the deeper the highlighted color, the more important the region is for a particular class prediction). Septum and LVOT segment were shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation (Figure 145).

Table 107. Performance of model for echocardiographic images classification (*MYH7* vs. ND), parasternal long axis view during ventricular diastole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

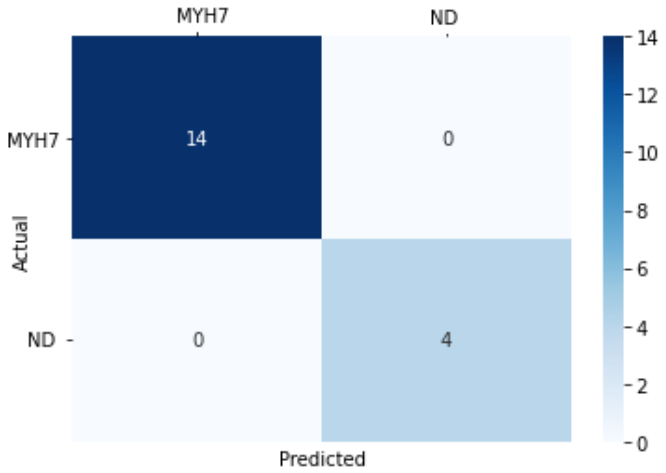

Figure 144. Confusion matrix, model for echocardiographic images classification (*MYH7* vs. ND), parasternal long axis view during ventricular diastole

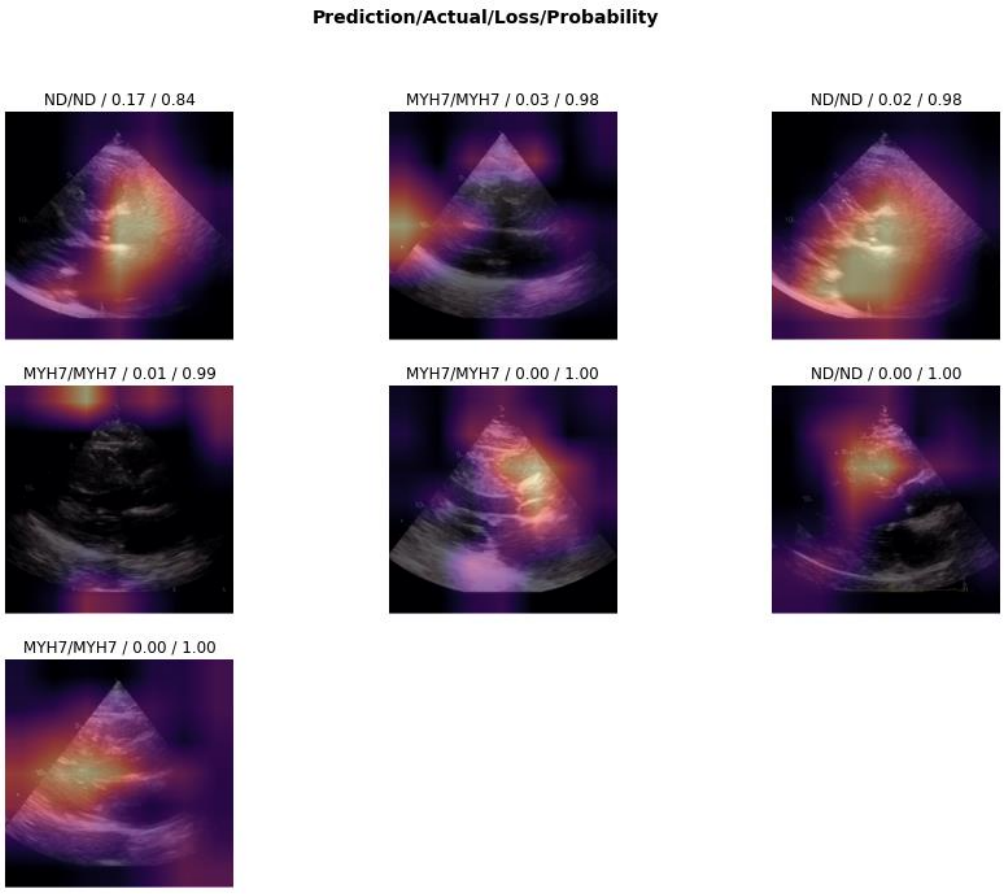

Figure 145. Discriminative areas for classification of a given image (*MYH7* vs. ND), parasternal long axis view during ventricular diastole

4.1.5.1.2. Parasternal long axis view, ventricular systole

A model for echocardiographic images classification (parasternal long axis view during ventricular systole) of patients with mutation in *MYH7* gene and without detected mutation was created, with performance shown in Table 108 and Figure 146. Anterior wall and apex were shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation (Figure 147).

Table 108. Performance of model for echocardiographic images classification (*MYH7* vs. ND), parasternal long axis view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

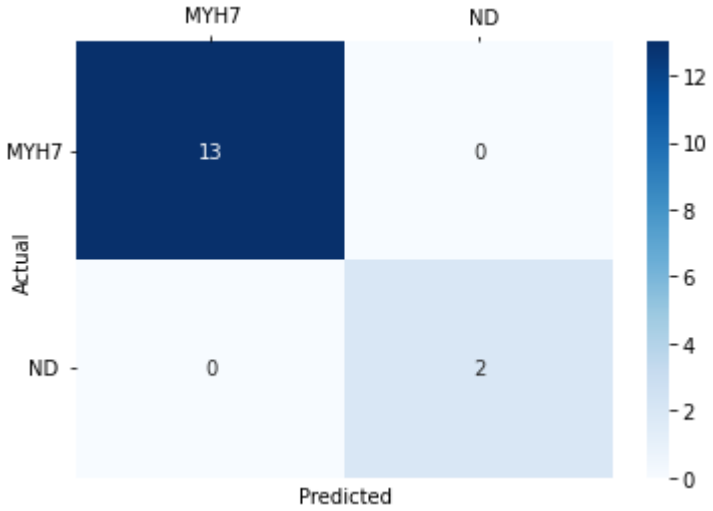

Figure 146. Confusion matrix, model for echocardiographic images classification (*MYH7* vs. ND), parasternal long axis view during ventricular systole

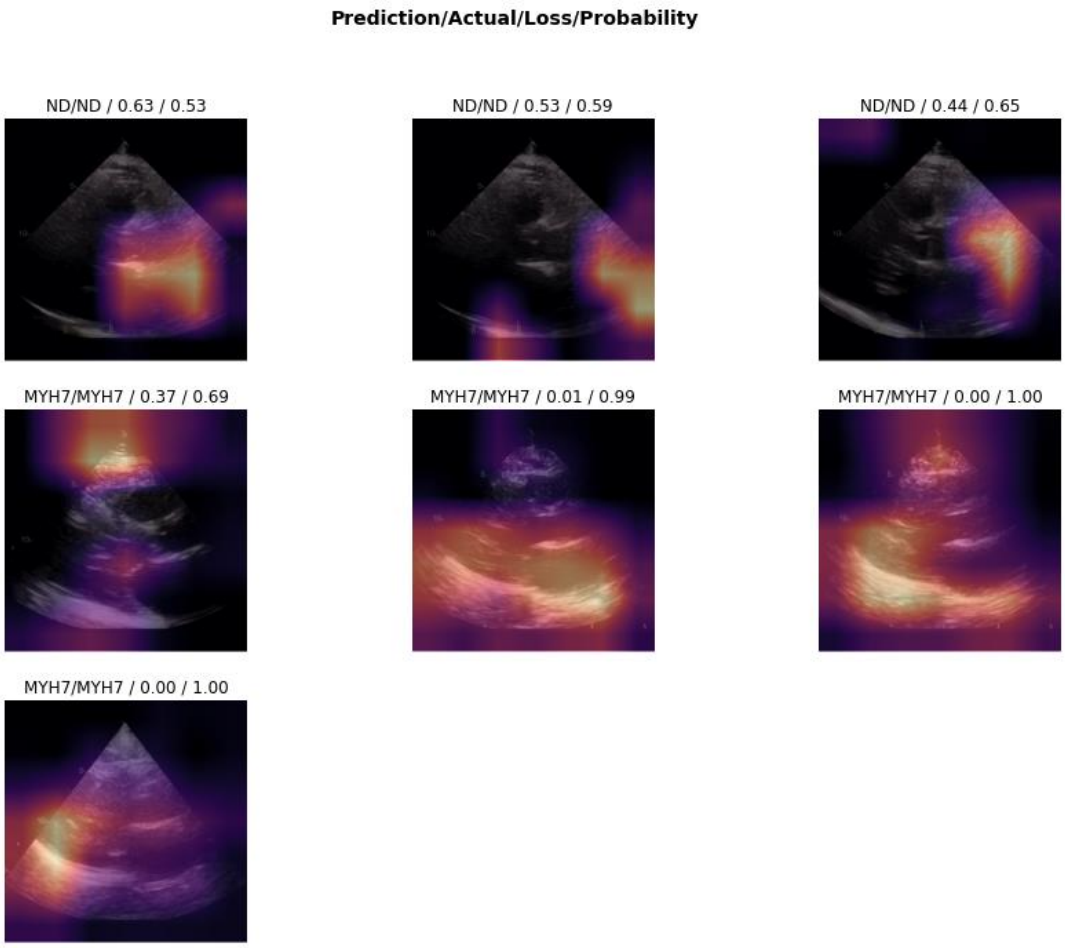

Figure 147. Discriminative areas for classification of a given image (*MYH7* vs. ND), parasternal long axis view during ventricular systole

4.1.5.1.3. Apical 2-chamber view, ventricular diastole

A model for echocardiographic images classification (apical 2-chamber view during ventricular diastole) of patients with mutation in *MYH7* gene and without detected mutation was created, with performance shown in Table 109 and Figure 148. No unequivocal discriminative areas for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation were found in this view (Figure 149).

Table 109. Performance of model for echocardiographic images classification (*MYH7* vs. ND), apical 2-chamber view during ventricular diastole

Metric	Value
Accuracy	0.928
Precision	1.000
Recall	0.875

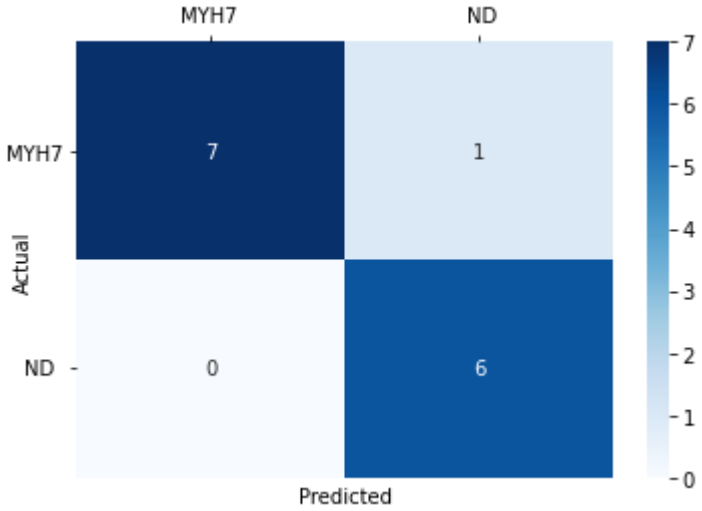

Figure 148. Confusion matrix, model for echocardiographic images classification (*MYH7* vs. ND), apical 2-chamber view during ventricular diastole

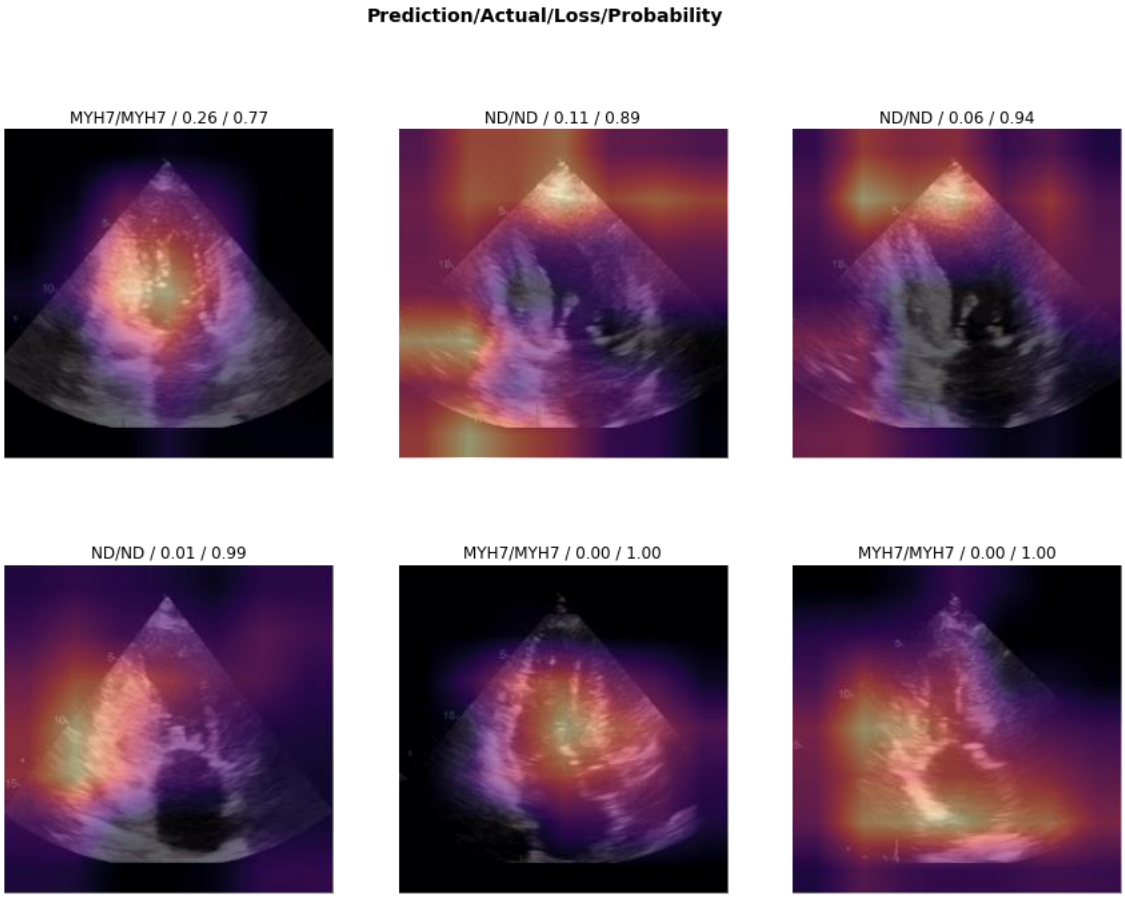

Figure 149. Discriminative areas for classification of a given image (*MYH7* vs. ND), apical 2-chamber view during ventricular diastole

4.1.5.1.4. Apical 2-chamber view, ventricular systole

A model for echocardiographic images classification (apical 2-chamber view during ventricular systole) of patients with mutation in *MYH7* gene and without detected mutation was created, with performance shown in Table 110 and Figure 150. No unequivocal discriminative areas for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation were found in this view (Figure 151).

Table 110. Performance of model for echocardiographic images classification (*MYH7* vs. ND), apical 2-chamber view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

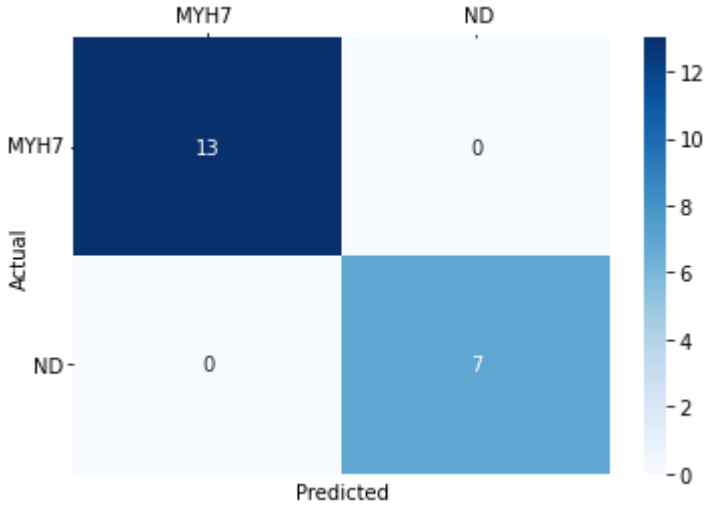

Figure 150. Confusion matrix, model for echocardiographic images classification (*MYH7* vs. ND), apical 2-chamber view during ventricular systole

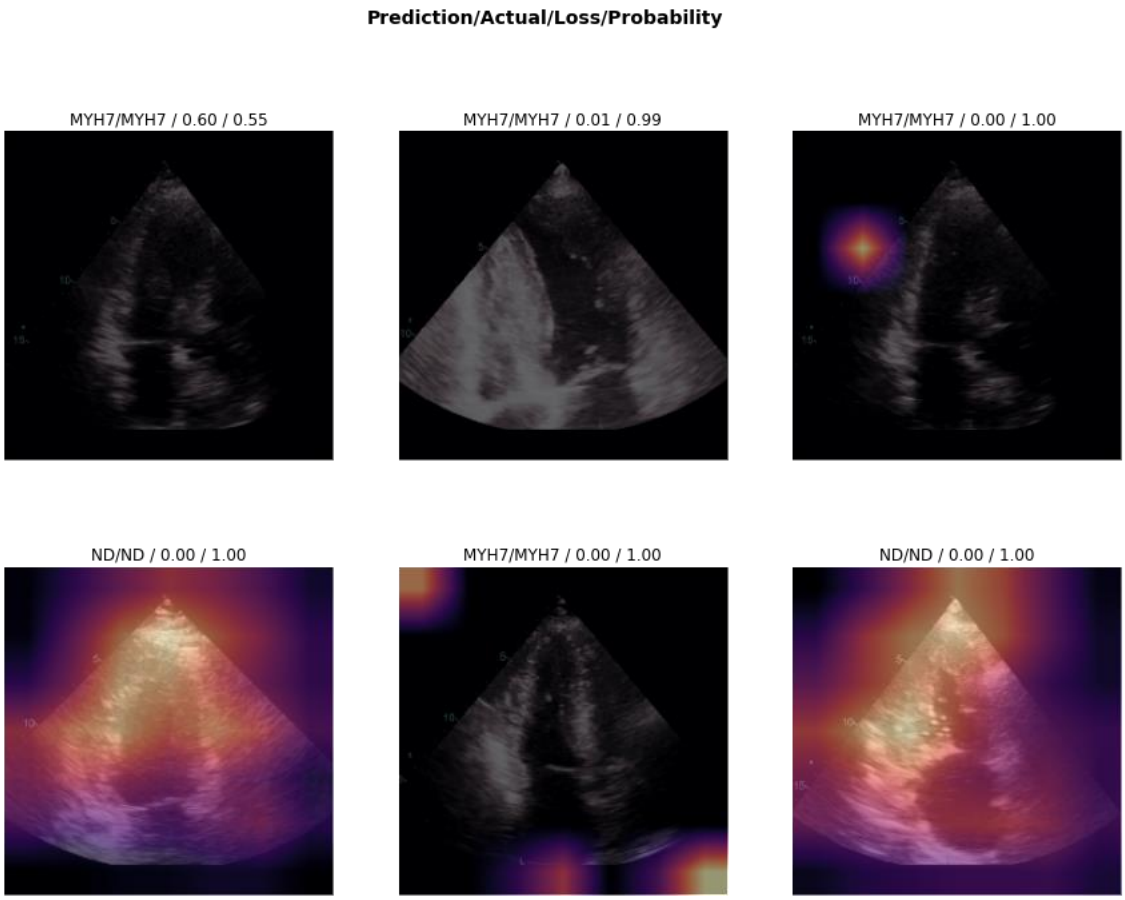

Figure 151. Discriminative areas for classification of a given image (*MYH7* vs. ND), apical 2-chamber view during ventricular systole

4.1.5.1.5. Apical 4-chamber view, ventricular diastole

A model for echocardiographic images classification (apical 4-chamber view during ventricular diastole) of patients with mutation in *MYH7* gene and without detected mutation was created, with performance shown in Table 111 and Figure 152. RV was shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation (Figure 153).

Table 111. Performance of model for echocardiographic images classification (*MYH7* vs. ND), apical 4-chamber view during ventricular diastole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

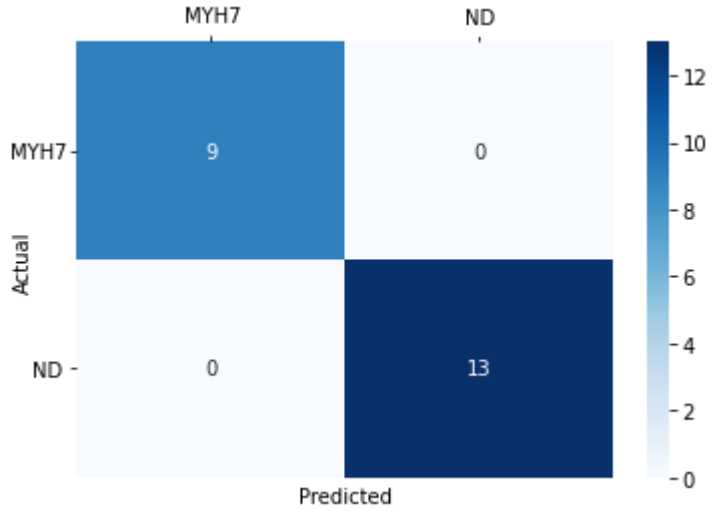

Figure 152. Confusion matrix, model for echocardiographic images classification (*MYH7* vs. ND), apical 4-chamber view during ventricular diastole

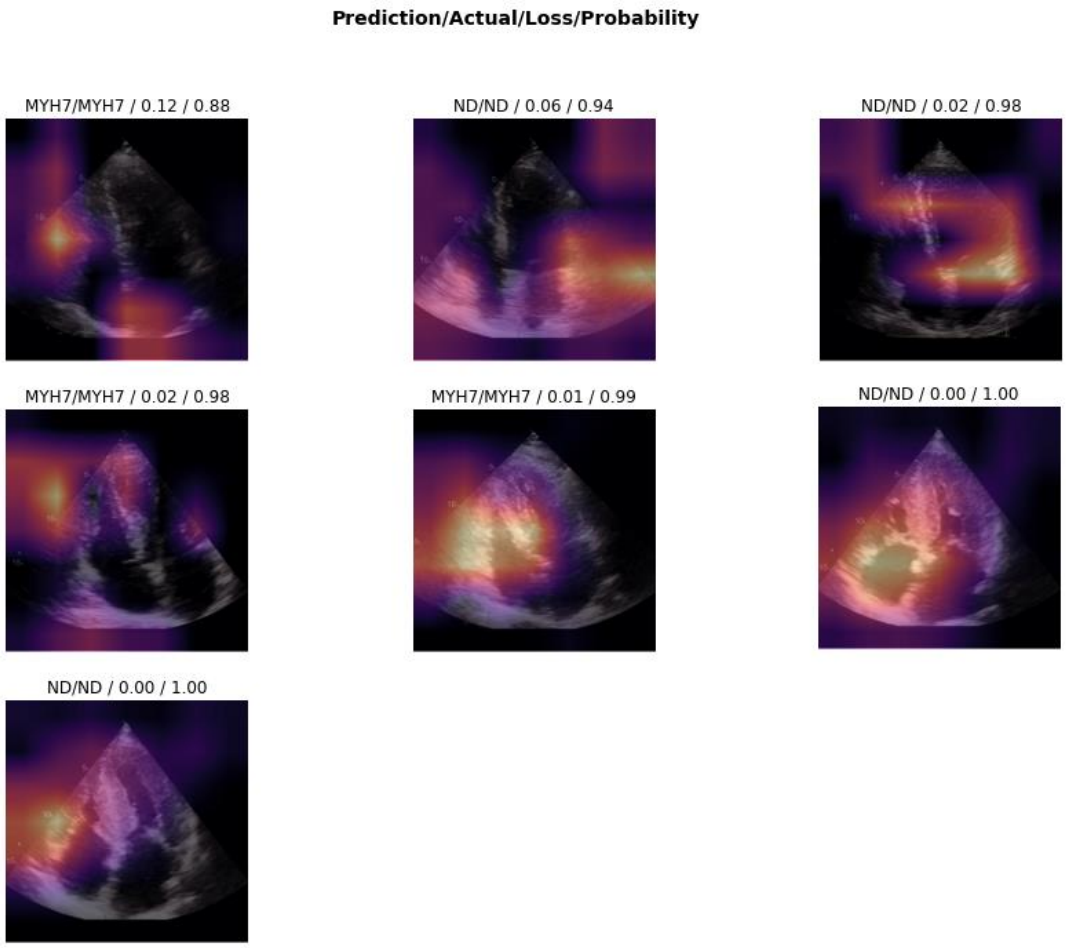

Figure 153. Discriminative areas for classification of a given image (*MYH7* vs. ND), apical 4-chamber view during ventricular diastole

4.1.5.1.6. Apical 4-chamber view, ventricular systole

A model for echocardiographic images classification (apical 4-chamber view during ventricular systole) of patients with mutation in *MYH7* gene and without detected mutation was created, with performance shown in Table 112 and Figure 154. Mitral apparatus was shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation (Figure 155).

Table 112. Performance of model for echocardiographic images classification (*MYH7* vs. ND), apical 4-chamber view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

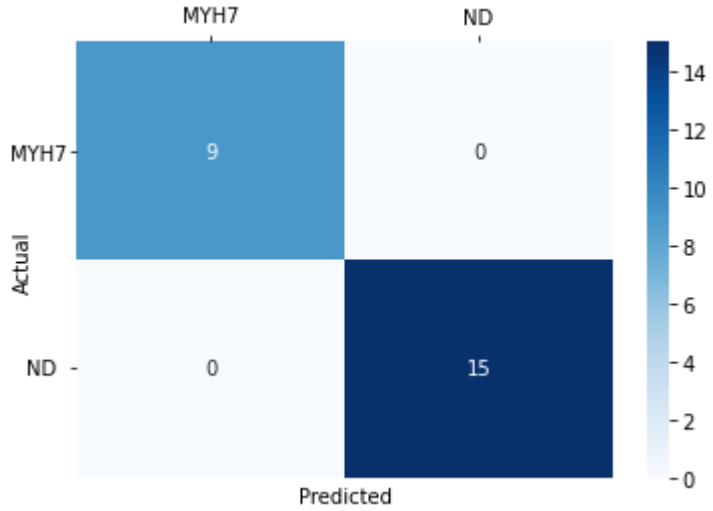

Figure 154. Confusion matrix, model for echocardiographic images classification (*MYH7* vs. ND), apical 4-chamber view during ventricular systole

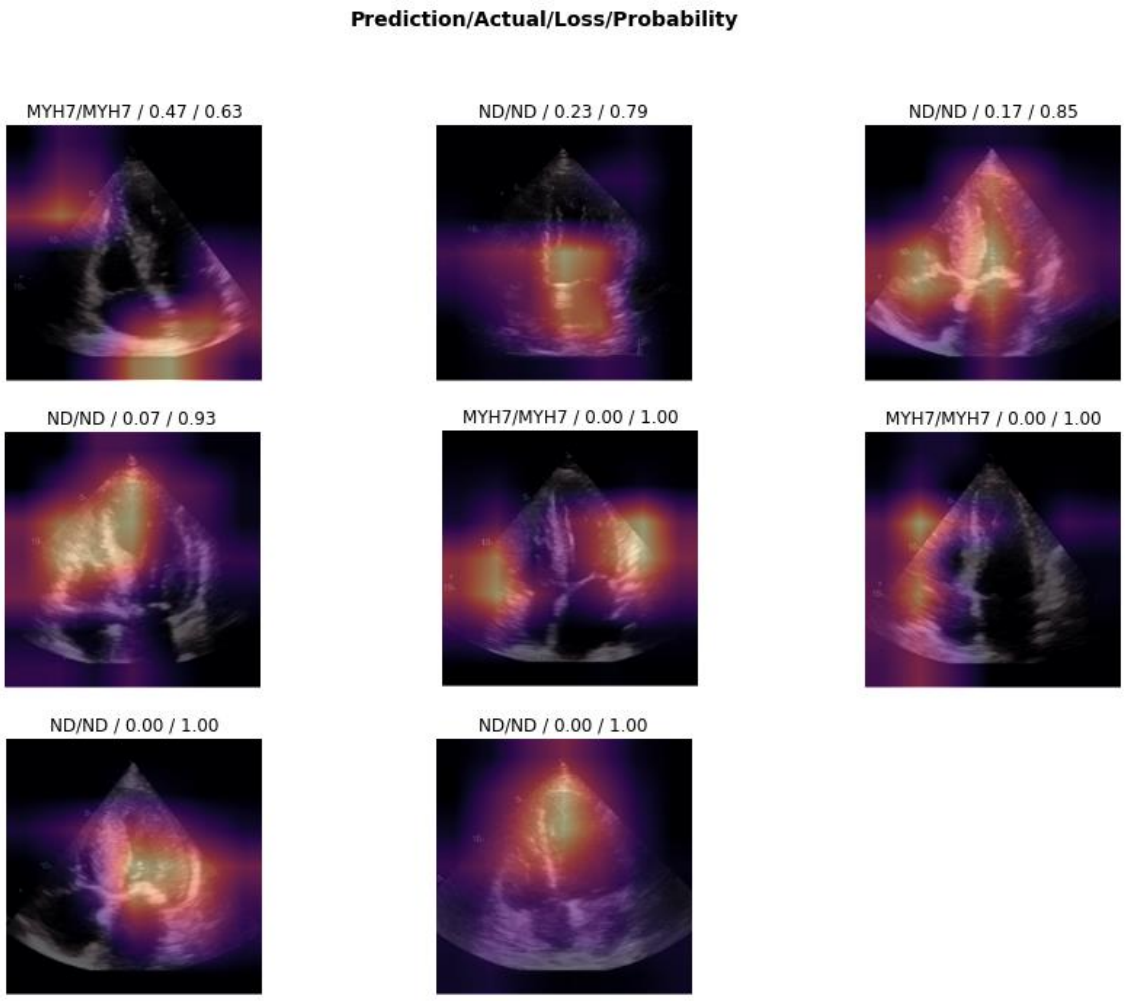

Figure 155. Discriminative areas for classification of a given image (*MYH7* vs. ND), apical 4-chamber view during ventricular systole

4.1.5.2. *TNNT2*-specific echocardiogram findings

4.1.5.2.1. Parasternal long axis view, ventricular diastole

A model for echocardiographic images classification (parasternal long axis view during ventricular diastole) of patients with mutation in *TNNT2* gene and without detected mutation was created, with performance shown in Table 113 and Figure 156. No unequivocal discriminative areas for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation were found in this view (Figure 157).

Table 113. Performance of model for echocardiographic images classification (*TNNT2* vs. ND), parasternal long axis view during ventricular diastole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

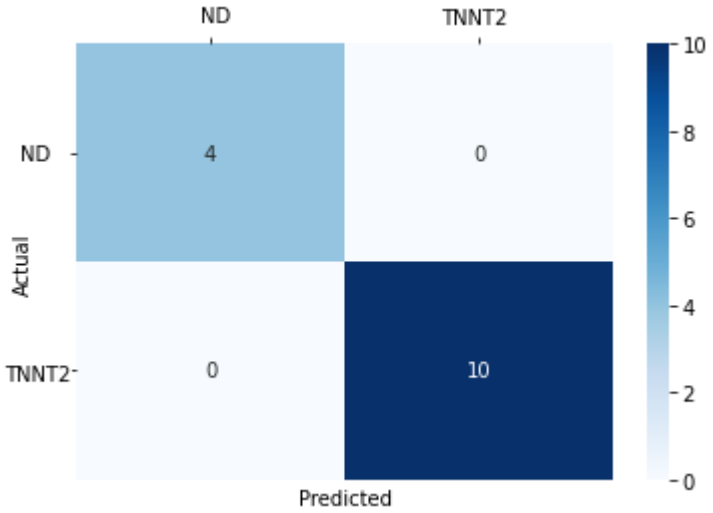

Figure 156. Confusion matrix, model for echocardiographic images classification (*TNNT2* vs. ND), parasternal long axis view during ventricular diastole

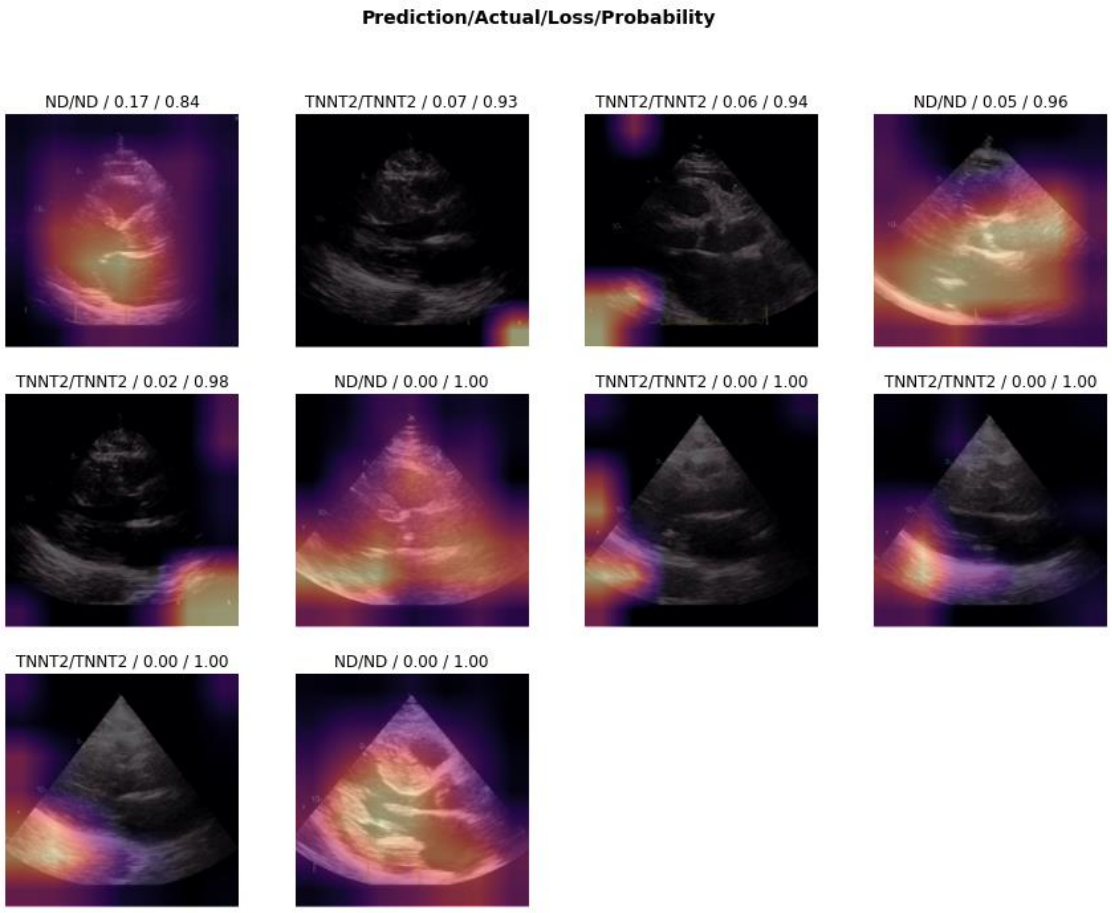

Figure 157. Discriminative areas for classification of a given image (*TNNT2* vs. ND), parasternal long axis view during ventricular diastole

4.1.5.2.2. Parasternal long axis view, ventricular systole

A model for echocardiographic images classification (parasternal long axis view during ventricular systole) of patients with mutation in *TNNT2* gene and without detected mutation was created, with performance shown in Table 114 and Figure 158. No unequivocal discriminative areas for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation were found in this view (Figure 159).

Table 114. Performance of model for echocardiographic images classification (*TNNT2* vs. ND), parasternal long axis view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

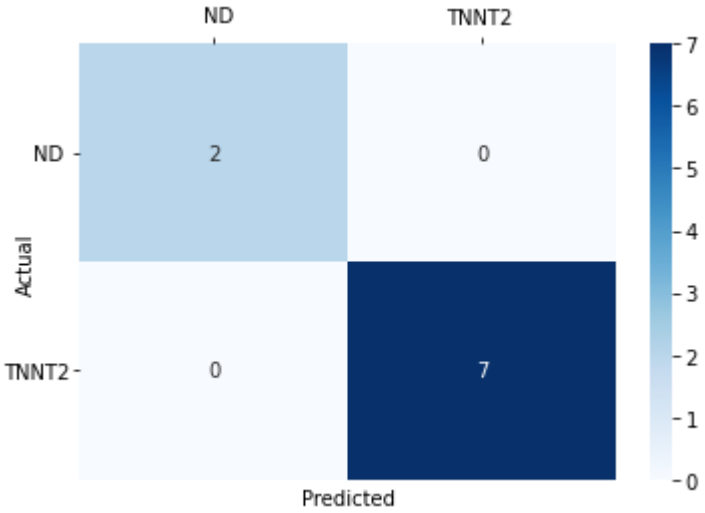

Figure 158. Confusion matrix, model for echocardiographic images classification (*TNNT2* vs. ND), parasternal long axis view during ventricular systole

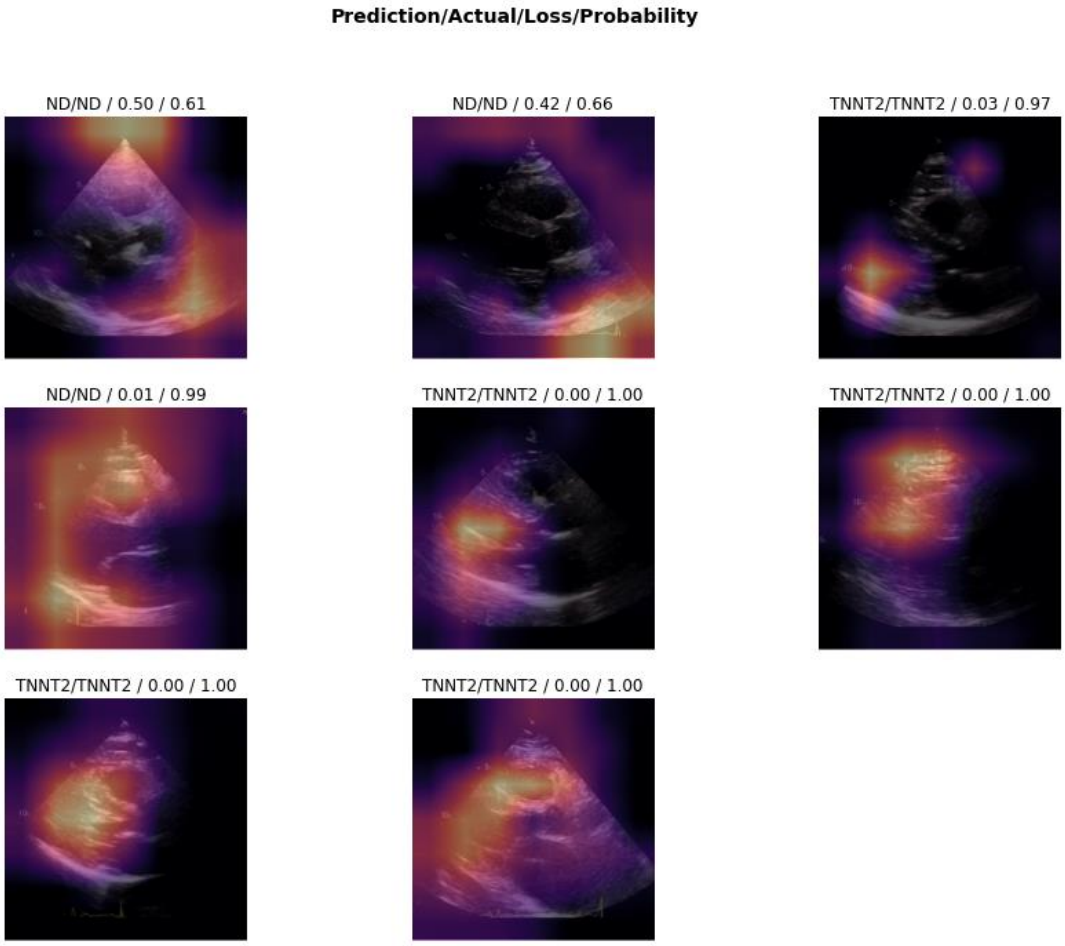

Figure 159. Discriminative areas for classification of a given image (*TNNT2* vs. ND), parasternal long axis view during ventricular systole

4.1.5.2.3. Apical 2-chamber view, ventricular diastole

A model for echocardiographic images classification (apical 2-chamber view during ventricular diastole) of patients with mutation in *TNNT2* gene and without detected mutation was created, with performance shown in Table 115 and Figure 160. Septum was shown to be discriminative for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation (Figure 161).

Table 115. Performance of model for echocardiographic images classification (*TNNT2* vs. ND), apical 2-chamber view during ventricular diastole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

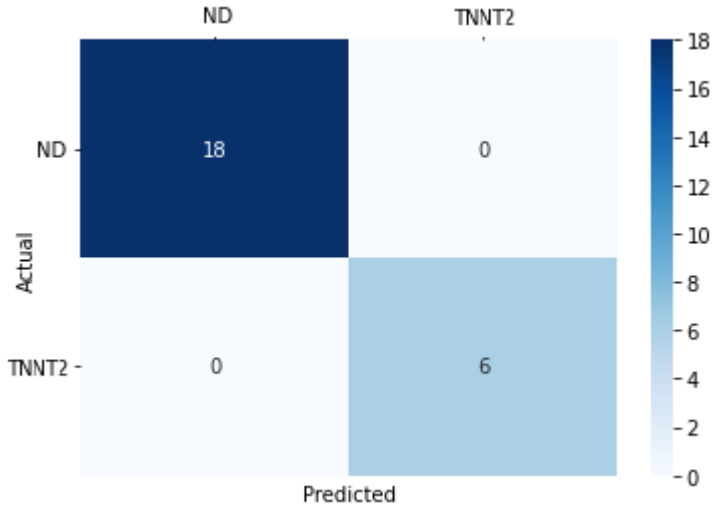

Figure 160. Confusion matrix, model for echocardiographic images classification (*TNNT2* vs. ND), apical 2-chamber view during ventricular diastole

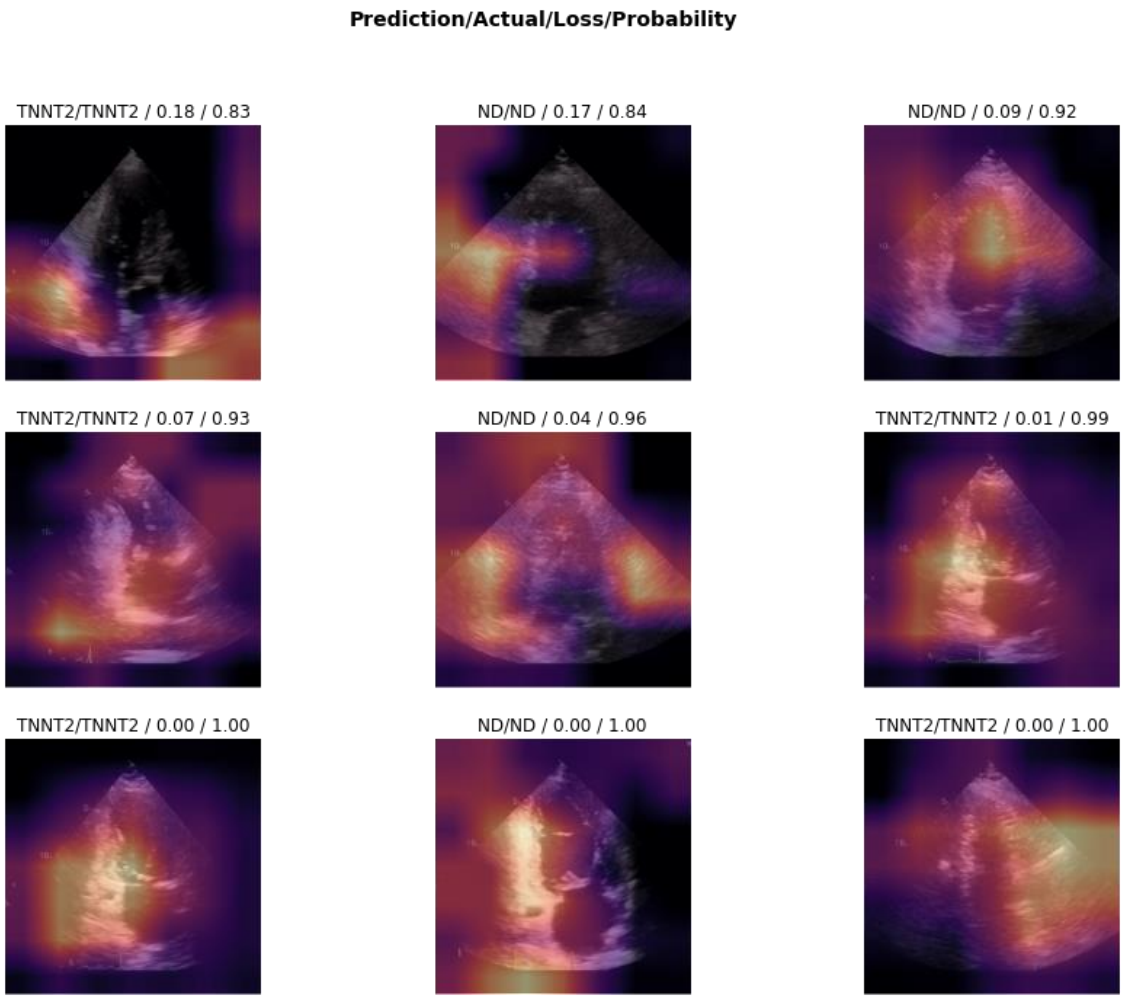

Figure 161. Discriminative areas for classification of a given image (*TNNT2* vs. ND), apical 2-chamber view during ventricular diastole

4.1.5.2.4. Apical 2-chamber view, ventricular systole

A model for echocardiographic images classification (apical 2-chamber view during ventricular systole) of patients with mutation in *TNNT2* gene and without detected mutation was created, with performance shown in Table 116 and Figure 162. Septum was shown to be discriminative for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation (Figure 163).

Table 116. Performance of model for echocardiographic images classification (*TNNT2* vs. ND), apical 2-chamber view during ventricular systole

Metric	Value
Accuracy	0.909
Precision	1.000
Recall	0.750

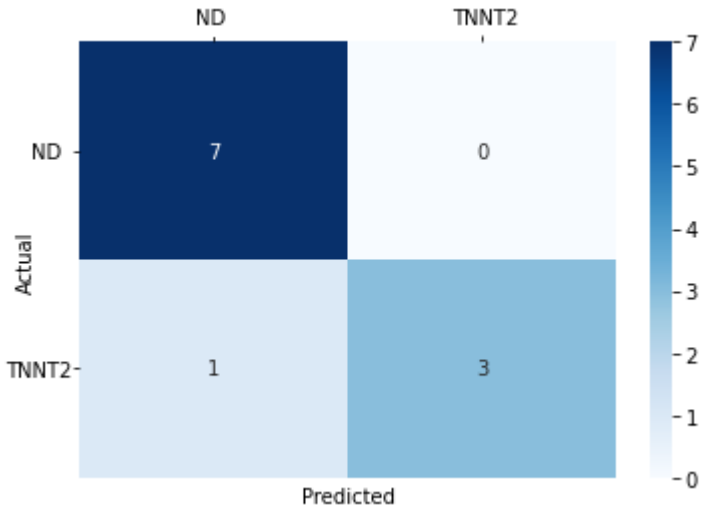

Figure 162. Confusion matrix, model for echocardiographic images classification (*TNNT2* vs. ND), apical 2-chamber view during ventricular systole

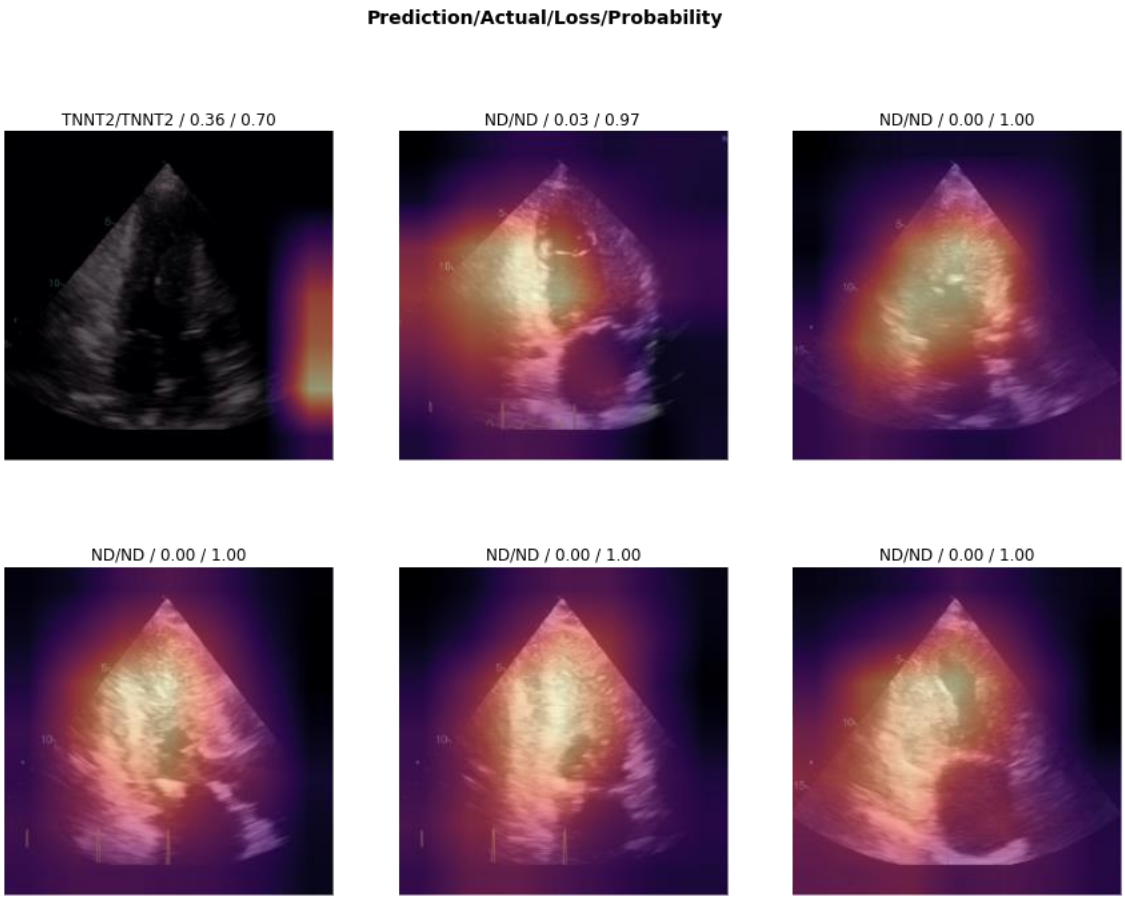

Figure 163. Discriminative areas for classification of a given image (*TNNT2* vs. ND), apical 2-chamber view during ventricular systole

4.1.5.2.5. Apical 4-chamber view, ventricular diastole

A model for echocardiographic images classification (apical 4-chamber view during ventricular diastole) of patients with mutation in *TNNT2* gene and without detected mutation was created, with performance shown in Table 117 and Figure 164. Septum was shown to be profoundly discriminative for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation (Figure 165).

Table 117. Performance of model for echocardiographic images classification (*TNNT2* vs. ND), apical 4-chamber view during ventricular diastole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

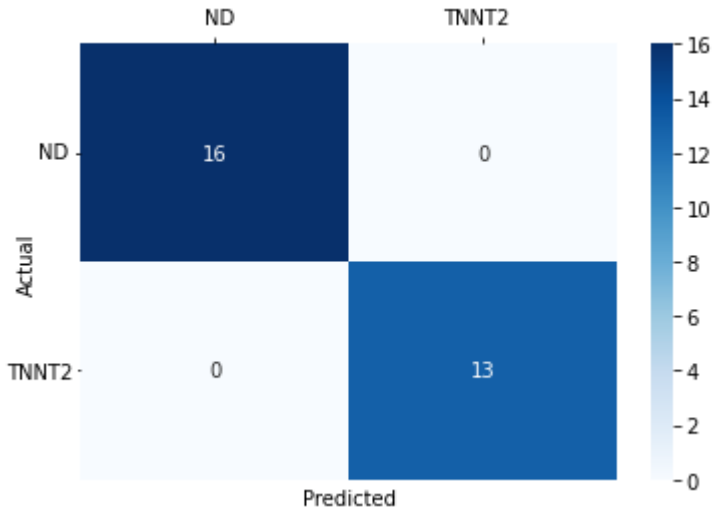

Figure 164. Confusion matrix, model for echocardiographic images classification (*TNNT2* vs. ND), apical 2-chamber view during ventricular systole

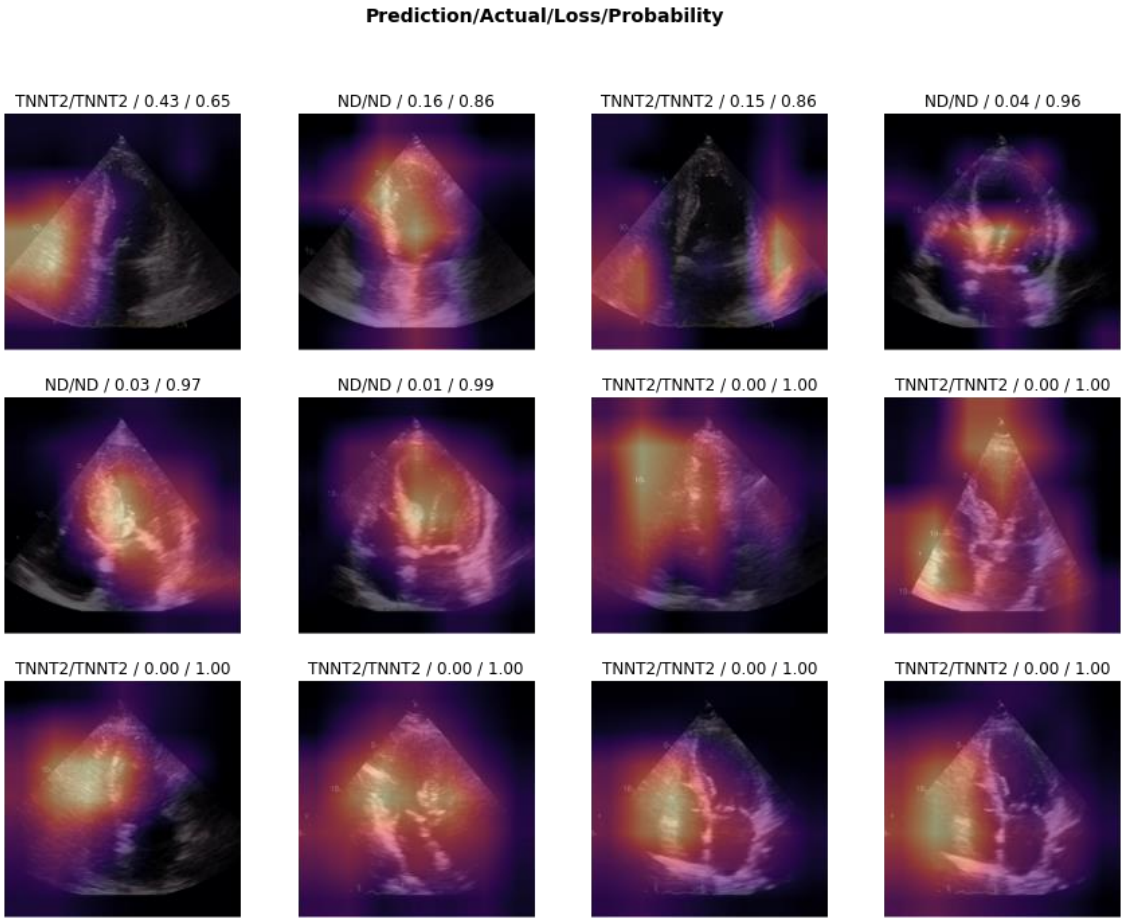

Figure 165. Discriminative areas for classification of a given image (*TNNT2* vs. ND), apical 2-chamber view during ventricular systole

4.1.5.2.6. Apical 4-chamber view, ventricular systole

A model for echocardiographic images classification (apical 4-chamber view during ventricular systole) of patients with mutation in *TNNT2* gene and without detected mutation was created, with performance shown in Table 118 and Figure 166. Septum and RV were shown to be discriminative for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation (Figure 167).

Table 118. Performance of model for echocardiographic images classification (*TNNT2* vs. ND), apical 4-chamber view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

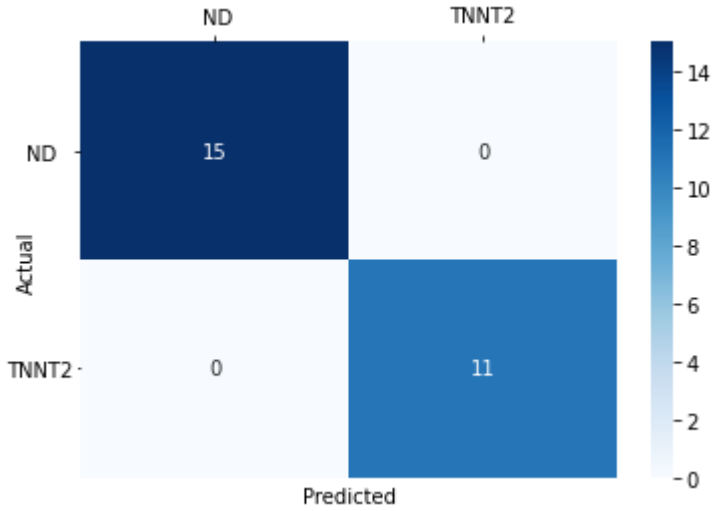

Figure 166. Confusion matrix, model for echocardiographic images classification (*TNNT2* vs. ND), apical 4-chamber view during ventricular systole

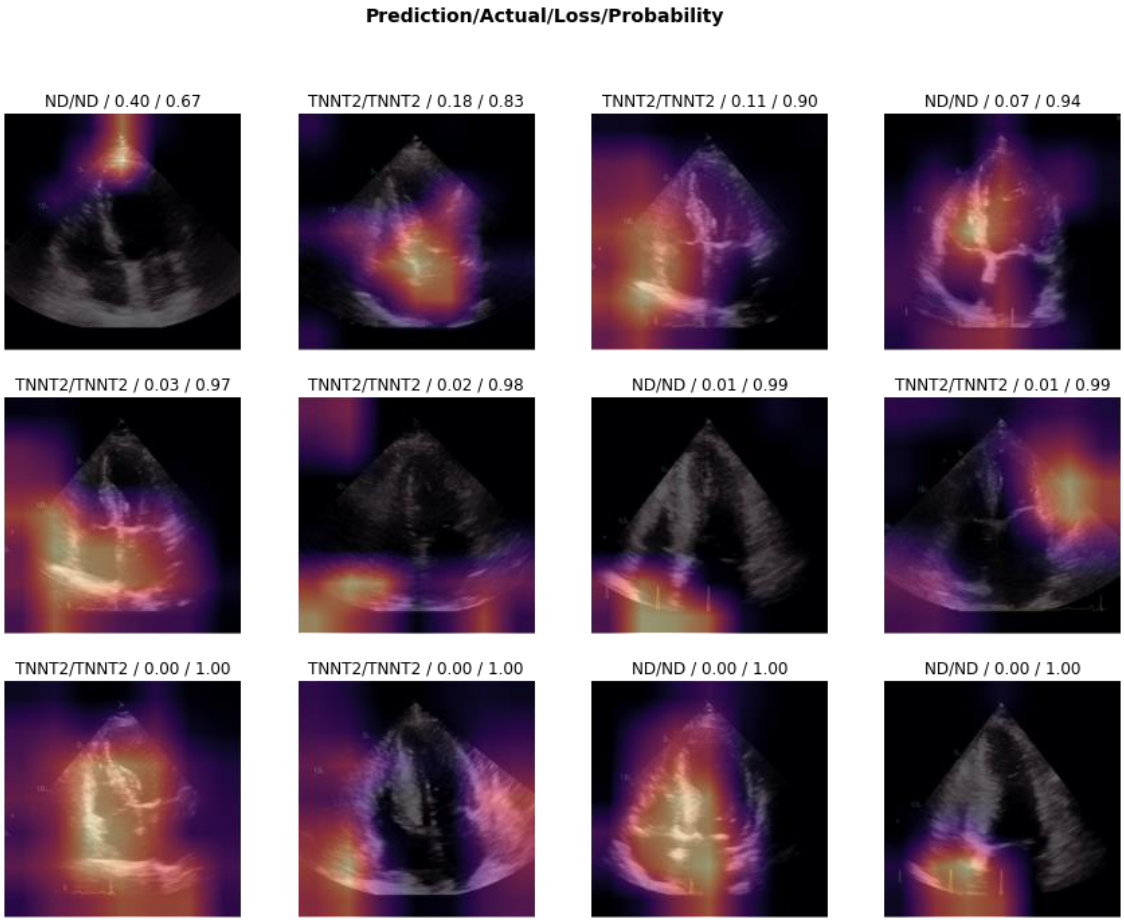

Figure 167. Discriminative areas for classification of a given image (*TNNT2* vs. ND), apical 4-chamber view during ventricular systole

4.1.5.3. *MYBPC3*-specific echocardiogram findings

4.1.5.3.1. Parasternal long axis view, ventricular diastole

A model for echocardiographic images classification (parasternal long axis view during ventricular diastole) of patients with mutation in *MYBPC3* gene and without detected mutation was created, with performance shown in Table 119 and Figure 168. No unequivocal discriminative areas for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation were found in this view (Figure 169).

Table 119. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), parasternal long axis view during ventricular diastole

Metric	Value
Accuracy	0.971
Precision	1.000
Recall	0.928

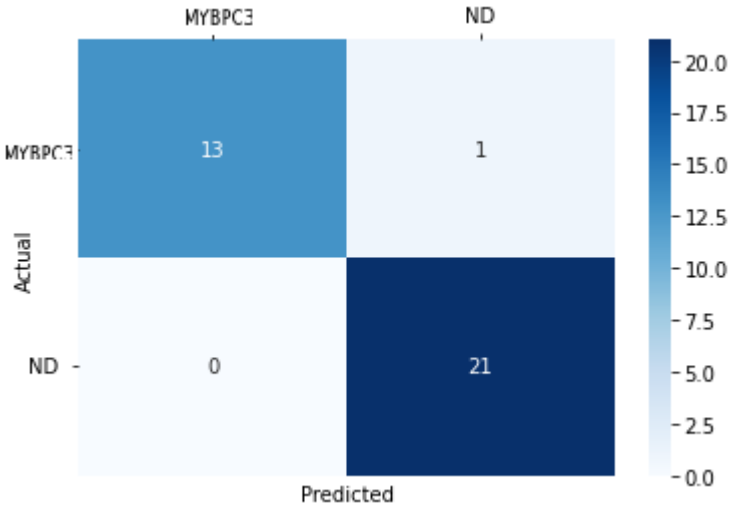

Figure 168. Confusion matrix, model for echocardiographic images classification (*MYBPC3* vs. ND), parasternal long axis view during ventricular diastole

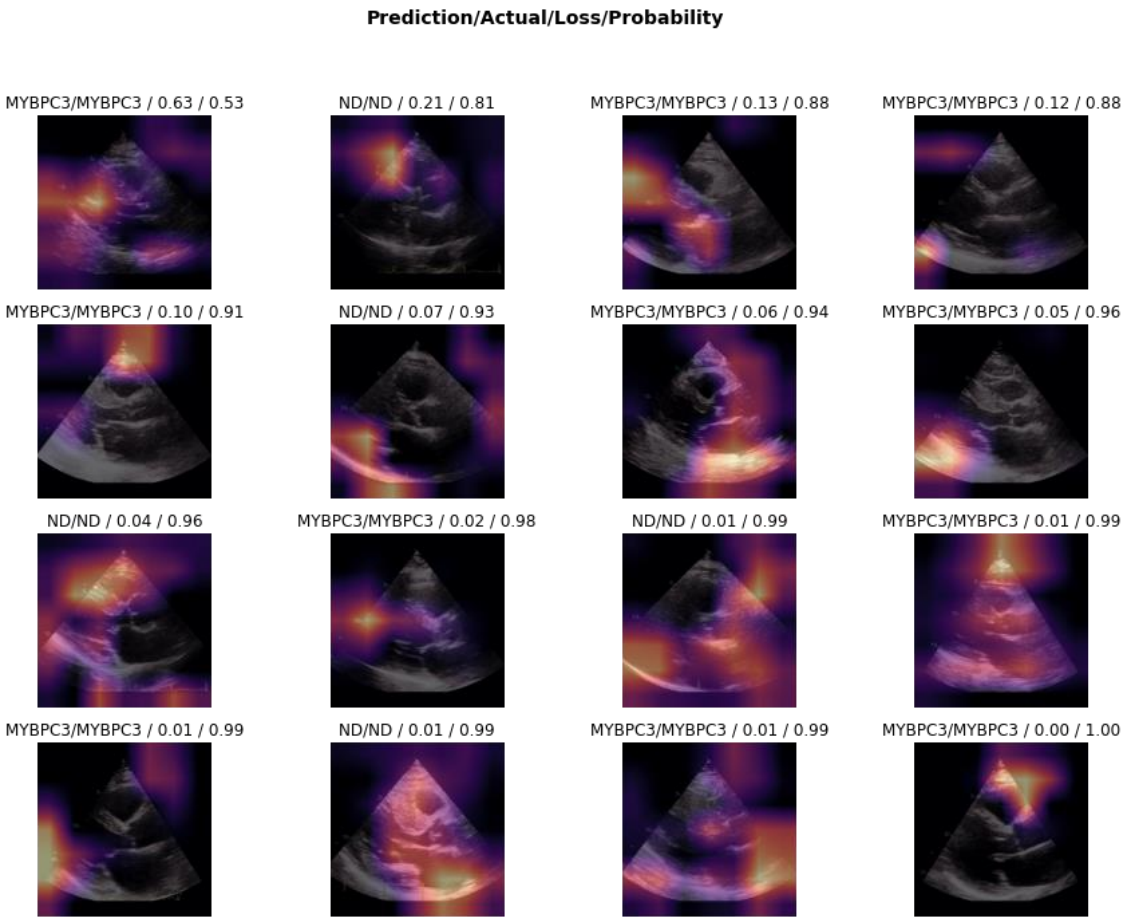

Figure 169. Discriminative areas for classification of a given image (*MYBPC3* vs. ND), parasternal long axis view during ventricular diastole

4.1.5.3.2. Parasternal long axis view, ventricular systole

A model for echocardiographic images classification (parasternal long axis view during ventricular systole) of patients with mutation in *MYBPC3* gene and without detected mutation was created, with performance shown in Table 120 and Figure 170. LV/chamber is potentially discriminative for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation (Figure 171).

Table 120. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), parasternal long axis view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

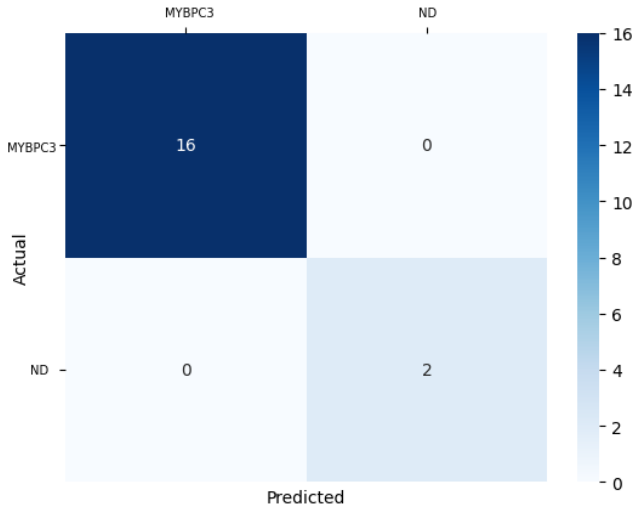

Figure 170. Confusion matrix, model for echocardiographic images classification (*MYBPC3* vs. ND), parasternal long axis view during ventricular systole

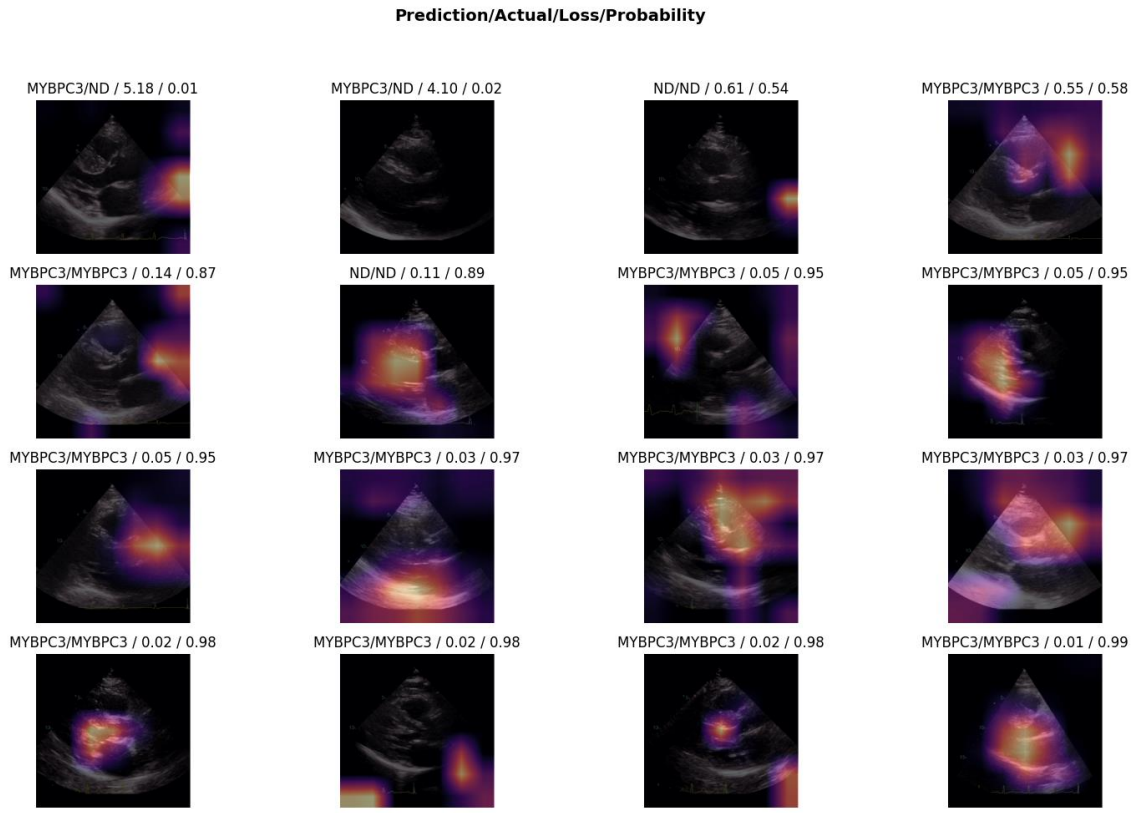

Figure 171. Discriminative areas for classification of a given image (*MYBPC3* vs. ND), parasternal long axis view during ventricular systole

4.1.5.3.3. Apical 2-chamber view, ventricular diastole

A model for echocardiographic images classification (apical 2-chamber view during ventricular diastole) of patients with mutation in *MYBPC3* gene and without detected mutation was created, with performance shown in Table 121 and Figure 172. LV was shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation (Figure 173).

Table 121. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), apical 2-chamber view during ventricular diastole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

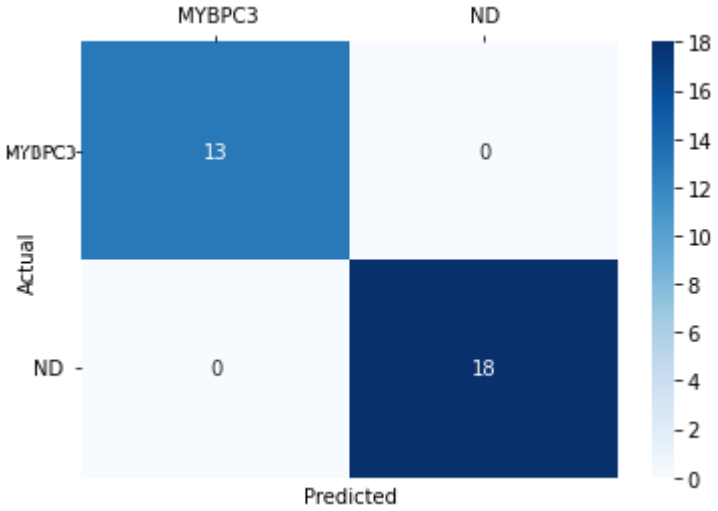

Figure 172. Confusion matrix, model for echocardiographic images classification (*MYBPC3* vs. ND), apical 2-chamber view during ventricular diastole

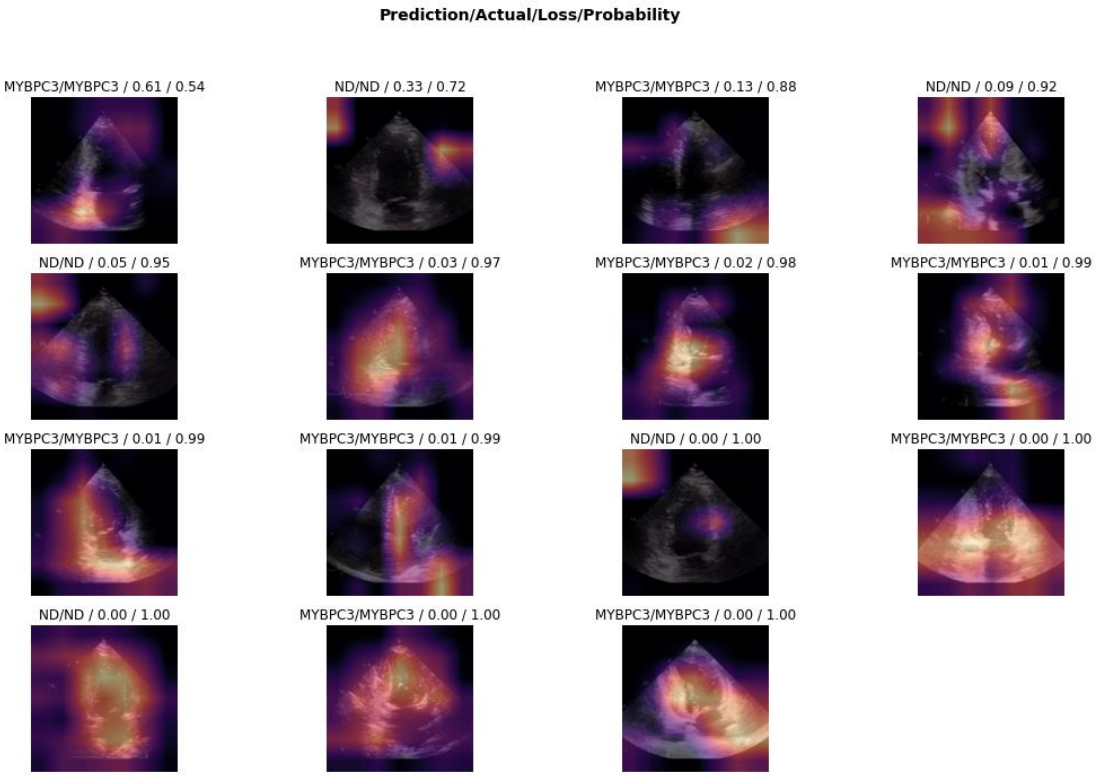

Figure 173. Discriminative areas for classification of a given image (*MYBPC3* vs. ND), apical 2-chamber view during ventricular diastole

4.1.5.3.4. Apical 2-chamber view, ventricular systole

A model for echocardiographic images classification (apical 2-chamber view during ventricular systole) of patients with mutation in *MYBPC3* gene and without detected mutation was created, with performance shown in Table 122 and Figure 174. No unequivocal discriminative areas for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation were found in this view (Figure 175).

Table 122. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), apical 2-chamber view during ventricular systole

Metric	Value
Accuracy	0.947
Precision	1.000
Recall	0.917

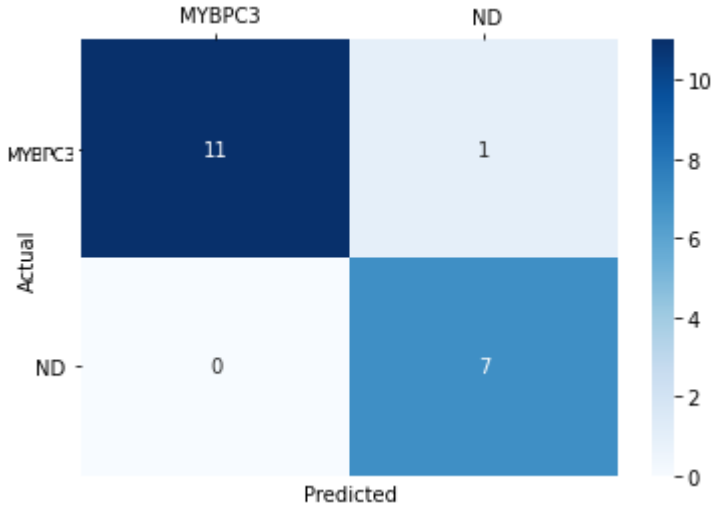

Figure 174. Confusion matrix, model for echocardiographic images classification (*MYBPC3* vs. ND), apical 2-chamber view during ventricular systole

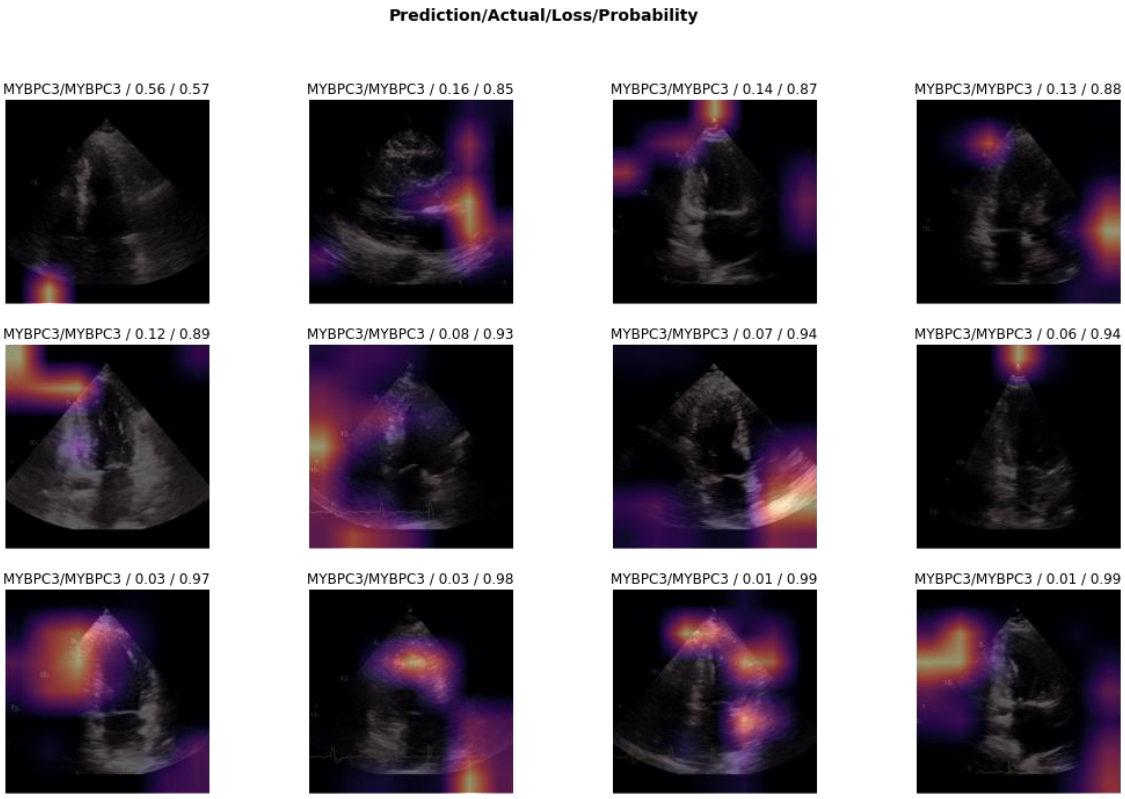

Figure 175. Discriminative areas for classification of a given image (*MYBPC3* vs. ND), apical 2-chamber view during ventricular systole

4.1.5.3.5. Apical 4-chamber view, ventricular diastole

A model for echocardiographic images classification (apical 4-chamber view during ventricular diastole) of patients with mutation in *MYBPC3* gene and without detected mutation was created, with performance shown in Table 123 and Figure 176. Septum was shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation (Figure 177).

Table 123. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), apical 4-chamber view during ventricular diastole

Metric	Value
Accuracy	0.969
Precision	0.941
Recall	1.000

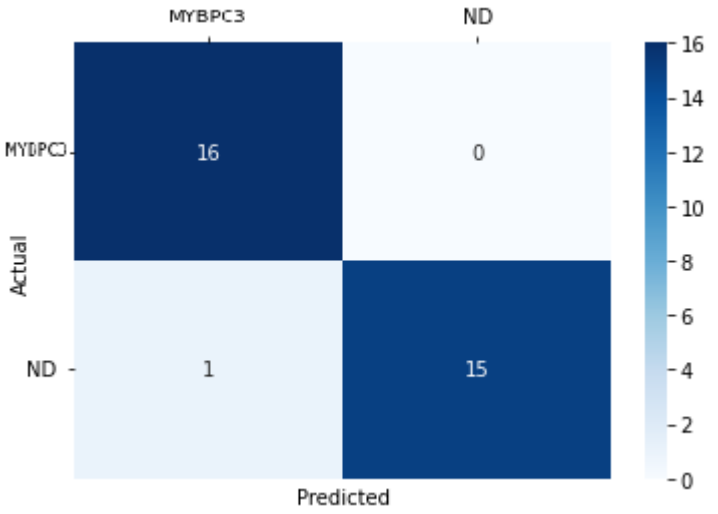

Figure 176. Confusion matrix, model for echocardiographic images classification (*MYBPC3* vs. ND), apical 4-chamber view during ventricular diastole

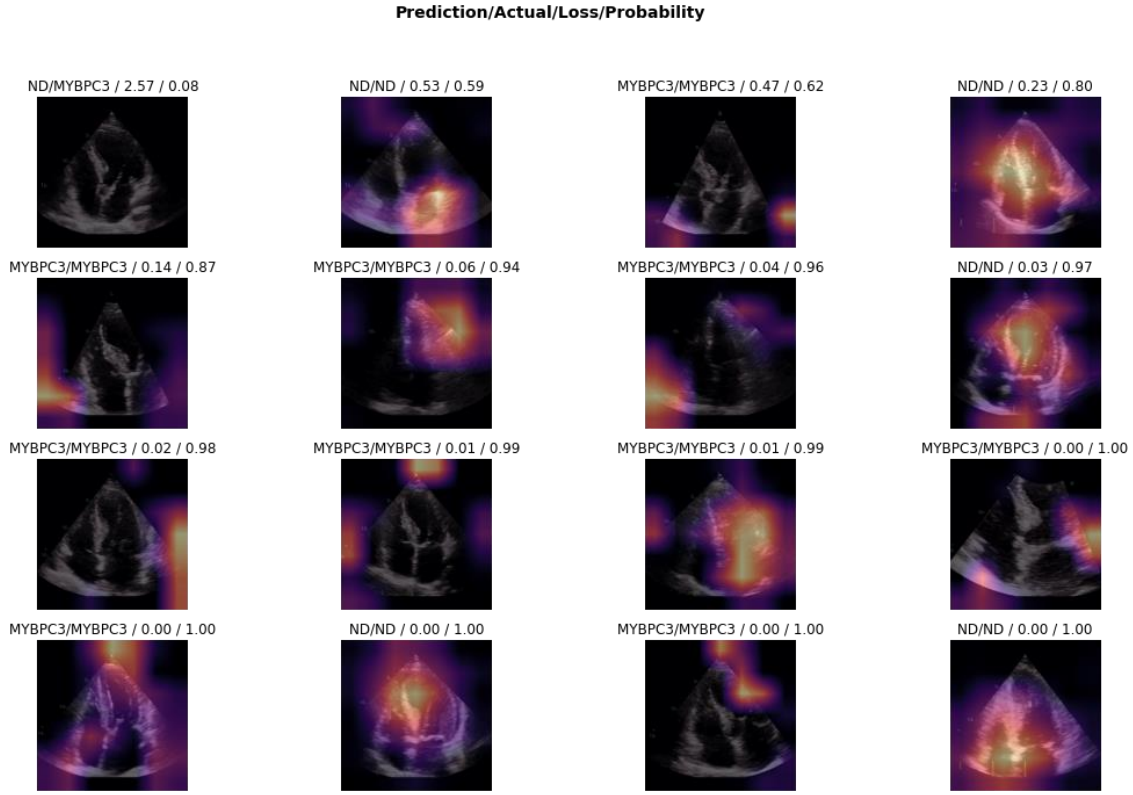

Figure 177. Discriminative areas for classification of a given image (*MYBPC3* vs. ND), apical 4-chamber view during ventricular diastole

4.1.5.3.6. Apical 4-chamber view, ventricular systole

A model for echocardiographic images classification (apical 4-chamber view during ventricular systole) of patients with mutation in *MYBPC3* gene and without detected mutation was created, with performance shown in Table 124 and Figure 178. Septum was shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation (Figure 179).

Table 124. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), apical 4-chamber view during ventricular systole

Metric	Value
Accuracy	1.000
Precision	1.000
Recall	1.000

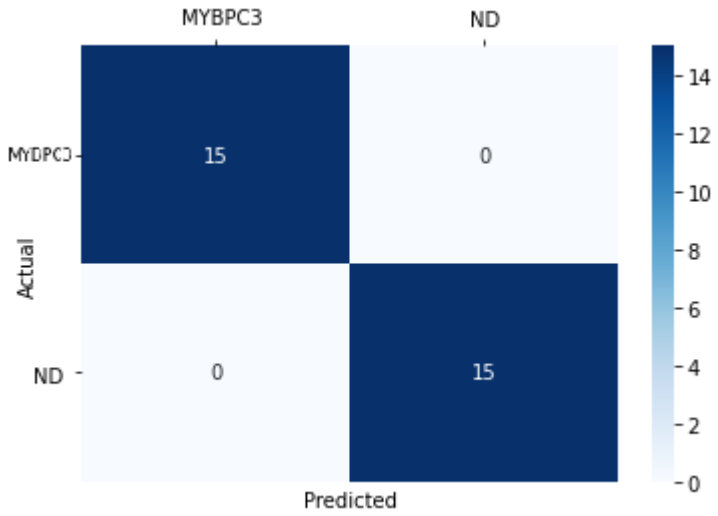

Figure 178. Confusion matrix, model for echocardiographic images classification (*MYBPC3* vs. ND), apical 4-chamber view during ventricular systole

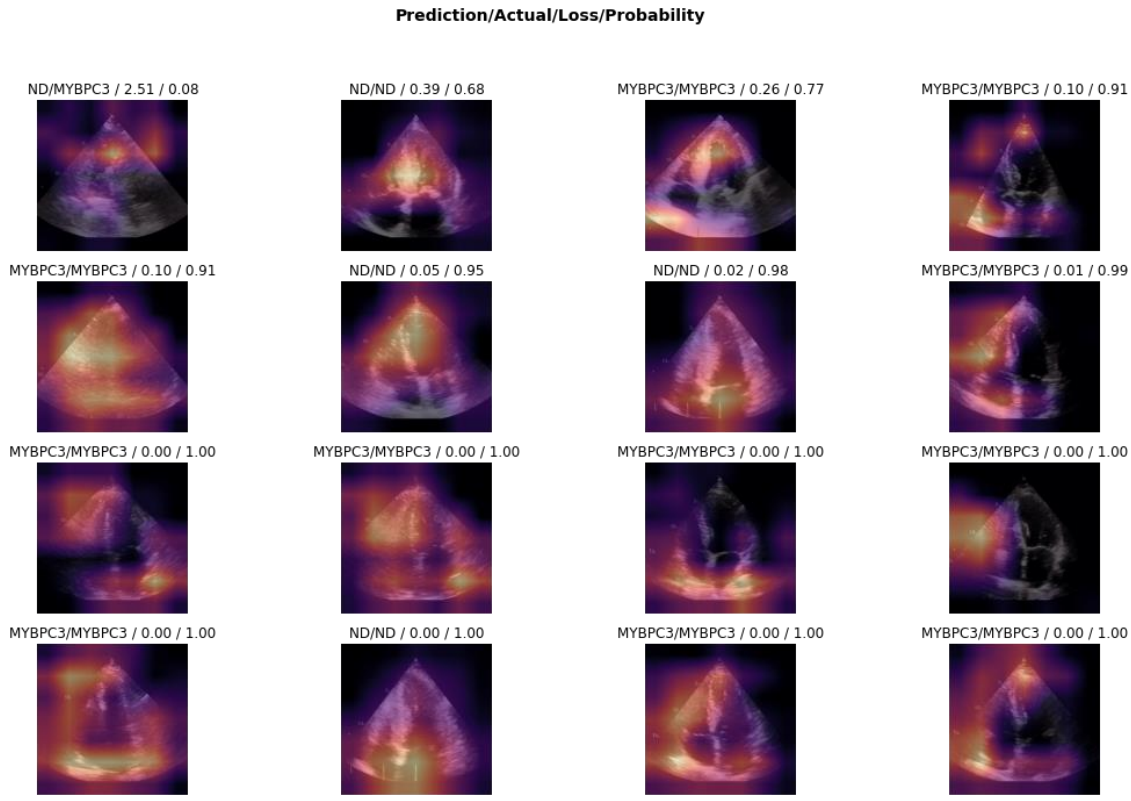

Figure 179. Performance of model for echocardiographic images classification (*MYBPC3* vs. ND), apical 4-chamber view during ventricular systole

4.1.5.4. Test set design

As a side result of accessory cardiac MR image classification, it has been shown that test sets containing MR images from patients who have other MR images in the training set produce too optimistic results: MR images of adjacent slices in the 2 different sets are so very much alike (Figure 180) that they produce data leakage effect. This is confirmed by the trend of training and test loss, as well as by areas that have mostly contributed to the classification decision — these were consistently the structures around the heart, and not the heart itself (Figure 181) (284).

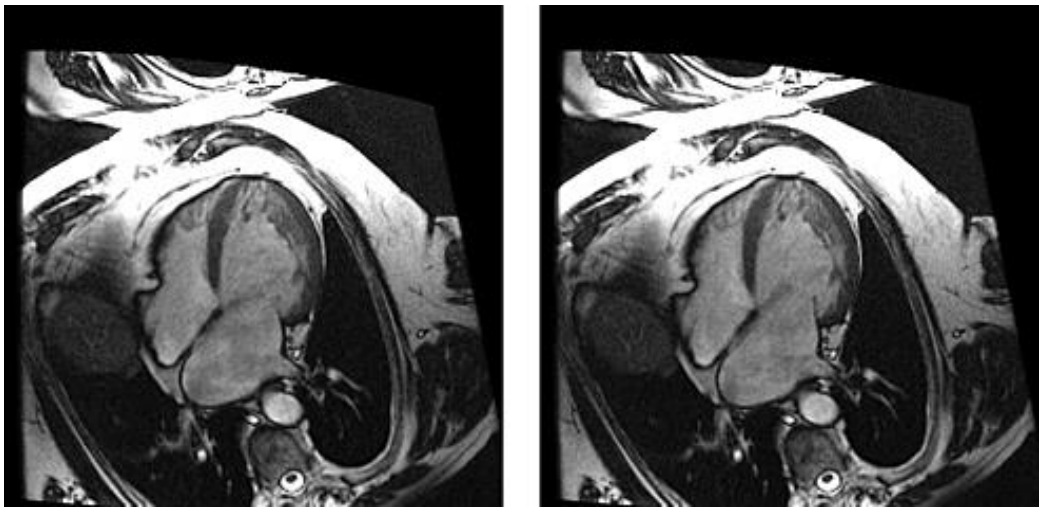

Figure 180. An example of neighboring slices. Image CAP_SCD0001401_MR__hrt_raw_20120813121634880_31 of patient 14 (left) and image CAP_SCD0001401_MR__hrt_raw_20120813121634905_32 of patient 14 (right) (284).

Figure 181. An example of top losses (284)

4.2. Molecular mechanisms of genotype-phenotype associations in HCM: Examination by machines using literature data

4.2.1. HCM Map: Interactive knowledge resource about genetic and molecular mechanisms of HCM

HCM Map is an interactive visual review of all principal HCM molecular pathways. It is a detailed visual and intuitive representation of molecular mechanisms of HCM (Figure 182) and a base for in silico analysis available at: https://silicofcm.eu/interactive-map/ (287).

HCM Map is the only publicly available knowledge resource on molecular mechanisms of HCM — it contains information about the entities represented and corresponding links to their descriptions in biomedical databases (Figure 183): PubMed IDs, Gene IDs, UniProt accession numbers, etc (287). The HCM Map interfaces with miRTarBase (224), CHEMBL (220), DrugBank (222), and CTDbase (223). HCM Map is both human- and computer-readable and can be exported into different formats for further analysis. Protein elements on the map that already have a 3D structure determined and are available in the Protein Data Bank can be further directly analyzed (287) via MOLeculAR structure annoTator (MolArt) (298) — integrated extension of the Minerva platform for the molecular structure annotation and visualization (Figure 184) (287).

Provided plugins enable onsite analysis. Focused molecular interaction network exploration can be performed using Map exploration plugin (286). Genes that have variants associated with HCM are marked by Disease-variant associations plugin (286). If there are pathway areas defined in the map, enrichment analysis can be performed using Gene set enrichment analysis (GSEA) plugin (286). Targets of drugs with identified adverse reactions are presented and filterable in the HCM map using Adverse drug reactions plugin (286).

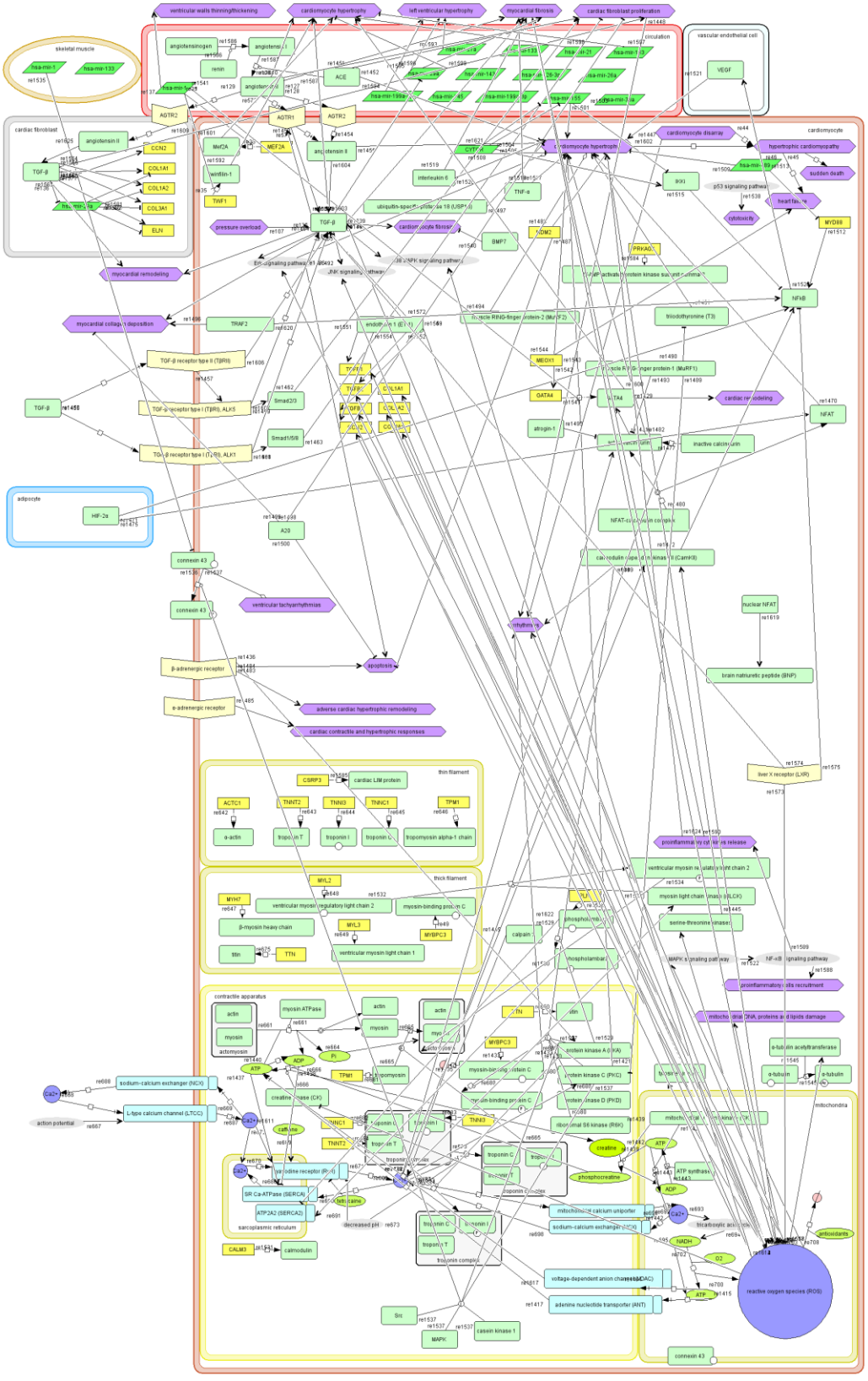

Figure 182. Visual appearance of entire HCM Map

Full name: myosin heavy chain 7

Symbol: MYH7

Former symbols: CMH1, MPD1

Synonyms: CMD1S

Muscle myosin is a hexameric protein containing 2 heavy chain subunits, 2 alkali light chain subunits, and 2 regulatory light chain subunits. This gene encodes the beta (or slow) heavy chain subunit of cardiac myosin. It is expressed predominantly in normal human ventricle. It is also expressed in skeletal muscle tissues rich in slow-twitch type I muscle fibers. Changes in the relative abundance of this protein and the alpha (or fast) heavy subunit of cardiac myosin correlate with the contractile velocity of cardiac muscle. Its expression is also altered during thyroid hormone depletion and hemodynamic overloading. Mutations in this gene are associated with familial hypertrophic cardiomyopathy, myosin storage myopathy, dilated cardiomyopathy, and Laing early-onset distal myopathy. [provided by RefSeq, Jul 2008]

Annotations:

Source: Ensembl

[1] Entrez Gene (4625)

[2] HGNC (7577)

[3] HGNC Symbol (7577)

Source: Entrez Gene

[4] Ensembl (ENSG00000092054)

[5] HGNC (7577)

Source: HGNC

[6] Ensembl (ENSG00000092054)

[7] Entrez Gene (4625)

[8] HGNC (7577)

[9] HGNC Symbol (MYH7)

[10] RefSeq (NM_000257)

[11] Uniprot (P12883)

Figure 183. Information about the entities represented in HCM Map (left) and corresponding links to their descriptions in biomedical databases (right) — example of element *MYH7* gene

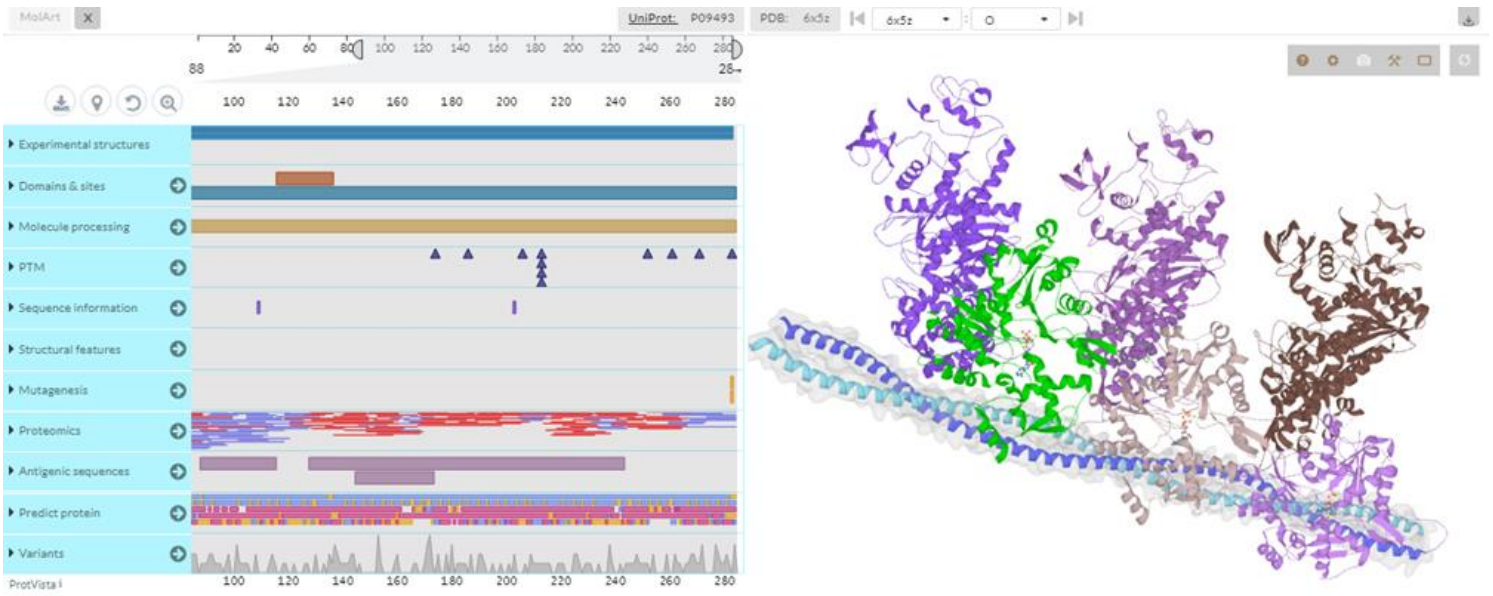

Figure 184. A MolArt exploration of an element of HCM Map — example of analysis of the tropomyosin alpha-1 chain element

4.2.2. Molecular mechanisms of genotype-phenotype associations in HCM collected and represented by machines

Four models of HCM molecular mechanisms were constructed automatically, using diverse computer-assisted approaches, and made publicly available (Table 125) (287).

Table 125. Automatically constructed models (287)

Model	Available at
INDRA-assembled PubMed HCM model	https://bit.ly/3blm2rB
INDRA-assembled PubMed+PathwayCommons HCM model	https://bit.ly/2OLxJQM
Truncated INDRA DB HCM model	https://bit.ly/2ZKypbD
INDRA DB HCM model	https://bit.ly/3upHsga

In subsequent network analysis, nodes' centrality scores and the most important nodes were determined. The top 10% elements ranked by several centrality measures were visualized for each network (Table 126). General agreement regarding the most important nodes was reached only with respect to one element (calcium). Each network was also displayed as a packed concentric ring sorted by k-shell and a gradient of the nodes' color used based on k-shell (Figures 185-188) (287).

Table 126. Elements ranked as top 10% by centrality measures for each network (287)

Model	Link to folder with top 10% elements for each of centrality measures for the model
INDRA-assembled PubMed HCM model	https://bit.ly/3k6Dmon
INDRA-assembled PubMed+PathwayCommons HCM model	https://bit.ly/3s9Wc0x
Truncated INDRA DB HCM model	https://bit.ly/3s6uqSL
INDRA DB model	https://bit.ly/37Kqlfc

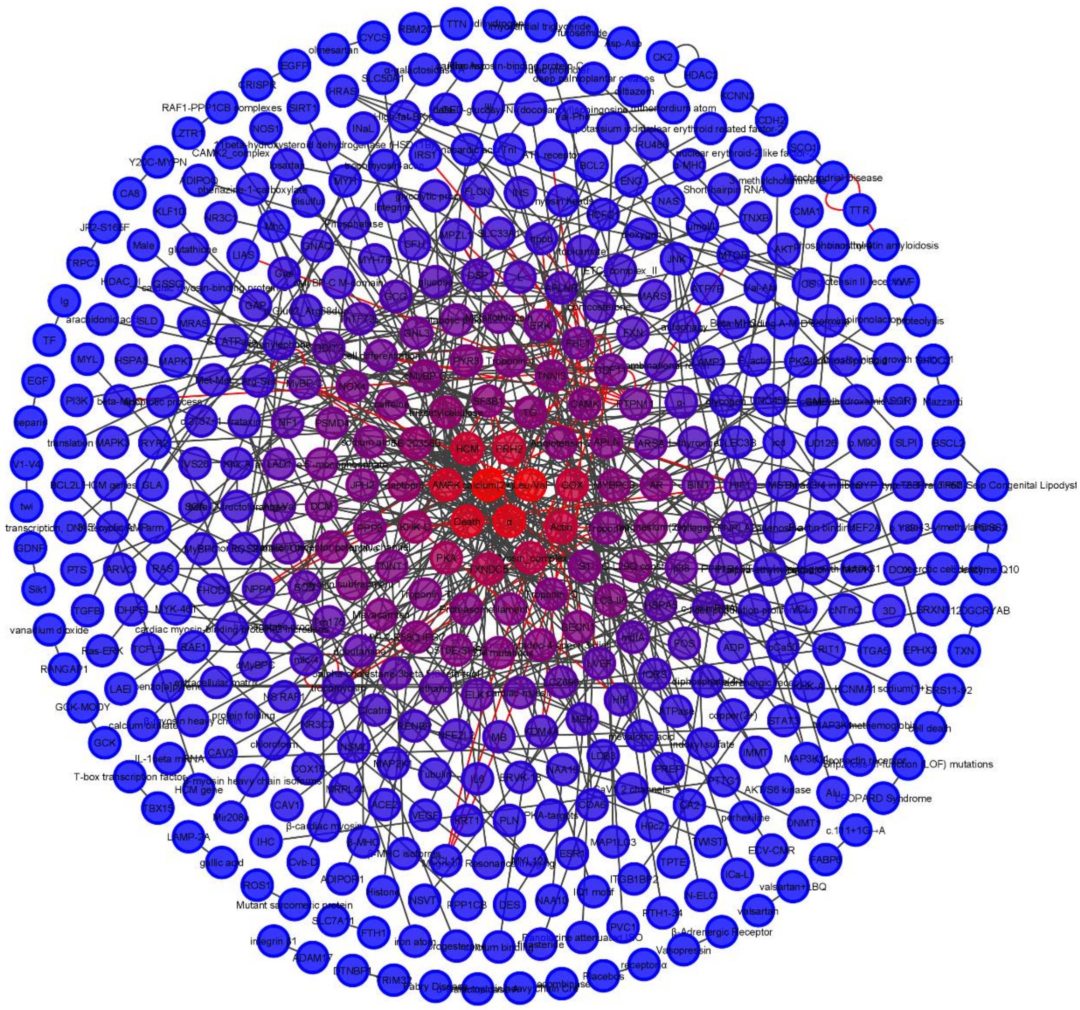

Figure 185. INDRA-assembled PubMed HCM model represented as a packed concentric ring sorted by k-shell (287)

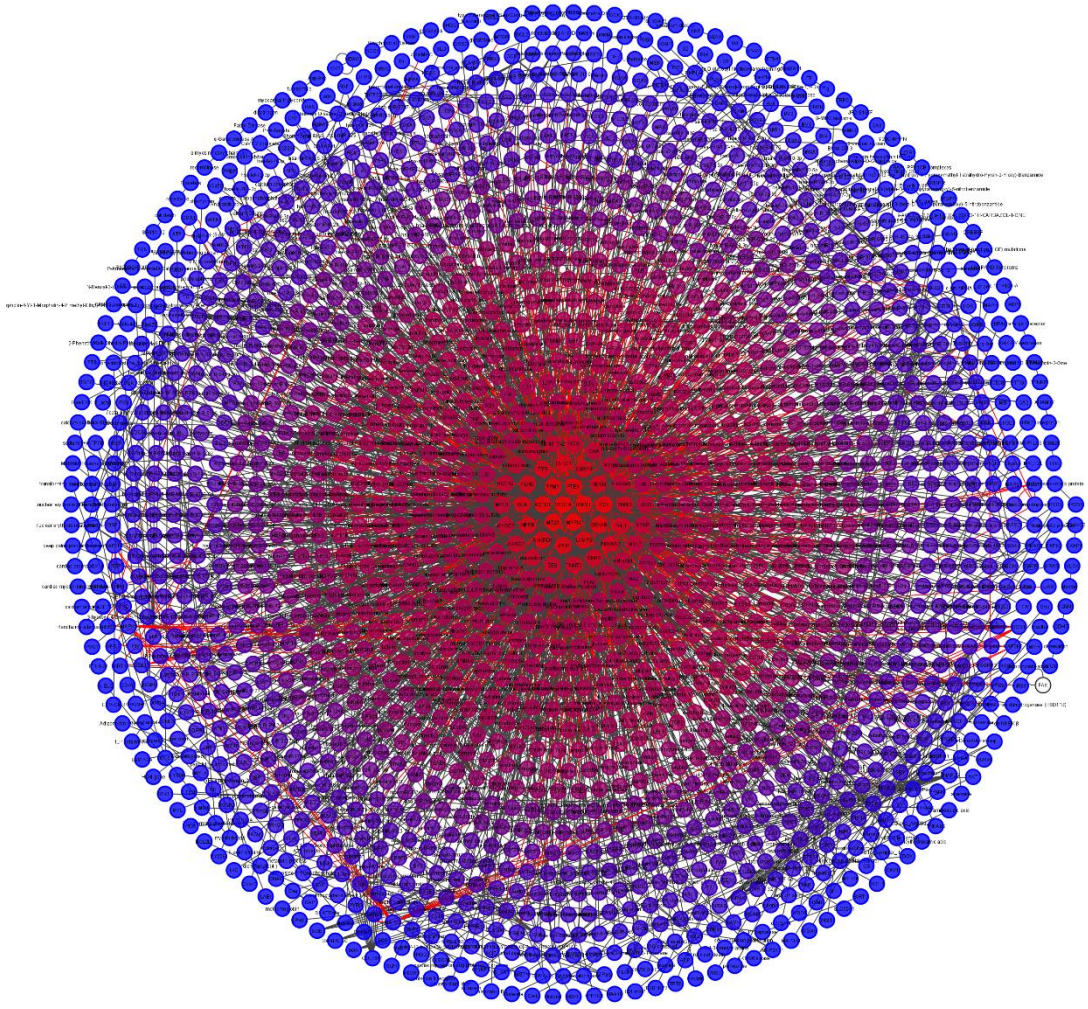

Figure 186. INDRA-assembled PubMed+PathwayCommons HCM model represented as a packed concentric ring sorted by k-shell (287)

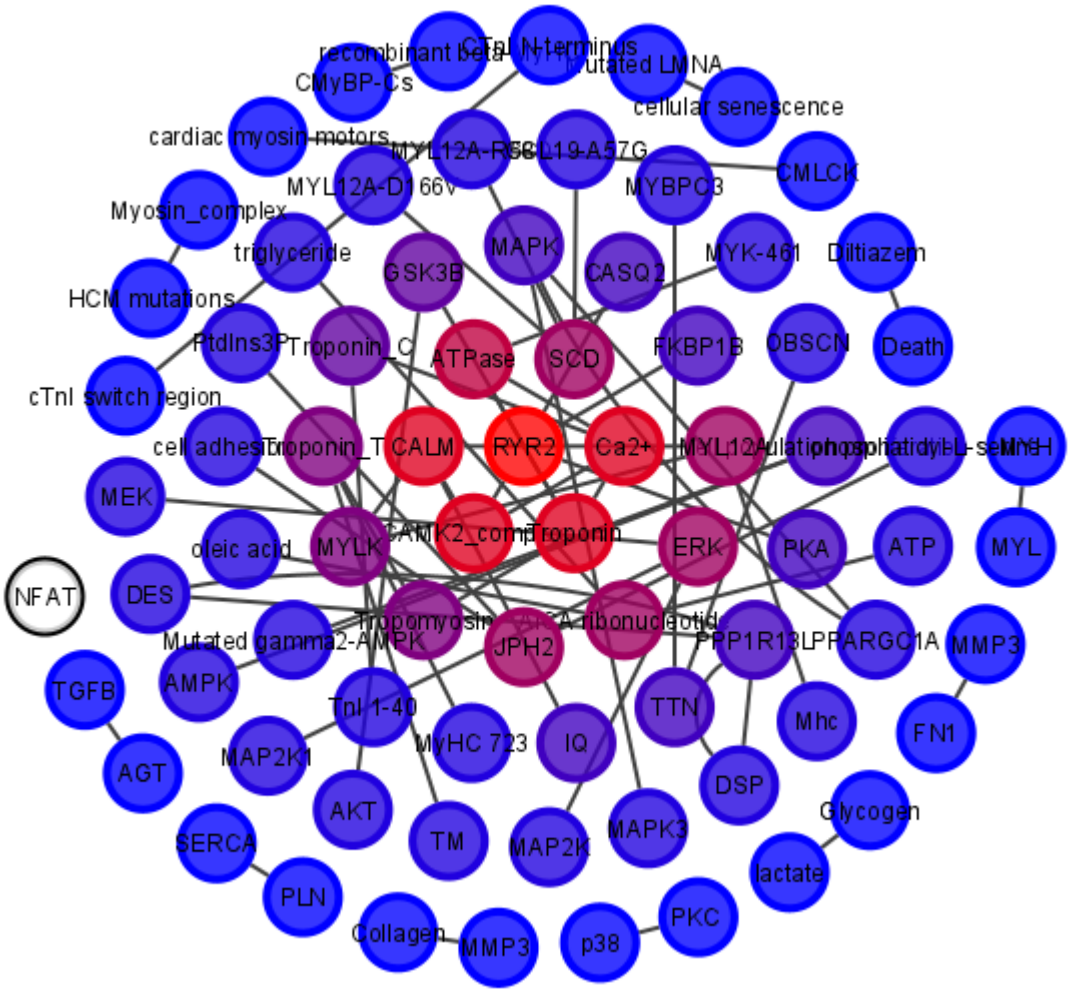

Figure 187. Truncated INDRA DB HCM model represented as a packed concentric ring sorted by k-shell (287)

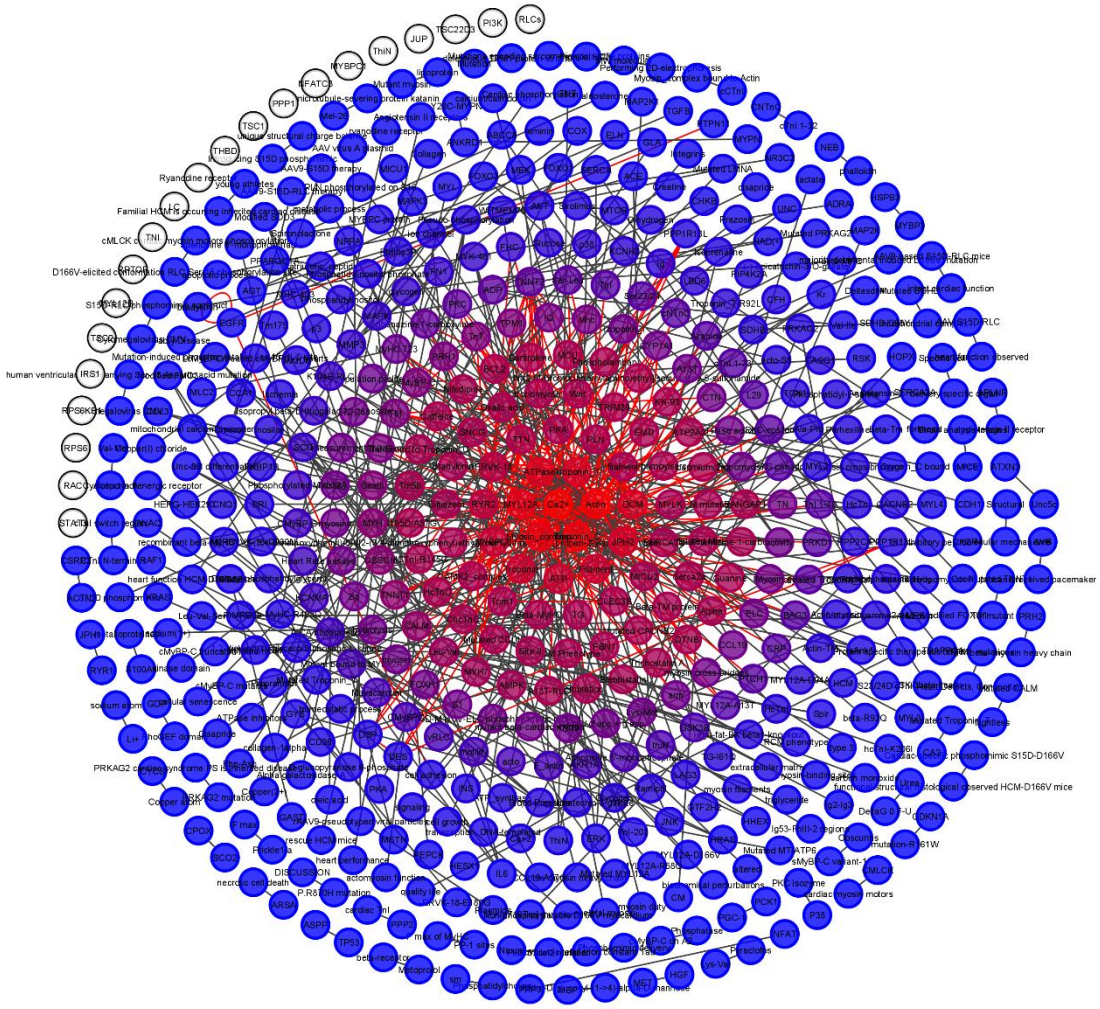

Figure 188. INDRA DB HCM model represented as a packed concentric ring sorted by k-shell (287)

Since computer-assisted approaches introduce noticeable level of noise into the models, models with lower levels of noise were created as well (Table 127) (287).

Table 127. Models with reduced level of noise (287)

Model	Models with reduced level of noise
INDRA-assembled PubMed HCM model	https://bit.ly/3bBKFkf
INDRA-assembled PubMed+PathwayCommons HCM model	https://bit.ly/3s6ALO3
Truncated INDRA DB HCM model	https://bit.ly/3k9iH2T
INDRA DB model	https://bit.ly/3pFqo1Y

4.2.4. Genetic and molecular mechanisms of different clinical presentations of HCM collected and represented by machines

Genetic and molecular HCM mechanisms were obtained from 230,072 articles on HCM and 19 HCM clinical presentations in the format of 182,167 INDRA statements (presentments of molecular mechanisms containing molecular subject, object, and their interaction) (Table 128) (92).

Table 128. Number of articles about HCM and its clinical presentations automatically read and number of INDRA statements extracted (92)

Pathophysiologic entity	Number of articles read	Number of INDRA statements extracted
hypertrophic cardiomyopathy	8,111	7,559
sudden cardiac death	10,060	6,770
HF	111,565	98,397
atrial fibrillation	54,117	25,842
major adverse cardiovascular events	4,700	1,713
rehospitalization	3,073	656
left ventricular outflow tract obstruction	1,023	177
myocardial ischemia	19,637	19,078
cardiac remodeling	4,572	5,432
myocardial fibrosis	3,634	4,978
myofibrillar disarray	51	356
cardiomyocyte disarray	11	22
impaired myocardial relaxation	31	33
impaired cardiac relaxation	12	28
cardiomyocyte hypertrophy	1,337	2,500
myocardial remodeling	967	1,500
myocardial hypercontractility	3	3
coronary microvascular dysfunction	569	522
myocardial stiffness	257	500
diastolic dysfunction	6,342	6,101

4.2.3.1. Network analysis

4.2.3.1.1. Intersection networks

Mutual genetic and molecular mechanisms of HCM and its clinical presentations are presented as intersection networks (Table 129) (92).

Table 129. Shared genetic and molecular mechanisms of HCM and its clinical presentations (92)

Pathophysiologic entities	Link to the network representing shared molecular mechanisms
hypertrophic cardiomyopathy, atrial fibrillation	https://bit.ly/3d31kyT
hypertrophic cardiomyopathy, cardiomyocyte disarray	https://bit.ly/3wJsmmy
hypertrophic cardiomyopathy, coronary microvascular dysfunction	https://bit.ly/31Xh2VN
hypertrophic cardiomyopathy, cardiomyocyte hypertrophy	https://bit.ly/39Yn90x
hypertrophic cardiomyopathy, cardiac remodeling	https://bit.ly/31ZG3Qh
hypertrophic cardiomyopathy, diastolic dysfunction	https://bit.ly/3wHxRCn
hypertrophic cardiomyopathy, heart failure	https://bit.ly/322UjI9
hypertrophic cardiomyopathy, impaired myocardial relaxation	https://bit.ly/322sU94
hypertrophic cardiomyopathy, left ventricular outflow tract obstruction	https://bit.ly/3dN8G8R
hypertrophic cardiomyopathy, major adverse cardiovascular events	https://bit.ly/3mvZZE1
hypertrophic cardiomyopathy, myocardial fibrosis	https://bit.ly/3fZX3hC
hypertrophic cardiomyopathy, myofibrillar disarray	https://bit.ly/2PRnPOz
hypertrophic cardiomyopathy, myocardial ischemia	https://bit.ly/31YlC6a
hypertrophic cardiomyopathy, myocardial remodeling	https://bit.ly/2Q8dDkD
hypertrophic cardiomyopathy, myocardial stiffness	https://bit.ly/3mxmecq
hypertrophic cardiomyopathy, rehospitalization	https://bit.ly/3myyx8t
hypertrophic cardiomyopathy, sudden cardiac death	https://bit.ly/3wIN5ao

The intersection of genetic and molecular interactions representing HCM and myocardial hypercontractility contains no molecular interactions. The intersection of HCM and impaired cardiac relaxation involves only phosphorylation of SMAD Family Member 2 (SMAD2) and therefore could not be represented as a network (92).

4.2.3.1.2. The most important genetic and molecular elements nodes of intersection networks by k-shell

All generated intersection networks of genetic and molecular elements of HCM and its clinical presentations are presented as packed concentric rings sorted by the most important nodes (by k-shell) (Figures 189-198). Gradient to the color of nodes corresponds to k-shell: the redder and closer to the center node is — the more important it is (92).

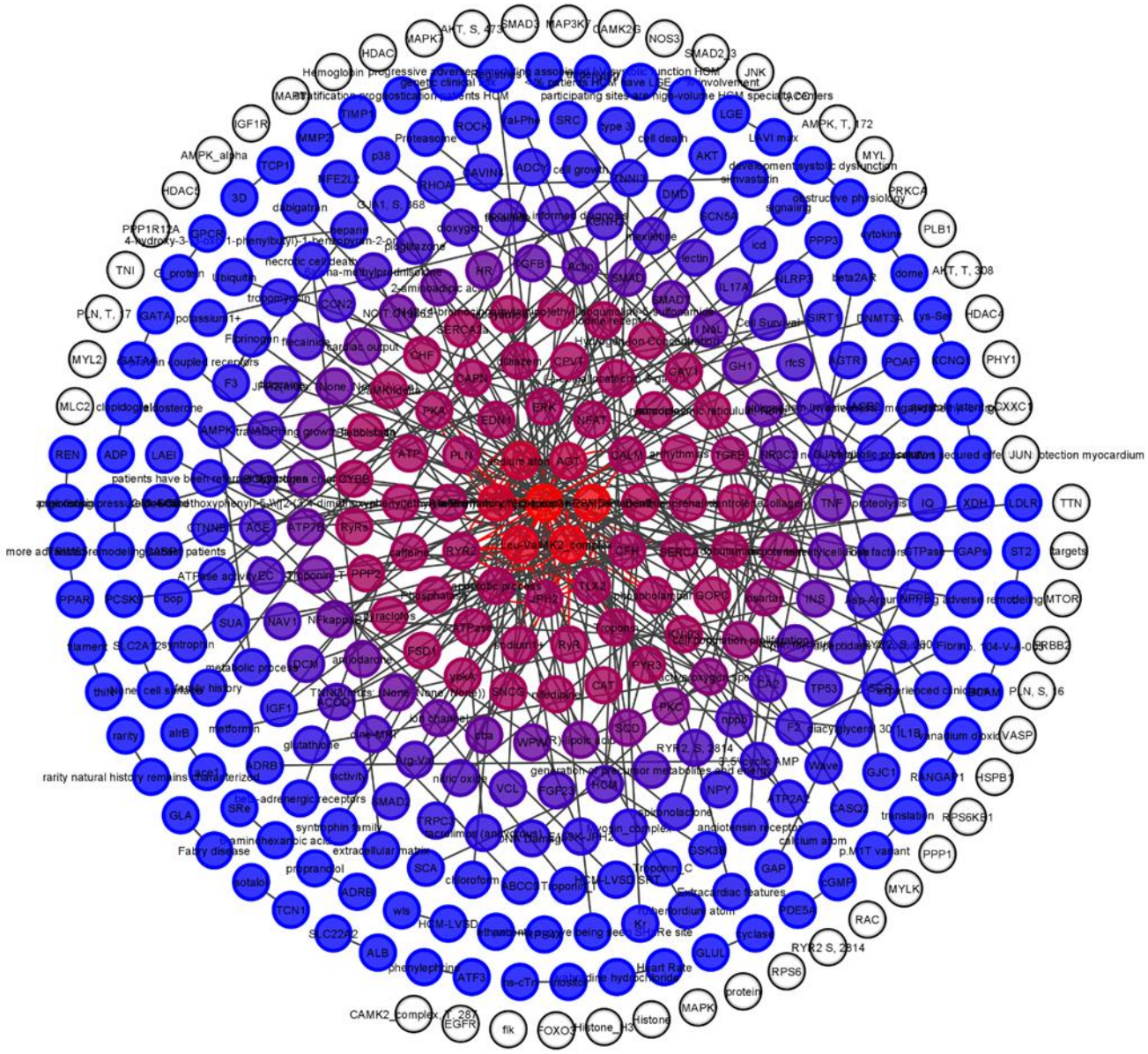

Figure 189. HCM ∩ AF (92)

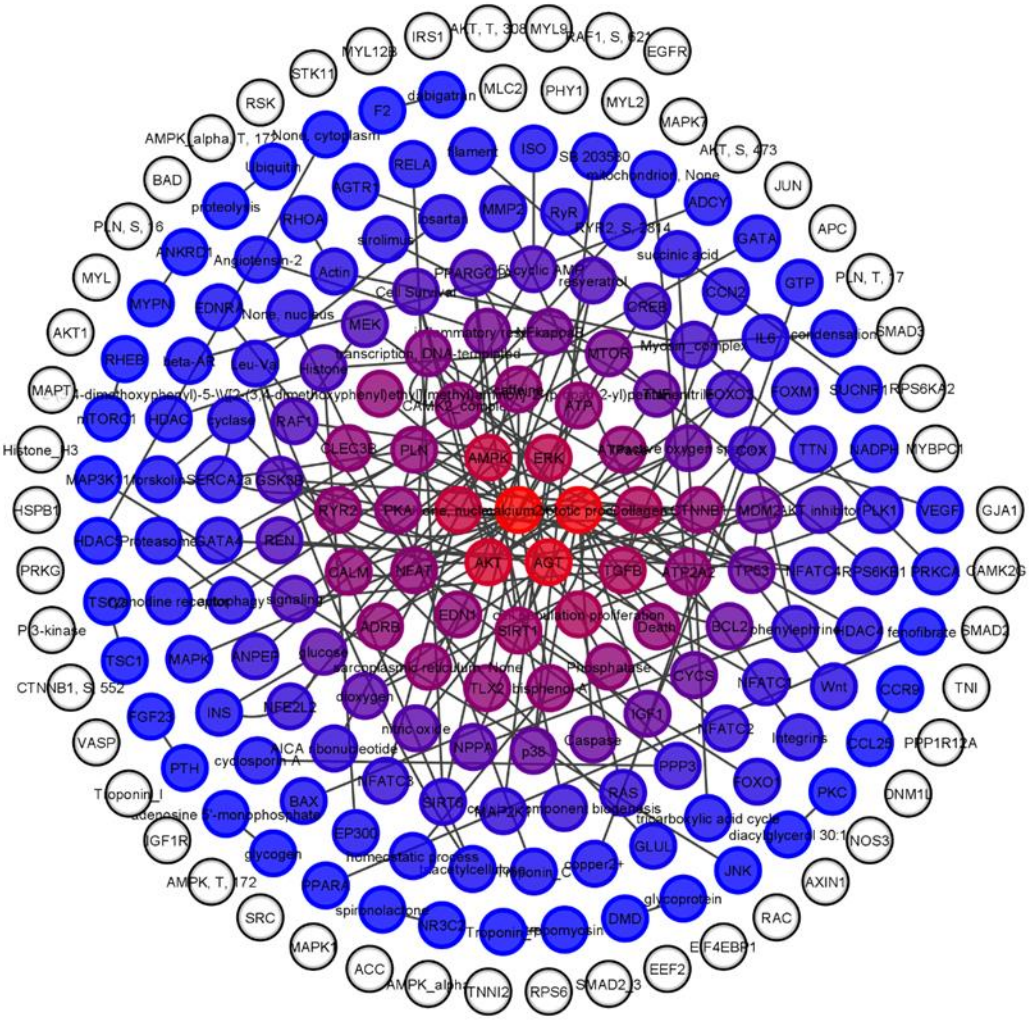

Figure 190. HCM ∩ cardiomyocyte hypertrophy (92)

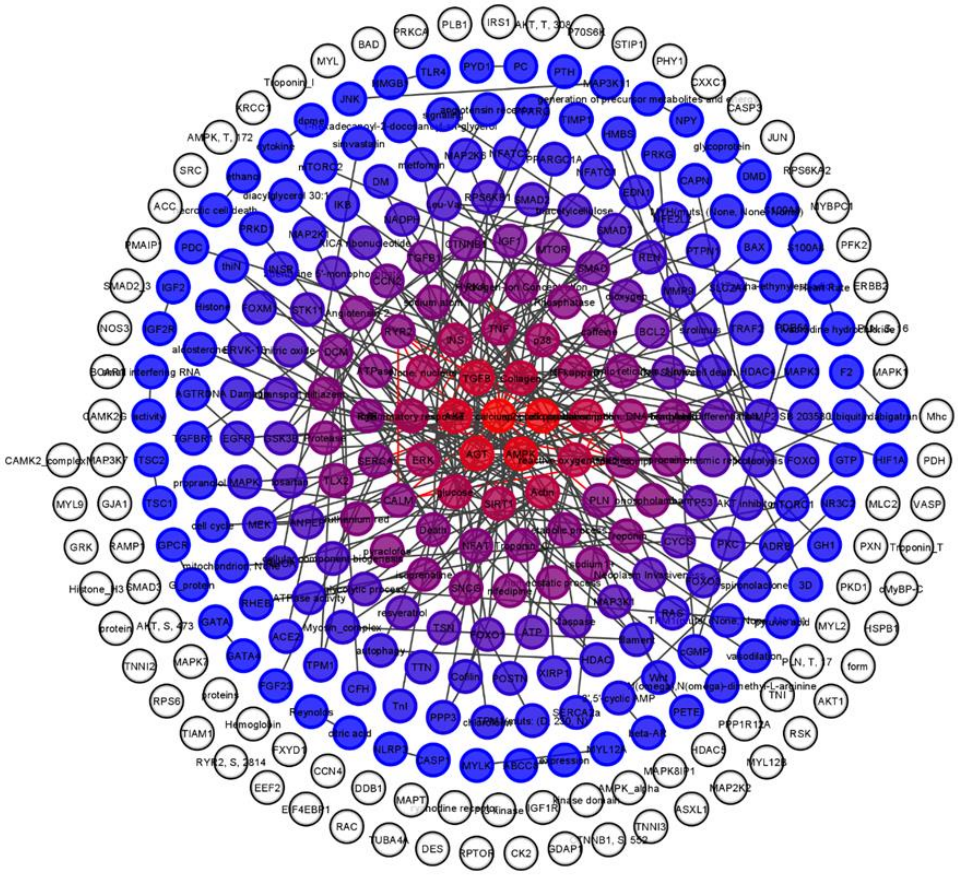

Figure 191. HCM ∩ cardiac remodeling (92)

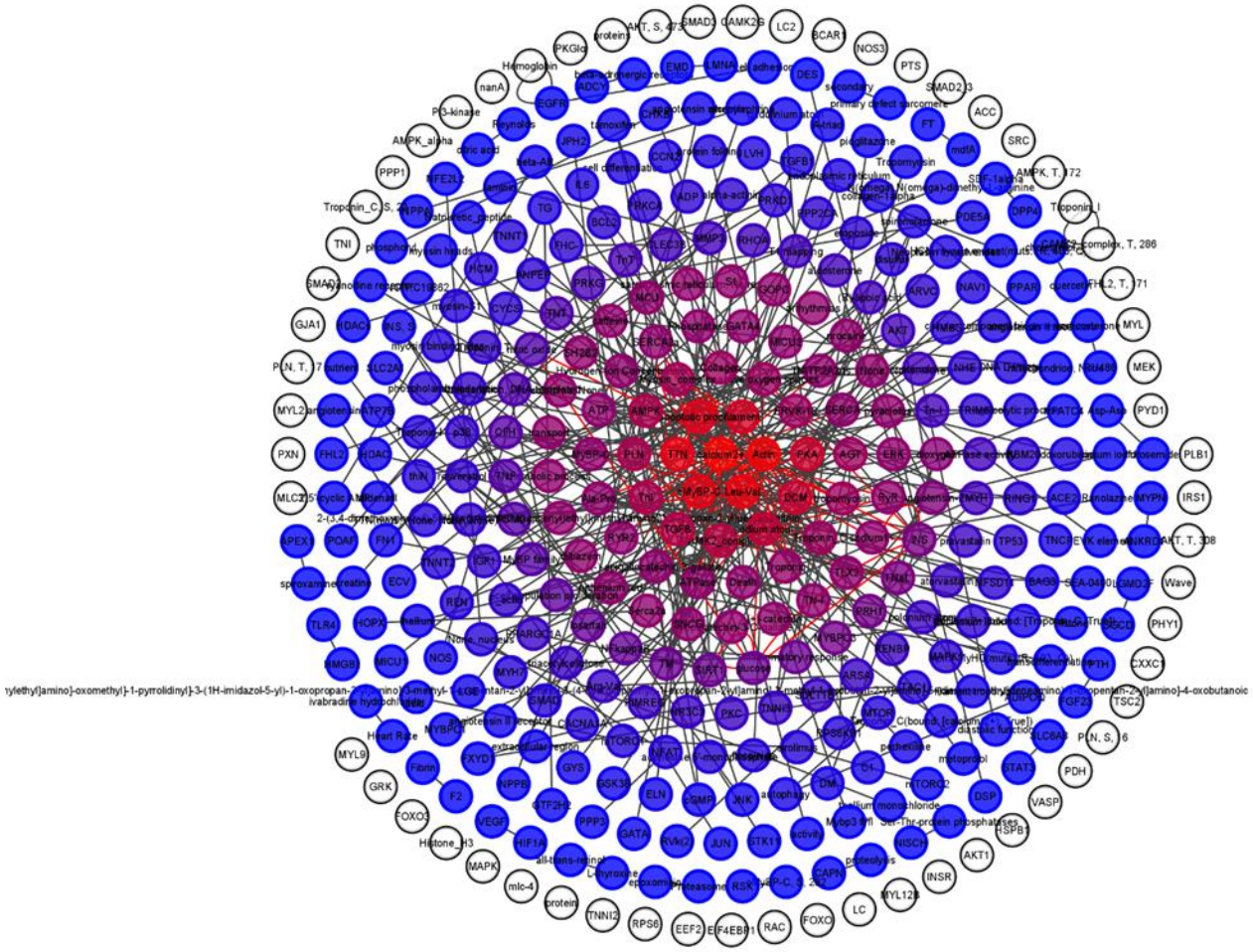

Figure 192. HCM ∩ diastolic dysfunction (92)

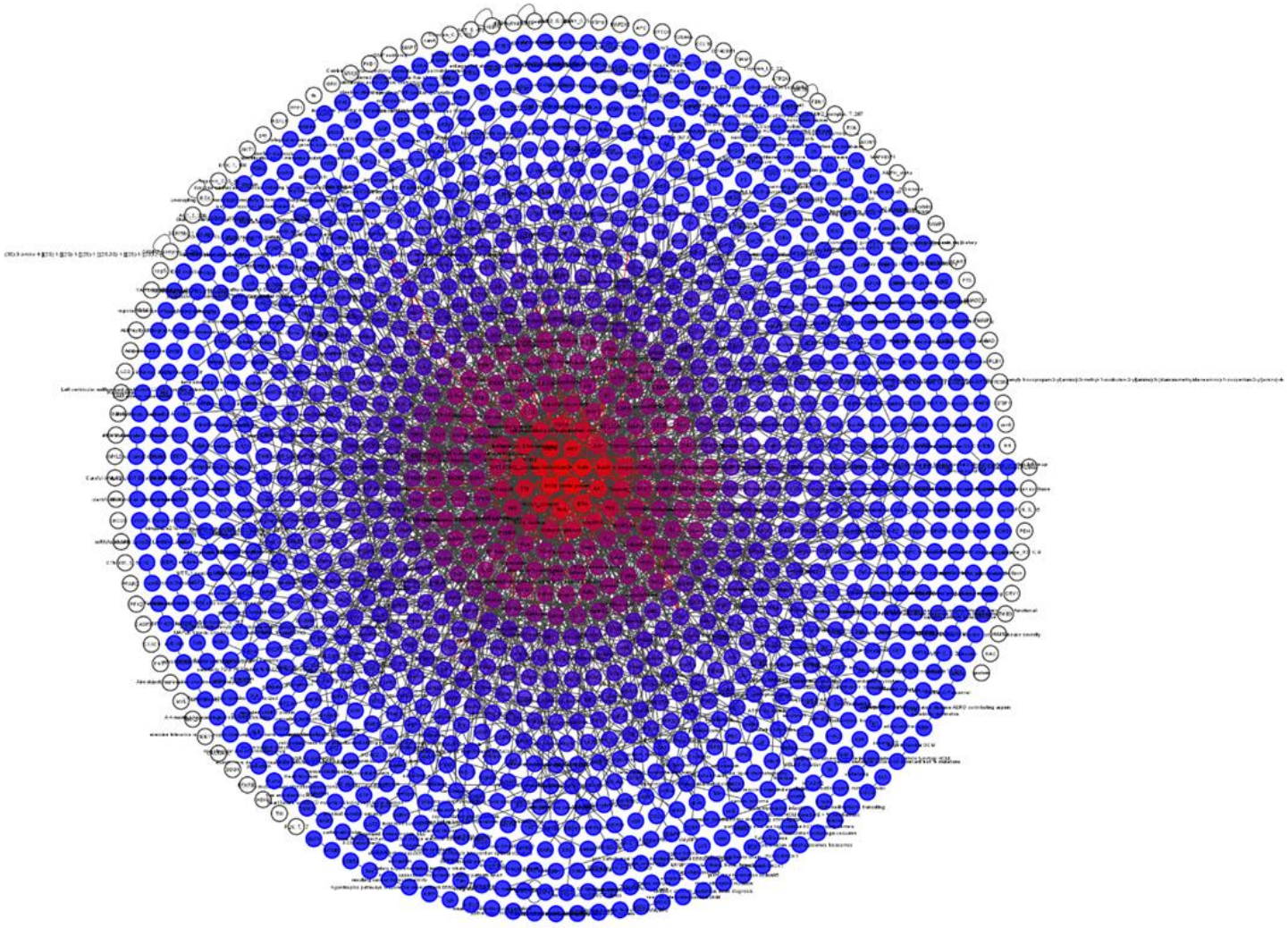

Figure 193. HCM ∩ HF (92)

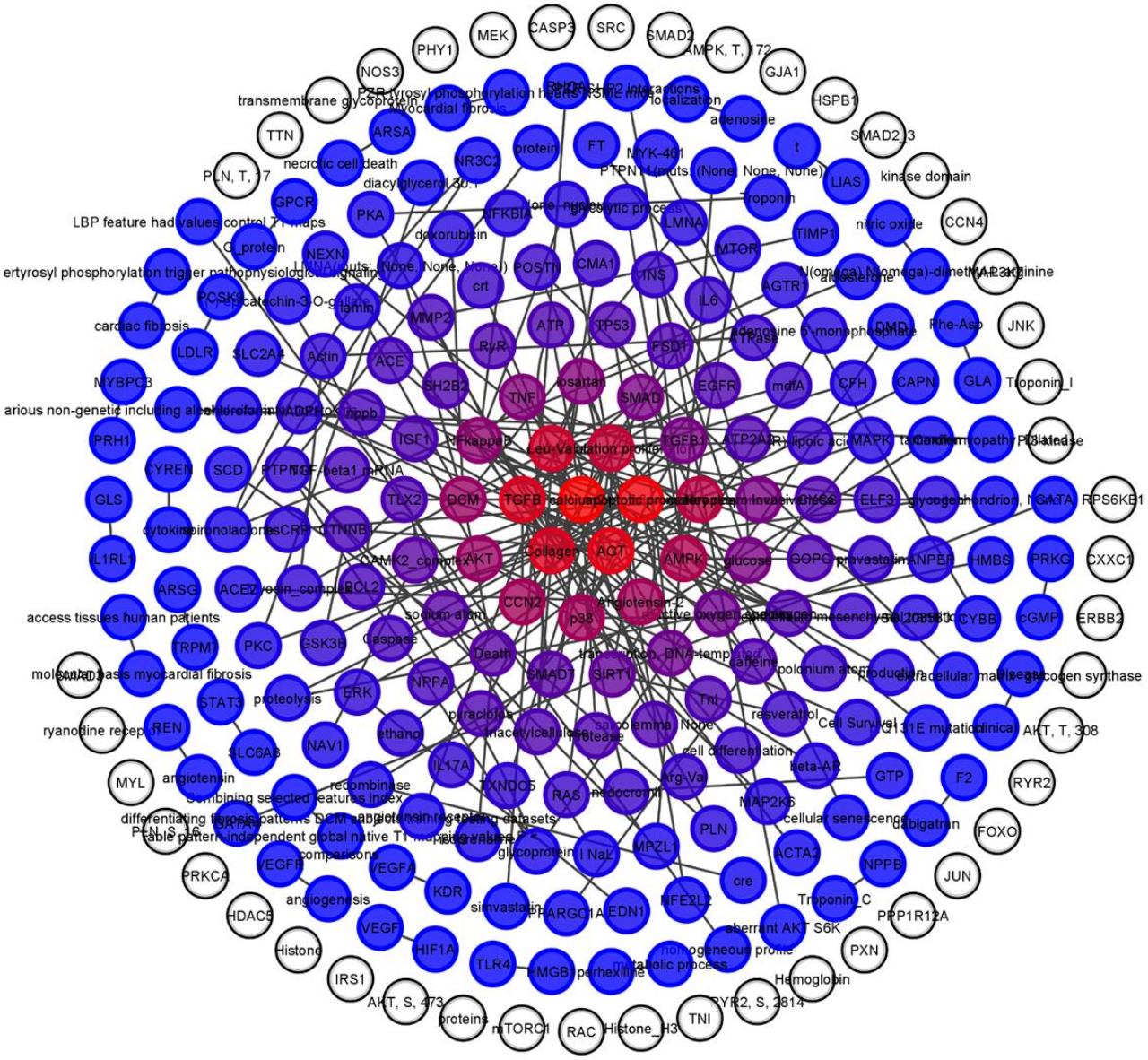

Figure 194. HCM ∩ myocardial fibrosis (92)

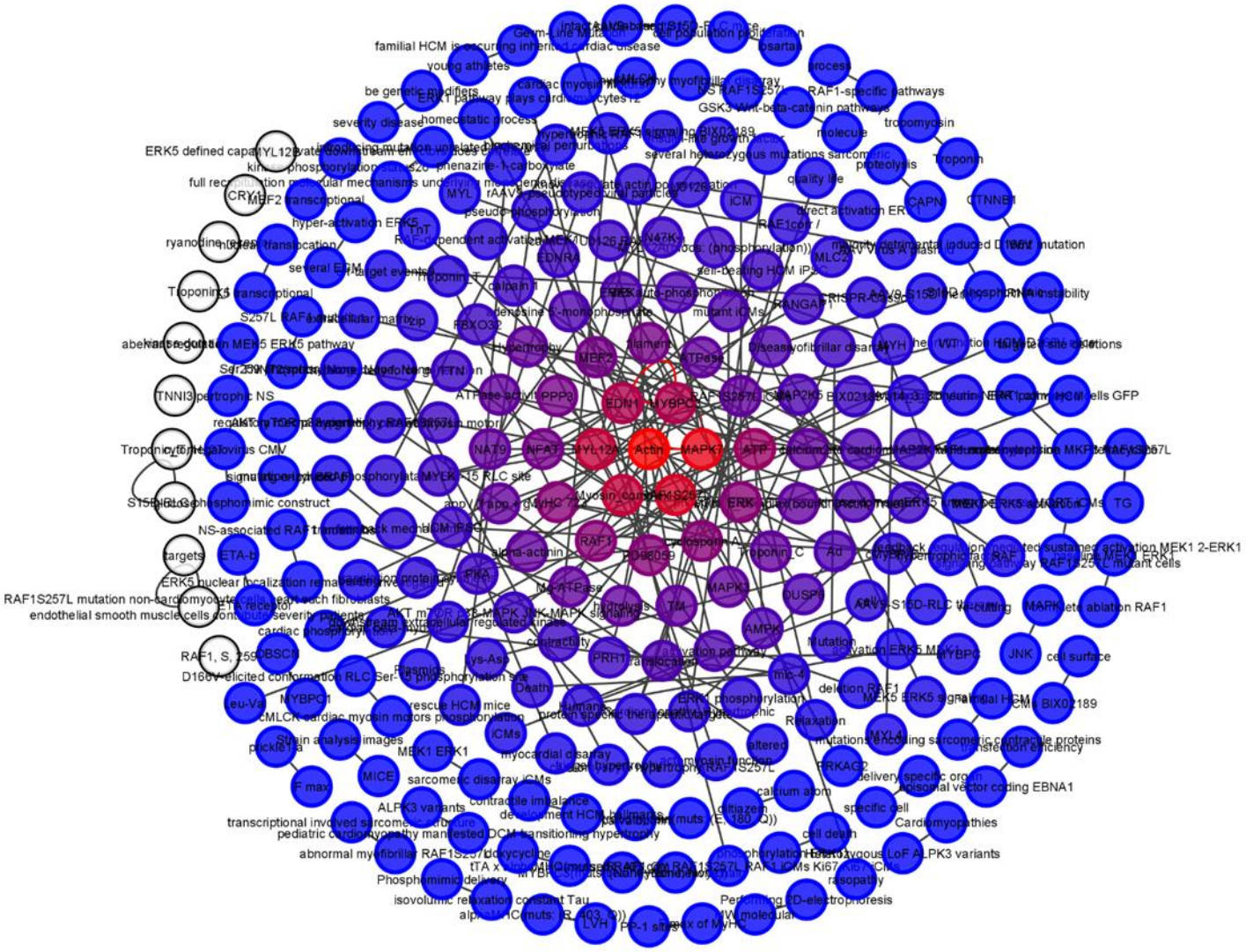

Figure 195. HCM ∩ myofibrillar disarray (92)

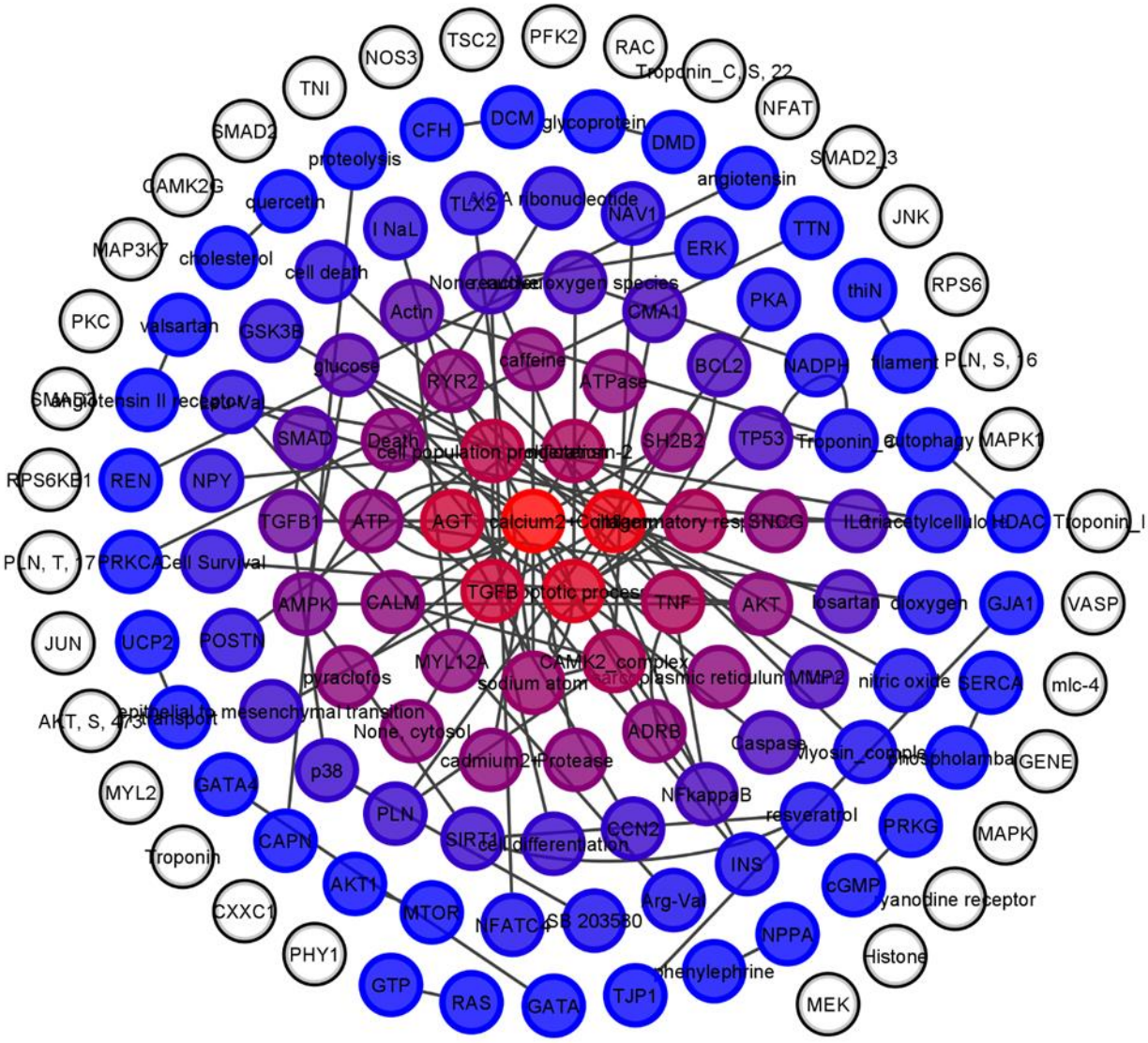

Figure 196. HCM ∩ myocardial remodeling (92)

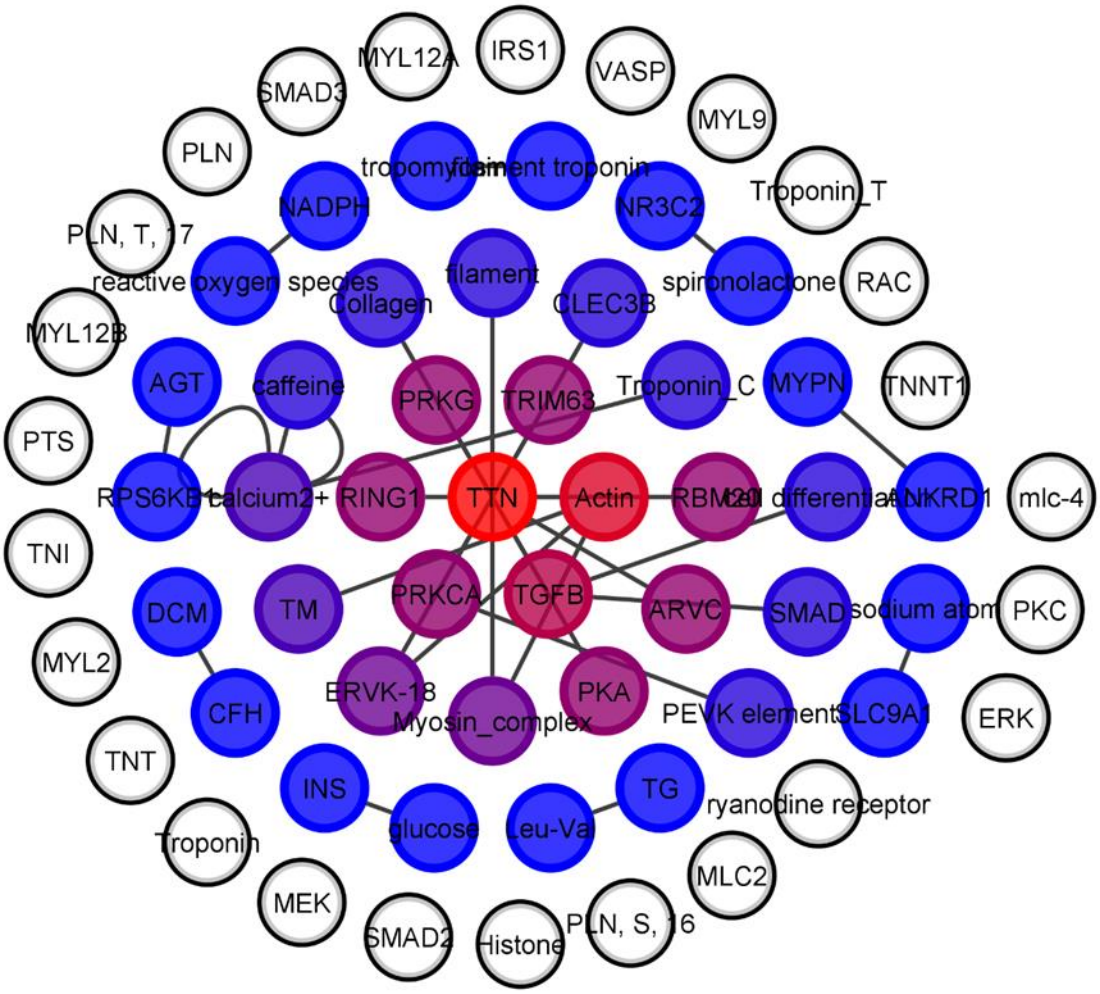

Figure 197. HCM ∩ myocardial stiffness (92)

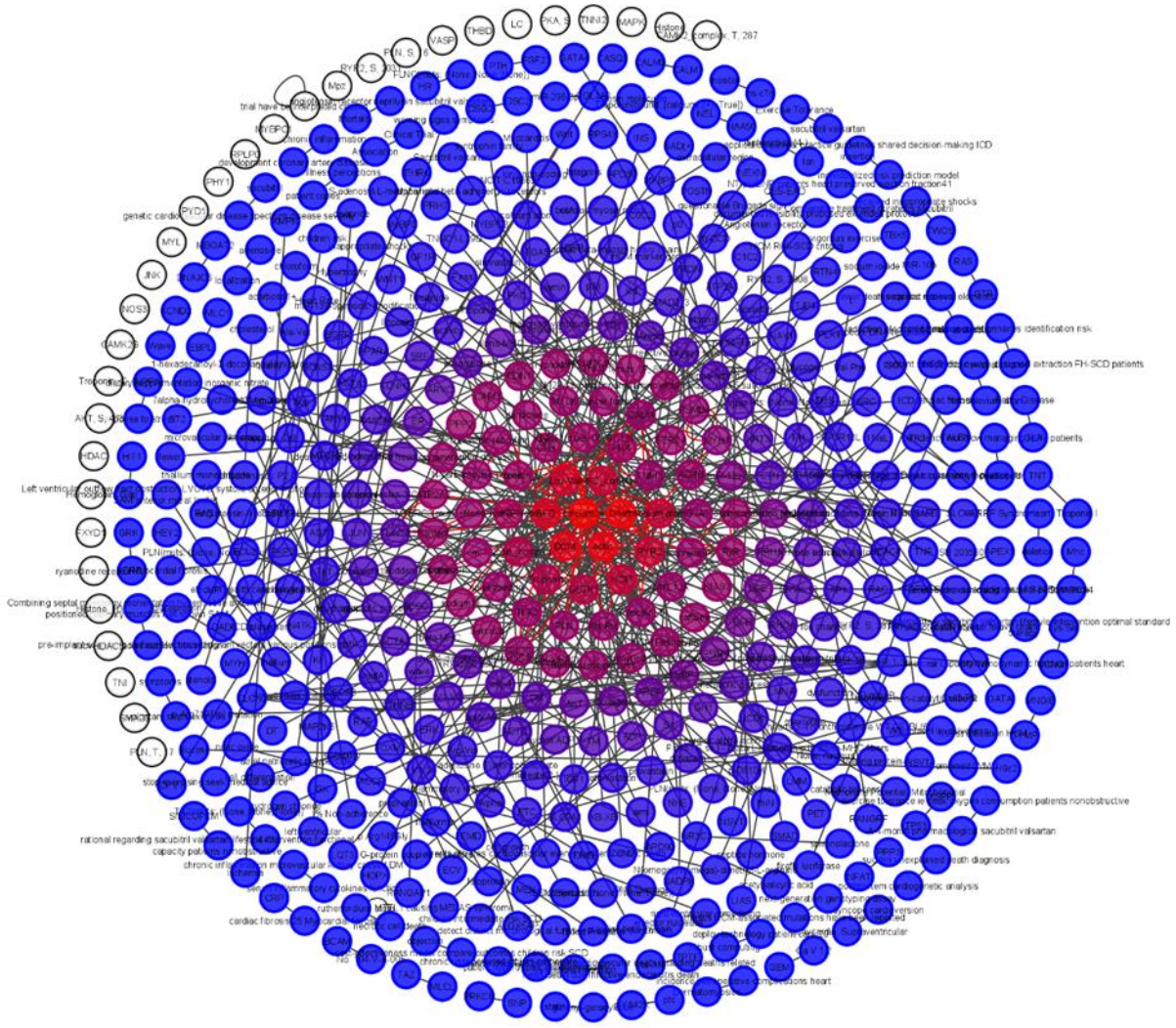

Figure 198. HCM ∩ SCD (92)

4.2.3.1.3. Top genetic and molecular nodes of intersection networks by centrality scores

Centrality scores for all nodes in the intersection networks were determined, and the top nodes by centrality scores were visualized (Table 130) (92).

Table 130. Top genetic and molecular nodes ranked by the centrality scores of each network (92)

Network	Top nodes ranked by centrality scores
hypertrophic cardiomyopathy, atrial fibrillation	https://bit.ly/2OhvNzE
hypertrophic cardiomyopathy, cardiomyocyte disarray	https://bit.ly/31LLUsi
hypertrophic cardiomyopathy, coronary microvascular dysfunction	https://bit.ly/3wiA5YR
hypertrophic cardiomyopathy, cardiomyocyte hypertrophy	https://bit.ly/3fCs1Mq
hypertrophic cardiomyopathy, cardiac remodeling	https://bit.ly/39CYWgj
hypertrophic cardiomyopathy, diastolic dysfunction	https://bit.ly/2PQuwju
hypertrophic cardiomyopathy, heart failure	https://bit.ly/3uwQiYP
hypertrophic cardiomyopathy, impaired myocardial relaxation	https://bit.ly/3dubAz7
hypertrophic cardiomyopathy, left ventricular outflow tract obstruction	https://bit.ly/3cJRWzY
hypertrophic cardiomyopathy, major adverse cardiovascular events	https://bit.ly/3fHsE7w
hypertrophic cardiomyopathy, myocardial fibrosis	https://bit.ly/3dc8HUA
hypertrophic cardiomyopathy, myofibrillar disarray	https://bit.ly/2OgKHpM
hypertrophic cardiomyopathy, myocardial ischemia	https://bit.ly/39Hexvk
hypertrophic cardiomyopathy, myocardial remodeling	https://bit.ly/3uj130t
hypertrophic cardiomyopathy, myocardial stiffness	https://bit.ly/2PpsZRM
hypertrophic cardiomyopathy, rehospitalization	https://bit.ly/3dzxLUy
hypertrophic cardiomyopathy, sudden cardiac death	https://bit.ly/3ugy2CI

4.2.3.1.4. Networks of genetic and molecular mechanisms with a reduced level of noise

Intersection networks of genetic and molecular mechanisms with a lower level of noise were created (Table 131) (92).

Table 131. Intersection networks of genetic and molecular mechanisms with a reduced level of noise (92)

Network	Link to networks with reduced level of noise
hypertrophic cardiomyopathy, atrial fibrillation	https://bit.ly/3mhXtRv
hypertrophic cardiomyopathy, cardiomyocyte disarray	https://bit.ly/3cMpoGd
hypertrophic cardiomyopathy, coronary microvascular dysfunction	https://bit.ly/3uhQQ4l
hypertrophic cardiomyopathy, cardiomyocyte hypertrophy	https://bit.ly/3sMXLm1
hypertrophic cardiomyopathy, cardiac remodeling	https://bit.ly/3dBWPdU
hypertrophic cardiomyopathy, diastolic dysfunction	https://bit.ly/3cNFNu8
hypertrophic cardiomyopathy, heart failure	https://bit.ly/3mi5i9U
hypertrophic cardiomyopathy, impaired myocardial relaxation	https://bit.ly/31MjKxv
hypertrophic cardiomyopathy, left ventricular outflow tract obstruction	https://bit.ly/31GacUC
hypertrophic cardiomyopathy, major adverse cardiovascular events	https://bit.ly/2QZt66N
hypertrophic cardiomyopathy, myocardial fibrosis	https://bit.ly/324nVoj
hypertrophic cardiomyopathy, myofibrillar disarray	https://bit.ly/39Ebt2N
hypertrophic cardiomyopathy, myocardial ischemia	https://bit.ly/31GmOLu
hypertrophic cardiomyopathy, myocardial remodeling	https://bit.ly/3duSh8V
hypertrophic cardiomyopathy, myocardial stiffness	https://bit.ly/3sMZnfz
hypertrophic cardiomyopathy, rehospitalization	https://bit.ly/2QYjedr
hypertrophic cardiomyopathy, sudden cardiac death	https://bit.ly/3wxjGzZ

4.2.3. Shared molecular pathways of HCM and its clinical presentations

The most important putative genetic and molecular elements shared by HCM and its clinical presentations are illustrated (Figure 199). Based on the automatically collected genetic and molecular mechanisms and subsequent analysis, the most important putative pathways are depicted as well (Figure 200) (92).

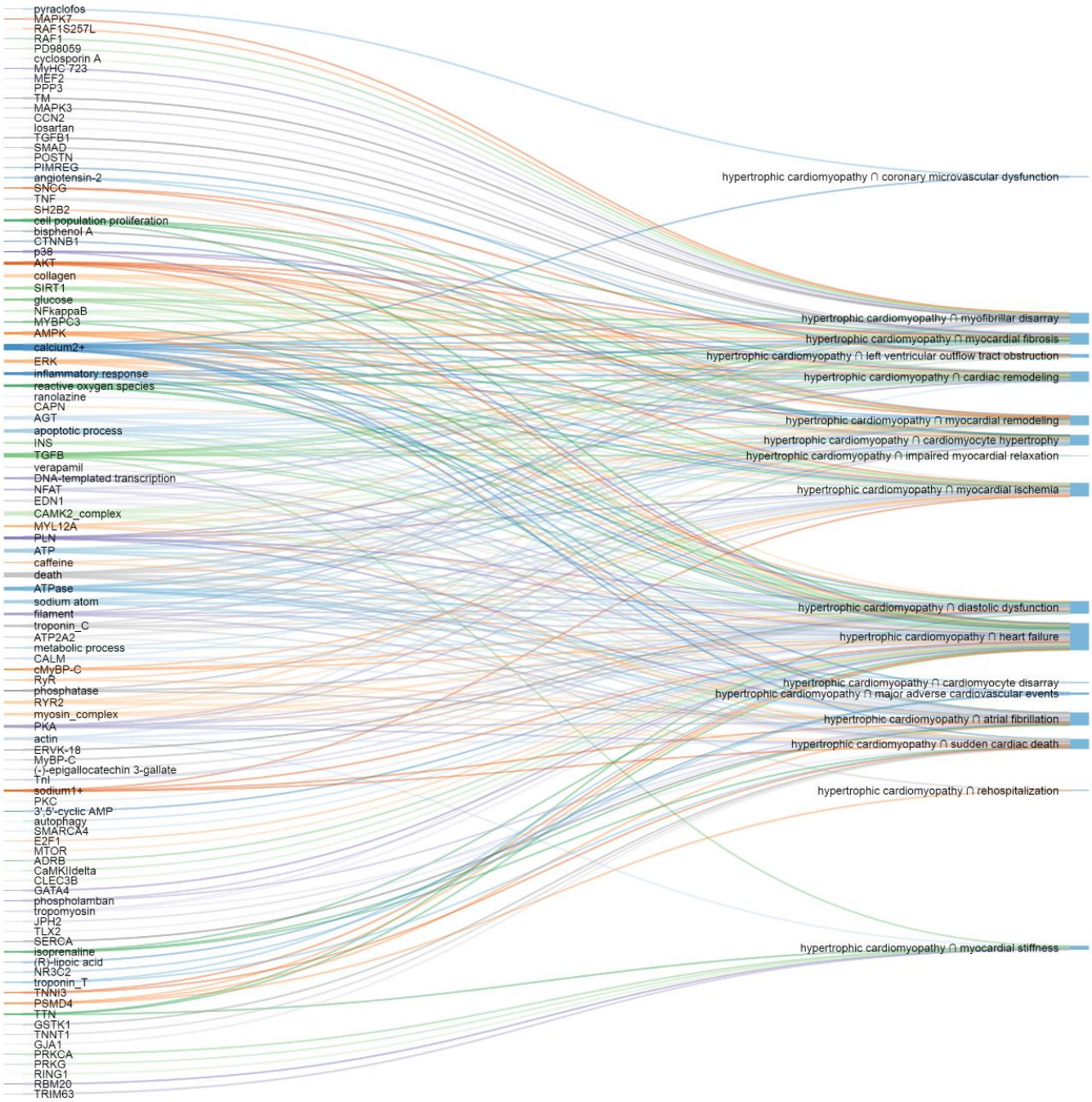

Figure 199. The most important putative genetic and molecular elements (left) and corresponding HCM presentations (right) (92)

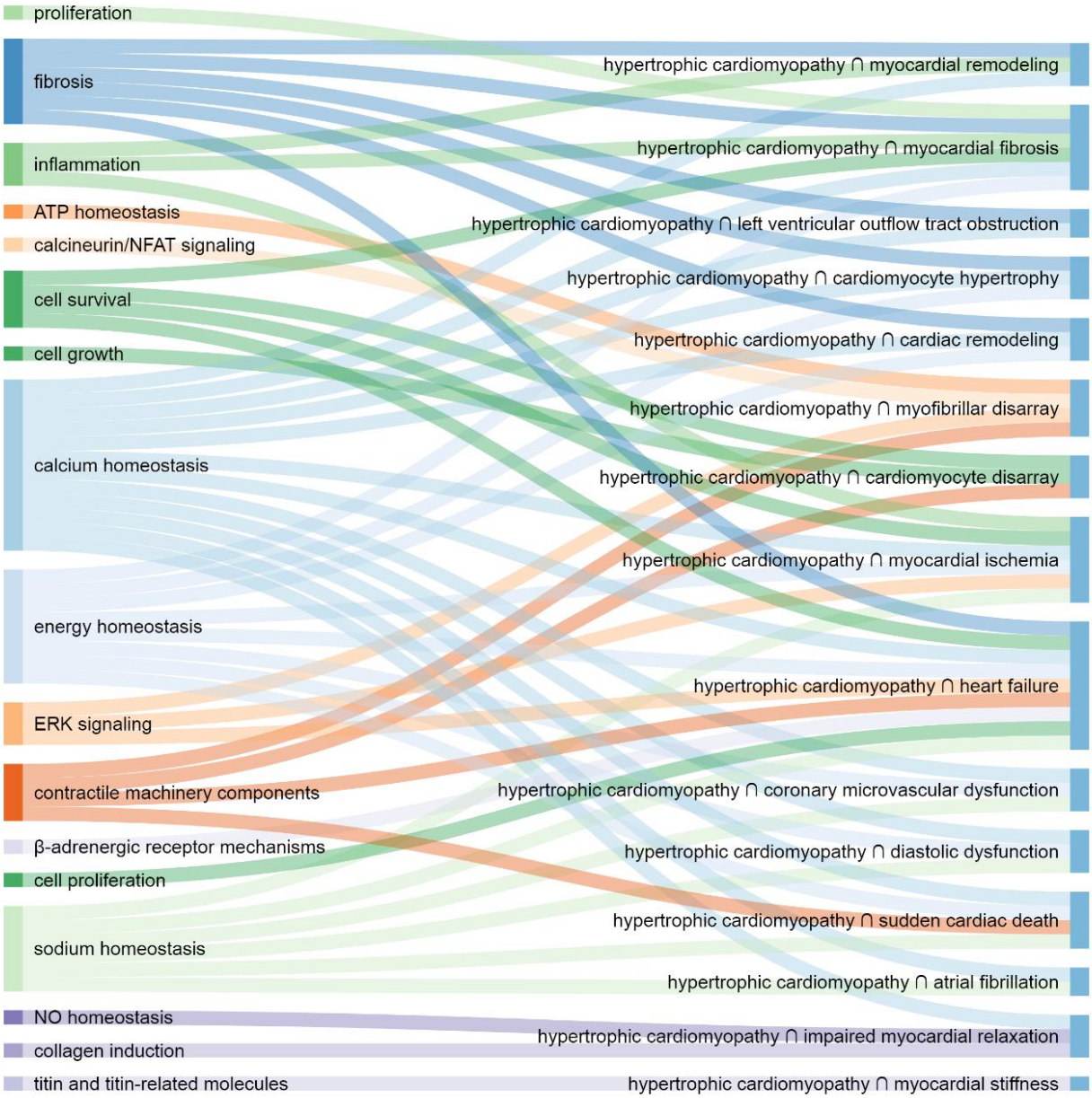

Figure 200. The most important putative pathways (left) and corresponding HCM presentations (right) (92)

4.2.4. HCM clinical course dynamics on lower scale

HCM Clinical interactive knowledge resource was built as an interactive knowledge resource for clinical aspects of HCM diagnosis (including differential genetic and clinical diagnosis), monitoring, and treatment (Figure 201). One section of HCM Clinical interactive knowledge resource is dedicated to HCM clinical course dynamics

represented on a lower scale (Figure 202). HCM Clinical interactive visual review resource is available at: https://silicofcm.eu/hcm-clinical-interactive/

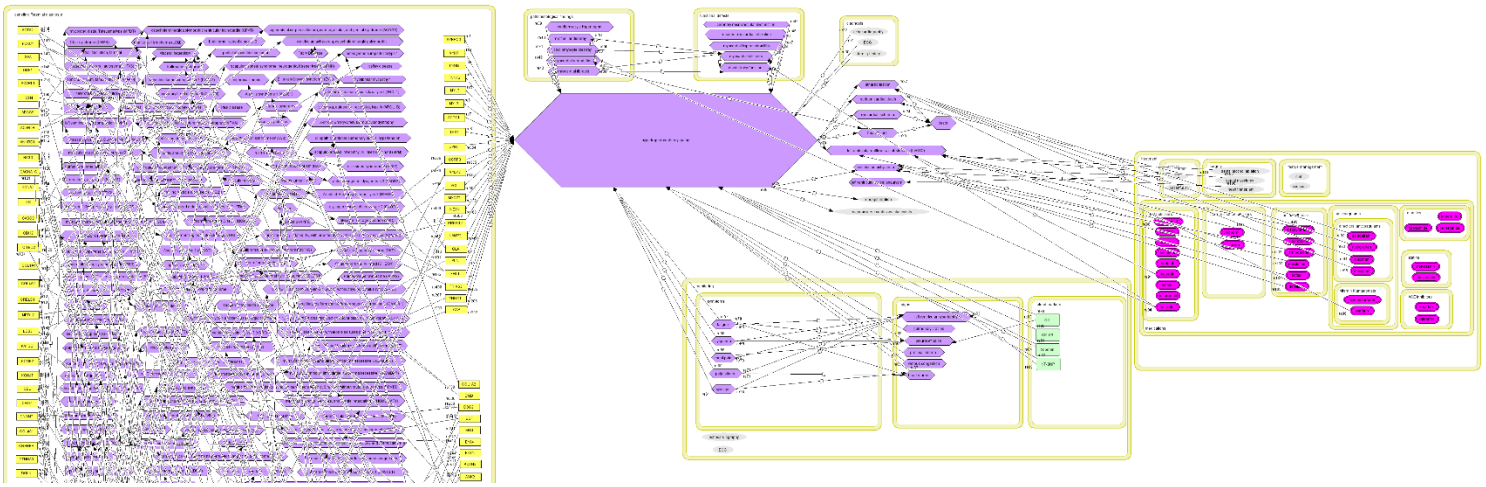

Figure 201. A section of HCM Clinical: An interactive knowledge resource for HCM diagnosis, monitoring, and treatment

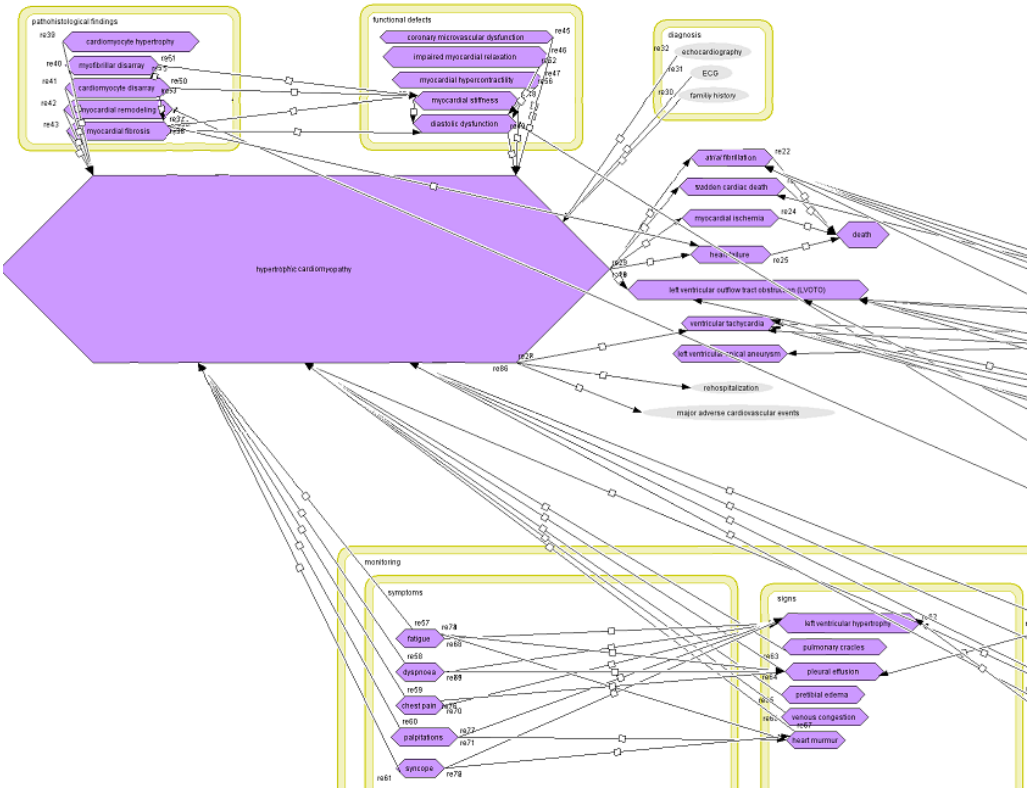

Figure 202. A section of HCM Clinical: An interactive knowledge resource dedicated to HCM clinical course dynamics on lower scale

5. DISCUSSION

5.1. Genotype-phenotype associations in HCM: Examination by machine learning algorithms using clinical data

5.1.1. Identification of HCM subphenotypes using clustering

Since 4 was determined as the optimal number of clusters, the 4-cluster setting will be discussed in more detail.

Cluster 0 consists of 55 patients, cluster 1 of 42 patients, cluster 2 of 17 patients, and cluster 3 of 29 patients. Clusters were determined based on 139 parameters (283).

HCM cases are traditionally categorized into subtypes according to morphologic sites of hypertrophy (e.g. basal, midventral, apical) (299). In general, there are 2 types of HCM: obstructive (70% of HCM cases, with LVOTO) and nonobstructive type (30% of HCM cases) (300). On the other hand, the ESC Working Group recommends morphological classification (32), while the AHA Working Group recommends that HCM should be specified genetically (301).

Maron *et al*. described 4 HCM types, based on morphologic sites of hypertrophy: type 1: basal septum; type 2: whole septum; type 3: septum, anterolateral, and anterior walls; type 4: apical LV (302,303). Helmy *et al*. proposed a HCM classification based on the hypertrophy patterns: septum alone; septum and adjacent segments; apical in combination with other LV segments; and apical (303,304). Syed *et al*. suggested at least 5 main subtypes of HCM based on the septal contour, extent and location of hypertrophy: apical form, mid-ventricular form, sigmoidal septum, neutral contour, and reverse curvature (303,304).

Kim *et al*. concluded that apical forms present milder myocardial fibrosis and diastolic dysfunction, and consequently better prognosis than other HCM types (305). Parato *et al*. reported that echocardiographic patterns significantly affect clinical course and prognosis of HCM (303). Tang *et al*. showed that different morphological types of HCM have different prognoses after surgical myectomy (299).

LDH has the highest values in cluster 0. Comorbiditic genetic disease is most frequent in cluster 0. Diastolic blood pressure is the highest in cluster 0. HCM in family history is most often present in cluster 1. In cluster 2, patients are more obese and have higher values of serum creatinine. Patients of cluster 2 have the highest measures of the

LV cavity. Cluster 3 mainly consists of women. Cluster 3 has the minimal diameters and volumes of the LV in both systole and diastole with the highest thickness of the interventricular septum among the clusters. Cluster 3 has the worst diastolic function and the highest left ventricular filling pressures expressed through E/e'. Left ventricular systolic function recorded through ejection fraction was largest in cluster 3 and smallest in cluster 2. SAM is most frequent in cluster 3 and absent in cluster 1. Patients in clusters 0 and 1 are younger than patients in clusters 2 and 3. In clusters 0 and 3, heart murmur exists in most cases, while it is a rare in clusters 1 and 2 (283).

Cluster 0 consists of younger patients with heart murmur, higher LDH values, and higher diastolic blood pressure (283).

Cluster 1 could be described as made of younger patients, often without heart murmur and SAM, but with positive HCM family history (283).

Cluster 2 is constituted of older, more obese males, often without heart murmur and with comparatively higher serum creatinine. It has the the lowest left ventricular systolic function and the highest measurements of the LV cavity (283).

Cluster 3 is mostly a female cluster, containing older patients, often with heart murmur and with SAM in approximately 60% of patients (283).

It has the smallest volumes and diameters of the LV and the greatest left ventricular systolic function (283).

These results suggest 4 HCM subtypes: cluster 0, distinguishable by "AHOLD" (word constructed from AO and LDH, from whose quantities the cluster is distinguishable) wherein values are mostly being AO > 30 mm, LDH > 300 U/L, LVOT Vmax < 2 m/s, MV maxPG < 2 mmHg, PLWD < 12 mm, and serum albumin > 44 g/L; cluster 1, distinguishable by "RVSP ASCAOVS" (RVSP, AscAO, and AOvs) wherein values are mainly being RVSP < 28 mmHg, AscAO < 31 mm, AOvs > 27 m/s, LVOT Vmax < 2 m/s, MV meanPG < 1 mmHg, PLWD < 12 mm, and serum albumin > 44 g/L; cluster 2, distinguishable by weight, wherein values are mainly being weight > 95 kg, LVOT Vmax > 2 m/s, PLWD > 12 mm, and serum albumin < 43 g/L; and cluster 3 distinguishable by "AV LVOT PG" (AV meanPG, AV maxPG, and LVOT maxPG) wherein values are mostly being AV maxPG > 15 mmHg, AV meanPG > 6 mmHg, LVOT

maxPG > 15 mmHg, LVOT Vmax > 2 m/s, MV maxPG > 5 mmHg, MV meanPG > 2 mmHg, PLWD > 12 mm, and serum albumin < 43 g/L (283).

Even though statistically significant differences were revealed among 4 clusters for many features, for some of them, the overlapping intervals of their values impede their usage for distinguishing the clusters from each other (283).

5.1.2. Genotype-cluster associations

The absence of association of cluster membership and genotype in some cluster settings does not mean that the genotype-phenotype association does not exist. It shows that clustering algorithms choose other characteristics of patients preferably to determine the optimal clusters of HCM patients and that these clusters do not correlate with the genotype.

5.1.3. Genotype-phenotype associations

5.1.3.1. Predicting phenotypic outcomes using genetic data only

In general, ML algorithms were struggling to predict phenotypic outcomes using genotypic data only — most of the created models have less than adequate prediction performance, especially when taking into account that AP is a more appropriate measure of performance than AUC for imbalanced datasets. In such cases, AutoML was applied both before and after oversampling, without significant success.

Shown relationships between genes (and their combinations) and phenotypic outcomes both explain and confirm that predicting HCM phenotype using genotypic data only is not an easy task, for both classical statistics and ML algorithms.

Two outcomes that can be predicted from mutated genes are absence or presence of sinus rhythm and absence or presence of myocardial injury — infarction. Models predicting absence or presence of sinus rhythm had similar performance when they were built using causative genes only and using all analyzed genes. The model with best performance for predicting absence or presence of sinus rhythm is based on Logistic Regression with oversampling, and has accuracy 0.83, F1-score 0.91 and AP 0.86 (in this case AP metric is appropriate due to imbalanced dataset (306,307), with more patients

having the predicted outcome). Several other models based on other ML algorithms have similar but slightly worse performance.

Models predicting myocardial injury — infarction had better performance when they were built using all analyzed genes (and not just causative ones). The model with best performance for predicting myocardial injury — infarction is based on Logistic Regression with oversampling, and has accuracy 0.73, F1-score 0.84 and AP 0.78 (in this case AP metric is more appropriate due to imbalanced dataset, with more patients having the predicted outcome). Several other models based on other ML algorithms have similar but slightly worse performance.

5.1.3.2. Predicting phenotypic outcomes using both genetic and other phenotypic data

5.1.3.2.1. Symptoms and signs

The associations between peak VO_2 and dyspnea, as well as VO_2 and dyspnea, are intuitive. Seiler *et al*. reported dyspnea as associated with left ventricular dilatation in HCM (308), while Karaye *et al*. found increased prevalence of moderate to severe dyspnea in hypertensive patients with decreased TAPSE (309).

Chida *et al*. found a high frequency of syncope in carriers of mutation in *MYBPC3* (310). However, Song *et al*. found more frequent syncope when mutations were in *MYH7* compared with mutations in *MYBPC3* (311).

In diabetes mellitus, thyroid and hepatic disorders, pretibial edema might be found (312–315).

5.1.3.2.2. Conduction and rhythm disorders

Palermo *et al*. reported a greater anaerobic threshold in HF patients with AF, than in HF patients with sinus rhythm due to a higher heart rate and a greater heart rate increase during exercise (316).

LA remodeling is a substrate for AF (283,317). LA dilatation has been shown to be useful in identifying the individuals that will present with AF, and it is suggested that all AF patients should be evaluated for LA size and anatomy (317). Additionally, LA enlargement is associated with AF (283,318).

PR interval might represent a predictor for AF (283,319). Kornej *et al*. reported that prolongation of PR interval shares similar characteristics with AF (283,320). LA fibrosis plays important role in AF pathogenesis and represents a risk factor for adverse outcomes in AF (283,321).

Wałek *et al*. proposed that LA dyskinesia registered in the LASct4c and 4c views represent independent risk factors for AF recurrence after direct current cardioversion (322).

Endocrine factors are important in AF pathogenesis (283,323).

In a study by Roshdy *et al*., a group of 49 patients with advanced-degree AV block had a significantly larger LA diameter than the 48 patients in the matched control group (324). Moreover, AV block is sometimes associated with MR (283,325–327).

Bay *et al*. reported that in the middle-aged population in Switzerland, the male sex is associated with RBBB (328). This is confirmed by the studies of Francissco-Pascual *et al*. on 443 patients (329), Trullàs *et al*. on 3,638 acute HF patients (330), and Bessem *et al*. on 1436 athletes, all reporting that RBBB is found more often in men (283,331).

5.1.3.2.3. Ischemia

Aguiar Rosa *et al*. found that ischemia is linked with a shorter time to anaerobic threshold in HCM patients (332). Chest pain is known to be frequently associated with ST-segment elevation.

In addition, cMyBP-C is associated with ischemia, as cMyBP-C proteolysis during ischemia impairs function of cells that survive ischemia-reperfusion injury. cMyBP-C phosphorylation by protein kinase A (PKA) controls myocardial function and provides protection of cardiac tissue against ischemia-reperfusion injury and resistance to proteolysis (333).

Moreover, post-ischemic proteolysis affects cardiac sarcomere proteins, as multiple proteins (cMyBP-C, cardiac troponin I, and myosin light chain) are post-translationally modified and degraded (333).

5.1.3.2.4. Technical and statistical aspects

Clinically illogical feature combinations were excluded in feature selection phase. Groups of combined features sufficient to "predict" a specific outcome are not typical predictors, but rather a set of co-expressions and associations with the "predicted" features. Some of these are previously reported, other are indicated in rare studies; however, there are some that are completely unknown (283).

5.1.4. Genotype-specific echocardiogram findings

5.1.4.1. *MYH7*

Septum, LVOT segment, anterior wall, apex, RV, and mitral apparatus were shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYH7* gene and those without detected mutation.

MYH7 is one of 2 most frequent causal genes in the usual type of HCM with involvement of the basal septum and apical HCM (29). Li *et al*. described a case of mutation in *MYH7*-induced HCM with significantly hypertrophied systolic interventricular septum (334). Van de Sande *et al*. described a case with myocardial crypt found in the anterior septum in carrier of the mutation in *MYH7* (335). In study of Gruner *et al*. in genetically tested cohort of 61 patients with apical HCM, sarcomere protein gene mutations were most often found in *MYH7* and *MYBPC3* genes (336). In study of Chung *et al*., among the 212 patients, pathogenic mutations in sarcomere-associated genes were more prevalent in non-apical HCM than in apical HCM and mutation frequency in *MYH7* gene was 9% in cases of apical HCM (337). In study of Velicki *et al*. involving 63 HCM *MYBPC3*- or *MYH7*-mutation positive patients, calcifications of mitral annulus were found only in patients with mutation in *MYH7* gene and mitral leaflet abnormalities were significantly more often in patients with mutation in *MYH7* gene (107). Waldmüller *et al*. reported a link between mutations in *MYH7* and a higher degree of mitral valve regurgitation in HCM (338).

5.1.4.2. *TNNT2*

Septum and RV were shown to be discriminative for classification of echocardiographic images of patients with mutation in *TNNT2* gene and those without detected mutation.

In study by Mori *et al.*, *TNNT2* p.Lys263Arg was shown to be associated with increased septum thickness (339).

5.1.4.3. *MYBPC3*

Septum, LV, and LV chamber were shown to be discriminative for classification of echocardiographic images of patients with mutation in *MYBPC3* gene and those without detected mutation.

MYBPC3 is one of 2 most frequent HCM causal genes in the usual type of HCM with involvement of the basal septum (29). Tarkiainen *et al*. reported that patients with the *MYBPC3*–Q10961X mutation have rised left ventricular septal convexity regardless of the presence of LVH (340). Waldmüller *et al*. reported a link between mutations in *MYBPC3* and a particularly large thickness of the interventricular septum (338).

5.2. Molecular mechanisms of genotype-phenotype associations in HCM: Examination by machines using literature data

5.2.1. HCM Map: Interactive knowledge resource about molecular mechanisms of HCM

HCM map is a quantum of knowledge on HCM mechanisms available today, still some molecular mechanisms have yet to be discovered. HCM Map is a resource for sharing and gathering molecular mechanisms of HCM — each component of the HCM Map can be commented by users and that will enable HCM Map to be continuously improved and updated involving the community of HCM molecular mechanisms experts (287).

5.2.2. Molecular mechanisms of genotype-phenotype associations in HCM collected and represented by machines

Machine-assembled representations of molecular mechanisms are a rich source of information, which generate detailed representations of diseases. Such an approach also brings a certain level of noise (incorporation of non-molecular elements) and inaccuracy (wrongly or misleadingly labeled elements) into the disease representations. Low consensus between models generated by different computer-assisted approaches regarding the top elements in terms of centrality measures, was to a certain degree a repercussion of diverse labeling among models and inconsistent labeling within models. The rare elements found as a consensus reflect the simultaneously fulfilled conditions of the same labeling principle and agreement about the highest values of centrality measures. General agreement regarding the most important nodes in each of created representations was reached only with respect to calcium (287).

5.2.3. Genetic and molecular mechanisms of different clinical presentations of HCM collected and represented by machines

These findings indicate that cardiomyocyte hypertrophy, myocardial fibrosis and remodeling, and cardiac remodeling; coronary microvascular dysfunction and myocardial ischemia; myocardial ischemia and HF; AF and SCD have similar molecular mechanisms (92).

The suggested shared pathways are in agreement with the clinical articles on HCM progression (341,342), ischemic character and interrelation of coronary microvascular dysfunction and myocardial ischemia (1,343,344), association of myocardial ischemia and HF in HCM (345,346), arrhythmic essence and link between AF and SCD (347–349). The obtained results propose a more distinctive character of myocardial stiffness, impaired myocardial relaxation, myofibrillar and cardiomyocyte disarray pathways. That is probably a result of the rarity of publications available about genetic and molecular mechanisms of these clinical contexts in HCM (and as a result, statements extracted). That reduces the overall ability of applied methodology to detect the most important molecular elements (92).

Automated extraction of genetic and molecular HCM mechanisms produces a lot of noise (e.g., elements that are not molecular mechanisms). Intersection operation removes much of the noise, because an avoidable element should exist in both intersected sets in order to occur in results. However, some noise still persists in shown results that are collected and represented by machines. The automatic extraction of genetic and molecular HCM mechanisms occasionally extracts gene products with the label of the corresponding genes. Consequently, each element with a gene label must be understood as the gene itself and/or its product (92).

Applied automated extraction of genetic and molecular HCM mechanisms was not intended to unravel definitive genetic and molecular basis of HCM genotype-phenotype associations, but to provide a foundation for further research (92).

5.2.4. HCM clinical course dynamics on lower scale

Although the HCM Clinical interactive knowledge resource was built with a different aim (as an interactive visual review of HCM diagnosis, monitoring, and treatment), one of its segments is showing HCM clinical course dynamics on lower scale. The HCM Clinical interactive knowledge resource will be updated with knowledge from the clinical guidelines and research about HCM diagnosis, monitoring, and treatment improvements. However, some major changes in its segment showing HCM clinical course dynamics are not expected.

5.3. Research hypotheses

All 5 research hypotheses have been supported.

5.4. Future perspectives

Identification of HCM subtypes established by complete phenotypic presentation is a step towards HCM precision medicine, which may ultimately enable the establishment of prevention and treatment strategies refined for specific subgroups of HCM patients (283). Each of the results shown here could be re-inspected in broader, more specific, and classical clinical studies.

5.5. Limitations

5.5.1. Genetic testing

Shown genetic variants are all that were available on genetic variant-level for patients included in this research study — other genetic results were available on mutated gene-level.

5.5.2. Identification of HCM subphenotypes using clustering

"Cut-offs" for distinguishing the clusters were determined for the examined dataset only and must be refined or confirmed before they could be utilized as cut-offs in any other context (283).

5.5.3. Phenotypic outcome prediction

All features with many null values were removed as a part of preparation for further analysis, to eliminate the possibility that the model learns a several such "by heart" and fuse them to deliver a favorable result. Features that may be able to provide direct solutions to the models (for the questions asked in predictions) were also removed, in order to exclude the possibility of data leakage (283).

In the prediction of HCM phenotypic outcomes, the models' performance was surprisingly outstanding. Additional analysis was done to examine the approach taken. For each of the shown phenotypic outcome predictions, models were generated utilizing distinct ML algorithms: logistic regression, linear SVC, and RBF kernel SVC, decision tree, random forest, ridge classifier. Only the best results are shown. Some of models that are not shown had performance metrics around 0.75 (283).

To verify the methodology and exclude the possibility that feature-overfitting is the cause of good performance, an artificial feature was created with the value of 0 for all even-index patients and the value of 1 for all odd-index patients in the dataset (patients were randomly ordered in the dataset). After application of the same methodology in such settings determined AUC values were around or below 0.5 (283).

A possible explanation for this unexpectedly good performance may be that this is a trivial task for ML, with some of these features (or the features as a whole) directly providing the correct answers to the models (if "predicted" phenotypic outcome will be

manifested or not). Some of them are still unknown as direct association or causative relation, especially since features are combined here (which is rarely the case in classical clinical research). Regardless of all the precautions described, these results must be inspected critically and perceived as possibilities rather than final results (283).

Demonstrated features in the HCM phenotypic outcomes predictions are not the best to predict particular outcomes; they are sufficient for each of the predictions and determined for this dataset, after elimination of all the features that may dis-reflect real relationships and associations. Some predicted phenotypic outcomes are not useful in clinical practice. Nevertheless, combinations of features sufficient to "predict" these outcomes may reveal some unknown relations between clinical presentations (283).

5.5.4. Genotype-specific echocardiogram findings

The examined sample is, to a certain extent, homogenic, and these findings need further confirmation using a more diverse sample or in clinical settings. Number of images used to train and validate the model is considerably small for image classification task.

5.5.5. General

Shown results reveal statistical distributions contained in the examined dataset, and require further validation using other datasets or additional research in clinical settings (283). Each of the shown results must be validated on other populations.

6. CONCLUSION

In this study, ML has been employed to facilitate the deciphering of genotype-phenotype associations in HCM.

Four HCM subtypes were identified based on the overall phenotypic appearance: cluster 0 ("AHOLD"), distinguishable by AO and LDH, with values mostly AO > 30 mm, and LDH > 300 U/L; cluster 1 ("RVSP ASCAOVS"), distinguishable by RVSP, AscAO, and AOvs, with the values of RVSP < 28 mmHg, AscAO < 31 mm, AOvs > 27 m/s; cluster 2 ("weight"), distinguishable by weight, with values mostly > 95 kg; and cluster 3 ("AV LVOT PG") distinguishable by AV meanPG, AV maxPG, and LVOT maxPG wherein values mostly being AV meanPG > 6 mmHg, AV maxPG > 15 mmHg, and LVOT maxPG > 15 mmHg.

Although none of the determined subtypes settings could be completely correlated with genotype, some significant correlations with genotype were found in the 3-subtypes setting (*TNNI3*), 4-subtypes setting (*TNNI3* and *MYBPC3*), 5-subtypes setting (*MYH7*), and 6-subtypes setting (*MYBPC3*).

In general, ML algorithms confirmed that the determination of genotype-phenotype associations in HCM is a cumbersome task. Two phenotypic outcomes that can be predicted from mutated genes are the absence or presence of sinus rhythm and the absence or presence of myocardial injury. Models predicting the absence or presence of sinus rhythm had similar performance when they were built using only causative genes and when using all analyzed genes, indicating potential importance of causative genes and irrelevance of non-causative genes for that outcome. On the other hand, models predicting myocardial injury — infarction had better performance when they were built using all analyzed genes (and not just causative ones), indicating a potentially significant role of non-causative genes in that outcome.

The ML algorithms were able to predict phenotypic outcomes — fatigue, dyspnea, chest pain, palpitations, syncope, heart murmur, pretibial edema, SAM, papillary muscle abnormalities, hypokinesia, AF, AV block I, LBBB, RBBB, left anterior hemiblock, ST segment abnormalities, and negative T wave — using genotypic and phenotypic data. Subsets of features that are sufficient for prediction of these phenotypic outcomes were identified and their relative importance was assessed. The combination of a mutation in

TNNT2 and peak RER contributed the most in predicting fatigue. The combination of a mutation in *MYBPC3* and peak VO_2 contributed the most in predicting dyspnea. The combination of a mutation in *TNNI3* and HDL level contributed the most in predicting chest pain. The combination of a mutation in *MYH7* and pacemaker/defibrillator implants in family history, as well as the combination of a mutation in *TNNT2* and LAV, contributed the most in predicting heart murmur. Lastly, the combination of a mutation in *MYBPC3* and MV maxPG aided the most in predicting negative T wave.

Genotype-specific ECHO findings were identified: for mutations in the *MYH7* gene (vs. mutation ND), the most discriminative structures are the septum, LVOT segment, anterior wall, apex, RV, and mitral apparatus; for mutations in the *TNNT2* gene (vs. mutation ND), the most discriminative structures are septum and RV; while for mutations in *MYBPC3* gene (vs. mutation ND) these are septum, LV, and LV chamber.

ML has thus been demonstrated to be useful in deciphering genotype-phenotype associations in HCM.

1. UVOD

1.1. Hipertrofična kardiomiopatija (HCM)

HCM je najčešća nasledna kardiomiopatija (1–5), sa prevalencom 1:500 u opštoj populaciji na globalnom nivou (3,4,6–11), a neke skorašnje studije procenjuju prevalencu HCM i do 1:200 (9,11–16). HCM predstavlja glavni uzrok iznenadne srčane smrti kod pacijenata mlađih od 35 godina (159).

Uglavnom se smatra da se HCM nasleđuje autozomno dominantno (1,11,12,24) i da su uzrok HCM mutacije gena koji kodiraju proteine sarkomere (1,12,25–27), zbog čega se HCM ponekad opisuje kao „bolest sarkomere“ (6,9,16). Međutim, skorašnja istraživanja sugerišu da je genetička podloga HCM daleko kompleksnija nego što se smatralo (12,28).

Dijagnoza HCM se postavlja na osnovu prisustva hipertrofije leve komore i odsustva bilo kog stanja koje bi moglo da uzrokuje hipertrofiju (29–31): kod odraslih osoba na osnovu debljine leve komore ≥ 15 mm (4,32). Uzroci sekundarne hipertrofije leve komore su sistemska hipertenzija, aortna stenoza i infiltrativne kardiomiopatije (3) i svi oni moraju da budu isključeni da bi se postavila dijagnoza HCM (3). Za srodnike prvog stepena pacijenata sa potvrđenom HCM, debljina leve komore ≥ 13 mm je dovoljna za postavljanje dijagnoze (32).

1.2. Sarkomera

Miokard se uglavnom sastoji od srčanih mišićnih ćelija (kardiomiocita) (9). Kardiomiociti sadrže longitudinalno pozicionirane miofibrile. Miofibrili su transverzalno dalje podeljeni na kontraktilne jedinice koje se nazivaju sarkomere (35,36). Sarkomera je ponavljajuća gradivna i kontraktilna jedinica prugastih (srčanih i skeletnih) mišića i ona koordiniše mišićnu kontrakciju (Slika 1) (33,37–40). Funkcija sarkomere je veoma osetljiva na poremećaje bilo koje vrste: narušavanje integriteta bilo koje komponente (postojanje pojedinačnog disfunkcionalnog proteina, promena u strukturi ili dinamici sarkomere, izmenjena ekspresija ili degradacija proteina) može da dovede do kardiomiopatije (39). Kompenzatorni mehanizmi koji se aktiviraju kako bi pomenuti defekti bili prevaziđeni nekada i sami postaju patološki (33).

1.3. Genotip HCM

Genotip je genetička osnova određenog organizma ili karakteristike, nasledni genetički materijal koji kodira sve procese u određenom organizmu. Odnosi se na pojedinačni gen ili kombinaciju alela jedinke (70).

Prva mutacija povezana sa HCM je identifikovana 1989. godine u *MYH7* genu (9,30,36). Do sada je identifikovano > 1400 mutacija povezanih sa HCM, od kojih je oko 90% pronađeno u genima koji kodiraju proteine tankih i debelih filamenata sarkomere (*MYH7*, *MYBPC3* i *TNNT2*) (9,14,31,71,72).

Nosioci mutacije su uglavnom heterozigoti (nose jedan mutantni i jedan zdrav alel). U retkim slučajevima nosioci mutacije mogu da budu i homozigoti, sa ranijim početkom (u detinjstvu) i težim oblicima bolesti. HCM se kod nosilaca mutacije koji su heterozigoti uglavnom javlja u uzrastu 20-50 godina (9). Oko 5% pacijenata ima dve (digenska) ili više (oligogenska) uzročnih mutacija u istom ili različitim genima (26,29).

Prevalenca HCM je postala veća od kada je uvedeno genetičko testiranje koje omogućava postavljanje molekularne dijagnoze pre postavljanja kliničke dijagnoze (15).

1.4. Fenotip HCM

Fenotip (grč. *phainein* — prikazati i *typos* — tip) je skup svih karakteristika organizma, kao što su morfološke, razvojne, biohemijske i fiziološke odlike (70). Kliničke manifestacije i prognoza HCM mogu da budu različite (1,10,11).

Patofiziološke karakteristike HCM su: hipertrofija (90–93) i dezorganizacija (91–94) kardiomiocita, remodelovanje (92,95,96) i fibroza (3,23,92,97–99) miokarda, disfunkcija koronarne mikrocirkulacije (1,92,93,100,101), ishemija (1,18,92,101,102) i hiperkontraktilnost (10,92,103–105) miokarda, poremećaj relaksacije miokarda (10,92,97,106) i dijastolna disfunkcija (92,94,107).

Kod većine pacijenata HCM ostaje asimptomatska ili blago simptomatska (148), dok se kod ostalih javljaju zamor, otežano disanje (u naporu), bol u grudima, palpitacije, presinkopa i sinkopa (1,3,10,25,101,148–151), pri čemu se otežano disanje javlja najčešće, a sinkopa najređe (149).

Kliničke prezentacije HCM su takođe veoma različite (2,3,8,17,144). Kod nekih pacijenata su odsutne (3,5,6,93,144) dok se kod drugih pacijenata HCM ispoljava kao

opstrukcija izlaznog trakta leve komore (17,18,31,92,105,144), atrijalna fibrilacija (3,11,17,92), iznenadna srčana smrt (3,5,8,92,144,146) ili srčana slabost (3,6–8,34,92). Iznenadna srčana smrt može da bude prva manifestacija HCM (1).

1.5. Povezanost genotipa i fenotipa HCM

HCM je heterogena bolest, kako u pogledu genetičkih mutacija, tako i kliničkog toka (6,12,34,86,172). Čak i članovi istih porodica koji nose istu mutaciju često imaju različite manifestacije, progresiju i komplikacije (kod nekih se manifestacije i ne javljaju) (6,15,80,84,168,175). Raznolikost mutacija, relativno mala učestalost svake od mutacija (12,84), kao i nekompletna penetrantnost (15,86,173,174) ometaju utvrđivanje univerzalnih korelacija genotipa i fenotipa. Međutim, neke pravilnosti su identifikovane (84). Prisustvo mutacija u nekom od gena koji kodira proteine sarkomere je povezano sa ranijim početkom HCM (175,176), naglašenijom hipertrofijom leve komore (175) i težim kliničkim formama (2,84).

Fenotip HCM je rezultat doprinosa i uzajamnog dejstva kako genetičkih mutacija tako i drugih faktora (npr. modifikatora bolesti i genske ekspresije, kao i faktora sredine) (29,80,168,174). Moguće je da poremećaj funkcije sarkomere predstavlja obavezan, ali ne nužno inicirajući događaj u patogenezi HCM (80).

1.6. Mašinsko učenje

Mašinsko učenje je grana veštačke inteligencije (198), pri čemu se kompjuterski algoritmi koriste za učenje kompleksnih šablona iz podataka (199). Algoritmi mašinskog učenja sa lakoćom identifikuju trendove i šablone koji ne mogu da budu uočeni golim okom ili korišćenjem klasičnih statističkih tehnika (200). Mašinsko učenje ima potencijal da bude primenjeno u više oblasti medicine (201), uključujući i kardiologiju (199).

Klasterizacija predstavlja podelu podataka na osnovu njihovih karakteristika u grupe (klastere) koji se sastoje od sličnih entiteta (199,202–204). Klasterizacijom se grupisanje vrši na osnovu kriterijuma da entiteti u jednom klasteru budu međusobno maksimalno slični, a u odnosu na entitete u drugim klasterima maksimalno različiti (203,205). Klasterizacija se uspešno koristi za pronalaženje struktura u medicinskim podacima (204). Na primer, identifikacija klastera može da dovede do definisanja novih

podtipova bolesti. Važno je naglasiti da identifikovani klasteri mogu da sugerišu grupe pacijenata, ali da svi uvidi moraju naknadno da budu klinički validirani (199).

Klasifikacija u mašinskom učenju predstavlja identifikaciju (predikciju) kategorije kojoj određeni entitet pripada (206). Klasifikacija se na ovaj način koristi i u biomedicini (207).

Reč „duboko“ u dubokom učenju se odnosi na broj slojeva kroz koji podaci u izvornom obliku bivaju transformisani (163). Struktura dubokih neuronskih mreža omogućava modelovanje nelinearnih zavisnosti (211,213) i integraciju kompleksnih podataka (210,211,213,214). Kroz učenje koje se dešava u slojevima, automatski se pronalaze relevantne karakteristike neophodne za dati zadatak (212,215). Duboko učenje je naročito korisno u klasifikaciji slika (210,214,215).

Uprkos velikim očekivanjima od primene veštačke inteligencije u medicini, postoje brojni izazovi koji moraju da budu prevaziđeni pre nego što se takvi alati mogu primeniti u kliničkoj praksi (212).

2. CILJEVI ISTRAŽIVANJA I HIPOTEZE

Ciljevi istraživanja:

1. Identifikacija subfenotipova hipertrofične kardiomiopatije primenom klasterizacije
2. Utvrđivanje korelacije između genotipova i subfenotipova hipertrofične kardiomiopatije
3. Utvrđivanje povezanosti genotipa i ishoda fenotipa hipertrofične kardiomiopatije uz pomoć algoritama mašinskog učenja
4. Izrada modela mašinskog učenja za predviđanje ishoda hipertrofične kardiomiopatije na osnovu genotipskih i fenotipskih podataka
5. Identifikacija genotip-specifičnih nalaza ehokardiograma hipertrofične kardiomiopatije uz pomoć algoritama mašinskog učenja

Hipoteze:

1. Subfenotipovi hipertrofične kardiomiopatije mogu se identifikovati klasterizacijom
2. Postoji korelacija između genotipova i subfenotipova hipertrofične kardiomiopatije
3. Povezanost genotipa i ishoda fenotipa hipertrofične kardiomiopatije može se utvrditi uz pomoć algoritama mašinskog učenja
4. Modeli mašinskog učenja mogu da predvide ishode hipertrofične kardiomiopatije na osnovu genotipskih i fenotipskih podataka
5. Genotip-specifični nalazi ehokardiograma hipertrofične kardiomiopatije mogu se identifikovati uz pomoć algoritama mašinskog učenja

3. MATERIJALI I METODE

3.1. Dizajn studije

Studija je bila multicentrična i retroprospektivna, pri čemu je retrospektivni period trajao 18, a prospektivni 12 meseci. U studiji je učestvovalo 6 institucija: Institut za kardiovaskularne bolesti Vojvodine (Republika Srbija), Klinički centar Srbije (Republika Srbija), Medicinski fakultet Univerziteta Njukasl (eng. *Newcastle University Medical School*) i Bolnica Njukasl na Tajnu fondacije Nacionalne zdravstvene službe (eng. *Newcastle upon Tyne Hospitals NHS Foundation Trust*) (Ujedinjeno Kraljevstvo), Univerzitetski medicinski centar Regensburg (eng. *Medical Centre Regensburg*) (Savezna Republika Nemačka) i Careggi Univerzitetska bolnica Firenca (eng. *Careggi University Hospital Florence*) (Italija).

Studija je dobila saglasnost etičkih komisija svake od institucija koje su učestvovale u istraživanju. Studija je vršena u skladu sa dobrom kliničkom praksom i Helsinškom deklaracijom.

3.2. Uzorak

Studija je obuhvatila 143 odrasla pacijenta (≥ 18 godina) sa potvrđenom dijagnozom HCM, koji su ispunili uključujuće i isključujuće kriterijume.

Uključujući kriterijumi:

1. potvrđena dijagnoza opstuktivne i/ili neopstruktivne HCM;
2. istorija neobjašnjene hipertrofije leve komore: ehokardiografijom određena maksimalna debljina zida ≥ 15 mm ili 13-14 mm (granična hipertrofija) kod pacijenata koji imaju bar jednog srodnika prvog stepena sa potvrđenom HCM.

Isključujući kriterijumi:

1. redukcija septuma (hirurška intervencija ili intervencija kateterom) u prethodna 3 meseca;
2. klinička dekompenzacija u prethodna 3 meseca; IV klasa simptoma kongestivne srčane slabosti na osnovu klasifikacije Njujorškog kardiološkog društva (*NYHA* klasifikacija);
3. krvni pritisak u mirovanju > 180/100 mmHg;
4. sistolni krvni pritisak < 100 mmHg;
5. gradijent u izlaznom traktu leve komore u mirovanju > 50 mmHg;
6. ejekciona frakcija leve komore određena ehokardiografijom < 50%;
7. pejsmejker ili kardioverter defibrilator implantiran ili zakazan u prethodna 3 meseca;
8. bubrežna insuficijencija sa brzinom glomerularne filtracije (GFR) < 30 mL/min/1.73m^2;
9. trudnoća ili planirana trudnoća;
10. očekivano trajanje života kraće od 12 meseci;
11. indeks telesne mase (BMI) > 40 kg/m^2;
12. istorija sinkope ili komorskih aritmija indukovanih vežbanjem;
13. istorija ili prisustvo bilo koje druge bolesti sa očekivanim trajanjem života kraćim od 3 godine;
14. istorija maligniteta bilo kog organskog sistema (osim lokalizovanog karcinoma bazalnih ili skvamoznih ćelija kože ili lokalizovanog karcinoma prostate), lečenog ili nelečenog, u prethodne 2 godine, bez obzira na to da li postoji dokaz o lokalnoj rekurenciji ili metastazama;
15. životno ugrožavajuća ili nekontrolisana disritmija, uključujući simptomatsku ili dugotrajnu komorsku tahikardiju, fibrilaciju ili lepršanje pretkomora uz komorsku frekvenciju u mirovanju > 110 otkucaja u minuti;

16. učešće u takmičarskim ili organizovanim sportskim aktivnostima (kao što su fudbal, košarka, itd.), eksplozivnim aktivnostima (kao što su sprint, sportovi sa reketom, itd.), ili naporno izometrijsko vežbanje (kao što je bodibilding) ili protivljenje uzdržavanju od istog tokom trajanja studije.

Ispunjenost kriterijuma za uključivanje u studiju utvrđivala se na osnovu uvida u elektronski karton pacijenta. Učesnici istraživanja su upoznati sa ciljevima doktorske disertacije, a potpisivanjem informisanog pristanka dali su saglasnost za učešće u istraživanju. Podaci i rezultati učesnika su preuzeti iz bolničkog informacionog sistema.

Svi učesnici studije su prošli kroz genetičko savetovanje, kada im je nedvosmisleno naglašen značaj daljeg kardiološkog skininga unutar porodice.

3.3. Protokol

3.3.1. Povezanost genotipa i fenotipa HCM: Ispitivanje korišćenjem algoritama mašinskog učenja i kliničkih podataka

3.3.1.1. Podaci

3.3.1.1.1. Demografski podaci

Evidentirani su pol i uzrast pacijenata.

3.3.1.1.2. Anamnestički podaci

Evidentirani su trenutni simptomi, znaci, *NYHA* klasa, sistolni i dijastolni krvni pritisak, kao i komorbiditeti. Takođe su evidentirane porodična istorija HCM, dilatativne kardiomiopatije, iznenadne srčane smrti u uzrastu < 40 godina, 40-59 godina i ≥ 60 godina, neobjašnjene srčane slabosti, implantacija pejsmejkera ili defibrilatora, kao sistemskih bolesti.

3.3.1.1.3. Antropometrijska merenja

Visina tela je određena stadiometrom, a težina tela vagom. Izračunat je BMI.

3.3.1.1.4. Uzorkovanje krvi

Uzorci krvi su dobijeni venepunkcijom antekubitalne vene, u jutarnjim časovima, nakon noćnog gladovanja.

3.3.1.1.5. Genetičko testiranje

Nakon izolacije DNK komercijalnim kitom QIAamp DNA Blood BioRobot MDx kit (QIAGEN GmbH, Hilden, Savezna Republika Nemačka), panel od 41 gena je analiziran polimeraza lančanom reakcijom (eng. *polymerase chain reaction*, PCR), kao i sekvencioniranjem naredne generacije na platformi Next Generation Sequencing (Illumina, Inc., San Diego, USA). Za ovaj panel, uzorak predstavlja DNK izolovana iz krvi pacijenata. Analizirani geni su deo komercijalno dostupnog panela TruSight Cardio Sequencing Kit (Illumina, Inc., San Diego, SAD), za ciljano sekvencioniranje u naučnoistraživačke svrhe.

3.3.1.1.5.1. Genetički panel

Osnova analiziranog panela se sastoji od: *ACTC1*, *ACTN2*, *ANKRD1*, *CSRP3*, *FHL1*, *GLA*, *LAMP2*, *MYBPC3*, *MYH7*, *MYL2*, *MYL3*, *PLN*, *PRKAG2*, *TNNI3*, *TNNT2* i *TPM1*. Ovi geni pripadaju panelu gena za HCM koji se analiziraju iz uzorka krvi u Medicinskoj genetičkoj laboratoriji Oksford (Oxford Medical Genetics Laboratory, OMGL). Ova osnova panela gena je dizajnirana tako da obuhvati gene za koje je definitivno pokazano da izazivaju HCM (*MYBPC3*, *MYH7*, *TNNI3*, *TNNT2*, *MYL2*, *MYL3*, *ACTC1*, *TPM1*) — prema Vodiču Fondacije Američkog koledža za kardiologiju (*American College of Cardiology Foundation, ACCF*) i Američke asocijacije za srce (*American Heart Association, AHA*) za dijagnozu i tretman HCM iz 2011. godine; zatim gene za diferencijalnu dijagnozu (*PRKAG2*, *GLA*, *FHL1*) i druge validirane gene povezane sa HCM (22). Pored toga analizirani su i geni: *ABCC9*, *CACNA1C*, *CTF1*, *DMD*, *DSC2*, *DSG2*, *DSP*, *ELN*, *FBN*, *HCN4*, *JPH2*, *LAMA4*, *MYH6*, *MYPN*, *NEXN*, *NKX2-5*, *NOTCH1*, *PDLIM3*, *PKP2*, *PTPN11*, *RBM2*, *RYR1*, *SDHA*, *TRPM4* i *TTN*. Svi ovi geni su deo panela Illumina TruSight Cardio Sequencing Panel, komercijalno dostupnog panela za ciljano sekvencioniranje gena koji su povezani sa naslednim

stanjima srca u naučnoistraživačke svrhe. Za sve analizirane gene, uzorak predstavlja DNK izolovana iz krvi pacijenata.

Ovo istraživanje se fokusiralo na gene za koje je pokazano da izazivaju HCM. Međutim, prilikom kreiranja modela za predikciju ishoda fenotipa korišćenjem isključivo genetičkih podataka, korišćeni su svi navedeni geni, kako bi se dobila kompletnija informacija koja potencijalno pruža bolju osnovu za predikciju.

3.3.1.1.6. Hematološke i biohemijske analize krvi

Kompletna krvna slika je urađena iz uzoraka pune krvi, određivanjem na hematološkom brojaču. Biohemijske analize su urađene iz krvnog seruma i plazme. Za dobijanje seruma, uzorci pune krvi su nakon potpune koagulacije bili centrifugirani na 3500 obrtaja/minuti tokom 15 minuta. Za dobijanje plazme, uzorci pune krvi su bili centrifugirani na 3500 obrtaja/minuti, tokom 15 minuta.

3.3.1.1.7. Transtorakalna ehokardiografija sa doplerom

Ehokardiogrami su snimljeni u realnom vremenu tokom 3 srčana ciklusa pregledanjem sledećih standardnih srčanih preseka: parastenalni (duga osa) i apikalni (4 šupljine, 2 šupljine i duga osa). Parasternalni prikazi kratke ose su zabeleženi u tri nivoa: bazalnom (na nivou mitralne valvule), srednjem papilarnom i apikalnom. Maksimalna brzina u izlaznom traktu leve komore bila je registrovana korišćenjem pulsnog doplera, u apikalnom 5-šupljinskom preseku, i korišćena za računanje gradijenta pritiska nad izlaznim traktom leve komore (eng. *left ventricular outflow tract*, LVOT), ukoliko on postoji.

3.3.1.1.8. Kardiopulmonalni test opterećenjem

Kardiopulmonalni test opterećenjem je izvršen korišćenjem bicikl-ergometra kako bi se evaluirala tolerancija na napor.

3.3.1.1.9. Elektrokardiografija i EKG-holter-monitoring

Elektrokardiogram (EKG) je zabeležen korišćenjem 12-kanalnog elektrokardiografa, u ležećem položaju. Za uočavanje sporadičnih aritmija, učesnici su nosili EKG-holter-monitor 24 časa i vodili dnevnik simptoma i aktivnosti.

3.3.1.2. Analiza podataka

3.3.1.2.1. Priprema podataka za analizu

Najveći deo pripreme i transformacije podataka iz izvornog oblika u format koji je pogodan za dalju analizu, kao i istraživanje podataka (eng. *data exploration*) izvršeno je korišćenjem biblioteke Pandas v. 1.4.3. programskog jezika Python (283).

3.3.1.2.2. Identifikacija subfenotipova HCM primenom klasterizacije

Pošto klasterizacija predstavlja podelu podataka u grupe (klastere) sličnih entiteta, a klasterizacija je u ovom istraživanju izvršena na osnovu fenotipskih parametara, grupe sličnih pacijenata su određene na osnovu parametara fenotipa. Stoga, identifikovani klasteri predstavljaju subfenotipove HCM, a termini klasteri i subfenotipovi su korišćeni sa istim značenjem.

Adekvatno pripremljeni parametri fenotipa su analizirani algoritmima za klasterizaciju (nenadgledano mašinsko učenje). Hijerarhijska klasterizacija (agglomerative clustering, affinity=”euclidean”, linkage=”ward”) je izvršena korišćenjem modula Scikit-learn v. 1.1.1. i biblioteke SciPy v. 1.8.1, a dendrogram je vizuelizovan korišćenjem MatPlotLib v. 3.5.2 (283).

Samo prve posete pacijenata su bile uključene u klasterizaciju. Podaci o drugoj poseti bi interferirali sa klasterizacijom, pa su izostavljeni iz ove analize. Parametri kojima je nedostajalo više od 30% vrednosti bili su uklonjeni iz skupa podataka, a druge nedostajuće vrednosti su imputirane pomoću Scikit-learn KNNImputer-a (n_neighbors = 12, weights = “uniform”). Numerički parametri su standardizovani pomoću Scikit-learn StandardScaler-a (283).

Da bi se dalja manipulacija podacima svela na minimum, a kako su korišćeni i numerički i kategorički parametri, za klasterizaciju je korišćen K-Prototype algoritam Kmodes paketa (pogodan za slučajeve kada skup podataka sadrži i kategoričke i

numeričke podatke). Za pronalaženje optimalnog broja klastera korišćena je metoda lakta, a određena vrednost je potvrđena korišćenjem Kneelocator-a (https://pypi.org/project/kneed/). Karakteristike klastera su vizuelizovane korišćenjem biblioteke Seaborn library v. 0.11.2 (283).

Pošto u ovom trenutku ne postoji metoda za direktnu interpretaciju K-Prototype klasterizacije, korišćene su dve indirektne metode. U prvoj je izvršena klasifikacija pomoću algoritma stablo odlučivanja (gde su klase dodeljene na način na koji su određene u klasterizaciji, a zatim je klasifikacija izvršena korišćenjem svih podataka koji su korišćeni u klasterizaciji) i generisan je dendrogram (korišćenjem sklearn.tree.plot_tree). Druga metodu je činilo izračunavanje značaja karakteristika za klasifikaciju pomoću algoritma slučajna šuma (korišćenjem atributa slučajne šume, feature_importances_), nakon klasterizacije (klase su dodeljene na način na koji su određene u klasterizaciji). Pošto stabla odlučivanja donose lokalno optimalne izbore, klasifikacija algoritmom slučajne šume je izvršena za postojaniju procenu globalnog značaja karakteristika (283).

3.3.1.2.3. Identifikacija povezanosti genotipa i ishoda fenotipa HCM i kreiranje modela za predikciju ishoda na osnovu parametara genotipa i fenotipa

Povezanost genotipa i ishoda fenotipa je evaluirana kreiranjem modela za predviđanje ishoda fenotipa korišćenjem Python modula Scikit-learn i SHapley Additive exPlanation (SHAP). Za generisanje veštačkih podataka za manjinsku klasu u klasifikaciji, korišćena je tehnika SMOTE (eng. *Synthetic Minority Oversampling Technique*) iz Imbalanced-learn biblioteke. Izbor karakteristika izvršen je korišćenjem Scikit-learn SelectKBest-a (score_func = f_classif), SelectPercentile-om, Scikit-learn VarianceThreshold-om (threshold = 0.02), rekurzivnom eliminacijom karakteristika, i na osnovu domenskog znanja. Doprinos svake od karakteristika predikcijama modela procenjen je upotrebom SHAP modula v. 0.41.0 (283).

Predikcija prisustva različitih ishoda fenotipa je izvršena kao zadatak klasifikacije. Ukupno 268 poseta je obuhvaćeno analizom. Za većinu pacijenata bili su dostupni podaci o 2 posete. Za manji broj pacijenata prikupljeni su samo podaci o prvoj poseti, zbog gubitka ovih pacijenata iz praćenja (283).

Vršen je inženjering karakteristika: numerički parametri su kombinovani međusobnim deljenjem, kategorički međusobnim množenjem, a kategorički i numerički međusobno množenjem. Množenjem ili sabiranjem karakteristika koje zajedno čine smisleni klinički entitet kreirane su i neke dodatne karakteristike (283).

Parametri kojima nedostajalo više od 30% vrednosti bili su isključeni iz dalje analize, a druge nedostajuće vrednosti su imputirane pomoću Scikit-learn KNNImputer-a (n_neighbors = 12, weights = "uniform"). Numerički parametri su standardizovani pomoću Scikit-learn StandardScaler-a. Imputacija i standardizacija su bile izvršene u okviru iste protične obrade podataka, a protočna obrada je izvršena posebno na podacima za trening i posebno na podacima za testiranje (283). Podaci za trening su se sastojali od 188 (75.80%) poseta. U slučajevima kada su bili dostupni podaci za obe posete, obe posete istog pacijenta su se nalazile ili u podacima za trening ili u podacima za testiranje (284).

Za svaki zadatak klasifikacije TPOT (eng. tree-based pipeline optimization) je takođe korišćena (285).

Primenjena je petostruka unakrsna validacija. Tačnost, preciznost, odziv, F1-skor, površina ispod ROC (eng. *Receiver Operating Characteristics*) krive i prosečna preciznost su sve korišćene kao mere performanse (283).

Globalni doprinos karakteristika predikcijama je određen kao srednja apsolutna vrednost Shapley vrednosti za svaku od karakteristika korišćenjem SHAP v. 0.41.0 (283).

3.3.1.2.4. Identifikacija genotip-specifičnih nalaza ehokardiograma HCM

Ultrazvučni pregledi u DICOM formatu su konvertovani u JPG format pomoću RadiAnt DICOM Viewer v. 2021.2.2. Slike su grupisane na osnovu preseka u kojima su snimljene. Ultrazvučne slike koje prikazuju početak P-talasa i T-talasa EKG-a (kao reprezentativne slike dijastole komora i sistole komora) bile su odvojene svaka u posebnu grupu slika. Kako bi se uklonio pozadinski šum na slikama (podaci o pacijentu, datum, EKG), korišćen je namenski kreirani prekrivač slika napravljen pomoću Python OpenCV biblioteke. Prekrivene ultrazvučne slike, koje prikazuju samo region od interesa (Slika 3), korišćene su u daljoj analizi.

Genotip-specifični nalazi ehokardiograma su identifikovani pomoću Python biblioteke za duboko učenje i računarski vid Fast AI. Ovo je ostvareno kreiranjem modela dubokog učenja za klasifikaciju ultrazvučnih slika, prema klasama koje su kreirane na osnovu podataka o genotipu i naknadnom analizom regiona slika koji su najviše uticali prilikom predikcije modela. Prilikom kreiranja ovih modela bio je konsultovan Autokeras. Modeli su kreirani prema sledećim podešavanjima: image_size = 224, batch_size = 4, validation_percentage = 0.15, bez augmentacije podataka, i primenom prethodno trenirane ResNet18.

3.3.2. Genetički i molekularni mehanizmi povezanosti genotipa i fenotipa HCM: Proučavanje pomoću mašina korišćenjem litarature

3.3.2.1. „HCM Map“: Interaktivni resurs o genetičkim i molekularnim mehanizmima HCM

„HCM Map“ je kreiran u SBML jeziku (eng. *Systems Biology Markup Language*) korišćenjem CellDesigner softvera v. 4.4.2. „HCM Map“ je vizuelizovan kao SBGN (eng. *Systems Biology Graphical Notation*) dijagram i javno dostupan pomoću MINERVA (eng. *Molecular Interaction NEtwoRks VisuAlization*) platforme (287).

„HCM Map“ interaktivni resurs je kreiran manuelno, unošenjem molekularnih interakcija koje su pronađene u naučnim radovima o molekularnim mehanizmima u humanoj HCM koji se nalaze u PubMed bazi podataka. Kreiranje je započelo opsežnom pretragom literature koja se nalazi u PubMed bazi, na osnovu relevantnih ključnih fraza poput „gene hypertrophic cardiomyopathy“, „signaling hypertrophic cardiomyopathy“, „micro RNA hypertrophic cardiomyopathy“, „noncoding RNA hypertrophic cardiomyopathy“, itd. Filter “10 years” (za obuhvatanje perioda od 2010-2020. godine) je primenjen prilikom odabira literature (287).

3.3.2.2. Genetički i molekularni mehanizmi povezanosti genotipa i fenotipa HCM prikupljeni i prikazani od strane mašina

Genetički i molekularni aspekti HCM i njenih kliničkih prezentacija koji se nalaze u naučnim radovima su automatski prikupljeni, a 4 modela su kreirana i analizirana.

PubMed HCM model sastavljen pomoću INDRA-e je automatski kreiran, korišćenjem INDRA PubMed klijenta za literaturu, termina za pretragu "hypertrophic cardiomyopathy" (major_topic = True), i odbacivanjem rezultata koji su stariji od 1. januara 2010. godine (287). Literatura je automatski pročitana od strane REACH (eng. *REading and Assembling Contextual and Holistic mechanisms from text*) sistema za čitanje (288). Izdvojene interakcije su zatim bile mapirane, deduplicirane i organizovane u hijerarhiju, a zatim sastavljene u model pomoću Cytoscape v. 3.8.2 (287).

PubMed+PathwayCommons HCM model sastavljen pomoću INDRA-e je automatski generisan na osnovu podataka u Pathway Commons bazi podataka preko BioPAX API INDRA-e i PubMed klijenta za literaturu INDRA-e, termina za pretragu "hypertrophic cardiomyopathy" (major_topic = True), i odbacivanjem rezultata koji su stariji od 1. januara 2010. godine (287). Literatura je automatski pročitana od strane REACH sitema za čitanje (288). Izdvojene interakcije su bile mapirane, deduplicirane i organizovane u hijerarhiju, a zatim sastavljene u model pomoću Cytoscape v. 3.8.2 (287).

Interakcije za skraćeni INDRA DB model su pronađene u INDRA bazi podataka primenom ograničenja upita na MeSH „Cardiomyopathy, Hypertrophic, Familial". Samo u potpunosti tačno izdvojene interakcije su unete u model. Kriterijum za tačnost je bio da su svi aspekti interakcije (oznake, subjekat, objekat, smer i tip interakcije) isti kao što bi bili na osnovu pažljivog maneulnog odabira. Interakcije su zatim manuelno prevedene u mrežnu tabelu u XLSX formatu (287).

Interakcije za INDRA DB model su pronađene u INDRA bazi podataka primenom ograničenja upita na MeSH „Cardiomyopathy, Hypertrophic, Familial". Sve pronađene interakcije su unete u model. Interakcije su prevedene u mrežnu tabelu u XLSX formatu (287).

3.3.2.3. Genetički i molekularni mehanizmi različitih kliničkih prezentacija HCM prikupljeni i prikazani od strane mašina

Genetički i molekularni mehanizmi HCM i njenih različitih kliničkih prezentacija su prikupljeni automatski iz INDRA baze podataka (227). Molekularni mehanizmi su pronađeni u svim naučnim radovima iz PubMed baze podataka koji su objavljeni počevši od 1. januara 2010. godine (296), posebno za HCM i svaku od njenih 19 kliničkih

prezentacija. Interakcije su pronađene u INDRA bazi podataka na osnovu PubMed identifikatora (PMID), a oni su prikupljeni pomoću INDRA PubMed klijenta (koji pretražuje članke na PubMed-u) (228) i termina za pretragu: „hypertrophic cardiomyopathy“, „myofibrillar disarray“, „cardiomyocyte disarray“, „cardiomyocyte hypertrophy“, „myocardial fibrosis“, „myocardial hypercontractility“, „impaired myocardial relaxation“, „myocardial stiffness“, „myocardial remodeling“, „cardiac remodeling“, „impaired cardiac relaxation“, „diastolic dysfunction“, „coronary microvascular dysfunction“, „myocardial ischemia“, „left ventricular outflow tract obstruction“, „atrial fibrillation“, „sudden cardiac death“, „heart failure“, „major adverse cardiovascular events“ i „rehospitalization“ (use_text_word = True, major_topic = True) (92).

Preseci skupova pronađenih interakcija za HCM i njene kliničke prezentacije su prevedeni u mrežnu tabelu, vizualizovani i analizirani pomoću Cytoscape v. 3.8.2 (92,289) i učinjeni javno dostupnim pomoću NDEx v 2.5.0 (290–292).

3.3.2.4. „HCM clinical“ interaktivni resurs o kliničkom toku HCM

„HCM Clinical“ interaktivni resurs je napravljen pomoću CellDesigner v. 4.4.2, pregledom literature i istraživanjem baza podataka, kao i objedinjavanjem pronađenih elemenata. Učinjen je javno dostupnim pomoću MINERVA platforme v. 15.1.2.

3.4. Statistička obrada podataka

Za statističku obradu podataka korišćen je licencirani program za statističku analizu SPSS v. 28.0.1.1. (IBM SPSS Statistics, Armonk, New York). Radi utvrđivanja statističke značajnosti, razlika u srednjim vrednostima kontinuiranih varijabli korišćena je ANOVA, a kategoričke varijable su bile upoređene chi-square testom. Za statistički značajnu odabrana je p vrednost < 0.05. Phi koeficijent je korišćen kao mera povezanosti između prisustva/odsustva mutacije u genima i prisustva/odsustva ishoda fenotipa. Procena slaganja između stvarnih i određenih klastera je vršena pomoću Cohen’s kappa koeficijenta.

4. REZULTATI

4.1. Povezanost genotipa i fenotipa HCM: Ispitivanje korišćenjem algoritama mašinskog učenja i kliničkih podataka

4.1.1. Identifikacija subfenotipova HCM primenom klasterizacije

Klasteri HCM pacijenata su određeni na osnovu parametara fenotipa (283).

4.1.1.1. Hierarhijska klasterizacija

Kao rezultat hijerarhijske klasterizacije generisan je dendrogram (dijagram u obliku drveta, na kome su nasličniji pacijenti smešteni na granama koje su u međusobnoj blizini) (Slika 4).

Klasteri se na osnovu dendrograma određuju pravljenjem preseka na različitim horizontalnim nivoima. U ovom slučaju, najprominentnije agregacije se uočavaju ako su pacijenti podeljeni u 2, 4 ili 6 klastera.

4.1.1.2. K-Prototype klasterizacija

Za ovaj skup podataka o fenotipu određeno je da je optimalan broj klastera 4. Značaj svake od karakteristika je indirektno procenjen (Slika 63, Tabela 12).

4.1.2. Povezanost genotipa i fenotipa HCM

Nije pronađena povezanost između klastera određenih hijerarhijskom klasterizacijom i genotipa (Slika 120).

Neke statistički značajne korelacije su pronađene između klastera određenih K-Prototype klasterizacijom i mutiranih gena (Slike 121-126). Iako niti jedna od određenih postavki subfenotipova ne korelira u potpunosti sa genotipom, neke značajne korelacije su pronađene u postavkama sa 3 subfenotipa (*TNNI3*, $p = 0.041$), 4 subfenotipa (*TNNI3*, $p = 0.045$ i *MYBPC3*, $p = 0.038$), 5 subfenotipa (*MYH7*, $p = 0.045$) i 6 subfenotipa (*MYBPC3*, $p = 0.044$).

4.1.3. Predikcija ishoda fenotipa isključivo na osnovu genetičkih podataka

Predviđanje dva ishoda fenotipa je moguće na osnovu podataka o mutiranim genima: prisustvo ili odsustvo sinusnog ritma (Tabele 49 i 51) kao i prisustvo ili odsustvo

oštećenja/infarkta miokarda (Tabele 102 i 104). Modeli koji vrše predikciju odsustva ili prisustva sinusnog ritma imali su slične performanse kada su kreirani samo na osnovu uzročnih gena i na osnovu svih analiziranih gena. Sa druge strane, modeli koji vrše predikciju oštećenja/infarkta miokarda imali su bolje performanse kada su kreirani na osnovu svih analiziranih gena (a ne samo na osnovu uzročnih gena).

4.1.4. Predikcija ishoda fenotipa na osnovu genetičkih podataka i drugih parametara fenotipa

Algoritmi mašinskog učenja mogu da vrše predikciju sledećih ishoda fenotipa — zamor, otežano disanje, bol u grudima, palpitacije, sinkopa, šum na srcu, pretibijalni edem, pokretanje mitralnog zalistka unapred, abnormalnosti papilarnih mišića, hipokinezija, atrijalna fibrilacija, AV blok prvog stepena, blok leve grane (LBBB), blok desne grane (RBBB), prednji levi hemiblok, abnormalnosti ST segmenta, i negativni T talas (Slike 127-143) — na osnovu podataka o genotipu i fenotipu, imajući pri tome odlične performanse (Tabela 106) (283).

Prilikom predviđanja zamora, najveći doprinos je imala kombinacija mutacije u *TNNT2* i maksimalnog odnosa disajne razmene (RER) (Slika 127). Prilikom predviđanja otežanog disanja najveći doprinos imala je kombinacija mutacije u *MYBPC3* i vršne potrošnje kiseonika (peak VO_2) (Slika 128). Prilikom predviđanja bola u grudima, najveći doprinos je imala kombinacija mutacije u *TNNI3* i koncentracije lipoproteina visoke gustine (eng. high-density lipoprotein, HDL) (Slika 129). Prilikom predviđanja šuma na srcu najveći doprinos imala je kombinacija mutacije u *MYH7* i podatka o implantiranju pejsmejkera/defibrilatora u porodičnoj istoriji, kao i kombinacija mutacije u *TNNT2* i zapremine leve pretkomore (LAV) (Slika 132). Prilikom predviđanja negativnog T talasa, najveći doprinos imala je kombinacija mutacije u *MYBPC3* i vrednosti transmitralnog maksimalnog gradijenta pritiska (MV maxPG) (Slika 143).

4.1.5. Identifikacija genotip-specifičnih nalaza ehokardiograma HCM

Identifikovani su genotip-specifični nalazi ehokardiograma: za mutaciju u *MYH7* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima), strukture koje najviše utiču na raspoznavanje su septum, izlazni trakt leve komore (LVOT), prednji zid,

vrh srca, desna komora i mitralni aparat (Slike 145, 147, 149, 151, 153, 155); za mutaciju u *TNNT2* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima) strukture koje najviše utiču na raspoznavanje su septum i desna komora (Slike 157, 159, 161, 163, 165, 167); dok su za mutaciju u *MYBPC3* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima) ove strukture septum, leva komora i šupljina leve komore (Slike 169, 171, 173, 175, 177, 179). Kreirani modeli imaju odlične performanse (Tabele 107-124).

4.2. Genetički i molekularni mehanizmi povezanosti genotipa i fenotipa HCM: Proučavanje pomoću mašina korišćenjem litarature

4.2.1. „HCM Map“: Interaktivni resurs o genetičkim i molekularnim mehanizmima HCM

„HCM Map“ predstavlja interaktivni pregled svih važnih molekularnih elemenata uključenih u HCM, detaljni vizuelni intuitivni prikaz genetičkih i molekularnih mehanizama HCM (Slika 182) i platformu za dalje *in silico* proučavanje, dostupan na: https://silicofcm.eu/interactive-map/. Predstavljeni elementi poseduju linkove ka njihovim opisima u biomedicinskim bazama podataka (Slika 183) (286).

4.2.2. Genetički i molekularni mehanizmi povezanosti genotipa i fenotipa HCM prikupljeni i prikazani od strane mašina

Četiri modela koja predstavljaju genetičke i molekularne mehanizme HCM su kreirana automatski, korišćenjem različitih komjuterski potpomognutih pristupa, i učinjeni su javno dostupnim (Tabela 125). U okviru analize mreže koja je usledila, određene su mere centralnosti čvorova, kao i najznačajniji elementi mreže (Tabela 126). Konsenzus o najvažnijim elementima je postignut samo u vezi sa jednim elementom (kalcijumom) (287).

4.2.3. Genetički i molekularni mehanizmi različitih kliničkih prezentacija HCM prikupljeni i prikazani od strane mašina

Genetički i molekularni mehanizmi su automatski prikupljeni iz 230.072 naučna rada o HCM i 19 kliničkih prezentacija HCM u formi 182.167 INDRA interakcija —

reprezentacije molekularnih mehanizama koje se sastoje od subjekta, objekta i i njihove interakcije (Tabela 128). Zajednički genetički i molekularni mehanizmi HCM i njenih kliničkih prezentacija predstavljeni su u vidu mreža koje su dobijene kao presek skupova interakcija HCM i svake od njenih kliničkih prezentacija (Tabela 129). Na osnovu automatski prikupljenih mehanizma i naknadne analize, prikazani su molekularni putevi koji su procenjeni kao najvažniji (Slika 200) (92).

4.2.4. „HCM clinical“ interaktivni resurs o kliničkom toku HCM

„HCM Clinical“ interaktivni resurs prikazuje kliničke aspekte HCM: dijagnozu (uključujući i diferencijalnu genetičku dijagnozu), monitoring i lečenje (Slika 201). Jedan deo „HCM Clinical“ u najvažnijim crtama prikazuje klinički tok HCM (Slika 202). „HCM Clinical“ interaktivni resurs je javno dostupan: https://silicofcm.eu/hcm-clinical-interactive/.

5. DISKUSIJA

5.1. Povezanost genotipa i fenotipa HCM: Ispitivanje korišćenjem algoritama mašinskog učenja i kliničkih podataka

5.1.1. Identifikacija subfenotipova HCM primenom klasterizacije

Slučajevi HCM se uglavnom kategorišu u podtipove na osnovu mesta na kojima se hipertrofija javila (npr. bazalna, apikalna) (299). Postoje 2 tipa HCM: opstruktivna (oko 70% slučajeva) i neopstruktivna (oko 30% slučajeva) (300). Radna grupa Evropskog kardiološkog društva (eng. *The European Society of Cardiology Working Group*) preporučuje klasifikaciju HCM na osnovu morfologije (32), dok Radna grupa Američkog udruženja za srce (eng. *American Heart Association Working Group*) sugeriše da bi HCM trebalo da bude definisana na osnovu njene genetičke osnove (301). 1981. godine, Maron *et al*. opisali su 4 tipa HCM, na osnovu pozicije hipertrofije (302,303). Helmy *et al*. predložili su klasifikaciju HCM na osnovu različitih šablona hipertrofije (303,304). Syed *et al*. su nagovestili postojanje najmanje 5 glavnih podtipova HCM koji se razlikuju na osnovu septalne konture, stepena i pozicije hipertrofije (303,304). Tang *et al*. pokazali su da različiti morfološki tipovi HCM imaju različite prognoze nakon hirurške mijektomije (299).

Rezultati ovog istraživanja sugerišu postojanje 4 podtipa HCM: klaster 0 (“AHOLD”), koji se razlikuje od ostalih na osnovu prečnika korena aorte (AO) i laktat dehidrogenaze (LDH), pri čemu su vrednosti AO > 30 mm i LDH > 300 U/L; klaster 1 (“RVSP ASCAOVS”), koji se razlikuje od ostalih na osnovu sistolnog pritiska desne komore (RVSP), dijametra ascedentne aorte (AscAO), i separacije aortnih kuspisa (AOvs), pri čemu su vrednosti AOvs > 27 m/s, AscAO < 31 mm i RVSP < 28 mmHg; klaster 2 (“weight”), koji se razlikuje od ostalih na osnovu telesne težine, sa vrednošću > 95 kg; i klaster 3 (“AV LVOT PG”) koji se razlikuje od ostalih na osnovu srednjeg gradijenta pritisaka nad aortnom valvulom (AV meanPG), maksimalnog gradijenta pritisaka nad aortnom valvulom (AV maxPG), i maksimalnog gradijenta pritisaka nad izlaznim traktom leve komore (LVOT maxPG), pri čemu su vrednosti AV maxPG > 15 mmHg, AV meanPG > 6 mmHg, i LVOT maxPG > 15 mmHg (283).

Iako je dosta statistički značajnih razlika pronađeno između klastera za mnoge parametre, za neke od njih, preklapanje intervala vrednosti ometa upotrebu u međusobnom razlikovanju klastera (283).

5.1.2. Korelacije genotipa i određenih klastera

Odsustvo korelacija između pripadnosti određenom klasteru i genotipa u nekim od postavki ne znači da povezanost genotipa i fenotipa ne postoji. Ono pokazuje da su se algoritmi prilikom klasterizacije najviše oslanjali na određene parametre kako bi odredili optimalne klastere HCM pacijenata i da na taj način određeni klasteri ne koreliraju u potpunosti sa genotipom.

5.1.3. Povezanost genotipa i fenotipa HCM

Algoritmi mašinskog učenja su uglavnom imali poteškoće prilikom vršenja predikcije ishoda fenotipa isključivo na osnovu genetičkih podataka — kreirani modeli su uglavnom imali neadekvatne performanse, naročito kada se uzme u obzir da je prosečna preciznost prikladnija mera performanse od površine ispod ROC (eng. *Receiver Operating Characteristics*) krive kada postoji neravnoteža u zastupljenosti klasa u skupu podataka (306,307). Kod takvih modela takođe je primenjen i AutoML kako pre, tako i posle naduzorkovanja (eng. *oversampling*), bez značajnog uspeha.

Modeli koji predviđaju prisustvo ili odsustvo sinusnog ritma imali su slične performanse kada su bili kreirani samo na osnovu uzročnih gena i kada su bili kreirani na osnovu svih analiziranih gena. Model sa najboljim performansama za predikciju prisustva ili odstustva sinusnog ritma se zasniva na logističkoj regresiji sa naduzorkovanjem i ima tačnost 0,83, F1-skor 0,91 i prosečnu preciznost 0,86 — u ovom slučaju prosečna preciznost je prikladna mera, jer je su klase različito zastupljene (306,307). Nekoliko drugih modela na bazi drugih algoritama mašinskog učenja imali su slične, ali nešto lošije performanse.

Modeli koji vrše predikciju oštećenja/infarkta miokarda imali su bolje performanse kada su kreirani korišćenjem svih analiziranih gena (a ne samo uzročnih gena). Model sa najboljim performansama za predikciju oštećenja/infarkta miokarda se zasniva na logističkoj regresiji sa naduzorkovanjem, i ima tačnost 0,73, F1-skor 0,84 i prosečnu preciznost 0,78 — u ovom slučaju prosečna preciznost je prikladna mera, jer je su klase različito zastupljene (306,307). Nekoliko drugih modela na bazi drugih algoritama mašinskog učenja imali su slične, ali nešto lošije performanse.

5.1.4. Identifikacija genotip-specifičnih nalaza ehokardiograma HCM

U studiji Chung *et al*., kod 212 pacijenata, patogene varijante u genima koji su povezani sa sarkomerom su češće kod ne-apikalnih formi HCM nego kod apikalnih formi HCM i učestalost mutacija u *MYH7* genu je bila 9% kod pacijenata sa apikalnom formom HCM (337). Waldmüller *et al*. opisali su vezu imeđu mutacija u *MYH7* i većeg stepena regurgitacije mitralne valvule u HCM (338). U studiji Mori *et al.*, pokazano je da je *TNNT2* p.Lys263Arg povezana sa povećanom debljinom septuma (339). Waldmüller *et al*. su pronašli vezu između mutacije u *MYBPC3* i naročito velike debljine interventrikularnog septuma (338).

5.2. Genetički i molekularni mehanizmi povezanosti genotipa i fenotipa HCM: Proučavanje pomoću mašina korišćenjem litarature

5.2.1. „HCM Map“: Interaktivni resurs o genetičkim i molekularnim mehanizmima HCM

„HCM map“ predstavlja skup trenutnog znanja o molekularnim mehanizmima HCM. Međutim, očekuje se da će mnoge molekularne interakcije HCM tek biti otkrivene. „HCM Map“ takođe predstavlja platformu za deljenje i prikupljanje molekularnih mehanizama HCM (287).

5.2.2. Genetički i molekularni mehanizmi povezanosti genotipa i fenotipa HCM prikupljeni i prikazani od strane mašina

Reprezentacije molekularnih mehanizama koje su kreirane automatski predstavljaju bogat i veoma detaljan izvor informacija. Međutim, ovaj pristup takođe donosi određeni nivo šuma (unošenje elemenata koji nisu molekuli) i netačnosti (pogrešno obeleženi elementi) u prikaze molekularnih aspekata bolesti (287).

5.2.3. Genetički i molekularni mehanizmi različitih kliničkih prezentacija HCM prikupljeni i prikazani od strane mašina

Prikupljeni i analizirani genetički i molekularni mehanizmi ukazuju da hipertrofija kardiomiocita, fibroza i remodelovanje miokarda; disfunkcija koronarne mikrocirkulacije i ishemija miokarda; ishemija miokarda i srčana slabost; atrijalna fibrilacija i iznenadna srčana smrt dele slične molekularne mehanizme (92). Sugerisani zajednički molekularni putevi su u skladu sa kliničkim istraživanjima o progresiji HCM (341,342), ishemijskim karakterom i povezanošću disfunkcije koronarne mikrocirkulacije i ishemije miokarda (1,343,344), povezanošću ishemije miokarda i srčane slabosti u HCM (345,346), aritmijskom prirodom i povezanošću atrijalne fibrilacije i iznenadne srčane smrti (347–349).

Automatsko izdvajanje genetičkih i molekularnih mehanizama povremeno izdvaja produkte gena pod nazivom gena koji ih kodira. Posledično, svaki element koji nosi ime gena mora da bude interpretiran kao taj gen i/ili njegov produkt (92). Primenjeno automatsko izdvajanje genetičkih i molekularnih mehanizama nije korišćeno sa ciljem da

se utvrdi definitivna genetička i molekularna osnova povezanosti genotipa i fenotipa HCM, nego sa ciljem da se postavi polazna osnova za buduća istraživanja (92).

5.3. Istraživačke hipoteze

Svih 5 istraživačkih hipoteza je prihvaćeno.

5.4. Buduće perspektive

Identifikacija HCM podtipova na osnovu ukupne fenotipske prezentacije predstavlja korak ka preciznoj medicini za HCM koja bi zatim mogla da omogući stvaranje strategija za preventivu i tretman određenih grupa HCM pacijenata (283). Svaki od rezultata prikazanih u ovom istraživanju zahteva proveru u okviru većih, specifičnih i klasičnih kliničkih studija.

6. ZAKLJUČAK

Četiri podtipa HCM su identifikovana na osnovu svih dostupnih podataka o fenotipu: klaster 0 („AHOLD“), koji se razlikuje od ostalih na osnovu AO i LDH, pri čemu su vrednosti AO > 30 mm i LDH > 300 U/L; klaster 1 („RVSP ASCAOVS“), koji se razlikuje od ostalih na osnovu RVSP, AscAO i AOvs, pri čemu su vrednosti AOvs > 27 m/s, AscAO < 31 mm i RVSP < 28 mmHg; klaster 2 („weight“), koji se razlikuje od ostalih na osnovu težine, pri čemu su vrednosti > 95 kg; i klaster 3 („AV LVOT PG“) koji se razlikuje od ostalih na osnovu AV meanPG, AV maxPG i LVOT maxPG pri čemu su vrednosti AV maxPG > 15 mmHg, AV meanPG > 6 mmHg i LVOT maxPG > 15 mmHg.

Iako niti jedna od određenih postavki subfenotipova ne korelira u potpunosti sa genotipom, neke značajne korelacije su pronađene u postavkama sa 3 subfenotipa (*TNNI3*, $p = 0.041$), 4 subfenotipa (*TNNI3*, $p = 0.045$ i *MYBPC3*, $p = 0.038$), 5 subfenotipa (*MYH7*, $p = 0.045$) i 6 subfenotipa (*MYBPC3*, $p = 0.044$).

Algoritmi mašinskog učenja su potvrdili da je utvrđivanje povezanosti genotipa i fenotipa HCM zahtevan zadatak. Predikciju isključivo na osnovu informacije o mutiranim genima je moguće izvršiti za dva ishoda fenotipa: prisustvo ili odsustvo sinusnog ritma i prisustvo ili odsustvo oštećenja/infarkta miokarda. Modeli koji vrše predikciju prisustva ili odsustva sinusnog ritma imali su podjednake performanse kada su

kreirani samo na osnovu uzročnih gena i kada su kreirani korišćenjem svih analiziranih gena, što ukazuje na mogući značaj uzročnih gena i irelevantnost gena koji nisu uzročni za ovaj ishod. Sa druge strane, modeli koji vrše predikciju oštećenja/infarkta miokarda imali su bolje performanse kada su kreirani na osnovu informacija o svim analiziranim genima (a ne samo o uzročnim genima), što sugeriše potencijalno značajnu ulogu gena koji nisu uzročni za ovaj ishod.

Algoritmi mašinskog učenja predvideli su sledeće ishode na osnovu podataka o genotipu i fenotipu: zamor, otežano disanje, bol u grudima, palpitacije, sinkopa, šum na srcu, pretibijalni edem, pokretanje mitralnog zalistka unapred (SAM), abnormalnosti papilarnih mišića, hipokinezija, atrijalna fibrilacija, AV blok prvog stepena, blok leve grane (LBBB), blok desne grane (RBBB), prednji levi hemiblok, abnormalnosti ST segmenta, i negativni T talas. Za svaki od ovih ishoda određeni su podskupovi parametara koji su dovoljni za vršenje predikcije i utvrđen je njihov relativni značaj u svakoj od predikcija. Prilikom predviđanja zamora, najveći doprinos je imala kombinacija mutacije u *TNNT2* i maksimalnog RER. Prilikom predviđanja dispneje najveći doprinos imala je kombinacija mutacije u *MYBPC3* i vršne potrošnje kiseonika (peak VO_2). Prilikom predviđanja bola u grudima, najveći doprinos je imala kombinacija mutacije u *TNNI3* i koncentracije HDL. Prilikom predviđanja šuma na srcu najveći doprinos imala je kombinacija mutacije u *MYH7* i podatka o implantiranju pejsmejkera/defibrilatora u porodičnoj istoriji, kao i kombinacija mutacije u *TNNT2* i vrednosti LAV. Prilikom predviđanja negativnog T talasa, najveći doprinos imala je kombinacija mutacije u *MYBPC3* i vrednost MV maxPG.

Identifikovani su genotip-specifični nalazi ehokardiograma: za mutaciju u *MYH7* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima), strukture koje najviše utiču na raspoznavanje su septum, izlazni trakt leve komore (LVOT), prednji zid, vrh srca, desna komora i mitralni aparat; za mutaciju u *TNNT2* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima) strukture koje najviše utiču na raspoznavanje su septum i desna komora; dok su za mutaciju u *MYBPC3* genu (nasuprot negativnom rezultatu na mutacije u analiziranim genima) ove strukture septum, leva komora i šupljina leve komore.

Mašinsko učenje je na ovaj način doprinelo u određenoj meri izučavanju povezanosti genotipa i fenotipa HCM.

7. LITERATURE

1. Aguiar Rosa S, Rocha Lopes L, Fiarresga A, Ferreira RC, Mota Carmo M. Coronary microvascular dysfunction in hypertrophic cardiomyopathy: pathophysiology, assessment, and clinical impact. Microcirculation. 2021;28(1):e12656.
2. 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(1):3–14.
3. Geske JB, Ommen SR, Gersh BJ. Hypertrophic cardiomyopathy: clinical update. JACC Heart Fail. 2018;6(5):364–75.
4. Firth J. Cardiology: hypertrophic cardiomyopathy. Clin Med (Lond). 2019;19(1):61–3.
5. Zegkos T, Tziomalos G, Parcharidou D, Ntelios D, Papanastasiou CA, Karagiannidis E, et al. Validation of the new American College of Cardiology/American Heart Association Guidelines for the risk stratification of sudden cardiac death in a large Mediterranean cohort with hypertrophic cardiomyopathy. Hell J Cardiol. 2022;63:15–21.
6. 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(1):3–14.
7. Cao Y, Zhang PY. Review of recent advances in the management of hypertrophic cardiomyopathy. Eur Rev Med Pharmacol Sci. 2017;21(22):5207–10.
8. Antunes MO, Scudeler TL. Hypertrophic cardiomyopathy. Int J Cardiol Heart Vasc. 2020;27:100503.
9. van der Velden J, Stienen GJM. Cardiac disorders and pathophysiology of sarcomeric proteins. Physiol Rev. 2019;99(1):381–426.
10. Borsari W, Davis L, Meiers E, Salberg L, McDonough B. Living with hypertrophic cardiomyopathy: a patient's perspective. Future Cardiol. 2022;18(1):43–50.
11. Maron BJ, Desai MY, Nishimura RA, Spirito P, Rakowski H, Towbin JA, et al. Management of hypertrophic cardiomyopathy: JACC state-of-the-art review. J

Am Coll Cardiol. 2022;79(4):390–414.

12. Bonaventura J, Polakova E, Vejtasova V, Veselka J. Genetic testing in patients with hypertrophic cardiomyopathy. Int J Mol Sci. 2021;22(19):10401.
13. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. J Am Coll Cardiol. 2015;65(12):1249–54.
14. Prondzynski M, Mearini G, Carrier L. Gene therapy strategies in the treatment of hypertrophic cardiomyopathy. Pflugers Arch. 2019;471(5):807–15.
15. Wolf CM. Hypertrophic cardiomyopathy: genetics and clinical perspectives. Cardiovasc Diagn Ther. 2019;9(Suppl 2):S388–415.
16. Younger J, Lo A, McCormack L, McGaughran J, Prasad S, Atherton JJ. Hypertrophic cardiomyopathy: challenging the status quo? Hear Lung Circ. 2020;29(4):556–65.
17. Tuohy CV, Kaul S, Song HK, Nazer B, Heitner SB. Hypertrophic cardiomyopathy: the future of treatment. Eur J Heart Fail. 2020;22(2):228–40.
18. Ariss RW, Khan Minhas AM, Nazir S, Patel MM, Nesheiwat Z, Mhanna M, et al. Outcomes and revascularization strategies of ST-elevation myocardial infarction in patients with hypertrophic cardiomyopathy. Curr Probl Cardiol. 2022;47(11):101102.
19. Novén J, Stagmo M, Wierup P, Nozohoor S, Bjursten H, Sjögren J, et al. Exercise echocardiography following septal myectomy for hypertrophic obstructive cardiomyopathy. Thorac Cardiovasc Surg. 2020;70(1):18–25.
20. Batzner A, Aicha D, Pfeiffer B, Neugebauer A, Seggewiss H. Age-related survival after alcohol septal ablation in hypertrophic obstructive cardiomyopathy. ESC Heart Fail. 2022;9(1):327–36.
21. Hayashi T. Hypertrophic cardiomyopathy: diverse pathophysiology revealed by genetic research, toward future therapy. Keio J Med. 2020;69(4):77–87.
22. Weissler-Snir A, Allan K, Cunningham K, Connelly KA, Lee DS, Spears DA, et al. Hypertrophic cardiomyopathy–related sudden cardiac death in young people in Ontario. Circulation. 2019;140(21):1706–16.
23. O'Hara RP, Binka E, Prakosa A, Zimmerman SL, Cartoski MJ, Abraham MR, et

al. Personalized computational heart models with T1-mapped fibrotic remodeling predict sudden death risk in patients with hypertrophic cardiomyopathy. Elife. 2022;11:e73325.

24. Hong Y, Su WW, Li X. Risk factors of sudden cardiac death in hypertrophic cardiomyopathy. Curr Opin Cardiol. 2022;37(1):15–21.
25. Huang H, Chen Y, Jin J, Du R, Tang K, Fan L, et al. *CSRP3*, p.Arg122*, is responsible for hypertrophic cardiomyopathy in a Chinese family. J Gene Med. 2022;24(1):e3390.
26. Zhang M, Sun X, Wu G, Wang D, Wang L, Zhang C, et al. Effect of cis-compound variants in *MYH7* on hypertrophic cardiomyopathy with a mild phenotype. Am J Cardiol. 2022;167:104–10.
27. Wu G, Liu J, Ruan J, Yu S, Wang L, Zhao S, et al. Deleterious rare desmosomal variants contribute to hypertrophic cardiomyopathy and are associated with distinctive clinical features. Can J Cardiol. 2022;38(1):41–8.
28. Tower-Rader A, Desai MY. Phenotype–genotype correlation in hypertrophic cardiomyopathy: less signal, more noise? Circ Cardiovasc Imaging. 2017;10(2):e006066.
29. Marian AJ, Braunwald E. Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. Circ Res. 2017;121(7):749–70.
30. Jordà P, Oudit GY, Tadros R. Unraveling the genetic substrate and phenotypic variability of hypertrophic cardiomyopathy: a role for desmosome gene variants? Can J Cardiol. 2022;38(1):3–5.
31. Pai SL, Chadha RM, Logvinov II, Brigham TJ, Watt KD, Li Z, et al. Preoperative echocardiography as a prognostic tool for liver transplant in patients with hypertrophic cardiomyopathy. Clin Transplant. 2022;36(2):e14538.
32. Elliott PM, Anastasakis A, Borger MA, Borggrefe M, Cecchi F, Charron P, et al. 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(39):2733–79.
33. Cimiotti D, Budde H, Hassoun R, Jaquet K. Genetic restrictive cardiomyopathy:

causes and consequences–an integrative approach. Int J Mol Sci. 2021;22(2):558.

34. Savariya U, Aponte MMP, Nathan S, Zhao B, Radovancevic R, de Armas IAS, et al. Hypertrophic cardiomyopathy with a complex clinical course leading to heart transplantation. Cardiovasc Pathol. 2022;58:107406.
35. Akhtar M, Elliott P. The genetics of hypertrophic cardiomyopathy. Glob Cardiol Sci Pract. 2018;2018(3):36.
36. Teekakirikul P, Zhu W, Huang HC, Fung E. Hypertrophic cardiomyopathy: an overview of genetics and management. Biomolecules. 2019;9(12):878.
37. Squire J. Special issue: the actin-myosin interaction in muscle: background and overview. Int J Mol Sci. 2019;20(22):5715.
38. Bassiouni W, Ali MAM, Schulz R. Multifunctional intracellular matrix metalloproteinases: implications in disease. FEBS J. 2021;288(24):7162–82.
39. Henderson CA, Gomez CG, Novak SM, Mi-Mi L, Gregorio CC. Overview of the muscle cytoskeleton. Compr Physiol. 2017;7(3):891–944.
40. Martin TG, Kirk JA. Under construction: the dynamic assembly, maintenance, and degradation of the cardiac sarcomere. J Mol Cell Cardiol. 2020;148:89–102.
41. Ribeiro M, Furtado M, Martins S, Carvalho T, Carmo-Fonseca M. RNA splicing defects in hypertrophic cardiomyopathy: implications for diagnosis and therapy. Int J Mol Sci. 2020;21(4):1329.
42. Solomon T, Filipovska A, Hool L, Viola H. Preventative therapeutic approaches for hypertrophic cardiomyopathy. J Physiol. 2021;599(14):3495–512.
43. Hayashi T. Hypertrophic cardiomyopathy: diverse pathophysiology revealed by genetic research, toward future therapy. Keio J Med. 2020;69(4):77–87.
44. Viola HM, Hool LC. Impaired calcium handling and mitochondrial metabolic dysfunction as early markers of hypertrophic cardiomyopathy. Arch Biochem Biophys. 2019;665:166–74.
45. Ommen SR, Semsarian C. Hypertrophic cardiomyopathy: a practical approach to guideline directed management. Lancet. 2021;398(10316):2102–8.
46. Ranta-Aho J, Olive M, Vandroux M, Roticiani G, Dominguez C, Johari M, et al. Mutation update for the *ACTN2* gene. Hum Mutat. In press. DOI: 10.1002/humu.24470.

47. Henrique JS, Braga PLG, de Almeida SS, Nunes NSP, Benfato ID, Arida RM, et al. Effect of the *ACTN-3* gene polymorphism on functional fitness and executive function of elderly. Front Aging Neurosci. 2022;14:943934.
48. Hou CR, Cortez D. Novel *ACTN2* missense variant is associated with idiopathic ventricular fibrillation: a case report. Eur Hear J Case Reports. 2022;6(7):ytac229.
49. Germain P, Delalande A, Pichon C. Role of muscle LIM protein in mechanotransduction process. Int J Mol Sci. 2022;23(17):9785.
50. Riaz M, Park J, Sewanan LR, Ren Y, Schwan J, Das SK, et al. Muscle LIM protein force-sensing mediates sarcomeric biomechanical signaling in human familial hypertrophic cardiomyopathy. Circulation. 2022;145(16):1238–53.
51. Herrero-Galán E, Martínez-Martín I, Sánchez-González C, Vicente N, Bonzón-Kulichenko E, Calvo E, et al. Basal oxidation of conserved cysteines modulates cardiac titin stiffness and dynamics. Redox Biol. 2022;52:102306.
52. Marcello M, Cetrangolo V, Savarese M, Udd B. Use of animal models to understand titin physiology and pathology. J Cell Mol Med. 2022;26(20):5103–12.
53. Kötter S, Krüger M. Protein quality control at the sarcomere: titin protection and turnover and implications for disease development. Front Physiol. 2022;13:914296.
54. Lin YH, Major JL, Liebner T, Hourani Z, Travers JG, Wennersten SA, et al. HDAC6 modulates myofibril stiffness and diastolic function of the heart. J Clin Invest. 2022;132(10):e148333.
55. Dassanayake Mudiyanselage SP, Gage MJ. Regulation of poly-E motif flexibility by pH, Ca^{2+} and the PPAK motif. Int J Mol Sci. 2022;23(9):4779.
56. Lamber EP, Guicheney P, Pinotsis N. The role of the M-band myomesin proteins in muscle integrity and cardiac disease. J Biomed Sci. 2022;29(1):18.
57. Williams TD, Rousseau A. Actin dynamics in protein homeostasis. Biosci Rep. 2022;42(9):BSR20210848.
58. Baltes C, Thalla DG, Kazmaier U, Lautenschläger F. Actin stabilization in cell migration. Front Cell Dev Biol. 2022;10:931880.

59. England J, Granados-Riveron J, Polo-Parada L, Kuriakose D, Moore C, Brook JD, et al. Tropomyosin 1: Multiple roles in the developing heart and in the formation of congenital heart defects. J Mol Cell Cardiol. 2017;106:1–13.
60. Pavadai E, Rynkiewicz MJ, Yang Z, Gould IR, Marston SB, Lehman W. Modulation of cardiac thin filament structure by phosphorylated troponin–I analyzed by protein-protein docking and molecular dynamics simulation. Arch Biochem Biophys. 2022;725:109282.
61. Cha YJ, Jeon SB, Oh J, Lee ST, Kim S, Kim H, et al. Derivation of YCMi005-A, a human-induced pluripotent stem cell line, from a patient with dilated cardiomyopathy carrying missense variant in *TPM1* (p. Glu192Lys). Stem Cell Res. 2022;60:102707.
62. Teekakirikul P, Zhu W, Xu X, Young CB, Tan T, Smith AM, et al. Genetic resiliency associated with dominant lethal *TPM1* mutation causing atrial septal defect with high heritability. Cell Rep Med. 2022;3(2):100501.
63. Chalovich JM, Zhu L, Johnson D. Hypertrophic cardiomyopathy mutations of troponin reveal details of striated muscle regulation. Front Physiol. 2022;13:902079.
64. Daniels MJ, Fusi L, Semsarian C, Naidu SS. Myosin modulation in hypertrophic cardiomyopathy and systolic heart failure: getting inside the engine. Circulation. 2021;144(10):759–62.
65. Day SM, Tardiff JC, Ostap EM. Myosin modulators: emerging approaches for the treatment of cardiomyopathies and heart failure. J Clin Invest. 2022;132(5):e148557.
66. Olivotto I, Oreziak A, Barriales-Villa R, Abraham TP, Masri A, Garcia-Pavia P, et al. Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet. 2020;396(10253):759–69.
67. Wood NB, Kelly CM, O'Leary TS, Martin JL, Previs MJ. Cardiac myosin filaments are maintained by stochastic protein replacement. Mol Cell Proteomics. 2022;21(10):100274.
68. Desai DA, Rao VJ, Jegga AG, Dhandapany PS, Sadayappan S. Heterogeneous

distribution of genetic mutations in myosin binding protein-C paralogs. Front Genet. 2022;13:896117.

69. Arif M, Nabavizadeh P, Song T, Desai D, Singh R, Bazrafshan S, et al. Genetic, clinical, molecular, and pathogenic aspects of the South Asian–specific polymorphic *MYBPC3*$^{\Delta 25bp}$ variant. Biophys Rev. 2020;12(4):1065–84.
70. Nussinov R, Tsai CJ, Jang H. Protein ensembles link genotype to phenotype. PLoS Comput Biol. 2019;15(6):e1006648.
71. Wijnker PJM, Sequeira V, Kuster DWD, Velden JV. Hypertrophic cardiomyopathy: a vicious cycle triggered by sarcomere mutations and secondary disease hits. Antioxid Redox Signal. 2019;31(4):318–58.
72. Cui Y, Liu C, Luo J, Liang J. Dysfunctional network and mutation genes of hypertrophic cardiomyopathy. J Healthc Eng. 2022;2022:8680178.
73. Yotti R, Seidman CE, Seidman JG. Advances in the genetic basis and pathogenesis of sarcomere cardiomyopathies. Annu Rev Genomics Hum Genet. 2019;20:129–53.
74. Alfares AA, Kelly MA, McDermott G, Funke BH, Lebo MS, Baxter SB, et al. Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. Genet Med. 2015;17(11):880–8.
75. Alders M, Jongbloed R, Deelen W, van den Wijngaard A, Doevendans P, Ten Cate F, et al. The 2373insG mutation in the *MYBPC3* gene is a founder mutation, which accounts for nearly one-fourth of the HCM cases in the Netherlands. Eur Heart J. 2003;24(20):1848–53.
76. Jääskeläinen P, Kuusisto J, Miettinen R, Kärkkäinen P, Kärkkäinen S, Heikkinen S, et al. Mutations in the cardiac myosin-binding protein C gene are the predominant cause of familial hypertrophic cardiomyopathy in eastern Finland. J Mol Med (Berl). 2002;80(7):412–22.
77. Adalsteinsdottir B, Teekakirikul P, Maron BJ, Burke MA, Gudbjartsson DF, Holm H, et al. Nationwide study on hypertrophic cardiomyopathy in Iceland: evidence of a *MYBPC3* founder mutation. Circulation. 2014;130(14):1158–67.
78. Kubo T, Kitaoka H, Okawa M, Matsumura Y, Hitomi N, Yamasaki N, et al.

Lifelong left ventricular remodeling of hypertrophic cardiomyopathy caused by a founder frameshift deletion mutation in the cardiac Myosin-binding protein C gene among Japanese. J Am Coll Cardiol. 2005;46(9):1737–43.

79. Dhandapany PS, Sadayappan S, Xue Y, Powell GT, Rani DS, Nallari P, et al. A common *MYBPC3* (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia. Nat Genet. 2009;41(2):187–91.
80. Chou C, Chin MT. Pathogenic mechanisms of hypertrophic cardiomyopathy beyond sarcomere dysfunction. Int J Mol Sci. 2021;22(16):8933.
81. Ogino S, Gulley ML, den Dunnen JT, Wilson RB; Association for Molecular Patholpogy Training and Education Committtee. Standard mutation nomenclature in molecular diagnostics : practical and educational challenges. J Mol Diagn. 2007;9(1):1–6.
82. Carrier L. Targeting the population for gene therapy with *MYBPC3*. J Mol Cell Cardiol. 2021;150:101–8.
83. Spudich JA. Three perspectives on the molecular basis of hypercontractility caused by hypertrophic cardiomyopathy mutations. Pflugers Arch. 2019;471(5):701–17.
84. Glazier AA, Thompson A, Day SM. Allelic imbalance and haploinsufficiency in *MYBPC3*-linked hypertrophic cardiomyopathy. Pflugers Arch. 2019;471(5):781–93.
85. Mazzarotto F, Olivotto I, Boschi B, Girolami F, Poggesi C, Barton PJR, et al. Contemporary insights into the genetics of hypertrophic cardiomyopathy: toward a new era in clinical testing? J Am Hear Assoc. 2020;9(8):e015473.
86. Maron BJ, Desai MY, Nishimura RA, Spirito P, Rakowski H, Towbin JA, et al. Diagnosis and evaluation of hypertrophic cardiomyopathy: JACC State-of-the-Art Review. J Am Coll Cardiol. 2022;79(4):372–89.
87. Ueda Y, Stern JA. A one health approach to hypertrophic cardiomyopathy. Yale J Biol Med. 2017;90(3):433–48.
88. Repetti GG, Kim Y, Pereira AC, Ingles J, Russell MW, Lakdawala NK, et al. Discordant clinical features of identical hypertrophic cardiomyopathy twins. Proc Natl Acad Sci U S A. 2021;118(10):e2021717118.

89. Sequeira V, Bertero E, Maack C. Energetic drain driving hypertrophic cardiomyopathy. FEBS Lett. 2019;593(13):1616–26.
90. Farrell ET, Grimes AC, de Lange WJ, Armstrong AE, Ralphe JC. Increased postnatal cardiac hyperplasia precedes cardiomyocyte hypertrophy in a model of hypertrophic cardiomyopathy. Front Physiol. 2017;8:414.
91. Ramachandra CJA, Mai Ja KPM, Lin YH, Shim W, Boisvert WA, Hausenloy DJ. Induced pluripotent stem cells for modelling energetic alterations in hypertrophic cardiomyopathy. Cond Med. 2019;2(4):142–51.
92. Glavaški M, Velicki L. Shared molecular mechanisms of hypertrophic cardiomyopathy and its clinical presentations: automated molecular mechanisms extraction approach. Life (Basel). 2021;11(8):785.
93. Zhang ZL, Xu YY, Qin Z, Lu YZ, Liu TD, Zhang L, et al. N-terminal pro-brain natriuretic peptide and adverse outcomes in Chinese patients with hypertrophic cardiomyopathy. Biosci Rep. 2022;42(1):BSR20212098.
94. MacIver DH, Clark AL. Contractile dysfunction in sarcomeric hypertrophic cardiomyopathy. J Card Fail. 2016;22(9):731–7.
95. Sukhacheva TV, Chudinovskikh YA, Eremeeva MV, Serov RA, Bockeria LA. Proliferative potential of cardiomyocytes in hypertrophic cardiomyopathy: correlation with myocardial remodeling. 2016;162(1):160–9.
96. Fernlund E, Gyllenhammar T, Jablonowski R, Carlsson M, Larsson A, Ärnlöv J, et al. Serum biomarkers of myocardial remodeling and coronary dysfunction in early stages of hypertrophic cardiomyopathy in the young. Pediatr Cardiol. 2017;38(4):853–63.
97. Ramachandra CJA, Kp MMJ, Chua J, Hernandez-Resendiz S, Liehn EA, Gan LM, et al. Inhibiting cardiac myeloperoxidase alleviates the relaxation defect in hypertrophic cardiomyocytes. Cardiovasc Res. 2022;118(2):517–30.
98. Coppini R, Ferrantini C, Mugelli A, Poggesi C, Cerbai E. Altered Ca^{2+} and Na^{+} homeostasis in human hypertrophic cardiomyopathy: implications for arrhythmogenesis. Front Physiol. 2018;9:1391.
99. Volpe M, Liuzzo G. VANISHing the progression of cardiac abnormalities in hypertrophic cardiomyopathy with early use of valsartan? Eur Heart J.

2022;43(3):181–2.

100. Yin L, Xu HY, Zheng SS, Zhu Y, Xiao JX, Zhou W, et al. 3.0 T magnetic resonance myocardial perfusion imaging for semi-quantitative evaluation of coronary microvascular dysfunction in hypertrophic cardiomyopathy. Int J Cardiovasc Imaging. 2017;33(12):1949–59.
101. Bakar SN, Hayman S, McCarty D, Thain AP, McLellan A, Wagner C, et al. Invasive assessment of microvascular resistance in hypertrophic cardiomyopathy with echocardiographic correlates. Hear Lung Circ. 2022;31(2):194–8.
102. Raphael CE, Cooper R, Parker KH, Collinson J, Vassiliou V, Pennell DJ, et al. Mechanisms of myocardial ischemia in hypertrophic cardiomyopathy: insights from wave intensity analysis and magnetic resonance. J Am Coll Cardiol. 2016;68(15):1651–60.
103. Argirò A, Zampieri M, Berteotti M, Marchi A, Tassetti L, Zocchi C, et al. Emerging medical treatment for hypertrophic cardiomyopathy. J Clin Med. 2021;10(5):951.
104. Toepfer CN, Wakimoto H, Garfinkel AC, McDonough B, Liao D, Jiang J, et al. Hypertrophic cardiomyopathy mutations in *MYBPC3* dysregulate myosin. Sci Transl Med. 2019;11(476):eaat1199.
105. Autore C, Ferrazzi P. Patients with hypertrophic cardiomyopathy are getting older. Int J Cardiol. 2022;353:73–4.
106. Cordts K, Seelig D, Lund N, Carrier L, Böger RH, Avanesov M, et al. Association of asymmetric dimethylarginine and diastolic dysfunction in patients with hypertrophic cardiomyopathy. Biomolecules. 2019;9(7):277.
107. Velicki L, Jakovljevic DG, Preveden A, Golubovic M, Bjelobrk M, Ilic A, et al. Genetic determinants of clinical phenotype in hypertrophic cardiomyopathy. BMC Cardiovasc Disord. 2020;20(1):516.
108. Su W, Huo Q, Wu H, Wang L, Ding X, Liang L, et al. The function of LncRNA-H19 in cardiac hypertrophy. Cell Biosci. 2021;11(1):153.
109. Liang T, Gao F, Chen J. Role of PTEN-less in cardiac injury, hypertrophy and regeneration. Cell Regen. 2021;10(1):25.
110. Shah AK, Bhullar SK, Elimban V, Dhalla NS. Oxidative stress as a mechanism

for functional alterations in cardiac hypertrophy and heart failure. Antioxidants. 2021;10(6):931.

111. Xie J, Wang Y, Ai D, Yao L, Jiang H. The role of the Hippo pathway in heart disease. FEBS J. 2022(19):5819–33.
112. Matsuura K. Toward the development of novel therapy for hypertrophic cardiomyopathy. Int Heart J. 2018;59(5):914–6.
113. Kraft T, Montag J. Altered force generation and cell-to-cell contractile imbalance in hypertrophic cardiomyopathy. Pflugers Arch. 2019;471(5):719–33.
114. Montag J, Kraft T. Stochastic allelic expression as trigger for contractile imbalance in hypertrophic cardiomyopathy. Biophys Rev. 2020;12(4):1055–64.
115. Schirone L, Forte M, Palmerio S, Yee D, Nocella C, Angelini F, et al. A review of the molecular mechanisms underlying the development and progression of cardiac remodeling. Oxid Med Cell Longev. 2017;2017:3920195.
116. Turner NA, Blythe NM. Cardiac fibroblast p38 MAPK: a critical regulator of myocardial remodeling. J Cardiovasc Dev Dis. 2019;6(3):27.
117. Zhou H, Wang B, Yang XY, Jia QJ, Zhang A, Qi ZW, et al. Long noncoding RNAs in pathological cardiac remodeling: a review of the update literature. Biomed Res Int. 2019;2019:7159592.
118. Liu X, Ma Y, Yin K, Li W, Chen W, Zhang Y, et al. Long non-coding and coding RNA profiling using strand-specific RNA-seq in human hypertrophic cardiomyopathy. Sci Data. 2019;6(1):90.
119. Chen C, Ponnusamy M, Liu C, Gao J, Wang K, Li P. MicroRNA as a therapeutic target in cardiac remodeling. Biomed Res Int. 2017;2017:1278436.
120. Garfinkel AC, Seidman JG, Seidman CE. Genetic pathogenesis of hypertrophic and dilated cardiomyopathy. Heart Fail Clin. 2018;14(2):139.
121. Webber M, Jackson SP, Moon JC, Captur G. Myocardial fibrosis in heart failure: anti-fibrotic therapies and the role of cardiovascular magnetic resonance in drug trials. Cardiol Ther. 2020;9(2):363–76.
122. Kotadia I, Whitaker J, Roney C, Niederer S, O'Neill M, Bishop M, et al. Anisotropic cardiac conduction. Arrhythm Electrophysiol Rev. 2020;9(4):202–10.

123. Giordano C, Francone M, Cundari G, Pisano A, d'Amati G. Myocardial fibrosis: morphologic patterns and role of imaging in diagnosis and prognostication. Cardiovasc Pathol. 2022;56:107391.
124. Zhu L, Li N, Sun L, Zheng D, Shao G. Non-coding RNAs: the key detectors and regulators in cardiovascular disease. Genomics. 2021;113(1):1233–46.
125. Liu MN, Luo G, Gao WJ, Yang SJ, Zhou H. miR-29 family: a potential therapeutic target for cardiovascular disease. Pharmacol Res. 2021;166:105510.
126. Carbone A, D'Andrea A, Sperlongano S, Tagliamonte E, Mandoli GE, Santoro C, et al. Echocardiographic assessment of coronary microvascular dysfunction: basic concepts, technical aspects, and clinical settings. Echocardiography. 2021;38(6):993–1001.
127. Musumeci B, Tini G, Russo D, Sclafani M, Cava F, Tropea A, et al. Left ventricular remodeling in hypertrophic cardiomyopathy: an overview of current knowledge. J Clin Med. 2021;10(8):1547.
128. Godo S, Suda A, Takahashi J, Yasuda S, Shimokawa H. Coronary microvascular dysfunction. Arterioscler Thromb Vasc Biol. 2021;41(5):1625–37.
129. Zhu H, Zhou H. Novel insight into the role of endoplasmic reticulum stress in the pathogenesis of myocardial ischemia-reperfusion injury. Oxid Med Cell Longev. 2021;2021:5529810.
130. Wang WL, Ge TY, Chen X, Mao Y, Zhu YZ. Advances in the protective mechanism of NO, H_2S, and H_2 in myocardial ischemic injury. Front Cardiovasc Med. 2020;7:588206.
131. Mamidi R, Li J, Doh CY, Verma S, Stelzer JE. Impact of the myosin modulator mavacamten on force generation and cross-bridge behavior in a murine model of hypercontractility. J Am Hear Assoc. 2018;7(17):e009627.
132. Ren X, Hensley N, Brady MB, Gao WD. The genetic and molecular bases for hypertrophic cardiomyopathy: the role for calcium sensitization. J Cardiothorac Vasc Anesth. 2018;32(1):478–87.
133. Sarkar SS, Trivedi DV, Morck MM, Adhikari AS, Pasha SN, Ruppel KM, et al. The hypertrophic cardiomyopathy mutations R403Q and R663H increase the number of myosin heads available to interact with actin. Sci Adv.

2020;6(14):eaax0069.

134. Voigt JU. Direct stiffness measurements by echocardiography: does the search for the holy grail come to an end? JACC Cardiovasc Imaging. 2019;12(7 Pt 1):1146–8.
135. Villemain O, Correia M, Mousseaux E, Baranger J, Zarka S, Podetti I, et al. Myocardial stiffness evaluation using noninvasive shear wave imaging in healthy and hypertrophic cardiomyopathic adults. JACC Cardiovasc Imaging. 2019;12(7 Pt 1):1135–45.
136. Li N, Hang W, Shu H, Zhou N. RBM20, a therapeutic target to alleviate myocardial stiffness via titin isoforms switching in HFpEF. Front Cardiovasc Med. 2022;9:928244.
137. Münch J, Abdelilah-Seyfried S. Sensing and responding of cardiomyocytes to changes of tissue stiffness in the diseased heart. Front Cell Dev Biol. 2021;9:642840.
138. Wijnker PJM, van der Velden J. Mutation-specific pathology and treatment of hypertrophic cardiomyopathy in patients, mouse models and human engineered heart tissue. Biochim Biophys Acta Mol Basis Dis. 2020;1866(8):165774.
139. Wenzl FA, Ambrosini S, Paneni F. Molecular underpinnings of myocardial stiffness in patients with hypertrophic cardiomyopathy. Int J Cardiol. 2021;343:80–2.
140. Nollet EE, Westenbrink BD, de Boer RA, Kuster DWD, van der Velden J. Unraveling the genotype-phenotype relationship in hypertrophic cardiomyopathy: obesity-related cardiac defects as a major disease modifier. J Am Hear Assoc. 2020;9(22):e018641.
141. Rajiah P, Fulton NL, Bolen M. Magnetic resonance imaging of the papillary muscles of the left ventricle: normal anatomy, variants, and abnormalities. Insights Imaging. 2019;10(1):83.
142. Hughes RK, Knott KD, Malcolmson J, Augusto JB, Mohiddin SA, Kellman P, et al. Apical hypertrophic cardiomyopathy: the variant less known. J Am Hear Assoc. 2020;9(5):e015294.
143. Popa-Fotea NM, Micheu MM, Bataila V, Scafa-Udriste A, Dorobantu L,

Scarlatescu AI, et al. Exploring the continuum of hypertrophic cardiomyopathy–from DNA to clinical expression. Medicina (Kaunas). 2019;55(6):299.

144. Chiang YP, Shimada YJ, Ginns J, Weiner SD, Takayama H. Septal myectomy for hypertrophic cardiomyopathy: important surgical knowledge and technical tips in the era of increasing alcohol septal ablation. Gen Thorac Cardiovasc Surg. 2018;66(4):192–200.
145. Bos JM, Towbin JA, Ackerman MJ. Diagnostic, prognostic, and therapeutic implications of genetic testing for hypertrophic cardiomyopathy. J Am Coll Cardiol. 2009;54(3):201–11.
146. Smole T, Žunkovič B, Pičulin M, Kokalj E, Robnik-Šikonja M, Kukar M, et al. A machine learning-based risk stratification model for ventricular tachycardia and heart failure in hypertrophic cardiomyopathy. Comput Biol Med. 2021;135:104648.
147. Díez-López C, Salazar-Mendiguchía J. Clinical presentations of hypertrophic cardiomyopathy and implications for therapy. Glob Cardiol Sci Pract. 2018;2018(3):19.
148. Veselka J, Anavekar NS, Charron P. Hypertrophic obstructive cardiomyopathy. Lancet. 2017;389(10075):1253–67.
149. Goff ZD, Calkins H. Sudden death related cardiomyopathies–hypertrophic cardiomyopathy. Prog Cardiovasc Dis. 2019;62(3):212–6.
150. Mateo JJS, Gimeno JR. Alcohol septal ablation in hypertrophic cardiomyopathy. Glob Cardiol Sci Pract. 2018;2018(3):30.
151. Kogut J, Popjes ED. Hypertrophic cardiomyopathy 2020. Curr Cardiol Rep. 2020;22(11):154.
152. Brieler J, Breeden MA, Tucker J. Cardiomyopathy: an overview. Am Fam Physician. 2017;96(10):640–6.
153. Maron BJ, Rowin EJ, Udelson JE, Maron MS. Clinical spectrum and management of heart failure in hypertrophic cardiomyopathy. JACC Hear Fail. 2018;6(5):353–63.
154. Nishimura RA, Seggewiss H, Schaff HV. Hypertrophic obstructive cardiomyopathy: surgical myectomy and septal ablation. Circ Res.

2017;121(7):771–83.

155. Marrakchi S, Kammoun I, Bennour E, Laroussi L, Kachboura S. Risk stratification in hypertrophic cardiomyopathy. Herz. 2020;45(1):50–64.
156. Savarimuthu S, Harky A. Alcohol septal ablation: a useful tool in our arsenal against hypertrophic obstructive cardiomyopathy. J Card Surg. 2020;35(8):2017–24.
157. Falasconi G, Pannone L, Slavich M, Margonato A, Fragasso G, Spoladore R. Atrial fibrillation in hypertrophic cardiomyopathy: pathophysiology, diagnosis and management. Am J Cardiovasc Dis. 2020;10(4):409–18.
158. Pelliccia F, Alfieri O, Calabrò P, Cecchi F, Ferrazzi P, Gragnano F, et al. Multidisciplinary evaluation and management of obstructive hypertrophic cardiomyopathy in 2020: towards the HCM Heart Team. Int J Cardiol. 2020;304:86–92.
159. Hindieh W, Adler A, Weissler-Snir A, Fourey D, Harris S, Rakowski H. Exercise in patients with hypertrophic cardiomyopathy: a review of current evidence, national guideline recommendations and a proposal for a new direction to fitness. J Sci Med Sport. 2017;20(4):333–8.
160. Garg L, Gupta M, Sabzwari SRA, Agrawal S, Agarwal M, Nazir T, et al. Atrial fibrillation in hypertrophic cardiomyopathy: prevalence, clinical impact, and management. Hear Fail Rev. 2019;24(2):189–97.
161. Vaidya K, Semsarian C, Chan KH. Atrial fibrillation in hypertrophic cardiomyopathy. Hear Lung Circ. 2017;26(9):975–82.
162. Yeung C, Enriquez A, Suarez-Fuster L, Baranchuk A. Atrial fibrillation in patients with inherited cardiomyopathies. Europace. 2019;21(1):22–32.
163. Patten M, Pecha S, Aydin A. Atrial fibrillation in hypertrophic cardiomyopathy: Diagnosis and considerations for management. J Atr Fibrillation. 2018;10(5);1556.
164. Kochi AN, Vettor G, Dessanai MA, Pizzamiglio F, Tondo C. Sudden cardiac death in athletes: from the basics to the practical work-up. Medicina (Kaunas). 2021;57(2):168.
165. Weissler-Snir A, Adler A, Williams L, Gruner C, Rakowski H. Prevention of

sudden death in hypertrophic cardiomyopathy: bridging the gaps in knowledge. Eur Heart J. 2017;38(22):1728–37.

166. Raiker N, Vullaganti S, Collins JD, Allen BD, Choudhury L. Myocardial tissue characterization by gadolinium-enhanced cardiac magnetic resonance imaging for risk stratification of adverse events in hypertrophic cardiomyopathy. Int J Cardiovasc Imaging. 2020;36(6):1147–56.
167. Jordà P, García-Álvarez A. Hypertrophic cardiomyopathy: sudden cardiac death risk stratification in adults. Glob Cardiol Sci Pract. 2018;2018(3):25.
168. Kitaoka H, Kubo T, Doi YL. Hypertrophic cardiomyopathy–a heterogeneous and lifelong disease in the real world. Circ J. 2020;84(8):1218–26.
169. Lee JM, Park HB, Song JE, Kim IC, Song JH, Kim H, et al. The impact of cardiopulmonary exercise-derived scoring on prediction of cardio-cerebral outcome in hypertrophic cardiomyopathy. PLoS One. 2022;17(1):e0259638.
170. Smith JG. Molecular epidemiology of heart failure: translational challenges and opportunities. JACC Basic Transl Sci. 2017;2(6):757–69.
171. Mandeş L, Roşca M, Ciupercă D, Popescu BA. The role of echocardiography for diagnosis and prognostic stratification in hypertrophic cardiomyopathy. J Echocardiogr. 2020;18(3):137–48.
172. Ji Q, Wang Y, Liu F, Yang Y, Xia L, Ding W, et al. Hypertrophic cardiomyopathy with latent obstruction: clinical characteristics and surgical results. J Cardiol. 2022;79(1):42–9.
173. Coppini R, Santini L, Olivotto I, Ackerman MJ, Cerbai E. Abnormalities in sodium current and calcium homoeostasis as drivers of arrhythmogenesis in hypertrophic cardiomyopathy. Cardiovasc Res. 2020;116(9):1585–99.
174. Butters A, Lakdawala NK, Ingles J. Sex differences in hypertrophic cardiomyopathy: interaction with genetics and environment. Curr Heart Fail Rep. 2021;18(5):264–73.
175. Norrish G, Field E, Kaski JP. Childhood hypertrophic cardiomyopathy: a disease of the cardiac sarcomere. Front Pediatr. 2021;9:708679.
176. Sedaghat-Hamedani F, Kayvanpour E, Tugrul OF, Lai A, Amr A, Haas J, et al. Clinical outcomes associated with sarcomere mutations in hypertrophic

cardiomyopathy: a meta-analysis on 7675 individuals. Clin Res Cardiol. 2018;107(1):30–41.

177. Marian AJ. Modifier genes for hypertrophic cardiomyopathy. Curr Opin Cardiol. 2002;17(3):242–52.
178. Orenes-Piñero E, Hernández-Romero D, Jover E, Valdés M, Lip GY, Marín F. Impact of polymorphisms in the renin-angiotensin-aldosterone system on hypertrophic cardiomyopathy. J Renin Angiotensin Aldosterone Syst. 2011;12(4):521–30.
179. Pradeep R, Akram A, Proute MC, Kothur NR, Georgiou P, Serhiyenia T, et al. Understanding the genetic and molecular basis of familial hypertrophic cardiomyopathy and the current trends in gene therapy for its management. Cureus. 2021;13(8):e17548.
180. Glavaški M, Stankov K. Epigenetics in disease etiopathogenesis. Genetika. 2019;51(3):975–94.
181. Dias KA, Link MS, Levine BD. Exercise training for patients with hypertrophic cardiomyopathy: JACC review topic of the week. J Am Coll Cardiol. 2018;72(10):1157–65.
182. Sun D, Schaff HV, McKenzie TJ, Nishimura RA, Geske JB, Dearani JA, et al. Safety of bariatric surgery in obese patients with hypertrophic cardiomyopathy. Am J Cardiol. 2022;167:93–7.
183. Becherucci F, Landini S, Cirillo L, Mazzinghi B, Romagnani P. Look alike, sound alike: phenocopies in steroid-resistant nephrotic syndrome. Int J Environ Res Public Health. 2020;17(22):8363.
184. Maron BJ, Rowin EJ, Arkun K, Rastegar H, Larson AM, Maron MS, et al. Adult monozygotic twins with hypertrophic cardiomyopathy and identical disease expression and clinical course. Am J Cardiol. 2020;127:135–8.
185. Maron BJ, Rowin EJ, Maron MS. Global burden of hypertrophic cardiomyopathy. JACC Hear Fail. 2018;6(5):376–8.
186. Makavos G, Kairis C, Tselegkidi ME, Karamitsos T, Rigopoulos AG, Noutsias M, et al. Hypertrophic cardiomyopathy: an updated review on diagnosis, prognosis, and treatment. Hear Fail Rev. 2019;24(4):439–59.

187. Chiang YP, Shimada YJ, Ginns J, Weiner SD, Takayama H. Septal myectomy for hypertrophic cardiomyopathy: important surgical knowledge and technical tips in the era of increasing alcohol septal ablation. Gen Thorac Cardiovasc Surg. 2018;66(4):192–200.
188. Chuang C, Collibee S, Ashcraft L, Wang W, Vander Wal M, Wang X, et al. Discovery of aficamten (CK-274), a next-generation cardiac myosin inhibitor for the treatment of hypertrophic cardiomyopathy. J Med Chem. 2021;64(19):14142–52.
189. Morelli C, Ingrasciotta G, Jacoby D, Masri A, Olivotto I. Sarcomere protein modulation: the new frontier in cardiovascular medicine and beyond. Eur J Intern Med. 2022;102:1–7.
190. Fitzgerald P, Kusumoto F. The effects of septal myectomy and alcohol septal ablation for hypertrophic cardiomyopathy on the cardiac conduction system. J Interv Card Electrophysiol. 2018;52(3):403–8.
191. Arévalos V, Rodríguez-Arias JJ, Brugaletta S, Micari A, Costa F, Freixa X, et al. Alcohol septal ablation: an option on the rise in hypertrophic obstructive cardiomyopathy. J Clin Med. 2021;10(11):2276.
192. Maron BJ, Rowin EJ, Maron MS, Braunwald E. Nonobstructive hypertrophic cardiomyopathy out of the shadows: known from the beginning but largely ignored … until now. Am J Med. 2017;130(2):119–23.
193. Liew AC, Vassiliou VS, Cooper R, Raphael CE. Hypertrophic cardiomyopathy–past, present and future. J Clin Med. 2017;6(12):118.
194. Ye Z, Smith MM, Jouni H, Geske JB, Carney SA, Urina-Jassir M, et al. Mitral valve cleft-like indentations in hypertrophic obstructive cardiomyopathy: insights from intraoperative three-dimensional transesophageal echocardiography. J Cardiothorac Vasc Anesth. 2022;36(2):429–36.
195. Douglas JS Jr. Current state of the roles of alcohol septal ablation and surgical myectomy in the treatment of hypertrophic obstructive cardiomyopathy. Cardiovasc Diagn Ther. 2020;10(1):36–44.
196. He M, Qiu J, Bai Y, Wang Y, Hu M, Chen G. Non-pharmaceutical interventions for hypertrophic cardiomyopathy: a mini review. Front Cardiovasc Med.

2021;8:695247.

197. Maron MS, Rastegar H, Dolan N, Carpino P, Koethe B, Maron BJ, et al. Outcomes over follow-up ≥10 years after surgical myectomy for symptomatic obstructive hypertrophic cardiomyopathy. Am J Cardiol. 2022;163:91–7.
198. MacEachern SJ, Forkert ND. Machine learning for precision medicine. Genome. 2021;64(4):416–25.
199. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC State-of-the-Art Review. J Am Coll Cardiol. 2021;77(3):300–13.
200. Edwards AS, Kaplan B, Jie T. A primer on machine learning. Transplantation. 2021;105(4):699–703.
201. Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction. IEEE Rev Biomed Eng. 2021;14:116–26.
202. Mahdi MA, Hosny KM, Elhenawy I. Scalable clustering algorithms for big data: a review. IEEE Access. 2021;9:80015–27.
203. Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci. 2021;2(3):160.
204. Garg A, Mago V. Role of machine learning in medical research: a survey. Comput Sci Rev. 2021;40:100370.
205. Golalipour K, Akbari E, Hamidi SS, Lee M, Enayatifar R. From clustering to clustering ensemble selection: a review. Eng Appl Artif Intell. 2021;104:104388.
206. Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, et al. Artificial intelligence and machine learning for medical imaging: a technology review. Phys Med. 2021;83:242–56.
207. Tchito Tchapga C, Mih TA, Tchagna Kouanou A, Fozin Fonzin T, Kuetche Fogang P, Mezatio BA, et al. Biomedical image classification in a big data architecture using machine learning algorithms. J Healthc Eng. 2021;2021:9998819.
208. Stankovic B, Kotur N, Nikcevic G, Gasic V, Zukic B, Pavlovic S. Machine learning modeling from omics data as prospective tool for improvement of inflammatory bowel disease diagnosis and clinical classifications. Genes (Basel).

2021;12(9):1438.

209. Wu S, Chen Y, Li Z, Li J, Zhao F, Su X. Towards multi-label classification: next step of machine learning for microbiome research. Comput Struct Biotechnol J. 2021;19:2742–9.
210. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med. 2021;13(1):152.
211. Baptista D, Ferreira PG, Rocha M. Deep learning for drug response prediction in cancer. Brief Bioinform. 2021;22(1):360–79.
212. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Adv Exp Med Biol. 2020;1213:3–21.
213. Muzio G, O'Bray L, Borgwardt K. Biological network analysis with deep learning. Brief Bioinform. 2021;22(2):1515–30.
214. Hallou A, Yevick HG, Dumitrascu B, Uhlmann V. Deep learning for bioimage analysis in developmental biology. Development. 2021;148(18):dev199616.
215. Wang H, Pujos-Guillot E, Comte B, de Miranda JL, Spiwok V, Chorbev I, et al. Deep learning in systems medicine. Brief Bioinform. 2021;22(2):1543–59.
216. Shi Q, Chen W, Huang S, Wang Y, Xue Z. Deep learning for mining protein data. Brief Bioinform. 2021;22(1):194–218.
217. Winnenburg R, Wächter T, Plake C, Doms A, Schroeder M. Facts from text: can text mining help to scale-up high-quality manual curation of gene products with ontologies? Brief Bioinform. 2008;9(6):466–78.
218. National Library of Medicine [Internet]. PubMed Overview [cited 2022 Jan 27]. Available from: https://pubmed.ncbi.nlm.nih.gov/about/
219. National Center for Biotechnology Information [Internet]. U.S. National Library of Medicine, MeSH [cited 2022 Jan 30]. Available from: https://www.ncbi.nlm.nih.gov/mesh/
220. Gaulton A, Hersey A, Nowotka ML, Patricia Bento A, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945–54.
221. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids

Res. 2011;39(Database issue):D685–90.

222. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014;42(Database issue):D1091–7.
223. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, et al. The comparative toxicogenomics database: update 2019. Nucleic Acids Res. 2019;47(D1):D948–54.
224. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. MiRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res. 2020;48(D1):D148–54.
225. Ostaszewski M, Gebel S, Kuperstein I, Mazein A, Zinovyev A, Dogrusoz U, et al. Community-driven roadmap for integrated disease maps. Brief Bioinform. 2019;20(2):659–70.
226. Hoyt CT, Domingo-Fernández D, Aldisi R, Xu L, Kolpeja K, Spalek S, et al. Re-curation and rational enrichment of knowledge graphs in Biological Expression Language. Database (Oxford). 2019;2019:baz068.
227. INDRA Database [Internet]. 2022 [cited 2022 Jan 21]. Available from: https://db.indra.bio/search
228. Gyori BM, Bachman JA, Subramanian K, Muhlich JL, Galescu L, Sorger PK. From word models to executable models of signaling networks using automated assembly. Mol Syst Biol. 2017;13(11):954.
229. Singh V, Kalliolias GD, Ostaszewski M, Veyssiere M, Pilalis E, Gawron P, et al. RA-map: building a state-of-the-art interactive knowledge base for rheumatoid arthritis. Database (Oxford). 2020;2020:baaa017.
230. Mazein A, Ostaszewski M, Kuperstein I, Watterson S, Le Novère N, Lefaudeux D, et al. Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms. npj Syst Biol Appl. 2018;4(1):21.
231. Kuperstein I, Bonnet E, Nguyen HA, Cohen D, Viara E, Grieco L, et al. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. Oncogenesis. 2015;4(7):e160.
232. Serhan CN, Gupta SK, Perretti M, Godson C, Brennan E, Li Y, et al. The Atlas

of Inflammation Resolution (AIR). Mol Aspects Med. 2020;74:100894.

233. Fujita KA, Ostaszewski M, Matsuoka Y, Ghosh S, Glaab E, Trefois C, et al. Integrating pathways of Parkinson's disease in a molecular interaction map. Mol Neurobiol. 2014;49(1):88–102.
234. Mizuno S, Iijima R, Ogishima S, Kikuchi M, Matsuoka Y, Ghosh S, et al. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. BMC Syst Biol. 2012;6:52.
235. Mazein A, Knowles RG, Adcock I, Chung KF, Wheelock CE, Maitland-van der Zee AH, et al. AsthmaMap: an expert-driven computational representation of disease mechanisms. Clin Exp Allergy. 2018;48(8):916–8.
236. Mazein A, Ivanova O, Balaur I, Ostaszewski M, Berzhitskaya V, Serebriyskaya T, et al. AsthmaMap: an interactive knowledge repository for mechanisms of asthma. J Allergy Clin Immunol. 2021;147(3):853–6.
237. Matsuoka Y, Matsumae H, Katoh M, Eisfeld AJ, Neumann G, Hase T, et al. A comprehensive map of the influenza A virus replication cycle. BMC Syst Biol. 2013;7:97.
238. National Library of Medicine [Internet]. *MYH7* gene [cited 2022 Feb 15]. Available from: https://www.ncbi.nlm.nih.gov/gene/4625
239. National Library of Medicine [Internet]. *MYBPC3* gene [cited 2022 Feb 15]. Available from: https://www.ncbi.nlm.nih.gov/gene/4607
240. National Library of Medicine [Internet]. *TNNT2* gene [cited 2022 Feb 18]. Available from: https://www.ncbi.nlm.nih.gov/gene/7139
241. National Library of Medicine [Internet]. *TPM1* gene [cited 2022 Feb 23]. Available from: https://www.ncbi.nlm.nih.gov/gene/7168
242. National Library of Medicine [Internet]. *MYL2* gene [cited 2022 Mar 2]. Available from: https://www.ncbi.nlm.nih.gov/gene/4633
243. National Library of Medicine [Internet]. *MYL3* gene [cited 2022 Mar 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/4634
244. National Library of Medicine [Internet]. *TNNI3* gene [cited 2022 Mar 17]. Available from: https://www.ncbi.nlm.nih.gov/gene/7137
245. National Library of Medicine [Internet]. *ACTC1* gene [cited 2022 Mar 26].

Available from: https://www.ncbi.nlm.nih.gov/gene/70

246. National Library of Medicine [Internet]. *ACTN2* gene [cited 2022 Apr 2]. Available from: https://www.ncbi.nlm.nih.gov/gene/88
247. National Library of Medicine [Internet]. *ANKRD1* gene [cited 2022 Apr 15]. Available from: https://www.ncbi.nlm.nih.gov/gene/27063
248. The Human Protein Atlas [Internet]. *ANKRD1* gene [cited 2022 Apr 20]. Available from: https://www.proteinatlas.org/ENSG00000148677-ANKRD1
249. National Library of Medicine [Internet]. *CSRP3* gene [cited 2022 May 23]. Available from: https://www.ncbi.nlm.nih.gov/gene/8048
250. The Human Protein Atlas [Internet]. *CSRP3* [cited 2022 May 27]. Available from: https://www.proteinatlas.org/ENSG00000129170-CSRP3/tissue
251. National Library of Medicine [Internet]. *FHL1* gene [cited 2022 Apr 3]. Available from: https://www.ncbi.nlm.nih.gov/gene/2273
252. National Library of Medicine [Internet]. *GLA* gene [cited 2022 Apr 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/2717
253. National Library of Medicine [Internet]. *LAMP2* gene [cited 2022 Apr 6]. Available from: https://www.ncbi.nlm.nih.gov/gene/3920
254. National Library of Medicine [Internet]. *PLN* gene [cited 2022 Apr 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/5350
255. National Library of Medicine [Internet]. *PRKAG2* gene [cited 2022 Apr 9]. Available from: https://www.ncbi.nlm.nih.gov/gene/51422
256. National Library of Medicine [Internet]. *JPH2* gene [cited 2022 Apr 14]. Available from: https://www.ncbi.nlm.nih.gov/gene/57158
257. National Library of Medicine [Internet]. *DSG2* gene [cited 2022 Apr 20]. Available from: https://www.ncbi.nlm.nih.gov/gene/1829
258. National Library of Medicine [Internet]. *TRPM4* gene [cited 2022 Apr 23]. Available from: https://www.ncbi.nlm.nih.gov/gene/54795
259. National Library of Medicine [Internet]. *TTN* gene [cited 2022 Apr 30]. Available from: https://www.ncbi.nlm.nih.gov/gene/7273
260. National Library of Medicine [Internet]. *RYR1* gene [cited 2022 May 3]. Available from: https://www.ncbi.nlm.nih.gov/gene/6261

261. National Library of Medicine [Internet]. *NEXN* gene [cited 2022 May 3]. Available from: https://www.ncbi.nlm.nih.gov/gene/91624
262. National Library of Medicine [Internet]. *DSC2* gene [cited 2022 May 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/1824
263. National Library of Medicine [Internet]. *ABCC9* gene [cited 2022 May 8]. Available from: https://www.ncbi.nlm.nih.gov/gene/10060
264. National Library of Medicine [Internet]. *DSP* gene [cited 2022 May 10]. Available from: https://www.ncbi.nlm.nih.gov/gene/1832
265. National Library of Medicine [Internet]. *FBN1* gene [cited 2022 May 12]. Available from: https://www.ncbi.nlm.nih.gov/gene/2200
266. National Library of Medicine [Internet]. *CTF1* gene [cited 2022 May 15]. Available from: https://www.ncbi.nlm.nih.gov/gene/1489
267. National Library of Medicine [Internet]. *CACNA1C* gene [cited 2022 May 18]. Available from: https://www.ncbi.nlm.nih.gov/gene/775
268. National Library of Medicine [Internet]. *ELN* gene [cited 2022 May 23]. Available from: https://www.ncbi.nlm.nih.gov/gene/2006
269. National Library of Medicine [Internet]. *NOTCH1* gene [cited 2022 May 26]. Available from: https://www.ncbi.nlm.nih.gov/gene/4851
270. National Library of Medicine [Internet]. *PTPN11* gene [cited 2022 May 30]. Available from: https://www.ncbi.nlm.nih.gov/gene/5781
271. National Library of Medicine [Internet]. *MYH6* gene [cited 2022 Jun 2]. Available from: https://www.ncbi.nlm.nih.gov/gene/4624
272. National Library of Medicine [Internet]. *RBM20* gene [cited 2022 Jun 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/282996
273. National Library of Medicine [Internet]. *DMD* gene [cited 2022 Jun 8]. Available from: https://www.ncbi.nlm.nih.gov/gene/1756
274. Fortunato F, Farnè M, Ferlini A. The DMD gene and therapeutic approaches to restore dystrophin. Neuromuscul Disord. 2021;31(10):1013–20.
275. Keegan NP. Pseudoexons of the DMD Gene. J Neuromuscul Dis. 2020;7(2):77–95.
276. National Library of Medicine [Internet]. *LAMA4* gene [cited 2022 Jun 10].

Available from: https://www.ncbi.nlm.nih.gov/gene/3910

277. National Library of Medicine [Internet]. *SDHA* gene [cited 2022 Jun 12]. Available from: https://www.ncbi.nlm.nih.gov/gene/6389
278. National Library of Medicine [Internet]. *HCN4* gene [cited 2022 Jun 18]. Available from: https://www.ncbi.nlm.nih.gov/gene/10021
279. National Library of Medicine [Internet]. *PKP2* gene [cited 2022 Jun 20]. Available from: https://www.ncbi.nlm.nih.gov/gene/5318
280. National Library of Medicine [Internet]. *PDLIM3* gene [cited 2022 Jun 25]. Available from: https://www.ncbi.nlm.nih.gov/gene/27295
281. National Library of Medicine [Internet]. *NKX2-5* gene [cited 2022 Jun 30]. Available from: https://www.ncbi.nlm.nih.gov/gene/1482
282. National Library of Medicine [Internet]. *MYPN* gene [cited 2022 Jul 3]. Available from: https://www.ncbi.nlm.nih.gov/gene/84665
283. Glavaški M, Preveden A, Jakovljević Đ, Filipović N, Velicki L. Subtypes and mechanisms of hypertrophic cardiomyopathy proposed by machine learning algorithms. Life (Basel). 2022;12(10):1566.
284. Glavaški M, Velicki L. More slices, less truth: effects of different test-set design strategies for magnetic resonance image classification. Croat Med J. 2022;63(4):370–8.
285. Le TT, Fu W, Moore JH. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. Bioinformatics. 2020;36(1):250–6.
286. Hoksza D, Gawron P, Ostaszewski M, Smula E, Schneider R. MINERVA API and plugins: opening molecular network analysis and visualization to the community. Bioinformatics. 2019;35(21):4496–8.
287. Glavaški M, Velicki L. Humans and machines in biomedical knowledge curation: Hypertrophic cardiomyopathy molecular mechanisms' representation. BioData Min. 2021;14(1):45.
288. Valenzuela-Escárcega MA, Babur Ö, Hahn-Powell G, Bell D, Hicks T, Noriega-Atala E, et al. Large-scale automated machine reading discovers new cancer-driving mechanisms. Database (Oxford). 2018;2018:bay098.

289. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
290. Pillich RT, Chen J, Rynkov V, Welker D, Pratt D. NDEx: a community resource for sharing and publishing of biological networks. Methods Mol Biol. 2017;1558:271-301.
291. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEx, the Network Data Exchange. Cell Syst. 2015;1(4):302–5.
292. Pratt D, Chen J, Pillich R, Rynkov V, Gary A, Demchak B, et al. NDEx 2.0: a clearinghouse for research on cancer pathways. Cancer Res. 2017;77(21):e58–61.
293. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. CytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4:S11.
294. Cytoscape App Store [Internet]. Wk-shell-decomposition [cited 2022 Jun 24]. Available from: http://apps.cytoscape.org/apps/wkshelldecomposition
295. Zaki N, Efimov D, Berengueres J. Protein complex detection using interaction reliability assessment and weighted clustering coefficient. BMC Bioinformatics. 2013;14:163.
296. Gyori BM, Bachman JA, Subramanian K, Muhlich JL, Galescu L, Sorger PK. From word models to executable models of signaling networks using automated assembly. Mol Syst Biol. 2017;13(11):954.
297. Tadaka S, Kinoshita K. NCMine: Core-peripheral based functional module detection using near-clique mining. Bioinformatics. 2016;32(22):3454–60.
298. Hoksza D, Gawron P, Ostaszewski M, Schneider R. MolArt: a molecular structure annotation and visualization tool. Bioinformatics. 2018;34(23):4127–8.
299. Tang B, Song Y, Cui H, Ji K, Yu Q, Zhu C, et al. Prognosis of adult obstructive hypertrophic cardiomyopathy patients with different morphological types after surgical myectomy. Eur J Cardiothorac Surg. 2018;54(2):310–7.
300. Prinz C, Farr M, Hering D, Horstkotte D, Faber L. The diagnosis and treatment of hypertrophic cardiomyopathy. Dtsch Arztebl Int. 2011;108(13):209–15.
301. Maron BJ, Towbin JA, Thiene G, Antzelevitch C, Corrado D, Arnett D, et al.

Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. Circulation. 2006;113(14):1807–16.

302. Maron BJ, Gottdiener JS, Epstein SE. Patterns and significance of distribution of left ventricular hypertrophy in hypertrophic cardiomyopathy: a wide angle, two dimensional echocardiographic study of 125 patients. Am J Cardiol. 1981;48(3):418–28.
303. Parato VM, Antoncecchi V, Sozzi F, Marazia S, Zito A, Maiello M, et al. Echocardiographic diagnosis of the different phenotypes of hypertrophic cardiomyopathy. Cardiovasc Ultrasound. 2016;14(1):30.
304. Syed IS, Ommen SR, Breen JF, Tajik AJ. Hypertrophic cardiomyopathy: identification of morphological subtypes by echocardiography and cardiac magnetic resonance imaging. JACC Cardiovasc Imaging. 2008;1(3):377–9.
305. Kim EK, Lee SC, Hwang JW, Chang SA, Park SJ, On YK, et al. Differences in apical and non-apical types of hypertrophic cardiomyopathy: a prospective analysis of clinical, echocardiographic, and cardiac magnetic resonance findings and outcome from 350 patients. Eur Hear J Cardiovasc Imaging. 2016;17(6):678–86.
306. Yuan Y, Su W, Zhu M. Threshold-free measures for assessing the performance of medical screening tests. Front Public Health. 2015;3:57.
307. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432.
308. Seiler C, Jenni R, Vassalli G, Turina M, Hess OM. Left ventricular chamber dilatation in hypertrophic cardiomyopathy: related variables and prognosis in patients with medical and surgical therapy. Br Heart J. 1995;74(5):508–16.
309. Karaye KM, Habib AG, Mohammed S, Rabiu M, Shehu MN. Assessment of right ventricular systolic function using tricuspid annular-plane systolic excursion

in Nigerians with systemic hypertension. Cardiovasc J Afr. 2010;21(4):186–90.

310. Chida A, Inai K, Sato H, Shimada E, Nishizawa T, Shimada M, et al. Prognostic predictive value of gene mutations in Japanese patients with hypertrophic cardiomyopathy. Heart Vessels. 2017;32(6):700–7.
311. Song L, Zou Y, Wang J, Wang Z, Zhen Y, Lou K, et al. Mutations profile in Chinese patients with hypertrophic cardiomyopathy. Clin Chim Acta. 2005;351(1–2):209–16.
312. Parlak S, Gulcek S, Kaplanoglu H, Altin L, Deveer M, Pasaoglu L. Hepatic artery pseudoaneurysm: a life-threatening complication of liver ransplantation. J Belgian Soc Radiol. 2015;99(2):61–4.
313. Kazama I, Mori Y, Baba A, Nakajima T. Pitting type of pretibial edema in a patient with silent thyroiditis successfully treated by angiotensin ii receptor blockade. Am J Case Rep. 2014;15:111–4.
314. Nigussie B, Abaleka FI, Gemechu T, Suhail M, Alikhan S. Severe pulmonary hypertension and cholestatic liver injury: two rare manifestations of Graves' disease. Cureus. 2020;12(7):e9236.
315. Balal M, Paydas S, Seyrek N, Karayaylali I, Gonlusen G. Other glomerular pathologies in three patients with diabetes mellitus. Ren Fail. 2004;26(2):185–8.
316. Palermo P, Magrì D, Sciomer S, Stefanini E, Agalbato C, Compagnino E, et al. Delayed anaerobic threshold in heart failure patients with atrial fibrillation. J Cardiopulm Rehabil Prev. 2016;36(3):174–9.
317. Delgado V, Di Biase L, Leung M, Romero J, Tops LF, Casadei B, et al. Structure and function of the left atrium and left atrial appendage: AF and stroke implications. J Am Coll Cardiol. 2017;70(25):3157–72.
318. Thomas L, Abhayaratna WP. Left atrial reverse remodeling: mechanisms, evaluation, and clinical significance. JACC Cardiovasc Imaging. 2017;10(1):65–77.
319. Schumacher K, Dagres N, Hindricks G, Husser D, Bollmann A, Kornej J. Characteristics of PR interval as predictor for atrial fibrillation: association with biomarkers and outcomes. Clin Res Cardiol. 2017;106(10):767–75.
320. Kornej J, Zeynalova S, Thiery J, Burkhardt R, Baber R, Engel C, et al.

Association between echocardiographic parameters and biomarkers in probands with atrial fibrillation and different PR interval lengths: insight from the epidemiologic LIFE Adult Study. PLoS One. 2019;14(2):e0212627.

321. Ma J, Chen Q, Ma S. Left atrial fibrosis in atrial fibrillation: mechanisms, clinical evaluation and management. J Cell Mol Med. 2021;25(6):2764–75.
322. Wałek P, Ciesla E, Gorczyca I, Wożakowska-Kapłon B. Left atrial wall dyskinesia assessed during contractile phase as a predictor of atrial fibrillation recurrence after electrical cardioversion performed due to persistent atrial fibrillation. Medicine (Baltimore). 2020;99(49):e23333.
323. Aguilar M, Rose RA, Takawale A, Nattel S, Reilly S. New aspects of endocrine control of atrial fibrillation and possibilities for clinical translation. Cardiovasc Res. 2021;117(7):1645–61.
324. Roshdy HS, Soliman MH, El-Dosouky II, Ghonemy S. Skin aging parameters: a window to heart block. Clin Cardiol. 2018;41(1):51–6.
325. Attar R, El-Tallawi KC. Diastolic mitral regurgitation. Methodist Debakey Cardiovasc J. 2021;17(5):89–90.
326. Aksu U, Topcu S, Gulcu O, Kalkan K, Tanboga IH. Diastolic mitral and tricuspid regurgitation in a patient with 2:1 AV block. Int J Cardiol. 2015;195:111–2.
327. Ishikawa T, Kimura K, Miyazaki N, Tochikubo O, Usui T, Kashiwagi M, et al. Diastolic mitral regurgitation in patients with first-degree atrioventricular block. Pacing Clin Electrophysiol. 1992;15(11 Pt 2):1927–31.
328. Bay M, Vollenweider P, Marques-Vidal P, Schläpfer J. Clinical factors associated with the intraventricular conduction disturbances in Swiss middle-aged adults: the CoLaus|PsyCoLaus study. Int J Cardiol. 2021;327:201–8.
329. Francisco-Pascual J, Rivas-Gándara N, Bach-Oller M, Badia-Molins C, Maymi-Ballesteros M, Benito B, et al. Sex-related differences in patients with unexplained syncope and bundle branch block: lower risk of AV block and lesser need for cardiac pacing in women. Front Cardiovasc Med. 2022;9:838473.
330. Trullàs JC, Aguiló O, Mirò Ó, Díez-Manglano J, Carrera-Izquierdo M, Quesada-Simón MA, et al. Prevalence and impact on prognosis of right-bundle branch block in patients with acute heart failure: findings from the RICA registry. Rev

Clin Esp (Barc). 2022;222(5):272–80.

331. Bessem BB, de Bruijn MMC, Nieuwland WW. Gender differences in the electrocardiogram screening of athletes. J Sci Med Sport. 2017;20(2):213–7.
332. Rosa SA, Thomas B, Fiarresga A, Papoila AL, Alves M, Pereira R, et al. The impact of ischemia assessed by magnetic resonance on functional, arrhythmic, and imaging features of hypertrophic cardiomyopathy. Front Cardiovasc Med. 2021;8:761860.
333. Barefield DY, McNamara JW, Lynch TL, Kuster DWD, Govindan S, Haar L, et al. Ablation of the calpain-targeted site in cardiac myosin binding protein-C is cardioprotective during ischemia-reperfusion injury. J Mol Cell Cardiol. 2019;129:236–46.
334. Li X, Tang J, Li J, Lin S, Wang T, Zhou K, et al. Genetic clues on implantable cardioverter-defibrillator placement in young-age hypertrophic cardiomyopathy: a case report of novel *MYH7* mutation and literature review. Front Cardiovasc Med. 2021;8:810291.
335. van de Sande DA, Hoogsteen J, Theunissen LJ. An unusual presentation of a myocardial crypt in hypertrophic cardiomyopathy. Case Reports Cardiol. 2014;2014:737052.
336. Gruner C, Care M, Siminovitch K, Moravsky G, Wigle ED, Woo A, et al. Sarcomere protein gene mutations in patients with apical hypertrophic cardiomyopathy. Circ Cardiovasc Genet. 2011;4(3):288–95.
337. Chung H, Kim Y, Cho SM, Lee HJ, Park CH, Kim JY, et al. Differential contributions of sarcomere and mitochondria-related multigene variants to the endophenotype of hypertrophic cardiomyopathy. Mitochondrion. 2020;53:48–56.
338. Waldmüller S, Erdmann J, Binner P, Gelbrich G, Pankuweit S, Geier C, et al. Novel correlations between the genotype and the phenotype of hypertrophic and dilated cardiomyopathy: results from the German Competence Network Heart Failure. Eur J Heart Fail. 2011;13(11):1185–92.
339. Mori AA, Castro LR, Bortolin RH, Bastos GM, Oliveira VF, Ferreira GM, et al. Association of variants in *MYH7*, *MYBPC3* and *TNNT2* with sudden cardiac death-related risk factors in Brazilian patients with hypertrophic cardiomyopathy.

Forensic Sci Int Genet. 2021;52:102478.

340. Tarkiainen M, Sipola P, Jalanko M, Heliö T, Jääskeläinen P, Kivelä K, et al. CMR derived left ventricular septal convexity in carriers of the hypertrophic cardiomyopathy-causing *MYBPC3*-Q1061X mutation. Sci Rep. 2019;9(1):5960.
341. Liu T, Song D, Dong J, Zhu P, Liu J, Liu W, et al. Current understanding of the pathophysiology of myocardial fibrosis and its quantitative assessment in heart failure. Front Physiol. 2017;8:238.
342. Schirone L, Forte M, Palmerio S, Yee D, Nocella C, Angelini F, et al. A review of the molecular mechanisms underlying the development and progression of cardiac remodeling. Oxid Med Cell Longev. 2017;2017:3920195.
343. Petersen SE, Jerosch-Herold M, Hudsmith LE, Robson MD, Francis JM, Doll HA, et al. Evidence for microvascular dysfunction in hypertrophic cardiomyopathy: new insights from multiparametric magnetic resonance imaging. Circulation. 2007;115(18):2418–25.
344. Cecchi F, Olivotto I, Gistri R, Lorenzoni R, Chiriatti G, Camici PG. Coronary microvascular dysfunction and prognosis in hypertrophic cardiomyopathy. N Engl J Med. 2003;349(11):1027–35.
345. Maron MS, Olivotto I, Maron BJ, Prasad SK, Cecchi F, Udelson JE, et al. The case for myocardial ischemia in hypertrophic cardiomyopathy. J Am Coll Cardiol. 2009;54(9):866–75.
346. Raphael CE, Mitchell F, Kanaganayagam GS, Liew AC, Di Pietro E, Vieira MS, et al. Cardiovascular magnetic resonance predictors of heart failure in hypertrophic cardiomyopathy: the role of myocardial replacement fibrosis and the microcirculation. J Cardiovasc Magn Reson. 2021;23(1):26.
347. Jordà P, García-Álvarez A. Hypertrophic cardiomyopathy: sudden cardiac death risk stratification in adults. Glob Cardiol Sci Pract. 2018;2018(3):25.
348. Waldmann V, Jouven X, Narayanan K, Piot O, Chugh SS, Albert CM, et al. Association between atrial fibrillation and sudden cardiac death: pathophysiological and epidemiological insights. Circ Res. 2020;127(2):301–9.
349. O'Mahony C, Elliott P, McKenna W. Sudden cardiac death in hypertrophic cardiomyopathy. Circ Arrhythm Electrophysiol. 2013;6(2):443–51.

План третмана података

Назив пројекта/истраживања
Утврђивање повезаности генотипа и фенотипа хипертрофичне кардиомиопатије применом машинског учења
Назив институције/институција у оквиру којих се спроводи истраживање
Институт за кардиоваскуларне болести Војводине, Република Србија
Назив програма у оквиру ког се реализује истраживање
Докторске академске студије, Претклиничка истраживања
1. Опис података

1.1 Врста студије

Укратко описати тип студије у оквиру које се подаци прикупљају

Ретропроспективна студија.

1.2 Врсте података

а) квантитативни

б) квалитативни

1.3. Начин прикупљања података

а) анкете, упитници, тестови

б) **клиничке процене, медицински записи, електронски здравствени записи**

в) генотипови: **ПЦР и НГС**

г) административни подаци

д) узорци ткива

ђ) снимци, фотографије: **слике ултразвучног прегледа**

е) текст: **доступна научна литература**

ж) мапа

з) остало: описати _____

1.3 Формат података, употребљене скале, количина података

1.3.1 Употребљени софтвер и формат датотеке:

а) **Excel фајл**

b) **SPSS фајл**

c) **PDF фајл**

d) **Текст фајл**

e) **JPG фајл**

f) Остало, датотека _____

1.3.2. Број записа (код квантитативних података)

а) број варијабли: велики број варијабли

б) број мерења (испитаника, процена, снимака и сл.): **143 испитаника**

1.3.3. Поновљена мерења

а) да

б) не

Уколико је одговор да, одговорити на следећа питања:

а) временски размак између поновљених мера је **4 месеца**

б) варијабле које се више пута мере односе се на **различите посете испитаника**

в) нове верзије фајлова који садрже поновљена мерења су именоване као _____

Напомене: _____

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?

а) Да

б) Не

Ако је одговор не, образложити _____

_____

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?

а) експеримент, навести тип _____

б) корелационо истраживање, навести тип _____

ц) анализа текста, навести тип _____

д) остало, навести шта: **Ретропроспективна студија, подаци сакупљени током истраживачког пројекта**

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

Велики број мерних инструмената.

2.2 Квалитет података и стандарди

2.2.1. Третман недостајућих података

а) Да ли матрица садржи недостајуће податке? **Да** Не

Ако је одговор да, одговорити на следећа питања:

а) Колики је број недостајућих података? **Различит у зависности од варијабле о којој се ради.**

б) Да ли се кориснику матрице препоручује замена недостајућих података? Да **Не**

в) Ако је одговор да, навести сугестије за третман замене недостајућих података

_____

2.2.2. На који начин је контролисан квалитет података? Описати

_____

_____

2.2.3. На који начин је извршена контрола уноса података у матрицу?

_____

_____

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у ***Репозиторијуму докторских дисертација Универзитета у Новом Саду.***

3.1.2. URL адреса ***https://cris.uns.ac.rs/searchDissertations.jsf***

3.1.3. DOI _____

3.1.4. Да ли ће подаци бити у отвореном приступу?

а) ***Да***

б) *Да, али после ембарга који ће трајати до* _____

в) *Не*

Ако је одговор не, навести разлог _____

3.1.5. Подаци неће бити депоновани у репозиторијум, али ће бити чувани.

Образложење

_____

_____

3.2 Метаподаци и документација података

3.2.1. Који стандард за метаподатке ће бити примењен? _____

3.2.1. Навести метаподатке на основу којих су подаци депоновани у репозиторијум.

_____

_____

Ако је потребно, навести методе које се користе за преузимање података, аналитичке и процедуралне информације, њихово кодирање, детаљне описе варијабли, записа итд.

_____

_____

_____

_____

3.3 Стратегија и стандарди за чување података

3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? **Трајно**

3.3.2. Да ли ће подаци бити депоновани под шифром? **Да** Не

3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? **Да** Не

3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена?

Да **Не**

Образложити

_____

_____

4. Безбедност података и заштита поверљивих информација

Овај одељак МОРА бити попуњен ако ваши подаци укључују личне податке који се односе на учеснике у истраживању. За друга истраживања треба такође размотрити заштиту и сигурност података.

4.1 Формални стандарди за сигурност информација/података

Истраживачи који спроводе испитивања с људима морају да се придржавају Закона о заштити података о личности *(https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html)* и одговарајућег институционалног кодекса о академском интегритету.

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? **Да** Не

Ако је одговор Да, навести датум и назив етичке комисије која је одобрила истраживање

16.07.2020., Медицински факултет Нови Сад

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Да **Не**

Ако је одговор да, наведите на који начин сте осигурали поверљивост и сигурност информација везаних за испитанике:

а) Подаци нису у отвореном приступу

б) Подаци су анонимизирани

ц) Остало, навести шта

_____

_____

5. Доступност података

5.1. Подаци ће бити

а) јавно доступни

б) доступни само уском кругу истраживача у одређеној научној области

ц) затворени

Ако су подаци доступни само уском кругу истраживача, навести под којим условима могу да их користе:

_____

_____

Ако су подаци доступни само уском кругу истраживача, навести на који начин могу приступити подацима:

_____

_____

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.

Ауторство – некомерцијално – делити под истим условима

6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података

Мила Главашки, milaglavaski@uns.ac.rs

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима

Мила Главашки, milaglavaski@uns.ac.rs

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима

Мила Главашки, milaglavaski@uns.ac.rs