Genetics in Medicine (2023) 25, 100012

ELSEVIER

Genetics in Medicine
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine

ARTICLE

Cardiomyopathy prevalence exceeds 30% in individuals with *TTN* variants and early atrial fibrillation

Kelly M. Schiabor Barrett[1,*], Elizabeth T. Cirulli[1], Alexandre Bolze[1], Chris Rowan[2,3], Gai Elhanan[2,4], Joseph J. Grzymski[2,4], William Lee[1], Nicole L. Washington[1]

ARTICLE INFO

Article history:
Received 25 August 2022
Received in revised form
4 January 2023
Accepted 5 January 2023
Available online 10 January 2023

Keywords:
Atrial fibrillation
Dilated cardiomyopathy
Population genomic screening
Rare variants
TTN

ABSTRACT

Purpose: *TTN* truncating variants (*TTN*tvs) represent the largest known genetic cause of dilated cardiomyopathies (DCMs), however their penetrance for DCM in general populations is low. More broadly, patients with cardiomyopathies (CMs) often exhibit other cardiac conditions, such as atrial fibrillation (Afib), which has also been linked to *TTN*tvs. This retrospective analysis aims to characterize the relationship between different cardiac conditions in those with *TTN*tvs and identify individuals with the highest risk of DCM.

Methods: In this work we leverage longitudinal electronic health record and exome sequencing data from approximately 450,000 individuals in 2 health systems to statistically confirm and pinpoint the genetic footprint of *TTN*tv-related diagnoses aside from CM, such as Afib, and determine whether vetting additional significantly associated phenotypes better stratifies CM risk across those with *TTN*tvs. We focused on *TTN*tvs in exons with a percentage spliced in >90% (hiPSI *TTN*tvs), a representation of constitutive cardiac expression.

Results: When controlling for CM and Afib, other cardiac conditions retained only nominal association with *TTN*tvs. A sliding window analysis of *TTN*tvs across the locus confirms that the association is specific to hiPSI exons for both CM and Afib, with no meaningful associations in percent spliced in ≤90% exons (loPSI *TTN*tvs). The combination of hiPSI *TTN*tv status and early Afib diagnosis (before age 60) found a subset of *TTN*tv individuals at high risk for CM. The prevalence of CM in this subset was 33%, a rate that was 3.5 fold higher than that in individuals with hiPSI *TTN*tvs (9% prevalence), 5-fold higher than that in individuals without *TTN*tvs with early Afib (6% prevalence), and 80-fold higher than that in the general population.

Conclusion: Our retrospective analyses revealed that those with hiPSI *TTN*tvs and early Afib (~1/2900) have a high prevalence of CM (33%), far exceeding that in other individuals with *TTN*tvs and in those without *TTN*tvs with an early Afib diagnosis. These results show that combining phenotypic information along with genomic population screening can identify patients at higher risk for progressing to symptomatic heart failure.

Kelly M. Schiabor Barrett and Elizabeth T. Cirulli contributed equally.

*Correspondence and requests for materials should be addressed to Kelly M. Schiabor Barrett, 101 S. Ellsworth Avenue, Suite 350, San Mateo, CA 94401.
E-mail address: kelly.barrett@helix.com
Affiliations are at the end of the document.

doi: https://doi.org/10.1016/j.gim.2023.100012

Introduction

Titin is a multiuse protein with spring-like function vital for cardiac contraction, and *TTN* gene variants have long been studied in relation to dilated cardiomyopathies (DCMs; OMIM 604145). Indeed, case-control studies have found rare *TTN* truncating variants (*TTN*tvs) in 15% to 20% of DCM cases, exceeding frequencies in matched controls.[1] Recently, the American College of Medical Genetics and Genomics upgraded *TTN* into the list of genes to screen for secondary findings in the context of clinical exome sequencing. This upgrade was based on evidence that *TTN*tvs found in the exons of heart-specific transcripts (percent spliced in >90% [hiPSI] *TTN*tvs) pose above average risk for DCM.[2-4] Although there is strong evidence linking *TTN*tvs to DCM, *TTN* has historically been left out of more broad population screening discussions because *TTN*tvs have variable penetrance in this context.[5] Furthermore, *TTN*tvs are pervasive (roughly 1%-2% of individuals) in unselected population cohorts and penetrance estimates for DCM are low, even when qualifying variants are confined to well-described hotspots, limiting the personal utility of this genetic information.[3,6] A more nuanced understanding of this gene–disease relationship is needed if we are to use *TTN* genetic data in population screening to quantify future heart failure (HF) risk and inform clinical care.

In gene–disease relationships in which loss of function (LoF) is a known mechanism of disease, it is often the case that within a definable range of the transcript, variants with similar molecular consequence confer similar disease risk. These patterns (akin to the PVS1 criterion in the American College of Medical Genetics and Genomics variant interpretation framework), once established and validated, allow for effective phenotype-agnostic variant interpretation and make population genomic screening of rare variation for disease risk possible.[7] For example, rare, LoF variants in *LDLR*, outside of the last exon, are routinely interpreted as pathogenic for familial hypercholesterolemia, even if the particular variant has not been observed clinically.[8,9]

TTN is genetically complex—a gene with 364 exons and 4 major structural regions (Z disk, I band, A band, and M band), each with largely independent functional roles and alternative splicing both within and across these domains. Early *TTN* association studies comparing DCM cases to population variant frequencies associated DCM with truncating variants specifically in the A-band region of the protein.[1] However, although odds ratios (ORs) remain strongest when associations are limited to A-band variants, subsequent studies have shown that DCM risk is not exclusive to this region. The percentage spliced in (PSI) index, a calculation of exon-based cardiac expression patterns, has been leveraged to bucket *TTN* exons across all domains by their potential relevance to cardiac phenotypes.[10] Population-level genetic associations remain significant when qualifying variants are defined as *TTN*tvs in constitutively expressed hiPSI exons (PSI > 90%).[3,4,10] Debate remains as to what extent risk extends to alternatively expressed, as opposed to constitutively expressed, cardiac exons, represented as non-hiPSI exons with intermediate PSI scores (PSI > 15%).[11] Establishing which regions of the *TTN* gene show statistically significant associations with relevant phenotypes and comparing these findings to expression-based expectations would help prioritize the genetic information that is relevant to return to patients in a genomic screening setting.

In heterogeneous population cohorts, as opposed to disease-specific cohorts, low penetrance estimates for those with *TTN*tvs may also be the result of applying a too-narrow disease phenotype to define cases. As we begin to sequence more broadly in populations, our understanding of genetic disease is changing—classic disease definitions may not completely capture population level clinical manifestations related to genetic variation at a locus.[12] Relevant diagnoses may also be different depending on when in the disease course a person is assessed and first treated. For example, patients with a *TTN*tv may present later in life with DCM-related HF. By including other common symptoms such as arrhythmias and other elements of stage B HF, it may be possible to detect the manifestations of this heart disease earlier and in a younger cohort. Indeed, *TTN*tv in hiPSI transcripts have also been associated with early onset atrial fibrillation (Afib), which may overlap with the well-established DCM association.[13-15] Determining how the phenotypic footprint of *TTN*tvs extends beyond DCM may improve population-based retrospective penetrance estimates.

In this work, we leveraged 2 population cohorts, the UK Biobank (UKB) and the Healthy Nevada Project (HNP), to evaluate the interplay between individually significant *TTN* gene–disease associations, including DCM and Afib. We then used statistical analysis to vet variant interpretation models for *TTN*tvs across the gene and finally quantify CM risk. Together, our study provides real-world evidence to support population screening for *TTN*tvs in relation to CM-driven HF.

Abbreviations

Afib – atrial fibrillation
CM – cardiomyopathy
DCM – dilated cardiomyopathy
EHR – Electronic Health Record
GTEx – genotype tissue expression
HF – heart failure
hiPSI – percent spliced in >90%
HNP – Health Nevada Project
ICD – International Classification of Diseases
LoF – loss of function
LOFTEE – Loss-Of-Function Transcript Effect Estimator
loPSI – percent spliced in ≤90%
PVS1 – pathogenic very strong criterion 1 from the ACMG variant interpretation rubric
*TTN*tvs – *TTN* truncating variants
UKB – UK Biobank

Materials and Methods

Subjects and genetic data

We used the UKB plink-formatted population level exome original quality functional equivalent exome files for N = 450,000 individuals (field 23155, with genotypes set to missing when read depth was <7 for single-nucleotide variations and <10 for indels, and variants excluded if there were no homozygotes or the maximum allelic balance was <0.15 for single-nucleotide variations or <0.2 for indels as per Backman et al[16]) and the imputed genotypes from genome-wide association studies genotyping (field 22801-22823) (UKB450K, n = *428,009 participants after mathcing to clinical records*). We also used N = 32,689 HNP samples (HNP32K) that were sequenced and analyzed at Helix using the Exome+ assay and matched to clinical records as previously described.[17] For the UKB cohort, participants ranged in age, as of 2021, from 50 to 87 and were 55% female, whereas the HNP age range was from 18 to 89+ and was 68% female. The UKB was of 83% British European ancestry, with another 10% other European ancestry and 7% other ancestries, and the HNP was 77% general European ancestry, 14% Hispanic ancestry, and 9% other ancestries. No filtering was applied to the cohorts based on ancestry in the main study, and confirmatory analyses were performed in the European ancestry subset.

Translating medical records to phecodes

HNP phenotypes were processed from Epic/Clarity electronic health records (EHRs) data as previously described and updated as of March 2021.[17] UKB data were provided from the UKB resource (accessed July 2021). For HNP, International Classification of Diseases (ICD), ninth (ICD-9) and tenth revision codes (ICD-10) and associated dates were collected from available diagnosis tables (from problem lists, medical histories, admissions data, surgical case data, account data, claims, and invoices). For UKB, ICD codes and associated dates (both ICD-9 and ICD-10) were collected from inpatient data (category 2000), cancer register (category 100092), and the first occurrences (category 1712), which records earliest occurrences of ICDs from the general practice, inpatient data, and death data at 3 digit resolution (ie, I42 instead of I42.0 or I42.9). This means that for some I42 codes, the full resolution ICD was not available. We used phecodes—curated groupings of ICD codes—to reduce phenotype complexity from >20,000 ICD codes to 1044 medically relevant phenotypes from available EHR records, specifically, ICD-9 (Phecode Map 1.2, used for both cohorts), ICD-10 (Phecode Map 1.2b to ICD-10 beta, used for UKB), and ICD-10-CM (Phecode Map 1.2b to ICD-10-CM beta, used for HNP).[18,19] When multiple ICDs from the same phecode were present for an individual, the earliest date (to year-level resolution) was used to represent that phecode.

For this study, the main phenotypes of interest were DCM, which is captured as part of phecode 425.1, and Afib, which is phecode 427.2. Again, phecodes are combinations of ICD codes that together are each meant to represent clinically meaningful diseases and conditions from EHR records.[18] This means that in our study, analysis of DCM also included some less-common CM categories (phecode 425.1 includes not only ICD10 code I42.0, DCM [n = 1181 cases; P value for association with *TTN*tv = $1.2e^{-66}$], but also I42.3, endomyocardial [eosinophilic] disease n = 2 cases; P = not applicable], I42.4, endocardial fibroelastosis [n = 16; P = NA]; I42.5, other restrictive CM [n = 19; P = NA]; I42.8, other CMs [n = 254; P = .01]; I42.9, unspecified CM [n = 1009; $P = 1.4e^{-19}$]; and the truncated I42 phenotype from the first occurrences table [n = 614 not in other I42 categories; P = .001]). To be as accurate as possible in our reporting we describe our findings in terms of CM risk instead of solely DCM. The ICD codes, I11.0 (hypertensive heart disease with HF) and 125.1(atherosclerotic heart disease of native coronary artery), which may capture DCM less specifically, did not show significant associations with *TTN* (P = .03 and P = .99, respectively) and were not used in this study.

For clinical analyses of diagnostic patterns, EHR ICD tables were updated for newest diagnosis data as of September 2022, the phecode mappings were updated, and the cohorts were further filtered down to only include individuals with more than one year of diagnosis history assessed by comparing the earliest and latest dates of any ICD code on record. To use as much ICD specificity as possible, 3 digit ICD entries from the first occurrences table were only used when the source of the code was the general practice table (where no better resolution was available). This improvement removed any I42 truncations attributed to ICD I42.1 (obstructive hypertrophic CM) and 142.2 (other hypertrophic CM) stemming from either the inpatient or death data. Analyses of the specific I42.1 and I42.2 codes showed no associations with *TTN*tv in these subsets, as expected (P = .94 and .29, respectively). These filtered cohorts included N = 32,689 HNP (HNP32K) and n = 396,690 UKB (UKB400K) participants, a more specific subset of the cohort used for initial genetic analyses.

Annotation and PSI

Variant annotation was performed using Ensembl Variant Effect Predictor-99.[20] Coding regions were defined according to Gencode version GENCODE 33, and the Ensembl canonical transcript ENST00000589042.5 was used to determine variant consequence.[21,22] However, the exon numbering we used included exon 48, which is not in the canonical transcript.[4] Variants from the exome were restricted to coding sequence regions plus essential splice

sites. Genotype processing for HNP data was performed in Hail 0.2.54-8526838bf99f (https://github.com/hail-is/hail/releases/tag/0.2.21).

We defined *TTN*tv as LoF variants (stop_lost, start_lost, splice_donor_variant, frameshift_variant, splice_acceptor_variant, or stop_gained). Variants were only included if their minor allele frequency was <0.1% in all Genome Aggregation Database populations and locally within each population analyzed.[8] The variants included in the analysis and their annotations are shown in Supplemental Table 1.

PSI data were obtained from cardiodb.org.[4] We also annotated LoF variants as low confidence (LC) or high confidence (HC) according to Loss-Of-Function Transcript Effect Estimator (LOFTEE).[8] LOFTEE flags LoF variants as LC if the LoF version is the ancestral state, if they are a stop gain or frameshift near the end of the gene or are in an exon with noncanonical splice sites around it, or if they are a splice variant that is not predicted to affect the splicing of a coding exon. In this analysis, LOFTEE flagged 247 of the 1609 *TTN*tvs as LC for having an ancestral allele and labeled the other variants as HC (Supplemental Table 1). Manually applying the 50 basepair rule to remove an additional 10 *TTN*tvs at the 3′ end of the gene yielded no effect on the results.

Power Window

We used the statistical power-based sliding window analysis technique, "Power Window," to create a continuous pathogenicity model across the entire *TTN* locus, as previously described (Cirulli ET, Schiabor Barrett KM, Bolze A, Grzymski JJ, Lee W, Washington NL. A power-based sliding window approach to evaluate the clinical impact of rare genetic variants. 2022. doi:10.1101/2022.07.29.22278171). In brief, this sliding window analysis groups variants located near each other into 1 unit and analyzes them together to improve power, much like a gene-based collapsing analysis but at a smaller scale. Rather than sizing the sliding window by the number of variants or bases covered, the sliding window is moved to maintain roughly the same number of people with a rare *TTN*tv and thus the statistical power, within each window. When enough people have a single variant for it to have as much statistical power as one of the windows, it is not combined into a window but undergoes regression analysis on its own, and the window slides past it, continuing to group surrounding variants as appropriate.

Given the overall sample size and rate of CM and Afib diagnosis in the UKB cohort, any *TTN* analysis window with at least 40 individuals with *TTN*tvs in which the true OR is 1 has, respectively, a 99.7% (CM) and 99.5% (Afib) probability of having <3 and <7 cases with *TTN*tvs, and thus an observed OR < 12 and OR < 3.5. An OR of 12 for CM and 3.5 for Afib was therefore chosen as the cutoff for a window to be considered associated with the trait. These OR cutoffs are also consistent with the overall OR observed for hiPSI exons in this analysis more generally (CM: 12.2, 95% CI = 10.2-14.5; Afib: 3.3, 95% CI = 2.9-3.8). With 40 individuals with *TTN*tvs within each window, the power for discovery is the same for each window as our analysis slides across the gene. A smaller window size, with fewer individuals with *TTN*tvs, would have less power to identify associations, whereas a larger window size, with more individuals with *TTN*tvs, would have less ability to home in specific regions of the gene. Our model tested 1336 windows. Each window includes a mean of 20.5 variants (median = 22, range = 1-35) and 7.2 exons (median = 7, range = 1-22). Variants are a mean of 100.3 coding bases apart (median = 63, range = 0-759).

Statistical analysis

We used regenie for genetic association analyses as previously described in our studies relating collapsed rare variants to phenotypes in large biobanks.[17,23,24] We used a representative set of 184,445 coding and noncoding linkage disequilibrium-pruned, high-quality common variants from the imputed genotypes to build the whole genome regression model for the first step of regenie, an analysis method that accounts for relatedness, population stratification, and case-control imbalance.[17,23,24] The covariates we included were age, sex, age*sex, age*age, sex*age*age, and bioinformatics pipeline version as appropriate. For step 2 of regenie, sample genotypes from the exome data were coded as a 1 if they had a qualifying *TTN*tv and a 0 otherwise for each set of variants analyzed: either collapsed across the entire *TTN* gene, just for hiPSI or loPSI exons, for the specific window being analyzed via power window, or for single variants with at least 40 heterozygotes. We analyzed all ancestries together for the main analysis; Supplemental Table 2 shows the hiPSI *TTN*tv counts by ancestry and Supplemental Table 3 and Supplemental Figure 2 show the results for an all-ancestry analysis with 40 principal components (field 22009) added to the covariates and a European ancestry-only analysis with 10 European ancestry-specific PCs added to the covariates. The results with and without PCs were essentially the same in this case because we were analyzing very rare variants (minor allele frequency < 0.1%) collapsed together and because step 1 of regenie builds its whole genome regression model to account for population stratification. Conditional analyses were run the same way but added the phenotype in question, CM or Afib, as a covariate.

Other statistical analyses were run using the statsmodel package in python 3.7.7. For binary variables, logistic regression was used; for quantitative variables, linear regression was used after rank-based inverse normal transformation. For time to event analyses, the lifelines package was used.[25]

Compound heterozygotes

Although we did not have phased data, we identified likely compound heterozygotes as individuals who were

heterozygous for 2 *TTN*tvs. In the UKB450K, we identified 105 such individuals, 83 of whom were unlikely to be true compound heterozygotes because the 2 variants were within 10 basepairs of each other or seen together in multiple individuals (although 15 kilobase [kb] apart, chr2:178661989:GTTTTC:G was observed in combination with chr2:178677634:TG:T all 7 times that it was seen in the UKB450K; and although 120 kb apart, chr2:178534380:T:A was observed with chr2:17865-3262:T:A both times it was seen in the UKB450K: all of these variants were in loPSI exons). The remaining potential compound heterozygotes had their 2 variants at least 4 kb apart from each other, and their specific combination of variants was seen only in 1 individual in the UKB450K.

Results

*TTN*tv associations with cardio phenotypes are driven by Afib and CM

Our previous work examined rare LoF variants in *TTN* (equivalent to and hereafter *TTN*tvs) with a gene-based collapsing approach in 2 population-based cohorts (~25,000 individuals from the HNP and ~200,000 from the UKB). Through these analyses we identified genome-wide significant associations ($P < 1 \times 10^{-9}$) with 7 heart phenotypes (phecodes, Materials and Methods): primary/intrinsic CMs, HF not otherwise specified (NOS), Afib and flutter, congestive HF NOS, nonrheumatic mitral valve disorders, mitral valve disease, and tachycardia NOS.[24] Although DCM and Afib are the most well-described population-level associations, *TTN*tvs have been implicated in a spectrum of heart disease and subclinical functional measures, including each of the phenotypes identified in our analyses.[13,14,26-28] Although Afib is well represented by the compilation of ICDs underlying its phecode, DCM is grouped into a slightly broader CM phenotype. We used this broader CM phenotype going forward in this study as opposed to only DCM specifically because there was an enrichment for those with *TTN*tv with this aggregate phenotype even after removing the strictly DCM cases ($P = 4.7e^{-10}$; see Materials and Methods for each independent ICD association in the CM phecode group). Importantly, the ICDs in the the larger CM phecode exclude codes for hypertrophic CM, secondary/extrinsic CMs, and CMs due to external agents (ie, alcohol, drugs, peripartum) all of which are represented by other phecodes.

Similar to previous reports, the penetrance estimates for each discrete phenotype, including the well-established DCM (as part of our CM phenotype) and Afib, were low (2%-9%), limiting the clinical utility of individual results for those with *TTN*tvs. Although each phenotypic association may represent a discrete disease, we wanted to test whether multiple associations reflect varying and compounding outcomes that result from underlying structural, functional, and metabolic changes stemming from *TTN*tvs.[3,13,15,29-31]

To test this hypothesis, we leveraged a larger data set of approximately 430,000 clinicogenomic records from the UKB (UKB450K). We found that 6 of the 7 previously associated individual phenotypes maintained genome-wide significant *P* values ($P < 1e^{-9}$) in an updated gene-based collapsing analysis, and the *P* value for tachycardia NOS was not statistically significant ($P = 1e^{-5}$; the *P* value of $5.1e^{-10}$ in our prior study was from a meta-analysis across cohorts) (Figure 1; Supplemental Table 3). The highest OR was for CM (OR = 5.5, $P = 3.5e^{-63}$), with the other ORs lying within the range 1.8 to 2.2. Given that DCM (as part of our CM phenotype) and Afib are already firmly established as associated with *TTN*tv,[3,14] we next conditioned on these 2 diagnoses to determine whether their associations were at least partially independent and whether other phenotypes had independent associations as well. We found that the association between CM and *TTN*tv was still statistically significant after controlling for Afib ($P = 1.9e^{-52}$), and the association between Afib and *TTN*tv was also still statistically significant after controlling for CM ($P = 4.2e^{-18}$). However, after controlling for both of these conditions, none of the other phenotypes maintained genome-wide significance. The best remaining association was with HF NOS ($P = 1e^{-5}$; other phenotype *P* values ranged from .02 to .002).

Overall, these conditional analyses showed that Afib and CM represent independent but overlapping associations with *TTN*tvs and support the hypothesis that these diagnoses may share underlying pathophysiology, such as impared sarcomere function, in individuals with *TTN*tvs.[13,14] Additional phenotypic associations (ie, mitral valve, HF, and tachycardia) while still enriched in those with *TTN*tvs, likely relate to *TTN* via the core diagnoses of CM and Afib. Future studies of *TTN*tvs with broad cardiac phenotyping may further our understanding of these trends. For the remainder of this study, we focused our analysis on the gene–disease relationship between CM, Afib, and *TTN*tvs in unselected population cohorts.

A sliding window statistical analysis confirms hiPSI variant prioritization for both Afib and CM risk

With confirmation of each association at the gene level, we next analyzed whether elevated Afib and CM risk localized to truncating variants in certain regions of the *TTN* locus. Given previous work on DCM, we expected that the CM signal would likely be localized to the constitutively expressed cardiac exons (hiPSI) and that variants in those exons would have a higher penetrance (Figure 2A). However, the hiPSI exon definition varies slightly based on the expression data set used to determine PSI values, with models derived from both Genotype-Tissue Expression (GTEx) data set and the left ventricular tissue from patients with DCM available (Supplemental Table 4).[3,4] Further, although debate remains around the mechanism of disease of *TTN*tvs in cardiac transcripts, there is evidence for

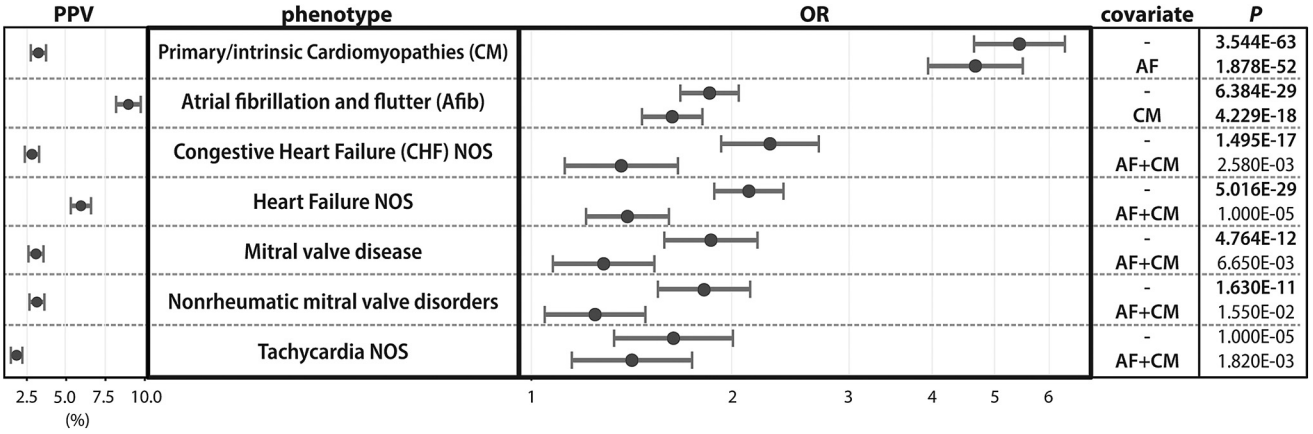

Figure 1 ***TTN* truncating variant genetic association analyses across phenotypes.** Plotted is the odds ratio (OR) and postivie predictive value (PPV) for the 7 phenotypes identified as significantly associated with *TTN* truncating variants in our prior studies ($P < 1 \times 10^{-9}$ in ~200,000 individuals).[26] The values shown and corresponding *P* values are updated based on the analysis of the 428,009 UK Biobank exomes with paired phenotypic data (UKB450K). Details of associations and European ancestry confirmation are presented in Supplemental Table 1. AF, atrial fibrillation; CM, cardiomyopathy; NOS, not otherwise specified; OR, odds ratio; PPV, positive predictive value; UKB450K, UK Biobank $n = 428{,}009$ participants.

haploinsufficiency and nonsense-mediated decay considerations (ie, LOFTEE) may improve *TTN*tv selection.[8,32]

To this end, we leveraged the clinical data to identify the regions of the gene where there was statistical evidence to support an association between *TTN*tvs and either CM or Afib. We identified these regions using a sliding window strategy to maintain roughly the same number of people with a rare *TTN*tv and thus the statistical power, within each window (Materials and Methods) (Cirulli ET, Schiabor Barrett KM, Bolze A, Grzymski JJ, Lee W, Washington NL. A power-based sliding window approach to evaluate the clinical impact of rare genetic variants. 2022. doi:10.1101/2022.07.29.22278171). Using this Power Window technique, an analysis of CM and Afib in the UKB450K cohort identified nearly all the hiPSI regions of the gene and almost no loPSI regions (Figure 2A). Specifically, in and near the Z band (exons 2-25, all hiPSI except 11 and 12), part of the I band (exons 29-53 and 217-252, all hiPSI except exons 217-219, 225, and 243), the A band (exons 253-358, all hiPSI), and part of the M band (part of hiPSI exon 359) all showed statistical enrichment for CM or Afib among the individuals with *TTN*tvs (Supplemental Table 4). This association was found despite the statistical analysis being blind to the PSI levels of each exon. This result showed that the signal for *TTN*tvs with both CM and Afib is driven by variants in the hiPSI exons across all major protein domains, matching the previously observed association between DCM and hiPSI *TTN*tvs.[3]

We also examined whether exons with PSI values indicative of alternative expression (PSI between 15% and 90%) were enriched for CM or Afib phenotypes using this sliding window method. We found that 77% and 90% of PSI > 90% exons overlapped a window with a significant association with CM and Afib phenotypes, respectively, whereas these numbers dropped to 3% to 15% of exons in other PSI ranges for these phenotypes (Supplemental Figure 1). We additionally found that the set of exons with PSI 15% to 90% as a whole were not significantly associated with Afib or CM ($P = .12$ and .04, respectively; Supplemental Figure 2). These results confirmed that a hiPSI cutoff of >90% is appropriate for exon selection for both CM and Afib.

As a further allelic characterization of the regions of *TTN* containing pathogenic variants, we looked at individuals who were homozygous for a *TTN*tv or were heterozygous for >1 *TTN*tv. In the UKB450K, we identified 16 individuals who were homozygous for a *TTN*tv, all in exons with PSI ≤ 90% (PSI range = 3-12%). Although we do not have phased data, we also identified individuals who were likely to be compound heterozygotes (Materials and Methods): we identified 9 individuals with *TTN*tvs in 2 different exons with PSI ≤ 90% (PSI range = 3-88%), and 13 individuals with 1 *TTN*tv in an exon with PSI > 90% (all were PSI 100%) and 1 in an exon with PSI ≤ 90% (PSI range = 5-88%; 4 at PSI ≥ 75%). However, no individuals with 2 *TTN*tvs in exons with PSI > 90% were observed despite the expectation, based on overall frequencies, that 10 such individuals would be found, further supporting the specificity of >90% as a hiPSI cutoff.

Next, we assessed whether hiPSI models calculated from expression data from the left ventricle of patients with DCM or from the heart tissue of GTEx participants were more predictive of CM or Afib at the population level. Using the now confirmed PSI >90% cutoff, we found that there was little difference in the magnitude and separation of ORs for hiPSI groupings stemming from each model (Figure 2B). With support for hiPSI models that are derived from either expression data source, we decided to use the maximum PSI value between these 2 sources for each exon to build a hiPSI model for population screening (Supplemental Table 4). Finally, we

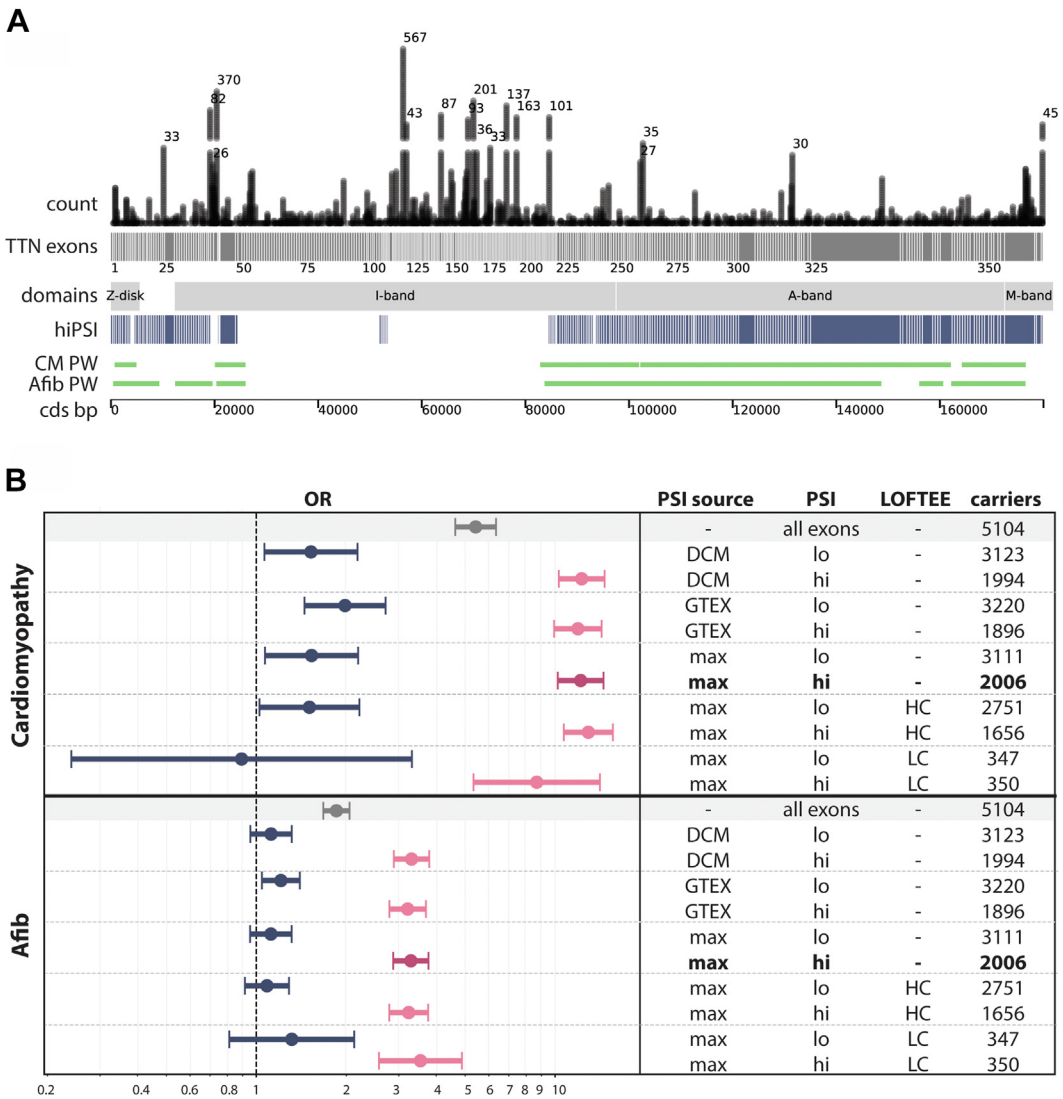

Figure 2 ***TTN* variants and regions that are associated with CM and Afib.** A. Individuals with *TTN* truncating variants (*TTN*tvs): shown above the *TTN* canonical exon track, each dot represents an individual and are stacked at a given position. For brevity, dots are trimmed to 35 and total number of individuals with *TTN*tvs is indicated when >40. Individuals are from the UKB450K cohort. *TTN* exons: canonical exons (gray; introns not to scale), with the exon number indicated below the track. Domains: major protein bands (Z disk, I band, A band, M band).[4] Cardio hiPSI: exons with PSI > 90% per exon, according to the maximum of either GTEx or patients with DCM for heart tissue (blue). PW: merger of all power windows with an OR outside the 99% confidence bound for a true OR of 1 when sampling 40 individuals with rare variants. Coding position indicated. B. Association between *TTN*tvs and CM or Afib. The whole gene is shown in gray, and the hiPSI exons (PSI > 90%) are shown in pink, whereas the loPSI exons (PSI ≤ 90%) are shown in blue. This was performed separately for PSI based on patients with DCM, GTEX (GTEx) data, or the maximum of those two. The variants are also broken up by whether LOFTEE categorized them as HC or LC. European ancestry confirmation is presented in Supplemental Figure 2. Afib, atrial fibrillation; bp, basepair; CDS, coding sequence; CM, cardiomyopathy; DCM, dilated cardiomyopathy; GTEX, Genotype-Tissue Expression; HC, high confidence; hiPSI, percent spliced in >90%; LC, low confidence; LOFTEE, Loss-Of-Function Transcript Effect Estimator; OR, odds ratio; PSI, percent spliced in; PW, Power Window; UKB450K, UK Biobank n = 428,009 participants.

assessed the utility of using LOFTEE in addition to a hiPSI exon filter in distinguishing pathogenic from benign *TTN*tvs. We found that hiPSI designation was the most important factor for pathogenicity, and that LOFTEE annotation did not meaningfully improve the discrimination of pathogenic *TTN*tvs from benign ones (Figure 2B, Supplemental Table 1).

Given these results, we performed our remaining analyses on all exons with >90% maximum PSI (hiPSI), disregarding LOFTEE status. Restricting to hiPSI exons removed 34% of the *TTN*tvs, assigning them to a likely benign status based on their location within loPSI exons and reclassified 61% of the individuals with *TTN*tvs because their variants were in loPSI exons. As expected, this refinement substantially improved both the P value (from $P = 3.5e^{-63}$ [CM] and $6.4e^{-29}$ [Afib] for the entire gene to $P = 1.1e^{-93}$ and $1.3e^{-52}$) and the positive predictive value (from 3% and 9% for the entire gene to 7% and 14%) (Figure 2; Supplemental Table 2).

Thirty-three percent of individuals with hiPSI *TTN*tvs with early Afib have a CM diagnosis

Although the penetrance for CM as an individual diagnosis, even when limiting to individuals with hiPSI *TTN*tvs, remained below 10% in the UKB400K, our genetic analyses suggested that the more prevalent Afib diagnosis (penetrance 14%) may also be relevant to consider. To that end, we next investigated whether accounting for Afib would identify a subset of hiPSI *TTN*tv individuals at highest risk for CM (basic cohort demographics and rates for CM and Afib phecodes used in this clinical analysis can be found in Supplemental Table 5).

We found that across all individuals with an Afib diagnosis, the prevalence of a CM diagnosis varied by *TTN* genotype, with the highest risk residing among those with hiPSI *TTN*tv and an Afib diagnosis before age 60 (early Afib).[13,14] Specifically, CM diagnoses in individuals with early Afib were recorded in 6% of individuals without *TTN*tvs and 4% of individuals with loPSI *TTN*tv as compared with 33% of individuals with hiPSI *TTN*tvs—a >5-fold increase in CM risk by genotype or a >80-fold increase in CM risk when compared with those with neither a hiPSI *TTN*tv nor an Afib diagnosis (the majority of the general population) (Figure 3A, $P = 1.7e^{-62}$ and $P = 1.4e^{-270}$, respectively, after controlling for age and sex). These trends replicated in the smaller HNP32K cohort—again, looking specifically at individuals with early Afib, those with hiPSI *TTN*tvs showed nearly a 4-fold increase in CM risk (44%) over those without *TTN*tvs (12%) and a >60-fold risk compared with those without hiPSI *TTN*tvs or Afib (Figure 3B, $P = 2.6e^{-4}$ and $P = 7.1e^{-21}$, respectively, after controlling for age and sex).

To determine whether the link between diagnoses is indicative of disease progression, we next used a time to event analysis to more thoroughly assess the temporal patterning of Afib and CM in individuals with and without a *TTN*tv from the UKB400K cohort. Assessed as individual phenotypes, both Afib and CM diagnoses showed an earlier onset and higher lifetime penetrance in individuals with a hiPSI *TTN*tv, with both sets of curves beginning to diverge as early as age 40 ($P < 1e^{-200}$ for Afib and $P = 1.42e^{-6}$ for CM, via log-rank test). Although there was a substantial hazard ratio (HR) for CM in individuals with hiPSI *TTN*tvs compared with individuals without a *TTN*tv (HR = 12.6, $P = 7.8e^{-195}$), as reported earlier in this article and previously in other work, the disease prevalence in both groups remains low across all ages—4% vs 0.5% at age 60 and 11% vs 1% at age 80 (Figure 4A, Supplemental Figure 3).

Aligning with the cumulative results presented in Figure 3, grouping individuals not only by *TTN*tv genetic status but also by Afib diagnosis amplifies the difference across the lifecourse between groups in age at onset, relative CM risk, and disease prevalence (Figure 4B). Confirming Afib as a general risk factor, those without a *TTN*tv but with an Afib diagnosis showed a roughly 8 times greater relative risk of CM than those without an Afib diagnosis (HR = 8.1 $P < 1e^{-200}$, Figure 4C). However, genetics further compounded the risk, because those with a hiPSI *TTN*tv and an Afib diagnosis showed an even greater, >50 times higher risk of CM compared with the same reference group (HR = 51.1, $P < 1e^{-200}$, Figure 4C). Strikingly, 27% of individuals with a hiPSI *TTN*tv had a recorded CM diagnosis by age 60, climbing to 37% by age 80, which may be used as an estimate for lifetime risk. When CM and Afib were diagnosed in the same person, Afib predated CM 44% of the time. Removing individuals for which CM occurred either before or concurrently with Afib produced similar results (Supplemental Figure 4, Supplemental Table 6).

Finally, these patterning trends remained when the analysis was expanded from an end point of CM to a more generalized end point of HF. Similar to CM, HF outcomes begin to diverge for individuals with a hiPSI *TTN*tv and an Afib diagnosis as early as age 40, with 31% exhibiting HF by age 60 and 54% by age 80, however the difference between hiPSI *TTN*tv cases with Afib and those without a *TTN*tv and with Afib was less pronounced (Supplemental Figure 5).

Taken together, our retrospective analysis of EHR diagnosis data supports a link between Afib and CM with likely progression to HF in those with hiPSI *TTN*tvs, a trend that may be especially relevant when Afib is diagnosed before age 60. Although *TTN*tv status alone has limited predictive value for CM, given the identified relationship with Afib, *TTN* genetic screening may still be a clinically meaningful screening tool, because monitoring those with hiPSI *TTN*tvs for signs of progressive HF such as Afib earlier in the life course may have clinical value.

Discussion

The link between *TTN* and DCM is well known, yet because of low penetrance estimates, *TTN* has not been suitable for general population screening. It is well-known that individuals often carry a range of structurally-related cardio diagnoses and that penetrance estimates across *TTN*tv-related diagnoses improve when relevant individuals are limited to only those with truncating variants in exons with constitutive expression in heart tissue (hiPSI exons). In this work, we focused on understanding the interplay between the various cardio diagnoses associated with *TTN*tvs and the underlying genetics for each association. We found that significant, population-level genetic associations remain for both DCM (as part of a CM phenotype) and Afib phenotypes when controlling for the other, whereas all other cardio associations can be largely explained by these 2 phenotypes. Using a statistical-based analysis across the locus we found that a hiPSI-based variant interpretation model faithfully represents both the CM and Afib associations and, in line with other reports, penetrance increases for both phenotypes (9% for CM and 14% for Afib) after subsetting to only those with hiPSI *TTN*tvs. Finally, to understand the connection between these diagnoses, we

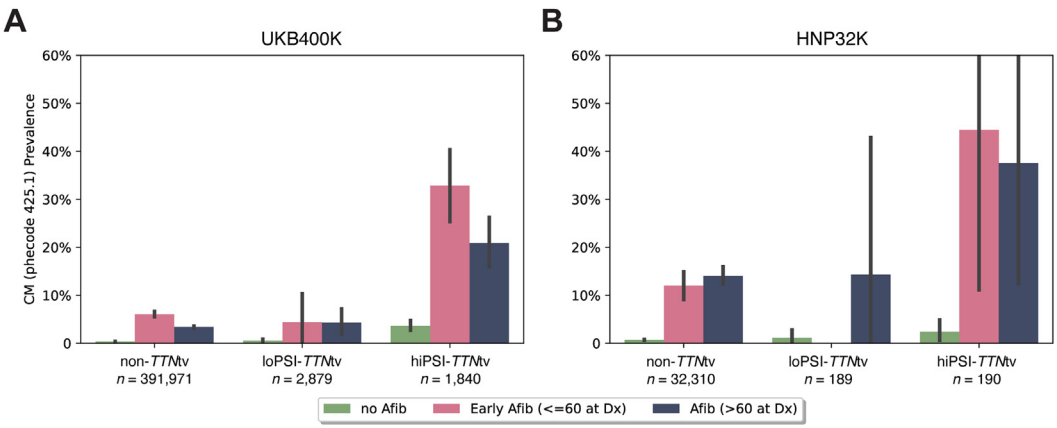

Figure 3 CM prevalence in UKB and HNP participants stratified by *TTN*tv status and Afib diagnosis. UKB (A) and HNP (B): all participants with International Classification of Diseases codes spanning >1 year in available electronic health records (UKB400K and HNP32K, respectively), separated by their *TTN* genotype status (non-*TTN*tv, loPSI, and hiPSI) and by the presence and, where relevant, onset of Afib. The penetrance and bootstrapped 95% CI for CM is plotted on the y-axis for each of these groupings. Although those with an early Afib diagnosis (pink bars) show the highest prevalence of CM, hiPSI *TTN*tvs are at roughly 5-fold greater risk for CM than any other group ($P = 1.7e^{-62}$ for UKB and $P = 2.6e^{-4}$ for HNP after controlling for age and sex). Afib, atrial fibrillation; CM, cardiomyopathy; hiPSI, percent spliced in >90%; HNP32K, Heatlhy Nevada Project $n = 32,689$ participants; loPSI, percent spliced in ≤90%; *TTN*tv, *TTN* truncating variant; UKB400K, UK Biobank $n = 396,690$ participants.

performed a retrospective analysis in individuals with and without hiPSI *TTN*tv. We found that when focusing on individuals with an early Afib diagnosis, 33% of those with hiPSI *TTN*tvs also had a diagnosis of CM, >5-fold enrichment compared with those without *TTN*tvs with early Afib and a 80-fold increase in CM risk compared with the general population. The temporal aspects of this relationship were then confirmed using a time to event style analysis.

Strategy for population screening for hiPSI *TTN*tvs

There are currently no screening recommendations for genetic CMs outside known family history. The high prevalence of CM in those with hiPSI *TTN*tvs with early Afib suggests that population screening for hiPSI *TTN*tvs (~0.5% of a cohort) and monitoring for Afib in these individuals may be one strategy to identify those at high risk of CM earlier in the disease course. Because early detection enables early intervention and because those with *TTN*tvs with DCM are known to respond well to standard of care interventions, [28] *TTN*tv screening in conjunction with routine monitoring may be an effective strategy to improve outcomes and to reduce the incidence of HF in the population.[32,33] Supporting this screening strategy, we found that Afib usually predates or is at least concurrently diagnosed with CM.

Afib and wearables

As wearables increase in popularity and function, awareness of Afib is likely to increase. Understanding heart disease risk in relation to an Afib finding from the wearable device will likely become a more common medical inquiry.[34] These findings can be hard to interpret as an individual data point in a patient, because they could represent one-off occurrences as opposed to a sign of elevated cardiovascular disease risk. Based on our findings surrounding clinically diagnosed Afib, CM, and *TTN*tvs, further studies comparing the relationship between Afib diagnoses collected via wearables and clinical-based assessments in those with hiPSI *TTN*tvs should be performed.

Limitations

One of the major limitations of this study (and most studies that rely on retrospective analysis of the EHR) is the lack of detailed phenotyping, especially of lower risk phenotypes, over a longitudinal timespan in the participants. Despite EHRs' comprehensive data sets, many data points are not recorded with the desired frequency, coding patterns vary by both physician and health system, and initiation dates of events are suboptimal. Penetrance estimates will likely be higher with more detailed evaluations. A lack of a standard screening strategy may at least partially explain the often concurrent documentation of Afib and CM diagnoses in our cohorts, a trend which was also reported in a recent epidemiologic study of all CMs.[29] For CM, penetrance and expressivity may also depend on environmental stressors including alcohol, chemotherapy, and pregnancy, which can put extra stress on the heart and could precipitate otherwise subclinical disease in some individuals with *TTN*tvs.[35] For Afib, there are many other risk factors such as environmental, lifestyle factors, and genetic effects from common variants. We were encouraged by the replication of the Afib and CM coding trends across 2 health system cohorts (Supplemenetal Table 5), and prospective follow-up studies with more detailed longitudinal phenotyping

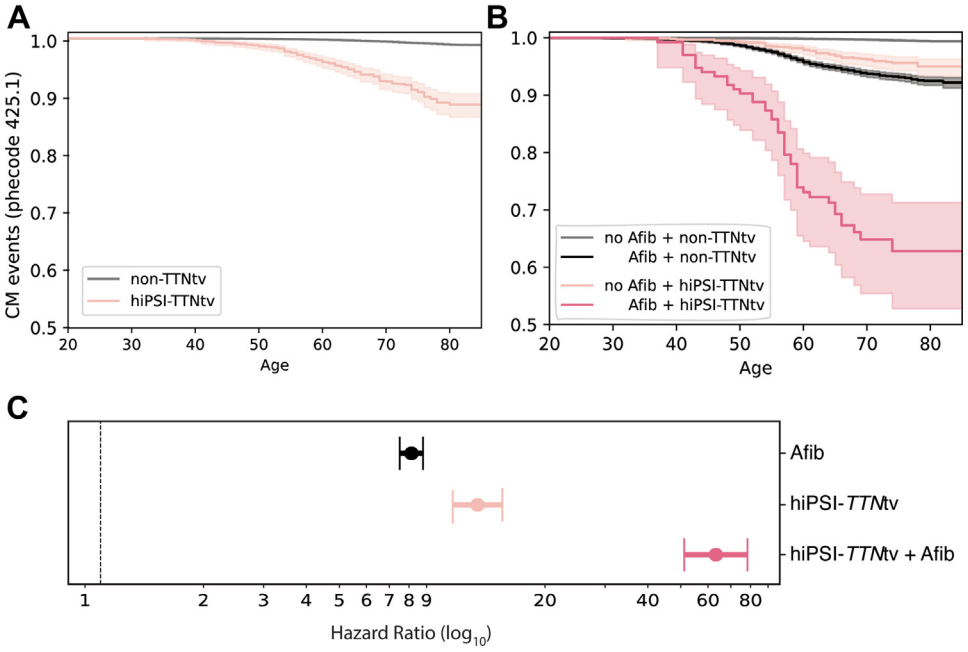

Figure 4 Time to event analysis and hazard ratios of CM events, stratified by hiPSI *TTN*tv and Afib status in UKB400K. Kaplan-Meier survival functions (A and B) for CM events by age at diagnosis (lines) with right censoring, including CIs (shadings) and hazard ratios (C) for the UKB with >1 year of International Classification of Diseases history (UKB400K). A. CM survival functions bifurcated by genotype hiPSI *TTN*tv (light pink) and those without a *TTN*tv (gray), with hazard ratios for those with a hiPSI *TTN*tv compared with those without a *TTN*tv (light pink) (panel C, $P < 7.83e^{-195}$). B. CM survival functions stratified by both genotypes, hiPSI *TTN*tv (pinks) vs non-*TTN*tvs (grays), and presence of any Afib diagnosis, which are further distinguished by color shades. (C) Hazard ratios for any Afib diagnosis (black) and hiPSI *TTN*tv with any Afib diagnosis (dark pink) compared with those with non-*TTN*tv and without Afib ($P < 1e^{-200}$ and $P < 1e^{-200}$, respectively). Afib, atrial fibrillation; CM, cardiomyopathy; hiPSI, percent spliced in >90%; *TTN*tv, *TTN* truncating variant; UKB400K, UK Biobank $n = 396{,}690$.

should be completed to confirm these retrospective trends. Finally, although our analysis showed similar trends between different ethnicities (Supplemenetal Table 2, Supplemenetal Figure 2), the limited sample sizes for non-European ancestries makes it difficult to definitively say whether there are ancestry-specific differences in the effects of variants in this gene, as previously suggested.[3]

Conclusion

Afib is a very common diagnosis (eg, 8% in UKB and 6% in HNP cohorts), and at current rates, it is difficult to vet what subset of individuals who present with this diagnosis are at risk for more severe heart-related complications in the future. Through this work, we provide evidence that an Afib diagnosis may represent early signs of a progressive HF phenotype, because individuals with *TTN*tv with early Afib (~1/2900) are enriched for subsequent CM and HF, often with earlier onset than nongenetic cases. All together, our analysis shows that population screening for *TTN* has the potential to contextualize Afib diagnoses and highlight those individuals who may warrant ongoing follow-up screening and care.

Data Availability

Statistics relating to the cardiomyopathy and atrial fibrillation Power Window analyses for all *TTN* exons calculated using the UKB450K are available in Supplemental Table 4. UK Biobank data are available for download (https://www.ukbiobank.ac.uk/) to qualified researchers. The Healthy Nevada Project data are available to qualified researchers upon request and with permission of the Institute for Health Innovation (IHI) and Helix. Researchers who would like to obtain the raw genotype data related to this study will be presented with a data user agreement, which requires that no participants will be reidentified and no data will be shared between individuals or uploaded onto public domains. The IHI encourages and collaborates with scientific researchers on an individual basis. Examples of restrictions that will be considered in requests to data access include but are not limited to (1) whether the request comes from an academic institution in good standing and will collaborate with our team to protect the privacy of the participants and the security of the data requested, (2) type and amount of data requested, (3) feasibility of the research suggested, and (4) amount of resource allocation for the IHI and Renown

Hospital required to support the collaboration. Any correspondence and data availability requests related to Healthy Nevada Project should be addressed to J.J.G. (Joe.Grzymski@dri.edu) or Craig Kugler (Craig.Kugler@dri.edu).

Acknowledgments

This research has been conducted using the UK Biobank resource under application number 40436. We acknowledge the entire Helix Bioinformatics team for their contributions to the production of exome sequencing pipeline. We thank all of the genomic representatives of the Healthy Nevada Project. We thank Renown Health and Desert Research Institute marketing for helping launch the Healthy Nevada Project.

Funding

Funding was provided to the Desert Research Institute by the Renown Institute for Health Innovation and the Renown Health Foundation.

Author Information

Conceptualization: K.M.S.B., E.T.C., N.L.W.; Formal Analysis: K.M.S.B., E.T.C., A.B.; Investigation: C.R., G.E.; Supervision: J.J.G., W.L., N.L.W.; Visualization: K.M.S.B., E.T.C., N.L.W.; Writing-original draft: K.M.S.B., E.T.C.; Writing-review and editing: K.M.S.B., E.T.C., N.L.W., A.B., C.R., G.E., J.J.G., W.L.

Ethics Declaration

The Healthy Nevada Project study was reviewed and approved by the University of Nevada, Reno Institutional Review Board (project 956068-12). The UK Biobank study was approved by the North West Multicenter Research Ethics Committee, United Kingdom. All participants gave their informed, written consent before participation. All data used for research were de-identified.

Conflict of Interest

K.M.S.B., E.T.C., A.B., W.L., and N.L.W. are employees of Helix. A patent has been filed by Helix for the Power Window analysis technique with E.T.C., K.M.S.B., and N.L.W. as inventors, and its current status is unpublished (application number 17575894). C.R., G.E., and J.J.G. claim no conflicts of interest.

Additional Information

The online version of this article (https://doi.org/10.1016/j.gim.2023.100012) contains supplementary material, which is available to authorized users.

Affiliations

[1]Helix, San Mateo, CA; [2]Renown Health, Reno, NV; [3]University of Nevada, School of Medicine, Reno, NV; [4]Center for Genomic Medicine, Desert Research Institute, Reno, NV

References

1. Herman DS, Lam L, Taylor MRG, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. 2012;366(7):619-628. http://doi.org/10.1056/NEJMoa1110186
2. Miller DT, Lee K, Chung WK, et al. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2021;23(8):1381-1390. Published correction appears in *Genet Med*. 2021;23(8):1582-1584. https://doi.org/10.1038/s41436-021-01172-3
3. Haggerty CM, Damrauer SM, Levin MG, et al. Genomics-first evaluation of heart disease associated with titin-truncating variants. *Circulation*. 2019;140(1):42-54. Published correction appears in *Circulation*. 2019;140(16):e727. https://doi.org/10.1161/CIRCULATIONAHA.119.039573
4. Roberts AM, Ware JS, Herman DS, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med*. 2015;7(270):270ra6. http://doi.org/10.1126/scitranslmed.3010134
5. McGurk KA, Zheng SL, Henry A, et al. Correspondence on "ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG)" by Miller et al. *Genet Med*. 2022;24(3):744-746. http://doi.org/10.1016/j.gim.2021.10.020
6. Akinrinade O, Koskenvuo JW, Alastalo TP. Prevalence of titin truncating variants in general population. *PLoS One*. 2015;10(12):e0145284. http://doi.org/10.1371/journal.pone.0145284
7. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018;39(11):1517-1524. http://doi.org/10.1002/humu.23626
8. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. Published correction appears in *Nature*. 2021;590(7846):E53. Published correction appears in *Nature*. 2021;597(7874):E3-E4. https://doi.org/10.1038/s41586-020-2308-7
9. Chora JR, Iacocca MA, Tichý L, et al. The Clinical Genome Resource (ClinGen) Familial hypercholesterolemia Variant Curation Expert Panel consensus guidelines for LDLR variant classification. *Genet Med*. 2022;24(2):293-306. http://doi.org/10.1016/j.gim.2021.09.012
10. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr Protoc Hum Genet*. 2015;87:11.16.1-11.16.14. http://doi.org/10.1002/0471142905.hg1116s87
11. Schafer S, de Marvao A, Adami E, et al. Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet*. 2017;49(1):46-53. http://doi.org/10.1038/ng.3719

12. Zeng C, Bastarache LA, Tao R, et al. Association of pathogenic variants in hereditary cancer genes with multiple diseases. *JAMA Oncol*. 2022;8(6):835-844. http://doi.org/10.1001/jamaoncol.2022.0373
13. Yoneda ZT, Anderson KC, Quintana JA, et al. Early-onset atrial fibrillation and the prevalence of rare variants in cardiomyopathy and arrhythmia genes. *JAMA Cardiol*. 2021;6(12):1371-1379. http://doi.org/10.1001/jamacardio.2021.3370
14. Choi SH, Weng LC, Roselli C, et al. Association between titin loss-of-function variants and early-onset atrial fibrillation. *JAMA*. 2018;320(22):2354-2364. http://doi.org/10.1001/jama.2018.18179
15. Butters A, Isbister JC, Medi C, et al. Epidemiology and clinical characteristics of atrial fibrillation in patients with inherited heart diseases. *J Cardiovasc Electrophysiol*. 2020;31(2):465-473. http://doi.org/10.1111/jce.14346
16. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599(7886):628-634. http://doi.org/10.1038/s41586-021-04103-z
17. Cirulli ET, White S, Read RW, et al. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun*. 2020;11(1):542. http://doi.org/10.1038/s41467-020-14288-y
18. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31(12):1102-1110. http://doi.org/10.1038/nbt.2749
19. Bastarache L, Hughey JJ, Hebbring S, et al. Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science*. 2018;359(6381):1233-1239. http://doi.org/10.1126/science.aal4043
20. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. http://doi.org/10.1186/s13059-016-0974-4
21. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47(D1):D766-D773. http://doi.org/10.1093/nar/gky955
22. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46(D1):D754-D761. http://doi.org/10.1093/nar/gkx1098
23. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103. http://doi.org/10.1038/s41588-021-00870-7
24. Schiabor Barrett KM, Bolze A, Ni Y, et al. Positive predictive value highlights four novel candidates for actionable genetic screening from analysis of 220,000 clinicogenomic records. *Genet Med*. 2021;23(12):2300-2308. http://doi.org/10.1038/s41436-021-01293-9
25. Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw*. 2019;4(40):1317. http://doi.org/10.21105/joss.01317
26. Pirruccello JP, Bick A, Chaffin M, et al. Titin truncating variants in adults without known congestive heart failure. *J Am Coll Cardiol*. 2020;75(10):1239-1241. http://doi.org/10.1016/j.jacc.2020.01.013
27. Jurgens SJ, Choi SH, Morrill VN, et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat Genet*. 2022;54(3):240-250. http://doi.org/10.1038/s41588-021-01011-w
28. Jansweijer JA, Nieuwhof K, Russo F, et al. Truncating titin mutations are associated with a mild and treatable form of dilated cardiomyopathy. *Eur J Heart Fail*. 2017;19(4):512-521. http://doi.org/10.1002/ejhf.673
29. Brownrigg JR, Leo V, Rose J, et al. Epidemiology of cardiomyopathies and incident heart failure in a population-based cohort study. *Heart*. 2022;108(17):1383-1391. http://doi.org/10.1136/heartjnl-2021-320181
30. Andreasen L, Bertelsen L, Ghouse J, et al. Early-onset atrial fibrillation patients show reduced left ventricular ejection fraction and increased atrial fibrosis. *Sci Rep*. 2020;10(1):10039. http://doi.org/10.1038/s41598-020-66671-w
31. Verdonschot JAJ, Hazebroek MR, Derks KWJ, et al. Titin cardiomyopathy leads to altered mitochondrial energetics, increased fibrosis and long-term life-threatening arrhythmias. *Eur Heart J*. 2018;39(10):864-873. http://doi.org/10.1093/eurheartj/ehx808
32. Fomin A, Gärtner A, Cyganek L, et al. Truncated titin proteins and titin haploinsufficiency are targets for functional recovery in human cardiomyopathy due to *TTN* mutations. *Sci Transl Med*. 2021;13(618):eabd3079. http://doi.org/10.1126/scitranslmed.abd3079
33. Akhtar MM, Lorenzini M, Cicerchia M, et al. Clinical phenotypes and prognosis of dilated cardiomyopathy caused by truncating variants in the *TTN* gene. *Circ Heart Fail*. 2020;13(10):e006832. http://doi.org/10.1161/CIRCHEARTFAILURE.119.006832
34. Williams BA, Chamberlain AM, Blankenship JC, Hylek EM, Voyce S. Trends in atrial fibrillation incidence rates within an integrated health care delivery system, 2006 to 2018. *JAMA Netw Open*. 2020;3(8):e2014874. http://doi.org/10.1001/jamanetworkopen.2020.14874
35. Giudicessi JR, Shrivastava S, Ackerman MJ, Pereira NL. Clinical impact of secondary risk factors in TTN-mediated dilated cardiomyopathy. *Circ Genom Precis Med*. 2021;14(2):e003240. http://doi.org/10.1161/CIRCGEN.120.003240