

RESEARCH ARTICLE

Prevalence of Titin Truncating Variants in General Population

Oyediran Akinrinade¹, Juha W. Koskenvuo^{2,3}, Tero-Pekka Alastalo^{1,2*}

1 Children's Hospital Helsinki, Institute of Clinical Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland, **2** Blueprint Genetics, Helsinki, Finland, **3** Department of Clinical Physiology and Nuclear Medicine, HUS Medical Imaging Center, Helsinki University Central Hospital and University of Helsinki, Finland

* tero-pekka.alastalo@helsinki.fi



OPEN ACCESS

Citation: Akinrinade O, Koskenvuo JW, Alastalo T-P (2015) Prevalence of Titin Truncating Variants in General Population. PLoS ONE 10(12): e0145284. doi:10.1371/journal.pone.0145284

Editor: Ralf Krahe, University of Texas MD Anderson Cancer Center, UNITED STATES

Received: June 16, 2015

Accepted: December 2, 2015

Published: December 23, 2015

Copyright: © 2015 Akinrinade et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by grants from the Sigrid Jusélius Foundation; the Finnish Foundation for Pediatric Research; the Finnish Foundation for Cardiovascular Research; the Finnish Cultural Foundation (00150083:OA); and the Ida Montin Foundation (OA). The funders Blueprint Genetics had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funder provided support in the form of salaries for authors [OA, JWK and TPA], but did not have any additional role in the study design, data collection and analysis, decision to publish, or

Abstract

Background

Truncating titin (*TTN*) mutations, especially in A-band region, represent the most common cause of dilated cardiomyopathy (DCM). Clinical interpretation of these variants can be challenging, as these variants are also present in reference populations. We carried out systematic analyses of *TTN* truncating variants (TTNtv) in publicly available reference populations, including, for the first time, data from Exome Aggregation Consortium (ExAC). The goal was to establish more accurate estimate of prevalence of different TTNtv to allow better clinical interpretation of these findings.

Methods and Results

Using data from 1000 Genomes Project, Exome Sequencing Project (ESP) and ExAC, we estimated the prevalence of TTNtv in the population. In the three population datasets, 52–54% of TTNtv were not affecting all *TTN* transcripts. The frequency of truncations affecting all transcripts in ExAC was 0.36% (0.32% - 0.41%, 95% CI) and 0.19% (0.16% - 0.23%, 95% CI) for those affecting the A-band. In the A-band region, the prevalences of frameshift, nonsense and essential splice site variants were 0.057%, 0.090%, and 0.047% respectively. *Cga/Tga* (arginine/nonsense-R/*) transitional change at CpG mutation hotspots was the most frequent type of *TTN* nonsense mutation accounting for 91.3% (21/23) of arginine residue nonsense mutation (R/*) at *TTN* A-band region. Non-essential splice-site variants had significantly lower proportion of private variants and higher proportion of low-frequency variants compared to essential splice-site variants ($P = 0.01$; $P = 5.1 \times 10^{-4}$, respectively).

Conclusion

A-band TTNtv are more rare in the general population than previously reported. Based on this analysis, one in 500 carries a truncation in *TTN* A-band suggesting the penetrance of these potentially harmful variants is still poorly understood, and some of these variants do not manifest as autosomal dominant DCM. This calls for caution when interpreting TTNtv in individuals and families with no history of DCM. Considering the size of *TTN*, expertise in

preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing Interests: The authors of this manuscript have read the journal's policy and have the following competing interest(s): JWK and TPA are co-founders of Blueprint Genetics. However, this does not alter the authors' adherence to PLOS ONE policies on sharing data and materials. The commercial affiliation [Blueprint Genetics] did not play any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DNA library preparation, high coverage NGS strategies, validated bioinformatics approach, accurate variant assessment strategy, and confirmatory sequencing are prerequisites for reliable evaluation of *TTN* in clinical settings, and ideally with the inclusion of mRNA and/or protein level assessment for a definite diagnosis.

Introduction

Titin (*TTN*) is a giant muscle protein expressed in the cardiac and skeletal muscles. It spans half of the sarcomere from Z-line to M-line. Titin is known to play a key role in muscle assembly, force transmission at the Z-line and maintenance of resting tension in the I-band region. The clinically relevant A-band of *TTN* binds to the thick filament, where it may regulate filament length and assembly, and is thought to be critical for biomechanical sensing and signaling. The C-terminal M-band contains a strain-sensitive kinase, which may have a role in cardiac signal transduction [1]. Before the era of next generation sequencing (NGS) only a small number of *TTN* gene mutations have been found to associate with cardiomyopathies [1–8]. This has largely been due to the difficulty to sequence large *TTN* gene with its ~363 exons. Consequently, *TTN* mutation frequency and therefore clinical impact were unknown.

Recently, using a combination of NGS and dideoxy sequencing, Herman et al. [9] estimated *TTN* truncating variants (TTNtv)—nonsense, frameshift and essential splice site, to be responsible for approximately 25% of familial cases of idiopathic dilated cardiomyopathy (DCM) and 18% of sporadic cases in a large cohort of subjects. Similarly to Herman et al. [9], in our recent study on Finnish patients with DCM, TTNtv were responsible for 20.6% of cases with family history of DCM. Furthermore, in both studies and others the mutations were not randomly distributed along the *TTN* gene as the bulk of the mutations were located predominantly in the A-band region and affecting all transcripts of *TTN* [10–14]. Another recent study confirmed that *TTN* truncations are highly enriched in DCM patients when compared to healthy controls [10]. Furthermore, identifying a truncating variant in the A-band region was estimated to have 93% risk of being disease causing. They also determined that C-terminal truncations affecting all transcripts were more pathogenic and mediated their effects through dominant negative mechanisms rather than haploinsufficiency. Roberts et al. [10] also confirmed that TTNtv-positive DCM patients manifest more severe clinical phenotypes than TTNtv-negative DCM patients. Recently, TTNtv have also been identified in patients with clinical features of both left ventricular non-compaction cardiomyopathy (LVNC) and dilated cardiomyopathy [15].

Utilization of reference population databases has significantly improved our ability to interpret genetic findings. Before the publishing of Exome Aggregation Consortium (ExAC) database in 2014, we have been dependent on the different versions of the 1000 Genomes Project and the National Heart Lung Institutes Exon Sequencing Projects (ESP) [16, 17]. Since the inception of 1000 Genomes in 2010, various phases, versions and releases of the project have evolved, with the phase 3 variants (which represents the final phase of the project) being based on data from 2535 individuals from 26 different populations around the world, with 60–100 individuals representing each population in the cohort. The quality and coverage of sequencing data in this database have varied significantly during the evolution of the database causing high prevalence of false positive insertion-deletions (INDELs) in earlier datasets. Recent effort by Golbus et al. [18] to estimate the prevalence of TTNtv in the population used the February 2012 release (phase 1 version 2) containing variants and phased genotypes across 1092

individuals from 14 different populations. In this study, the prevalence of frameshift INDELS that disrupt *TTN* was estimated to be as high as 3.2% (35/1092) among reference individuals. In the study by Roberts et al. [10], the prevalence of TTNtv was estimated 2% using a combination of 1000 Genomes call set (phase 1 version 3) and ESP call set.

Since the prevalence of TTNtv in reference populations is a critical determinant when interpreting genetic test results, we pursued to analyze this, for the first time, in the over 60 000 ExAC cohort together with the final version of 1000 Genomes project and ESP datasets. Our goal was also to define the specific prevalence of different TTNtv in various *TTN* domains to better understand the distribution and prevalence of different genetic findings. Our study reveals that identifying a frameshift, nonsense or an essential splice-site variant in the critical A-band of *TTN* is a rare event. In our study we conclude that in reference population only 1 in 1750 individuals carry a frameshift, 1 in 1100 carry a nonsense, and 1 in 2100 carry an essential splice variant in *TTN* A-band region that is estimated to affect all isoforms of the gene. We also determine the characteristics of nonsense variants identified and variability in the splice-site regions of *TTN* gene. We believe our results have a wide application in clinical interpretation of genetic test results and further emphasize that TTNtv in general population exceed significantly the prevalence of DCM associated with *TTN* truncations.

Methods

1000 Genomes Project Data

TTN gene variant data from the various phases and versions of 1000 Genomes Project were retrieved, and truncating variants comprising of splice, nonsense and frameshift variants were used for analysis.

Analysis of *TTN* Truncations in the Population

Considering the INDELS filtering on 1000 Genomes Project phase 1 variant call sets (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/README.phase1_integrated_release_version3_20120430) and the release of the phase 3 call set, we were keen to re-estimate the frequency of *TTN* truncating mutations in the population. We downloaded the exonic boundaries of *TTN* from the Ensembl Genome Browser (www.ensembl.org), and extracted exonic variants from the three versions of the phase 1 variants call sets (V1, V2, V3), and the version 5 (V5) of the phase 3 variants call set available in the 1000 Genomes Project database. Only calls that passed all quality filters were used for downstream analysis.

Exome Sequencing Project (ESP) and Exome Aggregation Consortium (ExAC) Data

In a bid to capture a complete map of population *TTN* truncations and to further filter likely false positives in the latest release of 1000 Genomes Project, we utilized variants reported in two other well-annotated databases—Exome Sequencing Project (ESP) [17] and ExAC [19], with over 6500 and 60700 individuals respectively. As 1000 Genomes Project cohorts and ESP cohorts were part of the ExAC project cohorts, we filtered *TTN* truncating variants unique to either 1000 Genomes Project or ESP call sets. Variants located in novex-specific exons and exons specific to few transcripts of *TTN* were filtered before filtering for sequencing coverage/depth. See below for more details on the workflow and [results](#).

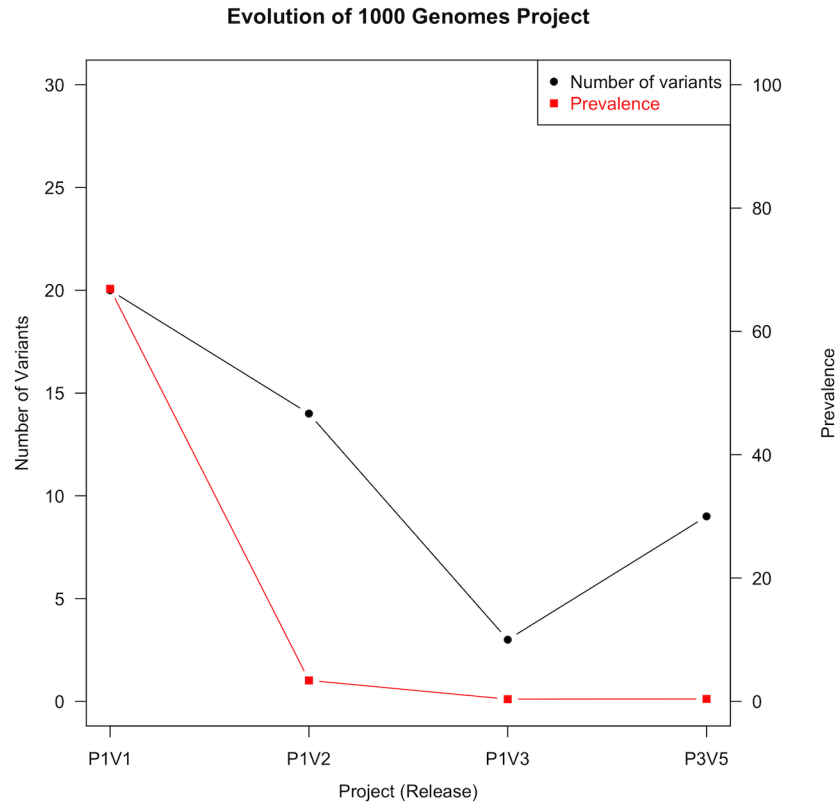


Fig 1. Influence of 1000 genomes project evolution on titin truncation prevalence. Abbreviations: P1V1 –phase 1 version 1; P1V2 –phase 1 version 2; P1V3 –phase 1 version 3; P3V5 –phase 3 version 5.

doi:10.1371/journal.pone.0145284.g001

Results

TTN Truncating Variants in the Evolution of 1000 Genomes Project

We extracted TTNtv that passed all quality filters as reported by the 1000 Genomes Project from various phases and versions of the project. Based on the total number of TTNtv in each version of the phase 1, and the version 5 of the phase 3 releases, we estimated the prevalence of TTNtv in the population. Interestingly, we observed a downward trend in both the number of TTNtv and TTNtv prevalence from the oldest to the latest release of the project (Fig 1). Version 2 of the phase 1 release contained predominantly rare TTNtv with high call qualities as against version 1 TTNtv composed of predominantly common TTNtv with very low qualities.

To demonstrate how the quality of sequencing and bioinformatic analysis contributes to our understanding of variant frequencies in general population, especially when estimating INDELS, we evaluated the detection rates of TTNtv in the evolution of 1000 Genomes project (see discussion) (Fig 1). The initial phase 1 (V1) integrated data contained 22 truncating mutations: 4 nonsense, 1 splice region, and 17 frameshift mutations (16 of which were single nucleotide (SN) INDELS) (S1 Table). After the first INDELS filtering on version 1 of the phase 1 integrated variant call set, six of the SN INDELS were filtered as false positives in the set due to technical artifacts introduced in the sequencing step. Version 2 phase 1 integrated variant call set contained 16 TTNtv: 4 nonsense, 1 splice region, and 11 frameshifts (10 SN INDELS) reported in Golbus et al. [18] study. All frameshift mutations as well as splice region variant

were filtered out as false positives in the version 3 of the phase 1 integrated variant call set (S2 and S3 Tables). We identified thirty-three TTNtv in the phase 3 integrated variant call set of the 1000 Genomes Project: 1 frameshift, 13 splice and 19 nonsense. Interestingly, a vast majority of the splice site and nonsense variants were located in either novex-3 specific exon, exons either not expressed or with low expression in human left ventricle (LV), described by Roberts et al. [10] as proportion spliced-in (PSI) values or in exons not affecting all transcripts of the gene. Furthermore, several of the nonsense variants were located in regions with low coverage, suggesting they are likely false positives. In this analysis we demonstrate a marked reduction in the prevalence of TTNtv during the evolution and improvement of 1000 Genomes Project (Fig 1). The variants of the final phase (phase 3) were used in the analyses as our 1000 Genomes dataset.

TTN Truncating Variants in General Population

In our goal to accurately estimate the prevalence of TTNtv in the general population, we queried TTNtv in 1000 Genomes, ESP and ExAC databases (Table 1). The workflow is described in Fig 2. With our strategy, we were able to filter likely false positives unique to either 1000 Genomes or ESP variants. Of note, only three splice-site variants were shared between 1000 Genomes and ESP TTN truncation call sets suggesting that ESP is enriched with population specific variants, as it is made up of primarily Americans either of European (EA) or African (AA) origin. Furthermore, at least 52% of TTNtv (1000Genomes: 52% - 17/33; ESP: 54% - 36/66; ExAC: 53% - 247/470) found in the populations were located in exons that are not present in all transcripts and estimated to have low probability of pathogenicity.

When comparing the prevalence of TTNtv that affect all transcripts (excluding Novex-3 specific and low PSI exon variants) in ExAC vs 1000 Genomes, we identified lower frequency of nonsense mutations in ExAC cohort compared to 1000 Genomes cohort (0.16% vs. 0.28%, respectively) (Table 1). This was also the trend in nonsense variants identified only in the A-band region (0.09% vs. 0.23%) (Table 1). Essential splice site variants were very rare in ExAC database (0.047%) and were absent in 1000 Genomes and ESP. When evaluating the TTNtv that affect all transcripts in ExAC reference population by TTN band, 11.0% (19/173) are located in the Z-band; 22.6% (39/173) in the I-band; 52.0% (90/173) in the A-band; and 14.5% (25/173) in the M-band (Table 1 and Fig 3). Based on cohort size, the ExAC database gave the most reliable result in this analysis as some subtypes of TTNtv were absent in either 1000 Genomes or ESP call sets (Table 1). In the ExAC database the total prevalence of TTNtv affecting all transcripts was 0.36% (0.315% - 0.413%, 95% CI). Given that low coverage regions are often enriched with false-positive INDELS, we rendered the histogram of the TTNtv coverage data from ExAC and found out that about 80% of the variants have coverage of at least 50x. Moreover, we estimated TTNtv prevalence without introducing coverage filter and found out that despite the little increase in the number of variants, the prevalence did not increase significantly (0.46%) and falls within the confidence interval of the prevalence estimate with coverage filter.

The prevalence of TTNtv in clinically relevant A-band was 0.19%, suggesting that 1 in 500 individual carries one of the subtypes of TTNtv. When evaluating by mutation subtype, we conclude that in reference population, only 1 in 1100 individuals carries a nonsense mutation; 1 in 1750 carries a frameshift, and 1 in 2100 carries an essential splice site variant in TTN A-band region that is estimated to affect all transcripts.

Distribution of TTN Truncating Variants by Ethnicity

Frequencies of TTNtv vary significantly between different ethnic groups (African population [AFR]: 0.46%, American population [AMR]: 0.50%, European population [EUR]: 0.27%, Asian

Table 1. Distribution and burden of TTN truncations in publicly available reference populations.

	Total TTN truncating Allele			Excluding novex-3 specific, low PSI and likely false positive variants		
	1KG	ESP	ExAC	1KG	ESP	ExAC
Population size	2504	6504	60706	2504	6504	60706
Total TTNtv variant	29	66	470	9	16	173
Individuals	46	612	813	10	19	219
Prevalence (%)						
Population	1.837	9.409	1.339	0.399	0.292	0.360
Mutation type # (Prevalence %)						
Frameshift	1 (0.120)	25 (2.550)	179 (0.420)	1 (0.079)	3 (0.090)	63 (0.110)
Nonsense	15 (1.120)	28 (0.510)	177 (0.420)	7 (0.280)	11 (0.170)	72 (0.160)
Splice site	13 (0.600)	13 (6.350)	114 (0.490)	1 (0.040)	2 (0.030)	38 (0.090)
Distribution by Domain # (Prevalence %)						
Z-band	1 (0.039)	2 (0.030)	28 (0.062)	-	1 (0.015)	19 (0.037)
I-band	13 (0.718)	49 (8.225)	289 (0.957)	1 (0.039)	7 (0.107)	39 (0.072)
A-band	12 (0.918)	9 (0.153)	122 (0.252)	7 (0.319)	4 (0.061)	90 (0.196)
M-band	3 (0.119)	6 (0.999)	31 (0.067)	1 (0.079)	4 (0.107)	25 (0.054)
Z-disk # (prevalence %)						
Frameshift	-	-	10 (0.016)	-	-	8 (0.013)
Nonsense	-	1 (0.015)	8 (0.013)	-	1 (0.015)	4 (0.006)
Splice site	1 (0.039)	1 (0.015)	10 (0.032)	-	-	7 (0.018)
I-band # (prevalence %)						
Frameshift	-	17 (0.261)	109 (0.303)	-	-	10 (0.016)
Nonsense	4 (0.276)	20 (0.307)	103 (0.265)	1 (0.039)	5 (0.076)	19 (0.034)
Splice site	9 (0.434)	12 (0.184)	77 (0.388)	-	2 (0.030)	10 (0.021)
A-band # (prevalence %)						
Frameshift	1 (0.079)	4 (0.061)	50 (0.082)	1 (0.079)	1 (0.015)	35 (0.057)
Nonsense	10 (0.789)	5 (0.076)	49 (0.112)	6(0.239)	3 (0.046)	37 (0.090)
Splice site	1 (0.039)	-	23 (0.057)	-	-	18 (0.047)
M-band # (prevalence %)						
Frameshift	-	4 (0.061)	10 (0.023)	-	2 (0.030)	10 (0.006)
Nonsense	1 (0.039)	2 (0.030)	17 (0.034)	-	2 (0.030)	12 (0.024)
Splice site	2 (0.079)	-	4 (0.009)	1 (0.039)	-	3 (0.006)

Titin truncations in 1000 Genomes project (Phase 3), Exome Sequencing Project, and ExAC databases. The number of alleles and population prevalence for each variant is shown. Variants were filtered to exclude variants with low probability of pathogenicity (location in novex-specific and other exons with low expressions in LV). Abbreviations: 1KG– 1000 Genomes project; ESP–Exome Sequencing Project; ExAC—Exome Aggregation Consortium; PSI–Proportion Spliced-In

doi:10.1371/journal.pone.0145284.t001

population [AS]: 0.53%; $P = 2.7 \times 10^{-5}$). AS has the highest frequency of TTNtv while the European population has the least. The frequency of TTNtv in East Asian population (EAS), though lower, is not significantly different from that of the South Asian Population (SAS) (EAS: 0.37% vs. SAS: 0.62%; $P = 0.091$). Similarly, the frequency of TTNtv in Finnish European population (FIN) is lower but statistically indistinct when compared with the non-Finnish Europeans (NFE) (FIN: 0.12% vs. NFE: 0.28%; $P = 0.125$).

Golbus et al. [18] recently reported that Asians have significantly more TTN protein altering variants (PAV) than all other ethnic groups (28.76 variants/individual). In this study, we have found a higher frequency of TTNtv in the Asians. Put together, this data confirms the

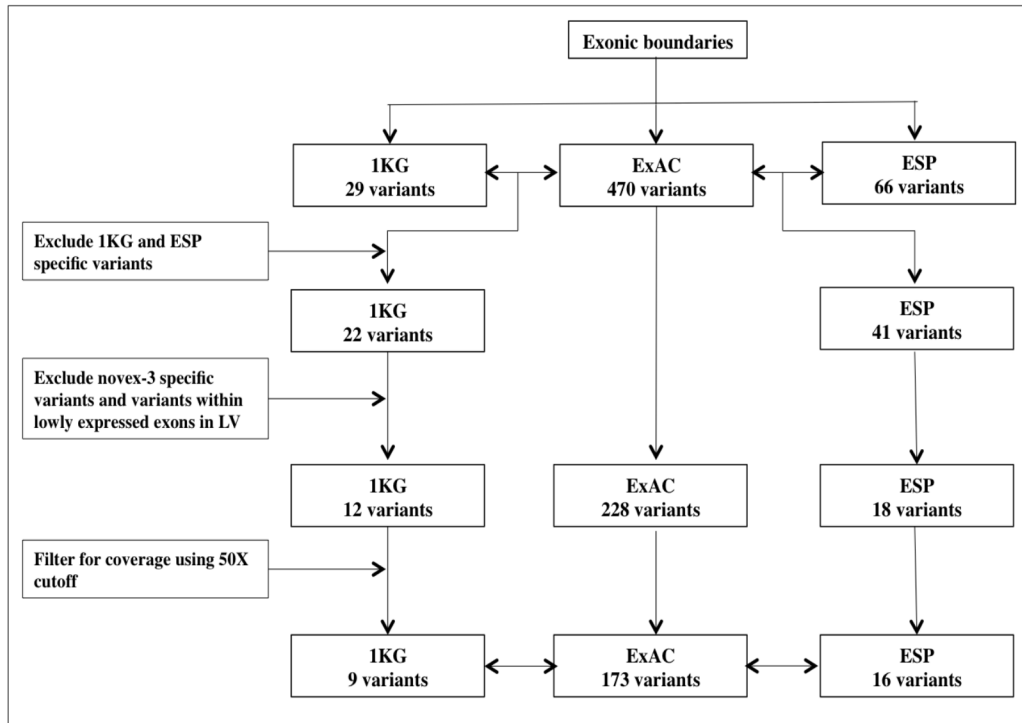


Fig 2. Flowchart for analysis of *TTN* truncations in publicly available reference populations. Abbreviations: 1KG—1000 Genomes project; ESP—Exome Sequencing Project; ExAC—Exome Aggregation Consortium; LV—Left ventricle.

doi:10.1371/journal.pone.0145284.g002

enrichment of not only *TTN* PAV but also *TTN*tv in the Asian population (SAS in particular) compared to other ethnic groups.

Nonsense *TTN* SNPs in Reference Population

We examined the patterns and the positions of the *TTN* nonsense SNPs in the population. Out of the twenty-three possible ways to change codons into stop codons (nine, seven and seven for the first, second and third positions, respectively), twenty-one were found in reference population, with tCg/tAg and tgT/tgA being absent. Nonsense SNPs were more frequent at the first codon position than at the second and third positions ($P = 5.2 \times 10^{-11}$, chi-square test). The most frequent type of nonsense mutation in the *TTN* gene was the change from Cga to Tga (R/*). Of note, the Cga/Tga transitional change at CpG mutation hotspots accounts for 91.3% (21/23) and 62% (21/34) of arginine residue nonsense mutations (R/*) at *TTN* A-band region and all *TTN* CpG hotspot nonsense mutations in the population data respectively. Interestingly, 36.1% (13/36) of R/* found in reference population have been reported in Catalogue of Somatic Mutations in Cancer (COSMIC) database [20].

Splice Site Variants in Reference Population

Of the total splice site/region variants, 86.9% (324/373) were non-essential splice variants located >2bp into the intronic region, with 52.8% (197/373) of the variants affecting *TTN* A-band exons. Of note, majority, 88.8% (175/197), of the splice variants affecting A-band exons were non-essential splice-site variants. When analyzing all essential splice site variants, we identified that 83.7% (41/49) were private. In contrast, non-essential splice-site variants had

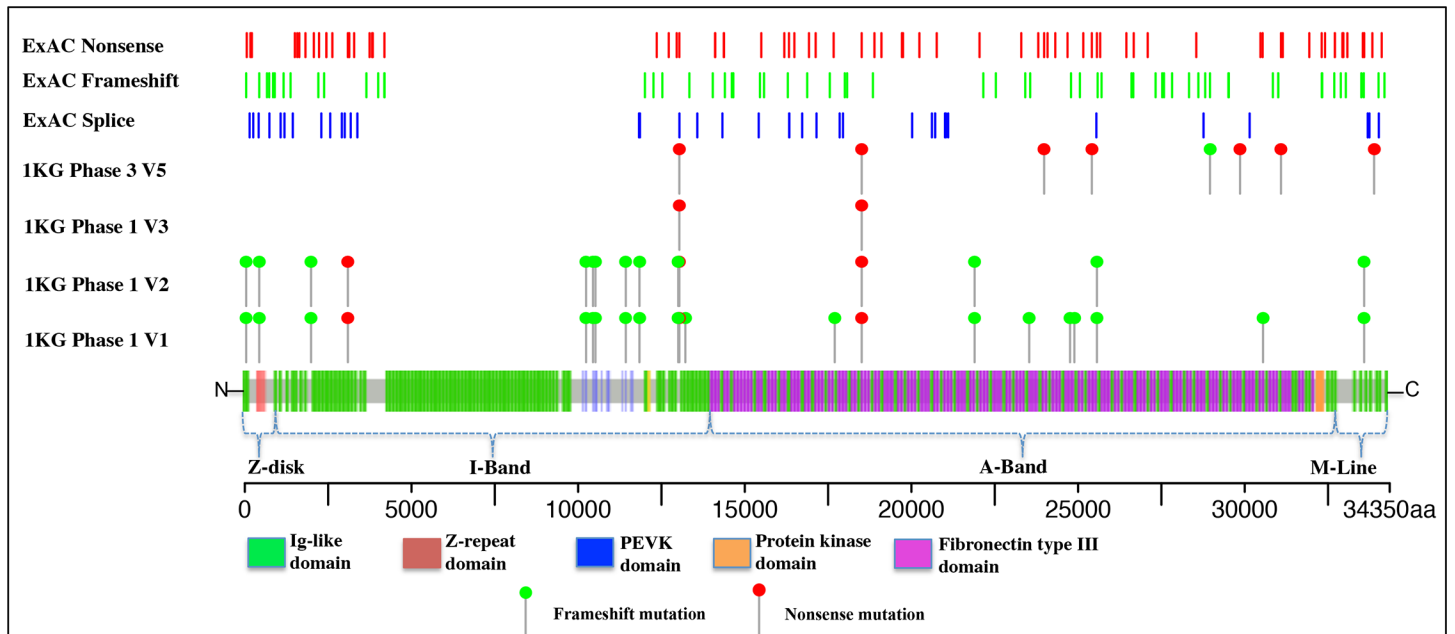


Fig 3. Spatial distribution of titin frameshift, nonsense and splice-site mutations in reference populations. Titin is linearly depicted with its 152 Ig-like domains in green and 132 fibronectin type III domains in purple. TTNtv are shown as lollipops and bars. Depicted are nonsense (red) and frameshift (green) mutations in various phases of the 1000 Genomes project. Dark grey bars represent the complete map of various *TTN* mutations from ExAC. Variants are shown relative to the titin Uniprot Sequence identifier Q8WZ42. The dashed lines below the protein schematic indicate the location of variants within the sarcomere. Abbreviations: 1KG—1000 Genomes project; ESP—Exome Sequencing project; ExAC—Exome Aggregation Consortium; V—version.

doi:10.1371/journal.pone.0145284.g003

significantly lower proportion of private variants and higher proportion of low-frequency variants when compared to essential splice-site variants ($P = 0.01$; $P = 5.1 \times 10^{-4}$, respectively). The significance increases with increasing distance away from essential splice site (Fig 4).

Discussion

By evaluating the ExAC reference population of over 60 000 individuals, we showed that TTNtv affecting all transcripts occur in 0.36% (0.315% - 0.413%, 95% CI) of the general population. These types of truncations in clinically relevant A-band can be detected in 0.19% of reference population. When evaluating the mutation types separately, the prevalence of each mutation type in A-band was lower than 0.1%, below a commonly used filtering criterion when searching for disease-causing mutation of rare dominantly inherited diseases. Furthermore, we provide insights on the position, type and frequency of nucleotide change(s) leading to nonsense mutations in *TTN*, and showed that non-essential splice-site variants in *TTN* have a tendency of being frequent in general population. In addition, we describe how increased quality of data analysis and larger number of reference individuals in 1000 Genomes Project has markedly reduced the prevalence of TTNtv in this database.

The 1000 Genomes Project published the pilot phase of the project in 2010. This release represents data from low-coverage (~2-fold– 6-fold) whole-genome sequencing of 179 individuals, high-coverage (~42-fold) whole-genome sequencing of 6 individuals in 2 trios, and exon-targeted sequencing (> 50-fold coverage) of 8140 exons in 697 individuals [16]. This was followed by Phase 1 project release that represents low coverage exome data available for the first 1092 samples. Phase 2 represents an expanded set of samples, around 1700 in number, and it was used for method development to both improve existing methods from phase 1. Furthermore, it

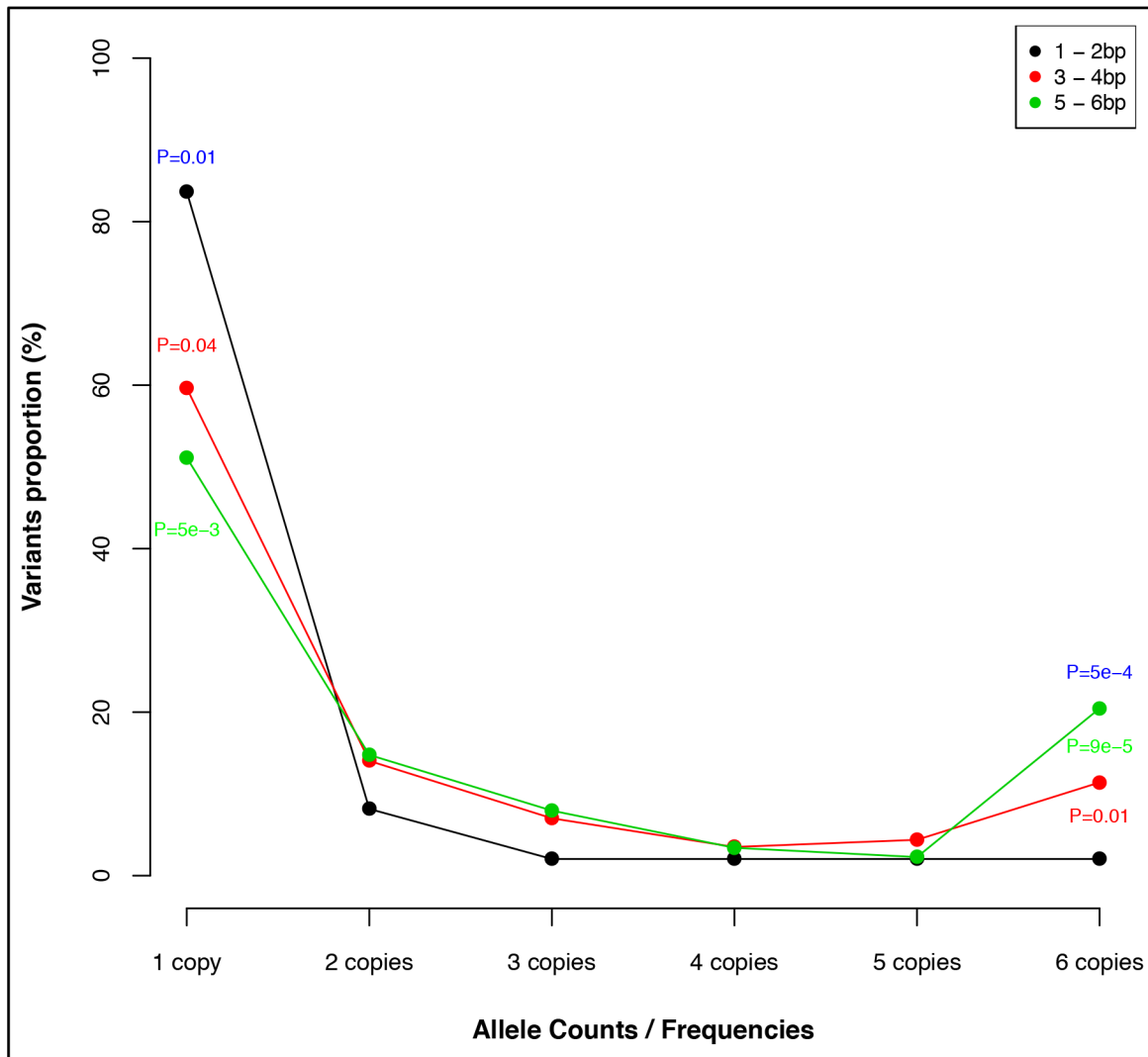


Fig 4. Titin population splice site/region allele frequency spectrum. *TTN* population non-essential splice-site variants (>2bp) have significantly lower proportion of private variants and higher proportion of low-frequency variants compared to essential splice-site variants ($P = 0.01$; $P = 5.1 \times 10^{-4}$, respectively). The P-values are shown for comparison between essential splice-site (1-2bp) and non-essential splice-site variants: 3-4bp (red), 5-6bp (green), and (blue) for combined comparison.

doi:10.1371/journal.pone.0145284.g004

was used to develop new methods to handle features like multi-allelic variant sites and true integration of complex variation and structural variants. The phase 1 integrated release, based on low coverage and exome data, contains phased genotypes for 1092 individuals from 14 populations, and the variants have been extensively used in several studies to draw conclusions on human genetic variations [18, 21, 22]. Although significantly improved, the data still contained higher fraction of frameshift INDELs in low-coverage regions suggesting high prevalence of false positive frameshifts. Subsequent filtering of these false positives resulted in version 2 (V2) and version 3 (V3) of the phase 1 integrated release. The false positive INDELs present in the February 2012 release used in Golbus et al. [18] study has led to high estimates of truncating *TTN* variants in reference population.

Nonsense mutations account for an appreciable proportion of single-basepair substitutions affecting gene-coding region reported in Human Gene Mutation Database (HGMD). Our

analysis shows that as in other genes, Cga/Tga (Arg/*; resulting from methylation-mediated deamination) nucleotide substitution at CpG hotspot is the most frequent nucleotide substitution leading to nonsense mutation in *TTN*. The cytosine-guanine (CpG) dinucleotide, though under-represented in vertebrate genomes, has been reported to be a hotspot for pathological mutations in the human genome [23, 24]. This hypermutability is related to its role as the major site of cytosine methylation with the attendant risk of spontaneous deamination of 5-methylcytosine (5mC) to yield thymine [25]. In this study, we found a significant over-representation of Arg/* CpG hotspot mutation at *TTN* A-band region. Interestingly, 62% of Arg/* CpG hotspot mutations are located at *TTN* A-band. Furthermore, a significant proportion of *TTN* Arg/* CpG hotspot mutations have been reported as germ line mutations identified in various cancer cells. This highlights the fact that the CpG is a critical hotspot for both germline and somatic mutations. Considering the high prevalence of Arg/* CpG hotspots in A-band, majority of exons coding for A-band present in all *TTN* transcripts, and that nonsense mutations in this region do not lead to mRNA decay but seem to work through dominant negative mechanism, it is not surprising that the A-band region is vulnerable to truncating mutations that are highly enriched among DCM patients.

Epigenetic mechanisms could underlie the elevated mutation frequency in the *TTN* A-band region occurring also in the germ line and manifesting hereditary cardiomyopathy. Epigenetic mechanisms are increasingly being recognized as causes and modulators of human disease. To date, there are few studies on the contribution of DNA methylation to disease onset and progression in cardiomyopathy [26–28]. In 2011, Movassagh et al. [26] reported differential DNA methylation patterns in CG dinucleotides located at the promoter, intragenic and gene bodies CpG islands in human end-stage cardiomyopathy. Furthermore, using a single gene model, Meurs & Kuan [28] evaluated the methylation of the CpGs within the exon regions of the skeletal muscle isoform of the myosin binding protein C gene (*MYBPC2*) and cardiac myosin binding protein C gene (*MYBPC3*), a common causal gene for hypertrophic cardiomyopathy. Interestingly, the mean methylation level of CpGs was significantly higher in *MYBPC3* than *MYBPC2* ($P < 0.0001$). These probably suggest that there are unique aspects of this cardiac gene that may result in increased genetic mutability. As mechanisms promoting mutability play a role in disease prevalence, evaluation of the methylation levels of *TTN*, and other DCM-associated genes are warranted.

Differential splicing in regions with variable sarcomere ultrastructures accounts for the variable structures of *TTN* I-band in different tissues. In our goal to characterize essential splice and other splice region variants that might contribute to disease development, we examined and compared the position (1 - 6bp into the intronic region, 1 - 2bp being essential splice site variants) and frequency of all *TTN* essential splice-site and splice-region variants in the population. While essential splice-site variants were predominantly private, we observed an increase in allele copies of non-essential splice-site variants as the distance increases from the essential splice-site. Robert et al. [10] recently reported lower enrichment of noncanonical splice variants in DCM compared to canonical splice variants. Our results confirm that clinical interpretation of non-essential splice site variants in *TTN* gene is difficult and these variants should not be claimed deleterious without RNA level evidence of splicing defect.

Several studies have reported M-line *TTN* mutations in various skeletal muscle diseases including centronuclear myopathy (CM, MIM #160150) [29], tibial muscular dystrophy (TMD, MIM #600334) [30, 31], early-onset myopathy (MIM #611705) [6], distal myopathy and limb girdle muscular dystrophy 2J (LGMD2J, MIM #608807) [32], with or without cardiac muscle involvement. In 2007, Carmignac et al. [6] reported the first M-line homozygous titin truncations causing congenital titinopathy involving both cardiac and skeletal muscle. While most heterozygous M-line titin mutations cause late-onset, dominant disorders involving

predominantly skeletal muscle, homozygous or compound heterozygous M-line *TTN* mutations cause early-onset, recessive muscle and cardiac disorder. M-line *TTN* mutations truncate only part of the M-band portion of *TTN* [6]. When such carboxy-terminal truncated titin proteins are integrated into the sarcomere, they cause recessive, early-onset skeletal and cardiac myopathy. In the case of DCM however, truncated titin proteins, though incorporated into the sarcomere, would not include the M-band residues, as TTNtv causing DCM were clustered in *TTN* A-band but absent from Z-disk and M-band regions of *TTN*. The position of TTNtv causing DCM suggests a dominant negative effect. In addition, Herman et al. [9] argued that we would expect a more uniform distribution of such mutations if more proximal TTNtv caused DCM through haploinsufficiency. Furthermore, using RNA sequencing, Roberts et al. [10] reported a comparable total *TTN* transcript levels in patients with or without TTNtv. They also found and reported a comparable allelic expression of TTNtv and SNPs, which does not support substantial nonsense-mediated decay. Using a combination of genetic data, *TTN* RNA and protein expression in LV tissues, Roberts et al. [10] also concluded that TTNtv cause DCM by a dominant negative effect.

As majority of TTNtv have been identified at the heterozygous state in patients with sporadic or dominant conditions, such mutations are typically considered autosomal dominant. In majority of the cases, dominant inheritance has been clearly established by co-segregation. To the best of our knowledge, only a few studies have suggested recessive inheritance with *TTN* mutations. In these studies, TTNtv co-occurred with a second hit but actually only eight out of 18 patients had two TTNtv, and in two of these cases, bi-allelic status was not confirmed by parental testing [29, 33, 34]. However, Ceyhan-Birsoy et al. [29] did not evaluate the parents' cardiac phenotype; Chauveau et al. [33] found normal echocardiography and ECG in all parents but they were relatively young (aged 38–55 years) and thus a genotype-positive yet a phenotype-negative status cannot be excluded. In the Evilä et al. [34] study, the other patient with bi-allelic TTNtv had an affected father with skeletal myopathy but the patient mother's cardiac phenotypes were not evaluated. Interestingly, all TTNtv reported in these studies were located in the M-line of *TTN*. These observations together with our prevalence estimates highlight that recessive cardiac disease caused by titin dysfunction is possible in rare occasions. Initial evidence exists for recessive inheritance of *TTN* mutations in neurological disease although clearly more comprehensive segregation studies are warranted to ease interpretation of TTNtv.

Concluding Remarks

Considering the accumulated evidence of *TTN* truncating mutations being the most common cause for hereditary DCM [9–11] together with our observations on low prevalence of potentially disease-causing TTNtv in general population, we emphasize that identifying a TTNtv, especially in the A-band region and affecting all transcripts, have a potentially higher risk of being disease causing than previously anticipated [10]. Further studies are required to elucidate the full phenotypic spectrum of TTNtv-associated myocardial disease. It is likely that the severe end-stage DCM or LVNC only represent the peak of the iceberg and milder disease forms that will never develop fulminant cardiomyopathy may exist e.g. transient peripartum cardiomyopathy. This information is needed as it affects our interpretation of genetic variants in reference populations and also the genetic counseling processes and clinical management. Evaluating the *TTN* gene reliably in clinical setting requires expertise in DNA library preparation, high coverage NGS strategies, validated bioinformatics tools, confirmatory sequencing, skilled interpretation team, and ideally the inclusion of mRNA and/or protein level assessment for definite diagnosis. Current low coverage exome and whole genome sequencing approaches may be sub-optimal for diagnostic work-up in cardiomyopathies as their efficacy for identifying

INDELS is weak compared to high coverage targeted sequencing strategies [35, 36]. Moreover, the size of *TTN* gene makes it more susceptible than an average sized gene to false positive truncating variant especially when using 'clinical' exome sequencing. It can be easily demonstrated by the fact there are four homozygous *TTN* variants (positions: 2:179610726, 2:179612587, 2:179571683, 2:179466515) in ExAC database and they are found as heterozygote in 1, 5, 10 and 10 individuals respectively, although we would expect to find 245 heterozygotes for one homozygote in this cohort. This indicates a high probability of detected homozygotes being false positives. Further studies are needed to evaluate the reliability of our reference data, and especially to find out how many of the TTNtv in reference populations are true positives. It should be kept in mind that one out of seven low quality score variants (QS<500) in exome sequencing are likely false positives (non-detectable by Sanger sequencing) [37].

Supporting Information

S1 Table. Truncating *TTN* mutations identified in 1000 Genomes Project Cohort (P1V1). Marked (†) are the false positives that were filtered out in the subsequent version. Abbreviations: P1V1 –phase 1 version 1; GMAF–Global minor allele frequency. (DOCX)

S2 Table. Truncating *TTN* mutations identified in 1000 Genomes Project Cohort (P1V2). Marked (†) are the false positives that were filtered out in the subsequent version. Abbreviations: P1V2 –phase 1 version 2; GMAF–Global minor allele frequency. (DOCX)

S3 Table. Truncating *TTN* mutations identified in 1000 Genomes Project Cohort (P1V3). Marked (†) are the variants that were not detected in the phase 3 integrated variant data. Abbreviations: P1V3 –phase 1 version 3; GMAF–Global minor allele frequency. (DOCX)

Acknowledgments

We would like to thank the 1000 Genomes Project, Exome Sequencing Project and Exome Aggregation Consortium; and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at <http://exac.broadinstitute.org/about>.

Author Contributions

Conceived and designed the experiments: OA JWK TPA. Performed the experiments: OA. Analyzed the data: OA. Contributed reagents/materials/analysis tools: OA JWK TPA. Wrote the paper: OA JWK TPA.

References

1. Itoh-Satoh M, Hayashi T, Nishi H, Koga Y, Arimura T, Koyanagi T, et al. Titin mutations as the molecular basis for dilated cardiomyopathy. *Biochemical and biophysical research communications*. 2002; 291(2):385–93. doi: [10.1006/bbrc.2002.6448](https://doi.org/10.1006/bbrc.2002.6448) PMID: [11846417](https://pubmed.ncbi.nlm.nih.gov/11846417/).
2. Gerull B, Atherton J, Geupel A, Sasse-Klaassen S, Heuser A, Frenneaux M, et al. Identification of a novel frameshift mutation in the giant muscle filament titin in a large Australian family with dilated cardiomyopathy. *Journal of molecular medicine*. 2006; 84(6):478–83. doi: [10.1007/s00109-006-0060-6](https://doi.org/10.1007/s00109-006-0060-6) PMID: [16733766](https://pubmed.ncbi.nlm.nih.gov/16733766/).
3. Matsumoto Y, Hayashi T, Inagaki N, Takahashi M, Hiroi S, Nakamura T, et al. Functional analysis of titin/connectin N2-B mutations found in cardiomyopathy. *Journal of muscle research and cell motility*. 2005; 26(6–8):367–74. doi: [10.1007/s10974-005-9018-5](https://doi.org/10.1007/s10974-005-9018-5) PMID: [16465475](https://pubmed.ncbi.nlm.nih.gov/16465475/).

4. Satoh M, Takahashi M, Sakamoto T, Hiroe M, Marumo F, Kimura A. Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochemical and biophysical research communications*. 1999; 262(2):411–7. doi: [10.1006/bbrc.1999.1221](https://doi.org/10.1006/bbrc.1999.1221) PMID: [10462489](https://pubmed.ncbi.nlm.nih.gov/10462489/).
5. Taylor M, Graw S, Sinagra G, Barnes C, Slavov D, Brun F, et al. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. *Circulation*. 2011; 124(8):876–85. doi: [10.1161/CIRCULATIONAHA.110.005405](https://doi.org/10.1161/CIRCULATIONAHA.110.005405) PMID: [21810661](https://pubmed.ncbi.nlm.nih.gov/21810661/); PubMed Central PMCID: PMC3167235.
6. Carmignac V, Salih MA, Quijano-Roy S, Marchand S, Al Rayess MM, Mukhtar MM, et al. C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy. *Annals of neurology*. 2007; 61(4):340–51. doi: [10.1002/ana.21089](https://doi.org/10.1002/ana.21089) PMID: [17444505](https://pubmed.ncbi.nlm.nih.gov/17444505/).
7. Yoskovitz G, Peled Y, Gramlich M, Lahat H, Resnik-Wolf H, Feinberg MS, et al. A novel titin mutation in adult-onset familial dilated cardiomyopathy. *The American journal of cardiology*. 2012; 109(11):1644–50. doi: [10.1016/j.amjcard.2012.01.392](https://doi.org/10.1016/j.amjcard.2012.01.392) PMID: [22475360](https://pubmed.ncbi.nlm.nih.gov/22475360/).
8. Gerull B, Gramlich M, Atherton J, McNabb M, Trombitas K, Sasse-Klaassen S, et al. Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nature genetics*. 2002; 30(2):201–4. doi: [10.1038/ng815](https://doi.org/10.1038/ng815) PMID: [11788824](https://pubmed.ncbi.nlm.nih.gov/11788824/).
9. Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of titin causing dilated cardiomyopathy. *The New England journal of medicine*. 2012; 366(7):619–28. doi: [10.1056/NEJMoa1110186](https://doi.org/10.1056/NEJMoa1110186) PMID: [22335739](https://pubmed.ncbi.nlm.nih.gov/22335739/); PubMed Central PMCID: PMC3660031.
10. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Science translational medicine*. 2015; 7(270):270ra6. doi: [10.1126/scitranslmed.3010134](https://doi.org/10.1126/scitranslmed.3010134) PMID: [25589632](https://pubmed.ncbi.nlm.nih.gov/25589632/).
11. Akinrinade O, Ollila L, Vattulainen S, Tallila J, Gentile M, Salmenperä P, et al. Genetics and Genotype-Phenotype Correlations in Finnish Patients with Dilated Cardiomyopathy *European heart journal*. 2015. doi: In press.
12. van Spaendonck-Zwarts KY, Posafalvi A, van den Berg MP, Hilfiker-Kleiner D, Bollen IA, Sliwa K, et al. Titin gene mutations are common in families with both peripartum cardiomyopathy and dilated cardiomyopathy. *European heart journal*. 2014; 35(32):2165–73. doi: [10.1093/eurheartj/ehu050](https://doi.org/10.1093/eurheartj/ehu050) PMID: [24558114](https://pubmed.ncbi.nlm.nih.gov/24558114/).
13. Pugh TJ, Kelly MA, Gowrisankar S, Hynes E, Seidman MA, Baxter SM, et al. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2014; 16(8):601–8. doi: [10.1038/gim.2013.204](https://doi.org/10.1038/gim.2013.204) PMID: [24503780](https://pubmed.ncbi.nlm.nih.gov/24503780/).
14. Haas J, Frese KS, Peil B, Kloos W, Keller A, Nietsch R, et al. Atlas of the clinical genetics of human dilated cardiomyopathy. *European heart journal*. 2014. doi: [10.1093/eurheartj/ehu301](https://doi.org/10.1093/eurheartj/ehu301) PMID: [25163546](https://pubmed.ncbi.nlm.nih.gov/25163546/).
15. Patricia Arscott M. Truncating Mutations in Titin Associated with Left Ventricular Noncompaction in Two Unrelated Families. *Heart Rhythm* 2014; May 9, 20142014.
16. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467(7319):1061–73. doi: [10.1038/nature09534](https://doi.org/10.1038/nature09534) PMID: [20981092](https://pubmed.ncbi.nlm.nih.gov/20981092/); PubMed Central PMCID: PMC3042601.
17. NHLBI. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA. Available: <http://evs.gs.washington.edu/EVS/>. Accessed 15 January 2015.
18. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. *Circulation Cardiovascular genetics*. 2012; 5(4):391–9. doi: [10.1161/CIRCGENETICS.112.962928](https://doi.org/10.1161/CIRCGENETICS.112.962928) PMID: [22763267](https://pubmed.ncbi.nlm.nih.gov/22763267/); PubMed Central PMCID: PMC3495587.
19. ExAC. Exome Aggregation Consortium (ExAC), Cambridge, MA. Available: <http://exac.broadinstitute.org>. Accessed 15 January 2015.
20. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *British journal of cancer*. 2004; 91(2):355–8. doi: [10.1038/sj.bjc.6601894](https://doi.org/10.1038/sj.bjc.6601894) PMID: [15188009](https://pubmed.ncbi.nlm.nih.gov/15188009/); PubMed Central PMCID: PMC2409828.
21. Pan S, Caleshu CA, Dunn KE, Foti MJ, Moran MK, Soyinka O, et al. Cardiac structural and sarcomere genes associated with cardiomyopathy exhibit marked intolerance of genetic variation. *Circulation Cardiovascular genetics*. 2012; 5(6):602–10. doi: [10.1161/CIRCGENETICS.112.963421](https://doi.org/10.1161/CIRCGENETICS.112.963421) PMID: [23074333](https://pubmed.ncbi.nlm.nih.gov/23074333/); PubMed Central PMCID: PMC3526690.
22. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. doi: [10.1038/nature11632](https://doi.org/10.1038/nature11632) PMID: [23128226](https://pubmed.ncbi.nlm.nih.gov/23128226/); PubMed Central PMCID: PMC3498066.
23. Rideout WM 3rd, Coetzee GA, Olumi AF, Jones PA. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science*. 1990; 249(4974):1288–90. PMID: [1697983](https://pubmed.ncbi.nlm.nih.gov/1697983/).

24. Krawczak M, Cooper DN. Single base-pair substitutions in pathology and evolution: two sides to the same coin. *Hum Mutat.* 1996; 8(1):23–31. doi: [10.1002/\(SICI\)1098-1004\(1996\)8:1<AID-HUMU3>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1098-1004(1996)8:1<AID-HUMU3>3.0.CO;2-Q) PMID: [8807332](https://pubmed.ncbi.nlm.nih.gov/8807332/).
25. Cooper DN, Mort M, Stenson PD, Ball EV, Chuzhanova NA. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Human genomics.* 2010; 4(6):406–10. PMID: [20846930](https://pubmed.ncbi.nlm.nih.gov/20846930/); PubMed Central PMCID: PMC3525222.
26. Movassagh M, Choy MK, Knowles DA, Cordeddu L, Haider S, Down T, et al. Distinct epigenomic features in end-stage failing human hearts. *Circulation.* 2011; 124(22):2411–22. doi: [10.1161/CIRCULATIONAHA.111.040071](https://doi.org/10.1161/CIRCULATIONAHA.111.040071) PMID: [22025602](https://pubmed.ncbi.nlm.nih.gov/22025602/); PubMed Central PMCID: PMC3634158.
27. Haas J, Frese KS, Park YJ, Keller A, Vogel B, Lindroth AM, et al. Alterations in cardiac DNA methylation in human dilated cardiomyopathy. *EMBO molecular medicine.* 2013; 5(3):413–29. doi: [10.1002/emmm.201201553](https://doi.org/10.1002/emmm.201201553) PMID: [23341106](https://pubmed.ncbi.nlm.nih.gov/23341106/); PubMed Central PMCID: PMC3598081.
28. Meurs KM, Kuan M. Differential methylation of CpG sites in two isoforms of myosin binding protein C, an important hypertrophic cardiomyopathy gene. *Environmental and molecular mutagenesis.* 2011; 52(2):161–4. doi: [10.1002/em.20596](https://doi.org/10.1002/em.20596) PMID: [20740642](https://pubmed.ncbi.nlm.nih.gov/20740642/).
29. Ceyhan-Birsoy O, Agrawal PB, Hidalgo C, Schmitz-Abe K, DeChene ET, Swanson LC, et al. Recessive truncating titin gene, TTN, mutations presenting as centronuclear myopathy. *Neurology.* 2013; 81(14):1205–14. doi: [10.1212/WNL.0b013e3182a6ca62](https://doi.org/10.1212/WNL.0b013e3182a6ca62) PMID: [23975875](https://pubmed.ncbi.nlm.nih.gov/23975875/); PubMed Central PMCID: PMC3795603.
30. Hackman P, Vihola A, Haravuori H, Marchand S, Sarparanta J, De Seze J, et al. Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *American journal of human genetics.* 2002; 71(3):492–500. doi: [10.1086/342380](https://doi.org/10.1086/342380) PMID: [12145747](https://pubmed.ncbi.nlm.nih.gov/12145747/); PubMed Central PMCID: PMC379188.
31. Udd B, Vihola A, Sarparanta J, Richard I, Hackman P. Titinopathies and extension of the M-line mutation phenotype beyond distal myopathy and LGMD2J. *Neurology.* 2005; 64(4):636–42. doi: [10.1212/01.WNL.0000151853.50144.82](https://doi.org/10.1212/01.WNL.0000151853.50144.82) PMID: [15728284](https://pubmed.ncbi.nlm.nih.gov/15728284/).
32. Hackman P, Marchand S, Sarparanta J, Vihola A, Penisson-Besnier I, Eymard B, et al. Truncating mutations in C-terminal titin may cause more severe tibial muscular dystrophy (TMD). *Neuromuscular disorders: NMD.* 2008; 18(12):922–8. doi: [10.1016/j.nmd.2008.07.010](https://doi.org/10.1016/j.nmd.2008.07.010) PMID: [18948003](https://pubmed.ncbi.nlm.nih.gov/18948003/).
33. Chauveau C, Bonnemann CG, Julien C, Kho AL, Marks H, Talim B, et al. Recessive TTN truncating mutations define novel forms of core myopathy with heart disease. *Human molecular genetics.* 2014; 23(4):980–91. doi: [10.1093/hmg/ddt494](https://doi.org/10.1093/hmg/ddt494) PMID: [24105469](https://pubmed.ncbi.nlm.nih.gov/24105469/); PubMed Central PMCID: PMC3954110.
34. Evila A, Vihola A, Sarparanta J, Raheem O, Palmio J, Sandell S, et al. Atypical phenotypes in titinopathies explained by second titin mutations. *Annals of neurology.* 2014; 75(2):230–40. doi: [10.1002/ana.24102](https://doi.org/10.1002/ana.24102) PMID: [24395473](https://pubmed.ncbi.nlm.nih.gov/24395473/).
35. Dewey FE, Grove ME, Pan C, Goldstein BA, Bernstein JA, Chaib H, et al. Clinical interpretation and implications of whole-genome sequencing. *Jama.* 2014; 311(10):1035–45. doi: [10.1001/jama.2014.1717](https://doi.org/10.1001/jama.2014.1717) PMID: [24618965](https://pubmed.ncbi.nlm.nih.gov/24618965/); PubMed Central PMCID: PMC4119063.
36. O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome medicine.* 2013; 5(3):28. doi: [10.1186/gm432](https://doi.org/10.1186/gm432) PMID: [23537139](https://pubmed.ncbi.nlm.nih.gov/23537139/); PubMed Central PMCID: PMC3706896.
37. Strom SP, Lee H, Das K, Vilain E, Nelson SF, Grody WW, et al. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genetics in medicine: official journal of the American College of Medical Genetics.* 2014; 16(7):510–5. doi: [10.1038/gim.2013.183](https://doi.org/10.1038/gim.2013.183) PMID: [24406459](https://pubmed.ncbi.nlm.nih.gov/24406459/); PubMed Central PMCID: PMC4079763.