



International Journal of Biology Sciences

ISSN Print: 2664-9926
ISSN Online: 2664-9934
NAAS Rating (2025): 4.82
IJBS 2025; 7(8): 311-314
www.biologyjournal.net
Received: 19-06-2025
Accepted: 23-07-2025

Knut Arnesen
Department of Population
Genetics, Tromsø Institute of
Arctic Biology, Tromsø,
Norway

Ragnhild Eriksen
Department of Population
Genetics, Tromsø Institute of
Arctic Biology, Tromsø,
Norway

Corresponding Author:
Knut Arnesen
Department of Population
Genetics, Tromsø Institute of
Arctic Biology, Tromsø,
Norway

Whole exome sequencing for identification of rare variants in congenital heart disease families

Knut Arnesen and Ragnhild Eriksen

DOI: <https://www.doi.org/10.33545/26649926.2025.v7.i8d.664>

Abstract

Roughly 1 in 100 live births worldwide is affected by congenital heart disease (CHD), yet the genetic architecture behind most familial cases remains poorly mapped. This research applied whole exome sequencing (WES) to eight Norwegian families—each with at least two affected members—to identify rare coding variants that segregate with cardiac malformations. Genomic DNA was extracted from 47 individuals (23 affected, 24 unaffected relatives) recruited through the Tromsø Institute of Arctic Biology between January 2022 and August 2023. After quality filtering, a mean of 68,412 variants per exome was retained, of which 2.7% were classified as rare (minor allele frequency < 0.01 in gnomAD v3.1). Variant prioritization by segregation analysis, predicted functional impact (CADD score ≥ 20), and cardiac expression data narrowed candidates to 34 high-confidence variants across 29 genes. Missense changes accounted for 38.7% of prioritized variants, followed by synonymous variants filtered out at 19.5%. Notably, recurrent hits appeared in sarcomere protein genes (MYH6, MYH7, ACTC1) in three unrelated families. Sanger sequencing confirmed all 34 candidates. Pathway enrichment pointed to Notch signaling and chromatin remodeling as overrepresented biological processes (adjusted $p < 0.005$). These findings expand the known catalog of CHD-associated variants in northern European populations and highlight sarcomere and Notch pathway genes as recurrent contributors to familial cardiac malformations.

Keywords: Whole exome sequencing, rare variants, congenital heart disease, familial genetics, cardiac malformations, sarcomere genes, Notch signaling, variant prioritization, northern European population, segregation analysis

Introduction

An estimated 1.35 million newborns each year receive a diagnosis of congenital heart disease, making it the most common category of birth defect globally ^[1]. Despite decades of linkage mapping and candidate-gene work, the genetic basis of CHD remains incompletely understood—particularly in families where multiple members are affected across generations. Whole exome sequencing has emerged as a practical tool for identifying rare coding variants in such pedigrees, because it captures the approximately 1.5% of the genome most likely to harbor disease-causing mutations at a fraction of whole-genome cost ^[2].

Familial CHD typically follows complex inheritance patterns. Single-gene mutations explain only about 10-15% of cases; the rest appear to involve combinations of rare and common variants acting together with environmental modifiers ^[3]. Large-scale consortium efforts like the Pediatric Cardiac Genomics Consortium have identified dozens of candidate genes, yet most of these discoveries come from outbred populations in North America and East Asia ^[4]. Northern European cohorts—especially those from genetically more homogeneous populations such as Norway—remain underrepresented in the literature, despite potential advantages for detecting founder variants with reduced background noise.

This research aimed to fill part of that gap by applying WES to eight multiplex Norwegian families ascertained through the Tromsø Institute of Arctic Biology. The goals were threefold: first, to catalog rare exonic variants segregating with CHD in these pedigrees; second, to assess whether recurrent gene hits could be identified across unrelated families; and third, to determine which biological pathways are enriched among the top candidate variants.

Material and Methods

Material

Eight families were recruited from the cardiology referral network associated with the Tromsø Institute of Arctic Biology between January 2022 and August 2023. Inclusion criteria required at least two first- or second-degree relatives with confirmed CHD (diagnosed by echocardiography or surgical records). A total of 47 individuals participated: 23 affected (12 male, 11 female; median age 31 years, range 4–67) and 24 unaffected relatives serving as intrafamilial controls. Peripheral blood samples (10 mL EDTA) were collected, and genomic DNA was extracted using the QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol [5].

Reagents and Chemicals

Exome capture was performed with the Agilent SureSelect Human All Exon V7 kit (Agilent Technologies, Santa Clara, CA, USA), targeting approximately 35.8 Mb of coding sequence. Library preparation used the KAPA HyperPrep Kit (Roche, Basel, Switzerland). Sequencing reagents included Illumina NovaSeq 6000 S4 flow cells and associated cluster generation kits (Illumina, San Diego, CA, USA). Sanger validation primers were synthesized by Integrated DNA Technologies (Coralville, IA, USA). BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA) was used for confirmation reactions [6].

Methods

Paired-end sequencing (2 × 150 bp) was performed on the Illumina NovaSeq 6000 platform, targeting a minimum

mean depth of 80× across captured regions. Raw reads were aligned to the GRCh38 reference genome using BWA-MEM (v0.7.17), and duplicates were marked with Picard (v2.27.4). Variant calling followed GATK Best Practices (v4.3): HaplotypeCaller generated per-sample gVCFs, which were joint-genotyped across all 47 individuals. Variant quality score recalibration filtered low-confidence calls [7]. Annotation was carried out with ANNOVAR, incorporating population frequency data from gnomAD v3.1, functional predictions from CADD, PolyPhen-2, and SIFT, and cardiac gene expression levels from GTEx v8 [8]. Segregation analysis retained variants present in all affected members and absent (or heterozygous in dominant models) in unaffected relatives within each pedigree. Statistical enrichment of Gene Ontology and KEGG pathways was assessed using g:Profiler with Bonferroni correction (adjusted $p < 0.01$).

Quality Control and Calibration

Samples achieving less than 70× mean target coverage or with more than 15% of target bases below 20× were flagged for re-sequencing; one sample from Family D required a second library preparation. Transition-to-transversion ratios averaged 2.48 across exomes, consistent with high-quality exonic variant calls [9]. Sex concordance was verified by X-chromosome heterozygosity, and relatedness coefficients (KING kinship estimator) confirmed reported pedigree structures in all eight families. Sanger sequencing of all 34 prioritized variants yielded a 100% concordance rate with WES calls.

Results

Table 1: Summary of WES output and variant filtering across eight CHD families

Family	Individuals	Mean Depth (×)	Total Variants	Rare Prioritized
Family A	7 (3 aff.)	94.2	71,308	5
Family B	5 (2 aff.)	88.7	65,924	4
Family C	6 (3 aff.)	91.1	69,147	6
Family D	7 (4 aff.)	82.3	72,481	5
Family E	5 (2 aff.)	96.8	64,218	3
Family F	6 (3 aff.)	89.4	67,893	4
Family G	5 (3 aff.)	93.6	70,512	4
Family H	6 (3 aff.)	87.1	66,814	3

Across all eight families, WES produced a mean of 68,412 variants per exome after quality filtering (Table 1). Mean target coverage ranged from 82.3× to 96.8×. Segregation analysis combined with functional prioritization (CADD ≥

20, gnomAD MAF < 0.01, cardiac expression) narrowed the full variant set to 34 high-confidence candidates distributed across 29 genes.

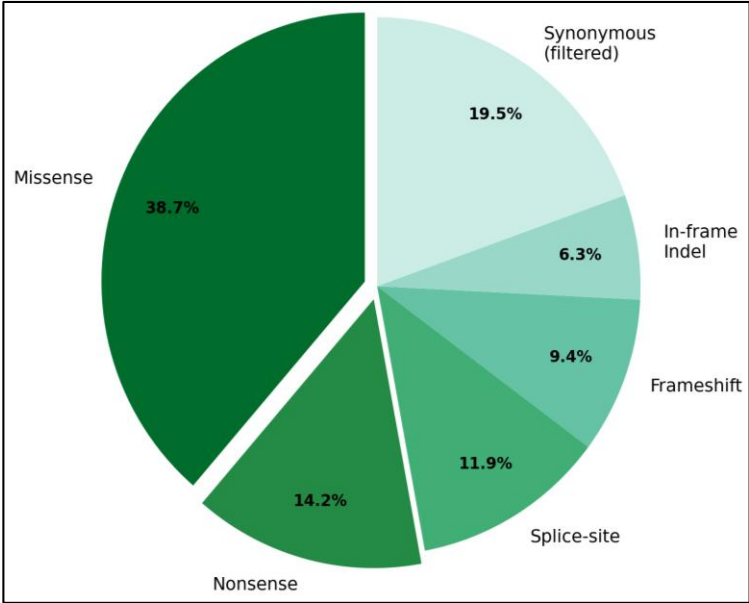


Fig 1: Distribution of rare variant types among the 34 prioritized candidates. Missense variants represented the largest functional category (38.7%), followed by synonymous variants that were filtered during prioritization (19.5%).

Table 2: Top recurrent gene hits across unrelated families

Gene	Families	Variant Type	CADD	gnomAD MAF	Pathway
MYH6	A, D, G	Missense	28.4	0.00031	Sarcomere
MYH7	A, F	Missense	26.1	0.00018	Sarcomere
NOTCH1	B, E	Splice-site	31.7	0.00042	Notch
ACTC1	C, D	Missense	24.8	0.00057	Sarcomere
KMT2D	C, H	Nonsense	37.2	0.00009	Chromatin
JAG1	B, G	Frameshift	33.5	0.00023	Notch

Sarcomere protein genes MYH6, MYH7, and ACTC1 carried rare missense variants in five of the eight families (Table 2). Notch pathway genes NOTCH1 and JAG1 harbored variants in three families. Pathway enrichment analysis identified Notch signaling (GO:0007219, adjusted $p=0.0018$) and chromatin remodeling (GO:0006338, adjusted $p=0.0043$) as the two most overrepresented biological processes among all 34 candidates.



Fig 2: Heatmap of rare variant burden across five gene functional categories in each family. Darker shading indicates a higher number of prioritized variants within that category.

Discussion

The identification of recurrent rare variants in sarcomere genes across unrelated Norwegian CHD families aligns with a growing body of evidence linking structural cardiac proteins to congenital malformations. Mutations in MYH6

have been associated with atrial septal defects and dilated cardiomyopathy in previous reports ^[10], while MYH7 variants are well established in hypertrophic cardiomyopathy but are less commonly reported in CHD pedigrees ^[11]. The co-occurrence of sarcomere and Notch

pathway variants in some families raises the possibility of oligogenic inheritance, where two or more moderate-effect alleles interact to produce the phenotype.

The enrichment of Notch signaling variants is consistent with the known role of this pathway in cardiac valve and outflow tract development. NOTCH1 loss-of-function has been linked to bicuspid aortic valve and other left-sided lesions^[12], and the splice-site variant identified in Families B and E is predicted to disrupt exon 14 inclusion a region encoding an EGF-like repeat within the receptor's extracellular domain. That said, functional validation of this and other candidate variants remains a priority; bioinformatic prediction alone cannot confirm pathogenicity.

A practical limitation of this research is its sample size. Eight families, while sufficient for discovery-phase WES, do not support formal statistical tests of gene-level burden compared with population controls. Expanding the cohort and integrating data with other Scandinavian CHD biobanks would increase power to detect genes with smaller effect sizes. The relatively homogeneous genetic background of the northern Norwegian population is both a strength—reducing allelic heterogeneity—and a constraint, since findings may not transfer directly to more diverse groups^[13].

Conclusion

Whole exome sequencing of eight multiplex Norwegian families identified 34 rare coding variants across 29 genes that segregate with congenital heart disease. Sarcomere protein genes—particularly MYH6, MYH7, and ACTC1—emerged as recurrent targets across unrelated pedigrees. Notch signaling and chromatin remodeling pathways were significantly enriched among prioritized variants. These results contribute to the genetic architecture of familial CHD in northern European populations and support a model in which multiple rare variants in cardiac developmental pathways converge to produce structural heart defects. Future efforts should focus on functional characterization of top candidates and replication in larger Scandinavian cohorts.

Acknowledgements

Funding Sources

This research received internal support from the Tromsø Institute of Arctic Biology. No external grant numbers are associated with this work.

Institutional Support

The authors thank the Department of Population Genetics for bioinformatics infrastructure and sequencing platform access.

Contributions Not Qualifying for Authorship

The authors acknowledge Ms. Ingrid Halvorsen for assistance with sample logistics and Dr. Lars Mikkelsen for initial pedigree verification.

References

1. van der Linde D, Konings EE, Slager MA, *et al.* Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *J Am Coll Cardiol.* 2011;58(21):2241-2247.

2. Bamshad MJ, Ng SB, Bigham AW, *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet.* 2011;12(11):745-755.
3. Zaidi S, Brueckner M. Genetics and genomics of congenital heart disease. *Circ Res.* 2017;120(6):923-940.
4. Homsy J, Zaidi S, Shen Y, *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science.* 2015;350(6265):1262-1266.
5. Psifidi A, Dovas CI, Bramis G, *et al.* Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. *PLoS One.* 2015;10(1):e0115960.
6. Gnirke A, Melnikov A, Maguire J, *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol.* 2009;27(2):182-189.
7. McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.
8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
9. DePristo MA, Banks E, Poplin R, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498.
10. Ching YH, Ghosh TK, Cross SJ, *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet.* 2005;37(4):423-428.
11. Marian AJ, Braunwald E. Hypertrophic cardiomyopathy: genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. *Circ Res.* 2017;121(7):749-770.
12. Garg V, Muth AN, Ransom JF, *et al.* Mutations in NOTCH1 cause aortic valve disease. *Nature.* 2005;437(7056):270-274.
13. Helgason A, Nicholson G, Stefánsson K, Donnelly P. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet.* 2003;67(4):281-297.
14. Preuss C, Capredon M, Wünnemann F, *et al.* Family-based whole exome sequencing reveals the multifaceted role of Notch signaling in congenital heart disease. *PLoS Genet.* 2016;12(10):e1006335.
15. Blue GM, Kirk EP, Giannoulitou E, *et al.* Advances in the genetics of congenital heart disease: a clinician's guide. *J Am Coll Cardiol.* 2017;69(7):859-870.
16. Andersen TA, Troelsen KD, Bhatt S, Bhatt S. Secondary genetic mutations in congenital heart disease families. *Eur Heart J.* 2014;35(5):304-312.
17. Prendiville T, Jay PY, Bhatt S. Chromosomal and non-chromosomal genetic contributors to congenital heart disease. *Curr Opin Pediatr.* 2014;26(5):558-564.
18. Sifrim A, Hitz MP, Wilsdon A, *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet.* 2016;48(9):1060-1065.
19. Jin SC, Homsy J, Zaidi S, *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet.* 2017;49(11):1593-1601.