

# Rare-variant collapsing and bioinformatic analyses for different types of cardiac arrhythmias in the UK Biobank reveal novel susceptibility loci and candidate amyloid-forming proteins

Bengt Zöller, MD, PhD,<sup>\*</sup> Eric Manderstedt, MSc,<sup>\*</sup> Christina Lind-Halldén, PhD,<sup>†</sup>  
Christer Halldén, PhD<sup>\*</sup>

From the <sup>\*</sup>Center for Primary Health Care Research, Department of Clinical Sciences, Lund University and Region Skåne, Malmö, Sweden, and <sup>†</sup>Department of Bioanalysis, Kristianstad University, Kristianstad, Sweden.

**BACKGROUND** Cardiac arrhythmias are a common health problem. Both common and rare genetic risk factors exist for cardiac arrhythmias. Cardiac amyloidosis is a rare disease that may manifest various arrhythmias. Few large-scale whole exome sequencing studies elucidating the contribution of rare variations to arrhythmias have been published.

**OBJECTIVE** To access gene collapsing analysis of rare variations for different types of cardiac arrhythmias in UK Biobank. Identified genes were analyzed *in silico* for probability to form amyloid fibrils.

**METHODS** We used 2 published UK Biobank portals (<https://azpewas.com/> and <https://app.genebass.org/>) to access gene collapsing analysis of rare variations for different types of cardiac arrhythmias. Diagnosis of arrhythmia was based on the International Classification of Diseases, 10th Revision (ICD-10) codes: conduction disorders (I44, I45), paroxysmal tachycardia (I47), atrial fibrillation (I48), and other arrhythmias (I49).

**RESULTS** Rare variations in 5 genes were linked to conduction disorders (*SCN5A*, *LMNA*, *SMAD6*, *HSPB9*, *TMEM95*). The *TTN* gene was

associated with both paroxysmal tachycardia and other arrhythmias. Atrial fibrillation was associated with rare variations in 8 genes (*TTN*, *RPL3L*, *KLF1*, *TET2*, *NME3*, *KDM5B*, *PKP2*, *PMVK*). Two of the genes linked to heart conduction disorders were potential amyloid-forming proteins (*HSPB9*, *TMEM95*), while none of the 8 genes linked to other types of arrhythmias were potential amyloid-forming proteins.

**CONCLUSION** Rare variations in 13 genes were associated with arrhythmias in the UK Biobank. Two of the heart conduction disorder-linked genes are potential amyloid-forming candidates. Amyloid formation may be an underestimated cause of heart conduction disorders.

**KEYWORDS** Epidemiology; Genetics; Mutation; Arrhythmias; Cardiac

(Cardiovascular Digital Health Journal 2023; ■:1–4) © 2023 Heart Rhythm Society. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Cardiac arrhythmias are a common challenge to human health.<sup>1</sup> The development of effective technologies for the treatment of cardiac arrhythmias has exceeded the knowledge of the underlying biology.<sup>1</sup> Both acquired and common and rare genetic and risk factors exist for cardiac arrhythmias.<sup>1,2</sup> Cardiac amyloidosis is a rare disease that may manifest various arrhythmias, such as atrioventricular nodal block, atrial fibrillation, and ventricular tachyarrhythmias.<sup>2</sup> Amyloidosis is characterized by extracellular deposits of amyloid in various organs.<sup>2</sup> Cardiac amyloidosis with cardiomyopathy is a

frequent feature of amyloidosis. Patients with cardiac amyloidosis frequently encounter various arrhythmias.<sup>2</sup> There are 2 main forms of amyloidosis that affect the heart: light chain amyloidosis and transthyretin amyloidosis.<sup>2</sup> However, there are other rare causes of heart amyloidosis. Genome-wide association studies (GWAS) have been successful in linking common variants to common disease (<https://www.ebi.ac.uk/gwas/>). Whole exome sequencing studies (WES) might be an interesting option to elucidate the genetics of cardiac arrhythmias. However, few large-scale WES of arrhythmias in the general population have been published.<sup>3</sup> We have used 2 published UK Biobank portals (<https://azpewas.com/> and <https://app.genebass.org/>) to access gene collapsing analysis of rare variations for different types of cardiac arrhythmias.<sup>4,5</sup> Identified genes were further analyzed using bioinformatic sources. The AMYPred-FRL<sup>6</sup> web server was used to analyze

**Address reprint requests and correspondence:** Prof Bengt Zöller, Center for Primary Health Care Research, CRC, Building 28, Floor 11, Jan Waldenströms gata 35, Skåne University Hospital, S-205 02 Malmö, Sweden. E-mail address: [bengt.zoller@med.lu.se](mailto:bengt.zoller@med.lu.se).

*in silico* whether identified proteins may be potential candidates to form amyloid fibrils.

## Methods

Wang and colleagues<sup>4</sup> reported the relationships between rare protein-coding variants and 17,361 binary and 1419 quantitative phenotypes using exome sequencing data from 269,171 UK Biobank participants of European ancestry (<https://azphewas.com/>). Recently, Karczewski and colleagues<sup>5</sup> determined gene-based associations for 4529 phenotypes in 394,841 UK Biobank exomes (<https://app.genebass.org/>). The 2 portals used different statistical methods for the analysis.<sup>4,5</sup> We used 2 published UK Biobank portals (<https://azphewas.com/> and <https://app.genebass.org/>) to access gene collapsing analysis of rare

variations for different types of cardiac arrhythmias (Table 1).<sup>4,5</sup> The significance levels used in the 2 portals and published studies were stringent.<sup>4,5</sup> In order not to discard potential candidate genes, we present genes with *P* values  $< .05/20,000$  genes =  $2.5 \times 10^{-6}$  commonly used for WES studies. In the AstraZeneca portal the UK Biobank 450k (v4) public version was used (<https://azphewas.com/>). In Table 1 only the genes with genome-wide significant results are shown with *P* values for the most significant model. Diagnosis of arrhythmia was based on the International Classification of Diseases, 10th Revision (ICD-10) codes (<https://azphewas.com/> and <https://app.genebass.org/>).<sup>4,5</sup> Other bioinformatic tools used were the GWAS catalog (<https://www.ebi.ac.uk/gwas/>), OMIM (<https://www.omim.org/>), the GeneCards suite (<https://www.genecards.org/>), and the AMYPred-FRL web server.<sup>6</sup> The AMYPred-FRL web server

**Table 1** Results of gene collapsing analysis of rare variants for heart arrhythmia according to 3-digit International Classification of Diseases, 10th Revision codes in 2 published UK Biobank portals (AstraZeneca; <https://azphewas.com/> and Genebass; <https://app.genebass.org/>)<sup>4,5</sup>

ICD-10 codes and name of diseases	Variant	GWAS catalog	OMIM/GeneCards	AMYPred-FRL amyloid probability <sup>6</sup>
I44 Atrioventricular and left bundle-branch block	<i>LMNA</i> 1.03e-6*	-	‡Cardiomyopathy dilated, muscular dystrophy	Non-AMY 0.069
	<i>SMAD6</i> 1.19e-9*†	-	Aortic valve disease	non-AMY 0.099
	<i>HSPB9</i> 2.13e-6*	-	Hereditary motor neuropathy (HSPB family)	AMY 0.829
	<i>SCN5A</i> 2.21e-8†	Brugada syndrome, atrial fibrillation	§Atrial fibrillation, heart block	Non-AMY 0.090
I45 Other conduction disorders	<i>SMAD6</i> 9.80e-9*†	-	Aortic valve disease	Non-AMY 0.099
	<i>TMEM95</i> 1.63e-6†	-	-	AMY 0.692
I47 Paroxysmal tachycardia	<i>TTN</i> 2.84e-9*	Atrial fibrillation, heart failure	Cardiomyopathy, muscular dystrophy, myopathy	Non-AMY 0.075
I48 Atrial fibrillation and flutter	<i>TTN</i> 8.05e-24*	Atrial fibrillation, heart failure	Cardiomyopathy, muscular dystrophy, myopathy	Non-AMY 0.075
	<i>RPL3L</i> 9.55e-13*†	Atrial fibrillation	Cardiomyopathy	Non-AMY 0.053
	<i>KLF1</i> 2.56e-8*	-	Dyserythropoietic anemia, blood group	Non-AMY 0.252
	<i>TET2</i> 2.85e-8*	Prostate carcinoma, COPD	Myelodysplastic syndrome, immunodeficiency	Non-AMY 0.183
	<i>NME3</i> 2.86e-7*†	-	Lipase deficiency, Heinz Body anemias	Non-AMY 0.134
	<i>KDM5B</i> 3.13e-12†	Mitral valve prolapse	Intellectual developmental disorder, autism	Non-AMY 0.072
	<i>PKP2</i> 1.62e-6†	Atrial fibrillation	Arrhythmogenic dysplasia and cardiomyopathy	Non-AMY 0.073
	<i>PMVK</i> 1.37e-7†	Atrial fibrillation, Parkinson disease	Porokeratosis	Non-AMY 0.278
I49 Other cardiac arrhythmias	<i>TTN</i> 6.14e-11*	Atrial fibrillation, heart failure	Cardiomyopathy, muscular dystrophy, myopathy	Non-AMY 0.075

COPD = chronic obstructive pulmonary disease; ICD-10 = International Classification of Diseases, 10th Revision.

Union was used to define phenotypes for [https://azphewas.com](https://azphewas.com/). Only genome-wide significantly associated genes are shown in the table (best model), ie,  $P < 2.5 \times 10^{-6}$  (\*significant in AstraZeneca portal [<https://azphewas.com/>]; †significant in Genebass portal [<https://app.genebass.org/>]). In the AstraZeneca portal the UK Biobank 450k (v4) public version was used (<https://azphewas.com/>). Three genes were genome-wide significant in the 2 databases, the others only in AstraZeneca portal or Genebass portal.<sup>4,5</sup>

Significance threshold was *P* value  $< .05/20,000$  genes =  $2.5 \times 10^{-6}$ , commonly used for WES studies. For the Genebass portal (<https://app.genebass.org/>) the highest value for SKAT-0, SKAT, or burden test are shown. For the AstraZeneca portal (<https://azphewas.com/>) the highest *P* value of the 12 tested models is shown.<sup>4,5</sup>

Other bioinformatic tools used were the GWAS catalog (<https://www.ebi.ac.uk/gwas/>), OMIM (<https://www.omim.org/>), the GeneCards suite (<https://www.genecards.org/>), and the AMYPred-FRL web server.<sup>6</sup> AMYPred-FRL results were obtained from the <http://pmlabstack.pythonanywhere.com/AMYPred-FRL> server using the respective protein sequence in FASTA format obtained from <https://www.uniprot.org/>.

‡Associated also with Hutchinson-Gilford progeria, Charcot-Marie-Tooth disease, Emery-Dreifuss muscular dystrophy, lipodystrophy, Malouf syndrome, restrictive dermopathy, mandibuloacral dysplasia, heart-hand syndrome.

§Associated also with Brugada syndrome, cardiomyopathy, long QT syndrome, sick sinus syndrome, ventricular fibrillation.

was used to analyze *in silico* whether the identified proteins may be potential candidates to form amyloid fibrils (<http://pmlabstack.pythonanywhere.com/AMYPred-FRL>).<sup>6</sup> The protein sequences in FASTA format were obtained from UniProt (<https://www.uniprot.org/>). Validation of the AMYPred software has shown an accuracy of 0.873, sensitivity of 0.848, and specificity of 0.883.<sup>6</sup>

## Results

One previously GWAS-linked gene (*SCN5A*) and 3 novel genes (*LMNA*, *SMAD6*, *HSPB9*) were identified for atrioventricular and left bundle branch block (ICD10 code I44). Only loss-of-function variants in the *SCN5A* were associated with I44. For other conduction disorders several novel associations were identified. The *SMAD6* and *TMEM95* genes were associated with ICD10 code I45. One gene (*TTN*) was associated with both paroxysmal tachycardia (ICD10 code I47) and other cardiac arrhythmias (ICD10 code I49). Atrial fibrillation (ICD10 code I48) was associated with rare variation in 8 genes (*TTN*, *RPL3L*, *KLF1*, *TET2*, *NME3*, *KDM5B*, *PKP2*, *PMVK*). Four of these genes have previously been linked in GWAS to atrial fibrillation (*TTN*, *RPL3L*, *PKP2*, *PMVK*).

Using AMYPred-FRL the *HSPB9* and *TMEM95* proteins were predicted to be novel amyloid-forming candidates with probabilities of 0.829 and 0.692, respectively. The other 11 proteins had low *in silico* probability to form amyloid fibrils (Table 1). Thus, 2 of the genes linked to heart conduction disorders were potential amyloid-forming proteins, while none of the 8 genes linked to other arrhythmias were potential amyloid-forming proteins.

## Discussion

The present study shows that rare variations in 13 genes are associated with different types of arrhythmias in UK Biobank. Eight of these associations were novel (*LMNA*, *SMAD6*, *HSPB9*, *TMEM95*, *KLF1*, *TET2*, *NME3*, *KDM5B*) and 5 genes have previously been linked to arrhythmias (*SCN5A*, *TTN*, *RPL3L*, *PKP2*, *PMVK*). Of special interest is that 2 of the genes linked to heart conduction disorders were potential amyloid-forming proteins (*HSPB9*, *TMEM95*), while none of the 8 genes linked to other arrhythmias were potential amyloid-forming proteins.

Previously genes coding for heat shock protein family B (*HSPB1*, *HSPB3*, *HSPB5*, *HSPB8*) closely related to the *HSPB9* gene have been associated with distal hereditary motor neuropathy (<https://www.genecards.org/>).<sup>7</sup> Neuropathy is a known manifestation of amyloidosis (<https://www.omim.org/>), which further strengthens our finding that *HSPB9* codes for a potential amyloid-forming protein. No association between the *TMEM95* gene and amyloid or arrhythmia has been reported. Another interesting finding is the link between atrial fibrillation and rare variations in the *TET2* gene. The *TET2* gene is one of the driver genes involved in clonal hematopoiesis of indeterminate potential (CHIP) and has been linked to cardiovascular diseases (<https://www.omim.org/>).

Clonal hematopoiesis of indeterminate potential involves proinflammatory macrophages and an inflammasome-dependent immune response (interleukin-1 and interleukin-6) in atherosclerotic plaques or directly in the myocardium.<sup>8</sup> Myocardial inflammation may induce cardiac fibrosis, even in the absence of atherosclerotic cardiovascular disease.<sup>8</sup>

## Strengths and limitations

A strength of our study is the large number of sequenced exomes in UK Biobank.<sup>4,5</sup> A limitation is the lack of validation studies of our findings. Other limitations are the lack of experimental evidence and the use of 3-digit ICD-10 codes for the diagnosis of cardiac arrhythmias in UK Biobank. Certain ICD-10 codes are heterogeneous in terms of electrophysiological mechanisms (ie, I44, I45, I47, I49). Moreover, owing to the large number of comparisons some associations might be spurious, although we applied the genome-wide threshold  $2.5 \times 10^{-6}$  commonly used for WES studies.

## Conclusion

In conclusion, rare variations in 13 genes were associated with different types of arrhythmias in UK Biobank. Two of the heart conduction disorders linked genes are potential amyloid-forming candidates. We therefore hypothesize that amyloid formation could be an underestimated cause of heart conduction disorders.

## Acknowledgments

We thank the free access to the Genebase and the AstraZeneca PheWAS portals that made this work possible (<https://azphewas.com/> and <https://app.genebase.org/>).<sup>4,5</sup>

## Funding Sources

This work was supported by a grant awarded to Dr Bengt Zöllner by ALF-funding from Region Skåne and by the Swedish Research Council. The funders had no role in the study.

## Disclosures

The authors declare no competing interests.

## Authorship

All authors attest they meet the current ICMJE criteria for authorship.

## Ethics Statement

Ethical statements are not required for the study as no human or animal studies are involved.

## Data Availability

Data are publicly available (<https://azphewas.com/>, <https://app.genebase.org/>, <https://www.ebi.ac.uk/gwas/>, <https://www.genecards.org/>, <https://www.omim.org/>, <https://www.uniprot.org/>, and <http://pmlabstack.pythonanywhere.com/AMYPred-FRL>).

## References

1. Grace AA, Roden DM. Systems biology and cardiac arrhythmias. *Lancet* 2012; 380:1498–1508.
2. Laptseva N, Rossi VA, Sudano I, et al. Arrhythmic manifestations of cardiac amyloidosis: challenges in risk stratification and clinical management. *J Clin Med* 2023;12:2581.
3. Kandola MS, Kulm S, Kim LK, et al. Population-level prevalence of rare variants associated with atrial fibrillation and its impact on patient outcomes. *JACC Clin Electrophysiol* 2023;9:1137–1146.
4. Wang Q, Dhindsa RS, Carss K, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* 2021;597:527–532.
5. Karczewski KJ, Solomonson M, Chao KR, et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom* 2022;2:100168.
6. Charoenkwan P, Ahmed S, Nantasenamat C, et al. AMYPred-FRL is a novel approach for accurate prediction of amyloid proteins by using feature representation learning. *Sci Rep* 2022;12:7697.
7. Sarparanta J, Jonson PH, Kawan S, Udd B. Neuromuscular diseases due to chaperone mutations: a review and some new results. *Int J Mol Sci* 2020;21:1409.
8. Sikking MA, Stroeks SLVM, Waring OJ, et al. Clonal hematopoiesis of indeterminate potential from a heart failure specialist's point of view. *J Am Heart Assoc* 2023;12:e030603.