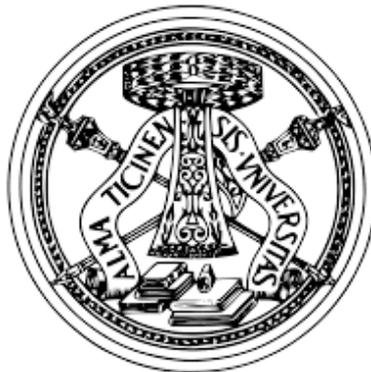


**Intelligenza Artificiale in Cardiogenetica:  
utilizzo dell'algoritmo predittivo Mutscore  
in 15 anni di pratica clinica**

**Artificial Intelligence in Cardiogenetics:  
the Mutscore Algorithm in a single center 15-year experience**

Dr. Alessandra Pia Porretta MD, PhD-student



Tutor:  
Prof. Lorenzo Cobianchi  
Università degli Studi di Pavia

Co-Tutor:  
Prof. Andrea Superti-Furga  
Università degli Studi di Losanna

Dottorato di Ricerca in Medicina Sperimentale XXXV ciclo  
2019-2022

## INDICE

<b>1. ABSTRACT.....</b>	<b>pag. 4</b>
<b>2. LA CARDIOGENETICA.....</b>	<b>pag. 8</b>
<b>2.1. Definizione.....</b>	<b>pag. 8</b>
<b>2.2. Le cardiomiopatie ereditarie.....</b>	<b>pag. 8</b>
<b>2.3. Le sindromi aritmiche ereditarie.....</b>	<b>pag. 9</b>
<b>2.4. Modello di trasmissione.....</b>	<b>pag. 9</b>
<b>2.5. Test genetico nella pratica clinica .....</b>	<b>pag. 10</b>
<b>2.6. Criteri di classificazione.....</b>	<b>pag. 11</b>
<b>2.7 Implicazioni pratiche del risultato genetico.....</b>	<b>pag. 14</b>
<b>2.8. Modelli predittivi computazionali.....</b>	<b>pag. 18</b>
<b>2.9 Mutscore e MutLand.....</b>	<b>pag. 19</b>
<b>3. OBIETTIVI DELLO STUDIO.....</b>	<b>pag. 26</b>
<b>3.1 Obiettivo primario.....</b>	<b>pag. 27</b>
<b>3.2. Obiettivo secondario.....</b>	<b>pag. 27</b>
<b>4. MATERIALI E METODI.....</b>	<b>pag. 28</b>
<b>4.1 Disegno dello studio.....</b>	<b>pag. 28</b>
<b>4.2 Popolazione dello studio.....</b>	<b>pag. 28</b>
<b>4.3 Raccolta dei dati.....</b>	<b>pag. 29</b>
<b>4.4 Analisi statistica.....</b>	<b>pag. 31</b>
<b>5. RISULTATI .....</b>	<b>pag. 32</b>
<b>5.1 Popolazione di studio.....</b>	<b>pag. 32</b>
<b>5.2 Varianti identificate nella popolazione di studio.....</b>	<b>pag. 33</b>
<b>5.3 Analisi delle varianti probabilmente patogene/patogene.....</b>	<b>pag. 39</b>
<b>5.4 Analisi delle varianti di significato incerto.....</b>	<b>pag. 41</b>
<b>5.5 Calcolo dello score predittivo Mutscore.....</b>	<b>pag. 41</b>

<b>6. DISCUSSIONE.....</b>	<b>pag. 46</b>
<b>6.1 Analisi dei risultati della popolazione di studio in relazione ai dati di letteratura..</b>	<b>pag.46</b>
<b>6.2 Applicazione dell’algoritmo predittivo Mutscore nella nostra coorte di studio.....</b>	<b>pag 50</b>
<b>7. CONCLUSIONE.....</b>	<b>pag. 51</b>
<b>8. BIBLIOGRAFIA.....</b>	<b>pag. 53</b>
<b>9. ANNESSI.....</b>	<b>pag. 55</b>

## **1. ABSTRACT**

### **Background**

Negli ultimi decenni, l'introduzione delle tecnologie di sequenziamento genetico di nuova generazione (NGS) ha cambiato profondamente la pratica della genetica medica. Se da un lato l'analisi dell'intero genoma, degli esoni o di specifici geni causali inclusi in pannelli multigenici, ha permesso un più ampio accesso all'analisi genetica nella pratica di routine, dall'altro ha sollevato nuove e difficili sfide nell'interpretazione dei risultati. A causa infatti dell'aumentato numero di test genetici, è divenuto sempre più frequente il riscontro di nuove varianti di sequenza che, differendo dalla sequenza standard della popolazione generale di riferimento, rimangono di significato incerto.

### **Obiettivi**

Lo studio attuale presenta l'esperienza quindicinale della consultazione congiunta di cardiogenetica dei servizi di Cardiologia e di Medicina Genetica dell'Ospedale Universitario di Losanna, relativa alla coorte di pazienti indirizzati per sospetto di malattia ereditaria cardiaca e arruolati nel registro prospettico delle malattie cardiache ereditarie del Servizio di Genetica Medica. Previa identificazione e analisi delle varianti probabilmente patogene/patogene e di significato incerto riscontrate, lo studio prevede l'applicazione del nuovo algoritmo computazionale predittivo Mutscore, basato su un approccio di apprendimento automatico ("machine learning"), al fine di testare, su un nuovo set di dati, la performance predittiva dell'algoritmo e la capacità di riclassificazione delle varianti di significato incerto.

### **Metodi**

Il nostro studio ha comportato l'analisi retrospettiva dei risultati di sequenziamento del DNA dei pazienti testati per sospetta malattia ereditaria cardiaca e indirizzati al nostro centro dal 1° luglio 2007 al 31 dicembre 2021. Per ogni variante probabilmente patogena/patogena e per le varianti di significato incerto abbiamo quindi applicato l'algoritmo Mutscore precedentemente descritto, utilizzando un approccio random forest.

## **Risultati**

Sui 488 pazienti testati, l'indicazione clinica più frequente è rappresentata dallo screening presintomatico (37,3%) in caso di variante familiare conosciuta, seguita dall'indicazione per sospetta sindrome aritmica (28,3%) e per cardiomiopatia ereditaria diagnosticata o sospetta (26,8%). Una variante patogena/probabilmente patogena è stata riscontrata in 198 pazienti per una resa diagnostica globale del 40,6% e del 23,6% nei soli casi indice. Il tasso di mutation-carriers identificati è risultato maggiore (45,8%) nei pazienti con cardiomiopatia rispetto a quelli (37%) affetti da sindromi aritmiche ereditarie. Le varianti patogene sono state identificate principalmente nei geni con chiara evidenza scientifica di associazione causale alla malattia. Abbiamo dimostrato l'ottima capacità predittiva del Mutscore sul nostro set di dati, come attestato da una AUC (Area Under the Curve) di 0,895 e dalla correlazione positiva statisticamente significativa ( $r= 0,65$ ,  $p\text{-value} = 0.000$ ) tra il Mutscore e l'interpretazione della variante annotata in ClinVar. Abbiamo infine documentato, sulla base del Mutscore, la possibile riclassificazione in varianti probabilmente benigne/patogene per il 47,2% delle varianti di significato incerto.

## **Conclusione**

Lo studio ha permesso di dimostrare, attraverso l'esperienza di 15 anni di consultazioni specializzate congiunte in Cardiogenetica, il ruolo fondamentale dell'analisi genetica nella conferma diagnostica (nel 40% dei pazienti testati) e nella presa in carico dei pazienti con sospetta malattia cardiaca ereditaria. Sulla base di questa solida esperienza clinica, abbiamo inoltre dimostrato, nella nostra coorte di pazienti, l'ottima performance predittiva dell'algorithmo computazionale Mutscore e la sua accuratezza nell'aiutare a riclassificare le varianti di significato incerto. Sebbene ulteriori studi di validazione siano ancora necessari, il Mutscore sembra essere un valido strumento ed una potente risorsa in grado di contribuire alla riclassificazione delle varianti di significato incerto e all'ulteriore consolidamento di un approccio orientato verso la medicina di precisione.

## **ABSTRACT**

### **Background**

In the last decades, the implementation of high-throughput next-generation (NGS) technologies has profoundly changed the landscape of human genome sequencing. The analysis of whole genome, exomes or deputed genes in the context of multigene panels, while being rapidly and extensively available in the routine practice, has raised new and difficult challenges in result interpretation. Among others, due to the increase of genetic specimen testing, novel sequence variants have become an increasingly frequent finding which, differing from the standard sequence of the referenced general population, remain of unknown/uncertain significance.

### **Objective**

We present hereby our single center experience over a 15-year period on a cohort of consecutive patients addressed for suspicion of cardiac hereditary disease to the specialized cardiogenetic outpatient clinic of the University Hospital of Lausanne and enrolled in the prospective registry of hereditary cardiac diseases of the Service of Medical Genetics. After the identification and careful analysis of likely pathogenic/pathogenic variants and of variants of uncertain significance, we applied the new computational predictive algorithm Mutscore, based on a machine-learning approach, to test, in a new dataset, Mutscore performance as well as its ability in reclassifying variants of uncertain significance.

### **Methods**

Our study implied a retrospective review of DNA sequencing results of patients tested for suspected cardiac hereditary disease and addressed at our center from July 1<sup>st</sup> 2007 through December 31<sup>st</sup> 2021. For each likely pathogenic/pathogenic variants and for variants of uncertain significance we then applied the previously described Mutscore algorithm using a random forest approach.

## **Results**

Among the 488 tested patients, the most frequent clinical indications were presymptomatic screening (37.3%) proposed in case of a known familial variant, followed by suspected arrhythmic syndrome (28.3%) and hereditary cardiomyopathy (26.8%). A likely pathogenic/pathogenic variant was found in 198 patients for an overall diagnostic yield of 40.6% and of 23.6% considering only index cases. The rate of identified mutation-carriers was higher among patients with cardiomyopathy (45.8%) than among those with inherited arrhythmic syndromes (37%). The pathogenic variants were mainly identified in genes with definitive scientific evidence of causal association with the disease. We demonstrated the excellent predictive performance of the Mutscore in our dataset, as attested by an AUC (Area Under the Curve) of 0.895 and by the statistically significant positive correlation ( $r = 0.65$ ,  $p\text{-value} = 0.000$ ) between the Mutscore and the variant interpretation reported in ClinVar. We finally documented, according to Mutscore, the possible reclassification into likely benign/pathogenic variants for the 47.2% of the variants of uncertain significance.

## **Conclusions**

The 15-year experience of our specialized cardiogenetic outpatient clinic has demonstrated the pivotal role of the genetic analysis in diagnostic confirmation (in 40% of the patients tested) and in therapeutic management of patients with suspected cardiac hereditary diseases. Based on this solid clinical experience, we have also demonstrated in our dataset the excellent predictive performance of the Mutscore computational algorithm and its accuracy in reclassifying variants of uncertain significance. Although further validation studies are still needed, the Mutscore algorithm seems to be a valuable tool and a powerful resource contributing to disambiguation of variants of uncertain significance and to the further build the pathway towards precision medicine.

## **2. LA CARDIOGENETICA**

### **2.1 Definizione**

La cardio-genetica rappresenta una branca emergente della medicina caratterizzata dalla diagnosi clinica, dall'analisi genetica previo counseling e dalla gestione terapeutica di patologie cardiache per le quali un'origine genetica è stata dimostrata<sup>1</sup>. Le malattie cardiache ereditarie includono, in aggiunta alle cardiopatie congenite, due principali gruppi di patologie gravate da una morbidità e una mortalità significativa a carico di soggetti di giovane età: le cardiomiopatie ereditarie e le sindromi aritmiche ereditarie. In virtù della diversità delle patologie in esame e della molteplicità di competenze necessarie alla loro presa in carico, la cardio-genetica si basa un approccio multidisciplinare integrato che vede il contributo di diverse professionalità, quali cardiologi adulti, cardiopediatri, genetisti, consiglieri in genetica e psicologi.

### **2.2 Le cardiomiopatie ereditarie**

Le cardiomiopatie ereditarie consistono in una alterazione strutturale e funzionale del miocardio di origine genetica e come tale non spiegabile dalle condizioni emodinamiche di carico del miocardio né dalla presenza di una malattia coronarica<sup>2,3</sup>. Da un punto di vista morfo-funzionale, esse sono classificate in cinque categorie maggiori che includono la cardiomiopatia dilatativa, la cardiomiopatia ipertrofica, la cardiomiopatia restrittiva, la cardiomiopatia aritmogena del ventricolo destro e le cardiomiopatie dette "non classificate" come la cardiomiopatia di tipo non-compaction del ventricolo sinistro. La loro prevalenza, precedentemente sottostimata, varia a seconda dei tipi da 1:200 a 1:5000 e la loro importante morbidità-mortalità è gravata da un impatto socio-economico maggiore. La cardiomiopatia ipertrofica rappresenta infatti la causa più frequente di morte improvvisa tra i giovani atleti<sup>2,3</sup> e la cardiomiopatia dilatativa è invece la causa più frequente di trapianto cardiaco<sup>2,4</sup>.

### **2.3 Le sindromi aritmiche ereditarie**

Le sindromi aritmiche ereditarie sono delle malattie puramente elettriche, caratterizzate dalla totale assenza di anomalie cardiache strutturali. Definite anche canalopatie, esse sono infatti causate dalla presenza di varianti patogene a livello di geni codanti per canali ionici implicati nella genesi del potenziale d'azione cardiaco. Le sindromi aritmiche ereditarie includono: la sindrome del QT lungo congenito, la Sindrome di Brugada, la sindrome del QT corto e la tachicardia ventricolare polimorfa catecolaminergica. Da un punto di vista epidemiologico, esse rappresentano delle patologie rare, la cui prevalenza è il frutto di una stima approssimativa, in assenza di studi sistematici basati sulla popolazione<sup>5</sup>. In generale, la prevalenza varia tra 1:2000-2500 per la Sindrome del QT lungo e la Sindrome di Brugada, a 1:10000 per la tachicardia ventricolare polimorfa catecolaminergica.

### **2.4 Modello di trasmissione**

Le cardiomiopatie e le sindromi aritmiche ereditarie sono tradizionalmente considerate delle malattie monogeniche Mendeliane, caratterizzate dall'ereditarietà di varianti genetiche patogene a carico di un singolo gene. In particolare, la forma di trasmissione può essere autosomica dominante, autosomica recessiva o legata al sesso. La trasmissione autosomica dominante rappresenta la modalità di trasmissione più frequente nella quale l'ereditarietà di un singolo allele mutato è sufficiente per determinare la patologia. Nella maggior parte dei casi, la variante patogena è ereditata dal padre o dalla madre, ma in una percentuale inferiore di casi la variante occorre "de novo" nel gene, in soggetti senza storia familiare di malattia<sup>6</sup>.

Un aspetto peculiare di tali patologie Mendeliane è rappresentato dal fenomeno della penetranza incompleta e dell'espressività variabile<sup>6</sup>. Il primo rappresenta un fenomeno "quantitativo" tale per cui non tutti i portatori di una data variante patogena esprimono il fenotipo. Il secondo invece rappresenta un fenomeno "qualitativo" per cui, tra i soggetti che esprimono il fenotipo, il quadro clinico può essere estremamente diverso, e come tale, gravato da manifestazioni di severità differente (anche tra soggetti portatori della stessa mutazione all'interno di una stessa famiglia). Sulla base di

recenti evidenze, la penetranza e l'espressione di tali patologie sembrano infatti essere determinate non solamente dalla mutazione monogenica presente, ma da una complessa interazione tra background genetico (e dunque la possibile co-ereditarietà di altri fattori genetici diversi dalla mutazione in questione) e fattori ambientali che contribuirebbero alla modulazione del fenotipo. Ecco perché, come recentemente statuito nel documento consensuale di esperti sullo stato del test genetico per le malattie cardiache, l'oggettivazione di una variante patogena presenta "un alto valore diagnostico ma una bassa utilità prognostica"<sup>6</sup>. Il contributo di fattori modificatori del fenotipo diversi dalla mutazione coinvolta ostacola infatti attualmente le correlazioni genotipo-fenotipo per la gran parte di queste patologie, impedendo di predire il decorso clinico di malattia nei portatori di mutazione.

Un'ulteriore caratteristica di tali patologie è l'eterogeneità genetica/allelica, dimostrata dall'identificazione di numerosi geni implicati in ciascun sottogruppo di cardiomiopatia, secondo il paradigma "one disease-many different genes"<sup>2,5</sup>. In aggiunta, gli avanzamenti recenti hanno permesso di evidenziare come uno stesso gene possa essere implicato nella manifestazione di fenotipi sovrapposti ("overlap syndromes"), secondo il più recente paradigma "one gene-many diseases"<sup>2,5</sup>.

## **2.5 Test genetico nella pratica clinica**

Secondo il recente documento consensuale di esperti sullo stato del test genetico per le malattie cardiache<sup>6</sup>, il principio di base di qualunque test genetico clinico è incentrato sull'analisi dei soli geni causalmente associati con forte evidenza scientifica alla patologia in esame. Qualora vengano testati geni con limitata evidenza scientifica in favore della loro associazione causale, i risultati di analisi rischerebbero infatti di essere gravati da un'interpretazione incorretta con conseguente diagnosi errata e ripercussioni terapeutiche inappropriate per i pazienti coinvolti. Per queste ragioni e al fine di uniformare la pratica genetica, uno sforzo scientifico notevole è stato sostenuto al fine di elaborare delle classificazioni "evidence-based" per le malattie cardiache ereditarie Mendeliane al fine di

distinguere i geni ritenuti causali con forte evidenza scientifica da quelli con evidenza limitata o ambigua<sup>6,8</sup>.

## 2.6 Criteri di classificazione

L'American College di Genetica e Genomica Medica (ACMG) e l'associazione di Patologia Molecolare hanno elaborato nel 2015<sup>9</sup> un documento consensuale di raccomandazioni per garantire l'utilizzo di una nomenclatura e di criteri standardizzati al fine di uniformare l'interpretazione delle varianti genetiche e di permettere l'effettiva condivisione dell'informazione genetica nella pratica clinica. Una prima importante modifica introdotta è rappresentata dalla sostituzione dei termini "mutazione" (definito come un cambiamento permanente della sequenza nucleotidica) e "polimorfismo" (definito come una variante con frequenza >1%) con quello di "variante" il cui significato clinico viene definito usando una classificazione, basata sulle evidenze scientifiche attuali, secondo il grado di probabilità di associazione causale della variante alla patologia in questione. In particolare, la classificazione si basa sulla valutazione integrata di molteplici criteri che includono dati di popolazione (frequenza dell'allele nella popolazione sana e con malattia), dati computazionali o ottenuti da algoritmi predittivi, dati ottenuti da studi funzionali o di segregazione familiare e informazioni inerenti alla localizzazione topografica entro domini funzionali rilevanti o meno della proteina<sup>6,9</sup>. In particolare, nessun singolo criterio risulta sufficiente per inferire la patogenicità di una variante che deve essere quindi corroborata da più criteri. **Figura 1.**

L'integrazione globale di tali criteri permette quindi la classificazione della variante in cinque possibili categorie:

1. Benigna
2. Probabilmente benigna
3. Variante di significato incerto
4. Probabilmente patogena
5. Patogena

**Figura 1. Criteri integrati nel sistema di classificazione delle varianti**

	Benign			Pathogenic		
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
<b>Population Data</b>	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
<b>Computational And Predictive Data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
<b>Functional Data</b>	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
<b>Segregation Data</b>	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
<b>De novo Data</b>				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
<b>Allelic Data</b>		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
<b>Other Database</b>		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
<b>Other Data</b>		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

La seguente tabella riassume i diversi criteri integrati nel sistema di classificazione. Ogni criterio è quindi valutato in funzione delle evidenze scientifiche attuali e della loro forza in supporto dell'associazione causale patogena (lato destro) o benigna (lato sinistro) alla patologia. Riprodotto da Richards S et al<sup>9</sup>. Abbreviazioni: BS: benigno forte; BP a supporto benigno; FH: storia familiare; LOF: loss-of-function; MAF: frequenza allelica minore; path: patogeno; PM: patogeno moderato; PP; a support patogeno; PS: patogeno forte; PVS: patogeno molto forte.

Piu in dettaglio, specifici criteri sono stati elaborati separatamente per le varianti di tipo 1-2 (benigne e probabilmente benigne) e quelle di tipo 4-5 (patogene e probabilmente patogene). Sulla base del peso e dell'entità delle evidenze scientifiche disponibili, i criteri di patogenicità sono stati ulteriormente graduati in “molto forti” (PVS1), forti (PS1-4), moderati (PM1-6) o in supporto (PP1-5). I criteri di benignità si dividono invece in “a sé stante” (BA1), “forti” (BS1-4) o “in supporto” (BP1-6). **Figura 2 e 3.**

**Figura 2. Criteri di classificazione delle varianti patogene**

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical <math>\pm 1</math> or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> <li>• Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>)</li> <li>• Use caution interpreting LOF variants at the extreme 3' end of a gene</li> <li>• Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact</li> <li>• Use caution in the presence of multiple transcripts</li> </ul>
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change</p> <p>Example: Val→Leu caused by either G&gt;C or G&gt;T in the same codon</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (<b>both</b> maternity and paternity confirmed) in a patient with the disease and no family history</p> <p>Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product</p> <p>Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls</p> <p>Note 1: Relative risk or OR, as obtained from case-control studies, is &gt;5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance.</p> <p>Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p> <p>PM2 Absent from controls (or at extremely low frequency if recessive) (<b>Table 6</b>) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium</p> <p>Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing.</p> <p>PM3 For recessive disorders, detected in <i>trans</i> with a pathogenic variant</p> <p>Note: This requires testing of parents (or offspring) to determine phase.</p> <p>PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants</p> <p>PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before</p> <p>Example: Arg156His is pathogenic; now you observe Arg156Cys</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.</p> <p>PM6 Assumed de novo, but without confirmation of paternity and maternity</p>
Supporting	<p>PP1 cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease</p> <p>Note: May be used as stronger evidence with increasing segregation data</p> <p>PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease</p> <p>PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)</p> <p>Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.</p> <p>PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology</p> <p>PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation</p>

LOF, loss of function; OR, odds ratio.

Riprodotta da Richards S et al<sup>9</sup>.

### Figura 3. Criteri di classificazione delle varianti benigne

Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see <a href="#">Table 6</a> )
	BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
	BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
	BS4 Lack of segregation in affected members of a family  Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease
	BP2 Observed in <i>trans</i> with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in <i>cis</i> with a pathogenic variant in any inheritance pattern
	BP3 In-frame deletions/insertions in a repetitive region without a known function
	BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)  Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
	BP5 Variant found in a case with an alternate molecular basis for disease
	BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
	BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Riprodotta da Richards S et al<sup>9</sup>.

Una volta definiti i criteri di interesse e il grado o “peso” di evidenza scientifica correlata, la definizione finale della variante avviene secondo un preciso schema classificativo che permette di assicurare, per ogni variante definita come probabilmente maligna o benigna, un grado di sicurezza >90% di essere effettivamente patogena o meno. **Figura 4**. Le varianti di significato incerto sono invece definite o dal mancato soddisfacimento dei criteri sopra menzionati o in caso di evidenze conflittuali in merito al loro ruolo causale. Esse sono rappresentate normalmente da nuove varianti per le quali non sono disponibili evidenze scientifiche in letteratura o analisi funzionali e con frequenze alleliche basse che non permettono di classificarle come polimorfismi<sup>2</sup>.

#### 2.7 Implicazioni pratiche del risultato genetico

In caso di riscontro di variante probabilmente patogena o patogena (4 o 5), la ripercussione pratica è duplice. Da un lato il risultato permette essenzialmente la conferma diagnostica della patologia e, in alcuni casi, fornisce informazioni prognostiche e terapeutiche rilevanti. Per altre patologie cardiache

ereditarie il dato genetico non è tuttavia in grado di modificare sostanzialmente la presa in carico o di fornire informazioni prognostiche a causa delle difficoltà ancora attuali nell'identificazione di solide correlazioni genotipo-fenotipo. **Figura 5.**

**Figura 4. Regole per la classificazione delle varianti basate sull'integrazione dei criteri previsti.**

Pathogenic	<ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) AND               <ul style="list-style-type: none"> <li>(a) <math>\geq 1</math> Strong (PS1–PS4) OR</li> <li>(b) <math>\geq 2</math> Moderate (PM1–PM6) OR</li> <li>(c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) OR</li> <li>(d) <math>\geq 2</math> Supporting (PP1–PP5)</li> </ul> </li> <li>(ii) <math>\geq 2</math> Strong (PS1–PS4) OR</li> <li>(iii) 1 Strong (PS1–PS4) AND               <ul style="list-style-type: none"> <li>(a) <math>\geq 3</math> Moderate (PM1–PM6) OR</li> <li>(b) 2 Moderate (PM1–PM6) AND <math>\geq 2</math> Supporting (PP1–PP5) OR</li> <li>(c) 1 Moderate (PM1–PM6) AND <math>\geq 4</math> supporting (PP1–PP5)</li> </ul> </li> </ul>
Likely pathogenic	<ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) AND 1 moderate (PM1–PM6) OR</li> <li>(ii) 1 Strong (PS1–PS4) AND 1–2 moderate (PM1–PM6) OR</li> <li>(iii) 1 Strong (PS1–PS4) AND <math>\geq 2</math> supporting (PP1–PP5) OR</li> <li>(iv) <math>\geq 3</math> Moderate (PM1–PM6) OR</li> <li>(v) 2 Moderate (PM1–PM6) AND <math>\geq 2</math> supporting (PP1–PP5) OR</li> <li>(vi) 1 Moderate (PM1–PM6) AND <math>\geq 4</math> supporting (PP1–PP5)</li> </ul>
Benign	<ul style="list-style-type: none"> <li>(i) 1 Stand-alone (BA1) OR</li> <li>(ii) <math>\geq 2</math> Strong (BS1–BS4)</li> </ul>
Likely benign	<ul style="list-style-type: none"> <li>(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) OR</li> <li>(ii) <math>\geq 2</math> Supporting (BP1–BP7)</li> </ul>
Uncertain significance	<ul style="list-style-type: none"> <li>(i) Other criteria shown above are not met OR</li> <li>(ii) the criteria for benign and pathogenic are contradictory</li> </ul>

Riprodotta da Richards S et al<sup>9</sup>. \* Le varianti di significato incerto sono definite nel caso in cui i criteri sopracitati non siano riempiti o in caso di criteri contraddittori.

Dall'altro lato, tale risultato genetico consente l'attuazione di uno screening presintomatico tra i membri della famiglia a rischio di aver ereditato la variante del caso indice. Questa pratica è raccomandata in particolare quando l'eventuale riscontro della variante nei familiari determina delle

ripercussioni clinico-terapeutiche nei portatori, permettendo invece di rassicurare e liberare dalla necessità di una sorveglianza clinica i non portatori.

**Figura 5. Ripercussione diagnostica, prognostica e terapeutica del test genetico**

Disease	Diagnostic	Prognostic	Therapeutic
<b>Arrhythmia syndromes</b>			
Long QT syndrome	+++	+++	+++
CPVT	+++	+	+
Brugada syndrome	+	+	+
Progressive cardiac conduction disease	+	+	+
Short QT syndrome	+	+	+
Sinus node disease	-	+	-
Atrial fibrillation	-	+	-
Early repolarization syndrome	-	-	-
<b>Cardiomyopathies</b>			
Hypertrophic cardiomyopathy	+++	++	++
Dilated cardiomyopathy	++	+++	++
Arrhythmogenic cardiomyopathy	+++	++	++
Left ventricular non-compaction	+	+	-
Restrictive cardiomyopathy	+	+	+
<b>Congenital heart disease</b>			
Syndromic CHD	+++	+	-
Non-syndromic CHD	+	-	-
Familial CHD	++	-	-

+++ : is recommended/is indicated or useful.  
 ++ : can be recommended/can be useful.  
 + : may be considered/may be useful.  
 - : is not recommended/is not indicated nor useful.

Riprodotta da Wilde et al<sup>8</sup>.

+++ : raccomandato/indicato o utile

++ : può essere raccomandato/utile

+ : potrebbe essere considerato/utile

- : non raccomandato/indicato o non utile

Il documento consensuale dell'American College di Genetica e Genomica Medica e dell'associazione di Patologia Molecolare sottolinea invece come in caso di riscontro di una variante di significato incerto, questa non debba essere utilizzata nel processo clinico decisionale<sup>9</sup>. A causa dell'assenza di solide evidenze scientifiche a sostegno della loro associazione causale con la malattia, l'identificazione di varianti di significato incerto non deve quindi essere utilizzata né per la conferma diagnostica né per lo screening cardiologico presintomatico di routine<sup>9, 10</sup>. Un'eventuale analisi di segregazione familiare può tuttavia essere proposta in caso di fenotipo patologico a carico di molteplici membri di una famiglia, al fine di dimostrare la co-segregazione della variante con il fenotipo di malattia<sup>6</sup>. Tale evidenza potrebbe infatti contribuire alla riclassificazione della variante di significato incerto in variante patogena o probabilmente patogena. In linea generale, tuttavia, vista l'assenza di ripercussione clinica, non esistono specifiche raccomandazioni che regolino la necessità e l'eventuale modalità di comunicazione ai pazienti in caso di identificazione di una variante di significato incerto. Un'importante autonomia decisionale è quindi lasciata ai diversi laboratori di analisi. Mentre alcuni di essi non comunicano tale risultato, altri informano i pazienti circa l'identificazione di una variante di significato incerto, lasciando talora il paziente in uno stato di confusione e timore sempre più chiaramente rilevato da recenti studi volti ad analizzare le implicazioni etiche e psicosociali associate al riscontro di varianti di significato incerto<sup>11</sup>. A causa quindi del persistente divario tra le competenze di lettura e sequenziamento genetico rispetto alla capacità di interpretazione dei risultati, futuri e ulteriori sforzi sono necessari al fine di favorire la riclassificazione delle varianti di significato incerto in varianti patologiche o benigne, con conseguente integrazione nel processo clinico decisionale. Allo stesso modo, secondo gli standard attuali per l'interpretazione delle varianti di sequenza<sup>8</sup>, i laboratori dovrebbero essere incoraggiati a rivalutare periodicamente le varianti di significato incerto attraverso un'indagine proattiva sulle nuove evidenze scientifiche disponibili.

## 2.8 Modelli predittivi computazionali

Come precedentemente descritto, l'analisi di una variante di sequenza attraverso l'utilizzo di programmi predittivi rappresenta un criterio fondamentale del processo interpretativo. Diversi algoritmi sono attualmente disponibili per predire l'effetto atteso della variante sia a livello nucleotidico che aminoacidico e il conseguente effetto strutturale e funzionale sul prodotto di trascrizione e sulla proteina finale<sup>8</sup>. Tali modelli valutano l'effetto di una variante missenso sulla base di uno o più criteri specifici quali il grado di conservazione evolutiva a livello nucleotidico e/o aminoacidico, il tipo di conseguenza biochimica indotta a livello aminoacidico dalla variante (in termini di dimensione, carica e idrofobia per esempio) o la localizzazione del cambio aminoacidico indotto nella proteina<sup>8</sup>. Dalla valutazione di uno o più di questi criteri deriva la capacità predittiva dell'algoritmo con un'accuratezza che si attesta attorno al 65-80% quando si valutano varianti note associate a malattia con i principali modelli attualmente disponibili (SIFT, PolzPhen2 etc...)<sup>8, 12</sup>.

Nel tentativo di elaborare uno strumento predittivo più performante, con maggiore potere discriminatorio nei confronti delle varianti di significato incerto, l'équipe del Prof. A. Superti-Furga e C. Rivolta ha recentemente elaborato un nuovo modello predittivo detto MutScore<sup>13</sup>. Attraverso l'analisi sistematica basata su apprendimento automatico (machine learning) della distribuzione intragene delle varianti missenso associate a patologie ereditarie riportate su ClinVar, gli autori hanno infatti dimostrato come le varianti missenso patogene tendano a localizzarsi in clusters ovvero in domini o regioni specifiche della proteina codificata spesso corrispondenti a noti domini funzionali. Sulla base di tale osservazione, Quinodoz et al. hanno quindi ipotizzato di poter predire la patogenicità di una variante sulla base della localizzazione topografica (in virtù del sopracitato fenomeno di clustering) costruendo un nuovo "positional score" che, integrato a score già esistenti, si caratterizzerebbe per un migliore potere discriminatorio<sup>13</sup>.

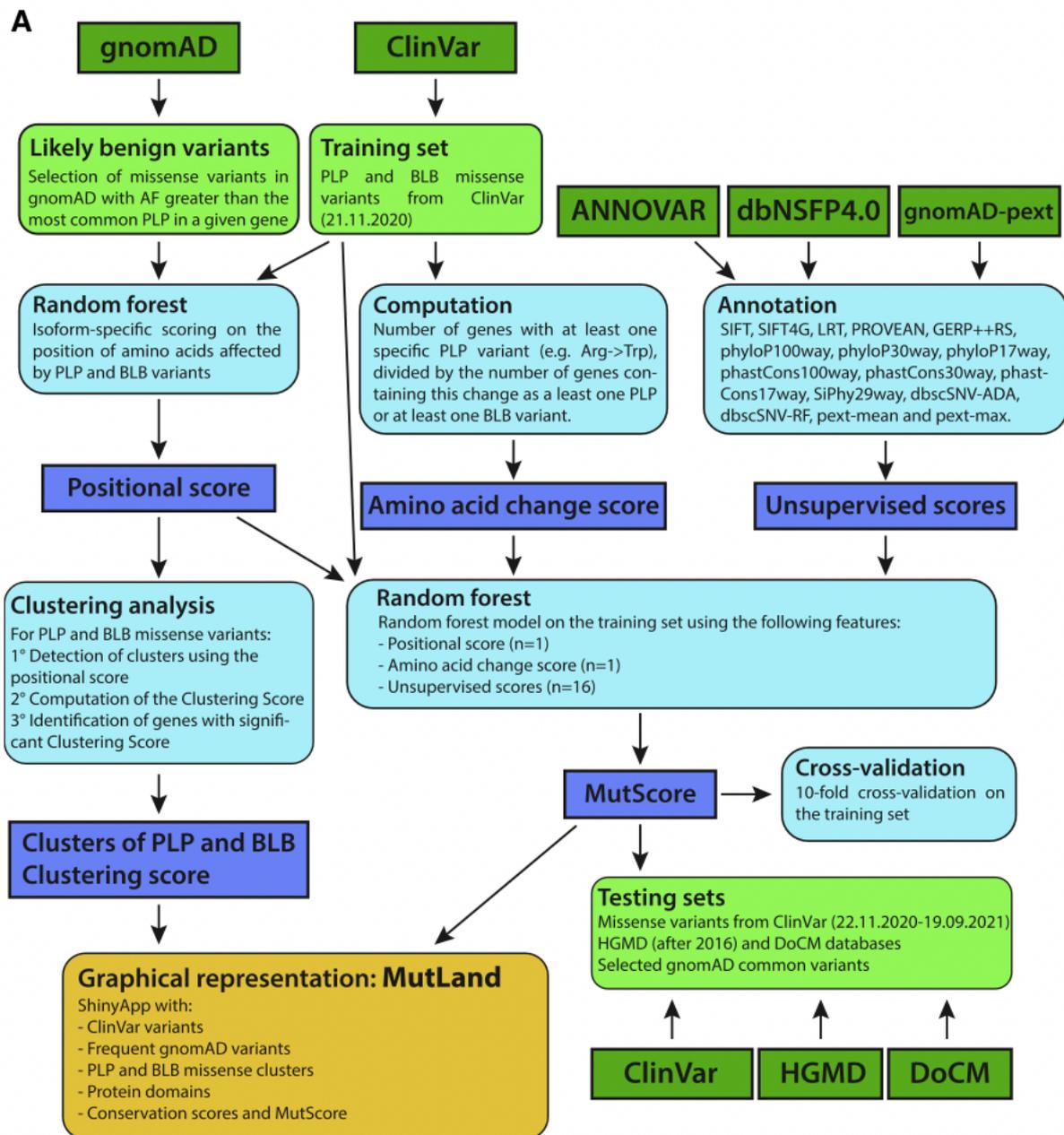
## 2.9 Mutscore e MutLand

Il Mutscore è un modello predittivo che si basa su un nuovo algoritmo incentrato sull'integrazione, a score già disponibili, di due nuovi score calcolati mediante un approccio di apprendimento automatico ("machine learning"):

- "Positional score" o score di posizione
- Score aminoacidico

Come rappresentato dalla **Figura 6**, a seguito dell'estrazione dal database di ClinVar di tutte le varianti missenso repertorate come patogene o probabilmente patogene (PLP) e delle varianti benigne o probabilmente benigne (BLB, queste ultime ulteriormente arricchite dalle varianti missenso repertorate in GnomAD come aventi una frequenza allelica maggiore rispetto alla frequenza allelica della variante PLP più comune in un dato gene), l'algoritmo prevede l'ottenimento di un "positional score" attraverso l'applicazione di un modello Random Forest per ogni trascritto individuale con almeno 10 PLP e 1 BLB. Per ogni variante, l'algoritmo prevede quindi la selezione del "positional score" appartenente all'isoforma con il più alto numero di PLP identificate. Segue l'analisi dei clusters ottenuta tramite l'identificazione delle regioni con "positional score" >0.05 e contenenti almeno 5 PLP. Un "clustering score" è analogamente calcolato e definito dal valore minimo tra la porzione di varianti missenso PLP localizzate nei clusters identificati e la porzione di trascritto non coperta da clusters, con valori compresi tra 0 e 1 (i.e. più il valore è prossimo a 1 maggiore è l'entità del clustering). Successivamente, il modello prevede il calcolo, per ogni variante missenso, di uno score aminoacidico ottenuto secondo la formula  $a / (a+b)$  dove a rappresenta il numero di geni contenenti la variante come PLP e b rappresenta il numero di geni contenenti la variante come BLB. Una volta calcolati gli score sopramenzionati, l'algoritmo comporta l'applicazione di un approccio Random Forest integrante il "positional score" e lo score aminoacidico con 16 altri score ottenuti dai più comuni software predittivi computazionali inclusi nella versione 4.0 di dbNSFP, database di predizione funzionale comunemente utilizzato per l'interpretazione delle varianti.

**Figura 6. Rappresentazione schematica delle procedure necessarie alla costruzione del Mutscore e del MutLand.**



Riprodotta da M. Quinodoz et al<sup>13</sup>. Lo schema rappresenta i differenti step necessari per generare il Mutscore e il MutLand (verde scuro: database; verde chiaro: set di varianti; blu chiaro: analisi computazionale; blu scuro: score; arancione: output grafico)

Più nel dettaglio, i 16 algoritmi predittivi usati includono SIFT, SIFT4G, LRT, PROVEAN, GERP++RS, phyloP100way, phyloP30way, phyloP17way, phast-Cons100way, phastCons30way, phastCons17way, Syphy29way, dbSCSNV-ADA and dbSCSNV-RF. Per ogni variante, il valore massimo (pext-max) e medio (pext) degli score di proporzione di espressione tra i trascritti (pext) ottenuti da GnomAD sono ugualmente inclusi e integrati nell'analisi. Per ogni variante, il Mutscore è quindi definito dal rapporto tra i voti a sostegno della patogenicità e il numero totale di voti calcolati dal modello Random Forest. Infine, dall'integrazione dei dati ottenuti da ClinVar, GnomAD, dagli score di dbNFSP4.0, dai clusters di varianti PLP e BLB e dal Mutscore, il modello elabora il cosiddetto MutLand, una rappresentazione grafica della topografia delle varianti per ogni gene con almeno una variante PLP repertoriata su ClinVar. La **Figura 7** rappresenta un esempio di output grafico MutLand relativo al gene *KCNH2*. Come facilmente comprensibile, le varianti PLP missenso tendono a presentarsi in clusters in zone specifiche principalmente in prossimità della porzione citoplasmatica N-terminale e in corrispondenza del segmento transmembranario. Le varianti PLP di tipo loss-of-function sono sparse lungo l'intera sequenza proteica. Analogamente, le varianti di significato incerto non si presentano raggruppate in cluster ma sono distribuite diffusamente sulla proteina. Tale modello fornisce quindi un argomento significativo in favore della possibile riclassificazione in “probabilmente benigne” per quelle varianti di significato incerto interessanti, per esempio, gli aminoacidi in prossimità dell'estremità C-terminale, localizzate al di fuori delle regioni cluster di PLP e con basso Mutscore.

Al fine di testare e validare l'efficacia predittiva del modello, Quinodoz et al. hanno quindi applicato l'algoritmo a tre set di dati rappresentati rispettivamente dall'insieme di varianti introdotte su ClinVar tra il 22.11.2020 e il 19.09.2021 e non utilizzate per l'elaborazione del “positional score” secondo quanto precedentemente descritto (set 1), dalle varianti patologiche missenso incluse nel “Human Gene Mutation Database” (set 2) a partire dal 2017 e nel database DoCM delle varianti cancerogene somatiche (set 3).

**Figura 7. Esempio di output grafico MutLand**

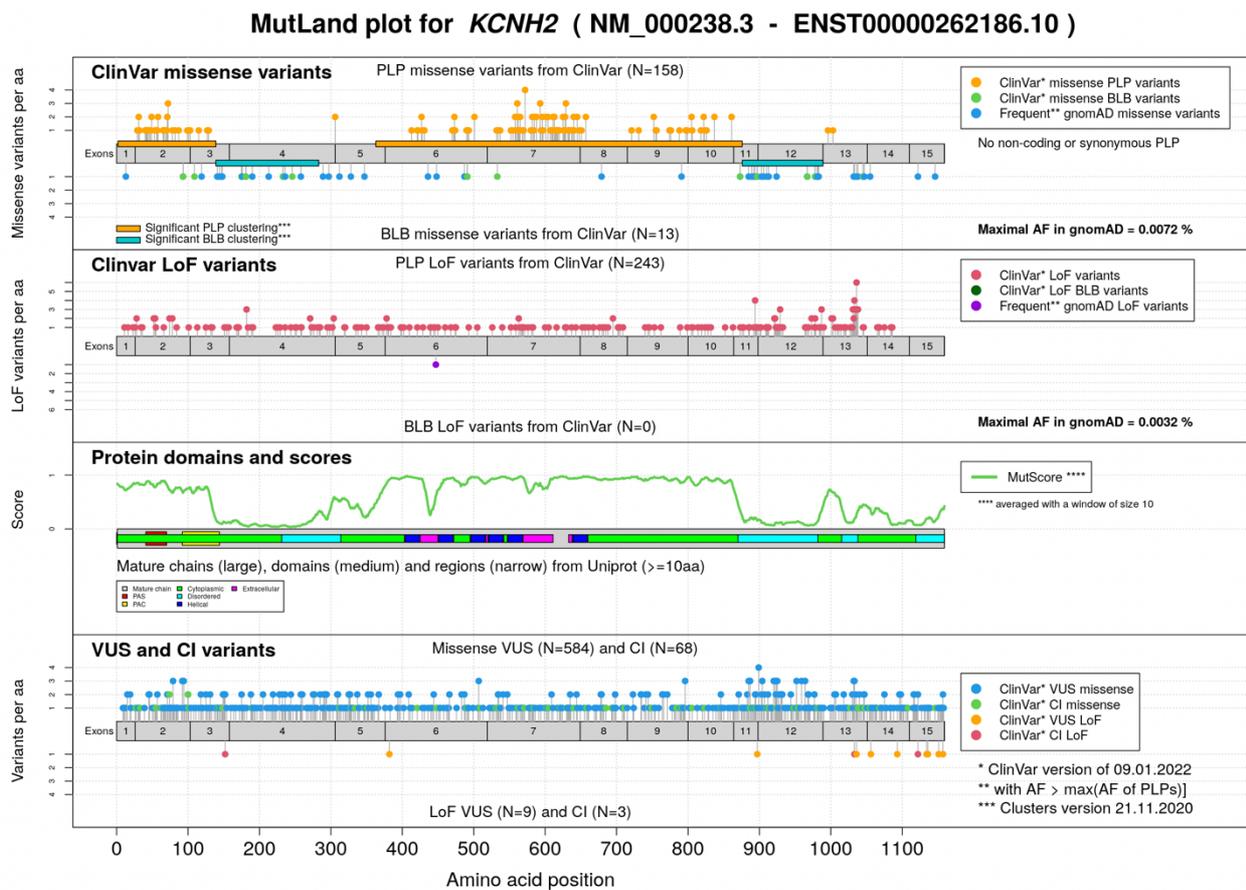


Grafico fornito per il gene *KCNH2*. Partendo dall'alto è possibile visualizzare la localizzazione topografica delle varianti missenso PLP, BLB e delle varianti loss-of-function (LoF) repertorate su ClinVar/GnomAD; l'andamento del Mutscore in relazione alla sede topografica delle varianti e la distribuzione topografica delle varianti di significato incerto (VUS) e di quelle con interpretazioni conflittuali (CI).

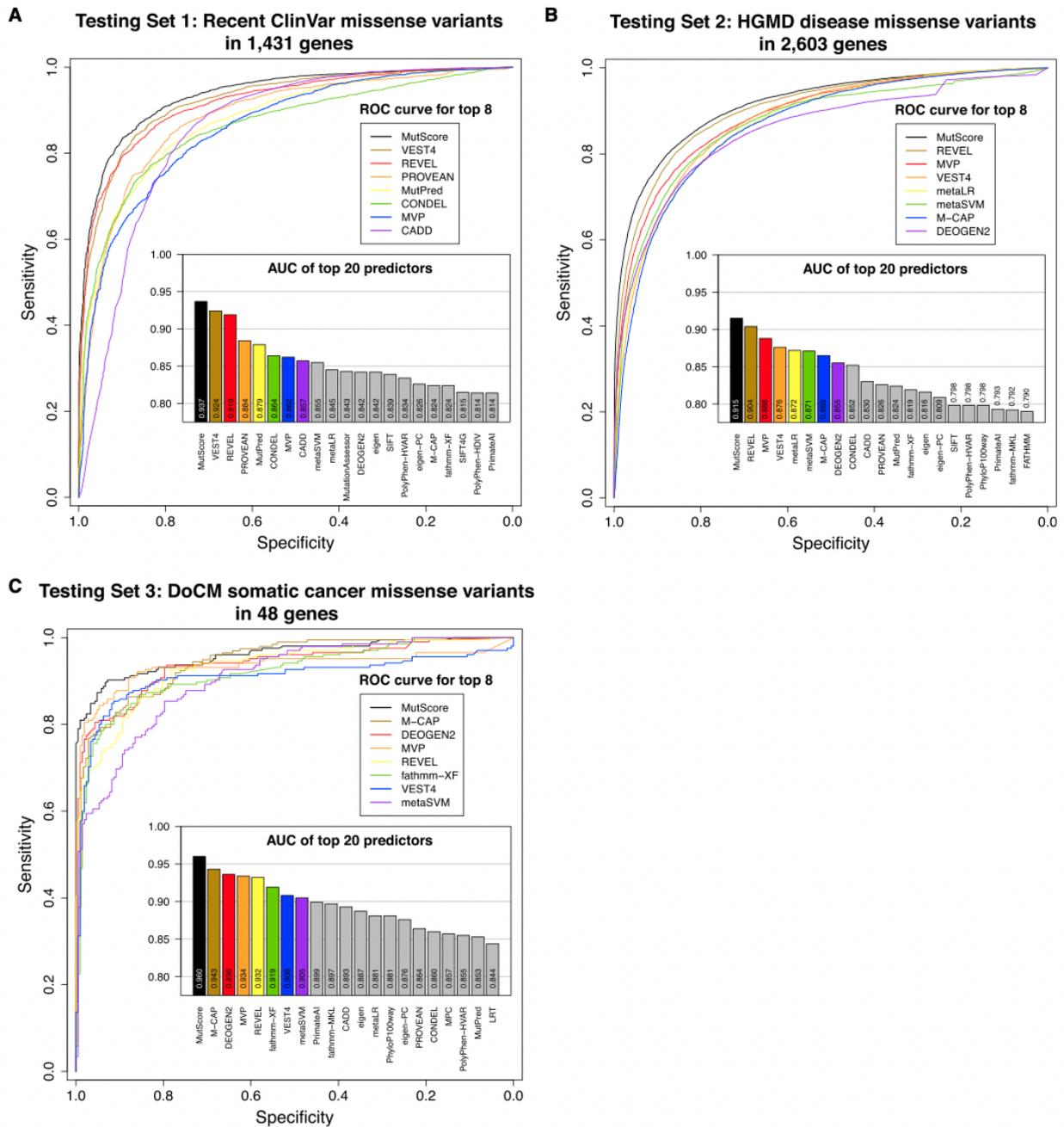
In tutti e tre i set il modello Mutscore si caratterizzava per la migliore performance predittiva con un'area sotto la curva (AUC) significativamente maggiore rispetto a tutti gli altri score predittivi (0.937, 0.915 e 0.960 rispettivamente per i set 1, 2 e 3). **Figura 8.**

In secondo luogo, gli autori hanno quindi valutato il grado di correlazione del Mutscore con la reale patogenicità della variante valutata sulla base delle evidenze scientifiche. Considerando per esempio lo score di Clinvar (da 1 a 4 stelle a seconda dell'entità di dati ed evidenze scientifiche disponibili a sostegno della patogenicità della variante), si evidenzia come l'AUC del Mutscore corredi bene con

il numero di stelle attribuite da Clinvar (AUC di 0.908, 0.946 e 0.963 per le varianti con rispettivamente 1, 2 e 3 stelle).

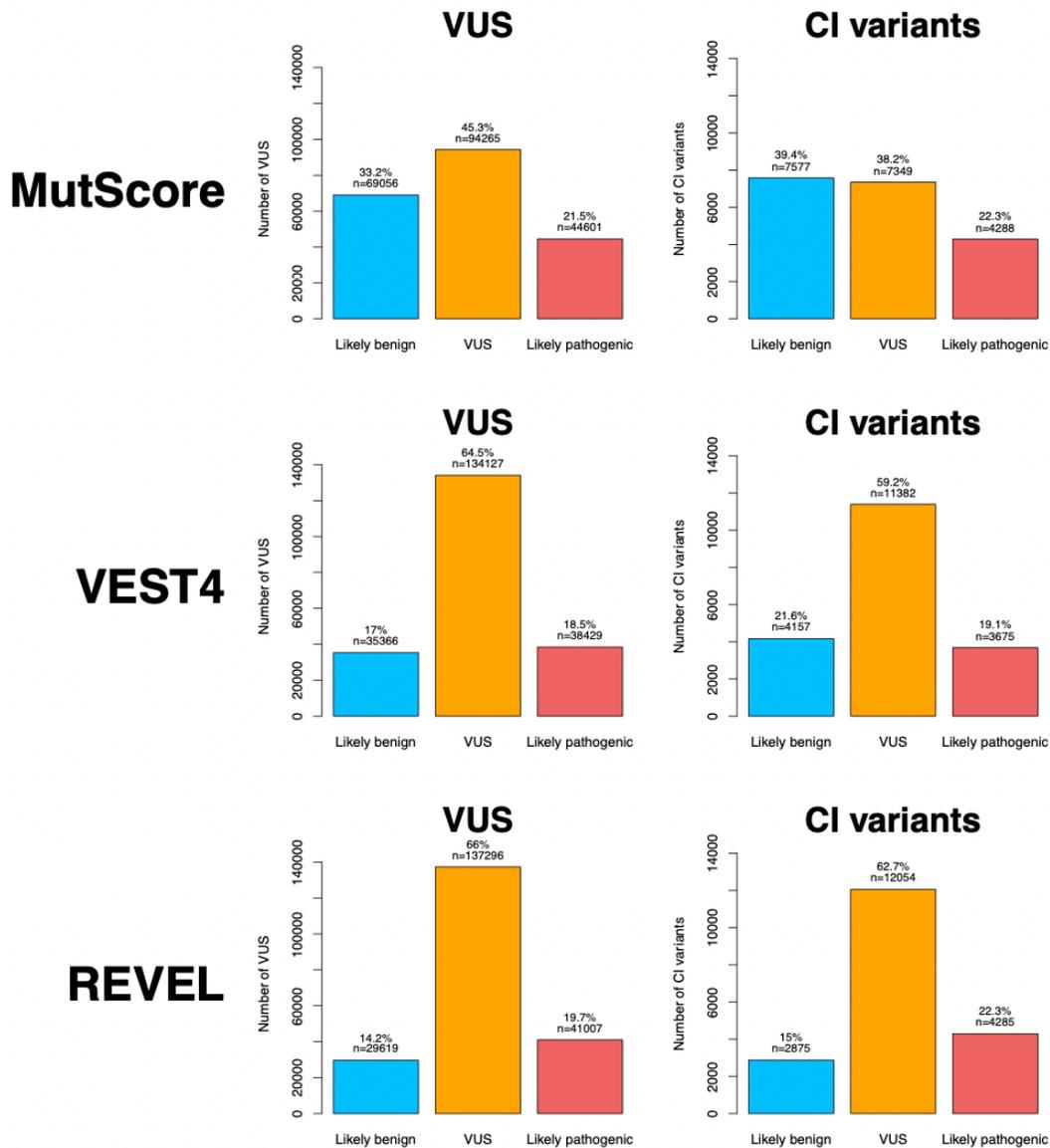
In terzo luogo, gli autori hanno infine determinato la capacità del Mutscore, in relazione agli altri due modelli predittivi più performanti emersi nei test precedenti (VEST4 e REVEAL), di riclassificare e discriminare le varianti di significato incerto. In particolare, gli autori definiscono la riclassificazione delle varianti sulla base di due valori soglia calcolati per ciascuno score. Una variante può essere riclassificata come probabilmente patogena se il suo score predittivo è maggiore dello score oltre il quale il 95% delle varianti del set di prova risulta costituito da PLP. Viceversa, una variante può essere riclassificata come probabilmente benigna se il suo score predittivo è minore dello score sotto il quale il 95% delle varianti del set di prova risulta costituito da BLB. Sulla base di tale premessa, gli autori dimostrano pertanto come il Mutscore permetta una riclassificazione del 54,7% delle varianti di significato incerto e del 61,7% delle varianti con interpretazioni contrastanti, rispetto al 35,5% e 33,9% delle varianti di significato incerto e al 40,7% e 37,3% delle varianti con interpretazioni contrastanti possibile rispettivamente grazie all'uso dei modelli predittivi VEST4 e REVEAL. **Figura 9.** In particolare, come evidenziato dalla Figura 9, rispetto al VEST4 e al REVEAL, il Mutscore sembra consentire una maggiore riclassificazione delle varianti in BLBs. Grazie al peculiare utilizzo del “positional score”, è infatti possibile integrare nella valutazione la localizzazione topografica della variante all'interno o meno delle regioni di clusters. Si potrebbe pertanto desumere che numerose varianti possano essere riclassificate in BLB proprio in virtù della loro localizzazione esterna ai comprovati clusters di patogenicità.

**Figura 8. Efficacia predittiva del Mutscore in relazione agli altri algoritmi predittivi nei tre set di varianti testate**



Riprodotta da Quindoz et al<sup>13</sup>. Pannello A. Set di test 1: varianti repertorate su ClinVar tra il 22.11.2020 e il 19.09.2021 (PLPs vs BLBs). Pannello B. Set di test 2: varianti patogene missenso (PLPs) incluse nel “Human Gene Mutation Database” dal 2017 vs varianti frequenti repertorate in GnomAD (BLBs). Pannello C. Set di test 3: varianti incluse nel database “DoCM” (PLPs) vs varianti frequenti repertorate in GnomAD (BLBs). Curve ROC per gli 8 algoritmi predittivi principali e istogrammi delle AUC per i principali 20 algoritmi predittivi.

**Figura 9. Riclassificazione delle varianti di significato incerto (VUS) e delle varianti con interpretazione conflittuale (CI) in varianti probabilmente benigne o probabilmente patogene in seguito all'uso di MutScore, VEST4 e REVEL.**



Riprodotta da Quindoz et al<sup>13</sup>. I valori rappresentati si riferiscono ai numeri assoluti e alle percentuali di varianti considerate (input: 207.922 VUS e 19.214 CI, corrispondenti entrambe al 100%). Le soglie utilizzate per riclassificare le varianti in BLBs e PLPs corrispondono specificamente a 0,140 e 0,730 per MutScore, 0,187 e 0.819 per VEST4 e 0.086 e 0.682 per REVEL.

### 3. OBIETTIVI DELLO STUDIO

Le varianti di significato incerto rappresentano attualmente la principale sfida nell'interpretazione dei test genetici correntemente realizzati nella pratica clinica. Negli ultimi decenni, la rapida espansione delle tecnologie di sequenziamento genetico di nuova generazione ("Next generation sequencing", NGS) ha infatti modificato profondamente il panorama del sequenziamento del genoma umano. Se da un lato tali tecniche hanno reso l'analisi dell'intero genoma, degli esoni o di geni specifici inclusi in pannelli multigenici facilmente accessibile nella pratica di routine, dall'altro lato hanno sollevato nuove difficoltà nell'interpretazione dei risultati. Tra queste, a causa dell'aumento massivo del numero di test genetici realizzati, è diventato sempre più frequente il riscontro di nuove varianti di sequenza che, differendo dalla sequenza standard della popolazione generale di riferimento, rimangono di significato sconosciuto/incerto. A causa della natura dubbia del relativo effetto strutturale e funzionale, le varianti di significato incerto non sono quindi integrate nell'iter decisionale diagnostico e la loro ricerca, salvo in casi particolari, non viene normalmente proposta tra i familiari asintomatici del caso indice. Molteplici casi di letteratura riportano infatti le ripercussioni deleterie, in termini diagnostico-terapeutici ed economici, derivanti dall'integrazione nel management clinico di varianti genetiche con limitate evidenze in supporto della relativa patogenicità<sup>14</sup>. D'altro canto, come emerso precedentemente grazie ai recenti sforzi di riclassificazione, molte varianti di significato incerto finiscono spesso per essere classate come varianti benigne o probabilmente benigne invece che essere repertorate come patogene o probabilmente patogene<sup>15</sup>.

Sulla base di quanto detto, il presente studio si propone di promuovere la riclassificazione delle varianti di significato incerto in varianti patogene/benigne nella nostra coorte di pazienti con malattie cardiache ereditarie accertate o sospette, attraverso l'applicazione del nuovo algoritmo predittivo Mutscore precedentemente descritto. L'impatto clinico del nostro studio, in conformità con il documento consensuale dell'ACMG e dell'associazione di Patologia Molecolare del 2015<sup>9</sup>, risulterebbe quindi maggiore per i pazienti. Le varianti di significato incerto riclassificate potrebbero

essere infatti integrate nel management clinico decisionale, contribuendo alla conferma o all'esclusione della diagnosi, alla stratificazione del rischio cardiovascolare ed eventualmente allo screening genetico familiare presintomatico. Il nostro studio rappresenta pertanto un ulteriore sforzo sul cammino della medicina di precisione. In secondo luogo, il presente studio permetterà di valutare, su un ulteriore set di dati reali, le performance predittive di questo nuovo algoritmo predittivo (Mutscore) basato su un approccio di apprendimento automatico (machine learning).

### **3.1 Obiettivo primario**

L'obiettivo primario dello studio consiste nel valutare la prevalenza delle varianti probabilmente patogene/patogene e di significato incerto nella coorte di pazienti con malattia cardiaca ereditaria accertata o sospetta, indirizzati dal 1° Gennaio 2007 al 31 Dicembre 2021, alla consultazione congiunta dei servizi di Cardiologia e di Genetica Medica dell'Ospedale Universitario di Losanna (CHUV).

### **3.2 Obiettivo secondario**

A seguito della relativa identificazione nella coorte di studio, ogni variante probabilmente patogena/patogena e di significato incerto sarà analizzata attraverso l'ausilio del nuovo modello predittivo computazione Mutscore. Questo approccio retrospettivo ha lo scopo di:

- fornire un nuovo set di dati reali per testare la performance predittiva del Mutscore
- verificare l'esistenza e l'eventuale grado di correlazione tra il Mutscore e la reale patogenicità delle varianti repertorate come probabilmente patogene/patogene mediante l'analisi integrata dei criteri definiti dal documento consensuale dell'ACMG e dell'associazione di Patologia Molecolare del 2015<sup>9</sup>.
- analizzare la capacità di riclassificazione delle varianti di significato incerto sulla base del Mutscore.

## **4. MATERIALI E METODI**

### **4.1 Disegno dello studio**

Il presente lavoro è uno studio retrospettivo monocentrico basato sui dati del registro prospettivo delle Malattie Cardiache Ereditarie del Servizio di Genetica dell'Ospedale Universitario di Losanna (Centre Hospitalier Universitaire Vaudois, CHUV). L'analisi sarà condotta sui pazienti inclusi a partire dal 1° Gennaio 2007 fino al 31 Dicembre 2021 e si articolerà in tre fasi:

1. Identificazione nella coorte di studio delle varianti repertorate come probabilmente patogene/patogene e come varianti di significato incerto;
2. Analisi delle varianti probabilmente patogene/patogene identificate, mediante l'ausilio del Mutscore e analisi della performance dell'algoritmo nella predizione di patogenicità ritenuta sulla base dei criteri integrati definiti dal documento consensuale dell'ACMG e dell'associazione di Patologia Molecolare<sup>9</sup>;
3. Analisi delle varianti di significato incerto identificate, mediante l'ausilio del Mutscore e analisi del grado di riclassificazione resa possibile sulla base del Mutscore.

### **4.2 Popolazione dello studio**

Il registro prospettivo delle Malattie Cardiache Ereditarie del Servizio di Genetica dell'Ospedale Universitario di Losanna include una coorte di pazienti con diagnosi clinica o sospetta di malattia cardiaca ereditaria in senso largo e/o i pazienti nei quali l'analisi genetica ha messo in evidenza una variante genetica probabilmente patogena/patogena associata ad un rischio di sviluppare una malattia cardiaca ereditaria nel contesto di uno screening presintomatico.

#### **4.2.1 Criteri di inclusione**

- Pazienti con diagnosi clinica o sospetta di malattia cardiaca ereditaria e/o portatori di variante genetica probabilmente patogena/patogena scoperta nel contesto di uno screening presintomatico e associata ad un rischio di sviluppare una malattia cardiaca ereditaria, tra cui:
  - a. Sindrome del QT lungo congenito
  - b. Sindrome del QT corto
  - c. Sindrome di Brugada
  - d. Difetto progressivo familiare della conduzione cardiaca
  - d. Tachicardia ventricolare polimorfa catecolaminergica
  - e. Fibrillazione ventricolare idiopatica
  - f. Sindrome del prolasso mitrale maligno familiare
  - f. Cardiomiopatia ipertrofica
  - g. Cardiomiopatia dilatativa
  - h. Cardiomiopatia restrittiva
  - i. Cardiomiopatia aritmogena
  - l. Cardiomiopatia ventricolare sinistra di tipo non-compaction
- Inclusione nel registro delle Malattie Cardiache Ereditarie del Servizio di Genetica dell'Ospedale Universitario di Losanna (CHUV)
- Consenso informato scritto ottenuto

#### **4.2.2. Criteri di esclusione**

- Indisponibilità del campione genetico
- Consenso informato rifiutato

### **4.3 Test genetici e interpretazione delle varianti**

L'analisi genetica prevede, a seconda del sospetto clinico, l'utilizzo di pannelli di geni basati sull'utilizzo delle tecniche di sequenziamento genetico di nuova generazione ("Next generation sequencing", NGS). Le varianti sono state quindi interpretate in accordo con le raccomandazioni del documento consensuale dell'ACMG e dell'associazione di Patologia Molecolare<sup>9</sup> e classificate in: patogene o probabilmente patogene, di significato incerto, benigne o probabilmente benigne.

### **4.4 Raccolta dei dati**

I dati medici dei pazienti sono stati ottenuti dai dossiers informatizzati repertoriati sul server istituzionale o dalle cartelle cartacee dei pazienti. Le informazioni di interesse sono state successivamente raccolte in un database istituzionale Excel e includono dati clinici (età, sesso, presentazione clinica, anamnesi familiare, terapia) e genetici (pannello genetico realizzato, risultato analisi genetica, nomenclatura dell'eventuale variante trovata).

#### **4.4.1 Calcolo dello score predittivo (Mutscore)**

Lo score predittivo Mutscore è stato successivamente calcolato per ogni variante probabilmente patogena/patogena e di significato incerto repertoriata nel database, mediante l'utilizzo del programma gratuito accessibile sul web all'indirizzo <https://mutscore-wg7hvakhq-ew.a.run.app>. L'interfaccia, specificamente creata dal gruppo di Quinodoz et al<sup>13</sup>, permette la visualizzazione di dati ottenuti da ClinVar, GnomAD e Uniprot, gli score di conservazione calcolati con altri algoritmi, l'esistenza e l'eventuale localizzazione di clusters di varianti probabilmente patogene/benigne e di significato incerto e l'output del Mutscore. Quest'ultimo in particolare è calcolato secondo l'algoritmo incentrato sull'integrazione, a score già disponibili, dei due nuovi score ("positional score" e score aminoacidico) calcolati mediante un approccio di apprendimento automatico ("machine learning") secondo quanto già precedentemente descritto al capitolo 1.7. In particolare, per quanto riguarda l'analisi delle varianti di significato incerto, sulla base dei risultati prodotti nello

studio di Quinodoz et al<sup>13</sup>., due valori soglia sono stati considerati per permettere la riclassificazione delle varianti in probabilmente patogene o probabilmente benigne sulla base del Mutscore. Una variante può essere riclassificata come probabilmente patogena se il suo Mutscore è maggiore dello score oltre il quale il 95% delle varianti del set di prova dello studio di Quinodoz et al<sup>13</sup>. risulta costituito da varianti probabilmente patogene/patogene. Viceversa, una variante può essere riclassificata come probabilmente benigna se il suo Mutscore è minore dello score sotto il quale il 95% delle varianti del set di prova risulta costituito da varianti probabilmente benigne/benigne.

#### **4.5 Analisi statistica**

Le variabili categoriche sono espresse in numero assoluto (N) e in termini percentuali (%) rapportati al numero assoluto totale di ogni variabile. Il test di Shapiro-Wilk è stato utilizzato per analizzare la distribuzione delle variabili continue: le variabili con distribuzione normale sono quindi espresse in termini di valore medio  $\pm$  deviazione standard mentre le variabili non normali sono descritte in termini di mediana e scarto interquantile [p25; p75].

Si è proceduto alla costruzione di una curva ROC (Receiver-operating characteristic) e all'analisi dell'area sotto la curva (AUC) al fine di determinare la performance del Mutscore nel discriminare tra le varianti patogene e le varianti di significato incerto nella nostra coorte di studio.

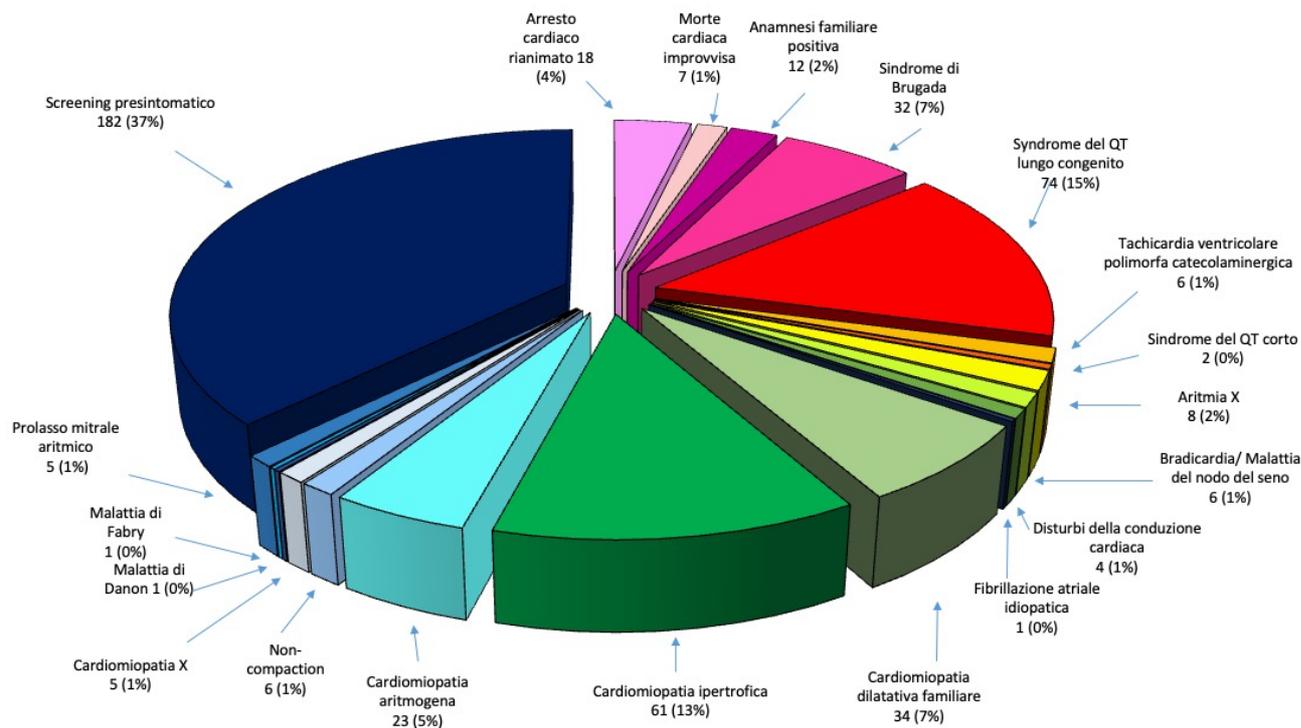
La relazione esistente tra i valori di Mutscore e il grado di patogenicità della variante repertoriato su ClinVar (1= benigna, 2= benigna, 3= significato incerto, 4= probabilmente patogena, 5= patogena) è stato analizzato prima mediante correlazione non parametrica di Spearman e successivamente mediante analisi di regressione lineare univariata. In particolare, per definire il grado di patogenicità, sono state considerate unicamente le varianti con evidenze univoche di patogenicità su ClinVar mentre sono state escluse le varianti con evidenze contrastanti. Un p-value  $<0,05$  è stato considerato statisticamente significativo. Tutte le analisi sono state eseguite con IBM SPSS Statistics 19.

## 5 RISULTATI

### 5.1 Popolazione di studio

La coorte di studio include 488 pazienti con diagnosi clinica o sospetta di malattia cardiaca ereditaria e/o portatori di variante genetica probabilmente patogena/patogena scoperta nel contesto di uno screening presintomatico e arruolati nel registro prospettivo delle Malattie Cardiache Ereditarie del Servizio di Genetica dell'Ospedale Universitario di Losanna (Centre Hospitalier Universitaire Vaudois, CHUV) tra il 1° Gennaio 2007 e il 31 Dicembre 2021. La popolazione include 251 (51,4%) maschi e 237 femmine (48,6%) con un'età media alla realizzazione del test genetico di  $37 \pm 20,3$  anni (range tra 0 e 88 anni). La **Figura 10** riassume le indicazioni cliniche alla realizzazione dell'analisi genetica nella popolazione in esame. Globalmente, 131 (26,8%) pazienti sono stati indirizzati per cardiomiopatia ereditaria diagnosticata o sospetta, 138 pazienti (28,3%) per sindrome aritmica ereditaria diagnosticata o sospetta, 182 (37,3%) per screening presintomatico in caso di variante familiare conosciuta e nei restanti 37 (7,6%) pazienti l'analisi genetica è stata eseguita a seguito di arresto cardiaco rianimato, morte improvvisa o per anamnesi familiare positiva. Più nel dettaglio, l'indicazione clinica più frequente dopo lo screening presintomatico, è rappresentata dalla sindrome del QT lungo congenito (15%) e dalla cardiomiopatia ipertrofica (13%).

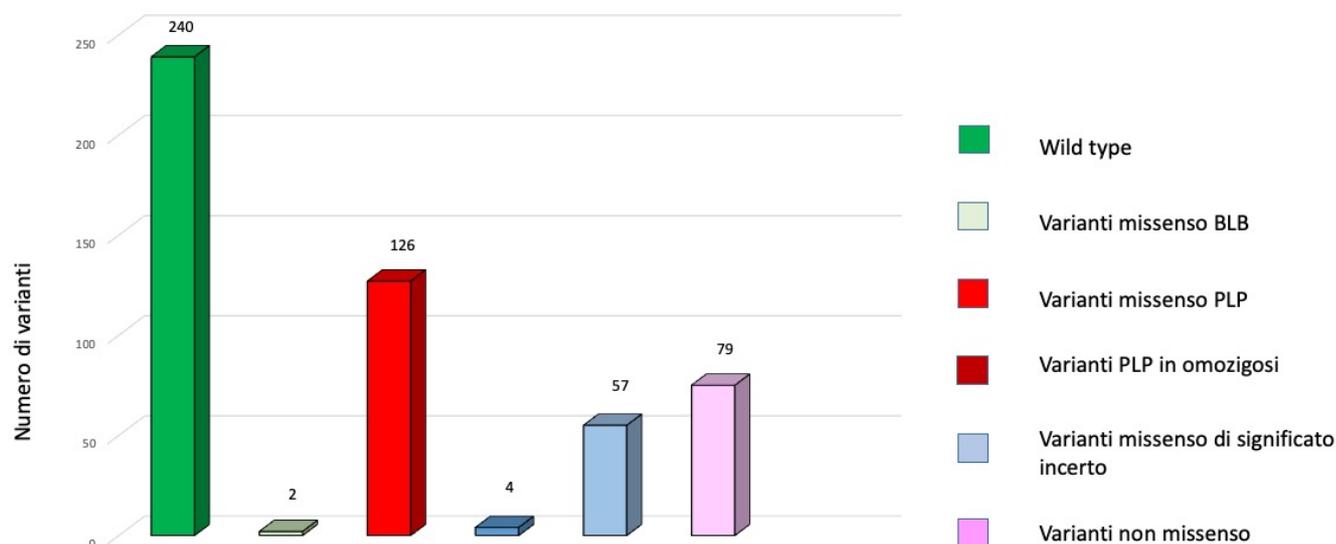
**Figura 10. Indicazioni cliniche al test genetico**



## 5.2 Varianti identificate nella popolazione di studio

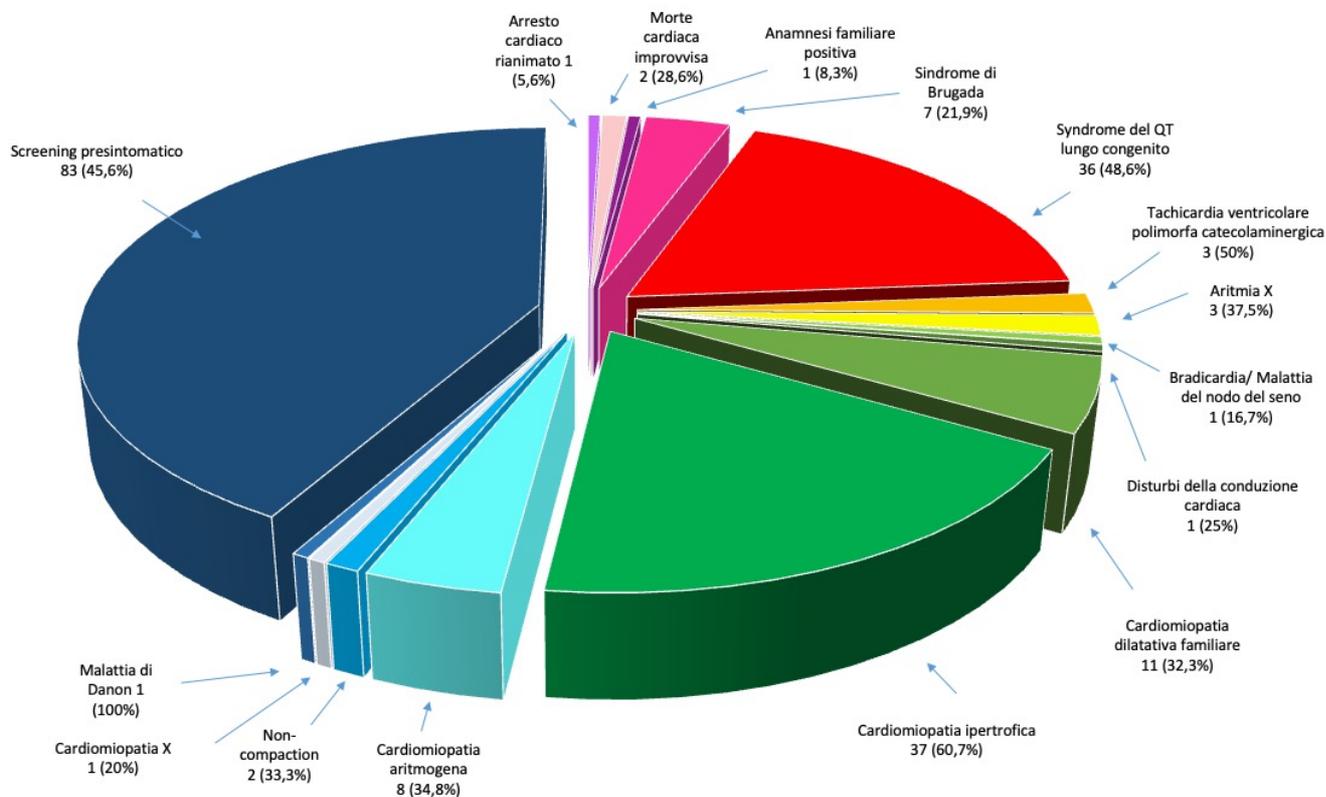
Nella popolazione in esame, non sono state identificate varianti in 240 (49%) pazienti. Nei restanti 249 pazienti sono state identificate 268 varianti rappresentate rispettivamente da: 2 (0,7%) varianti missenso probabilmente benigne, 57 (21,3%) varianti missenso di significato incerto, 126 (47%) varianti missenso patogene/probabilmente patogene, 4 (1,5%) varianti patogene/probabilmente patogene in omozigosi e 79 (29,5%) varianti non missenso. Queste ultime in particolare includono 72 varianti non missenso patogene/probabilmente patogene e 7 varianti non missenso di significato incerto. **Figura 11.**

**Figura 11. Varianti identificate nella popolazione di studio**



Su 488 pazienti testati, una variante patogena/probabilmente patogena, sia essa missenso o meno, è stata pertanto riscontrata in 198 pazienti per una resa diagnostica globale del 40,6%. Tra i 131 pazienti con indicazione di analisi rappresentata da una cardiomiopatia ereditaria generica, una variante patogena/probabilmente patogena è stata identificata in 60 (45,8%) pazienti mentre tra i 138 pazienti con indicazione di analisi rappresentata da una sindrome aritmica, una variante patogena/probabilmente patogena è stata identificata in 51 (37%) pazienti. In particolare, considerando la distribuzione delle varianti patogene/probabilmente patogene in relazione all'indicazione clinica di analisi, la maggiore resa diagnostica si è evidenziata tra i pazienti con diagnosi clinica o sospetta di cardiomiopatia ipertrofica (37 mutation carrier su 61 pazienti testati) e di malattia di Danon (1 mutation carrier su 1 paziente testato) per una resa diagnostica rispettivamente del 60,7% e del 100%. Seguono le rese diagnostiche osservate tra i pazienti con diagnosi clinica o sospetta di tachicardia polimorfa catecolaminergica e sindrome del QT lungo congenito, calcolate rispettivamente al 50% (3 mutation carriers su 6 pazienti testati) e al 48,6% (36 mutation carriers su 74 pazienti testati). **Figura 12.**

**Figura 12. Distribuzione delle varianti patogene/probabilmente patogene in relazione all'indicazione clinica di analisi**

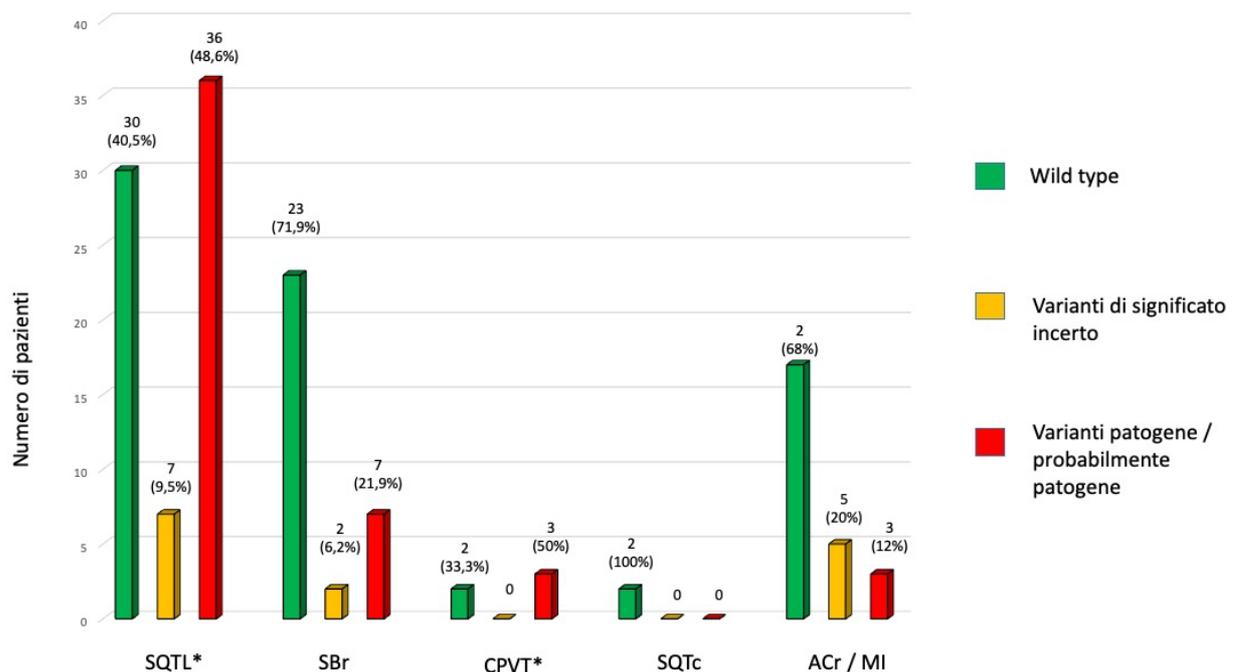


Più nel dettaglio, focalizzando l'analisi sulle principali sindromi aritmiche, si evidenzia che tra i pazienti (n= 74) con diagnosi o sospetto di Sindrome del QT lungo congenito, 30 (40,5%) presentano un genotipo wild-type, 7 (9,5%) presentano delle varianti di significato incerto mentre una variante patogene/probabilmente patogene è stata identificata in 36 (48,6%) pazienti (a cui si deve aggiungere il riscontro di una variante patogene allo stato omozigote in 1 paziente). In particolare, sono state identificate 10 varianti di significato incerto di cui 2 in pazienti già portatori di una variante patogene/probabilmente patogene mentre 1 è stata identificata in un paziente già portatore di un'altra variante di significato incerto.

Analogamente, tra i 32 pazienti con Sindrome di Brugada, 23 (71,9%) presentano un genotipo wild type, 2 (6,3%) esibiscono delle varianti di significato incerto e 7 (21,9%) una variante

patogena/probabilmente patogena. La **figura 13** riassume la distribuzione dei genotipi wild-type, delle varianti di significato incerto e delle varianti patogene/probabilmente patogene per le principali sindromi aritmiche.

**Figura 13. Distribuzione dei genotipi wild-type, delle varianti di significato incerto e delle varianti patogene/probabilmente patogene per le principali sindromi aritmiche.**

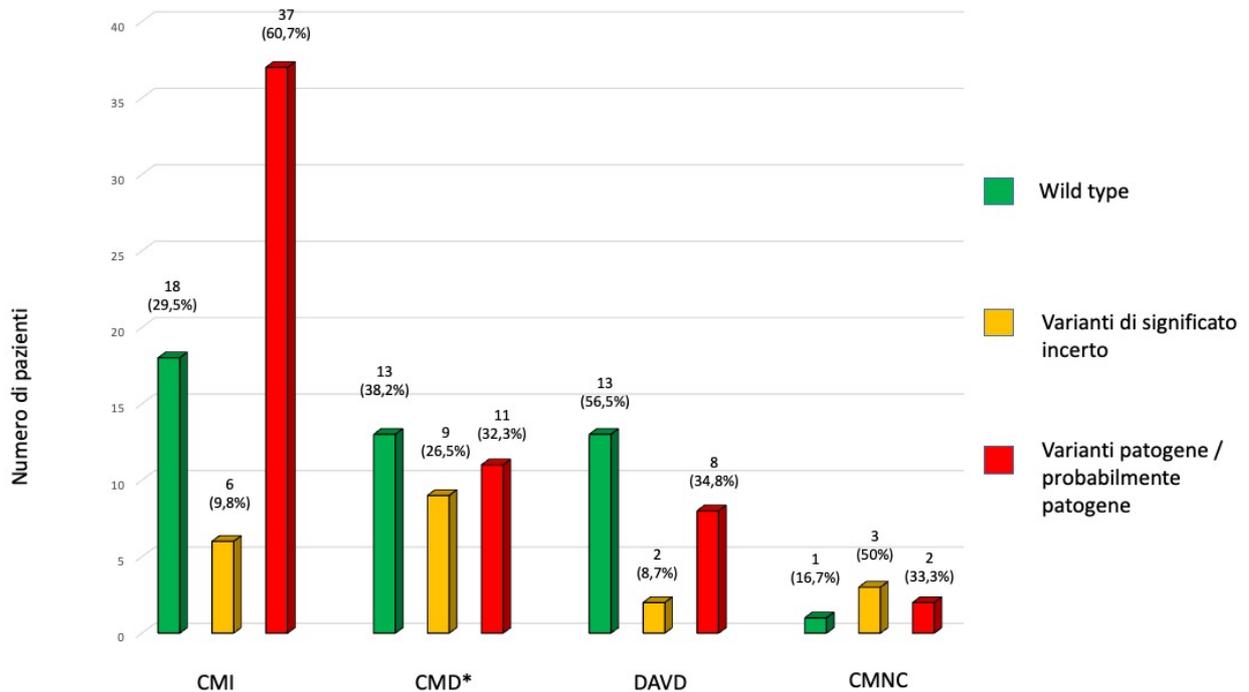


**Figura 13:** SQTl: Sindrome del QT lungo; SBr: Sindrome di Brugada; CPVT: Tachicardia ventricolare polimorfa catecolaminergica; SQTc: Sindrome del QT corto; ACr / MI: Arresto cardiaco rianimato / morte improvvisa. \* Tra i pazienti con SQTl e CPVT con PLP, si deve aggiungere il riscontro di una variante patogena allo stato omozigote in 1 paziente. Tra i pazienti con SQTl, 3 varianti di significato incerto sono state ulteriormente identificate di cui 2 in due pazienti già portatori di una variante patogena/probabilmente patogena e 1 in un paziente già portatore di un'altra variante di significato incerto (per un totale di 10 varianti di significato incerto identificate). Tra i pazienti con CPVT, 1 variante di significato incerto è stata identificata in 1 paziente già portatore di una variante patogena/probabilmente patogena (per un totale di 1 variante di significato incerto identificata). Tra i pazienti con ACr, 2 varianti di significato incerto sono state ulteriormente identificate in due pazienti già portatori di un'altra variante di significato incerto (per un totale di 7 varianti di significato incerto identificate).

All'analisi dei sottotipi di cardiomiopatie ereditarie, si evidenzia che tra i pazienti (n= 61) con diagnosi o sospetto di cardiomiopatia ipertrofica, 18 (29,5%) presentano un genotipo wild-type, 6 (9,8%) presentano delle varianti di significato incerto mentre una variante patogena/probabilmente patogena è stata identificata in 37 (60,7%) pazienti. Più nel dettaglio, 12 varianti di significato incerto sono state identificate di cui 4 in pazienti già portatori di una variante patogena/probabilmente patogena e 2 in pazienti già portatori di un'altra variante di significato incerto. Analogamente, tra i 34 pazienti con cardiomiopatia dilatativa familiare, 13 (38,2%) presentano un genotipo wild type, 9 (26,5%) esibiscono delle varianti di significato incerto mentre 11 (32,3%) una variante patogena/probabilmente patogena (a cui si deve aggiungere il riscontro di una variante patogena allo stato omozigote in 1 paziente). In particolare, si è identificato un totale di 12 varianti di significato incerto di cui 2 in due pazienti già portatori di una variante patogena/probabilmente patogena e 1 in un paziente già portatore di un'altra variante di significato incerto. La **figura 14** riassume la distribuzione dei genotipi wild-type, delle varianti di significato incerto e delle varianti patogene/probabilmente patogene per le principali cardiomiopatie ereditarie.

Analizzando infine i 182 pazienti con indicazione al test rappresentata dallo screening familiare presintomatico, 99 (54,4%) presentano un genotipo wild type e 83 (45,6%) presentano una variante patogena/probabilmente patogena, mentre 1 sola variante di significato incerto è stata identificata in un soggetto con una prima analisi risultata con genotipo wild-type.

**Figura 14. Distribuzione dei genotipi wild-type, delle varianti di significato incerto e delle varianti patogene/probabilmente patogene per le principali cardiomiopatie ereditarie.**



**Figura 14:** CMI: Cardiomiopatia ipertrofica; CMD: Cardiomiopatia dilatativa familiare; DAVD: Displasia aritmogena del ventricolo destro; CMNC: Cardiomiopatia ventricolare sinistra di tipo non-compaction; \* Tra i pazienti con CMD e PLP, si deve aggiungere il riscontro di una variante patogena allo stato omozigote in 1 paziente. Tra i pazienti con CMI, 6 varianti di significato incerto sono state identificate di cui 4 in pazienti già portatori di una variante patogena/probabilmente patogena e 2 in due pazienti già portatori di un'altra variante di significato incerto (per un totale di 12 varianti di significato incerto identificate). Tra i pazienti con CVDA, 1 variante di significato incerto è stata identificata in 1 paziente già portatore di una variante di significato incerto (per un totale di 3 varianti di significato incerto identificate). Tra i pazienti con CMNC, 2 varianti di significato incerto sono state identificate in due pazienti già portatori di una variante di significato incerto (per un totale di 5 varianti di significato incerto identificate).

### **5.3 Analisi delle varianti probabilmente patogene/patogene**

#### **5.3.1 SINDROME DEL QT LUNGO**

Tra i 74 pazienti testati, 36 (48,6%) presentano una variante patogena/probabilmente patogena. Come atteso, le varianti riscontrate si localizzano a carico dei 3 principali geni. In particolare, 15 (41,7%) pazienti presentano una variante sul gene *KCNQ1*, 15 (41,7%) sul gene *KCNH2* e 3 (8,3%) sul gene *SCN5A*. Per i restanti 3 pazienti, questi presentano rispettivamente 1 variante patogena/probabilmente patogena su *CACNA1C* e 2 su *CASQ2*. Tra le varianti patogene/probabilmente patogene riscontrate 26 (72,2%) sono missenso mentre 10 (27,8%) sono non missenso (rispettivamente 2 su *KCNQ1*, 6 su *KCNH2* e 2 su *CASQ2*).

#### **5.3.2 SINDROME DI BRUGADA**

Tra i 32 pazienti analizzati, 7 (21,9%) sono risultati portatori di una variante patogena/probabilmente patogena sul gene *SCN5A*. Tra queste ultime, 2 sono varianti missenso mentre 5 sono varianti non missenso.

#### **5.3.3 TACHICARDIA VENTRICOLARE POLIMORFA CATECOLAMINERGICA**

Tra i 6 pazienti analizzati, 3 (50%) soggetti presentano una variante patogena/probabilmente patogena di cui 2 missenso a carico rispettivamente del gene *RYR2* e *KCNJ2* e 1 non missenso a carico di *CASQ2*.

#### **5.3.4 ARRESTO CARDIACO RIANIMATO / MORTE IMPROVVISA**

Tra i 25 pazienti testati, 3 (12%) presentano una variante missenso patogena/probabilmente patogena identificata rispettivamente sui geni *KCNH2*, *RYR2* e *SCN5A*.

### **5.3.5 CARDIOMIOPATIA IPERTROFICA**

Tra i 61 pazienti analizzati, sono state identificate delle varianti patogene/probabilmente patogene in 37 (60,7%) individui, di cui 20 (54,1%) missenso e 17 (45,9%) non missenso. Le prime includono 3 (8,1%) sul gene *MYBPC3*, 13 (35,1%) sul gene *MYH7*, 1 (2,7%) sul gene *LMNA*, 1 (2,7%) sul gene *MYL3*, 1 (2,7%) su *PTPN11* e 1 (2,7%) sul gene *TNNT2*. Le varianti non missenso osservate includono 13 (35,1%) sul gene *MYBPC3*, 3 (8,1%) sul gene *MYH7* e 1 (2,7%) sul gene *CSRP3*.

### **5.3.5 CARDIOMIOPATIA DILATATIVA FAMILIARE**

Tra i 34 pazienti testati, 11 (32,3%) presentano una variante patogena/probabilmente patogena di cui missenso in 5 (45,5%) casi e non missenso nei restanti 6 (54,5%). Le prime comprendono 3 (27,3%) varianti patogene/probabilmente patogene sul gene *LMNA*, 1 (9,1%) su *TNNT2* e 1 (9,1%) sul gene *TTN*. Le varianti non missenso oggettivate includono 5 (45,4%) sul gene *TTN* e 1 (9,1%) sul gene *LMNA*.

### **5.3.6 CARDIOMIOPATIA ARITMOGENA**

Otto (34,8%) varianti patogene/probabilmente patogene sono state identificate tra i 23 pazienti analizzati per sospetta cardiomiopatia aritmogena. Queste includono 3 (37,5%) varianti missenso di cui 2 (25%) localizzate sul gene *PKP2* e 1 (12,5%) sul gene *ACTN2*. Le 5 (62,5%) varianti non missenso includono invece 2 (25%) localizzate sul gene *PKP2* e 1 rispettivamente sui geni *TTN*, *DSG2* (in omozigosi) e *DSP*.

### **5.3.7 CARDIOMIOPATIA VENTRICOLARE SINISTRA DI TIPO NON COMPACTION**

Tra i 6 pazienti testati, 2 (33,3%) presentano una variante patogena/probabilmente patogena di cui 1 (50%) missenso localizzata sul gene *MYH7* e 1 (50%) non missenso localizzata sul gene *TTN*.

## 5.4 Analisi delle varianti di significato incerto

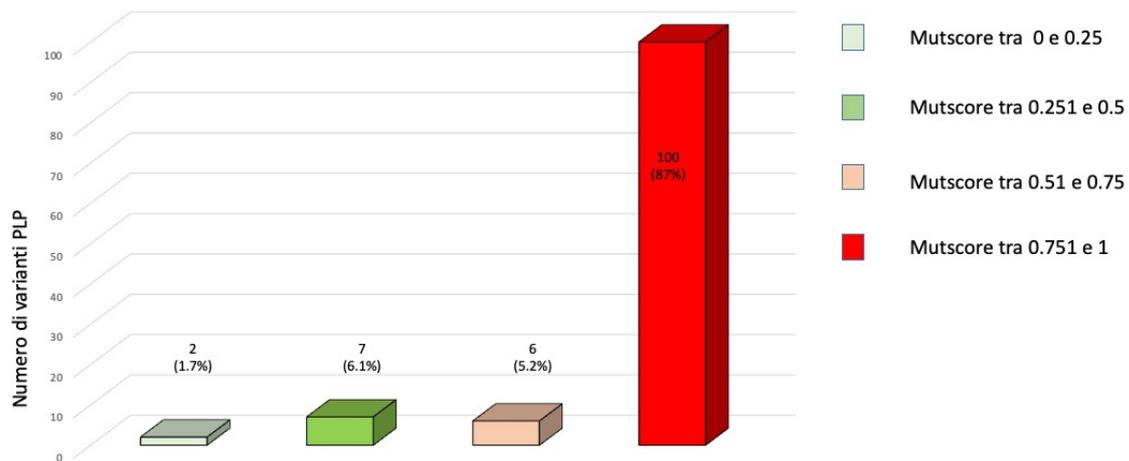
Nella coorte di studio sono state identificate un totale di 57 varianti missenso e 7 varianti non missenso di significato incerto in un totale di 44 (9%) pazienti. Più in dettaglio, infatti, si è riscontrata una singola variante missenso di significato incerto in 41 pazienti mentre 16 varianti missenso di significato incerto sono state identificate rispettivamente in 9 pazienti già portatori di una variante patogena/probabilmente patogena, in 6 pazienti già portatori di un'altra variante di significato incerto e in 1 paziente con pregresso genotipo wild-type. Per quanto riguarda le varianti non missenso di significato incerto, queste risultano isolate in 3 pazienti mentre le restanti 4 sono state identificate in soggetti già portatori di un'altra variante di significato incerto.

## 5.5 Calcolo dello score predittivo Mutscore

### 5.5.1 Varianti probabilmente patogene/patogene e analisi della performance dell'algoritmo

Si è quindi proceduto al calcolo del Mutscore secondo l'algoritmo precedentemente descritto per ognuna delle varianti missenso patogene/probabilmente patogene identificate. La **Figura 15** fornisce una rappresentazione grafica della distribuzione dei Mutscore ottenuti per ciascuna variante. Come indicato, sulle 115 varianti patogene/probabilmente patogene con Mutscore calcolabile, 100 (87%) presentano uno score  $>0,75$ . Se applichiamo in particolare i cut-off utilizzati da Quinodoz et al. nella popolazione di validazione dello studio<sup>13</sup>, corrispondenti rispettivamente a 0,14 (sotto il quale il 95% delle varianti del set di prova risulta costituito da varianti probabilmente benigne/benigne) e 0,73 (oltre il quale il 95% delle varianti del set di prova risulta costituito da varianti probabilmente patogene/patogene) osserviamo effettivamente che nessuna delle varianti patogene/probabilmente patogene identificate in questa coorte presenta un Mutscore  $<0,14$  e che 102 (88,7%) presentano un Mutscore  $> 0,73$ .

**Figura 15. Distribuzione dei Mutscore ottenuti per ciascuna variante missenso patogena/probabilmente patogena.**

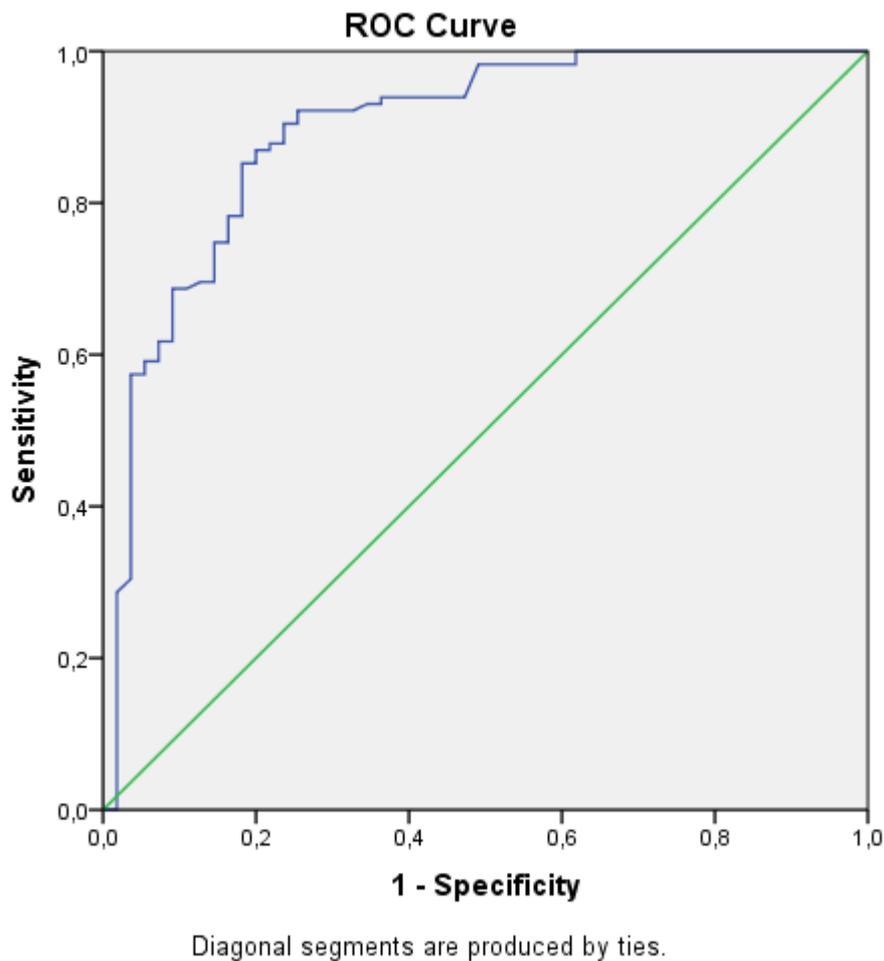


**Figura 15:** PLP: varianti probabilmente patogene/ patogene

Considerando le varianti patogene/probabilmente patogene vs le varianti di significato incerto della nostra coorte, abbiamo quindi valutato la capacità predittiva del Mutscore sul nostro set di dati effettuando l'analisi della curva ROC (Receiver Operating Characteristic) e ottenendo una AUC (Area Under the Curve) di 0,895. **Figura 16.**

Al fine di valutare la performance del Mutscore nel riflettere la reale patogenicità delle varianti, valutata secondo le evidenze scientifiche attuali, abbiamo successivamente verificato l'esistenza e l'eventuale grado di correlazione tra il Mutscore e l'interpretazione della variante annotata in ClinVar. Dall'analisi è quindi emersa l'esistenza di una correlazione positiva statisticamente significativa ( $r=0,65$ ,  $p\text{-value} = 0.000$ ) tra il Mutscore e il punteggio attribuito alla variante (1 per variante benigna, 2 per probabilmente benigna, 3 per variabile di significato incerto, 4 probabilmente patogena e 5 per patogena).

**Figura 16. Performance del Mutscore sul set di dati della coorte**



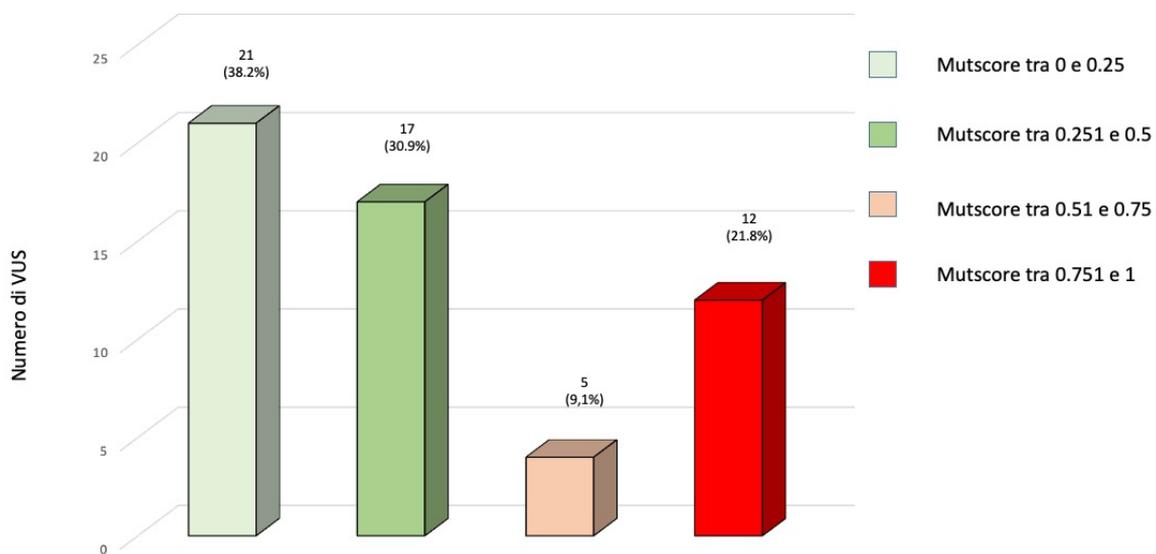
In particolare, attraverso un'analisi di regressione lineare univariata, è emerso che il Mutscore si configura come un predittore di patogenicità statisticamente significativo ( $R^2$  0,4; p-value = 0.000).

### **5.5.2 Analisi delle varianti di significato incerto e riclassificazione**

Questa parte dello studio si caratterizza per il potenziale clinico più rilevante. L'ipotesi di lavoro consiste infatti nel verificare se, applicando l'algoritmo Mutscore alle varianti finora classificate come di significato incerto, è possibile ottenere una riclassificazione (anche detta "disambiguazione") diagnosticamente utile. La **Figura 17** fornisce una rappresentazione grafica della distribuzione dei

Mutscore ottenuti per ciascuna variante di significato incerto. Sulle 55 varianti con Mutscore calcolabile, 21 (38,2%) presentano uno score tra 0 e 0,25; 17 (30,9%) uno score tra 0,251 e 0,5; 5 (9,1%) tra 0,51 e 0,75 e 12 (21,8%) presentano uno score tra 0,751 e 1. L'osservazione più interessante è che il Mutscore delle varianti di significato incerto presenta una distribuzione bimodale (nell'ipotesi nulla infatti la distribuzione sarebbe rimasta unimodale, o al massimo con un allargamento della distribuzione). Questa osservazione conferma che il calcolo del Mutscore aumenta significativamente la discriminazione fra varianti benigne e patogene.

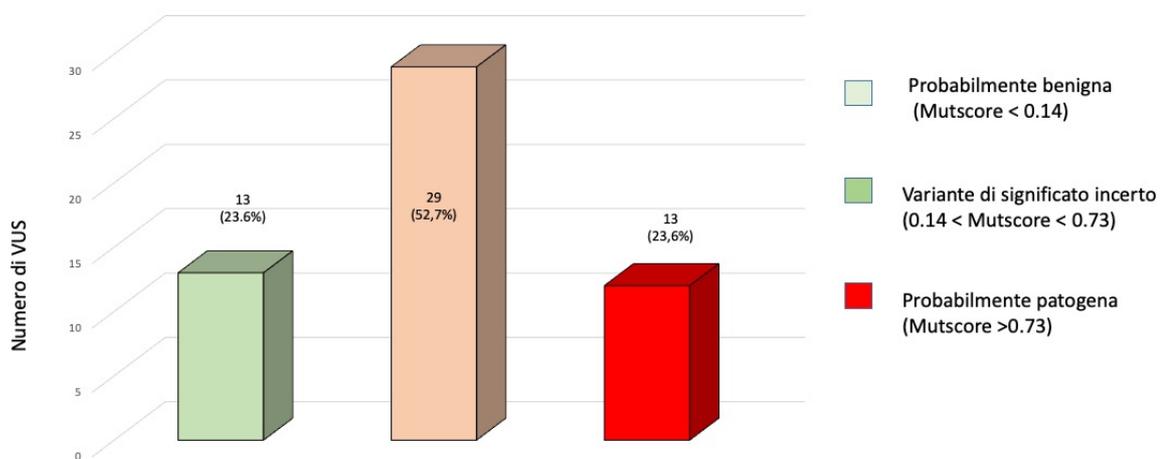
**Figura 17. Distribuzione dei Mutscore ottenuti per ciascuna variante missenso di significato incerto**



**Figura 17: VUS: Varianti di significato incerto**

Applicando i cut-off utilizzati da Quinodoz et al. nella popolazione di validazione dello studio<sup>13</sup> per la riclassificazione delle varianti di significato incerto, osserviamo in particolare che il Mutscore suggerisce la riclassificazione in varianti probabilmente patogene/patogene (Mutscore > 0,73) di 13 (23,6%) varianti di significato incerto e di altre 13 (23,6%) in varianti probabilmente benigne (Mutscore < 0,14). In totale, una disambiguazione sembra quindi possibile per 26 delle 55 varianti testate. **Figura 18.**

**Figura 18. Riclassificazione delle varianti di significato incerto secondo Mutscore**



**Figura 18:** VUS: Varianti di significato incerto

## **6. DISCUSSIONE**

Il presente studio ha permesso di valutare l'esperienza di 15 anni di consultazioni specializzate congiunte in Cardiogenetica, dei servizi di Cardiologia e di Genetica Medica dell'Ospedale Universitario di Losanna, e di dimostrare il ruolo chiave dell'analisi genetica nella conferma diagnostica e nella presa in carico dei pazienti. Abbiamo infatti analizzato per la prima volta la nostra coorte di pazienti con malattia cardiaca ereditaria accertata o sospetta, indirizzati alla nostra consultazione congiunta di Cardiogenetica e inclusi nel registro prospettivo delle Malattie Cardiache Ereditarie del Servizio di Genetica dell'Ospedale Universitario di Losanna. Sulla base di questa solida esperienza clinica, abbiamo quindi applicato il nuovo algoritmo predittivo Mutscore proposto da Quinodoz et al.<sup>13</sup> e calcolato mediante un approccio di apprendimento automatico ("machine learning"), al fine di testare la sua performance predittiva su un nuovo set di dati reali e la sua capacità di riclassificazione delle varianti di significato incerto.

### **6.1 Analisi dei risultati della popolazione di studio in relazione ai dati di letteratura**

Il nostro primo obiettivo di studio è consistito nell'analisi globale della coorte in termini di indicazioni diagnostiche e di risultati dei test genetici ottenuti. Globalmente, sui 488 pazienti testati, l'indicazione clinica più frequente (37,3%) è rappresentata dallo screening presintomatico in caso di variante familiare conosciuta. Per quanto riguarda invece le indicazioni diagnostiche tra i probandi, osserviamo un relativo equilibrio con 138 pazienti (28,3%) indirizzati per sindrome aritmica ereditaria e 131 (26,8%) per cardiomiopatia diagnosticata o sospetta. Questo risultato differisce dalle esperienze precedentemente pubblicate da altri centri<sup>16,17</sup> nei quali l'indicazione diagnostica è più frequentemente rappresentata dalle cardiomiopatie, probabilmente in virtù della loro più alta prevalenza rispetto alle sindromi aritmiche ereditarie. Il nostro risultato riflette tuttavia il peculiare quadro legale che definisce la pratica della Genetica Medica in Svizzera (Legge Federale sull'Analisi Genetica Umana (LAGH) del 2004 e messo a giorno nel 2018). Diversamente infatti dal sistema

italiano o francese in cui la presa in carico finanziaria delle analisi genetiche è pubblica, non implica costi per il paziente e la relativa prescrizione non è soggetta ad alcuna forma restrittiva, in Svizzera l'assunzione dei costi di analisi genetica non rientra tra le prestazioni rimborsate obbligatoriamente dalle società private di assicurazione-cassa malattia su cui è costruito il sistema sanitario e la cui stipula è obbligatoria per ogni cittadino svizzero. In caso di prescrizione di analisi genetica, il medico responsabile deve quindi indirizzare una domanda di presa in carico dei costi alla specifica cassa malattia del paziente. Ogni cassa-malattia è poi libera di decidere, caso per caso, se rimborsare i costi di analisi (sempre secondo la franchigia del paziente e una quota parte del 10%) o meno. Soprattutto nel passato, questo comportava che una domanda di analisi con la stessa indicazione diagnostica potesse talora determinare la presa in carico dei costi da parte di una cassa malattia e talora il rifiuto da parte di un'altra. Quest'ultimo caso determinava quindi delle ripercussioni maggiori sul seguire della presa in carico in quanto i pazienti finivano spesso con il rinunciare all'analisi genetica nel caso in cui i costi di analisi (per pannelli genetici intorno a 3-4'000 CHF) fossero stati completamente a loro carico, a seguito del rifiuto della cassa-malattia. Solo recentemente, la situazione è migliorata e i rifiuti di presa in carico dei costi di analisi genetica sono quindi estremamente sporadici. Tale contesto spiega tuttavia perché, soprattutto inizialmente, l'accesso all'analisi genetica fosse limitato e la prescrizione di analisi spesso confinata ai casi più severi, con prognosi apparentemente più preoccupante o a quelle patologie in cui l'analisi genetica potesse rappresentare una svolta chiave nel processo diagnostico e terapeutico. In tal senso, pertanto, mentre nei casi indice la diagnosi di cardiomiopatia ereditaria è inevitabilmente basata sugli esami di imaging e il dato genetico non è in grado di modificare sostanzialmente (salvo in rare eccezioni) la presa in carico, nelle sindromi aritmiche ereditarie la diagnosi clinica è significativamente più complessa. Il dato genetico può risultare dunque fondamentale sia per la conferma diagnostica che per la successiva presa in carico (per esempio nella Sindrome del QT Lungo). Tutto ciò spiega quindi nella nostra popolazione la più frequente indicazione al test in caso di sindrome aritmica (28,3% vs 18,2% nello studio di Janin et

al<sup>16</sup>) e la meno frequente in caso di cardiomiopatia ereditaria (26,8% vs 77,3% nello studio di Janin et al<sup>16</sup>).

Un altro dato particolarmente significativo è rappresentato dall'elevato rendimento diagnostico dell'analisi genetica nella nostra popolazione. Su 488 pazienti testati, una variante patogena/probabilmente patogena, sia essa missenso o meno, è stata infatti riscontrata in 198 pazienti per una resa diagnostica globale del 40,6%. Tale resa rimane ancora significativa scorpendo la quota di analisi effettuata per screening presintomatici, a fronte del riscontro di una variante patogena/probabilmente patogena nel 23,6% dei casi indice. Il rendimento diagnostico nella nostra coorte si attesta quindi come il più alto rispetto a quelli riportati in letteratura (20,3%<sup>16</sup> e 19,9%<sup>17</sup>) e riflette probabilmente, ancora una volta nel panorama specificamente svizzero, il maggior grado di sospetto clinico necessario a giustificare la domanda di presa in carico dei costi di analisi genetica presso una cassa malattia. Questo aspetto potrebbe aver quindi determinato un bias di selezione involontario legato alla realizzazione del test genetico nei casi fenotipicamente più severi di malattia, a causa dell'accesso ancora precedentemente limitato all'analisi genetica.

Globalmente, inoltre, il tasso di mutation-carriers identificati è risultato maggiore (45,8%) nei pazienti con cardiomiopatia rispetto a quelli (37%) affetti da sindromi aritmiche ereditarie. Più nel dettaglio, per quanto riguarda i rendimenti diagnostici specifici, questi variano notevolmente a seconda delle patologie. Tra le principali sindromi aritmiche osserviamo, tra i pazienti con Sindrome di Brugada, un rendimento diagnostico del 21,9%, in linea con i dati di letteratura<sup>6,18</sup>. Per i pazienti affetti da Sindrome del QT Lungo, il tasso di probandi mutation-carriers è stato calcolato a 50%, nettamente inferiore al 75-80% atteso secondo i dati di letteratura<sup>16,19</sup>. Questo risultato potrebbe tuttavia essere spiegato dal fatto che, sebbene sempre richiesti, gli esami relativi alla precisa fenotipizzazione del paziente (ECG, Holter, ergometrie etc...) non sono stati sistematicamente ottenuti. Questo aspetto, associato all'estrema complessità di diagnosi della malattia, potrebbe quindi aver portato a testare pazienti con sospetto di malattia troppo basso o non correttamente diagnosticati. Per quanto concerne invece le cardiomiopatie ereditarie, si è calcolato un rendimento diagnostico del

60,7% per i pazienti con cardiomiopatia ipertrofica, del 32,3% per quelli con cardiomiopatia dilatativa e del 33,3% per quelli con cardiomiopatia ventricolare di tipo non compaction, relativamente in linea con i dati riportati in letteratura<sup>16,20</sup>. Per la cardiomiopatia aritmogena, il tasso di mutation-carriers è stato calcolato a 34,8%, relativamente inferiore ai dati di letteratura che stimano il rendimento diagnostico dell'analisi genetica tra il 50-60%<sup>6</sup>. Anche in questo caso, tuttavia, l'assenza di accesso sistematico agli esami necessari per la fenotipizzazione unito alla complessità di diagnosi potrebbero spiegare questo risultato.

Per quanto riguarda il tipo di varianti patogene riscontrate, come atteso, queste sono state identificate sui geni con solida evidenza scientifica di associazione causale alla patologia in esame. Tra i pazienti affetti da sindrome del QT lungo, per esempio, il 92,3% delle varianti probabilmente patogene/patogene sono state identificate a livello dei 3 geni principalmente interessati (*KCNQ1*, *KCNH2* e *SCN5A*), in linea con i dati di letteratura<sup>6</sup>. Tra l'altro, a questo soggetto, è interessante notare come in 2 casi con sospetta diagnosi clinica di Sindrome del QT lungo si siano identificate delle varianti probabilmente patogene/patogene a carico del gene *CASQ2*, che hanno permesso quindi di modificare la diagnosi finale ritenuta in tachicardia ventricolare polimorfa catecolaminergica. Tra i pazienti con cardiomiopatia ipertrofica invece il 43,2% delle varianti probabilmente patogene/patogene sono state identificate sul gene *MYBPC3* e un altro 43,2% sul gene *MYH7*. In particolare, come atteso, l'81,2% di quelle sul gene *MYBPC3* sono rappresentate da varianti troncanti non missenso mentre per gli altri geni identificati le varianti sono perlopiù missenso<sup>6</sup>. Tra i pazienti con cardiomiopatia dilatativa familiare tutte le mutazioni identificate si localizzano sui geni causalmente associati con maggiore evidenza scientifica (*TTN*, *LMNA*, *TNNT2*). In particolare, le varianti patogene/probabilmente patogene riscontrate sul gene *TTN* sono le più frequenti (54,5%) e, in accordo con i dati di letteratura<sup>6</sup>, sono prevalentemente troncanti non missenso (83,3%). Analizzando i pazienti con sospetta cardiomiopatia aritmogena, come atteso<sup>6</sup>, *PKP2* è risultato essere il gene più frequentemente mutato (50%), seguito da altri geni desmosomali come *DSP*, *DSG2* e non desmosomali come *ACTN2*. Tra i pazienti con cardiomiopatia ventricolare sinistra di tipo non

compaction, infine, le due varianti identificate sono situate rispettivamente su *MYH7* e *TTN*, che assieme a *MYBPC3*, rappresentano i tre geni più frequentemente implicati in questo tipo di cardiomiopatia<sup>6</sup>.

## **6.2 Applicazione dell' algoritmo predittivo Mutscore nella nostra coorte di studio**

A seguito dell'esame dettagliato della nostra popolazione di studio, abbiamo quindi analizzato ogni variante probabilmente patogena/patogena e di significato incerto repertoriata, attraverso l'ausilio del nuovo modello predittivo computazione Mutscore. Il nostro obiettivo di studio successivo è consistito infatti nel verificare l'esistenza e l'eventuale grado di correlazione tra il Mutscore e la reale patogenicità delle varianti identificate, al fine di testare la performance predittiva del Mutscore su un nuovo set di dati reali. In particolare, applicando i cut-off identificati da Quinodoz et al. nella popolazione di validazione dello studio<sup>13</sup>, corrispondenti rispettivamente a 0,14 (sotto il quale il 95% delle varianti del set di prova risulta costituito da varianti probabilmente benigne/benigne) e 0,73 (oltre il quale il 95% delle varianti del set di prova risulta costituito da varianti probabilmente patogene/patogene), abbiamo effettivamente osservato che, tra le varianti patogene/probabilmente patogene identificate nella nostra coorte, nessuna presenta un Mutscore  $< 0,14$  mentre l'88,7% presenta un Mutscore  $> 0,73$ . Abbiamo inoltre dimostrato un'ottima capacità predittiva del Mutscore sul nostro set di dati, come attestato da una AUC (Area Under the Curve) di 0,895. Un'ulteriore comprova della performance del Mutscore nel riflettere la reale patogenicità delle varianti, è stata infine ottenuta dimostrando l'esistenza di una correlazione positiva statisticamente significativa ( $r = 0,65$ ,  $p\text{-value} = 0.000$ ) tra il Mutscore e l'interpretazione della variante annotata in ClinVar. Nella nostra coorte di studio, abbiamo inoltre evidenziato che il Mutscore si configura come un predittore di patogenicità statisticamente significativo ( $R^2 0,4$ ;  $p\text{-value} = 0.000$ ).

Analizzando la capacità di riclassificazione delle varianti di significato incerto secondo i cut-off identificati da Quinodoz et al. nella popolazione di validazione dello studio<sup>13</sup>, abbiamo infine dimostrato che il Mutscore suggerisce la riclassificazione del 47,3% delle varianti di significato

incerto identificate. Tuttavia, diversamente dallo studio di Quinodoz et al.<sup>13</sup> in cui il Mutscore sembrava presentare una particolare peculiarità soprattutto nella riclassificazione in varianti probabilmente benigne (33,7% vs 21,5%), nella nostra popolazione di studio osserviamo una capacità riclassificativa in varianti probabilmente benigne e probabilmente patogene del tutto sovrapponibile (23,6% vs 23,6%).

## **7. CONCLUSIONE**

In conclusione, il nostro studio ha permesso di valutare l'esperienza di 15 anni di consultazioni specializzate congiunte in Cardiogenetica, dei servizi di Cardiologia e di Genetica Medica dell'Ospedale Universitario di Losanna, con un totale di 488 pazienti testati. Questo lavoro rappresenta pertanto un contributo fondamentale nel panorama della letteratura internazionale dove finora, a nostra conoscenza, solo due altri studi<sup>16,17</sup> hanno riportato l'esperienza acquisita nella propria popolazione di pazienti testati per sindromi aritmiche o cardiomiopatie ereditarie. Il nostro lavoro ha inoltre l'importante peculiarità di riflettere le particolari specificità della pratica della Genetica Medica in Svizzera caratterizzata, soprattutto ai suoi albori, da una presa in carico finanziaria non completamente scevra da costi per il paziente (soprattutto in caso di mancato rimborso da parte delle casse-malattia) e gravata da restrizioni più o meno variabili imposte dalle casse-malattia. Questi aspetti permettono quindi di spiegare l'accesso limitato all'analisi genetica e/o la prescrizione di analisi confinata ai casi più severi o a quelle patologie in cui l'analisi genetica potesse rappresentare una svolta chiave nel processo diagnostico-terapeutico. Nonostante queste caratteristiche, il nostro studio ha permesso di dimostrare il ruolo fondamentale dell'analisi genetica nella conferma diagnostica (nel 40% dei pazienti testati) e nella presa in carico dei pazienti, come dimostrato dai casi in cui la diagnosi genetica molecolare ha fornito informazioni prognostico-terapeutiche (per es. nella Sindrome del QT lungo) o ha permesso di sovvertire completamente la diagnosi clinica e la conseguente presa in carico (si pensi ai casi documentati con sospetta diagnosi clinica di Sindrome

del QT lungo e con diagnosi molecolare accertata di tachicardia ventricolare polimorfa catecolaminergica).

Sulla base di questa solida esperienza clinica, abbiamo infine validato in un nuovo set di dati reali l'utilità del nuovo algoritmo predittivo Mutscore, incentrato su un approccio di apprendimento automatico ("machine learning"). Abbiamo infatti dimostrato nella nostra coorte di pazienti la sua ottima performance predittiva e la sua accuratezza nell'aiutare a riclassificare le varianti di significato incerto. Sebbene ulteriori studi di validazione siano ancora necessari, il Mutscore sembra essere un valido strumento ed una potente risorsa in grado di contribuire all'ulteriore sviluppo e consolidamento di un approccio orientato verso la medicina di precisione.

## 8. BIBLIOGRAFIA

1. Charron P, Elliott P. Time for education in cardiogenetics. *Cardiogenetics* 2011; 1:e2
2. Richard P, Ader F, Charron P. Génétique des cardiomyopathies héréditaires. *EMC - Cardiologie*, 13; 3: Agosto 2018
3. Maron BJ, Haas TS, Ahluwalia A, Murphy CJ, Garberisch RF. Demographics and epidemiology of sudden deaths in young competitive athletes: from the United States National Registry. *Am J Med* 2016; 129: 1170-7
4. Barry J Maron, Jeffrey A Towbin, Gaetano Thiene, Charles Antzelevitch, Domenico Corrado, Donna Arnett, Arthur J Moss, Christine E Seidman, James B Young, American Heart Association; Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; 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 Apr 11;113(14):1807-16.
5. Schwartz PJ, Ackerman MJ, Antzelevitch C, Bezzina CR, Borggrefe M, Cuneo BF, Wilde AM. Inherited cardiac arrhythmias. *Nature Reviews* 2020; 6:58.
6. Wilde AAM, Semsarian C, Márquez MF, Sepeshri Shamloo A, Ackerman MJ, Ashley EA, Sternick EB, Barajas-Martinez H, Behr ER, Bezzina CR, Breckpot J, Charron P, Chockalingam P, Crotti L, Gollob MH, Lubitz S, Makita N, Ohno S, Ortiz-Genga M, Sacilotto L, Schulze-Bahr E, Shimizu W, Sotoodehnia N, Tadros R, Ware JS, Winlaw DS, Kaufman ES; ESC Scientific Document Group. European Heart Rhythm Association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) Expert Consensus Statement on the state of genetic testing for cardiac diseases. *Europace*. 2022 Apr 4:euac030.

7. Limongelli G. A new era in cardiogenetics. *Cardiogenetics* 2011; 1:e1.
8. Walsh R, Adler A, Amin AS, Abiusi E, Care M, Bikker H, Amenta S, Feilotter H, Nannenber EA, Mazzarotto F, Trevisan V, Garcia J, Hershberger RE, Perez MV, Sturm AC, Ware JS, Zareba W, Novelli V, Wilde AAM, Gollob MH. Evaluation of gene validity for CPVT and short QT syndrome in sudden arrhythmic death. *Eur Heart J.* 2022 Apr 14;43(15):1500-1510.
9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015 May;17(5):405-24.
10. Uncertain Results Related to Patient Symptoms: Variants of Uncertain Significance. [www.ashg.org/education/csertoolkit/uncertainresults.html](http://www.ashg.org/education/csertoolkit/uncertainresults.html)
11. Werner-Lin A, Mccoyd JLM, Bernhardt BA. Actions and Uncertainty: How Prenatally Diagnosed Variants of Uncertain Significance Become Actionable. *Hastings Cent Rep.* 2019 May;49
12. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011 Apr;32(4):358-68.
13. Quinodoz M, Peter VG, Cisarova K, Royer-Bertrand B, Stenson PD, Cooper DN, Unger S, Superti-Furga A, Rivolta C. Analysis of missense variants in the human genome reveals widespread gene-specific clustering and improves prediction of pathogenicity. *Am J Hum Genet.* 2022 Mar 3;109(3):457-470.
14. Ackerman JP, Bartos DC, Kapplinger JD, Tester DJ, Delisle BP, Ackerman MJ. The Promise and Peril of Precision Medicine: Phenotyping Still Matters Most. *Mayo Clin Proc.* 2016 Oct 8:S0025-6196(16)30463-3.
15. Catherine M Otto, Jainy J Savla, Fuki M Hisama. Cardiogenetics: a primer for the clinical cardiologist. *Heart.* 2020 Jun;106(12):938-947.

16. Janin A, Januel L, Cazeneuve C, Delinière A, Chevalier P, Millat G. Molecular Diagnosis of Inherited Cardiac Diseases in the Era of Next-Generation Sequencing: A Single Center's Experience Over 5 Years. *Mol Diagn Ther*. 2021 May;25(3):373-385.
17. Dellefave-Castillo LM, Cirino AL, Callis TE, Esplin ED, Garcia J, Hatchell KE, Johnson B, Morales A, Regalado E, Rojahn S, Vatta M, Nussbaum RL, McNally EM. Assessment of the Diagnostic Yield of Combined Cardiomyopathy and Arrhythmia Genetic Testing. *JAMA Cardiol*. 2022 Sep 1;7(9):966-974.
18. Kapplinger JD, Tester DJ, Alders M, Benito B, Berthet M, Brugada J, Brugada P, Fressart V, Guerchicoff A, Harris-Kerr C, Kamakura S, Kyndt F, Koopmann TT, Miyamoto Y, Pfeiffer R, Pollevick GD, Probst V, Zumhagen S, Vatta M, Towbin JA, Shimizu W, Schulze-Bahr E, Antzelevitch C, Salisbury BA, Guicheney P, Wilde AA, Brugada R, Schott JJ, Ackerman MJ. An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart Rhythm*. 2010 Jan;7(1):33-46.
19. Adler A, Novelli V, Amin AS, Abiusi E, Care M, Nannenber EA, Feilotter H, Amenta S, Mazza D, Bikker H, Sturm AC, Garcia J, Ackerman MJ, Hershberger RE, Perez MV, Zareba W, Ware JS, Wilde AAM, Gollob MH. An International, Multicentered, Evidence-Based Reappraisal of Genes Reported to Cause Congenital Long QT Syndrome. *Circulation*. 2020 Feb 11;141(6):418-428.
20. Ommen S.R., Mital S., Burke M.A., Day S.M., Deswal A. et al. 2020 AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2020; Vol. 142 (No. 25): 558-631.

## 9. ANNESSI

**Articoli pubblicati (o under review) durante il dottorato (2019-2022) sulla popolazione di studio oggetto della tesi.**

- **In press in the European Journal of Medical Genetics (IF 2.7)**

1 **The « Amish » NM\_000256.3:c.3330+2T>G splice variant in *MYBPC3* associated**  
2 **with hypertrophic cardiomyopathy is an ancient Swiss mutation.**

3

4 Claire Redin<sup>1,@</sup>, Despina Christina Pavlidou<sup>2,3</sup>, Zahurul Bhuiyan<sup>2,3</sup>, Alessandra Pia  
5 Porretta<sup>4,5</sup>, Pierre Monney<sup>3,4</sup>, Nicola Bedoni<sup>2,3</sup>, Fabienne Maurer<sup>2,3</sup>, Nicole Sekarski<sup>6</sup>,  
6 Isis Atallah<sup>2,3</sup>, Émeline Davoine<sup>2,3</sup>, Xavier Jeanrenaud<sup>4</sup>, Etienne Pruvot<sup>4</sup>, Jacques  
7 Fellay<sup>1</sup>, Andrea Superti-Furga<sup>2,3@</sup>

8

9

10 1. Precision Medicine Unit, Lausanne University Hospital (CHUV), 1011 Lausanne,  
11 Switzerland.

12 2. Division of Genetic Medicine, Lausanne University Hospital (CHUV), 1011 Lausanne,  
13 Switzerland.

14 3. University of Lausanne, 1011 Lausanne, Switzerland.

15 4. Service of Cardiology, Lausanne University Hospital (CHUV), 1011 Lausanne, Switzerland.

16 5. Department of Clinical-Surgical Diagnostic and Pediatric Sciences, University of Pavia,  
17 Pavia, Italy.

18 6. Pediatric cardiology, Women-Mother-Child department, Lausanne University Hospital  
19 (CHUV), 1011 Lausanne, Switzerland.

20 @. Corresponding authors: correspondence to:

21 Prof. Andrea SUPERTI-FURGA,

22 Division of Genetic Medicine

23 Hôpital Nestlé

24 Av. Pierre Decker 5

25 1011 Lausanne

26 Tel: +4121 314 35 60

27 [Andrea.Superti-Furga@chuv.ch](mailto:Andrea.Superti-Furga@chuv.ch), or

28

29 Dr Claire Redin, Translational Research Lead

30 LVBC - CHUV

31 Chemin des Roches 1a/1b

32 1011 Lausanne

## ABSTRACT

MYBPC3 is the most frequently mutated gene in hypertrophic cardiomyopathy (HCM). Several loss-of-function founder variants have been reported in MYBPC3 from various geographic regions, altogether suggestive of a modest or absent effect of these variants on reproductive fitness. One of them, a MYBPC3 splice variant, NM\_000256.3:c.3330+2T>G, was first described at homozygosity in newborns presenting with a severe, recessive form of HCM among the Amish population and was later associated with adult-onset dominant HCM in heterozygous carriers. We here report this splice variant at heterozygosity in eight unrelated Swiss families with HCM, making it the most prevalent cardiomyopathy variant in western Switzerland. This variant was identified in patients using targeted (n=5) or full-genome sequencing (n=3). Given the prevalence of this variant in the Old Order Amish, Mennonites and Swiss populations, and given that both Amish and Mennonites founders originated from the Bern Canton in Switzerland, the MYBPC3, NM\_000256.3:c.3330+2T>G variant appears to be of Swiss origin. Neighboring regions that hosted the first Amish settlements (Alsace, South Germany) should be on the lookout for that variant. The existence of MYBPC3 founder variants in different populations suggests that individuals with early-onset clinical disease may be the tip of the iceberg of a much larger number of asymptomatic carriers. Alternatively, reproductive fitness could even be slightly increased in some variant carriers to compensate for the reduction of fitness in the more severely affected ones, but this remains to be investigated.

Keywords: MYBPC3, hypertrophic cardiomyopathy, founder variant, phasing

## INTRODUCTION

Inherited cardiomyopathies are associated with a significant risk for heart failure, arrhythmias, sudden cardiac death (SCD) and stroke. Among these, HCM is the most common with a revised prevalence of up to 1 in 200 individuals in the general population (ref. 1). HCM is mostly inherited as an

autosomal dominant trait but with variable penetrance and expressivity. Moreover, it is genetically heterogeneous: variants in at least 11 genes, mostly encoding sarcomeric proteins, have been associated with the disease (ref. 2-4). MYBPC3 is the most prevalent HCM-associated gene, accounting for over 40% of all reported HCM variants. Most MYBPC3 mutations (~75%) are loss-of-function (LoF) variants, all leading to absent or truncated protein and therefore lacking the major binding domains to other sarcomeric proteins. Interestingly, MYBPC3 mutations are associated with a relatively favorable disease outcome and a lower disease penetrance as only 58% of adult carriers below 50 are symptomatic; overall survival also seems to be increased compared to other genes associated with HCM, such as MYH7 (ref. 5,6). Such an attenuated disease prognosis is compatible with more favorable reproductive fitness and therefore less stringent negative selection and higher carrier frequency rates. In line with the latter, a palette of founder mutations in MYBPC3 has been described with an allelic frequency among local populations of up to 0.36% (ref. 7). A few cases with biallelic MYBPC3 mutations have also been reported with a much more severe, neonatal HCM phenotype (ref. 8-10). We here focus on the MYBPC3 splice site mutation NM\_000256.3:c.3330+2T>G running in eight unrelated Swiss families with dominant HCM, which was originally reported at the homozygous state in Amish newborns with severe HCM (ref. 9). Based on our inpatient clinic population, we find that this is likely the most frequent MYBPC3 and even HCM-causing mutation in western Switzerland. Haplotype reconstruction indicates that NM\_000256.3:c.3330+2T>G is a variant of Swiss origin whose common ancestor dates back to at least 200 years ago and traveled to the US with the first Amish founders. We also highlight that, unlike with other MYBPC3 variants, the associated HCM phenotype is not particularly mild given that two of our patients underwent heart transplantation before age 50.

## SUBJECTS AND METHODS

### Patients

Five unrelated probands were referred because of HCM to our multidisciplinary cardiogenetic clinic. All probands underwent a thorough clinical evaluation including electrocardiogram, transthoracic echocardiography, and MRI. A confirmed diagnosis of HCM (left ventricular wall thickness > 15mm; ref. 11,12) by the cardiology division prompted genetic investigations (see Table 1). Three other unrelated patients, with a confirmed diagnosis of HCM, were enrolled based on a familial history of primary cardiomyopathy as part of an independent research study on sudden cardiac death (SCD) susceptibility (Table 1). For all eight individuals, a dedicated, written informed consent – for either clinical diagnostic or research purposes, respectively – was obtained, as approved by local ethical review boards (Ethics Board of the Lausanne University Hospital, Research Ethics Board from Vaud Canton) and in compliance with Swiss laws.

#### Exome and whole genome sequencing

Genetic testing strategies varied based on the date and channel through which each patient was recruited. For the five patients seen in the cardiogenetics clinic, blood was obtained for molecular analysis in the context of the diagnostic pathway. Genomic DNA was isolated from peripheral blood. Library preparation was performed using Illumina TruSeqOne or Twist Exome capture kit, and sequencing was performed on an Illumina NextSeq platform. Raw sequencing data was then processed on-site using routine diagnostic bioinformatic pipelines. For patient 3, a 20-year old girl, segregation studies were performed that showed the presence of the variant in her brother, who also had HCM, and in her mother who was clinically asymptomatic and had a normal cardiac ultrasound. The remaining three patients had been enrolled in a research study and their DNA samples were available at the CHUV institutional biobank. DNA extraction, normalization and quality controls were performed at the CHUV biobank laboratory. Whole-genome sequencing (WGS) was achieved at the Health2030 Genome Center in Geneva using Illumina TruSeq PCR-free library preparation protocols, which include: DNA fragmentation, end-repair, ligation of sequencing adapters and barcodes, size selection of 600-bp fragments, quantitation and normalization of individual libraries.

Normalized libraries were loaded to a NovaSeq machine from Illumina for a 150-bp paired117 end sequencing run, aiming for a minimal 30X final depth of coverage. Raw sequencing data was then sent to us for further processing. All MYBPC3 variants were subsequently confirmed using Sanger sequencing at our clinical-grade diagnostic laboratory, prior to disclosure for research participants.

### Bioinformatic pipelines

For patients enrolled from the clinic, raw data processing, variant annotation and variant prioritization were performed as described previously (ref. 13). For the remaining three patients with full genome investigation, the analytic pipeline was standard and followed GATK best practices (ref. 14). For each sample, raw read pairs were converted to unmapped BAMs and then aligned to the human reference genome version hg38 using BWA-MEM v0.7.17 (ref. 15), independently per sequencing lane and flowcell. Resulting BAM files were merged and read duplicates were marked using PicardTools v2.21.8 MarkDuplicates. Deduplicated BAM files were sorted and indexed using PicardTools v2.21.8 SortSam. Base quality score recalibration was modeled and run using GATK v4.1.3 BaseRecalibrator and GATK v4.1.3 ApplyBQSR. Variant calling was performed individually using GATK v4.1.3 HaplotypeCaller, producing one gVCF file per sample. Genotyping was performed jointly for the full cohort of >100 samples recruited 134 for this research study using GATK v4.1.3 GenotypeGVCFs (ref. 16), producing a unique multi-sample gVCF. Indels and single-nucleotide variants (SNVs) were recalibrated using GATK v4.1.3 VariantRecalibrator and ApplyVQSR. Genotypes were refined with GATK CalculateGenotypePosteriors, using 1KG VCF files as a gold standard. Low quality variants (read depth <10 and genotype quality < 20) were flagged. FASTQ, BAM and VCF files metrics were gathered using PicardTools v2.21.8 CollectQualityYieldMetrics, CollectMultipleMetrics and CollectWgsMetrics, GATK v4.1.3 CollectVariantCallingMetrics, and FastQC (ref. 17). Coverage metrics were obtained using BEDTools v 2.29.2 and Mosdepth (ref. 18,19). Sample relatedness, sex-check and ancestry inference were performed using Peddy (ref. 20).

The final list of variants was annotated with ANNOVAR and vcfanno (ref. 21,22) using external databases. Variants were then ranked and prioritized based on population and cohort allelic frequencies, predicted gene impact, and association to disease. Only variants in genes previously associated with a cardiac phenotype (approximately 600 genes) were screened to minimize the risk of incidental findings.

### Haplotype reconstruction

We first used whatsHap to construct haplotype blocks of nearby heterozygous variants using sequencing reads from the bam/cram files aligned on hg19 for WES data (patients 1, 2, 3a, 7, 8), or on hg38 for WGS data (patient 4, 5, 6) (ref. 23). Pre-phased genotypes in VCF files were then passed to shapeit4.0 and compared to population-based phasing calls in order to reconstruct reliable haplotype estimates for chromosome 11 (ref. 24). Genetic recombination maps for hg19 and hg38 were obtained from shapeit4. We used phasing data from the 1000 Genomes Project phase 3 as a reference (3'202 samples total). Final phased data from patients 1, 2, 3a, 7, 8 was then lifted-over to hg38 to be collated to phased data from patients 4-6. WES data from the first three patients were much scarcer and the vast majority of SNP genotypes were missing (see Table 2). 160 Final, reconstructed haplotypes containing the MYBPC3 variant were then extracted and compared between patients in order to extract the shared region of identity-by-descent (IBD).

### Estimation of mutation age by IBD

We assumed that the shared IBD haplotype is derived from a common ancestor across the six patients. To estimate the age of that common ancestor, we used the size of this shared haplotype as a proxy for the number of meiotic recombination steps that occurred since this ancestor. We used the following formula (ref. 25):  $L_{hap} = 200 / K = 200 / (A \sim c)$  where  $L_{hap}$  is the length of the shared IBD region in centimorgans,  $K$  is the number of meiosis in the descent tree until coalescence,  $A$  = age in generations,  $c$  = number of copies of the shared haplotype. Solving for  $A$  gives an estimate of the age

of the common ancestor. We used genetic recombination maps for hg38 to extract the haplotype length into centimorgans.

With the scarcer data yet from eight patients, the smallest IBD region is of 4.9 Mb (first to last common genotypes across all eight patients), and the largest IBD region is of 11.3 Mb (first to last discording genotypes between patients). As there are increased IBD sharing signals on both sides of the centromere, this yields to minor differences in terms of centimorgans and therefore in age estimation:

- $A_{AllPatients\_LowerBound} = 200 / (3.63 \text{ cM} \times 8) = 6.9$  generations
- $A_{AllPatients\_UpperBound} = 200 / (3.86 \text{ cM} \times 8) = 6.5$  generation.

With exhaustive genotyping data but from only three patients:

- $A_{ThreePatients} = 200 / (3.8 \times 3) = 17.5$ .

Assuming 25 years per generation, this traces the last common ancestor back to ~160-170 years ago, therefore in the 19th century.

## RESULTS

Same MYBPC3 splice variant identified in eight unrelated Swiss individuals with HCM Analysis of WGS data from 100 cases with either primary cardiomyopathies (HCM: n=30, dilated cardiomyopathy – DCM: n=11, arrhythmogenic right ventricular dysplasia - ARVD: n=11), arrhythmias (LQTS: n=13, Brugada syndrome: n=8), or early-onset SCD (n=27) identified the same heterozygous variation in MYBPC3 (NM\_000256.3:c.3330+2T>G) in three unrelated male patients (cf. Fig. 1). A parallel screening of our patient database from the cardiogenetic clinic identified five additional, unrelated heterozygous carriers, all presenting with an HCM phenotype (cf. Table 1). The affected sibling of the index case from Family 3 as well as the clinically unaffected mother was later confirmed to also carry the MYBPC3 variant. Interestingly, mutation carriers from all eight families express a variable disease course, with members of Family 1 and 3 presenting with a more severe phenotype than typically reported for MYBPC3 variants. Indeed, among them, two members

underwent heart transplantation at 46 and 18 years old respectively. The MYBPC3 variant was clinically reported as pathogenic to all patients and families. This splice site variant was originally reported at the homozygous state in newborns with severe, lethal HCM disease among the Amish population (ref. 9,10). Some probands' relatives presented some HCM-related symptoms such as chest pain, fatigue and palpitation, or even SCD suggesting that heterozygous carriers could be at risk of cardiac disease (ref. 9). A subsequent study specifically investigating heterozygous carriers of this mutation within the Amish population of Ohio showed that they indeed also harbored HCM features (ref. 26). Transcript studies demonstrated that this variant results in the introduction of a premature stop codon via skipping of exon 30, thus leading to a frameshift variant (NP\_000247.2:p.Asp1064Glyfs\*38), confirming its classification as a Pathogenic variant for dominant HCM according to ACMG guidelines (applied criteria: PVS1, PP1, PM2).

Most frequent MYBPC3, HCM-associated variant in 212 western Switzerland

Approximately 170 individuals with HCM have visited the Cardiogenetics clinic at our hospital. Among those who underwent genetic testing, only a few pathogenic variants were observed more than once in our cohort (e.g. MYH7: NM\_000257.4:c.2652\_2654del (p.Lys884del) and MYBPC3: NM\_000256.3:c.2905C>T (p.Gln969Ter) observed four and three times respectively in presumably unrelated families). The MYBPC3, NM\_000256.3:c.3330+2T>G splice variant is thus remarkable because it was observed in five apparently unrelated families from our clinic, as well as in three unrelated individuals from our parallel research study. This is the most common disease-causing variant in our HCM patient population (MYBPC3 splice variant allelic frequency in our inpatient clinic cohort:  $5/340 = 1.47E-02$ , expected carrier frequency:  $2.92E-02$ ), leading to an extrapolated carrier frequency of about  $1.46E-04$  and corresponding allelic frequency of  $7.3E-05$  within the Swiss population when considering an HCM prevalence of 1 in 200 (ref. 1). This is likely an underestimate given the incomplete penetrance of the mutation, and given that not all HCM patients have undergone genetic testing; however, it matches closely the allelic frequency reported in Europeans from gnomAD (MAF:  $7E-05$  in non-Finnish Europeans). The variant allelic frequency is below the MAF

threshold for ACMG BA1 criteria for MYBPC3 variants using conservative disease attributes, which was estimated at  $1.7E-4$  (HCM prevalence of 1:200, variant penetrance of 60%, MYBPC3 gene contribution in HCM of 40%, maximum allelic heterogeneity contribution of 10%, computed using <https://www.cardiodb.org/allelefrequencyapp/>). Hypothesis of a founder mutation of Swiss origin. The majority of the homozygous carriers of the NM\_000256.3:c.3330+2T>G mutation originate from the Geauga settlement of Ohio (ref. 27,28). Yet, multiple cases from other Amish families across the US were reported as presumable carriers, suggesting the mutation came from a common ancestor. The NM\_000256.3:c.3330+2T>G mutation was also found in one family from the Mennonite community (ref. 9). 238 Both high prevalence and large geographic distribution are suggestive of more distant common ancestors. Interestingly, Old Order Amish and Old Order Mennonite populations share a similar geographic origin, as both emigrated from Switzerland to the US. Indeed, from 1683–1880, the ancestors of the present-day Old Order Mennonites and Old Order Amish emigrated in eastern Pennsylvania from Switzerland and South Germany (ref. 29). The first Mennonite settlement in Pennsylvania dates back to the 1680s, while the first Amish settlement dates back to 1714 (ref. 30). Before their migration, they were mostly located in Bern Canton (Switzerland), Palatinate (southwestern Germany) and Alsace (eastern France). Notably, all the mutation-carrying individuals and families from this study had at least one branch in their pedigrees originating from Central or Western Switzerland (cf. Fig. 2). From public databases, this variant is listed in ClinVar ([VCV000008621.10](https://www.ncbi.nlm.nih.gov/clinvar/variant/MYBPC3/3330+2T>G)), reported from diagnostic laboratories from the United States but also Germany and Switzerland. In gnomAD, the variant was exclusively observed within the non-Finnish European population. Haplotype reconstruction of this variant in our patients identified an IBD region of 3.6-3.9 centimorgans across patients (cf. Table 2), dating the last common ancestor to less than 200 years ago. Altogether, these data suggest that the NM\_000256.3:c.3330+2T>G variant is of Swiss/South-German origin, and spread through a common ancestor that we can date back to the 19<sup>th</sup> century.

## DISCUSSION

In this study, we report the same variant, NM\_000256.3:c.3330+2T>G, in MYBPC3 identified either by exome or whole-genome sequencing, in eight unrelated probands with HCM. Clinical examination revealed typical non-obstructive HCM phenotypes. Interestingly, two affected siblings including the patient with the earliest-onset and more severe phenotype inherited the variant from their mother, who is asymptomatic 263 and has a normal cardiac ultrasound at age 50 years, suggesting a reduced penetrance.

This variant, with an estimated allelic frequency of 7.3E-05 in the Swiss population, appears to be the most common variant associated with dominant HCM in Switzerland. Its initial report in the Amish and Mennonite populations along with its predominance in Europeans from both diagnostic (ClinVar) and population databases (gnomADv3) all suggest that it is a founder variant whose common ancestor originated in Switzerland/Southern-Germany back in the 1800s. Given the putatively reduced penetrance of this variant, its frequency in the general population may be underestimated, meaning that many carriers are left undiagnosed. Some associated recessive cases may arise. Neighboring regions that hosted the first Amish settlements (Alsace, South Germany) should be on the lookout for that variant.

Interestingly, numerous other founder mutations in MYBPC3 have been reported in various geographical regions across the globe: USA, Iceland, Poland, Spain, France, Finland, Japan (ref. 31-38). The MYBPC3 gene itself does not seem to be a mutational hotspot (there are not many recurrent variants) but once mutations arise in a given population, they appear to be fixed in spite of their pathogenic potential. Such a significant prevalence of founder mutations in different populations raises interesting hypotheses: (1) the phenotypic penetrance of these variants must be reduced and/or age-dependent, indicating that many carriers are probably not diagnosed; (2) the independent occurrence of MYBPC3 LoF variants in various geographic regions and their persistence across generations indicate that overall genetic fitness is at most mildly reduced. If this is true, severe phenotypes as observed in two of our patients must be rare, perhaps being the tip of the iceberg of a much larger number of individuals who remain asymptomatic. (3) Finally, reproductive fitness could

even be slightly increased in some young adults to compensate for the reduction of fitness in the more severely affected individuals.

Does a MYBPC3 HCM variant confer a transient period of cardiac « overperformance » before becoming clinically relevant? This hypothesis is purely speculative 288 but deserves to be kept in mind.

#### FIGURE TITLES AND LEGENDS

Figure 1: Pedigrees of all eight families.

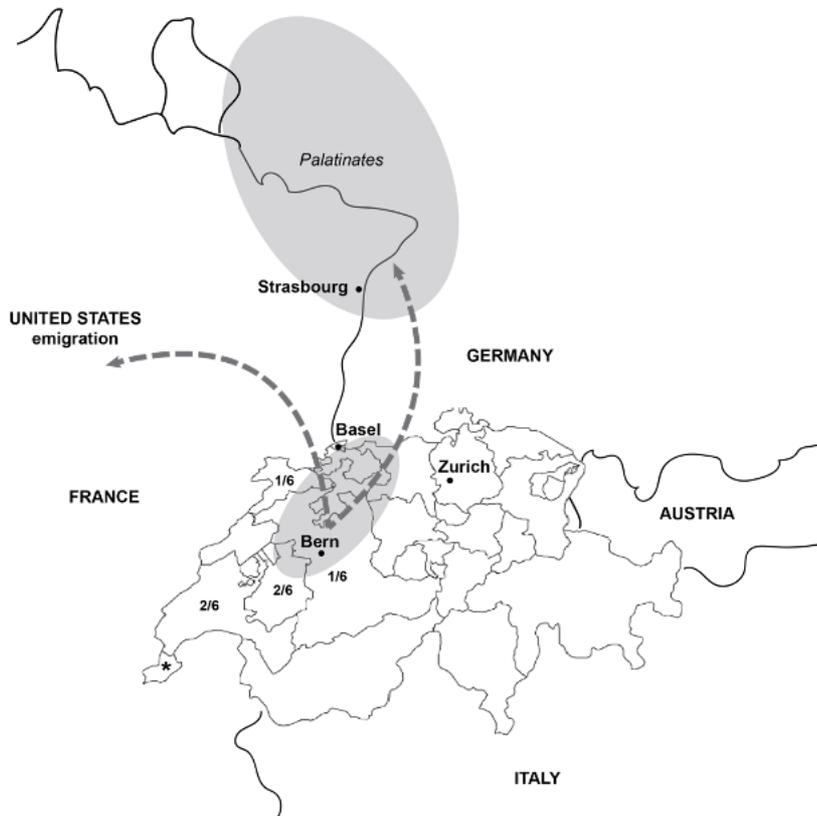
CAD: coronary artery disease; HF: heart failure; MI: myocardial infarct; NCD: non-cardiac death; PM: pacemaker; SUD: sudden unexplained death; asymp.: asymptomatic; mut: MYBPC3: NM\_000256.3:c.3330+2T>G, wt: MYBPC3 wild-type NM\_000256.3:c.3330+2T.

Figure 2: Distribution of the MYBPC3 NM\_000256.3:c.3330+2T>G carriers in Switzerland and migration of the variant in Europe.

The figures indicate the distribution of the eight MYBPC3 NM\_000256.3:c.3330+2T>G carrier families in Swiss cantons (\* denotes one additional patient from the Geneva canton listed in ClinVar). All patients originate from Central/Western Switzerland, which is from where the Old14 Order Amish and Old-Order Mennonites populations (greyed areas) emigrated 314 to Palatinates, Strasbourg region but also the United States (Pennsylvania and Ohio in particular, where >40 mutation carriers have been identified).



**Figure 2.**



TABLES TITLES AND LEGENDS

Table 1: Clinical information for all herein reported individuals carrying the MYBPC3 NM\_000256.3:c.3330+2T>G variant.

Table 2: Haplotypes associated with the MYBPC3 NM\_000256.3:c.3330+2T>G variant across patients.

**Table 1**

	<i>Fam 1</i>	<i>Fam 2</i>	<i>Fam. 3</i>		<i>Fam. 4</i>	<i>Fam. 5</i>	<i>Fam. 6</i>
	<i>III.2</i>	<i>III.3</i>	<i>IV.1</i>	<i>IV.2</i>	<i>III.1</i>	<i>III.3</i>	<i>III.1</i>
<i>Age</i>	49	65	18	15	65	55	76
<i>Sex</i>	M	M	F	M	M	M	M
<i>Age at diagnosis</i>	41	63	1.5	6	45	50	42
<i>First symptoms</i>	Dyspnea on exertion	Dyspnea upon awakening	Dyspnea on exertion	<u>Nihil?</u>	<u>Syncope</u>	Dizziness on exertion	Dyspnea on exertion
<i>Diagnosis</i>	NO-HCM	NO-HCM	NO-HCM	NO-HCM	NO-HCM	NO-HCM	NO-HCM
<i>LV WT (mm)</i>	16	18	28	14	18	19	27
<i>LV EF</i>	34	58	Lower bound	60	75	66	65
<i>LVOT &gt;30 mmHg</i>	No	No	Yes	No	No	No	No
<i>Syncope</i>	Yes	No	No	No	Yes	No	No
<i>Arrhythmia</i>	Yes	Yes	No	No	No	No	Yes
<i>Genetic test</i>	CMH panel	CMH panel	CMH panel	Sanger	WGS	WGS	WGS

LV WT: left ventricular wall thickness; LV EF: left ventricular ejection fraction; LVOT: left ventricular outflow tract gradient; NO-HCM: non-obstructive hypertrophic cardiomyopathy; WGS: whole genome sequencing

## REFERENCES

- 1 Semsarian, C, Ingles, J, Maron, MS, Maron, BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol.* 2015; 65:1249-1254.
- 2 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. *Genetics in Medicine.* 2015; 17:880-888.
- 3 Ingles, J, Goldstein, J, Thaxton, C, Caleshu, C, Corty, EW, Crowley, SB et al. Evaluating the Clinical Validity of Hypertrophic Cardiomyopathy Genes. *Circ-Genom Precis Me.* 2019; 12.
- 4 Walsh, R, Thomson, KL, Ware, JS, Funke, BH, Woodley, J, McGuire, KJ et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genetics in Medicine.* 2017; 19:192-203.
- 5 Christiaans, I, Birnie, E, van Langen, IM, van Spaendonck-Zwarts, KY, van Tintelen, JP, van den Berg, MP et al. The yield of risk stratification for sudden cardiac death in hypertrophic cardiomyopathy myosin-binding protein C gene mutation carriers: focus on predictive screening. *Eur Heart J.* 2010; 31:842-848.
- 6 Niimura, H, Patton, KK, McKenna, WJ, Soultis, J, Maron, BJ, Seidman, JG et al. Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation.* 2002; 105:446-451.

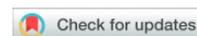
- 7 Carrier, L. Targeting the population for gene therapy with MYBPC3. *J Mol Cell Cardiol.* 2021; 150:101-108.
- 8 Deprez, RHL, Muurling-Vlietman, JJ, Hruda, J, Baars, MJH, Wijnaendts, LCD, Stolte- Dijkstra, I et al. Two cases of severe neonatal hypertrophic cardiomyopathy caused by compound heterozygous mutations in the MYBPC3 gene. *Journal of Medical Genetics.* 2006; 43:829-832.
- 9 Xin, B, Puffenberger, E, Tumbush, J, Bockoven, JR, Wang, H. Homozygosity for a novel splice site mutation in the cardiac myosin-binding protein c gene causes severe neonatal hypertrophic cardiomyopathy. *Am J Med Genet A.* 2007; 143a:2662-2667.
- 10 Zahka, K, Kalidas, K, Simpson, MA, Cross, H, Keller, BB, Galambos, C et al. Homozygous mutation of MYBPC3 associated with severe infantile hypertrophic cardiomyopathy at high frequency among the Amish. *Heart.* 2008; 94:1326-1330.
- 11 Ommen, SR, Mital, S, Burke, MA, Day, SM, Deswal, A, Elliott, P 381 et al. 2020 AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy: Executive Summary A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation.* 2020; 142:e533-e557.
- 12 Authors/Task Force, m, Elliott, PM, Anastasakis, A, Borger, MA, Borggrefe, M, Cecchi, F 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; 35:2733-2779.
- 13 Royer-Bertrand, B, Castillo-Taucher, S, Moreno-Salinas, R, Cho, TJ, Chae, JH, Choi, M et al. Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci Rep.* 2015; 5:17154.
- 14 Van der Auwera, GA, Carneiro, MO, Hartl, C, Poplin, R, Del Angel, G, Levy-Moonshine, A et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013; 43:11 10 11-11 10 33.

- 15 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint. 2013.
- 16 Poplin, R, Ruano-Rubio, V, DePristo, MA, Fennell, TJ, Carneiro, MO, Van der Auwera, GA et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018:201178.
- 17 Andrew, S. FastQC: a quality control tool for high throughput sequence data. Online tool. 2010.
- 18 Quinlan, AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014; 47:11 12 11-34.
- 19 Pedersen, BS, Quinlan, AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018; 34:867-868.
- 20 Pedersen, BS, Quinlan, AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. *Am J Hum Genet*. 2017; 100:406-413.
- 21 Wang, K, Li, MY, Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010; 38.
- 22 Pedersen, BS, Layer, RM, Quinlan, AR. Vcfanno: fast, flexible annotation of genetic variants. *Genome Biol*. 2016; 17:118.
- 23 Patterson, M, Marschall, T, Pisanti, N, van Iersel, L, Stougie, L, Klau, GW et al. WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J Comput Biol*. 2015; 22:498-509.
- 24 Delaneau, O, Zagury, JF, Robinson, MR, Marchini, JL, Dermitzakis, ET. Accurate, scalable and integrative haplotype estimation. *Nat Commun*. 2019; 10:5436.
- 25 Budde, BS, Namavar, Y, Barth, PG, Poll-The, BT, Nurnberg, G, Becker, C et al. tRNA splicing endonuclease mutations cause pontocerebellar hypoplasia. *Nat Genet*. 2008; 40:1113-1118.
- 26 De, S, Borowski, AG, Wang, H, Nye, L, Xin, BZ, Thomas, JD et al. Subclinical echocardiographic abnormalities in phenotype-negative carriers of myosin-binding protein C3 gene mutation for hypertrophic cardiomyopathy. *Am Heart J*. 2011; 162:262-U282.

- 27 McKusick, VA, Hostetler, JA, Egeland, JA. Genetic Studies of the Amish, Background and Potentialities. *Bull Johns Hopkins Hosp.* 1964; 115:203-222.
- 28 Greksa, LP. Population growth and fertility patterns in an Old Order 428 Amish settlement. *Ann Hum Biol.* 2002; 29:192-201.
- 29 Puffenberger, EG. Genetic heritage of the Old Order Mennonites of southeastern Pennsylvania. *Am J Med Genet C Semin Med Genet.* 2003; 121C:18-31.
- 30 Guss, J. *Persecution, Division, and Opportunity: The Origins of the Old Order Amish.* 2007.
- 31 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; 46:1737-1743.
- 32 Jaaskelainen, P, Miettinen, R, Karkkainen, P, Toivonen, L, Laakso, M, Kuusisto, J. Genetics of hypertrophic cardiomyopathy in eastern Finland: few founder mutations with benign or intermediary phenotypes. *Ann Med.* 2004; 36:23-32.
- 33 van Velzen, HG, Schinkel, AFL, Oldenburg, RA, van Slegtenhorst, MA, Frohn-Mulder, IME, van der Velden, J et al. Clinical Characteristics and Long-Term Outcome of Hypertrophic Cardiomyopathy in Individuals With a MYBPC3 (Myosin-Binding Protein C) Founder Mutation. *Circ Cardiovasc Genet.* 2017; 10.
- 34 Sabater-Molina, M, Saura, D, Garcia-Molina Saez, E, Gonzalez-Carrillo, J, Polo, L, Perez-Sanchez, I et al. A Novel Founder Mutation in MYBPC3: Phenotypic Comparison With the Most Prevalent MYBPC3 Mutation in Spain. *Rev Esp Cardiol (Engl Ed).* 2017; 70:105-114.
- 35 Calore, C, De Bortoli, M, Romualdi, C, Lorenzon, A, Angelini, A, Basso, C et al. A founder MYBPC3 mutation results in HCM with a high risk of sudden death after the fourth decade of life. *J Med Genet.* 2015; 52:338-347.

- 36 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; 24:1848-1853.
- 37 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; 130:1158-1167.
- 38 Adalsteinsdottir, B, Burke, M, Maron, BJ, Danielsen, R, Lopez, B, Diez, J et al. Hypertrophic cardiomyopathy in myosin-binding protein C (MYBPC3) Icelandic founder mutation carriers. *Open Heart.* 2020; 7:e001220.

# SCN5A overlap syndromes: An open-minded approach



Alessandra P. Porretta, MD,<sup>\*†</sup> Vincent Probst, MD,<sup>‡</sup> Zahurul A. Bhuiyan, MD, PhD,<sup>§</sup>  
Emeline Davoine, MSc, CGC,<sup>§</sup> Antoine Delinière, MD,<sup>¶</sup> Patrizio Pascale, MD,<sup>\*</sup>  
Juerg Schlaepfer, MD,<sup>\*</sup> Andrea Superti-Furga, MD,<sup>§||</sup> Etienne Pruvot, MD<sup>\*</sup>

From the <sup>\*</sup>Service of Cardiology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland, <sup>†</sup>Department of Clinical-Surgical Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy, <sup>‡</sup>l'Institut du thorax, Service de Cardiologie du CHU de Nantes, Hôpital Nord, Nantes Cedex, France, <sup>§</sup>Division of Genetic Medicine, Lausanne University Hospital, Lausanne, Switzerland, <sup>¶</sup>Service of Rhythmology, Hôpital Cardiologique Louis Pradel, Hospices Civils de Lyon, Lyon, France, and <sup>||</sup>University of Lausanne, Lausanne, Switzerland.

SCN5A overlap syndromes are clinical entities that express a phenotype combining aspects of different canonical SCN5A-related arrhythmia syndromes or a variable arrhythmic phenotype among individuals carrying the same SCN5A mutation. Here we review the literature addressing SCN5A overlap syndromes as well as the principal mechanisms currently proposed. Among others, a multifactorial determination encompassing an interaction between SCN5A

variant(s), other genetic polymorphisms, and possibly environmental factors seems the most plausible hypothesis.

**KEYWORDS** Cardiogenetics; Channelopathies; Polygenic inheritance; SCN5A gene; SCN5A overlap syndrome

(Heart Rhythm 2022;19:1363–1368) © 2022 Heart Rhythm Society. All rights reserved.

## Introduction

The SCN5A gene, located in the short arm of chromosome 3 (3p21), codes for the  $\alpha$ -subunit of the cardiac sodium channel (Na<sub>v</sub>1.5), which contains 4 homologous domains (each consisting of 6 transmembrane segments) that, by interacting with the ancillary  $\beta$ -subunit and with specific regulatory proteins, mediates the fast upstroke (phase 0) of the cardiac action potential.<sup>1–4</sup>

Genetic variants in the SCN5A gene have been implicated in several cardiac diseases, including congenital long QT syndrome, Brugada syndrome (BS), sinus nodal dysfunction (SND), progressive familial cardiac conduction disorder, isolated atrial fibrillation, multifocal ectopic Purkinje-related premature contractions, dilated cardiomyopathy, and arrhythmogenic right ventricular cardiomyopathy. In addition, patients with SCN5A-related SND and atrial arrhythmias may show poor pacemaker capture (Figure 1).<sup>2–7</sup>

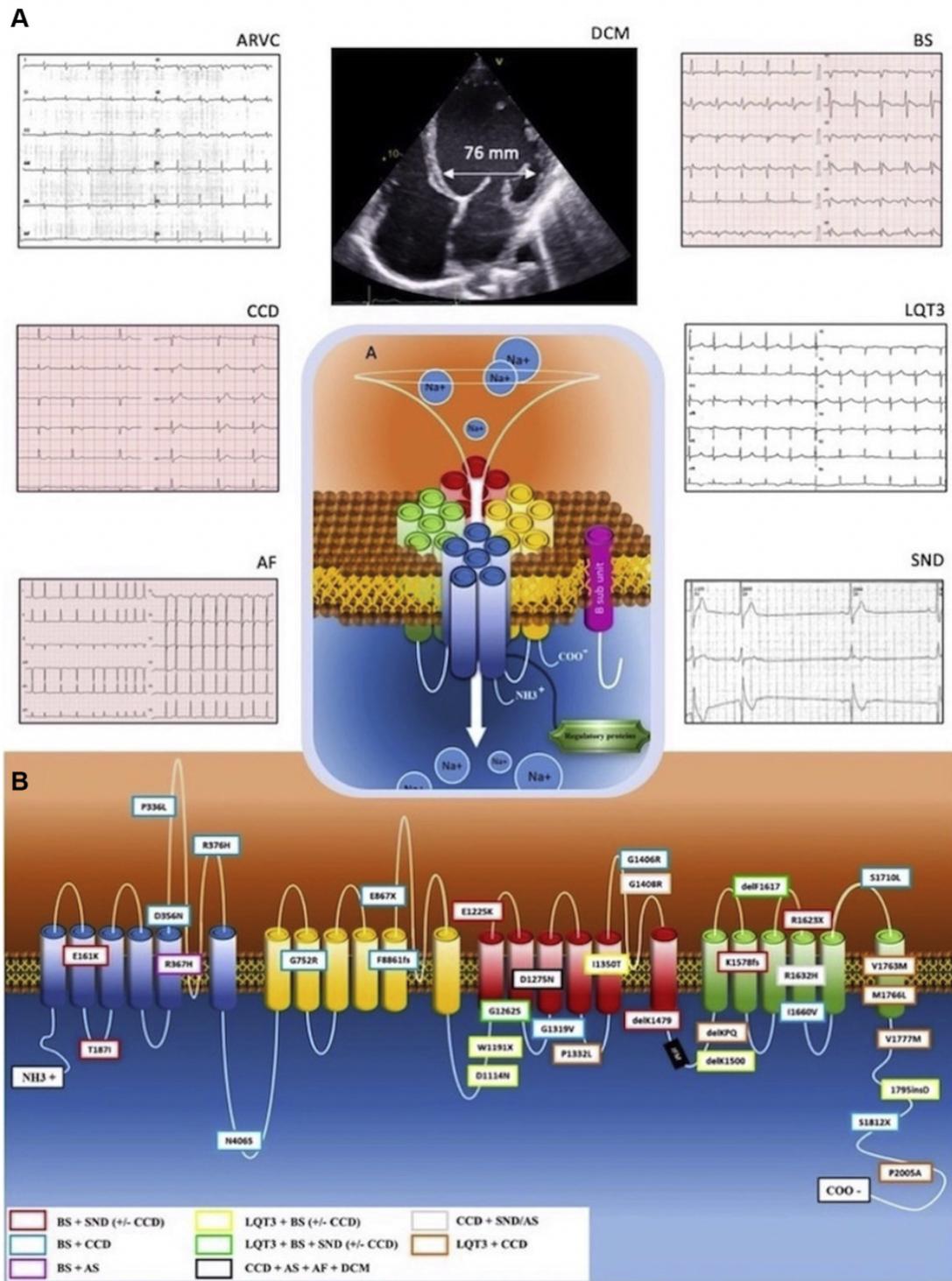
Originally, SCN5A-related diseases were considered separate clinical entities, characterized by distinctive and isolated manifestations resulting from specific biophysical features induced by the SCN5A mutation on the voltage-gated sodium channel.<sup>3,4</sup> However, many studies have

described individuals with SCN5A variants who do not have these “canonical” entities but rather overlap syndromes expressing a mixed phenotype between different SCN5A-related arrhythmic syndromes. In addition, SCN5A phenotypes may not breed true within families, as individuals carrying the same variant may have different phenotypes.<sup>2</sup> Of note, such clinical overlap may induce the simultaneous onset, in the same patient, of atrial and ventricular arrhythmias. Among others, Ghovanloo et al<sup>8</sup> recently described a novel SCN5A mutation associated with atrial tachyarrhythmias and polymorphic ventricular ectopy occurring at rest and under adrenergic stimulation, coupled with a heavy family history of sudden cardiac death. This particular aspect highlights a further level of clinical complexity that makes prescription of antiarrhythmic drugs difficult.

## SCN5A overlap syndromes: The evolution of a concept

The first published description dates back to 1999, when Bezina et al<sup>3,9</sup> identified a SCN5A-1795insD mutation that was associated with an extreme phenotypic variability characterized by the expression of long QT syndrome type 3 (LQT3), BS, conduction disorders, and/or SND of variable severity, either isolated or in association within the same family. Subsequently, several other papers reported single SCN5A mutations, with a clinical phenotype combining BS and conduction disorders.<sup>3,10–15</sup> Further studies then described overlap syndromes involving BS in association with SND,<sup>3,13,15–17</sup> atrial standstill,<sup>18</sup> or LQT3.<sup>19</sup> Remme et al<sup>3</sup>

**Funding Sources:** The authors have no funding sources to disclose. **Disclosures:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. **Address reprint requests and correspondence:** Dr Alessandra Pia Porretta, Service of Cardiology, Heart and Vessel Department, Centre Hospitalier Universitaire Vaudois (CHUV), Rue du Bugnon 46, 1011 Lausanne, Switzerland. E-mail address: [Alessandra-pia.porretta@chuv.ch](mailto:Alessandra-pia.porretta@chuv.ch).



**Figure 1** A: Representation of the cardiac sodium channel made of the pore-forming  $\alpha$ -subunit ( $Na_v1.5$ ) interacting with regulatory proteins and  $\beta$ -subunits. The main clinical phenotypes associated with *SCN5A* pathogenic variants also are displayed. B: Schematic representation of the location of the main *SCN5A* pathogenic variants causally associated with overlap syndromes. AF = atrial fibrillation; ARVC = arrhythmogenic right ventricular cardiomyopathy; AS = atrial standstill; BS = Brugada syndrome; CCD = cardiac conduction disorder; DCM = dilated cardiomyopathy; LQT3 = long QT syndrome type 3; SND = sinus nodal dysfunction.

summarized published data, providing an extensive list of *SCN5A* mutations associated with overlap syndromes. What is the mechanistic basis of such variability? Several mechanisms have been proposed to explain the variable expression of *SCN5A* syndromes, and different clinical factors have been reported as modulating phenotypic expression.

## Clinical modifiers of phenotypic expression

### Age

Beaufort-Krol et al<sup>3,20</sup> demonstrated that the 1795insD mutation is associated with an age-dependent penetrance of electrocardiographic (ECG) alterations as the detection of QT prolongation and conduction disorders occurred from birth, whereas that of Brugada pattern occurred only from the age of 5 years. However, this observation remains unexplained to date. It may involve an age-dependent increase in the transient outward potassium repolarization current ( $I_{to}$ ) at the epicardial level, which, in association with the *SCN5A* mutation-related reduction in the depolarizing sodium current, could lead to epicardial loss of the plateau phase of the action potential. This may consequently lead to the age-dependent development of Brugada pattern.<sup>20</sup> An age-dependent reduction of the sodium current might be similarly involved.<sup>20</sup>

### Gender

Gender might represent a second pivotal phenotypic modifier. For the G1406R mutation, Kyndt et al<sup>10</sup> detected a possible gender-dependent expression. Although the Brugada phenotype was more frequently detected among males, isolated conduction disorders more frequently affected females. Aizawa et al<sup>21</sup> pointed out a sex-related phenotypic distribution among several familial BS patients carrying *SCN5A* mutations, with only female members developing SND and male patients carrying a higher susceptibility to sudden cardiac death.

Of note, regardless of the presence of an overlapping phenotype, several studies have already postulated a gender difference affecting the clinical expression of BS. Benito et al<sup>22</sup> demonstrated that male gender presents more frequently with a type I ECG pattern and that, among all patients with the spontaneous type I pattern, ST-segment elevation is definitively more pronounced in males than in females. From a clinical point of view, males seem to develop a significantly higher number of cardiac events during follow-up, with a higher rate of ventricular fibrillation than women.

From a pathophysiological point of view, several hypotheses have been proposed to explain the ionic/cellular mechanism of such a gender-dependent clinical variability. Di Diego et al<sup>23</sup> suggested the existence of specific gender-dependent differences affecting the ionic currents, with significantly greater density of the transient outward current ( $I_{to}$ ) in the male ventricular epicardium. A second hypothesis pertains to the effect of sex hormones. Shimizu et al<sup>24</sup> demonstrated that higher testosterone levels associated with lower

visceral fat may contribute to male predominance in BS, perhaps because of the effect of testosterone on active currents during phase 1 of the action potential. Bai et al<sup>22,25</sup> observed that testosterone might reduce the duration of the action potential by potentiating the slow component of the delayed rectifier potassium current ( $I_{Ks}$ ). Similarly, the down-regulation of repolarizing potassium channels might explain the regression of the typical ECG pattern in men with BS following castration.<sup>22,26</sup>

### Other factors

Other factors, such as temperature or medications and drugs, have been linked to phenotypic variability.

Keller et al<sup>3,27</sup> demonstrated temperature-induced shortening of the action potential duration through modification of the sodium channel gating properties among carriers of *SCN5A*-L325R and R535X mutations.

Other studies have illustrated the potential “rescue” of misprocessed channels by pharmacological sodium channel blockers such as mexiletine.<sup>3</sup> Pathogenic *SCN5A* variants may generate their effect through different functional mechanisms, including either changes in kinetics and channel gating properties or defects in channel synthesis and trafficking to the cell surface membrane. Valdivia et al<sup>28</sup> demonstrated that mexiletine and quinidine both were able to rescue defective channel trafficking associated with specific *SCN5A* variants. That is, by acting as pharmacological chaperones, these drugs may restore the correct localization of channels to the cell surface membrane, suggesting some future potential mutation-specific therapeutic options.<sup>3,29</sup>

## Genetic modifiers of phenotypic expression

In an attempt to explain the incomplete penetrance with variable expression typical of BS, recent studies questioned the monogenic model of mendelian inheritance historically proposed. A more complex inheritance model, possibly polygenic, has been suggested, encompassing an interaction among the *SCN5A* variants, the so-called genetic background, and the environment.

### Biallelic and compound variants

Bezzina et al<sup>3,29</sup> have shown that the simultaneous inheritance of more than one *SCN5A* variants (“compound mutations”) might cause a more severe phenotype compared to that resulting from an isolated mutation.

### Heterogeneous biophysical properties

According to recent evidence, specific *SCN5A* mutations may account for heterogeneous biophysical properties, that is, they determine at the same time “loss-of-function” and “gain-of-function” effects.<sup>2,3</sup> Benson et al<sup>30</sup> observed a “constellation of biophysical defects” associated with the delF1617 mutation, which, in both LQT3 and BS, causes a delay in the inactivation of the mutant channel, with consequent increase of the inward sodium current (gain of function). In contrast, the same delF1617 variant induces a

disease inheritance.<sup>39,41</sup> Several other functional SNPs have been described that, by impacting cardiac conduction, contribute to the genetic background modulating the mutation-related expressivity in *SCN5A* overlap syndromes.

### Emerging evidence

Several recent studies have disrupted the traditional paradigm of functional independence of ion channels.<sup>42</sup> A complex interplay among voltage-gated ion channel subunits seems to lead to reciprocal functional regulation and modulation of the global biophysical phenotype.<sup>42</sup>

### Homotypic protein interactions

Clatot et al<sup>43</sup> provided new insights into the structure and function of the sodium channel by suggesting that voltage-gated sodium  $\alpha$ -subunits do not act as single monomers but reciprocally interact. Indeed,  $\alpha$ -subunits “assemble, function and gate” as a 14-3-3 protein-mediated dimer.<sup>44</sup> In the context of such deep interactions between  $\text{Na}_v1.5$   $\alpha$ -subunits, Poelzing et al<sup>38</sup> showed that coexpression of the *H558R-SCN5A* polymorphism (on a separate gene) with the BS mutation *R282H-SCN5A* was able to normalize electrophysiological properties by interacting with the mutant subunits and by restoring their membrane trafficking.

### $\text{Na}_v1.5$ heterotypic protein interactions

Other groups have pointed out the complex interplay between  $\text{Na}_v1.5$  and different ion channel subunits in the context of macromolecular complexes called “channelosomes.”<sup>42</sup>

#### *Na<sub>v</sub>1.5 and Kir2.1-2.2 channel interaction*

An iconic example is represented by the synergistic cooperation, mediated by  $\alpha 1$ -syntrophin, between  $\text{Na}_v1.5$  and the Kir2.1-2.2 channels, the latter conveying the inward potassium rectifier current ( $I_{K1}$ ) during the late phase of repolarization.  $\text{Na}_v1.5$  allows for the positive modulation of  $I_{Kir2.1}$  and  $I_{Kir2.2}$  through a chaperone-like effect, which increases channel densities on the cell membrane.<sup>45</sup> However, Pérez-Hernández et al<sup>46</sup> found that such interplay requires normal trafficking of  $\text{Na}_v1.5$  channels toward the cell membrane. For this reason,  $\text{Na}_v1.5$  pathogenic variants leading to trafficking defects do not positively modulate  $I_{Kir2.1}$  and  $I_{Kir2.2}$  but cause a functional impairment of wild-type Kir2.1-2.2 channels. Such mechanisms explain the so-called “dominant negative effect” of certain *SCN5A* variants on normal Kir2.1-2.2 channels leading to  $I_{K1}$  reduction while predisposing to a higher risk of ventricular arrhythmias (due to the concomitant  $I_{Na}$  and  $I_{K1}$  impairment).<sup>46</sup>

#### *Na<sub>v</sub>1.5 and K<sub>v</sub>4.3 channel interaction*

Another substantial body of evidences concerns the interplay between  $\text{Na}_v1.5$  and the  $\text{K}_v4.3$  channel, which conveys the transient outward potassium current ( $I_{to}$ ) during phase 1 of the action potential. Portero et al<sup>47</sup> and Clatot et al<sup>42</sup> showed that, due to their reciprocal gating and trafficking regulation,

gain-of-function  $\text{K}_v4.3$  pathogenic variants (leading to  $I_{to}$  increase) are responsible for a concomitant impairment of the sodium depolarizing current ( $I_{Na}$ ). This observation may thus explain the development of a BS phenotype in patients with mutated gain-of-function  $\text{K}_v4.3$  channels.

Thus, such homo- and heterotypic protein interactions add a further level of complexity in determining the phenotypic expression of single gene variants. Akin to the “polygenic risk score” approach developed for conditions such as breast cancer and coronary disease, it should be possible to develop a “genomic” approach to cardiac channelopathies that integrates dominant single-gene variants in *SCN5A* to genetic variations in functionally related genes to explain the phenotypic variability and the spectrum of overlapping phenotypes.

### Conclusion

*SCN5A* overlap syndromes represent a novel clinical entity expressing either a mixed phenotype (including aspects of various *SCN5A*-related arrhythmia syndromes) or a variable arrhythmic phenotype between individuals carrying the same *SCN5A* mutation.<sup>2</sup> Since the first description in 1999 by Bezzina et al,<sup>3,9</sup> a growing number of studies and case reports have been published describing similar examples of phenotypic variability associated with an ever-increasing number of *SCN5A* mutations. Here we reviewed the literature addressing *SCN5A* overlap syndromes as well as the mechanisms contributing to variable expression. Different clinical factors have been advocated as modulating phenotypic expressions, such as age or gender, as well as exogenous factors including temperature and specific drugs. Recent studies have called into question the monogenic model of mendelian inheritance and have suggested a more complex inheritance model, possibly polygenic, encompassing a multifactorial interaction between environment and genetic background. Further studies are needed to further clarify phenotypic modifiers and their pathways to support diagnosis, risk stratification, and therapeutic management of *SCN5A* overlap syndromes.

### References

1. Wilde AA, Brugada R. Phenotypical manifestations of mutations in the genes encoding subunits of the cardiac sodium channel. *Circ Res* 2011;108:884–897.
2. Robyns T, Nuyens D, Van Casteren L, et al. Reduced penetrance and variable expression of *SCN5A* mutations and the importance of co-inherited genetic variants: case report and review of the literature. *Indian Pacing Electrophysiol J* 2014;14:133–149.
3. Remme CA, Wilde AA, Bezzina CR. Cardiac sodium channel overlap syndromes: different faces of *SCN5A* mutations. *Trends Cardiovasc Med* 2008;18:78–87.
4. Remme CA, Wilde AA. *SCN5A* overlap syndromes: no end to disease complexity? *Europace* 2008;10:1253–1255.
5. Chiang DY, Kim JJ, Valdes SO, et al. Loss-of-function *SCN5A* mutations associated with sinus node dysfunction, atrial arrhythmias, and poor pacemaker capture. *Circ Arrhythm Electrophysiol* 2015;8:1105–1112.
6. Laurent G, Saal S, Amarouch MY, et al. Multifocal ectopic Purkinje-related premature contractions: a new *SCN5A*-related cardiac channelopathy. *J Am Coll Cardiol* 2012;60:144–156.
7. Ter Bekke RMA, David M, Krapels IPC, Crijns HJGM, Volders PGA. Beauty and the beat: a complicated case of multifocal ectopic Purkinje-related premature contractions. *HeartRhythm Case Rep* 2018;4:429–433.

reduction of the peak current density, as well as impaired recovery from the inactivation phase with consequent mutant channel loss of function.<sup>4,30</sup>

Another iconic example is represented by the dual biophysical properties associated with the 1795insD mutation. Among 1795insD mutation carriers, QT-interval prolongation has been observed during slow heart rate, whereas distinctive ST-segment elevation characteristic of BS has been demonstrated during rapid heart rate (exercise).<sup>31</sup> Veldkamp et al<sup>31</sup> showed that, during slow heart rate, 1795insD mutation disrupts fast inactivation and leads to persistent Na<sup>+</sup> current throughout the action potential plateau phase, prolonging cardiac repolarization and the QT interval. In contrast, during rapid heart rate (in the same patient), the mutation accelerates the inactivation of the I<sub>Na</sub> current and delays the recovery of Na<sup>+</sup> channel availability between stimuli, leading to Na<sup>+</sup> current reduction (which is the pathologic basis of BS).<sup>31</sup>

## Splicing variants

### Alternative splicing

Alternative splicing is the phenomenon by which a primary RNA transcript can, through variable cutting and ligation phases, result in mature mRNA molecules that differ in respect to the presence of certain exons.<sup>2,3</sup> Tan et al<sup>32</sup> reported a correlation between the expression level of a given mutant *SCN5A* splice variant and the degree of dysfunction of the mutant sodium channel. The *SCN5A* gene encodes 2 main splice variant transcripts characterized by the absence and presence of a glutamine at position 1077 (Q1077del and Q1077), representing approximately 65% and 35% of all *SCN5A* transcripts, respectively.<sup>3</sup> The authors observed a significantly higher degree of sodium channel dysfunction when the BS-causal G1406R mutation is associated with the Q1077 spliced variant. As postulated by Remme et al,<sup>3</sup> the clinical phenotype (associated with the same *SCN5A* mutation) might therefore vary according to the variability in the ratio between the 2 splice variants.

### Structural variations: Copy number variations

Splicing variants may be the result of copy number variations (CNVs), structural variations characterized by duplications or deletions of a significant number of base pairs, which may affect gene expression and regulation. Although single nucleotide variants and small indels are the most frequent genetic rearrangement described in BS,<sup>33</sup> in 2011 Eastaugh et al<sup>34</sup> detected for the first time a CNV implying a large deletion (exons 9 and 10) of *SCN5A* responsible for haploinsufficiency, because no functional protein was predicted to appear in the membrane. Since then, a few reports have investigated the role of CNVs in *SCN5A*-related arrhythmic syndromes. More recently, Sonoda et al<sup>35</sup> reported the case of 4 unrelated probands carrying *SCN5A*-CNVs that determined severe haploinsufficiency. In particular, 1 proband expressed an overlap syndrome including sick sinus syndrome, symptomatic BS

(with several episodes of ventricular fibrillation requiring implantable cardioverter-defibrillator insertion), and cardiac conduction defects. Of note, his *SCN5A*-CNV was similarly detected in his sister, brother, and mother, who all were asymptomatic (first-degree atrioventricular block was described in the mother and the sister). Moreover, in the study by Sonoda et al,<sup>35</sup> clinical and ECG features of the probands with CNVs were similar to those of the probands with *SCN5A* truncating mutations as well as to the BS patients harboring missense mutations causing peak I<sub>Na</sub> reduction >90%.<sup>35,36</sup> Additionally, episodes of ventricular fibrillation or cardiopulmonary arrest were reported more in the *SCN5A* mutation-positive (missense, truncating, CNV) BS patients than in the *SCN5A*-negative patients.<sup>35</sup>

## Coinheritance of other genetic variants

Recently, interest has also focused on the coinheritance of rare or common genetic variants either in *SCN5A* or in other relevant genes that might modulate the phenotype.

### Genetic variant definition

A genetic variant can be defined as any modification of the standard genetic sequence triggered by any of several mechanisms, including structural (inversions, translocations, splice variants), epigenetic (methylations, prions), and/or DNA sequence variations. According to detection frequency in the general population, genetic variants can be classified into polymorphisms (with an allelic frequency in the general population >1%) and rare variants (allelic frequency <1%).<sup>2</sup>

### Role of concurrent polymorphisms

H558R represents the single nucleotide polymorphism (SNP) most frequently detected in *SCN5A* in the general population (between 9% and 36%, depending on the ethnic group<sup>3,37</sup>). Several authors demonstrated that, independently from H558R polymorphism location (in cis or trans with other *SCN5A* variants), its interaction with specific *SCN5A* variants can exert a “rescue” action, which attenuates phenotypic severity by improving the function of the mutant sodium channel.<sup>3,38,39</sup>

Furthermore, Bezzina et al<sup>40</sup> identified in a genome-wide association study 3 SNPs significantly associated with BS, the simultaneous coexpression of which might explain up to 7% of the variance in disease susceptibility through a cumulative effect. The first SNP (rs11708996) is located on *SCN5A* and affects cardiac conduction by modulating PR interval and QRS duration variability. The second SNP (rs10428132) is located on the *SCN10A* gene, which encodes the Na<sub>v</sub>1.8 sodium channel isoform expressed in the myocardium as well as in the specialized conduction system. The third SNP (rs9388451) is located on the *HEY2* gene and codes for a transcriptional regulator involved in cardiac electrical function. They demonstrated a cumulative effect of the 3 identified loci as the risk of disease expression increased significantly with the increasing number of co-inherited risk alleles according to a typically multifactorial model of

8. Ghovanloo MR, Atallah J, Escudero CA, Ruben PC. Biophysical characterization of a novel SCN5A mutation associated with an atypical phenotype of atrial and ventricular arrhythmias and sudden death. *Front Physiol* 2020;11:610436.
9. Bezzina CI, Veldkamp MW, van Den Berg MP, et al. A single Na(+) channel mutation causing both long-QT and Brugada syndromes. *Circ Res* 1999; 85:1206–1213.
10. Kyndt F, Probst V, Potet F, et al. Novel SCN5A mutation leading either to isolated cardiac conduction defect or Brugada syndrome in a large French family. *Circulation* 2001;104:3081–3086.
11. Six I, Hermida JS, Huang H, et al. The occurrence of Brugada syndrome and isolated cardiac conductive disease in the same family could be due to a single SCN5A mutation or to the accidental association of both diseases. *Europace* 2008;10:79–85.
12. Schulze-Bahr E, Eckardt L, Breithardt G, et al. Sodium channel gene (SCN5A) mutations in 44 index patients with Brugada syndrome: different incidences in familial and sporadic disease. *Hum Mutat* 2003;21:651–652. Erratum in: *Hum Mutat* 2005;26:61.
13. Smits JP, Koopmann TT, Wilders R, et al. A mutation in the human cardiac sodium channel (E161K) contributes to sick sinus syndrome, conduction disease and Brugada syndrome in two families. *J Mol Cell Cardiol* 2005;38:969–981.
14. Rossenbacker T, Carroll SJ, Liu H, et al. Novel pore mutation in SCN5A manifests as a spectrum of phenotypes ranging from atrial flutter, conduction disease, and Brugada syndrome to sudden cardiac death. *Heart Rhythm* 2004;1:610–615.
15. Nakajima S, Makiyama T, Hanazawa K, et al. A novel SCN5A mutation demonstrating a variety of clinical phenotypes in familial sick sinus syndrome. *Intern Med* 2013;52:1805–1808.
16. Hayashi H, Sumiyoshi M, Nakazato Y, Daida H. Brugada syndrome and sinus node dysfunction. *J Arrhythm* 2018;34:216–221.
17. Makiyama T, Akao M, Tsuji K, et al. High risk for bradyarrhythmic complications in patients with Brugada syndrome caused by SCN5A gene mutations. *J Am Coll Cardiol* 2005;46:2100–2106.
18. Takehara NI, Makita N, Kawabe J, et al. A cardiac sodium channel mutation identified in Brugada syndrome associated with atrial standstill. *J Intern Med* 2004; 255:137–142.
19. Priori SG, Napolitano C, Gasparini M, et al. Clinical and genetic heterogeneity of right bundle branch block and ST-segment elevation syndrome: a prospective evaluation of 52 families. *Circulation* 2000;102:2509–2515.
20. Beaufort-Krol GC1, van den Berg MP, Wilde AA, et al. Developmental aspects of long QT syndrome type 3 and Brugada syndrome on the basis of a single SCN5A mutation in childhood. *J Am Coll Cardiol* 2005;46:331–337.
21. Aizawa Y, Fujisawa T, Katsumata Y, et al. Sex-dependent phenotypic variability of an SCN5A mutation: Brugada syndrome and sick sinus syndrome. *J Am Heart Assoc* 2018;7:e0009378.
22. Benito B, Sarkozy A, Mont L, et al. Gender differences in clinical manifestations of Brugada syndrome. *J Am Coll Cardiol* 2008;52:1567–1573.
23. Di Diego JM, Cordeiro JM, Goodrow RJ, et al. Ionic and cellular basis for the predominance of the Brugada syndrome phenotype in males. *Circulation* 2002; 106:2004–2011.
24. Shimizu W, Matsuo K, Kokubo Y, et al. Sex hormone and gender difference—role of testosterone on male predominance in Brugada syndrome. *J Cardiovasc Electrophysiol* 2007;18:415–421.
25. Bai CX, Kurokawa J, Tamagawa M, Nakaya H, Furukawa T. Nontranscriptional regulation of cardiac repolarization currents by testosterone. *Circulation* 2005; 112:1701–1710.
26. Matsuo K, Akahoshi M, Seto S, Yano K. Disappearance of the Brugada-type electrocardiogram after surgical castration: a role for testosterone and an explanation for the male preponderance. *Pacing Clin Electrophysiol* 2003;26(7 Pt 1):1551–1553.
27. Keller DI, Rougier JS, Kucera JP, et al. Brugada syndrome and fever: genetic and molecular characterization of patients carrying SCN5A mutations. *Cardiovasc Res* 2005;67:510–519.
28. Valdivia CR, Tester DJ, Rok BA, et al. A trafficking defective, Brugada syndrome-causing SCN5A mutation rescued by drugs. *Cardiovasc Res* 2004; 62:53–62.
29. Bezzina CR, Rook MB, Groenewegen WA, et al. Compound heterozygosity for mutations (W156X and R225W) in SCN5A associated with severe cardiac conduction disturbances and degenerative changes in the conduction system. *Circ Res* 2003;92:159–168.
30. Benson DW, Wang DW, Dymment M, et al. Congenital sick sinus syndrome caused by recessive mutations in the cardiac sodium channel gene (SCN5A). *J Clin Invest* 2003;112:1019–1028.
31. Veldkamp MW, Viswanathan PC, Bezzina C, Baartscheer A, Wilde AA, Balser JR. Two distinct congenital arrhythmias evoked by a multidysfunctional Na(+) channel. *Circ Res* 2000;86:E91–E97.
32. Tan B-H, Valdivia CR, Song C, Makielski JC. Partial expression defect for the SCN5A missense mutation G1406R depends on splice variant background Q1077 and rescue by mexiletine. *Am J Physiol Heart Circ Physiol* 2006; 291:H1822–H1828.
33. Mademont-Soler I, Pinsach-Abuin ML, Riuro H, et al. Large genomic imbalances in Brugada syndrome. *PLoS One* 2016;11:e0163514.
34. Eastaugh LJ, James PA, Phelan DG, Davis AM. Brugada syndrome caused by a large deletion in SCN5A only detected by multiplex ligation-dependent probe amplification. *J Cardiovasc Electrophysiol* 2011;22:1073–1076.
35. Sonoda K, Ohno S, Ozawa J, et al. Copy number variations of SCN5A in Brugada syndrome. *Heart Rhythm* 2018;15:1179–1188.
36. Meregalli PG, Tan HL, Probst V, et al. Type of SCN5A mutation determines clinical severity and degree of conduction slowing in loss-of-function sodium channelopathies. *Heart Rhythm* 2009;6:341–348.
37. Ackerman MJ, Splawski I, Makielski JC, et al. Spectrum and prevalence of cardiac sodium channel variants among black, white, Asian, and Hispanic individuals: implications for arrhythmogenic susceptibility and Brugada/long QT syndrome genetic testing. *Heart Rhythm* 2004;1:600–607.
38. Poelzing S, Forleo C, Samodell M, et al. SCN5A polymorphism restores trafficking of a Brugada syndrome mutation on a separate gene. *Circulation* 2006; 114:368–376.
39. Shinlapawittayatorn K, Du XX, Liu H, Ficker E, Kaufman ES, Deschênes I. A common SCN5A polymorphism modulates the biophysical defects of SCN5A mutations. *Heart Rhythm* 2011;8:455–462.
40. Bezzina CR, Barc J, Mizusawa Y, et al. Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat Genet* 2013;45:1044–1049.
41. Vohra J, Rajagopalan S; CSANZ Genetics Council Writing Group. Update on the diagnosis and management of Brugada syndrome. *Heart Lung Circ* 2015; 24:1141–1148.
42. Clatot J, Neyroud N, Cox R, et al. Inter-regulation of Kv4.3 and voltage-gated sodium channels underlies predisposition to cardiac and neuronal channelopathies. *Int J Mol Sci* 2020;21:5057.
43. Clatot J, Ziyadeh-Isleem A, Maugeen S, et al. Dominant-negative effect of SCN5A N-terminal mutations through the interaction of Na(v)1.5  $\alpha$ -subunits. *Cardiovasc Res* 2012;96:53–63.
44. Clatot J, Hoshi M, Wan X, et al. Voltage-gated sodium channels assemble and gate as dimers. *Nat Commun* 2017;8:2077.
45. Matamoros M, Pérez-Hernández M, Guerrero-Serna G, et al. Nav1.5 N-terminal domain binding to  $\alpha$ 1-syntrophin increases membrane density of human Kir2.1, Kir2.2 and Nav1.5 channels. *Cardiovasc Res* 2016;110:279–290.
46. Pérez-Hernández M, Matamoros M, Alfayate S, et al. Brugada syndrome trafficking-defective Nav1.5 channels can trap cardiac Kir2.1/2.2 channels. *JCI Insight* 2018;3:e96291.
47. Portero V, Wilders R, Casini S, Charpentier F, O Verkerk A, Remme A. Kv4.3 expression modulates Nav1.5 sodium current. *Front Physiol* 2018;9:178.

- **Under Review in Heart Rhythm**

**A polygenic risk score to predict sudden cardiac arrest  
in patients with coronary artery disease**

Eleonora Porcu<sup>1,\*</sup>, Christian W. Thorball<sup>1</sup>, Alessandra Pia Porretta<sup>2,3</sup>, Etienne Pruvot<sup>2,4</sup>, Kim Wiskott<sup>5</sup>, Federica Gilardi<sup>5</sup>, Aurelien Thomas<sup>5</sup>, Claire Redin<sup>1</sup>, Zoltán Kutalik<sup>5,7,8</sup>, Tony Fracasso<sup>5</sup>, Olivier Müller<sup>2</sup>, Jacques Fellay<sup>1,6,9,\*</sup>

<sup>1</sup> Precision Medicine Unit, Biomedical Data Science Center, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

<sup>2</sup> Cardiovascular Department, Division of Cardiology, Lausanne University Hospital (CHUV), Lausanne, Switzerland

<sup>3</sup> Department of Clinical-Surgical Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy.

<sup>4</sup> Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

<sup>5</sup> University Center of Legal Medicine, Geneva and Lausanne University Hospitals, Switzerland

<sup>6</sup> Swiss Institute of Bioinformatics, Lausanne, Switzerland

<sup>7</sup> University Center for Primary Care and Public Health, Lausanne, Switzerland

<sup>8</sup> Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

<sup>9</sup> School of Life Sciences, École Polytechnique Fédérale de Lausanne, Station 19, 1015, Lausanne, Switzerland

## Abstract

Cardiovascular disease (CVD) is a leading health problem and the main cause of death globally. Even when underlying causative factors are known, it is unclear why a cardiovascular condition causes premature death in a victim while others can live longer with the same condition. Here we propose a combined polygenic risk score (metaPRS) based on coronary artery disease (CAD), myocardial infarction (MI), low-density lipoprotein (LDL) cholesterol, body mass index (BMI) and type 2 diabetes (T2D) to predict the risk of sudden cardiac arrest (SCA) in patients affected by severe cardiovascular conditions.

We collected 2,114 patients with reported history of acute coronary syndrome from the Centre Hospitalier Universitaire Vaudois (CHUV) Genomic Biobank (BGC) and extracted data from the UK Biobank (UKB) on 13,696 participants with similar medical history. Among them, 303 and 932 had a further reported diagnosis of SCA or ventricular tachycardia/fibrillation according to the International Classification of Diseases (ICD-10) codes in BGC and UKB, respectively. We demonstrate that the metaPRS is significantly associated with SCA in both cohorts ( $OR_{BGC} = 1.28$ ,  $P_{BGC} = 8.39 \times 10^{-05}$  and  $OR_{UKB} = 1.14$ ,  $P_{UKB} = 7.07 \times 10^{-05}$ ). Furthermore, using the diagnosis based on the International Classification of Diseases (ICD-10) codes available in the UKB, the metaPRS exhibits a strong association with the presence of aortocoronary bypass graft ( $OR_{UKB} = 1.31$ ,  $P_{UKB} = 6.93 \times 10^{-33}$ ) and coronary angioplasty implant ( $OR_{UKB}=1.14$ ,  $P_{UKB}=1.46 \times 10^{-12}$ ).

These results show that a combined genetic risk score for CVD and associated risk factors has the potential to predict the occurrence of SCA in patients with myocardial infarction, hence to identify patients who could benefit from further preventive measures.

## Introduction

Sudden cardiac arrest (SCA) is a leading health problem representing the main cause of cardiac mortality. While in the young population (< 35 years old), SCA is mainly due to inherited cardiomyopathies or congenital/acquired diseases, in the older population SCA largely occurs in the setting of sequela of coronary artery disease, namely acute ischemia, acute myocardial infarction (MI), or structural alterations such as scar formation or ventricular dilatation after ischemia or infarction [1].

Family-based studies showed that SCA has a genetic component [2-4] that is independent of all other well-known risk factors - including smoking, inactivity, obesity, alcohol consumption, hypertension, diabetes and high cholesterol. This has led to a considerable interest in identifying factors underlying such heritability.

Over the last decade, numerous genome-wide association studies (GWAS) have focused on elucidating the genetic basis of several cardiovascular diseases (CVD) [5-7]. However, while the GWAS approach proved successful for many CVD, mapping their polygenic architecture and identifying dozens of associated single nucleotide polymorphisms (SNPs), only few GWAS have reported genome-wide significant results for SCA [8-11]. Somewhat surprisingly, these SNP associations were not replicated in a larger GWAS meta-analysis performed in 2018 by Ashar et al [12]. This lack of replication may reflect an association specific to the populations studied, or can be due to differences in the study design.

GWAS performed on CVD have confirmed the “common variant-common disease” hypothesis, *i.e.* genetic risk for CVD is mainly due to the effect of multiple variants commonly observed in the population. Thus, although these variants typically have a small individual effect, their integration in a polygenic risk score (PRS) representing the genetic susceptibility, has been proposed to improve cardiovascular risk prediction. Since the first CAD GWAS, PRS showed its potential to improve CAD risk prediction. In particular, studies leveraging PRS have improved the prediction of incident and prevalent disease risk compared with individual traditional risk factors, and

individuals with the highest PRS appear to have a similar CAD risk as individuals with familial hypercholesterolemia, despite having relatively normal cholesterol levels and no other traditional risk factors [13]. These findings suggest that PRS can capture residual risk not accounted for by traditional risk factors and add predictive value beyond them.

Although many studies have focused on PRS to predict the risk of developing several CVD, no PRS has been proposed to calculate the risk of developing SCA. This lack of investigation could be due to the lack of GWAS on SCA. Given that ~80% of SCA cases are due to CAD [14], here we propose a CAD-PRS to predict the risk of SCA hypothesizing an additive effect of CAD-associated variants on SCA.

Although several risk factors for SCA are highly correlated with CAD, they could have an effect on SCA through different biological pathways. For example, several studies showed that type 2 diabetes is strongly associated with CAD [15, 16] but non-coronary atherosclerotic pathophysiologic processes associated with diabetes, like micro-vascular disease and autonomic neuropathy, have the potential to independently influence SCA among patients with diabetes [17-20]. As recent analytical advances have enabled more powerful PRS construction, through the integration of multiple sets of GWAS summary statistics (metaPRS) [21], here we used such strategy to predict SCA by incorporating to CAD-PRS the PRSs for four additional SCA risk factors: myocardial infarction, low-density lipoprotein cholesterol, body mass index and type 2 diabetes.

We applied the PRS in two cohorts of patients with acute coronary syndrome collected from the Centre Hospitalier Universitaire Vaudois (CHUV) Genomic Biobank (BGC) and the UK Biobank (UKB). Furthermore, using the diagnosis reported in the UKB electronic medical records based on the International Classification of Diseases (ICD-10), we performed phenome-wide association analysis to determine the relationship between a high polygenic score and ~5,000 clinical outcomes.

## Methods

### *Cohorts description*

BGC: The CHUV Genomic Biobank (BGC) was created by the Centre Hospitalier Universitaire Vaudois (CHUV) and the University of Lausanne in 2013 under the name of Institutional Biobank of Lausanne (BIL). The BGC collects biological samples, genetic and clinical data from CHUV patients who have agreed to participate by signing a broad informed consent form. BGC samples were genotyped with the Illumina Global Screening Array version 2 + multi disease (including ~650,000 genetic markers). Genotypes for >10 million genetic markers were inferred by imputation methods using the 1000 Genomes phase 3 reference panel [22]. All genotypes were quality checked using standard quality controls based on call rate, minor allele frequency, imputation quality, Hardy-Weinberg equilibrium and strand errors.

UKB: The UK Biobank (UKB) is a volunteer-based cohort of ~500'000 individuals from the general UK population. Individuals were aged 40-69 years at recruitment and underwent microarray-based genotyping as well as extensive phenotyping, which is constantly extended and includes physical measurements, blood biomarker analyses, socio-demographic and health-related questionnaires, as well as linkage to medical health records [23]. Participants signed a broad informed consent form and data were accessed through the application number 16389. Details of genotype data acquisition and quality control have been previously described [23]. Briefly, UKB participants were genotyped on two similar arrays (95% probe overlap): 438,427 samples were genotyped with the Applied Biosystems UK Biobank Axiom Array (825,927 probes) and 49,950 samples were genotyped with the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix (807,411 probes). Genotypes for >70 million genetic markers were inferred by imputation methods using the UK10K haplotype reference panel [24] merged together with the 1000 Genomes Phase 3 reference panel [22].

	ICD-10 code	BGC (N)*	UKB (N)*
controls (& cases)	I20.0 Unstable angina	609	6,869
	I21.0 Acute transmural myocardial infarction of anterior wall	402	2,283
	I21.1 Acute transmural myocardial infarction of inferior wall	443	2,785
	I21.2 Acute transmural myocardial infarction of other sites	98	341
	I21.3 Acute transmural myocardial infarction of unspecified site	79	171
	I21.4 Acute subendocardial myocardial infarction	1,305	2,168
	I21.9 Acute myocardial infarction, unspecified	59	4,789
	I22.0 Subsequent myocardial infarction of anterior wall	1	117
	I22.1 Subsequent myocardial infarction of inferior wall	2	205
	I22.8 Subsequent myocardial infarction of other sites	4	90
	I22.9 Subsequent myocardial infarction of unspecified site	5	545
	I23 Certain current complications following acute myocardial infarction	NA	73
	I25.2 Old myocardial infarction	1,704	12,045
	cases	I46.0 Cardiac arrest with successful resuscitation	88
I46.1 Sudden cardiac death, so described		NA	13
I46.9 Cardiac arrest, unspecified		27	634
I47.2 Ventricular tachycardia		391	1,307
I49.0 Ventricular fibrillation and flutter		121	516
	*not unique counts		

**Table 1.** List of ICD-10 codes used to define cases and controls in the CHUV Genomic Biobank (BGC) and UK Biobank (UKB) cohorts.

### Sample selection

From BGC and UKB, participants are selected from the diagnosis reported in the electronic medical records based on the International Classification of Diseases (ICD-10) codes defining acute coronary syndrome (myocardial infarction or unstable angina). These patients are considered as controls. Among them, cases are defined as patients with an additional reported diagnosis of SCA or ventricular tachycardia/fibrillation according to ICD-10 codes (Table 1). In our analysis, we considered only European unrelated samples and in total we collected 303 cases and 1,811 controls in BGC and 932 cases and 12,764 controls in UKB (for cohort description see Table 2).

	BGC		UKB	
	Controls	Cases	Controls	Cases
<b>Sex, N (%)</b>				
Female	544 (30%)	41 (14%)	3452 (27%)	154 (17%)
Male	1267 (70%)	262 (86%)	9312 (83%)	778 (83%)
<b>Age</b>	76.07 (12.81)	72.40 (11.02)	61.33 (6.33)	61.53 (5.95)
<b>BMI, mean (SD)</b>	26.87 (5.21)	27.63 (5.00)	29.08 (4.87)	29.11 (4.84)
<b>Diastolic blood pressure, mean (SD)</b>	73.61 (14.26)	71.69 (14.21)	81.18 (11.31)	80.65 (11.40)
<b>Systolic blood pressure, mean (SD)</b>	130.88 (22.39)	123.68 (20.79)	142.81 (20.41)	141.62 (21.56)
<b>Total cholesterol, mean (SD)</b>	4.76 (1.36)	4.89 (1.54)	5.02 (1.29)	4.96 (1.32)
<b>HDL cholesterol, mean (SD)</b>	1.21 (0.40)	1.11 (0.35)	1.24 (0.33)	1.21 (0.33)
<b>LDL cholesterol, mean (SD)</b>	2.63 (1.12)	2.62 (1.17)	3.11 (0.97)	3.09 (0.99)
<b>Triglycerides, mean (SD)</b>	1.21 (0.40)	1.11 (0.35)	2.02 (1.15)	1.98 (1.12)
<b>Ever smoked</b>				
Yes	528	101	9102	700
No	896	145	3585	229
Missing	387	57	77	3
<b>Diabetes (ICD10 code)</b>				
Type 1 (E10)	8	5	458	44
Type 2 (E11)	498	94	2676	213
Other (E12-E14)	77	14	425	36

**Table 2.** Descriptive statistics of BGC and UKB cohorts. The table shows descriptive statistics of the two cohorts included in the study. Diastolic and systolic blood pressure are reported in mmHg. Total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides are reported in mmol/L. SD, standard deviation.

### Polygenic risk score (PRS) analysis

PRS based on SCA-risk factors GWAS data were investigated with respect to SCA status using the software PRSice-2 [25]. We used large publicly available GWAS summary statistics as baseline data describing the strength and significance of the association of each SNP with coronary artery disease (CAD) [26], myocardial infarction (MI) [26], low-density lipoprotein (LDL) cholesterol (LDL summary statistics are from the Neale Lab - <http://www.nealelab.is/uk-biobank/>), body mass index (BMI) [27] and type 2 diabetes (T2D) [28]. Clumping was performed in PRSice-2 to account for linkage disequilibrium (LD) between SNPs using genotype data from the UK10K reference panel [24]. Scores were then calculated as the sum of effect alleles for all SNPs weighted by their reported regression coefficients. The p-values provided in the GWAS summary were used to

determine which SNPs to include in the PRS calculation. Various significance thresholds for SNP inclusion in the PRS were tested, and the threshold giving the best predictive value was used for follow-up analyses.

These multiple PRS were combined into one meta-score (metaPRS) using the formula described in [21] and reported below:

$$PRS_i^{meta} = \frac{\beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3}}{\sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2 + 2\beta_1\beta_2\rho_{1,2} + 2\beta_1\beta_3\rho_{1,3} + 2\beta_2\beta_3\rho_{2,3}}}$$

where  $Z_{i1}, Z_{i2}, Z_{i3}$  are the (zero-mean and unit-variance standardised) PRS for the  $i^{\text{th}}$  individual for the risk factors 1, 2 and 3 respectively,  $\beta_1, \beta_2, \beta_3$  are the univariate effects for each score, and  $\rho_{i,j}$  is the Pearson correlation between the  $i^{\text{th}}$  and  $j^{\text{th}}$  scores.

The association between the PRSs and SCA status was evaluated by fitting logistic regression models. To compare the proportion of SCA cases between each decile and the first decile, we used Fisher's test. All the statistical analysis were performed using R version 3.6.0 software (The R Foundation).

### Phenome-wide association analysis

To assess UKB patients' disease status, ICD-10 codes were used (#41270).

Logistic regression models were used to evaluate association of metaPRS with ICD-10 diagnoses. P-values, odds ratios (ORs) and 95% confidence intervals (CI) were estimated using R version 3.6.0 software (The R Foundation).

## **Results**

Given that ~80% of SCA cases are due to CAD, it is very likely that the two diseases share genetics – this hypothesis is supported by the significant genetic correlation between the two traits ( $r_g=0.46$ ). Using PRSice-2 software, we generated a set of CAD-PRSs using summary statistics derived from a large GWAS performed on 60,801 CAD cases and 123,504 controls [26] at different

significance thresholds. We observed the strongest association of CAD PRS and SCA when using 2,629 ( $P_{GWAS} < 3.40 \times 10^{-03}$ ) and 3,416 ( $P_{GWAS} < 3.75 \times 10^{-03}$ ) independent variants in BGC and UKB samples, respectively. In particular: in BGC samples we observed an OR of 1.16 [95% confidence interval (CI) 1.04–1.29] per SD increment in CAD-PRS ( $P = 0.014$ ); in UKB samples we observed an OR of 1.09 [1.02–1.15] per SD increment in CAD-PRS ( $P = 0.015$ ) (**Table 2**).

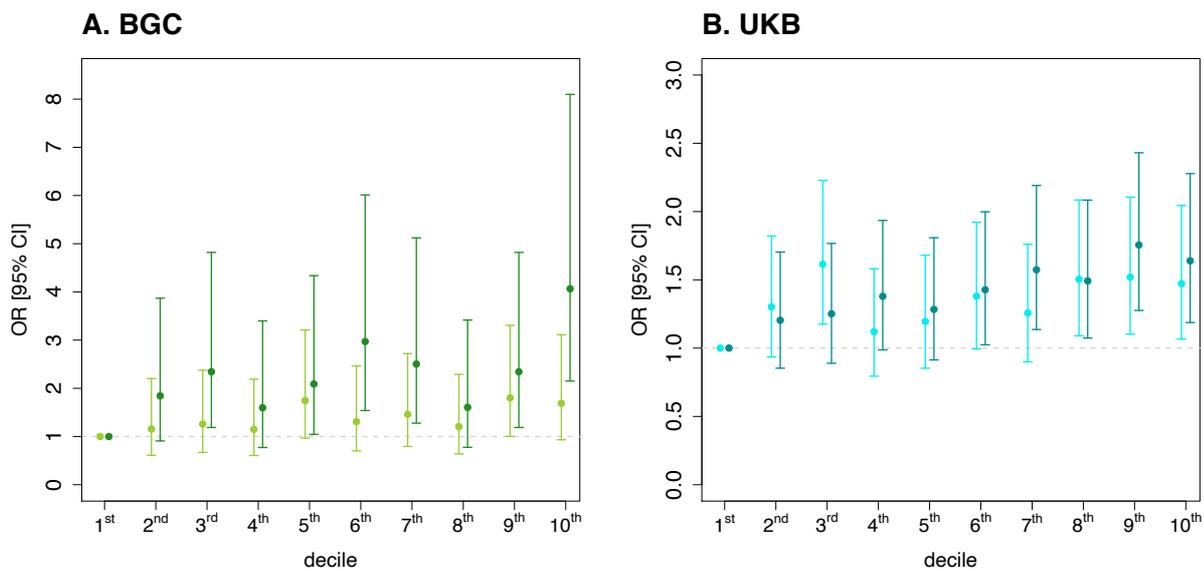
Similarly, in both cohorts we calculated a PRS for four SCA risk factors, i.e. myocardial infarction (MI), low-density lipoprotein (LDL) cholesterol, body mass index (BMI) and type 2 diabetes (T2D), using effect sizes from large GWAS metanalysis. While in BGC samples, only LDL-PRS shows a positive significant effect on SCA, in UKB samples we observed significant effect for PRS-MI and PRS-BMI (**Table 2**). Of note, the PRSs were correlated with each other to different degrees (**Supplementary Table 1**).

Trait	BGC				UKB			
	N SNPs	Threshold	OR [95%CI]	Pvalue	N SNPs	Threshold	OR [95%CI]	Pvalue
CAD	2629	$3.40 \times 10^{-03}$	1.16 [1.04-1.29]	0.014	3416	$3.75 \times 10^{-03}$	1.09 [1.02-1.15]	0.015
MI	223	$5.01 \times 10^{-05}$	1.11 [0.99-1.23]	0.095	3515	$2.70 \times 10^{-03}$	1.09 [1.03-1.16]	$9.48 \times 10^{-03}$
LDL	102543	0.15	1.16 [1.04-1.28]	0.017	1940	$1.00 \times 10^{-04}$	0.97 [0.91-1.04]	0.419
BMI	72594	$1.93 \times 10^{-03}$	1.09 [0.98-1.20]	0.131	12737	0.041	1.07 [1.01-1.14]	0.039
T2D	3122	$1.20 \times 10^{-03}$	1.13 [1.01-1.24]	0.051	47147	0.028	1.07 [1.00-1.14]	0.056
metaPRS	-	-	1.28 [1.16-1.40]	$8.39 \times 10^{-05}$	-	-	1.14 [1.08-1.21]	$7.07 \times 10^{-05}$

**Table 2.** Logistic regression results for individual PRS and metaPRS

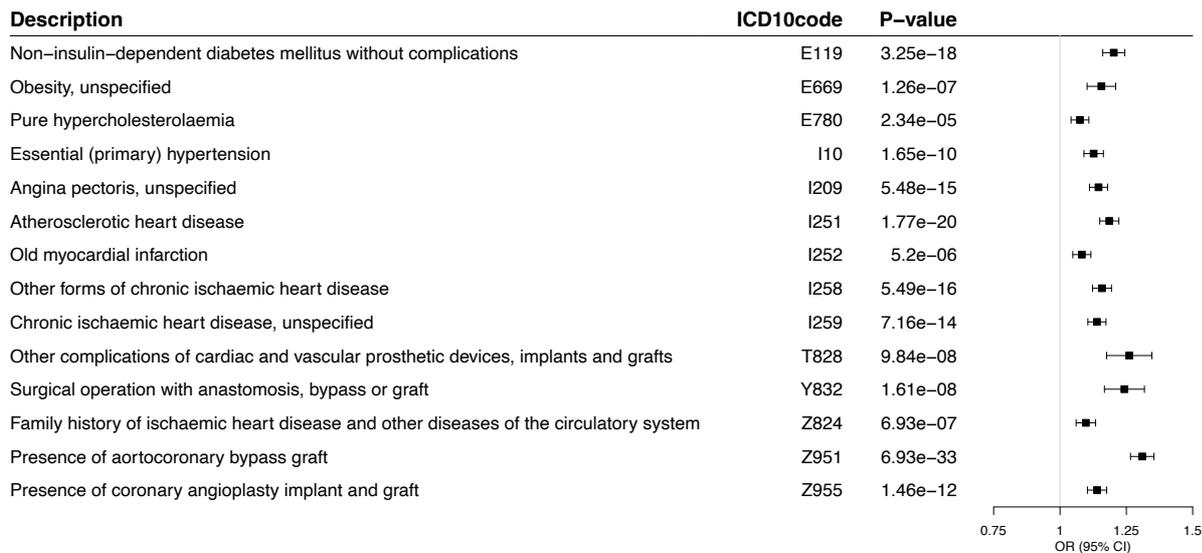
We then used a meta-analytic strategy to generate a metaPRS by integrating the five individual PRSs (see Methods). As expected, in both cohorts, the metaPRS had the greatest association with SCA than any other individual PRS with the OR of 1.28 [1.16–1.40] per SD increment in BGC samples ( $P = 8.39 \times 10^{-05}$ ) and the OR of 1.14 [1.08–1.21] per SD increment in UKB samples ( $P = 7.07 \times 10^{-05}$ ). We next stratified the populations according to PRS decile and also observed a more marked gradient of SCA risk across deciles of metaPRS than CAD-PRS (**Figure 1**). In BGC

samples, the metaPRS has a higher OR of 4.07 [2.15-8.10] ( $P = 1.71 \times 10^{-06}$ ) in the top decile vs. the bottom decile than the CAD-PRS (OR: 1.69 [0.93-3.11],  $P = 0.069$ ) (**Supplementary Table 2**). We observed the same trend in UKB: the metaPRS has a higher OR of 1.64 [1.19-2.28] ( $P = 1.82 \times 10^{-03}$ ) in the top decile vs. the bottom decile than the CAD-PRS (OR: 1.47 [1.07-2.04],  $P = 0.015$ ) (**Supplementary Table 3**).



**Figure 1.** Decile plots (OR) by metaPRS within each decile for BGC (A) and UKB (B) samples. Data are shown as ORs and 95% confidence limits (error bars) in each decile of the CAD-PRS (in light) and metaPRS (in dark) in comparison with the first decile.

We next determined the relationship between metaPRS and all ICD-10 codes reported in UKBB. Such association was determined in a logistic regression model within the UKB cohort of individuals affected by acute coronary syndrome (**Supplementary Table 4**). Beyond expected association with T2D, obesity and hypercholesterolaemia, metaPRS exhibits a strong association with the presence of aortocoronary bypass graft ( $OR_{UKB} = 1.31$ ,  $P_{UKB} = 6.93 \times 10^{-33}$ ) and coronary angioplasty implant ( $OR_{UKB}=1.14$ ,  $P_{UKB}=1.46 \times 10^{-12}$ ) (**Figure 2**).



**Figure 2.** The significant association of metaPRS with the prevalence of 14 ICD-10 codes reported in UKB. ORs and 95% confidence limits (error bars) were estimated using logistic regression on the metaPRS.

## Discussion

Sudden cardiac arrest (SCA) is a major global public health problem. Worldwide, SCA is the most common cause of death accounting for the 25% of the ~17 million annual deaths due to cardiovascular diseases [<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>]. Although in the adult population, the cause of SCA is often related to coronary artery diseases (CAD), knowledge remains incomplete. Indeed, even when causative factors are known, it remains unclear why the same condition causes premature death in a patient, while others can live longer with it. Today, among patients with coronary artery disease, arrhythmic risk stratification still remains very challenging and relies mainly upon clinical criteria. Among them, left ventricle ejection fraction (LVEF) represents the current recommended tool to identify high-risk patients needing ICD implantation for primary prevention of SCA. However, several studies have clearly pointed out that the greater proportion of patients experiencing SCA do not present with reduced LVEF. Moreover, in about half of cases, SCA represents the first overt expression of an underlying cardiac disease. This

context clearly underlines the dramatic need for refining SCA prediction and prevention, through a multidisciplinary approach encompassing different and early predictors of SCA.

Over the past years, family studies and monogenic forms of the disease have demonstrated that SCA, like other cardiovascular diseases, has a genetic component [2-4]. Here we proposed a PRS to predict which patients with acute coronary syndrome are more likely to develop SCA. Although the optimal scenario would be to have a PRS built on genetic variants associated with SCA, the lack of significant results in the largest GWAS performed to date on SCA using 3,939 cases and 25,989 controls [12] makes it unfeasible.

To overcome such limitation here we propose a PRS built on several SCA risk factors: CAD, myocardial infarction, LDL-cholesterol, BMI and type 2 diabetes. Our results show that CAD-PRS has the best predictive value. However, the other individual risk-factor PRS, although correlated with each others, contain independent information about SCA risk, so their combination in a metaPRS results in a more powerful prediction. The robustness of our findings is proven by their validation in two independent cohorts of patients affected by acute coronary syndrome: BGC and UKB.

Polygenic risk predictors have important potential implications for clinical medicine because they represent early predictive markers, which may identify high-risk individuals before that the potentially life-threatening condition manifests. In particular, in addition to predicting people at risk of SCA, our metaPRS shows a strong association with the presence of aortocoronary bypass graft and coronary angioplasty implant. We recognize that compared with PRS for other common diseases the performance of our metaPRS for SCA is limited but its ability to identify high-risk patients may facilitate targeted therapies, interventions and appropriate preventive strategies. In this context, our study adds a further contribution, which lays the groundwork to the application of a precision medicine approach in coronary artery disease. The complex interplay between the organic substrate, the transient ongoing triggers and the genetic background represents indeed the keystone to unravel risk prediction in coronary artery disease whose variable expression may encompass SCA as the most severe phenotype.

Our study has several limitations. The metaPRS score was built only on SCA-risk factors due to the lack of significant variants associated with SCA. This lack of findings is likely due to the small number of SCA cases and the large heterogeneity with respect to underlying pathology and case definitions. Such previous unsuccessful GWAS highlighted the importance of stratifying patients according to their similar clinical presentation.

Furthermore, the cohorts studied here were of European ancestry and future studies are needed to investigate what performance is achievable in individuals of different ancestry. In the future, successful development of PRS in non-Europeans will require both GWAS summary statistics from non-Europeans as well as the inclusion of non-European SCA patients on which to validate the PRS.

Taken together, despite limitations, our study presents the first PRS developed to predict the risk of SCA in patients affected by acute coronary syndrome and assesses its potential for risk stratification in the context of targeted strategies for prevention of SCA and other cardiovascular complications. Our work highlights the value of combining multiple individual PRSs and how a careful study of individuals at the tails of a PRS distribution might shed light on the pathogenesis of diseases.

## References

1. Hayashi, M., W. Shimizu, and C.M. Albert, *The spectrum of epidemiology underlying sudden cardiac death*. *Circ Res*, 2015. **116**(12): p. 1887-906.
2. Dekker, L.R., et al., *Familial sudden death is an important risk factor for primary ventricular fibrillation: a case-control study in acute myocardial infarction patients*. *Circulation*, 2006. **114**(11): p. 1140-5.
3. Friedlander, Y., et al., *Family history as a risk factor for primary cardiac arrest*. *Circulation*, 1998. **97**(2): p. 155-60.
4. Jouven, X., et al., *Predicting sudden death in the population: the Paris Prospective Study I*. *Circulation*, 1999. **99**(15): p. 1978-83.

5. Aro, A.L., et al., *Prognostic significance of prolonged PR interval in the general population*. Eur Heart J, 2014. **35**(2): p. 123-9.
6. Jouven, X., et al., *Resting heart rate as a predictive risk factor for sudden death in middle-aged men*. Cardiovascular Research, 2001. **50**(2): p. 373-378.
7. Laukkanen, J.A., et al., *T-Wave Inversion, QRS Duration, and QRS/T Angle as Electrocardiographic Predictors of the Risk for Sudden Cardiac Death*. American Journal of Cardiology, 2014. **113**(7): p. 1178-1183.
8. Aouizerat, B.E., et al., *GWAS for discovery and replication of genetic loci associated with sudden cardiac arrest in patients with coronary artery disease*. BMC Cardiovasc Disord, 2011. **11**: p. 29.
9. Arking, D.E., et al., *Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in European ancestry individuals*. PLoS Genet, 2011. **7**(6): p. e1002158.
10. Bezzina, C.R., et al., *Genome-wide association study identifies a susceptibility locus at 21q21 for ventricular fibrillation in acute myocardial infarction*. Nat Genet, 2010. **42**(8): p. 688-691.
11. Huertas-Vazquez, A., et al., *Novel loci associated with increased risk of sudden cardiac death in the context of coronary artery disease*. PLoS One, 2013. **8**(4): p. e59905.
12. Ashar, F.N., et al., *A comprehensive evaluation of the genetic architecture of sudden cardiac arrest*. Eur Heart J, 2018. **39**(44): p. 3961-3969.
13. Levin, M.G. and D.J. Rader, *Polygenic Risk Scores and Coronary Artery Disease: Ready for Prime Time?* Circulation, 2020. **141**(8): p. 637-640.
14. Myerburg, R.J., *Sudden cardiac death: exploring the limits of our knowledge*. J Cardiovasc Electrophysiol, 2001. **12**(3): p. 369-81.
15. Aronson, D. and E.R. Edelman, *Coronary artery disease and diabetes mellitus*. Cardiol Clin, 2014. **32**(3): p. 439-55.

16. Grant, P.J., F. Cosentino, and N. Marx, *Diabetes and coronary artery disease: not just a risk factor*. Heart, 2020. **106**(17): p. 1357-1364.
17. Ewing, D.J., I.W. Campbell, and B.F. Clarke, *The natural history of diabetic autonomic neuropathy*. Q J Med, 1980. **49**(193): p. 95-108.
18. Rathmann, W., et al., *Mortality in diabetic patients with cardiovascular autonomic neuropathy*. Diabet Med, 1993. **10**(9): p. 820-4.
19. Targher, G., et al., *Diabetic retinopathy is associated with an increased incidence of cardiovascular events in Type 2 diabetic patients*. Diabet Med, 2008. **25**(1): p. 45-50.
20. Wheeler, S.G., J.H. Ahroni, and E.J. Boyko, *Prospective study of autonomic neuropathy as a predictor of mortality in patients with diabetes*. Diabetes Res Clin Pract, 2002. **58**(2): p. 131-8.
21. Inouye, M., et al., *Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention*. J Am Coll Cardiol, 2018. **72**(16): p. 1883-1893.
22. Genomes Project, C., et al., *A global reference for human genetic variation*. Nature, 2015. **526**(7571): p. 68-74.
23. Bycroft, C., et al., *The UK Biobank resource with deep phenotyping and genomic data*. Nature, 2018. **562**(7726): p. 203-209.
24. Huang, J., et al., *Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel*. Nat Commun, 2015. **6**: p. 8111.
25. Choi, S.W. and P.F. O'Reilly, *PRSice-2: Polygenic Risk Score software for biobank-scale data*. Gigascience, 2019. **8**(7).
26. Nikpay, M., et al., *A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease*. Nat Genet, 2015. **47**(10): p. 1121-1130.
27. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology*. Nature, 2015. **518**(7538): p. 197-206.

28. Scott, R.A., et al., *An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans*. *Diabetes*, 2017. **66**(11): p. 2888-2902.

**Supplementary Table 1.** Correlation between PRS in BGC and UKBB

Trait-PRS	CAD	MI	LDL	BMI	T2D
<b>CAD</b>	-	0.51 ( $<2.23 \times 10^{-308}$ )	0.12 ( $6.09 \times 10^{-47}$ )	0.01 (0.51)	0.05 ( $1.54 \times 10^{-09}$ )
<b>MI</b>	0.44 ( $2.93 \times 10^{-102}$ )	-	0.09 ( $6.79 \times 10^{-27}$ )	0.00 (0.79)	0.03 ( $5.65 \times 10^{-05}$ )
<b>LDL</b>	0.09 ( $4.35 \times 10^{-05}$ )	0.09 ( $4.24 \times 10^{-05}$ )	-	-0.02 ( $5.82 \times 10^{-03}$ )	0.00 (0.33)
<b>BMI</b>	0.02 (0.36)	-0.08 ( $2.04 \times 10^{-04}$ )	-0.04 (0.56)	-	0.09 ( $3.81 \times 10^{-26}$ )
<b>T2D</b>	0.04 (0.06)	-0.06 ( $9.78 \times 10^{-03}$ )	-0.03 (0.22)	0.37 ( $6.90 \times 10^{-69}$ )	-

Each cell contains the correlation (and the P-value) between the individual PRSs in BGC (green cells) and UKB (blue cells)

**Supplementary Table 2.** Fisher's Test results for each decile of the CAD-PRS and metaPRS in comparison with the first decile in BGC samples.

decile	CAD-PRS				metaPRS			
	Cases	Controls	OR [CI]	P	Cases	Controls	OR [CI]	P
1 <sup>st</sup>	22	189	-	-	15	196	-	-
2 <sup>nd</sup>	26	185	1.15 [0.61-2.20]	0.65	26	185	1.84 [0.91-3.87]	0.072
3 <sup>rd</sup>	28	183	1.26 [0.67-2.38]	0.46	32	179	2.34 [1.19-4.82]	0.0086
4 <sup>th</sup>	26	186	1.15 [0.61-2.19]	0.76	23	189	1.60 [0.77-3.40]	0.23
5 <sup>th</sup>	37	174	1.75 [0.97-3.21]	0.052	29	182	2.09 [1.04-4.34]	0.026
6 <sup>th</sup>	29	182	1.31 [0.70-2.46]	0.38	39	172	2.97 [1.54-6.01]	$4.4 \times 10^{-04}$
7 <sup>th</sup>	32	180	1.46 [0.79-2.72]	0.25	34	178	2.50 [1.28-5.12]	0.0058
8 <sup>th</sup>	27	184	1.21 [0.64-2.29]	0.55	23	188	1.60 [0.78-3.42]	0.18
9 <sup>th</sup>	38	173	1.80 [1.00-3.31]	0.039	32	179	2.34 [1.19-4.82]	0.0086
10 <sup>th</sup>	36	175	1.69 [0.93-3.11]	0.069	50	161	4.07 [2.15-8.10]	$1.7 \times 10^{-06}$

**Supplementary Table 3.** Fisher’s Test results for each decile of the CAD-PRS and metaPRS in comparison with the first decile in UKB samples.

decile	CAD-PRS				metaPRS			
	Cases	Controls	OR [CI]		Cases	Controls	OR [CI]	P
1 <sup>st</sup>	71	1298	-	-	68	1301	-	-
2 <sup>nd</sup>	91	1278	1.30 [0.93-1.82]	0.11	81	1288	1.20 [0.85-1.70]	0.27
3 <sup>rd</sup>	111	1258	1.61 [1.17-2.23]	0.0021	84	1285	1.25 [0.89-1.77]	0.18
4 <sup>th</sup>	79	1290	1.12 [0.79-1.58]	0.50	92	1277	1.38 [0.99-1.93]	0.051
5 <sup>th</sup>	84	1285	1.20 [0.85-1.68]	0.28	86	1283	1.28 [0.91-1.81]	0.14
6 <sup>th</sup>	96	1273	1.38 [0.99-1.92]	0.046	95	1274	1.43 [1.02-2.00]	0.029
7 <sup>th</sup>	88	1281	1.26 [0.90-1.76]	0.17	104	1265	1.57 [1.14-2.19]	0.0046
8 <sup>th</sup>	104	1265	1.50 [1.09-2.08]	0.010	99	1270	1.49 [1.07-2.08]	0.013
9 <sup>th</sup>	105	1264	1.52 [1.10-2.11]	0.0081	115	1254	1.76 [1.28-2.43]	0.00031
10 <sup>th</sup>	102	1267	1.47 [1.07-2.04]	0.015	108	1261	1.64 [1.19-2.28]	0.0018