Deciphering the Molecular Mechanisms of ALPK3 and FLNC Cardiomyopathy

Citation

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37365925

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Deciphering the molecular mechanisms of

ALPK3- and FLNC-cardiomyopathy

A dissertation presented

by

Radhika Agarwal

to

The Division of Medical Sciences

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

March 2020
© 2020 Radhika Agarwal

All rights reserved.
Abstract

Cardiomyopathies represent a spectrum of diseases characterized by pathologic cardiac enlargement that can lead to heart failure. A significant proportion of non-ischemic cardiomyopathies occur as a consequence of rare damaging genetic variants. Understanding how these variants induce cardiomyocyte dysfunction can provide insight into the pathogenesis of heart failure – a disease for which few therapies exist. CRISPR/Cas9-mediated genome editing in human induced pluripotent stem cells (hiPSCs), coupled with methods to differentiate hiPSCs into cardiomyocytes (hiPSC-CMs), allow the effects of mutated cell lines to be compared directly to isogenic non-edited parent lines. Harnessing these technologies, we characterized the molecular mechanisms by which damaging variants in two genes – FLNC (filamin C) and ALPK3 (α-kinase-3) – cause cardiomyopathy.

To investigate FLNC pathogenesis, we engineered four mutant hiPSC lines – one with homozygous null alleles, two with different heterozygote variants, and one with a homozygous missense variant of unknown significance. We found that some FLNC expression was necessary for normal sarcomere thin filament gene expression and assembly – these processes were deficient in FLNC homozygous null hiPSC-CMs. However, hiPSC-CMs carrying heterozygote variants, modeling human disease, displayed an additional, distinct mechanism of pathogenicity characterized by toxic protein accumulation of FLNC-binding partners and increased lysosome abundance.
We suggest that strategies to ameliorate protein accumulation may be therapeutic in FLNC-cardiomyopathy.

To investigate ALPK3 pathogenesis, we engineered two mutant hiPSC lines – one containing biallelic ALPK3-null alleles (modeling human disease variants), and the other containing homozygote missense variants predicted to conformationally alter the ALPK3 kinase domain. Both perturbations increased the protein levels of myomesins (MYOM1, MYOM2) and sarcomere thick filament proteins in hiPSC-CMs, effects that we also observed in Alpk3−/− mice. Furthermore, ALPK3 co-localized with MYOM1 and MYOM2 at the sarcomere M-band and the nuclear envelope respectively, and myomesin assembly into these structures was deficient in ALPK3-null hiPSC-CMs. Surprisingly, genetic inhibition of the ALPK3 kinase domain did not detectably alter phosphorylation across the cardiomyocyte phosphoproteome, thus defining this protein as a pseudokinase. Given these results, we propose a model in which ALPK3 serves as a critical scaffold for myomesin incorporation and thick filament turnover, the dysfunction of which results in cardiomyopathy.
# Table of Contents

Abstract........................................................................................................................................iii
Acknowledgements ..................................................................................................................vi

Chapter 1: Introduction................................................................................................................1
Cardiomyopathy: definitions and classifications .................................................................... 2
The genetic architecture of cardiomyopathy ............................................................................ 5
Structure and function of the cardiomyocyte .......................................................................... 10
Molecular mechanisms of cardiomyopathy and therapeutic avenues .................................... 24
Overview of dissertation .......................................................................................................... 32

Chapter 2: Filamin C truncating variants that cause dilated cardiomyopathy result in toxic protein accumulation and increased lysosome abundance in hiPSC-CMs .............................................. 33
Abstract: ......................................................................................................................................34
Introduction .................................................................................................................................35
Results ..........................................................................................................................................40
Discussion ................................................................................................................................. 60
Methods .......................................................................................................................................67
Contributions ...............................................................................................................................77

Chapter 3: Deficiency of α-kinase 3 (ALPK3), a pseudokinase, results in impaired nuclear and sarcomeric myomesin turnover .......................................................................................... 78
Abstract ......................................................................................................................................79
Introduction .................................................................................................................................80
Results ..........................................................................................................................................84
Discussion: .................................................................................................................................107
Methods .......................................................................................................................................114
Contributions ...............................................................................................................................125

Chapter 4: Reflections and Future Directions.............................................................................126
Filamin C and Z-discopathies as protein aggregation diseases of the heart ......................... 127
α-kinase-3: nucleus-to-M-band signaling and mechanisms of sarcomere protein turnover .................................................................................................................................133
A final note: ................................................................................................................................137

References...................................................................................................................................138

Appendix 1: Material related to Chapter 2 ...............................................................................152

Appendix 2: Material related to Chapter 3 ...............................................................................160
Acknowledgements

Graduate school has taught me a lot about myself. I have learned that I love being part of the scientific community – a vibrant group of doers and dreamers. What makes science exciting is that it is all about discovery. But discoveries can lead one in unexpected directions, and it can be confusing to figure out where to go next. These crossroads, of frustratingly endless paths and possibilities and no clear directions, were challenging for me. I have been lost many times.

I am lucky to have had people who have helped me find my way, time and time again. I am immensely grateful to my mentors – Kricket and Jon Seidman – whose vision and seemingly indefatigable persistence to keep searching and pushing forward, have kept me inspired throughout my graduate career. It is rare to find such people, who are so genuinely curious about nearly every aspect of biology, that the hurdles and challenges of taking on big problems begin to look small by comparison. When I feel weary, it only takes a short meeting with Kricket and Jon for their enthusiasm to permeate into my bones, energizing me once again to keep going. I think their philosophy is that no discovery is interesting, because all discoveries are interesting. And if you haven’t figured out why, well, maybe you just need to keep searching.

To my family, I owe more than words can convey. In my childhood, my parents and my grandmother instilled in me the value of an ancient Hindu teaching – that the desire for the fruits of one’s labor must never be one’s motive in working – a mantra that reminds me to enjoy each day, and sustains me through times when the fruits feel far away. My mother has listened patiently to my every gripe, and has proudly celebrated each of my successes. To be on the receiving end of her love is the greatest blessing
and biggest joy of my life. My father taught me how to climb mountains, both physical and intellectual. In my second year of graduate school, we hiked the Annapurna Circuit together, a whirlwind adventure that peaked when we climbed over the Thorang La mountain pass at 17,716 feet. I am lucky to have parents who have taught me to aim high and dream big, who believe in my potential, and who lift me up and walk alongside me as I reach towards my goals.

Finally, I must thank the army of friends and colleagues who have supported me throughout my graduate career. To my friends and colleagues in the Seidman lab, you make it a joy to come to the NRB each day. To Steve Gygi and Joao Paulo, thank you for taking a chance on me as a rotation student, who never quite went away. To Thomas Michel, Rick Mitchell, Loren Walensky, Amy Cohen, members of my PQE, DAC, and thesis committees, and my incredible medical and graduate school friends and colleagues, thank you for all your support and for making my HST and MD/PhD training experiences outstanding. To Ann Ogletree, thank you for being my home-away-from-home, and for living the kind of life of service and gratitude that I strive towards. To Ally Freedy, thank you for being a true, kind, and loyal friend since the very beginning; this journey would have been a lot less fun without you in it. To each person who has helped me to find my calling, I am so grateful.
Chapter 1: Introduction
**Cardiomyopathy: definitions and classifications**

In the United States alone, about 6.5 million adults are living with heart failure, a leading cause of cardiac transplantation and premature death (Chamberlain, 2019). One prevalent cause of heart failure is cardiomyopathy. Cardiomyopathies are disorders of the heart muscle that occur in the absence of underlying causes such as coronary artery disease, hypertension, valvular, or congenital heart disease (Elliott *et al.*, 2008). In the absence of these secondary causes, cardiomyopathies reflect a group of genetic disorders with Mendelian patterns of inheritance (Towbin, 2000; Seidman and Seidman, 2001; Richard *et al.*, 2003).

Based on their effects on cardiac morphology and function, cardiomyopathies have historically been subclassified into several subtypes: hypertrophic (HCM), dilated (DCM), arrhythmogenic right ventricular (ARVC), restrictive (RCM), and otherwise unclassified cardiomyopathies (Richardson *et al.* 1996). The characteristic features of the major subtypes are summarized here (Fig 1.1).

- **HCM** is characterized by an increase in left ventricular (LV) wall thickness. The distribution of hypertrophy is usually asymmetric with disproportionate involvement of the interventricular septum as compared to the LV free wall; this results in dynamic LV outflow tract obstruction in ~70% of patients (Maron *et al.*, 2006). However, other patterns of hypertrophy, including concentric and apical hypertrophy, can also occur. The LV volume is normal-to-reduced and diastolic dysfunction is usually present (Elliott *et al.*, 2008). Arrhythmias are
prevalent in HCM, with atrial fibrillation occurring in ~2% of patients (Ware et al., 2018).

- **DCM** is characterized by left ventricular dilatation and global systolic dysfunction (ejection fraction (EF) < 50%). DCM is associated with an increase in total cardiac mass, typically with normal or decreased LV wall thickness. Ventricular and/or atrial arrhythmias can be prevalent based on the underlying cause (Cecchi, Tomberli and Olivotto, 2012).

- **ARVC** is defined by a specific histopathology; the myocardium of the RV free wall (and frequently the LV as well) is partially replaced by fibrous and/or fibrofatty tissue. RV function is abnormal with regional akinesis or dyskinesis, and LV function can be abnormal as well, producing global systolic dysfunction, similar to DCM. Ventricular arrhythmias are characteristic of this disease (Cecchi, Tomberli and Olivotto, 2012).

- **RCM** is characterized by nondilated, stiffened ventricles that impair ventricular filling, resulting in diastolic dysfunction, similar to HCM. LV wall thickness is usually normal or mildly increased and systolic function is typically preserved in the early stages of this disease (Elliott et al., 2008).
Furthermore, some cardiomyopathies may be “mixed,” with clinical phenotypes that reflect features of multiple classes. This is likely a reflection of the fact that cardiomyopathy subtypes often share common genetic etiologies, suggesting that cardiomyopathies represent a spectrum of diseases, rather than distinct entities.

Cardiomyopathies are relatively common disorders. The population prevalence of left ventricular hypertrophy in the United States has been estimated at ~1 in 500 individuals, based on epidemiologic data from CARDIA, an echocardiographic study of over 4,000 individuals aged 25-39 years (Maron et al., 1995). Extrapolation from smaller studies, and the frequency of DCM diagnoses in clinical trials of heart failure, have predicted the prevalence of DCM to lie between ~1:400 and 1:250 individuals.
Furthermore, the clinical consequences of cardiomyopathies can be severe. DCM is associated with frequent progression to heart failure and a 1-year mortality of 25-30% (Dec and Fuster, 1994). Additionally, ~20% of HCM patients progress to heart failure (Ware et al., 2018).

Current therapies for heart failure primarily rely on blocking neurohormonal activation with inhibitors of the renin-angiotensin pathway, β-blockers, and/or aldosterone receptor blockers. These agents result in some reversal of cardiac remodeling in approximately 1/3 of patients (Merlo et al., 2011). However, these pharmacologic agents have not substantially altered the natural course of this disease, and further therapeutic options for heart failure remain limited to cardiac transplantation. Understanding the pathogenesis of damaging genetic variants that cause cardiomyopathy can reveal disease mechanisms and identify novel therapeutic targets, with the potential to impact the large number of individuals who suffer from heart failure.

The genetic architecture of cardiomyopathy
DNA sequencing technologies can now identify pathogenic and likely pathogenic variants in ~30-40% of unselected probands, and in over 60% of patients with a positive family history of cardiomyopathy (Alfares et al., 2015; van Lint et al., 2019). Over the past two decades, linkage studies and next generation sequencing studies of large patient cohorts have revealed major genetic causes of cardiomyopathy.

HCM is now understood to be a genetic disease of cardiomyocyte contractile proteins (Table 1.1). In adults, approximately 70% of cases occur due to variants in just two genes: myosin binding protein C (MYPBC3) and β-myosin heavy chain (MYH7) (Richard et al., 2003). Many additional cases can be explained by variants in six additional genes: cardiac troponin T (TNNT2), troponin I (TNNI3), α-tropomyosin (TPM1), myosin regulatory light chain (MYL2), myosin essential light chain (MYL3), and cardiac α-actin (ACTC1). However, the cause(s) of HCM remain unknown in up to 30% of cases. HCM variants are generally deleterious missense variants with autosomal dominant patterns of inheritance, with the exception of MYBPC3 variants, which are generally loss-of-function (Ingles, 2019).

The main genetic causes of RCM are damaging missense variants in MYH7, ACTC1, TNNI3, and TNNT2, indicating that RCM and HCM share a genetic pathogenesis. As both entities are associated with similar clinical features (decreased LV compliance and diastolic dysfunction), it has been suggested that RCM is a “severe form” of HCM (Kostareva et al., 2016).

Table 1.1. Major genetic causes of HCM. Pathogenic variants in 8 genes explain >75% of HCM cases. The majority of HCM variants are deleterious missense variants in
genes that encode components of the cardiomyocyte contractile apparatus, and are
inherited in an autosomal dominant (AD) manner. Expanded diagnostic gene panels
have not substantially increased the yield of variant detection (Ingles, 2019).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Inheritance Pattern</th>
<th>Variant Type</th>
<th>Approximate Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MYH7</td>
<td>AD</td>
<td>Missense</td>
<td>28%</td>
</tr>
<tr>
<td>TNNT2</td>
<td>AD</td>
<td>Missense, splice</td>
<td>5%</td>
</tr>
<tr>
<td>TPM1</td>
<td>AD</td>
<td>Missense</td>
<td>2%</td>
</tr>
<tr>
<td>MYL2</td>
<td>AD</td>
<td>Missense</td>
<td>2%</td>
</tr>
<tr>
<td>MYL3</td>
<td>AD</td>
<td>Missense, splice</td>
<td>1%</td>
</tr>
<tr>
<td>ACTC1</td>
<td>AD</td>
<td>Missense</td>
<td>1%</td>
</tr>
</tbody>
</table>

ARVC is a genetic disease of desmosome proteins, with ~50% of cases involving
dominant pathogenic variants in 1 or more of the 5 major desmosomal genes
(plakophilin (PKP2), desmoplakin (DSP), desmoglein-2 (DSG2), desmocollin-2 (DSC2),
and plakoglobin (JUP)) (Table 1.2). Most variants cause disease by an autosomal
dominant mode, with the exception of JUP variants, which are typically autosomal
recessive (Asimaki, Kleber and Saffitz, 2015; Walsh et al., 2017).

**Table 1.2. Major genetic causes of ARVC.** Pathogenic variants in 5 genes explain
~50% of ARVC cases. The majority of ARVC variants are damaging missense or LOF
variants that encode components of the desmosome, and are inherited in an autosomal
dominant (AD) manner, with the exception of JUP varaints, which are typically
autosomal recessive (AR). Data from the Atlas of Cardiac Genetic Variation (Walsh et
al., 2017).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Inheritance Pattern</th>
<th>Variant Type</th>
<th>Approximate Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKP2</td>
<td>AD</td>
<td>LOF</td>
<td>28%</td>
</tr>
<tr>
<td>DSP</td>
<td>AD</td>
<td>LOF, Missense</td>
<td>12%</td>
</tr>
<tr>
<td>DSG2</td>
<td>AD</td>
<td>Missense, LOF</td>
<td>5%</td>
</tr>
<tr>
<td>DSC2</td>
<td>AD</td>
<td>Missense</td>
<td>2.5%</td>
</tr>
<tr>
<td>JUP</td>
<td>AR</td>
<td>Missense</td>
<td>8%</td>
</tr>
</tbody>
</table>
The genetic architecture of DCM is more heterogeneous (Table 1.3). Genetic causes are identified in only 50% of cases, and causal DCM variants have been identified in over 30 genes (Hershberger, Hedges and Morales, 2013). Truncating variants in the giant sarcomere protein titin (TTN) account for ~25% of DCM cases, and represent the most common genetic cause of DCM (Herman et al., 2012). Other damaging variants are found in genes that encode diverse components of the cardiomyocyte contractile machinery, linkers that connect myofibrils to the plasma membrane, and nuclear proteins (Dellefave and McNally, 2010).

**Table 1.3. Genetic causes of DCM.** The genetic architecture of DCM is heterogeneous, and genetic causes are currently identified in only ~50% of cases. Rare damaging variants are found in a large number of genes that encode proteins that function in diverse compartments of the cardiomyocyte. The most prevalent causes of DCM are truncating variants in TTN, which account for ~25% of DCM cases (Herman et al., 2012). LMNA variants account for ~5% of cases. Rare pathogenic variants in other genes each account for <5% of DCM cases. Data curated from the Atlas of Cardiac Genetic Variation (Walsh et al., 2017) and other sources (Dellefave and McNally, 2010; Hershberger, Hedges and Morales, 2013; McNally and Mestroni, 2017).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Localization</th>
<th>Inheritance Pattern</th>
<th>Variant Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTN</td>
<td>sarcomere</td>
<td>AD</td>
<td>Truncating, missense</td>
</tr>
<tr>
<td>MYH7</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>TNNT2</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>TPM1</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>MYL3</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>TNNI3</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>ACTC1</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>MYL2</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>MYBP3</td>
<td>sarcomere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>ACTN2</td>
<td>Z-disc</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>FLNC</td>
<td>Z-disc</td>
<td>AD</td>
<td>Truncating, Missense</td>
</tr>
<tr>
<td>LDB3</td>
<td>Z-disc</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>TCAP</td>
<td>Z-disc</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>CSRP3</td>
<td>Z-disc</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>MYPN</td>
<td>Z-disc</td>
<td>AD</td>
<td>Missense</td>
</tr>
</tbody>
</table>
Table 1.3 (continued)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Localization</th>
<th>Mode of Inheritance</th>
<th>Type of Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAG3</td>
<td>Z-disc</td>
<td>AD</td>
<td>Truncating</td>
</tr>
<tr>
<td>CRYAB</td>
<td>Z-disc</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>SGCD</td>
<td>costamere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>DES</td>
<td>costamere</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>VCL</td>
<td>costamere</td>
<td>AD</td>
<td>Truncating, Missense</td>
</tr>
<tr>
<td>DMD</td>
<td>costamere</td>
<td>X-linked, Duchenne/Becker muscular dystrophy</td>
<td>Truncating</td>
</tr>
<tr>
<td>LMNA</td>
<td>nucleus</td>
<td>AD</td>
<td>Missense, truncating, splice</td>
</tr>
<tr>
<td>EMD</td>
<td>nucleus</td>
<td>X-linked</td>
<td>LOF</td>
</tr>
<tr>
<td>TMEM43</td>
<td>nucleus</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>ALPK3</td>
<td>nucleus (?)</td>
<td>AR</td>
<td>LOF</td>
</tr>
<tr>
<td>RBM20</td>
<td>spliceosome</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>PLN</td>
<td>sarcoplasmic reticulum</td>
<td>AD</td>
<td>Missense, truncating</td>
</tr>
<tr>
<td>SCN5A</td>
<td>sarcolemma</td>
<td>AD</td>
<td>Missense, truncating</td>
</tr>
<tr>
<td>TAZ</td>
<td>mitochondria</td>
<td>X-linked, Barth syndrome</td>
<td>Missense, truncating</td>
</tr>
<tr>
<td>LAMP2</td>
<td>Lysosome</td>
<td>X-linked, Danon disease</td>
<td>Missense</td>
</tr>
<tr>
<td>GLA</td>
<td>Lysosome</td>
<td>X-linked, Fabry disease</td>
<td>Missense</td>
</tr>
<tr>
<td>PRKAG2</td>
<td>Cytosol</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>DSP</td>
<td>Desmosome</td>
<td>AD</td>
<td>LOF, Missense</td>
</tr>
<tr>
<td>DSC2</td>
<td>Desmosome</td>
<td>AD</td>
<td>Missense</td>
</tr>
<tr>
<td>DSG2</td>
<td>Desmosome</td>
<td>AD</td>
<td>Missense, LOF</td>
</tr>
<tr>
<td>JUP</td>
<td>Desmosome</td>
<td>AR</td>
<td>Missense</td>
</tr>
<tr>
<td>PKP2</td>
<td>Desmosome</td>
<td>AD</td>
<td>LOF</td>
</tr>
</tbody>
</table>

To understand the pathogenic mechanisms of damaging cardiomyopathy variants, it is necessary to first review the normal structure and function of the cardiomyocyte (Fig 1.2).
Figure 1.2. Cardiomyocyte structure, excitation-contraction coupling, and genetic causes of cardiomyopathy. Pacemaker cells spontaneously produce action potentials which travel along the sarcolemma. Sarcolemmal depolarization results in activation of dihydropyridine (DHPR) receptors (also called L-type calcium channels, LTCC), allowing Ca\textsuperscript{2+} to pass into the cardiomyocyte. Ca\textsuperscript{2+} binds to the ryanodine receptor (RyR2) located on the sarcoplasmic reticulum (SR). RyR2 opening results in Ca\textsuperscript{2+}-induced-Ca\textsuperscript{2+}-release from the SR. DHPRs close, preventing further Ca\textsuperscript{2+} influx. Ca\textsuperscript{2+} influx triggers a chemomechanical cycle of sarcomere contraction (Bers, 2012). Sarcomeres are connected via Z-discs to costameres, specialized focal adhesion complexes in the sarcolemma. Cardiomyocytes are connected to one another via specialized junctional complexes called intercalated discs. Select causal
(Figure 1.2 continued) cardiomyopathy genes are indicated in red (see also Fig 1.4). HCM variants primarily affect sarcomere proteins. ARVC variants primarily affect desmosome components of the intercalated disc. DCM variants affect a heterogeneous set of proteins, including proteins of the sarcomere, Z-disc, costamere, nucleus, sarcoplasmic reticulum, lysosome, and mitochondrion (Hershberger, Hedges and Morales, 2013).

Cardiomyocyte structure

The sarcomere

The sarcomere is the fundamental contractile unit of a myocyte. It is composed of interdigitating arrays of myosin-rich thick filaments and actin-rich thin filaments. The arrays of thick and thin filaments produce a characteristic appearance by polarizing or electron microscopy. The A-band (dark) contains the entire lengths of thick filaments, while the I-band (light) contains thin filaments that are not superimposed by thick filaments. Thick filaments are anchored at a structure called the M-band at the center of the sarcomere, while thin filaments are anchored at Z-discs at sarcomere ends (Fig 1.3, Fig 1.4). Sarcomere contraction is achieved by myosin-driven movement of thick filaments along actin-binding sites on thin filaments, which brings Z-discs closer together (Gautel and Djinovic-Carugo, 2016).
Figure 1.3. Electron micrograph of sarcomeres in a mouse cardiomyocyte. Sarcomeres consist of interdigitating arrays of thick and thin filaments. The A-band (dark) contains the entire lengths of thick filaments. The I-band (light) contains thin filaments that are not superimposed by thick filaments. The boundaries of each sarcomere are marked by Z-discs, specialized structures where thin filaments are anchored. Thick filaments are anchored at a central structure called the M-band. Sarcomere contraction is achieved by the myosin-driven movement of thick filaments along actin-binding sites on thin filaments, which brings Z-discs closer together. Sarcomere contraction thus shortens the I-band, without changing the length of the A-band. Mitochondria (mito) are intimately associated with sarcomeres.
Figure 1.4. Schematic of the cardiomyocyte sarcomere. Sarcomeres consist of interdigitating arrays of thick and thin filaments, which are anchored at the M-line and at Z-discs, respectively. Pathogenic variants in all annotated sarcomere genes are genetic causes of cardiomyopathy, with the exception of \textit{MYOM1}, for which there is limited evidence.

Thin filaments are composed of $\alpha$-actin (ACTC1) filaments and the calcium-sensitive troponin-tropomyosin regulatory apparatus, which includes $\alpha$-tropomyosin (TPM1) and the troponin complex (troponin C (TNNC1), troponin T (TNNT2), troponin I (TNNI3)) (Garfinkel, Seidman and Seidman, 2018) (Fig 1.4).

Thick filaments are composed of cardiac $\beta$-myosin heavy chain (MYH7) which is the molecular motor of the thick filament, and other proteins with regulatory functions.
(MYL2, MYL3, MYBPC3) (Fig 1.4). The α-helical tails of myosin heavy chain pack together to form a cylindrical thick filament backbone, from which pairs of myosin heads protrude outwards at regular intervals in a helical manner. These heads (denoted S1) contain binding sites for actin, nucleotide-binding pockets with ATP hydrolase activity, and regulatory domains. The regulatory domains interact with the myosin regulatory and essential light chains (RLC and ELC, encoded by MYL2 and MYL3 respectively). The S1 heads are linked to the myosin tail backbone via an S2 fragment, which binds additional proteins, including cardiac myosin binding protein C (MYBPC3) and titin (TTN) (Garfinkel, Seidman and Seidman, 2018).

TTN (~3820 kDa) is the largest human protein and spans an entire half-sarcomere from the Z-disc to the M-line, serving as a giant elastic scaffold. TTN has 4 functional domains: the N-terminal Z-disc region participates in sarcomere assembly and stabilization; an I band region acts as a bidirectional spring, providing elasticity to restore resting sarcomere length after contraction and passive tension to limit sarcomere stretch during relaxation; an A band region binds myosin and MYBPC3; and a pseudokinase-containing M-band region which may function to scaffold additional proteins (Bogomolovas et al., 2014; Garfinkel, Seidman and Seidman, 2018). The size of TTN and its biomechanical properties have led to hypotheses that TTN may serve as a molecular ruler to determine the length of the sarcomere, and also may have key roles in strain-sensitive signaling (Tonino et al., 2017). Alternative splicing of TTN in the heart produces two predominant isoforms: the N2BA isoform is longer and more compliant, whereas the N2B isoform is shorter and relatively more stiff (Hidalgo and Granzier,
2013). Tuning the isoform composition of TTN contributes greatly to the diastolic stiffness of the heart (Linke and Hamdani, 2014).

**The Z-disc**

Z-discs are the sites at which antiparallel thin filaments from adjoining sarcomeres are crosslinked to one another, a critical early step in sarcomere assembly (Fig 1.3, Fig 1.4). Actin-crosslinking is mediated by a number of dimeric proteins, including α-actinin (ACTN2) and filamin C (FLNC) (Korenbaum, 2002; Knoll and Buyandelger, 2011). CapZ is a heterodimeric protein in the Z-disc that caps the barbed ends of these actin filaments (Papa et al., 1999; Pappas et al., 2008).

In addition to actin filaments, the N-terminus of TTN and the C-terminus of the thin-filament-associated protein nebulette (NEBL), each attach directly to the Z-disc (Prill and Dawson, 2020). The insertion of TTN into Z-discs, and its interactions with MYH7 have been demonstrated to be essential for transmitting forces that are required for ACTN2-mediated actin-crosslinking (Chopra et al., 2018). These observations indicate that the biomechanical properties of the thick filament can critically influence assembly of the thin filament and actin-crosslinking at the Z-disc.

Furthermore, the Z-disc is a critical signaling hub in muscle cells. Mature Z-discs are associated with over two hundred additional proteins with diverse roles in sarcomere assembly and repair, protein turnover, and sensing and responding to mechanical forces ('mechanotransduction'). Additionally, the Z-discs are intimately connected to other structures in the cardiomyocyte, including costameres and intercalated discs; these connections are critical for mechanotransduction (Knoll and Buyandelger, 2011).
**The M-band**

M-bands are the sites at which antiparallel thick filaments are crosslinked at the center of the sarcomere (Fig 1.3, Fig 1.4). Compared to the Z-disc, relatively little is known about the organization, protein components, and functions of the M-band. Recent studies have indicated that one function of the M-band may be to absorb force imbalances generated by opposing myosin thick filaments, thus serving as a sarcomeric “shock absorber” during active contraction (Agarkova and Perriard, 2005; Tskhovrebova and Trinick, 2012; Xiao and Grater, 2014).

The members of the myomesin gene family (MYOM1, MYOM2, MYOM3) are important for linking the thick filaments. These proteins share a common domain architecture, with an intrinsically unstructured head domain, followed by twelve immunoglobulin (Ig) and fibronectin type III (Fn) domains (Lange et al., 2019). Specific myomesin proteins and isoforms have been associated with certain developmental stages and disease. A particular splice isoform of MYOM1 (embryonic heart myomesin, EH-myomesin) is highly upregulated in DCM hearts (Schoenauer et al., 2011). Additionally, MYOM2 expression is upregulated only later during heart development, around birth (Grove et al., 1985; Carlsson et al., 1990; Schoenauer et al., 2011). It is speculated that these different proteins and isoforms influence the passive elasticity of the sarcomere similar to the N2A and N2B isoforms of TTN.

Several other proteins have been found to localize to the M-band. The C-terminus of TTN inserts into the M-band and contains a Ser/Thr-kinase domain. This domain likely lacks catalytic activity and rather, provides scaffolding for MuRF family proteins, E3-ubiquitin ligases that are important for the turnover of myosin heavy chain
Obscurin, is another giant M-band protein, consisting of Ig domains, linker domains, Fn domains, and two C-terminal kinase domains, that may be involved in the regulation of cell adhesion (Hu and Kontrogianni-konstantopoulos, 2001). Other protein kinases at the M-band include the muscle specific isoform of creatine kinase (M-CK), which binds the central domains of myomesin proteins. Phosphorylation also appears to regulate myomesin binding; MYOM1 phosphorylation by PKA inhibits its binding to titin and MYOM2 phosphorylation by PKA inhibits its binding to myosin in vitro (Obermann et al., 1998; Lange et al., 2019). The importance of phosphorylation at the M-band is further underscored by the M-band phosphatase PP2A, which is important for sarcomere formation in C. elegans (Qadota et al., 2009).

**The costamere and cytoskeletal elements**

Costameres are specialized membrane complexes that physically connect the Z-disc, the sarcolemma, and the ECM. Costameres are constructed from focal adhesion protein complexes, including the integrin-vinculin-talin complex, the dystrophin-glycoprotein complex, and the sarcoglycan complex (Anastasi et al., 2004). Costameres are the sites at which sarcomere assembly (sarcomerogenesis) initiates (Chopra et al., 2018). These sites are also critical for mechanotransduction and the repair and maintenance of sarcomeres in response to pressure-overload and stress (Sequeira et al., 2014; Samarel, 2020).
The intercalated disc

Cardiomyocytes are electrically and mechanically coupled with one another through specialized intercellular junctions called intercalated discs. Intercalated discs are composed of three main junctional complexes: desmosomes, fascia adherens, and gap junctions (Manring et al., 2018).

**Gap junctions:** Each gap junction is composed of 12 connexin proteins, 6 from each neighboring cardiomyocyte, which couple to form a channel pore. In the mammalian ventricular cardiomyocyte, connexin-43 is the most prominent connexin. An assembly of 5-500 gap junctions constitutes a gap junction plaque, which facilitates the propagation of electrical signals between cardiomyocytes (Asimaki, Kleber and Saffitz, 2015).

**Adherens junctions:** The adherens junction protein complex is composed of transmembrane cadherins and cytosolic catenins. Through its extracellular domain, N-cadherin bridges to a neighboring cardiomyocyte via the formation of a calcium-dependent homodimer. The intracellular portion of N-cadherin is linked via catenins with the actin cytoskeleton. Thus, adherens junctions couple the sarcolemma with the cardiomyocyte cytoskeleton and facilitate the transmission of contractile forces between cardiomyocytes (Asimaki, Kleber and Saffitz, 2015).

**Desmosomal junctions:** The desmosomal complex is composed of desmosomal cadherins (desmoglein, desmocollin), which similar to other cadherins, link neighboring cardiomyocytes via dimerization. On the intracellular side, proteins of the armadillo/catenin (plakoglobin and plakophilin) and plakin (desmoplakin) family link desmosomal cadherins to the intermediate filament desmin. These tight connections are
thought to be necessary for cardiomyocytes to withstand stress generated by contractile forces (Asimaki, Kleber and Saffitz, 2015).
Cardiomyocyte function

Excitation-contraction coupling

Excitation-contraction coupling is an elaborate process that converts an electrical action potential into a Ca\(_{2+}\)-dependent chemical signal to trigger sarcomere contraction. This process occurs in a highly-regulated sequence (Bers, 2012).

1. Pacemaker cells spontaneously produce action potentials, which are propagated along the sarcolemma. The sarcolemma contains specialized invaginations called T-tubules which penetrate into the interior of the cardiomyocyte, and facilitate synchronous activation of the entire cardiomyocyte (Ibrahim et al., 2011).

2. An action potential activates dihydropyridine receptors (DHPR) (also called L-type Ca\(_{2+}\) channels LTCC) on T-tubule membranes, allowing Ca\(_{2+}\) to pass into the cell.

3. Ca\(_{2+}\) binds to ryanodine receptors (RYR2) localized on the sarcoplasmic reticulum (SR). The SR is a specialized structure that stores Ca\(_{2+}\), that is similar to the endoplasmic reticulum of other cells.

4. RYR2 receptors open, triggering Ca\(_{2+}\)-release from the SR (Ca\(_{2+}\)-induced-Ca\(_{2+}\)-release).

5. The rise in intracellular Ca\(_{2+}\) concentration causes DHPRs to close, preventing further Ca\(_{2+}\) influx.

6. Ca\(_{2+}\) binds the troponin complex, triggering conformational changes in the troponin-tropomyosin regulatory apparatus which exposes actin-binding sites for
myosin, and initiates the process of sarcomere contraction (refer to
"chemomechanical cycle" below).

7. For relaxation to occur, the concentration of intracellular Ca\textsuperscript{2+} must decline.

Several proteins are involved in this process (Bers, 2012).

a. SERCA2A is a Ca\textsuperscript{2+}-ATPase located on the SR membrane. SERCA2A
sequesters Ca\textsuperscript{2+} back into the SR after contraction. SERCA2A activity is
regulated by phospholamban (PLN), which inhibits SERCA2A via
phosphorylation-dependent mechanisms (Gustavsson \textit{et al.}, 2013).

b. Na\textsuperscript{+}/Ca\textsuperscript{2+} exchangers and sarcolemmal Ca\textsuperscript{2+}-ATPases on the sarcolemma
transport Ca\textsuperscript{2+} out of the cardiomyocyte (Bers, 2012).

c. Mitochondrial Ca\textsuperscript{2+}-uniporters sequester Ca\textsuperscript{2+} in the mitochondrial matrix
(Kwong, 2017).

\textbf{The chemomechanical cycle of sarcomere contraction}

The process by which chemical energy is converted into the mechanical process of
sarcomere contraction is called the chemomechanical cycle. The chemomechanical
cycle involves several steps (Garfinkel, Seidman and Seidman, 2018):

(1) In low Ca\textsuperscript{2+} conditions, tropomyosin sterically inhibits myosin binding sites on the
actin thin filament. High intracellular Ca\textsuperscript{2+} concentration – triggered by an action
potential – results in Ca\textsuperscript{2+} binding to troponin C, resulting in conformational
changes in Troponin I and Troponin T. Conformational changes in the troponin
complex releases tropomyosin-mediated steric inhibition of the thin filament, thus
exposing myosin binding sites.
(2) ATP binds to an actin-bound myosin head, causing myosin dissociation from the thin filament.

(3) The S1 myosin ATPase hydrolyzes ATP, resulting in a conformational change in the myosin head (‘pre-power stroke’ conformation).

(4) Release of Pi results in myosin binding at a new position along the actin filament (crossbridge formation), resulting in sarcomere shortening (power stroke).

(5) Release of ADP allows new ATP binding, initiating another chemomechanical cycle, and continued sarcomere shortening.

Several additional factors regulate crossbridge formation and thus influence sarcomere power output. Firstly, the power output of sarcomeres is length-dependent; while a moderate amount of stretch in diastole increases power output during contraction, sarcomeres that are either ‘too short’ or ‘too long’ have decreased power output due to complex physiologic mechanisms (Frank-Starling mechanism). Titin is particularly important for maintaining passive tension and sarcomere elasticity, thus reducing the length-dependency of the sarcomere. Additionally, several additional proteins and post-translational modifications can influence sarcomere contractility. In particular, dephosphorylated MYBPC3 limits thick filament sliding, thereby acting as a functional 'brake' on the sarcomere (Coulton and Stelzer, 2012; Garfinkel, Seidman and Seidman, 2018). In contrast, phosphorylation of the myosin regulatory light chain (MYL2) induces movement of myosin toward the thin filament and facilitates contraction (Yu et al., 2016).
**Sarcomere relaxation**

Once a contractile cycle is complete, the sarcomere returns to its relaxed state. Relaxation is an active process that involves distinct structural conformations of myosins, which are associated with different rates of ATP hydrolysis (Alamo et al., 2017).

Paired myosin heads can adopt distinct structures during relaxation. The interacting-heads-motif (IHM) is a configuration of paired myosin heads in relaxed muscle, in which actin-binding and ATP turnover are inhibited by the interaction of each head with the other. When both heads dock onto the thick filament and interact through the IHM, ATPase activity is maximally inhibited (super-relaxed state, SRX). Alternatively, one of the two heads can bind onto the thick filament, leaving the other head free (disordered-relaxed state, DRX) (Alamo et al., 2017; Garfinkel, Seidman and Seidman, 2018).

SRX myosins hydrolyze ATP slowly, while DRX myosins have a 5-fold higher rate of ATP turnover (Hooijman, Stewart and Cooke, 2011). Thus, the proportion of heads in SRX versus DRX can profoundly affect cardiomyocyte energy demand. Studies in cardiomyocytes have shown that under physiologic conditions, approximately 60% of heads are in the SRX conformation, while 40% are in the DRX conformation. Hibernation increases the SRX/DRX ratio, suggesting that the SRX conformation may have evolved to reduce cardiac metabolic demand. Proteins that interact with myosin, including MYBPC3, modulate the SRX/DRX ratio, thereby influencing cardiac diastolic function (Toepfer, Wakimoto, et al., 2019).
**Molecular mechanisms of cardiomyopathy and therapeutic avenues**

Studies over the past decade have begun to elucidate the mechanisms by which damaging genetics variants induce cardiac dysfunction. While a comprehensive review of this topic is beyond the scope of this dissertation, we review a few key emerging mechanistic understandings of cardiomyopathy.

**Deleterious missense variants in MYH7 and loss-of-function variants in MYBPC3 alter myosin IHM conformations and perturb cardiomyocyte relaxation**

The most common variants associated with HCM (i.e. deleterious missense variants in MYH7 and loss-of-function variants in MYBPC3) perturb the IHM and are associated with a decreased SRX/DRX ratio (Bernstein et al., 2019; Toepfer, Wakimoto, et al., 2019). Furthermore, phosphorylation of MYBPC3 has been shown to release myosin heads from the SRX state (Mcnamara, Singh and Sadayappan, 2019). While MYH7 and MYBPC3 variants generally enhance sarcomere contractility, perturbation of the IHM impairs relaxation. MYH7 pathogenic missense variants that affect IHM residues also predict an increased risk for heart failure and atrial fibrillation (Garfinkel, Seidman and Seidman, 2018). Perturbation of the IHM, and a consequent decrease in the SRX/DRX ratio, may be a unifying mechanism that explains the diastolic dysfunction and increased metabolic demand that is characteristic of HCM. Whether non-thick-filament HCM variants, and other DCM variants, also perturb IHM conformations remains an area of active investigation.

Molecular strategies to normalize the SRX/DRX ratio may be therapeutic for patients with MYH7 and MYBPC3-related HCM. MYK-461 is a new small molecule allosteric modulator of the myosin ATPase (Green et al., 2016). MYK-461 administration in mice, and hiPSC-CMs carrying damaging MYH7 or MYBPC3 variants, restores
normal sarcomere contractility and rescues relaxation defects (Green et al., 2016). This molecule also normalizes the SRX/DRX ratio, suggesting that the proportion of myosin heads in the SRX vs DRX states is directly related to the pathogenesis of cardiac dysfunction in HCM (Toepfer, Wakimoto, et al., 2019; Toepfer, 2020). Thus, for variants that perturb the SRX/DRX ratio, administration of MYK-461, or other small molecules that can normalize this ratio, may be therapeutic. Importantly, the SRX/DRX ratio can be assayed using fresh or frozen cardiac biopsy tissue, indicating that this approach may have direct clinical utility.

**Deleterious missense variants in MYH7 perturb myosin motor domain function**

Unlike MYH7 HCM variants, damaging MYH7 missense variants in DCM have modest effects on IHM interactions, and minimal impact on cardiomyocyte relaxation. However, these variants appear to directly impair the myosin motor domain ATPase function, and thus substantially reduce sarcomere contractile function (Garfinkel, Seidman and Seidman, 2018). Whether other DCM variants similarly impair sarcomere contractility via effects on the myosin motor domain remains an area of active investigation.

Therapeutic strategies to directly activate the myosin ATPase may be beneficial for DCM variants that impair motor activity. A new small molecule, omecamtiv mecarbil, is one such allosteric regulator of myosin that has been shown to improve cardiac performance in animal models of heart failure (Malik et al., 2019). It was originally believed that this molecule allosterically increased the transition rate of myosin into the strongly actin-bound force generating state (Malik et al., 2019). However, newer studies suggest that in fact, this molecule suppresses myosin’s power stroke, prolongs
actomyosin attachment, and indirectly increases thin-filament Ca\textsuperscript{2+} sensitivity via unclear mechanisms (Woody et al., 2018). Additional therapeutic strategies to either (1) allosterically activate the myosin ATPase or (2) sensitize the regulatory proteins to calcium, thereby exposing myosin-binding sites on the thin filament, may be therapeutic in DCM.

**Pathogenic variants in desmosome genes impair trafficking of intercalated disc components and inhibit electrical propagation of action potentials**

Approximately 60% of ARVC probands harbor dominant (usually missense or loss-of-function) variants in genes that encode desmosome proteins. The mechanisms of these pathogenic variants likely relate to haploinsufficiency, as earlier-onset and more severe cases are often caused by compound or digenic heterozygosity (Asimaki, Kleber and Saffitz, 2015).

In the majority of cases of ARVC, the amount of immunoreactive signal for plakoglobin is depressed at intercalated discs, though the protein level is not changed, suggesting an abnormality in protein trafficking. These changes are generally not observed in patients with ischemic heart disease or other forms of cardiomyopathy, indicating that the redistribution of plakoglobin is a specific hallmark of ARVC (Asimaki, Kleber and Saffitz, 2015). Most cases of ARVC also show reduced immunoreactive signal for Cx43, the major ventricular gap junction protein (Fidler et al., 2009; Asimaki, Kleber and Saffitz, 2015). Given the role of gap junctions in electrical propagation, it is possible that these changes increase arrhythmogenesis.

The specific effects of desmosomal variants on cell-cell adhesion, and cardiomyocyte signaling, remain an area of active investigation. However, recent studies have reported that plakoglobin redistribution leads to aberrant activation of the
Hippo pathway and suppression of the canonical Wnt signaling (Asimaki, Kleber and Saffitz, 2015). In a transgenic zebrafish model carrying a two base pair deletion in the gene encoding plakoglobin, a damaging variant that in humans causes a cardiocutaneous syndrome called Naxos disease, drug screening revealed a compound that was able to prevent bradycardia and contractile deficits in mutant fish and normalize survival (Asimaki et al., 2014). This compound, SB216763, has been previously annotated as an inhibitor of the GSK-3β, and hence an activator of the Wnt signaling pathway. This compound was also able to normalize abnormal action potentials in neonatal rat ventricular myocytes expressing mutant forms of plakophilin (Asimaki et al., 2014). Treatment of patient-derived hiPSC-CMs harboring damaging plakophilin variants with a different Wnt activator, 6-bromoindirubin-3'-oxime, has been shown to result in plakoglobin redistribution, and ameliorate calcium-handling defects (Lombardi, Cabreira-hansen and Bell, 2012). These observations suggest that ARVC may be related to a common disease mechanism of inhibited Wnt and Hippo signaling, which interferes with the proper trafficking of key intercalated disc proteins (Asimaki, Kleber and Saffitz, 2015). Therapeutics that restore the proper trafficking of these proteins may have clinical utility in ARVC. The precise mechanisms by which intercalated disc components are trafficked and assembled remains an area of active investigation.

**TTNtvvs and damaging variants in costamere components result in defects in sarcomere assembly and maintenance**

TTN truncating variants (TTNtvvs) in DCM patients are markedly enriched in the A-band region compared to the I-band region of TTN. Recent studies have demonstrated that alternative splicing of TTN excludes I-band exons from the major
N2A and N2B transcripts, thus reducing the functional consequences of I-band TTNtv variants (Herman et al., 2012). Three dimensional cardiac microtissues (CMTs) constructed from hiPSC-CMs with A-band TTNtv variants show defects in sarcomere assembly and reduced contractile force (Hinson et al., 2015). Further studies have since demonstrated that these effects are likely due to the critical role of TTN in transmitting myosin-generated tension from the costamere, which is essential for sarcomere assembly and stability at the Z-disc (Chopra et al., 2018).

Pathogenic missense and loss-of-function variants in other costamere components are also genetic causes of DCM. These include damaging variants in vinculin, δ-sarcoglycan, and dystrophin, which also cause skeletal muscular dystrophies (Sequeira et al., 2014). As costameres are critical for sarcomere assembly and force transmission, it is possible that genetic loss of these components results in sarcomere insufficiency phenotypes, similar to those induced by TTNtv variants.

Titin, dystrophin, and other costamere components often encode very large protein products, which has complicated efforts for gene replacement, as viral delivery vectors have limited cargo sizes. One important observation from studies of TTNtv variants, and dystrophin variants in Duchenne Muscular Dystrophy (DMD) and DCM, is that partially truncated forms of these proteins can retain and restore function. For DMD, many of the thousands of pathogenic variants cluster between exon 45 and 50 of the dystrophin gene. The development of antisense oligonucleotides to promote exon skipping, and thus restore the dystrophin open reading frame, has proven to be a successful strategy in this disease (Repetti et al., 2019). It is possible that similar strategies to restore appropriate open reading frames may be therapeutic for TTNtv variants,
and other DCM variants. Furthermore, the isoform composition of TTN and certain other sarcomere proteins, can exert large effects on the biomechanical properties of sarcomeres. Genetic strategies to modulate TTN isoform composition – preferentially upregulating stiffer or more compliant isoforms – can affect the passive tension of sarcomeres, and may be especially therapeutic for cardiomyopathies associated with diastolic dysfunction.

**Deleterious missense and truncating variants in Z-disc proteins induce protein aggregation and can impair normal autophagic pathways**

Missense and truncating variants in Z-disc proteins, including desmin, αβ-crystallin, myotilin, and ZASP, can lead to both skeletal muscle myofibrillar myopathy (MFM) and/or DCM (Fürst et al., 2013). Variants in these genes often cause protein aggregation of Z-disc components (Mclendon and Robbins, 2019). MFM results in progressive skeletal muscle disintegration with characteristic large cytoplasmic aggregates, which are thought to impair autophagic processes in cardiomyocytes over time, thereby resulting in progressive cardiomyocyte dysfunction (Ruparelia et al., 2016). Damaging missense and truncating variants in the Z-disc component, FLNC, have recently been identified as novel causes of HCM/RCM and DCM (Valdés-Mas et al., 2014; Brodehl et al., 2016; Ortiz-Genga et al., 2016; Begay et al., 2018). I further explore the pathogenic mechanisms of FLNC variants in chapter 2 of this dissertation.

**Loss of LINC-complex components and altered nuclear membrane integrity results in aberrant cardiomyocyte gene expression**

Damaging variants (missense, truncating, splice) in LMNA account for 5-10% of DCM cases, and are associated with a high incidence of arrhythmia and sudden cardiac death. LMNA encodes the intermediate filament proteins lamins A and C, which are
generated by alternative splicing, and polymerize with B-type lamins to form lattice structures of the nuclear lamina (NL) (Brayson and Shanahan, 2017). The nuclear lamina lies directly adjacent to the inner nuclear membrane (see Fig 1.5) and is intimately connected to the cytoskeleton through the Linkers of the Nucleoskeleton to Cytoskeleton (LINC) complex, which contains several proteins including emerin, SUN domain proteins, and nesprins (Meinke, Nguyen and Wehnert, 2011).

Several hypotheses have been proposed to explain the pathogenicity of LMNA variants. One hypothesis suggests that LMNA and the LINC complex is critical for proper chromatin organization and DNA transcription, and that loss of LINC complex integrity results in aberrant gene expression. Consistent with this hypothesis, patient-derived hiPSC-CMs containing damaging LMNA variants show aberrant activation of the PDGF signaling pathway. Pharmacological and molecular inhibition of the PDGF signaling pathway ameliorates the arrhythmic phenotypes of mutant hiPSC-CMs in vitro (Lee et al., 2019).

The importance of nuclear structure is emphasized by other pathogenic DCM variants in genes that encode components of the LINC complex. A deleterious missense variant in TMEM43, which encodes a protein that interacts with lamins, causes severe DCM, characterized by a high incidence of premature SCD (Baskin et al., 2013; Christensen et al., 2015). Variants in emerin, another key component of the LINC complex also cause a highly arrhythmogenic form of DCM. Human iPSC-CMs bearing emerin variants result in alterations in nuclear calcium transients, suggesting that the LINC complex may have critical roles in both determining nuclear structure and in Ca_{2+} handling (Shimojima et al., 2017).
Furthermore, recent studies have indicated that components of the LINC complex are crucial for the proper assembly and stability of the sarcomere network. It has been suggested that analogous to costameres, there may exist similar nascent complexes that co-localize with the nucleus and determine aspects of sarcomere assembly. In *Drosophila*, the Z-disc protein, ZASP, assembles into complexes that colocalize with the nucleus before ZASP organizes throughout myocytes; this positioning was dependent on the LINC complex (Auld and Folker, 2016). Thus, it may be possible that damaging variants in genes that disrupt the LINC complex or nuclear envelope integrity may additionally perturb sarcomere assembly, force transmission, and/or mechanosensitive gene expression. The mechanisms of signaling between the nucleus and the sarcomere are further explored in chapter 3 of this dissertation.
Overview of dissertation

Molecular characterizations of causal cardiomyopathy variants have revealed that diverse mechanisms can induce cardiomyocyte dysfunction. Understanding these mechanisms can inform therapeutic targets for heart failure, a disease for which there are few effective therapies.

Recently, damaging variants in two genes – *FLNC* (filamin C) and *ALPK3* (α-kinase-3) – have been identified as novel genetic causes of cardiomyopathy. *FLNC* encodes an actin-crosslinking Z-disc protein, while *ALPK3* encodes a putative nuclear kinase of unknown function. The mechanisms by which damaging variants in these genes cause cardiomyopathy are unknown. This dissertation employs powerful technologies - CRISPR/Cas9 genome editing, hiPSC-CMs and mouse models, and unbiased molecular characterizations via RNAseq, proteomics and phosphoproteomics - to elucidate the mechanisms of damaging variants in *ALPK3* and *FLNC*. Our studies demonstrate the power of studying multiple layers of gene regulation – extending from transcription, translation, to phosphorylation – in revealing protein function, disease mechanisms, and novel therapeutic targets.
Chapter 2:

*Filamin C* truncating variants that cause dilated cardiomyopathy result in toxic protein accumulation and increased lysosome abundance in hiPSC-CMs

This chapter represents a manuscript being prepared for submission:

*Filamin C* truncating variants that cause dilated cardiomyopathy result in toxic protein accumulation and increased lysosome abundance in human cardiomyocytes. Radhika Agarwal, Joao A. Paulo, Christopher N. Toepfer, Jourdan K. Ewoldt, Anant Chopra, Joshua Gorham, Steven R. DePalma, Christopher S. Chen, Steven P. Gygi, Christine E. Seidman, J. G. Seidman.
Abstract:

The mechanisms by which dominant truncating variants in FLNC cause dilated cardiomyopathy (DCM) are poorly understood. We used CRISPR/Cas9 technology to create a series of FLNC mutant human induced pluripotent stem cell lines, differentiated these cell lines into cardiomyocytes, and assessed the consequences of FLNC mutations on sarcomere structure, contractile function, and protein expression. We found that some FLNC expression was necessary for the coordinated expression of actin-binding and sarcomere thin filament genes, and for actin-crosslinking at sarcomere Z-discs, the lack of which drastically impaired sarcomere assembly in FLNC-null hiPSC-CMs. However, sarcomere assembly was not significantly impaired in FLNC-heterozygote hiPSC-CMs. Rather, FLNC heterozygote variants resulted in accumulation of FLNC-binding partners and increased cellular lysosome content. Our studies indicate that the mechanisms by which FLNC variants cause DCM relate to toxic protein accumulation. These data link the pathophysiology of FLNC DCM variants to other cardiomyopathies with inadequate lysosome-dependent degradation of proteins, underscoring the critical role of protein turnover in cardiomyocyte health, and its dysfunction as a pathway for cardiomyopathy.
Introduction

Dilated cardiomyopathy (DCM), characterized by pathologic cardiac enlargement and contractile dysfunction, represents a leading cause of heart failure. A genetic etiology for non-ischemic DCM can be identified in over 20% of patients (Schultheiss et al., 1984). Linkage studies and exome and whole genome sequencing of non-ischemic DCM cohorts have revealed that the genetic architecture of DCM is complex, with variants in over 40 genes predicted to cause disease (Cho, Lee and Kim, 2016). Variants in one of these genes, FLNC, encoding filamin C, represents one of the more common genetic causes of DCM, with a prevalence of 3-4% in non-ischemic DCM cohorts (Ortiz-Genga et al., 2016; Janin et al., 2017; Begay et al., 2018; Ader et al., 2019). Dominant truncating variants in FLNC cause a highly penetrant and malignant form of DCM that presents in early to mid-adulthood and is frequently complicated by ventricular arrhythmias and sudden cardiac death (Ortiz-Genga et al., 2016; Begay et al., 2018; Peters, Kumar and Fatkin, 2019). The mechanisms by which FLNC variants cause DCM are poorly understood.

Filamins are a class of large dimeric proteins, consisting of 270-290kDa subunits, that crosslink filamentous (F-) actin (Hartwig and Stossel, 1981; Thompson et al., 2000). Filamins have a modular structure consisting of an aminoterminal actin-binding domain (ABD) followed by 24 repeats, all of which adopt an immunoglobulin-like (Ig) fold. The last of these Ig repeats mediates protein dimerization (Flier and Sonnenberg, 2001; Pudas et al., 2005). Although FLNC has the same overall structure as the other two filamin isoforms (FLNA, FLNB), it is unique in that it is expressed highly and specifically in cardiac and skeletal muscle (Ven et al., 2000; Razinia et al., 2012).
Within striated muscle, FLNC has important roles in maintaining the structure of sarcomeres, the fundamental contractile units of muscle cells. Each sarcomere consists of overlapping parallel arrays of myosin-containing thick filaments, and actin-containing thin filaments (Fig 2.1). Thin filaments are anchored at specialized structures called Z-discs, located at the ends of each sarcomere (Squire, 2016). A unique insertion in FLNC Ig-like domain 20 targets FLNC to Z-discs (Ven et al., 2000), where it crosslinks actin, keeping thin filaments stably anchored (González-Morales, Holenka and Schöck, 2017). FLNC also localizes, in smaller part, to costameres, specialized focal adhesion complexes that anchor sarcomeres to the muscle cell membrane (Thompson et al., 2000; Kiema et al., 2006) (Fig 2.1).

At both Z-discs and costameres, FLNC interacts with a large number of proteins, primarily through its C-terminal Ig domains (Ig 16-24). These include interactions with the dystrophin-glycoprotein-complex, γ-sarcoglycan (SGCG) (Thompson et al., 2000), β1-integrin tail (ITGB1) (Kiema et al., 2006), costamere-associated protein (SORBS1) (Zhang et al., 2007), migfilin (FBLIM1) (Tu et al., 2003), aciculin (PGM5) (Molt et al., 2014), xin proteins (XIRP1, XIRP2) (Molt et al., 2014), synaptopodin 2 (SYNPO2) (Linnemann et al., 2010), myotilin (MYOT) (Ven et al., 2000), myozenis (MYOZ1, MYOZ2) (Takada et al., 2001), titin (TTN) (González-Morales, Holenka and Schöck, 2017), and heat shock proteins (HSPB1(Collier et al., 2019), HSPB7(Juo et al., 2016)).

The consequences of FLNC mutations on striated muscle biology have been explored in several studies. Prior to its recognition as a causal cardiomyopathy gene, FLNC dominant truncating variants were found to cause myofibrillar myopathy (MFM), a skeletal muscle disease that results in myofibrillar destruction with characteristic large
cytoplasmic protein aggregates (Kley et al., 2007). However, such aggregates have not been previously identified in cardiac biopsies of DCM patients, leading to speculation that FLNC truncating mutations cause cardiomyopathy via alternative disease mechanisms (Ortiz-Genga et al., 2016). Genetic ablation of Flnc in mouse, Drosophila, and medaka results in abnormal Z-disc morphology, suggesting that Flnc depletion may compromise sarcomere structural integrity, especially under conditions of mechanical strain (Dalkilic et al., 2006; Fujita et al., 2012; González-Morales, Holenka and Schöck, 2017). However, the precise pathogenic mechanisms of FLNC heterozygote variants in human DCM have remained elusive.

In this study, we employed human induced pluripotent stem-cell-derived cardiomyocytes (hiPSC-CMs) to investigate the endogenous role of FLNC and the mechanisms by which damaging FLNC variants cause DCM. Using CRISPR/Cas9, we introduced a series of variants in FLNC, resulting in four mutant isogenic hiPSC lines: one with homozygous null alleles, two lines with different heterozygote variants, and one with a homozygous missense (G1674S) variant of unknown significance (VUS) that was identified in a cardiomyopathy patient. Consistent with prior studies, we found that some FLNC expression was necessary for sarcomere assembly and Z-disc integrity, as well as for the coordinated expression of actin-binding and sarcomere thin-filament genes. However, sarcomere gene expression and assembly was not significantly impaired in hiPSC-CMs carrying FLNC heterozygote variants. Rather, FLNC heterozygote variants caused protein accumulation of FLNC-binding partners and increased cellular lysosome content. Our investigations establish FLNC-DCM as a
disease of toxic protein accumulation and suggest shared disease mechanisms between FLNC-DCM and FLNC-MFM.
Figure 2.1. Sarcomere structure and FLNC localization in hiPSC-CMs. Schematic representation of the cardiac sarcomere (the unit of contraction that spans 2 neighboring Z-discs). The sarcomere is a highly organized structure consisting of parallel arrays of thick and thin filaments. Myosin-containing thick filaments are anchored at the M-band, while actin-containing thin filaments are anchored at Z-discs. FLNC localizes primarily to Z-discs, where it crosslinks actin contained in thin filaments. FLNC also localizes to costameres, specialized focal adhesion complexes that tether sarcomeres to the muscle cell membrane. The giant protein titin (TTN) spans an entire half-sarcomere length with the N-terminus inserting in the Z-disc and the C-terminus in the M-band.
Results

CRISPR/Cas9 genome engineering of FLNC mutant hiPSC lines

Using CRISPR/Cas9, we created a series of FLNC mutations in the human induced pluripotent stem cell (hiPSC) line PGP1-TTN-GFP which contains an endogenous eGFP tag on the N-terminus of the sarcomere protein titin (eGFP-TTN) (Sharma, Toepfer, Ward, et al., 2018) that allows functional analyses of hiPSC-CMs (Sharma, Toepfer, Schmid, et al., 2018; Toepfer, Sharma, et al., 2019b). We targeted FLNC exon 30, a region that encoded the 15th immunoglobulin domain (Ig 15) (Fig 2.2a). Using homology-directed repair, we also introduced a FLNC variant of unknown significance (VUS), G1674S, which was identified in a human patient with HCM. We produced wild-type (WT) and four isogenic FLNC mutant hiPSC lines: FLNC-/-, FLNC+/-, FLNC+/∆7aa, and FLNG1674SG1674S (Fig 2.2b), with variants on one or both alleles. The FLNC+/∆7aa occurred from spurious CRISPR mutagenesis and deleted amino acids 1668-1674 from within the 15th immunoglobulin domain on one allele. All hiPSC lines were subcloned and genotypes were confirmed via next-generation sequencing (METHODS; Fig S2.1).

Using an established small-molecule-based differentiation protocol (Lian et al., 2013), we differentiated WT and each of the FLNC mutant cell lines into hiPSC-CMs. Differentiations of all cell lines produced beating cardiomyocytes. To assess whether cell lines achieved comparable states of cardiomyocyte maturation, we sequenced RNA from WT and FLNC mutant hiPSC-CMs at day 30 of cardiomyocyte differentiation, and applied principal component analysis to compare expression profiles to those of
cardiomyocytes at earlier stages of maturation (Fig S2.2a). These data indicated that WT and all FLNC mutant hiPSC lines achieved similar states of cardiomyocyte maturation after 30 days in culture, and expressed appropriate cardiomyocyte markers (Fig S2.2a-b).

To investigate protein expression in WT and FLNC mutant hiPSC-CMs, we applied quantitative mass-spectrometry-based proteomics, specifically tandem-mass-tag (TMT), to protein extracts from each hiPSC-CM line. In this approach, equal amounts of protein from each sample are digested into peptides. Peptides are subsequently tagged with unique, sample-specific isobaric chemical labels, which allows the relative abundance of peptides amongst samples to be determined by mass spectrometry. Up to 11 samples can be analyzed in a single TMT experiment (‘11-plex’) by mass spectrometry, which typically results in the relative quantification of several thousand proteins.

Two TMT experiments were conducted: one 10-plex and one 11-plex. In the 10-plex experiment, the relative expression of 7,123 proteins from WT (n=5) and FLNC+/− hiPSC-CMs (n=5) was determined. In the 11-plex experiment, the relative expression of 6,863 proteins from WT (n=3), FLNC+/- (n=3), FLNC+Δ7aa (n=2), and FLNCG1674S/G1674S (n=3) hiPSC-CMs was determined. Each replicate (n) represents protein lysate harvested from a separate differentiation of hiPSC-CMs, and was analyzed independently.

To determine sample quality, hierarchical clustering was applied to the protein expression data from the 10-plex and 11-plex TMT experiments to determine the influence of genotypes and potential confounders including batch effects. The protein
expression of hiPSC-CMs clustered by genotype in all but one case, which revealed a batch effect; we excluded the outlier WT and FLNC−/− sample from subsequent analysis (Fig S2.2c-d).

Using RNAseq and proteomic data, we first examined FLNC transcript and protein expression in each hiPSC-CM line. Two mutant hiPSC-CM lines (FLNC+/Δ7aa and FLNCG1674S/G1674S) had comparable FLNC transcript levels to WT (Fig 2.2b). By contrast, these transcripts were 33% and 7% of WT levels in FLNC+/- and FLNC−/− hiPSC-CMs. FLNC+/Δ7aa and FLNCG1674S/G1674S lines had comparable levels of FLNC protein to WT hiPSC-CMs, while FLNC+/- and FLNC−/− lines had approximately 50% and 6% of normal levels (Fig 2.2b). These effects on FLNC protein expression in FLNC mutant hiPSC-CMs were confirmed by Western blotting with both an N-terminal and C-terminal FLNC antibody (Fig 2.2c).

Our mass spectrometry analyses identified over 100 peptides that mapped to FLNC, which was significantly higher than the average number (~8) of mapped peptides per protein. The number of mapped peptides per protein relates to both protein size (larger proteins = more peptides) and abundance (more abundant proteins = more peptides). To determine the absolute abundance of FLNC, intensity-based absolute quantification (iBAQ) was applied, an estimate which corrects for the influence of protein size (Arike et al., 2012). The iBAQ score for FLNC ranked in the 97.8th percentile, indicating that FLNC is a highly abundant protein in hiPSC-CMs. The five proteins with the highest iBAQ scores were striated muscle actin (ACTC1/ACTA1), histone H4 (H4), myosin heavy chain 7 (MYH7), tubulin alpha 1b (TBA1B), and myosin regulatory light chain 2 (MYL2).
Figure 2.2. CRISPR/Cas9 genome engineering of FLNC mutant hiPSC lines and effects on FLNC expression.

(A) Structure of FLNC, depicting an N-terminal actin-binding domain (red) and 24 immunoglobulin-like (Ig) domains (blue), the last of which mediates protein dimerization. The CRISPR/Cas9 guide RNA targeted FLNC exon 30, corresponding to a region in the 15th Ig domain. Symbol (scissors) indicates relative position of CRISPR/Cas9 cut site.

(B) The effects of FLNC mutations on RNA and protein levels in hiPSC-CMs were quantified by RNAseq and mass spectrometry respectively (refer to METHODS). For RNAseq, the mean RPKM is shown. Symbol (*) highlights nonsense-mediated decay of the mutant mRNA transcript. For mass spectrometry, data are displayed as mean ± 95% confidence interval. The number of replicates for each analysis is indicated in parentheses.

(C) Western blots of FLNC mutant day 30 hiPSC-CM protein extracts probed with N-terminal (top panel) and C-terminal (bottom panel) FLNC antibodies. Numbers below blots indicate quantification of the relative intensity of FLNC to TUBB, normalized to wild-type. Loading controls: β-tubulin (TUBB) and cardiac troponin T2 (TNNT2).
**FLNC is necessary for the expression of actin and thin filament genes, F-actin crosslinking, and sarcomere assembly**

We first examined gene expression in FLNC−/− hiPSC-CMs, which lacked both FLNC RNA and protein expression. Comparison of the proteomes of WT and FLNC−/− hiPSC-CMs revealed 136 differentially expressed (DE) proteins (Fig 2.3a), with significant enrichment of actin-binding and sarcomere proteins (Fig 2.3b). Overall, expression of these proteins was decreased in FLNC−/− hiPSC-CMs. To determine whether decreased protein expression was a consequence of transcriptional or post-transcriptional mechanisms, we compared RNA-level-expression between WT and FLNC−/− hiPSC-CMs. There was a strong correlation (r=0.91) between DE protein and RNA expression, indicating that the majority of DE protein expression in FLNC−/− hiPSC-CMs was attributable to changes in gene transcription (Fig 2.3c). Included amongst these DE transcripts and proteins were the FLNC-binding partner synaptopodin (SYNPO), thin filament-associated sarcomere components (tropomyosin (TPM2), troponin I (TNNI3), and troponin T (TNNT1)), as well as Z-disc components (synaptopodin 2-like (SYNPO2L), ankyrin repeat domain 1 (ANKRD1), PDZ and LIM domain 3 (PDLIM3), leomodin 2 (LMOD2), cysteine and glycine rich protein 3 (CSRP3)).

Given the effects of the FLNC−/− mutation on the expression of actin and sarcomere genes, we next directly examined actin and sarcomere structure in FLNC−/− hiPSC-CMs. We stained hiPSC-CMs for F-actin and co-visualized sarcomeres using endogenous expression of eGFP-TTN at the Z-discs. In comparison to WT, FLNC−/−
hiPSC-CMs showed reduced F-actin content (Fig 2.3d-e). Furthermore, FLNC⁻/⁻ hiPSC-CMs showed deficits in actin crosslinking, demonstrated by the variable size and breakage of actin bundles (Fig 2.3d, panels 3-4). These defects co-localized with regions of fragmented Z-discs, leading to overall deficiencies in sarcomere content (Fig 2.3e).

We considered whether cell culture and plating conditions were contributing factors by studying FLNC⁻/⁻ hiPSC-CMs plated onto surfaces with organized fibronectin patterns, as these surfaces normally promote sarcomere alignment and organization (Chopra et al., 2018). However, even on patterned surfaces, FLNC⁻/⁻ hiPSC-CMs still had significantly fewer striated sarcomeres at two and seven days post-plating (Fig 2.3f, Fig S2.3a-b). Furthermore, this phenotype was not a consequence of altered expression of the eGFP-TTN marker, as eGFP-TTN protein expression in FLNC⁻/⁻ hiPSC-CMs, assessed by both flow cytometry and mass spectrometry, was not significantly different from WT (Fig S2.3c-d). We deduced that the loss of appropriate actin and sarcomere gene expression, as well as defects in FLNC-mediated actin-crosslinking, contributed to the impaired assembly of sarcomeres in FLNC⁻/⁻ hiPSC-CMs.

Prior work has demonstrated that sarcomere assembly initiates from costameres (Chopra et al., 2018). As FLNC is partially localized to costameres (Zhang et al., 2007), we considered if these sites were aberrant in FLNC⁻/⁻ hiPSC-CMs. Using an antibody to paxillin, a major costamere component, we found appropriate immunofluorescent signals, focally localized at the cell periphery in patterned hiPSC-CMs (Fig 2.3f). From these data, we deduced that FLNC was not essential for costamere formation, though
loss of *FLNC* expression may have other effects on costamere composition and signaling.
Figure 2.3. *FLNC*−/− hiPSC-CMs have impaired expression of sarcomere and actin-related genes and loss of actin-crosslinking resulting in deficient sarcomere assembly

(A) Quantitative proteomics was used to determine the relative expression levels of 7,123 unique proteins in WT and *FLNC*−/− hiPSC-CMs. A total of 136 differentially expressed (DE) proteins were identified including 99 downregulated and 37 upregulated proteins (DE criteria: $|\log_2(\text{fold-change})| > 0.67$ and Benjamini-Hochberg adjusted $p$-value < 0.1). Each replicate column in heatmap represents an independent differentiation of each cell line.

(B) Enrichment analysis of DE proteins in *FLNC*−/− hiPSC-CMs.

(C) The correlation between *FLNC*−/− DE protein and RNA changes for DE sarcomere and actin-binding genes is strong ($r=0.91$). RNA changes were quantified by RNA sequencing (WT: $n=3$, *FLNC*−/−: $n=3$).

(D) F-actin staining of WT and *FLNC*−/− hiPSC-CMs. Scale bar = 25µm.

(E) Co-staining of F-actin (red) and eGFP-TTN marking sarcomere Z-discs (green), in WT and *FLNC*−/− hiPSC-CMs. Scale bar = 25µm.

(F) WT and *FLNC*−/− hiPSC-CMs, replated onto bioengineered surfaces containing fibronectin patterns (green = eGFP-TTN marking sarcomere Z-discs; purple = paxillin marking costameres; yellow = DAPI marking nuclei). Scale bar = 25µm.
**FLNC** heterozygote mutations do not significantly affect sarcomere assembly in hiPSC-CMs.

Having established the effects of **FLNC-null** mutations in hiPSC-CMs, we next asked whether these effects were shared with **FLNC-heterozygote** hiPSC-CMs. Harnessing the endogenous eGFP-**TTN**, we examined both sarcomere structure and contractile function in WT and all **FLNC** mutant hiPSC-CM lines (Fig 2.4). We recorded the contraction of thousands of fluorescent sarcomeres in WT and each **FLNC** mutant hiPSC-CM lines, using video microscopy and assessed sarcomere contractile function using a high-throughput analysis software, SarcTrack (Toepfer, Sharma, et al., 2019b).

Some sarcomeres were still contractile in **FLNC-/--** hiPSC-CMs, despite severe abnormalities in structure, and interestingly, several sarcomeres exhibited a conspicuous ‘bending’ motion during cardiomyocyte relaxation. Analysis of >12,000 **FLNC-/--** sarcomeres showed a reduced percent of sarcomere shortening (a measurement of cardiomyocyte contraction), indicating that **FLNC-/--** sarcomeres are functionally deficient (Fig 2.4b). In contrast to **FLNC-/--** hiPSC-CMs, sarcomere structure and function in **FLNC+/-**, **FLNC+/-Δ7aa**, and **FLNCG1674S/G1674S** hiPSC-CMs were comparable to WT (Fig 2.4a).

Taken together, these results indicated that some FLNC expression was required for sarcomere structure and function; however 50% normal FLNC expression did not significantly impair these processes in hiPSC-CMs in vitro. As such, these experiments inferred that additional mechanism(s) likely contributed to the pathogenesis of **FLNC** heterozygote mutations in DCM.
Figure 2.4. FLNC heterozygote mutations do not significantly affect sarcomere structure or contractile function in hiPSC-CMs.

(A) Widefield images of hiPSC-CMs containing an endogenous N-terminal eGFP tag on the sarcomere protein, TTN, marking sarcomere Z-discs (green). FLNC⁻/⁻ sarcomeres are fragmented, while FLNC⁺/-, FLNC⁺/-, and FLNC⁻/- myofibrils are intact. Scale bar = 25μm.

(B) The contractile function of WT and FLNC mutant sarcomeres, assessed using a high-throughput analysis platform (SarcTrack). FLNC⁻/- sarcomeres show significantly impaired contractile function (p < 0.0001, unpaired t-test, compared to wild-type). FLNC⁺/-, and FLNC⁺/-, and FLNC⁻/- sarcomeres are not significantly affected. Data are displayed as mean ± SEM. The number of differentiations (n), number of videos (v), and number of total tracked sarcomeres (s) are indicated as follows: WT (n=6, v=181, s=47,498); FLNC⁺/- (n=2, v=44, s=12,934); FLNC⁺/- (n=3, v=59, s=13,006); FLNC⁺/- (n=2, v=20, s=3,571); FLNC⁻/- (n=3, v=83, s=12,138).
**FLNC** heterozygote mutations result in increased lysosome protein expression and intracellular accumulation of membrane and secreted proteins

To define the effects of other **FLNC** variants, we compared the proteomes of WT, **FLNC**+/-, **FLNC**+/Δ7aa, and **FLNC**G1674S/G1674S hiPSC-CMs. Comparison of the proteomes of WT, **FLNC**+/-, and **FLNC**+/Δ7aa hiPSC-CMs showed striking differences amongst both **FLNC** heterozygote mutants, with 212 and 568 DE proteins, respectively (Fig 2.5a). More than 50% of DE proteins were shared by both **FLNC** heterozygote mutants. The similarities of these responses implied that both the **FLNC**+/- and **FLNC**+/Δ7aa lines carried a functionally-null allele. The **FLNC**G1674S/G1674S showed no DE proteins, suggesting that this variant is benign (Fig 2.5a). Comparison of the DE proteins in the **FLNC**- heterozygote and **FLNC**-null lines showed distinct patterns of expression, suggesting that partial and total loss of **FLNC** protein had distinct effects on cardiomyocyte biology (Fig S2.4a).

We conducted enrichment analyses of the shared DE proteins between **FLNC**+/- and **FLNC**+/Δ7aa hiPSC-CMs, specifying the 6,863 proteins quantified in these experiments as a background dataset (Fig 2.5b). Amongst the 350 shared DE proteins, there was significant enrichment for lysosome proteins. The levels of lysosomal membrane proteins, proteins that mediate lysosomal acidification, and lysosomal enzymes including proteases, glycosidases, sulfatases, phospholipases, and ceramidases were all increased in **FLNC** heterozygote mutants (Fig 2.5c, Fig S2.4b). In contrast, **FLNC**-/- hiPSC-CMs did not show changes in lysosomal protein expression (Fig S2.4b). Additional enrichment was observed amongst proteins containing signal peptides, disulfide bonds, glycoproteins, and secreted proteins (Fig 2.5b). As this latter
group of proteins are incorporated into cell membranes or secreted, we collectively referred to these as ‘trafficked proteins’. The increase in lysosome protein levels and trafficked protein levels was even higher in \(FLNC^{+\Delta7aa}\) hiPSC-CMs compared to \(FLNC^{+/}-\) hiPSC-CMs (Fig 2.5c-d).

RNAseq data showed that the majority of transcripts encoding lysosomal proteins were not significantly changed (Fig S2.4c), including the expression of \(TFEB\), a known master regulator of lysosomal biogenesis (Fig S2.4d) (Sardiello et al., 2019). Hence, we deduce that the mechanisms that accounted for lysosomal protein elevations were largely post-transcriptional.
Figure 2.5. FLNC heterozygote mutations result in increased levels of lysosome proteins and intracellular accumulation of membrane and secreted proteins in hiPSC-CMs.

(A) Quantitative proteomics was used to determine the expression levels of 6,863 unique proteins in WT, FLNC+/−, FLNC<sub>G1674S/G1674S</sub>, and FLNC+/∆7aa hiPSC-CMs. A total of 684 unique proteins were differentially expressed (DE) amongst all mutants DE criteria: |log<sub>2</sub>(fold-change)| > 0.67 and Benjamini-Hochberg adjusted p-value < 0.1). The FLNC+/−, FLNC<sub>G1674S/G1674S</sub>, and FLNC+/∆7aa mutants had 212, 0, and 568 DE proteins respectively. Each replicate column in heatmap (scaled by row) represents an independent differentiation of each cell line.

(B) 350 DE proteins (160 upregulated, 190 downregulated) overlapped between FLNC+/− and FLNC+/∆7aa hiPSC-CMs (overlap criteria: |log<sub>2</sub>(fold-change)| > 0.515). Enrichment analysis was performed on the 350 overlapping proteins. Proteins with signal peptides, disulfide bonds, secreted proteins, and glycoproteins, are collectively referred to as ‘trafficked’ proteins.

(C-D) Heatmaps of enriched protein groups, scaled by row.
**FLNC heterozygote mutations result in increased lysosome abundance and accumulation of FLNC-binding proteins**

We directly assessed lysosome abundance and dynamics in WT and *FLNC* mutant hiPSC-CMs, using a LysoTracker red fluorescent membrane-permeable dye and live-cell imaging. In comparison to WT, *FLNC*+/∆7aa and *FLNC*+/- cells displayed increased LysoTracker staining (Fig 2.6a), indicating increased lysosome expression in these mutants. There was dynamic motility of lysosomes along linear paths in all cell lines, suggesting that *FLNC* mutations did not severely impact lysosome movement along microtubules (Guardia et al., 2016).

Given the even higher levels of lysosomal proteins in *FLNC*+/∆7aa compared to *FLNC*+/ hiPSC-CMs, we hypothesized that the *FLNC*∆7aa allele encoded a stable dysfunctional protein that might produce FLNC aggregates. We tested this hypothesis using FLNC immunofluorescence (Fig 2.6b). In WT and *FLNC*+/- hiPSC-CMs, FLNC was labeled only at sarcomere Z-discs, while *FLNC*+/∆7aa hiPSC-CMs contained FLNC-staining both in aggregates and in Z-discs. Based on these observations, we suggest that the presence of FLNC aggregates contributed to the increased severity of phenotype in *FLNC*+/∆7aa hiPSC-CMs, as evidenced by the greater degree of lysosomal and trafficked protein accumulation.

As FLNC is a large dimeric protein that has previously been shown to interact with a large number of proteins, we considered if heterozygous *FLNC* variants altered the stoichiometry of these interactions; this could lead to an accumulation of protein complexes that could underlie the increase in lysosome abundance in *FLNC* heterozygote hiPSC-CMs.
We therefore examined the protein levels of FLNC-binding partners and Z-disc proteins in $FLNC^{+/\Delta 7aa}$ and $FLNC^{+/-}$ hiPSC-CMs, using $FLNC^{G1674S/G1674S}$ hiPSC-CMs – which did not show increases in lysosome expression – as a negative control. The intracellular protein levels of FLNC-interacting proteins and selective Z-disc proteins were significantly increased in the heterozygote FLNC hiPSC-CMs, including $\beta_1$-integrin (ITGB1), synaptopodin (SYNPO), and myoferlin (MYOF) (Fig 2.6c). The levels of other sarcomere proteins were not significantly affected in FLNC heterozygote mutants including thick filament components (TTN, MYH6, MYH7, MYBPC3) and most thin filament components (TNNC1, TNNT2, TPM2), but there were increased protein levels of TNNI3. The degree of protein accumulation of FLNC-interacting and Z-disc proteins was greater in $FLNC^{+/\Delta 7aa}$ than in $FLNC^{+/-}$ hiPSC-CMs (Fig 2.6c). We deduced that mutant FLNC$^{\Delta 7aa}$ protein as well as FLNC-interacting proteins aggregated in $FLNC^{+/\Delta 7aa}$ hiPSC-CMs.

Cardiomyocytes dispose of accumulating, damaged, misfolded proteins through several mechanisms, including macroautophagy and chaperone-assisted selective autophagy (CASA)(Arndt et al., 2010). Macroautophagy involves the formation of a double-membrane sequestering compartment termed the phagophore, which matures into an autophagosome. The autophagosome delivers material to the lysosome for degradation and recycling (Pohl and Dikic, 2019). Several protein kinases regulate macroautophagy including mTOR Complex 1 and protein kinase A (Stephan et al., 2009). Muscle cells have been shown to have additional forms of autophagy that are selective for the degradation of particular proteins. Chaperone-mediated selective autophagy (CASA) is an autophagic pathway in muscle cells that facilitates the
degradation of damaged Z-disc components, including FLNC (Arndt et al., 2010). Upon contraction-induced damage, proteins are recognized by a Z-disc localized chaperone machinery comprising the small heat shock protein HspB8, BAG3 and Hsc70, and subsequently ubiquitinated and engulfed by autophagosomes for lysosomal degradation (Arndt et al., 2010). Given the intracellular accumulation of FLNC-interacting proteins, Z-disc proteins, and trafficked proteins, we predicted that the expression of proteins involved in macroautophagy and/or CASA might be abnormal in heterozygous FLNC hiPSC-CMs.

We tested this hypothesis by examining the expression of 28 proteins involved in various stages of macroautophagy, regulators of macroautophagy, and CASA in FLNC heterozygote lines, using the FLNC<sub>G1674S/G1674S</sub> line, which showed no protein accumulation as a negative control mutant line (Arndt et al., 2010). A significant number of proteins involved in macroautophagy had decreased levels in both FLNC<sup>+</sup>/ and FLNC<sup>+</sup>/∆7aa hiPSC-CMs, but not in FLNC<sub>G1674S/G1674S</sub> hiPSC-CMs (Fig 2.6d). Proteins with decreased levels in both FLNC heterozygote mutants included those involved in early phagophore formation (BECN1, ATG13), proteins involved in autophagosome formation (ATG2B, MAP1LC3A, ATG7, ATG12, ATG5, ATG16L1) and regulators of autophagy including PRKACA and PRKACB (Stephan et al., 2009; Dikic and Elezar, 2018; Pohl and Dikic, 2019). Overall, levels of these proteins were even lower in FLNC<sup>+</sup>/∆7aa hiPSC-CMs than in FLNC<sup>+</sup>/ hiPSC-CMs, suggesting that the severity of changes observed in macroautophagy-related proteins was directly associated with the degree of cardiomyocyte protein accumulation.
Figure 2.6. *FLNC* heterozygote mutations result in increased lysosome content and accumulation of FLNC-binding partners and Z-disc proteins in hiPSC-CMs.

(A) LysoTracker labelling in hiPSC-CMs (LysoTracker = red, nuclei = blue). *FLNC*+/− and *FLNC*+/Δ7aa mutant hiPSC-CMs have increased lysosome content; lysosomes frequently cluster in the perinuclear region. Two representative fields are shown per cell line (scale bar = 100 μm).

(B) FLNC immunofluorescence in hiPSC-CMs. FLNC is predominantly localized to Z-discs in WT and *FLNC*+/− hiPSC-CMs. *FLNC*+/Δ7aa hiPSC-CMs have FLNC aggregates (arrows), in addition to some Z-disc staining (asterisks). (scale bar = 25 μm)

(C) Quantitative proteomics was used to determine the levels of twenty-eight proteins - either known FLNC binding partners, Z-disc proteins, actin, or other thin/thick filament proteins – amongst WT and *FLNC* mutant hiPSC-CMs. Differentially expressed (DE) proteins are indicated in red font. (DE criteria: |log2fold-change| ≥ 0.4 and p-value < 0.05). The number of DE proteins in the *FLNC*+/− mutant (8) and the *FLNC*+/Δ7aa mutant (13) was significantly greater than expected by chance (binomial test p = 4.9E-5, p=2.2E-10 respectively), while the *FLNG1674S/G1674S* mutant (1) was not significantly affected. DE proteins are predominantly FLNC-binding partners and Z-disc proteins.

(D) Quantitative proteomics was used to determine the levels of twenty-eight proteins involved in macroautophagy (phagophore formation, autophagosome elongation, autophagosome maturation, regulation of autophagy) and chaperone-assisted-selective-autophagy (CASA) amongst WT and *FLNC* mutant hiPSC-CMs. DE proteins are indicated in red font. (DE criteria: |log2fold-change| ≥ 0.3 and p-value < 0.05) The number of DE proteins in the *FLNC*+/− mutant (12) and the *FLNC*+/Δ7aa mutant (12) was significantly greater than expected by chance (binomial test p = 3.5E-9), while the *FLNG1674S/G1674S* mutant (0) was not significantly affected.
Discussion

From our studies of a series of FLNC variants, we uncovered the importance of FLNC in human cardiomyocyte biology. We found that some FLNC expression was necessary for the coordinated transcription and translation of actin-binding and sarcomere thin-filament genes. Disruption of this gene network and deficient actin-crosslinking at Z-discs resulted in severe impairment in sarcomere assembly in FLNC-null hiPSC-CMs. However, FLNC heterozygote mutations did not severely impair sarcomere assembly in hiPSC-CMs, indicating that additional mechanisms accounted for the pathogenesis of FLNC heterozygote variants. Indeed, FLNC heterozygote variants resulted in accumulation of FLNC-binding partners and Z-disc proteins, and increased cellular lysosome content, presumably a response to turnover accumulating proteins. FLNC heterozygote truncating variants identified in human DCM patients can (1) result in stable truncated FLNC proteins that themselves aggregate or (2) reduce FLNC levels to half-normal. Importantly, our data indicated that both of these consequences disrupted the stoichiometry of FLNC protein-interactions, and profoundly disturbed cardiomyocyte proteostasis (Fig 2.7). These findings have key implications for therapeutic design.

Previous genetic studies have outlined a critical role for FLNC in maintaining muscle structure and function. Biallelic deletion of Flnc in mice results in neonatal lethality due to respiratory muscle insufficiency, and causes loss of Z-disc structure in skeletal muscle fibers (Dalkilic et al., 2006). Knockout of the FLNC ortholog cheerio in Drosophila results in widened, fractured, and indistinct Z-discs in indirect flight muscle (González-Morales, Holenka and Schöck, 2017). Deletion of Flnc in medaka results in
progressive skeletal muscle degeneration, as well as cardiac enlargement and rupture, associated with focal Z-disc destruction (Fujita et al., 2012). Common to all these studies are structural Z-disc abnormalities, and poor tissue integrity, an observation that underscores Z-disc functions in sarcomere assembly and the maintenance of muscle structure under biomechanical forces. Here, show that FLNC is necessary to coordinate the transcription and translation of a network of sarcomere and actin-associated genes including SYNPO, TPM2, TNNI3, TNNT1, SYNPO2L, ANKRD1, PDLIM3, LMOD2, and CSRP3, and for actin-crosslinking at Z-discs. Deficiencies of these processes in FLNC-null hiPSC-CMs resulted in impaired sarcomere assembly and compromised Z-disc integrity, concordant with phenotypes observed in other model systems.

Given the severe defects in sarcomere assembly in FLNC-null hiPSC-CMs, we initially predicted that partial loss of FLNC, in the context of FLNC heterozygote variants, would also impair sarcomere formation and function in hiPSC-CMs. Unexpectedly, we discovered that partial loss of FLNC expression had additional, largely distinct consequences. In contrast to FLNC-null hiPSC-CMs, FLNC heterozygote hiPSC-CMs did not have significant deficits in sarcomere structure or function in vitro. As hiPSC-CM models do not recapitulate in vivo hemodynamics, we cannot exclude the potential for half-normal FLNC protein levels to result in defective sarcomere assembly under conditions of increased biomechanical load, thereby contributing to human DCM pathophysiology. But in addition, we show that FLNC heterozygote hiPSC-CMs resulted in severe protein accumulation and increased cellular lysosome content. We suggest that toxic protein accumulation is the predominant cause of human DCM pathogenesis.
The FLNC+/− and FLNC+/Δ7aa variants evoked similar but graded effects, with more profound dysregulation of protein expression in FLNC+/Δ7aa hiPSC-CMs. This was explained by the finding that the FLNCΔ7aa protein was stably expressed and aggregated in hiPSC-CMs, while the FLNC+/− mutation resulted in half-normal FLNC expression. However, both types of mutations resulted in the accumulation of FLNC-binding partners and Z-disc proteins, and increased lysosome expression. The magnitude of these protein accumulations would be considerable as FLNC is a highly abundant protein (in the top 3% of proteins expressed in hiPSC-CMs).

Heterozygous truncating variants in FLNC also cause myofibrillar myopathy (MFM), a progressive skeletal muscle disease resulting disintegration of muscle fibers that include characteristic giant protein aggregates (Kley et al., 2007). The presence of protein aggregates in FLNC MFM suggests shared mechanisms with FLNC DCM. Furthermore, FLNC MFM patients often develop DCM later in their disease course (Kley et al., 2007). Why some FLNC truncating variants result in predominantly skeletal versus cardiac disease remains unknown. We speculate that FLNC truncations that produce more severe protein aggregation affect skeletal (and cardiac) muscle, but those that produce less severe aggregation affect cardiac muscle alone. This may explain why large protein aggregates have not yet been observed in cardiac biopsies of patients with isolated DCM (Begay et al., 2018).

In addition to FLNC variants, damaging variants in other Z-disc-associated genes also cause both MFM and DCM including desmin, αβ-cystallin, myotilin, Z-band alternatively spliced PDZ motif-containing protein (ZASP), bcl-2-associated anthanogene 3 (BAG3), four and a half LIM domain 1 (FHL1), titin, and actin (Ruparelia
et al., 2016). Damaging variants in some of these genes also trigger protein accumulation, suggesting that the efficient turnover of proteins at Z-discs is crucial for myocyte function. Protein turnover, mediated by lysosomes, is particularly important, as demonstrated by damaging variants in LAMP-2 (Chr X), which depletes the encoded lysosomal-associated membrane protein-2, and causes multisystem disease, including a profound cardiomyopathy in affected males (Samad et al., 2017). LAMP2 deficiency disrupts autophagosome-lysosome fusion in cardiomyocytes, thus decreasing the efficiency of normal autophagic pathways to turnover accumulating proteins (Chi et al., 2019). Furthermore, electrostimulation of soleus muscles isolated from lamp2−/− mice results in the disintegration of Z-discs (Arndt et al., 2010), providing further evidence that lysosome-mediated degradation is essential for the maintenance and turnover of Z-disc components in contractile myocytes. As the levels of proteins involved in autophagosome formation and maturation were decreased in FLNC heterozygote mutant hiPSC-CMs, we suggest that autophagy is a key component of FLNC degradative pathways. Saturation of normal autophagic pathways and resultant ER/Golgi stress in FLNC heterozygote hiPSC-CMs may account for the observed intracellular accumulation of a wide range of proteins that are normally trafficked to the sarcolemma or secreted from the cell.

Further studies are needed to identify the precise effects of Z-disc protein accumulation on proteasome-mediated degradation and autophagic flux. Chaperone proteins may be particularly important for the efficient turnover of Z-disc components. The BAG domain-containing cochaperone (BAG3) has been shown to concentrate at Z-discs and cooperate with the phagophore-interacting ubiquitin adaptor p62 to dispose of
ubiquitinated FLNC via chaperone-mediated selective autophagy (Arndt et al., 2010). Notably, damaging variants in BAG3 also cause DCM and MFM (Knezevic et al., 2015). Our studies showed that FLNC heterozygote hiPSC-CMs also had depleted levels of two additional heat shock proteins – HSPB1 and HSPB7 – previously shown to interact with FLNC. The absence of HSPB7 disrupts FLNC localization and results in FLNC aggregation in myofibers (Juo et al., 2016). Phosphorylation of HSPB1 results in exposure of its FLNC binding site and prevents FLNC from further unfolding (Collier et al., 2019). Thus, the regulation of FLNC folding and turnover is related to the function of at least three chaperone proteins including BAG3, HSPB1, and HSPB7. Chaperones may be broadly important for Z-disc protein folding, as αβ-crystallin (CRyAB)-mediated folding of desmin, another Z-disc component, is crucial for muscle function (Mclendon and Robbins, 2019). We propose that proteostasis regulators and replenishment of chaperone proteins may be therapeutic for patients with FLNC DCM and other Z-disc myopathies. These approaches have been applied in other organs affected by genetic disorders of protein aggregation (Ugarte, A and Pérez, 2018). In addition, interventions to increase autophagic flux, including lifestyle modifications, such as endurance exercise and caloric restriction, may have clinical value (Galluzzi et al., 2017). We recognize several limitations in our investigations. Human iPSC-CMs are immature and cannot entirely recapitulate features of damaging FLNC variants in adult cardiomyocytes or appropriately model in vivo hemodynamics (Zhu et al., 2014). Furthermore, we only studied a few damaging FLNC variants, focusing on the effects of truncating alleles. Additional classes of FLNC variants, including heterozygous missense variants, may have additional/distinct effects that account for hypertrophic
cardiomyopathy and restrictive cardiac phenotypes (Valdés-Mas et al., 2014; Brodehl et al., 2016; Gómez et al., 2017; Tucker et al., 2017; Kiselev et al., 2018). Despite these limitations, our studies define one mechanism by which FLNC variants cause DCM. We expect further analyses of additional classes of variants will yield further mechanistic insights.

In summary, our studies demonstrate that FLNC truncating variants cause DCM by reducing FLNC protein levels, resulting in toxic protein accumulation and increased lysosome expression. These studies add to a growing recognition that lysosome-dependent turnover of excess, misfolded and/or damaged proteins is critical for maintaining cardiomyocyte function. These preclinical studies provide key mechanisms that can enable translational studies. We suggest that strategies to increase chaperone expression and/or autophagic flux may ameliorate FLNC disease phenotypes, and more broadly, benefit patients with cardiomyopathies that are associated with pathologic protein accumulations. Genetically-engineered hiPSC-CMs with pathogenic variants and isogenic control lines provide important models to test the efficacy of these strategies. As pathologic protein aggregation occurs in several forms of DCM and MFM, a better understanding of associated mechanisms and development of appropriate interventions has the potential to benefit a large subset of patients with Z-disc-associated myopathies.
Figure 2.1. Model of the effects of FLNC mutations in human cardiomyocytes.

(A) Complete loss of FLNC expression in cardiomyocytes results in diminished expression of actin-binding and sarcomere thin-filament related genes including SYNPO, TPM2, TNNI3, TNNI3, SYNPO2L, ANKRD1, PDLIM3, LMOD2, and CSRP3. Additionally, FLNC-mediated actin-crosslinking at Z-discs is lost. Both these effects result in deficient sarcomere assembly in cardiomyocytes.

(B) Heterozygote loss of FLNC expression disrupts the proper stoichiometry of FLNC-mediated protein-protein interactions, resulting in the accumulation of FLNC-binding partners and Z-disc components. To degrade accumulating proteins, lysosome expression is increased. Over time, autophagic degradative processes are compromised, resulting in further protein accumulation in the ER/Golgi apparatus. This impairs protein trafficking to the sarcolemma and secretion of proteins from the cell. Eventually, saturation of normal protein handling pathways results in cardiomyocyte dysfunction.
**Methods**

**Culture and maintenance of hiPSCs.** PGP1 hiPSCs were engineered to contain an endogenous eGFP tag on the sarcomere gene *TTN*, as previously described (Sharma, Toepfer, Ward, *et al.*, 2018). 6-well tissue culture dishes (Fisher Scientific, 07-200-83) were coated with BD Matrigel (BD Biosciences 354277) at a 1:320 dilution in DMEM F12 (Invitrogen 11330057). iPSCs were seeded on Matrigel-coated 6-well tissue-culture dishes at approximately 30% confluency and maintained in stem cell maintenance media (mTESR1, STEMCELL Technologies). iPSCs were passaged every 2-3 days in the presence of 5μM ROCK inhibitor Y-27632 (R&D Systems, 125410). mTESR media was replenished daily. Cells were cultured under sterile conditions, and maintained at 37°C, 5% CO₂.

**CRISPR/Cas9-mediated gene editing of hiPSCs** The Benchling® design tool was used for guideRNA and HDR design. The guideRNA (ATGCCAAGGCAGCCGGTGAG) targeted FLNC exon 30. To create the G1674S mutation, an 127bp HDR arm was designed

(GAGACGGTGATCACCGGTGATGCCAAGGCAGCCAGTGAGGGAAAGGTGACATGCACGGTGTTCCACGGGATGGGGGAGAGCTCGATGTGGATGTGGTTGAGAACCAGACGGGTACCTTTGACATCT) and synthesized as an Ultramer from IDT. A 455bp fragment bearing all components necessary for gRNA expression (U6 promoter + target sequence + guide RNA scaffold + termination signal) was synthesized as a gBlock from IDT and cloned into the TOPO plasmid using the Zero Blunt Topo II PCR cloning kit (Invitrogen K2800-02). Adherent iPSCs were dissociated into single cells from the tissue
culture plate using accutase. One million cells were centrifuged for 5 minutes at 1000rpm. For nucleofection of CRISPR/Cas9 components, cells were resuspended in 82μL of human stem cell nucleofector solution and 18μL of supplement 1 (Lonza). A combination of 2μg of Cas9 plasmid (pSpCas9(BB)-2A-Puro (PX459) V2.0 - Addgene Plasmid #62988), 4μg of guideRNA TOPO plasmid, and 4μg of HDR arm was added, and cells were nucleofected using program B-016. 500μL of mTESR media containing ROCK inhibitor were added to the nucleofection cuvette and transferred dropwise at varying dilutions onto a matrigel coated 6-well plate using a sterile plastic pipette. 48 hours post-nucleofection, mTESR media containing puromycin at a concentration of 0.5μg/mL was added. Puromycin selection was continued for 2 days. Six days post-nucleofection, single colonies were picked under a dissection microscope using sterile pipette tips into a matrigel-coated 96-well plate. Genomic DNA was extracted using PrepGEM (VWR PUN0500) and colonies were screened for the presence of gene editing by PCR amplification with the following primers (CCTGGCCTCACACTCTTCTCTC, AAGTGGGATTCCTTGGGGACAG) followed by Sanger sequencing with the following primer (GGCTACAGAACTTACCAGCACG). Selected colonies with evidence of gene editing were subcloned to verify clonality of the edited population. Clonality of population was further verified by next generation sequencing analysis (MiSeq).

**Differentiation of hiPSC-CMs.** hiPSCs were differentiated into cardiomyocytes in monolayers using a previously described chemically-defined protocol with minimal modification (Lian et al., 2013). hiPSCs seeded in Matrigel™-coated 6-well culture plates were grown to 90% confluency prior to initiating the hiPSC-CM differentiation
protocol. To initiate differentiation, mTESR media was removed and cells were washed 2X with sterile PBS. B27 without insulin (Thermo Fisher Scientific A1895601) in RPMI media (Invitrogen 11875-119) supplemented with 18μM CHIR-99021 (Tocris 44-231-0) was added to hiPSCs for 2 days (day 0-1). The media was changed to RPMI with B27 without insulin for one day (day 2). Cultures were then treated with RPMI with B27 without insulin for two days supplemented with 2μM Wnt-C59 (Biorbyt orb181132), and then incubated for two days (day 3-5). Cells were maintained in RPMI with B27 without insulin for 2 additional days (day 6-7). On day 7, media was changed to RPMI with B27 containing insulin (Invitrogen 17505-055) and the media was replenished every other day until day 11. Cells underwent a 6-day glucose-deprivation process to enrich for cardiomyocytes. Cells were cultured in RPMI media devoid of glucose (Life Technologies 11879-020) containing B27 with insulin, which was replaced every other day until day 17. On day 17, cells were again cultured in full growth media (RPMI + B27 containing insulin) with media replacement every other day until day 30.

**Replating hiPSC-CMs for imaging.** For all imaging studies, hiPSC-CMs were replated onto Matrigel-coated glass-bottom imaging-optimized 12-well plates (Mat-Tek). hiPSC-CMs were cultured in RPMI B27 plus insulin containing 5μM ROCK inhibitor and 20% FBS for 2 days post-replating, after which they were cultured in RPMI B27 plus insulin for at least 1 week prior to imaging analysis.

**Contractility analysis of hiPSC-CMs.** Replated hiPSC-CMs were imaged between day 30 and day 40 of hiPSC-CM differentiation. Five-second video imaging of live cells (typically clusters of 2-4 cells each) was performed in a heated and humidified chamber (37C, 5% CO₂) using a 100X objective of a fluorescence microscope (Keyence BZ X-
with an acquisition rate of 30 frames per second. Analysis of sarcomere contraction was performed with SarcTrack software, as previously described (Toepfer, Sharma, et al., 2019b)

**Harvest of Cell Lysate for Western Blotting and Mass Spectrometry** iPSC-CMs were harvested on day 30 of cardiomyocyte differentiation. Cells were washed quickly 1X in cold PBS. Adherent cells were scraped from all wells of a 6-well tissue culture dish, pelleted by spinning at 1000RPM for 5 minutes, and lysed with cold RIPA buffer in the presence of protease inhibitors (Sigma-Aldrich 11873580001) and phosphatase inhibitors (Sigma-Aldrich 4906837001). To increase protein yield and ensure lysis of nuclei, the lysate was subjected to additional mechanical disruption by passage 20 times through a 21G needle. The mixture was shaken gently on ice for 15 minutes, and then centrifuged at 14,000g for 15 minutes to pellet cell debris. The supernatant was transferred to a new tube and protein concentration was quantified using the BCA assay kit (Pierce 23227).

**Western Blotting** For each sample, a total of 12.5μg of protein lysate was loaded and resolved under denaturing conditions on a NuPAGE 3-8% Tris-Acetate Protein Gel. Lysate was transferred to a low-fluorescence PVDF membrane overnight at 40mA at 4°C. Membrane was blocked for 1 hour at room temperature in 5% skim milk in TBST buffer, before addition of primary antibodies. Primary antibodies included an N-terminal FLNC antibody (1:500, mybiosource, Rb pAb, MBS2026155), a C-terminal FLNC antibody (1:500, abcam, Rb mAb, ab180941), TNNT2 (1:500, abcam, Ms mAb, ab8295), TUBB (1:500, abcam, Rb pAb, ab6046). Primary antibodies were diluted in 5% skim milk and incubated at 4°C overnight. Secondary antibodies for infrared fluorescent
western blotting included Goat anti-Rabbit IgG (IRDye®680RD) and Goat anti-Mouse IgG (IRDye®800CW). Blots were imaged on a LI-COR Biosciences Odyssey CLx imaging system.

**Tandem Mass Tag (TMT) Proteomics of hiPSC-CMs** Samples were prepared for tandem mass tag quantitative proteome profiling using a previously published protocol (Navarrete-perea et al., 2018). **Protein digestion** For each sample, 100μg of total protein was subjected to disulfide bond reduction with 5 mM tris (2-carboxyethyl) phosphine (room temperature, 30 min) and alkylation with 10 mM iodoacetamide (room temperature, 30 min in the dark). Excess iodoacetamide was quenched with 10 mM dithiotreitol (room temperature, 15 min in the dark). Methanol-chloroform precipitation was performed prior to protease digestion. Precipitated proteins were resuspended in 200 mM EPPS, pH 8.5 and digested at room temperature for 13 h with Lys-C protease at a 100:1 protein-to-protease ratio. Trypsin was then added at a 100:1 protein-to-protease ratio and the reaction was incubated for 6 h at 37°C. **Tandem Mass Tag Labeling** TMT reagents (0.8 mg) were dissolved in anhydrous acetonitrile (40 μL) of which 10 μL was added to the peptides (100 μg) with 30 μL of acetonitrile to achieve a final acetonitrile concentration of approximately 30% (v/v). Following incubation at room temperature for 1 h, the reaction was quenched with hydroxylamine to a final concentration of 0.3% (v/v). The TMT-labeled samples were pooled at in equal amounts across all samples. The pooled sample was vacuum centrifuged to near dryness and subjected to C18 solid-phase extraction (SPE) (Sep-Pak, Waters). **Off-line basic pH reversed-phase (BPRP) fractionation** The pooled TMT-labeled peptide sample was fractionated using BPRP HPLC. An Agilent 1200 pump was used equipped with a
degasser and a photodiode array (PDA) detector (set at 220 and 280 nm wavelength) from ThermoFisher Scientific (Waltham, MA). Peptides were subjected to a 50-min linear gradient from 5% to 35% acetonitrile in 10 mM ammonium bicarbonate pH 8 at a flow rate of 0.6 mL/min over an Agilent 300Extend C18 column (3.5 μm particles, 4.6 mm ID and 220 mm in length). The peptide mixture was fractionated into a total of 96 fractions, which were consolidated into 24, from which 12 non-adjacent samples were analyzed. Samples were subsequently acidified with 1% formic acid and vacuum centrifuged to near dryness. Each consolidated fraction was desalted via StageTip, dried again via vacuum centrifugation, and reconstituted in 5% acetonitrile, 5% formic acid for LC-MS/MS processing. **LC-MS/MS analysis** All samples were analyzed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific). Peptides were separated on a 100 μm inner diameter microcapillary column packed with 35 cm of Accucore C18 resin (2.6 μm, 150 Å, ThermoFisher). For each analysis, we loaded approximately 2 μg onto the column. Peptides were separated using a 150min gradient of 3 to 25% acetonitrile in 0.125% formic acid with a flow rate of 450 nL/min. Each analysis used an MS3-based TMT method, which has been shown to reduce ion interference compared to MS2 quantification. Prior to starting our analysis, we perform two injections of trifluoroethanol (TFE) to elute any peptides that may be bound to the analytical column from prior injections to limit carry over. The scan sequence began with an MS1 spectrum (Orbitrap analysis, resolution 120,000, 350–1400 Th, automatic gain control (AGC) target 5E5, maximum injection time 100 ms). The top ten precursors were then selected for MS2/MS3 analysis. MS2 analysis
consisted of: collision-induced dissociation (CID), quadrupole ion trap analysis, automatic gain control (AGC) 2E4, NCE (normalized collision energy) 35, q-value 0.25, maximum injection time 120 ms), and isolation window at 0.7. Following acquisition of each MS2 spectrum, we collected an MS3 spectrum in which multiple MS2 fragment ions are captured in the MS3 precursor population using isolation waveforms with multiple frequency notches. MS3 precursors were fragmented by HCD and analyzed using the Orbitrap (NCE 65, AGC 1.5E5, maximum injection time 150 ms, resolution was 50,000 at 400 Th). For MS3 analysis, we used charge state-dependent isolation windows: For charge state z=2, the isolation window was set at 1.3 Th, for z=3 at 1 Th, for z=4 at 0.8 Th, and for z=5 at 0.7 Th. **Data analysis** For iBAQ estimation of absolute protein abundance, mass spectra were processed using MaxQuant (Version 1.6.10.43). Otherwise, mass spectra were processed using a Sequest-based pipeline. Spectra were converted to mzXML using a modified version of ReAdW.exe. Database searching included all entries from the human UniProt database. This database was concatenated with one composed of all protein sequences in the reversed order. Searches were performed using a 50 ppm precursor ion tolerance for total protein level analysis. The product ion tolerance was set to 0.9 Da. TMT tags on lysine residues and peptide N termini (+229.163 Da) and carbamidomethylation of cysteine residues (+57.021 Da) were set as static modifications, while oxidation of methionine residues (+15.995 Da) was set as a variable modification. Peptide-spectrum matches (PSMs) were adjusted to a 1% false discovery rate (FDR). PSM filtering was performed using a linear discriminant analysis (LDA), as described previously, while considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor
mass accuracy. For TMT-based reporter ion quantitation, we extracted the summed signal-to-noise (S:N) ratio for each TMT channel and found the closest matching centroid to the expected mass of the TMT reporter ion. For protein-level comparisons, PSMs were identified, quantified, and collapsed to a 1% peptide false discovery rate (FDR) and then collapsed further to a final protein-level FDR of 1%, which resulted in a final peptide level FDR of <0.1%. Moreover, protein assembly was guided by principles of parsimony to produce the smallest set of proteins necessary to account for all observed peptides. Proteins were quantified by summing reporter ion counts across all matching PSMs, as described previously. PSMs with poor quality, MS3 spectra with more than eight TMT reporter ion channels missing, MS3 spectra with TMT reporter summed signal-to-noise of less than 100, or having no MS3 spectra were excluded from quantification. Each reporter ion channel was summed across all quantified proteins and normalized assuming equal protein loading of all samples. Protein quantification values were exported for further analysis in R (histograms, heatmaps, scatterplots) and bargraphs (GraphPad Prism).

**Live imaging and staining of lysosomes in hiPSC-CMs.** LysoTracker® Red DND-99 (Thermo Fisher Scientific L7528), a red fluorescent membrane-permeable dye for labeling and tracking acidic organelles in live cells, was used to visualize lysosomes in hiPSC-CMs. hiPSC-CMs were replated on glass-bottom 12-well dishes (MatTek). One week later, pre-warmed (37°C) RPMI B27 plus insulin media containing LysoTracker® probe (75nM) was added to the cells and incubated for 1 hour, after which media was removed and fresh RPMI B27 plus insulin was added. Videos were acquired at an acquisition rate of 30 frames per second using a 100X objective of a widefield
fluorescence microscope (Keyence BZ X-710) while maintained in heated and humidified incubator chamber (37°C, 5% CO₂). Lysosomes were imaged using a 100X objective of a spinning disk confocal microscope. Image data was processed using Fiji software.

**RNA sequencing of hiPSC-CMs**

Cells were lysed in Trizol Reagent (Life Technologies) and RNA was extracted by conventional methods. RNA quality (RIN) and quantity were assessed on the TapeStation 2200 (Agilent). Two rounds of mRNA purification were performed on total RNA (1 ug) using Dynabeads mRNA DIRECT Kit (Invitrogen). The Superscript III First-Strand Synthesis System (Invitrogen) was used to generate double-stranded cDNA. cDNA libraries were constructed using the Nextera XT DNA Library Preparation Kit (Illumina). Libraries were sequenced on the Illumina NextSeq500 platform. 75 bp paired-end reads were aligned to the human reference genome hg38 using Spliced Transcripts Alignment to a Reference (STAR). Raw reads were normalized to the total number of reads per kilobase of transcript per million (RPKM).

**Sarcomere and actin cytoskeletal imaging of iPSC-CMs**

Replated hiPSC-CMs were cultured to between day 30 and day 40 of cardiomyocyte differentiation and subsequently fixed with 4% PFA, and permeabilized with 0.2% Triton-100. An F-actin probe (molecular probes ActinRed 555, Thermo Fisher R37112) was applied to cells (2 drops in 1mL PBS). Cardiomyocytes were washed once with PBS before imaging using an 100X objective (Keyence BZ X-710).

**Patterning of hiPSC-CMs and quantification of sarcomere striation**

iPSC-CMs at day 47 of differentiation were replated onto 2000 μm² fibronectin coated rectangular
micropatterns on polydimethylsiloxane. The patterned iPSC-CMs were fixed in 4% PFA, permeabilized using 0.2% Triton, and stained to visualize nuclei (NucBlue) and costameres using a paxillin antibody (BD Biosciences 610569, 1:100 dilution) on cells 2 and 8 days after patterning. Images were acquired of costameres (paxillin), sarcomere Z-discs (eGFP-TTN), and nuclei (NucBlue). eGFP-TTN images were then processed in MATLAB as previously described (Hinson et al., 2015; Chopra et al., 2018). In brief, the 2D Fast-Fourier transform (FFT) of each image was taken, the sum of the intensity as a function of distance from the center was calculated to convert the 2D transforms into 1D spectra, and the 1D spectra were normalized. The 1D spectra were then fit with a function containing an aperiodic and periodic component, representing irregular Z-disc structures and regularly spaced Z-disc structures, respectively. The area of the first peak in the normalized 1D FFT was summed to represent the regularity of sarcomere striations. This normalized sum was divided by the total area of the 1D FFT, with the area of the harmonic peak subtracted, to quantify the fraction of the FFT spectra that represents regularly striated sarcomeres.

**Flow cytometry of eGFP-TTN expression in hiPSC-CMs** hiPSC-CMs were dissociated from the culture plate using TrypLE Select (10X), spun at 1,000RPM for 5 minutes, resuspended in phenol-free RPMI B27 plus media, and strained through a 100µm cell strainer prior to data collection with a FACSCanto flow cytometer. hiPSC-CMs without an eGFP-TTN tag were analyzed and compared with eGFP-TTN positive hiPSC-CMs to set appropriate gates for the identification of GFP-positive cells. For GFP-positive cells, the GFP (FITC-A) intensity per cell was recorded and resultant frequency distributions were plotted using FlowJo software.
Contributions
Radhika Agarwal, Christine Seidman, and Jonathan Seidman designed the study. R.A. drafted the manuscript and performed all experiments and computational analyses unless otherwise noted. Joao Paulo and Steve Gygi provided resources and technical expertise for proteomic and phosphoproteomic studies. Joshua Gorham and Steve DePalma assisted with preparation of MiSeq and RNA libraries and sequencing. Jourdan Ewoldt, Anant Chopra, and Christopher Chen, assisted with patterning of cardiomyocytes and quantification of sarcomere striation. Christopher Toepfer assisted with methods and analysis of sarcomere contractility.
Chapter 3:

Deficiency of $\alpha$-kinase 3 (ALPK3), a pseudokinase, results in impaired nuclear and sarcomeric myomesin turnover

This chapter represents a manuscript being prepared for submission:

Deficiency of the $\alpha$-kinase (ALPK3), a pseudokinase, results in impaired nuclear and sarcomeric myosin turnover. Radhika Agarwal, Hiroko Wakimoto, Joao A. Paulo, Christopher Toepfer, Qi Zhang, Arun Sharma, Angela Tai, Mingyue Lun, Joshua Gorham, Steven R. DePalma, Steven P. Gygi, J. G. Seidman and Christine E. Seidman
Abstract

ALPK3 (α-kinase-3) encodes a protein with unknown functions that is developmentally expressed in striated myocytes and harbors recessive damaging variants that cause a lethal neonatal cardiomyopathy, with ventricular dilatation and systolic dysfunction. We investigated the endogenous role of ALPK3 and pathogenic mechanisms of ALPK3 mutations using isogenic human induced pluripotent stem cell derived cardiomyocytes (hiPSC-CMs) and mouse models. In both experiment systems, homozygous ALPK3 null alleles (ALPK3−/−) increased the levels of myomesin (MYOM1 and MYOM2) and additional thick filament proteins. hiPSC-CMs with missense variants that conformationally-altered the ALPK3 α-kinase domain showed similar effects, though surprisingly, did not detectably alter phosphorylation across the cardiomyocyte proteome, thus establishing ALPK3 as a pseudokinase. We found that ALPK3 co-localized with MYOM1 and MYOM2 at the sarcomere M-band and nuclear envelope respectively, and that myomesin assembly into these structures was deficient in ALPK3−/− hiPSC-CMs. We propose a model in which ALPK3 serves as a scaffold for myomesin incorporation and thick filament protein turnover, the dysfunction of which results in cardiomyopathy. Importantly, the ALPK3 α-kinase fold is crucial for these activities, independent of catalysis.
Introduction

Primary cardiomyopathies are primary disorders of the heart muscle that occur in the absence of coronary artery disease, hypertension, valvular, or congenital heart disease (Elliott et al., 2008). Dilated cardiomyopathy (DCM) is characterized by ventricular dilatation and diminished contractile function. DCM occurs with an estimated population prevalence of 1:250 to 1:3000 adults, with variable and usually gradual disease progression that may result in heart failure many years after diagnosis (Merlo et al., 2011; Lee et al., 2017). Dominant damaging variants are identified in ~40% of adults with DCM (Hershberger, Hedges and Morales, 2013). These alter more than 30 genes that encode protein components of the cardiomyocyte contractile apparatus (the sarcomere), linker proteins that connect sarcomeres to the cell membrane, and proteins of the nuclear envelope, among others (Dellefave and McNally, 2010). Pediatric-onset DCM occurs less commonly (estimated incidence of 1.1 to 1.5 per 100,000 children) and when diagnosed in infancy, has a devastating prognosis: 40% death or transplantation within two years (Lee et al., 2017). Damaging variants in adult DCM genes can also cause DCM early in childhood. But among affected infants and preadolescent children, dominant and recessive variants are identified in many additional genes including those involved in developmental syndromes, neuromuscular disorders, inborn errors of metabolism, and fatty acid oxidation defects (Rojnueangnit et al., 2019). As these genes are uncommonly identified in adult DCM, they provide additional insights into critical molecules and pathways of cardiomyocyte biology.

Recessive homozygous and compound heterozygous damaging (loss-of-function or deleterious missense) variants in ALPK3 (α-kinase-3) were recently identified to
cause prenatal or neonatal cardiomyopathy (Almomani et al., 2016) with systolic dysfunction and thinned ventricular walls, typical of DCM. Remarkably, some infants with typical DCM rapidly developed ventricular hypertrophy, a most unusual sequence in either adult or pediatric cardiomyopathies. ALPK3 cardiomyopathies can occur in isolation or with skeletal muscle abnormalities, joint contractures, scoliosis, and/or subtle facial dysmorphisms, and often results in rapidly progressive heart failure that necessitates cardiac transplantation or causes death (Almomani et al., 2016; Phelan et al., 2016; Çağlayan et al., 2017; Jaouadi et al., 2018; Al Senaidi et al., 2019). ALPK3 has not emerged as the cause of adult-onset cardiomyopathy; however exome-wide association studies of adults with sporadic DCM identified the ALPK3 locus (Esslinger et al., 2017). Experimental models have recapitulated some of these human ALPK3 phenotypes. Alpk3/- mice develop ventricular hypertrophy with reduced cardiac output within a month of birth and show disordered intercalated discs, structures which allow mechanical and electrical coupling of cardiomyocytes (Van Sligtenhorst et al., 2012). ALPK3-deficient patient-derived human cardiomyocytes derived from induced pluripotent stem cells (hiPSC-CMs) display disorganization or sarcomeres and intercalated discs, and abnormalities in calcium handling, features that are observed in DCM (Phelan et al., 2016).

Despite evidence that ALPK3 ablation induces cardiomyocyte dysfunction, an understanding of the molecular function(s) of ALPK3 in cardiomyocyte biology and the mechanism(s) by which ALPK3 deficiency cause disease, remain poorly understood.

ALPK3 encodes a protein kinase, molecules that typically catalyze the transfer of $\gamma$-phosphates from ATP molecules to protein substrates at serine and threonine or at
tyrosine residues. However, protein kinases can also signal through non-catalytic mechanisms, and may therefore have important catalytic and/or non-catalytic functions (Manning et al., 2002; Miranda-saavedra, Barton and Alessi, 2006; Kung and Jura, 2016).

Based upon sequence similarity of the kinase domain, protein kinases are classified into conventional protein kinases (CPKs, 90%) and atypical protein kinases (APKs, 10%) (Manning et al., 2002). One subfamily of APKs, the $\alpha$-kinases, appeared more recently in evolution and are found only in eukaryotes (Drennan and Ryazanov, 2004). ALPK3 is one of six human $\alpha$-kinases and is expressed early in cardiomyocyte development, and its expression remains restricted to striated muscle (Hosoda et al., 2001), suggesting that this particular $\alpha$-kinase has evolved to fulfill unique, muscle-specific roles in eukaryotes.

Elucidation of the crystal structure of the $\alpha$-kinase domain from the cation channel kinase, TRPM7, has shown that the structures of CPK domains and $\alpha$-kinase domains are remarkably similar, despite their lack of sequence similarity (Yamaguchi et al., 2001). Similar to CPKs, $\alpha$-kinases have an N-terminal lobe (N-lobe) that folds into a curved $\beta$-sheet and contains a phosphate binding loop (P-loop). The C-terminal lobe (C-lobe) consists mainly of $\alpha$-helices, and contains the activation loop, which has key phosphorylated residues that can enhance catalytic potential (Adams, 2003). Between the N-lobe and C-lobe lies the interlobe cleft, which contains key residues for ATP positioning and catalysis. In contrast to CPKs, $\alpha$-kinases also contain a zinc-finger motif which coordinates a zinc ion to stabilize the tertiary structure of the catalytic core (Yamaguchi et al., 2001). While knowledge of the $\alpha$-kinase structure can inform
experiments to elucidate $\alpha$-kinase functions, the biological roles of $\alpha$-kinases, including ALPK3, remain understudied.

We interrogated the kinase function of ALPK3 to better understand the pathogenesis of ALPK3-cardiomyopathy. By integrating sequence conservation, proteomic, and phosphoproteomic analyses of hiPSC-CMs with ALPK3-variants that vary gene dosage or conformationally-alter the kinase domain, we show that ALPK3 is catalytically inactive, and thus a pseudokinase. However, we demonstrate that expression of ALPK3 and its conserved $\alpha$-kinase fold is required for the assembly of myomesins (MYOM1 and MYOM2), which co-localize with ALPK3 at the sarcomere M-band and nuclear envelope, and for the appropriate turnover of additional sarcomere thick filament proteins. These results implicate ALPK3 as a critical scaffolding protein with crucial non-catalytic functions in cardiomyocyte biology.
Results

Structure and conservation of the ALPK3 α-kinase domain

*ALPK3* encodes a 201kDa protein with two immunoglobulin-like domains and a C-terminal α-kinase domain that spans 236 amino acids (**Fig 3.1a**). We compared the conservation of the α-kinase domain amongst all α-kinases by sequence alignment of the six human α-kinases proteins. This alignment identified 16 invariant residues in all six proteins (**Fig 3.1b**). Additionally, there were four invariant residues in five human α-kinases, that were different in ALPK3. We denote these four amino acids as ALPK3-variant residues. Notably, these four residues were also invariant in α-kinase orthologs in mouse, chicken, frog, and zebrafish, as well in evolutionarily ancient α-kinase domains from the slime mold, *dictyostelium discoideum* (**Fig S3.1**).

The functions of α-kinase domain residues have been established based on structural homology of the TRPM7 α-kinase domain to protein kinase A (PKA), a well-studied conventional protein kinase (**Fig 3.1c-d**). Several important, conserved residues within the α-kinase domain lie in the P-loop and interlobe cleft and participate in positioning ATP. Others residues within the C-lobe are involved in chelating zinc and hydrophobic packing that stabilizes the tertiary structure of the domain (Yamaguchi *et al.*, 2001; Drennan and Ryazanov, 2004). Notably, of the four ALPK3-variant residues, two (K1729, Q1769) position ATP, and one (N1797) is predicted to be involved in substrate binding (**Fig 3.1c-d**). Thus, while ALPK3 retains the majority of the conserved and presumptive functional residues in the α-kinase domain, it also shows evolutionary divergence at a few residues that may be critical for ATP-binding and catalysis.
Figure 3.1. Structure of ALPK3 and key residues of the ALPK3 α-kinase domain

(A) Structure of human ALPK3 (14 exons, 3 protein domains).
(B) Multiple sequence alignment of the α-kinase domains from all six human α-kinase proteins. Amino acids are numbered by their position in the ALPK3 α-kinase domain. The 16 invariant residues are indicated (red). Four invariant residues in five α-kinases and their orthologs, but not in ALPK3 (ALPK3-variant residues) are indicated (green). Two invariant residues (G1777 and D1784), chosen for CRISPR/Cas9 mutagenesis, are highlighted (lightening symbol).
(C) Function of key residues in the TRPM7 α-kinase domain (see Drennan and Ryazanov, 2004) and the homologous residues in ALPK3. Color and symbols denote annotations detailed in (B).
(D) Crystal structure of the TRP M7 α-kinase domain (see Yamaguchi et al., 2001). Residues in red were invariant in all α-kinases; color and symbols denote annotations detailed in (B). The locations of the N-terminal lobe (N-lobe) containing the phosphate-binding loop (P-loop), C-terminal lobe (C-lobe) containing the activation loop, and the interlobe cleft are indicated. The interlobe cleft contains conserved residues for binding and positioning ATP. A zinc-finger motif in the C-lobe contains conserved residues for coordinating zinc and stabilizing the tertiary structure of the α-kinase domain.
CRISPR/Cas9 genome engineering of ALPK3 mutant hiPSC cell lines and effects on ALPK3 expression

Using CRISPR/Cas9 mutagenesis strategies, we introduced ALPK3 variants into the human induced pluripotent stem cell (hiPSC) line PGP1-TTN-GFP (Fig 3.2a). This stem cell line contains an endogenous eGFP tag on the N-terminus of the sarcomere protein titin (eGFP-TTN), which marks sarcomere ends (Z-discs), and enables functional analyses of sarcomere contractility (Sharma, Toepfer, Schmid, et al., 2018; Toepfer, Sharma, et al., 2019a). The ALPK3+/− hiPSC line carried two frameshift variants, one on each allele (c.2442_2443insA is predicted to result in an insertion of one missense residue at position 815, followed by a premature stop codon at position 839 and c.2428_2458del31 is predicted to delete residues 809-819, followed by a premature stop codon at position 853). The hiPSC line ALPK3G1777ED1784A/G1777ED1784A carried homozygous missense variants at two residues that were invariant in the α-kinase domain of all family members. Based on the position of the homologous residues in the TRPM7 α-kinase domain, the G1777 and D1784 residues are predicted to lie in the interlobe cleft of the α-kinase domain, in close proximity to ATP, and D1784 is predicted to directly participate in ATP binding (Fig 3.1d). Thus, the homozygous compound missense variants (G1777E, D1784A) are predicted to alter the conformation of the interlobe cleft of the α-kinase domain and destabilize ATP-binding. We denote the ALPK3G1777ED1784A/G1777ED1784A mutant line as ALPK3-active-site. Each mutant hiPSC line was subcloned and sequence-validated (Fig S3.2a-b).

Using an established small-molecule-based 30-day differentiation protocol (Lian et al., 2013), WT and ALPK3 mutant hiPSC lines were differentiated into
cardiomyocytes. All differentiations produced beating cardiomyocytes. To determine whether loss of \textit{ALPK3} gene expression affected cardiomyocyte maturation, we conducted RNAseq of independent differentiations of WT and \textit{ALPK3}\textsuperscript{-/-} hiPSC-CMs, and applied principal component analysis to compare their transcriptional profiles at various stages of differentiation (\textbf{Fig S3.2c}). There were no significant differences in transcriptional profiles throughout the 30-day differentiation protocol that yielded enriched, maximally mature hiPSC-CMs. Similar to WT, \textit{ALPK3}\textsuperscript{-/-} hiPSC-CMs achieved high level expression of prototypic cardiomyocyte genes (\textbf{Fig S3.2c-d}).

We investigated protein expression and phosphorylation in WT and \textit{ALPK3} mutant hiPSC-CMs using quantitative mass-spectrometry-based (tandem-mass-tag; TMT) phosphoproteomics of protein extracts from each hiPSC-CM line (Navarrete-perea \textit{et al.}, 2018). This approach allows sample multiplexing by labeling peptides (digested from \(\sim 100\mu\text{g}\) of protein per line) with sample-specific isobaric tandem mass tags (TMT). Phosphopeptides are subjected to an additional centrifugation-based enrichment procedure, and both the phosphopeptide-enriched fractions and flow-through fractions are analyzed by mass spectrometry to determine the relative abundance of both phosphopeptides and proteins among samples (refer to METHODS).

We quantified 3,714 phosphopeptides and 7,341 proteins in WT, \textit{ALPK3}-active-site, and \textit{ALPK3}-null hiPSC-CMs. We first analyzed this dataset to compare ALPK3 protein expression amongst cell lines. Twenty peptides mapped specifically to ALPK3, and comparative quantification indicated that \textit{ALPK3}-null hiPSC-CMs had 5\% of WT ALPK3 protein levels, with no evidence for expression of a truncated protein (\textbf{Fig 3.2b}). Concurrent comparative analyses of RNAseq data from these lines showed a
concordant decrease in ALPK3 mRNA levels (Fig S3.2c) in ALPK3-null hiPSC-CMs. Unexpectedly, ALPK3 protein expression in ALPK3-active-site hiPSC-CMs was approximately twice the levels found in WT hiPSC-CMs (Fig 3.2b).

Figure 3.2. CRISPR/Cas9 generation of isogenic ALPK3 mutant hiPSC models

(A) Targeted mutagenesis of the hiPSC line PGP1-TTN-GFP yielded WT, ALPK3-null, and ALPK3-active site mutant lines (see also Fig S3.1 for sequence validation). (B) ALPK3 protein expression in WT and mutant hiPSC-CMs, assessed by mass spectrometry. X-axis indicates positions of peptides in the ALPK3 protein sequence. ALPK3 protein level in ALPK3-null hiPSC-CMs is <5% of WT levels. ALPK3 protein level in ALPK3-active-site hiPSC-CMs is ~two-fold higher than WT levels. Dotted line indicates relative position of the CRISPR guideRNA cut site that was used for engineering of the ALPK3-null mutant. Data are depicted as mean±SEM. Replicates indicate independent cardiomyocyte differentiations (WT: n= 4, ALPK3-active-site: n=4, ALPK3-null: n=3) of hiPSC-CMs.
**ALPK3 does not have significant catalytic activity**

As ALPK3 was predicted to be a catalytically-active kinase, we expected that loss of ALPK3 expression or destabilization of its ATP-binding function would significantly decrease the phosphorylation of target substrates. Thus, we compared the phosphoproteomes of WT and mutant hiPSC-CMs to identify phosphosites with decreased phosphorylation in ALPK3-null and ALPK3-active-site hiPSC-CMs. ALPK3-null hiPSC-CMs had 225 differentially phosphorylated (DP) sites, but remarkably ~75% of DP sites had increased phosphorylation relative to WT (Fig 3.3a). We deduced from these data that most of these phosphosites did not represent direct ALPK3 substrates, but were altered indirectly as a consequence of loss of ALPK3 expression. By contrast, ALPK3-active-site hiPSC-CMs had only two phosphosites with significantly decreased phosphorylation relative to WT hiPSC-CMs, and both these phosphosites occurred on the ALPK3 protein itself (S1384, S1855) (Fig 3.3b).

With our prior observations that key residues in the α-kinase domain of ALPK3 showed evolutionary divergence, and phosphoproteomic evidence for remarkably few differentially phosphorylated sites in ALPK3-active-site hiPSC-CMs, we deduced that ALPK3 is catalytically-inactive, or retains only minimal catalytic activity or possibly autophosphorylative function.

To confirm these findings, we transiently overexpressed ALPK3 in HEK293 cells, and compared the phosphoproteome of transfected and control cells (Fig 3.3c-d). TMT phosphoproteomics identified 6,596 unique proteins and 1,158 unique phosphopeptides amongst transfected and control lines. ALPK3 protein levels were greater than 16-fold higher in transfected versus control cells; Western blot confirmed substantial over-
expression (Fig S3.3). The expression of other proteins was not significantly different, however, nine phosphopeptides had increased abundance in ALPK3-transfected versus control cells (Fig 3.3d). All nine phosphopeptides mapped to phosphosites on the ALPK3 protein (S430, T728, S1209, S1384, S1394, S1406, S1424, S1855, S1865). The abundance of these phosphopeptides is directly correlated with the increased protein expression of ALPK3, indicating that this protein undergoes phosphorylation by unknown kinases, or possibly by self. As ALPK3 overexpression did not significantly alter the phosphorylation of any other phosphosites in transfected cells, we concluded that ALPK3 is a pseudokinase, with minimal, if any, catalytic activity.
Figure 3.3. Phosphoproteomics of ALPK3 inhibition and overexpression demonstrates lack of significant ALPK3 catalytic activity

(A) Quantitative phosphoproteomics was used to determine the phosphorylation state of 3,714 phosphosites, which mapped to 1,630 unique proteins, in WT, ALPK3-active-site, and ALPK3-null hiPSC-CMs. The relative phosphorylation of each phosphosite was determined by normalizing phosphopeptide abundance to protein abundance. Heatmap shows differentially phosphorylated phosphosites, scaled by row. Relative to WT hiPSC-CMs, there were 225 differentially phosphorylated (DP) sites in ALPK3-null hiPSC-CMs (54 down, 171 up) and 3 differentially phosphorylated sites in ALPK3-active-site hiPSC-CMs (2 down and 1 up). DP criteria: (|log₂-fold change| > 0.67 & p-value < 0.01, unpaired t-test). Replicates represent independent differentiations (WT: n=4; ALPK3-active-site: n=4; ALPK3-/-: n=3) of each hiPSC-CM line.
(B) Relative phosphorylation of ALPK3 phosphosites in WT and ALPK3-active-site hiPSC-CMs. The (only) two phosphosites with significantly reduced phosphorylation in ALPK3-active-site hiPSC-CMs mapped to the ALPK3 protein itself (S1394, S1855).

(Figure 3.3 continued)

(C) Log-log plot of the abundance of 6,596 proteins in ALPK3-transfected versus non-transfected HEK293 cells, determined via quantitative proteomics. Each point represents a single protein, and the density of points in a given region of the plot is indicated with color. Replicates represent independent transient transfections and their paired controls (transfected: n=2; control, n=2).

(D) Log-log plot of the abundance of 1,158 phosphopeptides in ALPK3-transfected versus non-transfected HEK293 cells. Each point represents a single phosphopeptide, and the density of points in a given region of the plot is indicated with color. Replicates represent independent transient transfections and their paired controls (transfected: n=2; control: n=2). The levels of nine phosphopeptides (highlighted in grey box) were highly increased in ALPK3-transfected versus control cells. All nine phosphopeptides mapped to phosphorylation sites on the ALPK3 protein.
**ALPK3-null and α-kinase domain destabilizing mutations result in increased myomesin protein levels in hiPSC-CMs**

As pseudokinases can function as chaperones or scaffolds for orchestrating the assembly and localization of signaling complex (Miranda-saavedra, Barton and Alessi, 2006), we further examined DE proteins in ALPK3 mutant hiPSC-CMs. There were 346 DE proteins in ALPK3-null and 28 DE proteins in ALPK3-active-site hiPSC-CMs (Fig 3.4a). Ten DE proteins were shared amongst both ALPK3 mutants (Fig 3.4b). The presence of shared proteins between ALPK3-active-site and ALPK3-null hiPSC-CMs suggested that conformationally altering the α-kinase domain fold conveyed partial loss-of-function, which occurred despite no apparent ALPK3 catalytic activity.

Five DE proteins that were shared amongst ALPK3-null and ALPK3-active-site hiPSC-CMs encoded proteins of the sarcomere thick filament. Four of these proteins, myomesin-1 (MYOM1), myomesin-2 (MYOM2), myosin heavy chain (MYH6), myosin light chain (MYL3), had increased levels; one protein, TRIM63, also called MuRF1, a known sarcomere M-band and Z-disc associated E3 ubiquitin-protein ligase, had decreased levels. Notably, one major substrate of MuRF1 is myosin heavy chain (Clarke et al., 2007). In both ALPK3 mutants, the steady-state levels of myomesins and myosin thick filament proteins were increased, whereas the levels of other sarcomere proteins, including proteins of the thin-filament, were not significantly affected (Fig 3.4c). With the exception of MYH6, corresponding increases in mRNA levels were not observed in ALPK3+/− hiPSC-CMs, suggesting that the observed changes in protein levels were due predominantly to post-transcriptional mechanisms (Fig 3.4d). Amongst DE protein in ALPK3-null hiPSC-CMs, there was significant enrichment for GO terms
related to the M-band and myosin complexes, cell-cell adherens junctions, and cadherin-mediated adhesion (**Fig S3.4**).

The dysregulated expression of key proteins of the sarcomere prompted evaluation of contractile function in *ALPK3* mutant hiPSC-CMs. Harnessing GFP-labeled sarcomeres in hiPSC-CMs, we measured sarcomere lengths and contractile velocities (Sharma, Toepfer, Schmid, *et al.*, 2018). Contraction of over 40,000 sarcomeres per cell line was recorded using fluorescent video microscopy, and processed using SarcTrack, a high-throughput image analysis platform to assess sarcomere contractile function (Toepfer, Sharma, *et al.*, 2019). These analyses showed that sarcomere lengths and overall contractile function in both *ALPK3*-null and *ALPK3*-active-site hiPSC-CMs were not significantly affected (**Fig 3.4e**).
Figure 3.4. *ALPK3* null and kinase domain destabilizing mutations in hiPSC-CMs result in increased myomesin and thick filament protein levels.

(A) Volcano plots of DE proteins in *ALPK3*-null and *ALPK3*-active-site hiPSC-CMs. Each point depicts the expression level of one of 5,583 total proteins, determined via quantitative proteomics. There were 347 DE proteins in *ALPK3*-null and 29 DE proteins in *ALPK3*-active-site hiPSC-CMs relative to WT. The most significant DE proteins are highlighted; red text indicates ALPK3 and proteins of the thick filament, and other significant DE proteins are indicated in blue. (DE criteria: $|\log_2$-fold change$| > 0.67$ & unpaired t-test p-value $< 0.01$). Replicates represent independent differentiations of hiPSC-CMs (*WT*: $n=4$, *ALPK3*-active-site: $n=4$; *ALPK3*-/: $n=7$).

(B) Overlap of DE proteins from *ALPK3*-null and *ALPK3*-active-site hiPSC-CMs.

(C) Protein levels of shared DE proteins in *ALPK3*-null and *ALPK3*-active-site hiPSC-CMs with the levels of other sarcomere proteins included for comparison. Data are depicted as mean±SEM. Numbers in parentheses indicate numbers of peptides identified by mass spectrometry, the levels of which were averaged for protein quantification.

(D) RNA levels of sarcomere genes in hiPSC-CMs, determined via RNAseq. Data are depicted as mean±95% confidence interval (*WT*: $n=3$; *ALPK3*-null, $n=3$).

(E) Sarcomere contractility (contraction %) in WT and *ALPK3* mutant hiPSC-CMs, assessed via SarcTrack (Toepfer, Sharma, *et al.*, 2019a). Each data point represents an averaged estimate of sarcomere contractility from one video, with mean±SEM indicated. The total number of sarcomeres analyzed ($n$) from number of videos ($v$) and number of differentiations ($d$) were as follows (*WT*: $n=43,761, v=94, d=2$; *ALPK3*-active-site: $n=46,720, v=95, d=2$; *ALPK3*-null: $n=47,977, v=95, d=2$). Comparisons to WT were assessed using unpaired t-tests.
Alpk3−/− mice develop DCM in early postnatal life and have increased left ventricular myomesin protein levels

As cultured hiPSC-CMs cannot fully recapitulate in vivo biology, we coupled these analyses with studies of the functional and molecular properties of C57BL/6N Alpk3-deficient mice, distributed by the MRC Harwell Institute (Blake et al., 2010). Alpk3+/- mice were bred to produce Alpk3−/− mice, which were born at the expected Mendelian ratio of ~25%.

Cardiac function was assessed by echocardiography at birth (postnatal day 0, P0), and intermittently throughout 52 weeks. WT and Alpk3+/- mice had comparable cardiac morphology and birth and throughput life. At birth, Alpk3−/− mice had comparable cardiac function as WT mice, but during the first week of life rapidly developed progressive features of DCM including increased left ventricular (LV) dimensions, and severely diminished contractile function (Fig 3.5a). While hypertrophy can emerge in some human patients with biallelic ALPK3 deficiency, left ventricular wall thickening was not observed in young Alpk3−/− mice. However, the severe, progressive DCM in these mice resulted in early mortality, and precluded in vivo cardiac studies beyond 14 weeks of age.

Cardiac contractility in vivo is a function of sarcomere performance, but is also influenced by other in vivo parameters, including neurohormonal factors, heart rate, hemodynamic load, and myocardial geometry. To consider if intrinsic cardiomyocyte properties or extrinsic factors accounted for severely depressed cardiac function, we studied isolated WT and Alpk3 mutant cardiomyocytes. While we expected that ex vivo analyses would demonstrate reduced contractility in Alpk3−/− cardiomyocytes, we found
the opposite; \textit{Alpk3}\textsuperscript{-/-} cardiomyocytes had significantly greater contractility than WT cardiomyocytes. The contractility of WT and \textit{Alpk3}\textsuperscript{+/-} cardiomyocytes was similar (\textbf{Fig 3.5b}).

To investigate whether the expression of \textit{Alpk3} influenced the cardiac transcriptome, we performed RNAseq on LV tissue from WT, \textit{Alpk3}\textsuperscript{+/-} and \textit{Alpk3}\textsuperscript{-/-} mice at P0 and P8. At birth, the RNA expression profiles of hearts from both mutant lines were indistinguishable from WT. At P8, the transcriptome of \textit{Alpk3}\textsuperscript{+/-} mice remained indistinguishable from WT, while \textit{Alpk3}\textsuperscript{-/-} mice had 273 differentially expressed (DE) RNAs relative to WT (\textbf{Fig 3.5c}). DE RNAs included prototypic stress-responsive molecules (Nppa) and showed enrichment in GO terms associated with organ development and morphogenesis (\textbf{Fig S3.5a}).

We combined these transcriptional analyses with assessment of cardiac proteomes. Using TMT proteomics, we quantitated the relative expression of 5,789 proteins from LV tissues from WT, \textit{Alpk3}\textsuperscript{+/-} and \textit{Alpk3}\textsuperscript{-/-} mice at P8. The proteomes of \textit{Alpk3}\textsuperscript{+/-} mice were indistinguishable from WT, while 64 DE proteins were identified in \textit{Alpk3}\textsuperscript{-/-} mice (\textbf{Fig 3.5d}).

Comparison of the DE proteins from \textit{Alpk3}\textsuperscript{-/-} mice and ALPK3-mutant hiPSC-CMs revealed three shared proteins (myomesin-1 (MYOM1), myomesin-2 (MYOM2) and myosin isoforms (Myh7 in mice; MYH6 in humans). \textit{Myom1} has two splice isoforms (\textbf{Fig S3.6a}); Western blotting showed that the protein level of both splice isoforms was increased in \textit{Alpk3}\textsuperscript{-/-} mice (\textbf{Fig S3.6b}). The corresponding mean LV RNA levels were unchanged or modestly increased (but below DE criteria) in the \textit{Alpk3}\textsuperscript{-/-} compared to WT mice, but not in proportion to the large increase in protein levels (\textbf{Fig 3.5e-f}), suggesting
post-transcriptional mechanisms. Notably, the protein levels of many other sarcomere components, including proteins of the thin filament, were not significantly affected (Fig 3.5e).
Figure 3.5. *Alpk3*−/− mice develop DCM in early postnatal life and have increased left ventricular myomesin protein levels

(A) Metrics of left ventricular (LV) function, assessed by echocardiogram, in mice at birth (P0) to 52 weeks. Cardiac function was not assessed in *Alpk3*−/− mice after 14 weeks due to early lethality. (LVID;d = LV internal dimension at end-diastole, LVPW = LV posterior wall thickness).

(B) Sarcomere contractility (contraction %) of isolated cardiomyocytes *ex vivo*, paced at 1Hz. Each data point represents an individual cardiomyocyte (*WT* n = 104, *Alpk3*+/− n = 100, *Alpk3*−/− n = 79). Cardiomyocytes were isolated from 3 mice per genotype; mice were between 7-16 weeks of age.

(C) Heatmap of differentially expressed (DE) RNAs, identified by RNAseq, in mouse LV tissues obtained at postnatal day 0 (P0) and postnatal day 8 (P8), scaled by row. Each column represents RNAseq of a different mouse. Numbers of DE RNAs are displayed below heatmap. (DE criteria: |log2fold-change > 0.67| & Benjamini-Hochberg-adjusted p-value < 0.1).

(D) Heatmap of DE proteins in *WT* and *Alpk3* mutant mouse P8 LV tissues, scaled by row. Each column represents proteomics of a different mouse. Numbers of DE proteins are indicated in parentheses. Bold stars indicate proteins that were also dysregulated in *ALPK3*-null and *ALPK3*-active-site hiPSC-CMs (There were 5,789 total proteins identified through TMT proteomics; DE criteria: |log2fold-change > 0.515| & p-value < 0.1).

(E) Protein levels of shared DE proteins and other sarcomere proteins in *Alpk3*+/− and *Alpk3*−/− P8 mice relative to *WT*. Data are depicted as mean±SEM. Numbers in parentheses indicate numbers of peptides identified by mass spectrometry, the levels of which were averaged for protein quantification.

(F) Mean LV RNA levels (±95% confident interval) of sarcomere genes from P8 *Alpk3* mice, determined by RNAseq. RNA was obtained from *WT*, n=3, *Alpk3*+/− (n=2) and *Alpk3*−/− (n=3) mice.
ALPK3 and myomesin proteins co-localize at the nuclear envelope and sarcomere M-band in hiPSC-CMs.

The protein levels of MYOM1 and MYOM2 were significantly increased in vitro in ALPK3-null and ALPK3-active-site hiPSC-CMs and in the LV of Alpk3−/− mice. Prior studies demonstrated the nuclear localization of ALPK3 in transfected COS cells (Hosoda et al., 2001) and MYOM1 in neonatal rat cardiomyocytes (Reddy et al., 2008), and both MYOM1 and MYOM2 at the sarcomere M-band, a central structure where the C-terminus of TTN and thick filaments are anchored (Grove et al., 1985; Obermann et al., 1996). Given these associations, we explored the localization of ALPK3, MYOM1, MYOM2 in hiPSC-CMs.

Given the lack of specific ALPK3 antibodies, we transiently transfected hiPSC-CMs with a plasmid encoding the ALPK3 ORF with a C-terminal FLAG epitope tag. Anti-FLAG immunofluorescence showed that ectopically expressed ALPK3 protein localized to two specific structures: the nuclear envelope and the sarcomere M-band (Fig 3.6a).

We identified some diffuse MYOM1 in some nuclei of WT hiPSC-CMs (data not shown), but less than was described in 3-day-old rat cardiomyocytes (Reddy et al., 2008), and most MYOM localized distinctly to the M-band. ALPK3−/− hiPSC-CMs showed more MYOM1 staining, within the M-band, but also diffusely in the cytoplasm, suggesting misassembly (Fig 3.6b, panels 1-4).

MYOM2 antibodies did not stain the M-band in WT hiPSC-CMs, as has been previously reported in studies of adult myocytes (Grove et al., 1985; Obermann et al., 1996). In hiPSC-CMs, MYOM2 antibodies prominently staining the nuclear envelope. The specificity of the MYOM2 antibody used in these studies was validated by Western
Blotting, which showed protein levels MYOM2 in WT and ALPK3 mutant hiPSC-CMs that were consistent with those measured by mass spectrometry (Fig S3.6c). While some MYOM2 staining was present in the nuclei of ALPK3−/− hiPSC-CMs, specific localization at the nuclear envelope was largely absent (Fig 3.6b, panels 7-8).
Figure 3.6. ALPK3 and myomesin proteins co-localize at the nuclear envelope and sarcomere M-band in hiPSC-CMs.

(A) Immunofluorescence of ALPK3-FLAG construct transfected in hiPSC-CMs. eGFP-TTN marks sarcomere Z-discs. ALPK3-FLAG localizes to the M-band, which lies between Z-discs (panel 2). ALPK3-FLAG also localizes to the nuclear envelope (panel 3, white arrows; green arrows depicts ALPK3-FLAG signal at sarcomeres proximal to the nucleus). Scale bars: 25 \( \mu \text{m} \) in panel 1, 5 \( \mu \text{m} \) in panels 2-3.

(B) Immunofluorescence staining of MYOM1 and MYOM2 in WT and ALPK3-null hiPSC-CMs. Panels 1-2 show MYOM1 localization to the M-band. Panels 3-4 show MYOM1 staining in ALPK3-null hiPSC-CMs both at the M-band, and also diffusely in the cytoplasm (arrows). Panels 5-6 show MYOM2 localization to the nuclear envelope. Panels 7-8 show less prominent MYOM2 staining at the nuclear envelope (arrows) in in ALPK3-null hiPSC-CMs. Scale bars = 25 \( \mu \text{m} \).
Discussion:

The molecular functions of ALPK3 remain largely uncharacterized, even though the expression of a putative kinase early in cardiomyocyte development, and the profound myocardial dysfunction evoked by its depletion signify critical physiologic roles in cardiac biology. Here, we demonstrated several new insights into this enigmatic protein. We show that ALPK3 lacked significant catalytic activity, thus establishing it as a pseudokinase. Gene variants that abolished ALPK3 expression, or altered the conformation of the $\alpha$-kinase domain yielded comparable deleterious effects and thereby establish the pathogenicity of both null and missense alleles. Strikingly, these damaging variants increased levels of myomesin-1 (MYOM1) and myomesin-2 (MYOM2) proteins in hiPSC-CMs and Alpk3$^{-/-}$ mice, and also perturbed the expression of other thick filament components, including myosin heavy chains and myosin light chain. We found that ALPK3 localized to two structures in hiPSC-CMs – the M-band, along with MYOM1, and the nuclear envelope, along with MYOM2. Moreover, depletion or dysfunction of ALPK3 in hiPSC-CMs perturbed incorporation of MYOM1 and MYOM2 at the M-band and nuclear envelope, respectively. Collectively, these results indicate that ALPK3 functions as a critical scaffold for myomesin protein assembly at two important subcellular locations in cardiomyocytes, which is essential for the protein homeostasis and appropriate turnover of myomesins and thick filament proteins (Fig 3.7).

The M-band is a critical sarcomere structure at which antiparallel thick filaments are cross-linked. Although the full repertoire of M-band functions remains elusive, recent studies suggest that the M-band maintains thick filaments in register, and absorbs force
imbalances generated by opposing thick filaments during contraction (Xiao and Grater, 2014; Lange et al., 2019). Prior studies have demonstrated that both MYOM1 and MYOM2 are major constituents of the M-band in adult cardiomyocytes (Grove et al., 1985; Obermann et al., 1996). The C-terminal domains of myomesins are highly distensible, with α-helices that can rapidly unfold and refold, so as to provide both extensibility and force buffering at the M-band (Xiao and Grater, 2014).

Consistent with earlier observations, we found that MYOM1 was robustly expressed at the M-band in hiPSC-CMs. Previous studies also reported that MYOM2 is absent from the M-band in embryonic cardiomyocytes (Grove et al., 1985); while, we also did not find MYOM2 at the M-band in hiPSC-CMs, cells that are relatively immature in comparison to adult cardiomyocytes (DeLaughter et al., 2016), we identified that MYOM2 localized specifically to the nuclear envelope, suggesting a role for this myomesin protein in early development. Furthermore, prior studies of neonatal rat cardiomyocytes reported that MYOM1 shuttled between the nucleus and cytosol via a sumoylation-dependent mechanism (Reddy et al., 2008). While we observed MYOM1 is some hiPSC-CM nuclei, MYOM2 was far more prominent. Whether alternative myomesin expression at these two intracellular structures reflects species differences, antibody specificity, or other parameter is unknown. Regardless, these observations suggest that ALPK3, MYOM1, and MYOM2 are all part of a developmental signaling network that connects the nucleus and the M-band. We speculate that nuclear myomesin absorbs mechanical forces, as it does at the M-band (Pernigo et al., 2017), and thereby helps to maintain nuclear structural integrity despite close contact with contracting sarcomeres.
Our studies demonstrated that damaging *ALPK3* variants resulted in increased levels of both MYOM1 and MYOM2 and other thick filament proteins including myosin. Moreover, immunohistochemistry of MYOM1, which is more abundant than MYOM2, appeared to accumulate outside the M-band in hiPSC-CMs. The N-termini of MYOM1 and MYOM2 mediate myosin binding, while the C-termini mediate protein dimerization (Obermann *et al.*, 1996; Auerbach *et al.*, 1999). As MYOM1 directly interacts with myosin heavy chain, it is possible that loss of ALPK3-myomesin interactions broadly affects the turnover and organization of thick filament proteins.

Unexpectedly, missense variants that are predicted to destabilize ATP-binding and conformationally alter the α-kinase domain resulted in an nearly a two fold-increase in ALPK3 protein levels and also increased myomesin protein levels. These results implied a direct interaction between ALPK3 and myomesin proteins, possibly involving the α-kinase domain itself, and that loss of this interaction results in the altered turnover of ALPK3-myomesin protein complexes. Notably, the immunoglobulin (Ig) domains of both ALPK3 and MYOM1 share >30% sequence identity by protein BLAST, indicating that elements outside the ALPK3 α-kinase domain may also mediate interactions. The development of antibodies that specifically recognize ALPK3 would allow further investigation of ALPK3 protein interactions, and further elucidate the role of this protein in cardiomyocytes.

Given that ALPK3 encodes a protein kinase, the lack of apparent ALPK3 catalytic activity, assessed by phosphoproteome-wide analyses, was surprising. While recognizing that this experimental platform might fail to detect some changes in phosphorylation, and that the expression of co-factors or stimuli required for kinase
activation may not be faithfully recapitulated by developmentally-immature hiPSC-CMs, comparative sequence data supported our conclusion that ALPK3 is a pseudokinase. Human ALPK3 and its orthologs are divergent at several residues in the \(\alpha\)-kinase interlobe cleft – residues that are conserved in all other \(\alpha\)-kinase family members down to primitive slime molds – hinting that ALPK3 lost its catalytic activity early in its evolution. Approximately 10% of all protein kinases encoded by the human genome lack one of three conserved catalytic residues (corresponding to TRPM7 K1648, D1767, D1777), and as such, are bioinformatically predicted to be catalytically-inactive pseudokinases. (Manning et al., 2002). Though these particular residues were conserved in ALPK3, four other highly-conserved residues (G1725, K1729, Q1769, N1797) were variant in the ALPK3 \(\alpha\)-kinase domain. Thus, one, or a combination of these residues, is likely essential for \(\alpha\)-kinase catalytic function. Despite the apparent lack of ALPK3 \(\alpha\)-kinase catalytic activity, our studies demonstrated that the \(\alpha\)-kinase domain had critical roles in determining myomesin levels, suggesting that this domain retains critical scaffolding roles independent of catalytic function.

These data add to the growing evidence that one major function of the M-band is to scaffold proteins that allow for the dynamic regulation of thick filament protein turnover in the sarcomere. Recent data indicate that like ALPK3, the kinase domain of TTN that resides in the M-band is also a pseudokinase, and functions in scaffolding (Bogomolovas et al., 2014). Notably, the M-band kinase domain of TTN binds ubiquitin E3 ligases, including TRIM63 (MuRF2), which was downregulated in ALPK3-active-site hiPSC-CMs, supporting functional linkages between TTN- and ALPK3-scaffolding. Evidence that MuRF2 binds serum response factor (SRF) and shuttles across the
nuclear membrane where MYOM2 colocalizes further supports a model whereby force-dependent signaling by the M-band modulates the expression and protein homeostasis of force-generating thick filament proteins (Lange et al., 2005).

We assessed how these perturbations affected cardiomyocyte contractility. Notably WT and mutant ALPK3 hiPSC-CMs were indistinguishable, while newborn Alpk3−/− mice developed rapidly progressive DCM with severely decreased cardiac contractile function (~10% fractional shortening). Remarkably, however, ex vivo Alpk3−/− cardiomyocytes had more robust contraction than WT or Alpk3+/− littermates. Earlier histopathologic studies of Alpk3−/− mouse hearts and hiPSC-CMs report morphologic abnormalities in intercalated discs, structures that mechanically couple cardiomyocytes to enable coordinated in vivo contraction that is not assessed in isolated cells. While we did not restudy intercalated disc morphology in these studies, we note that dysregulated proteins in ALPK3-null hiPSC-CMs were enriched for cadherins and cell-cell adhesion components. Furthermore, the expression of two protein components of intercalated discs, FRX1 and XIRP1, were dysregulated in Alpk3−/− mouse hearts. Furthermore, in vivo cardiomyocytes are subjected to hemodynamic forces and loads that are not present in isolated ex vivo cardiomyocytes. Nuclear-M-band-structure and signaling may be particularly relevant to in vivo sarcomere physiology. Additionally, we suggest that in vivo deficits may have triggered compensatory responses that resulted in increased contractility of ex vivo Alpk3−/− mouse cardiomyocytes. Further investigations of ALPK3 cardiomyocytes in vitro under conditions that simulate in vivo conditions may help to understand the discrepancy between in vitro and in vivo phenotypes.
In summary, our studies identify ALPK3 as a catalytically-inactive pseudokinase that scaffolds myomesin assembly at the M-band and nuclear envelope. We identify a signaling network between the nucleus and the M-band – involving ALPK3, MYOM1, and MYOM2, and perhaps TTN – with roles in the bidirectional sensing, transmission, and response to biophysical forces from the sarcomere to the nucleus. Further studies to investigate the effects of loss of myomesin expression will likely further elucidate the role of this network in cardiomyocyte biology.
Figure 3.7. Model of ALPK3 function in cardiomyocytes and mechanism of ALPK3 loss-of-function in DCM.

(A) In WT cardiomyocytes, ALPK3 localizes to the nuclear envelope and to the sarcomere M-band. ALPK3 functions to scaffold myomesin-1 (MYOM1) at the M-band, and myomesin-2 (MYOM2) at the nuclear envelope. ALPK3-scaffolding of these proteins is important for force buffering during sarcomere contraction, and for mediating the appropriate turnover of thick filament proteins.

(B) In ALPK3−/− cardiomyocytes, MYOM2 at the nuclear envelope and MYOM1 at the M-band are misassembled. This results in altered turnover of several thick filament proteins. The lack of myomesin-mediated force buffering functions, and altered turnover of thick filament proteins, results in impaired sarcomere force transmission.
Methods

Cell lines

**hiPSCs**

PGP1-eGFP-TTN is a male human pluripotent stem cell line containing an endogenous GFP tag on the N-terminus of TTN for visualization of sarcomeres (Sharma, Toepfer, Ward, *et al.*, 2018). Tissue culture dishes (6-well) (Fisher Scientific, 07-200-83) were coated with BD Matrigel (BD Biosciences 354277) at a 1:320 dilution in DMEM F12 (Invitrogen 11330057). hiPSCs were seeded on Matrigel-coated tissue-culture dishes at approximately 30% confluency and media (mTESR1, STEMCELL Technologies) was replaced daily. hiPSCs were passaged every 3-4 days, using 0.5mM EDTA in PBS, so as to not exceed a confluency of 80%. Cell viability after passaging was improved by including 10μM ROCK inhibitor Y-27632 (R&D Systems, 125410) in culture media for 1 day post-passage. Cells were cultured under sterile conditions, and maintained at 37°C, 5% CO₂.

**HEK293T**

HEK293T cells were grown in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal calf serum (FBS) and maintained at 37°C, 5% CO₂.

Animals

Experimental animals were kept in a pathogen-free facility at the New Research Building at Harvard Medical School, fed *ad libitum* on a 14/10 h light/dark cycle and checked daily by veterinary staff. All *in vivo* experiments were carried out in accordance with the Institutional Animal Care and Use Committee (IUCAC) guidelines reviewed and
approved under the IUCAC guidelines at Harvard Medical School. Adult C57BL/6N-Alpk3tm1b(EUCOMM)Hgmu/H were obtained from MRC Harwell and all mice were at least doubly housed. Mice between postnatal day 0 and 35 weeks were used for echocardiograms and mice at postnatal day 0 and postnatal day 8 were sacrificed, and a four-chamber cardiac dissection was performed to obtain left ventricular tissue for subsequent RNaseq and proteomic analysis. C57BL/6N mice were bred in facility and both male and female pups were sacrificed. The sex of all animals was determined either by visual inspection or by Sry PCR for P0 pups. The sex ratio of litters was balanced and both males and females were used for all experiments.

**Multiple sequence alignment of the α-kinase domain and structural models**

Clustal Omega (Sievers et al., 2011) was used for multiple sequence alignment of α-kinase domains from all human α-kinases, and their orthologs in other species. UniProt accession of sequences used for alignment are included in Fig S3. Visualization and analysis of multiple sequence alignment was performed in JalView (Waterhouse et al., 2009).

The crystal structure of the α-kinase domain from human TRPM7 (PDB 1IA9_A) (Yamaguchi et al., 2001) was used for annotation of functionally important residues (Drennan and Ryazanov, 2004) in UCSF Chimera (Pettersen et al., 2004).

**CRISPR/Cas9 genome editing**

All guide RNAs and homology-directed-repair arms were designed using Benchling. For the ALPK3 exon 6 guideRNA, a fragment bearing all components necessary for gRNA expression (U6 promoter + target sequence (AGAAGAATGTGCAGGCAGAT) + guide RNA scaffold + termination signal) was
synthesized as a gBlock from IDT and cloned into the TOPO plasmid using the Zero Blunt Topo II PCR cloning kit (Invitrogen K2800-02). A combination of 4μg of guideRNA TOPO plasmid and 2μg of Cas9 plasmid (pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene 62988) was used for genome editing.

The ALPK3 exon 13 CRISPR RNA was synthesized as a custom ALT-R CRISPR crRNA in IDT. The crRNA (targeting sequence: CTCTCTTTTCAGGGGTTGAC) was annealed with the 5' ATTO 550 trcRNA to create the guideRNA, which was incubated together with the ALT-R S.p HiFi Cas9 3NLS to create the function ribonucleoprotein (RNP) complex per IDT protocol. The ALPK3 exon 13 ssHDR template was synthesized as an Ultramer ssDNA oligo from IDT:

(AGTATATCTTCTGCTGGAGATGTGGGGGCTTAGAGCCAGCCAGACTGGCATCAACTCCCAACTTTCTCTCTTTTCAGAAGTTGACTGGAAGATGACTGCTGTGCAGATTGCTACCAAACTCCGAGGGTGAGTGAGTTCTTCTACCAAACCTCCGAGGGTGAGTGAGTTCTT )

For nucleofection of CRISPR components, adherent hiPSCs were dissociated into single cells using accutase and washed 1X with PBS. 1x10⁶ hiPSCs were resuspended in 82μL of human stem cell nucleofector solution and 18μL of supplement 1 (Lonza). CRISPR components (either guideRNA plasmid + Cas9 plasmid or CRISPR RNP complex + 1uL of 100uM ssHDR template) were added, and the mixture was transferred into an electroporation cuvette, and nucleofected using program B-016 on the Amaxa 2B nucleofector (Lonza). For plasmid-based CRISPR, puromycin selection at a concentration of 0.5μg/mL was continued for 2 days post-nucleofection.

After approximately one week, single rounded colonies were picked under a dissection microscope using sterile pipette tips into a new Matrigel-coated 96-well plate.
in the presence of 5μM ROCK inhibitor Y-27632 (R&D Systems, 125410). Surviving colonies were split and expanded.

Genomic DNA was extracted using PrepGEM (VWR PUN0500) and colonies were screened for the presence of gene editing by PCR amplification followed by Sanger sequencing. Colonies with evidence of gene editing were subcloned to verify clonality. Clonality of populations containing indels were further confirmed by next generation sequencing analysis (MiSeq).

**Differentiation of hiPSC-CMs**

hiPSCs were differentiated into cardiomyocytes in monolayers using a previously described chemically-defined protocol with minimal modification (Lian et al., 2013). hiPSCs seeded in Matrigel-coated 6-well culture plates were grown to 90% confluency prior to initiating the hiPSC-CM differentiation. To initiate differentiation, 18μM CHIR-99021 (Tocris 44-231-0) in RPMI media (Invitrogen) supplemented with B27 minus insulin (Thermo Fisher Scientific) was added to hiPSCs for 1 day (day 0-1). The media was changed to RPMI with B27 minus insulin for two days (day 1-2). Cultures were then treated with 2μM Wnt-C59 (Biorbyt orb181132) in RPMI supplemented with B27 minus insulin for two days (day 3-5). Cells were maintained in RPMI with B27 minus insulin for 2 additional days (day 6-7). On day 7, media was changed to RPMI with B27 plus insulin (Invitrogen) and the media was replenished every other day until day 11, at which point cells underwent a 7-day glucose-deprivation protocol to enrich for cardiomyocytes; for enrichment, cells were cultured in RPMI media devoid of glucose (Life Technologies) supplemented with B27 with insulin, which was replaced every other day until day 17. On day 17, cells were again cultured growth media containing glucose.
(RPMI supplemented with B27 plus insulin) with media replacement every other day. All proteomics and RNAseq experiments were performed with hiPSC-CMs at day 30-31 of differentiation. All other assays (sarcomere contractility, immunofluorescence) were performed with hiPSC-CMs that were between day 30-day 40 of cardiomyocyte differentiation.

**Evaluation of sarcomere contractile function in hiPSC-CMs**

hiPSC-CMs were lifted and replated onto Matrigel-coated glass-bottom imaging-optimized 12-well plates (Mat-Tek) at day 20 of cardiomyocyte differentiation via previously described methods (Sharma, Toepfer, Schmid, et al., 2018). hiPSC-CMs were cultured in RPMI B27 plus insulin containing 5μM ROCK inhibitor and 20% FBS for 2 days post-replating, after which they were cultured in RPMI B27 plus insulin until day 30-32 when video microscopy was performed. For (GFP) video microscopy, hiPSC-CMs were maintained in a heated and humidified chamber (37C, 5% CO2) attached to the stage of a fluorescence microscope (Keyence BZ X-710). Five-second videos of sarcomeres in beating hiPSC-CMs (typically clusters of 2-4 cells each) were acquired using a 100X oil objective at an acquisition rate of 30 frames per second. Analysis of sarcomere contraction was performed using SarcTrack software (Toepfer, Sharma, et al., 2019a).

**HEK293T transfection**

HEK293T cells were transiently transfected with an ALPK3 ORF plasmid encoding the human ALPK3 cDNA sequence with a C-terminal FLAG tag (Origene RC220076). HEK293T cells were grown in 100mm culture dishes to 80% confluency and transfected with 18μg of ALPK3 ORF plasmid using Lipofectamine 3000 (Invitrogen
L3000001) per manufacturer instructions. Expression of ALPK3 ORF plasmid was evaluated with Western Blotting (anti-FLAG) and mass spectrometry.

**Harvest of cell lysate for western blotting and mass spectrometry**

Cells (HEK293T cells/hiPSC-CMs at day 30 of cardiomyocyte differentiation) were washed quickly 1X in cold PBS. Adherent cells were scraped from tissue culture dish (100mm plate for HEK293T cells, all wells of a 6-well culture dish for hiPSC-CMs) and pelleted by spinning at 1000RPM for 5 minutes. Cells were lysed with cold RIPA buffer in the presence of protease inhibitors (Sigma-Aldrich 11873580001) and phosphatase inhibitors (Sigma-Aldrich 4906837001). To increase protein yield and ensure lysis of nuclei, the lysate was subjected to additional mechanical disruption by passage 20 times through a 21G needle. The mixture was shaken gently on ice for 15 minutes, and then centrifuged at 14,000g for 15 minutes to pellet cell debris. The supernatant was transferred to a new tube and protein concentration was quantified using the BCA assay kit (Pierce).

**Western blotting**

Protein lysates were loaded and resolved under denaturing conditions on a NuPAGE 3-8% Tris-Acetate Protein Gel. Lysate was transferred to a PVDF membrane overnight at 40mA at 4°C. 5% skim milk in TBST buffer was used as a blocking agent. Primary antibodies include mouse anti-FLAG 1:1000 (Sigma-Aldrich, F1804), rabbit anti-MYOM1 1:100 (Abcam, ab201228), anti-MYOM2 1:100 (Abcam, ab93915) and mouse anti-Cardiac Troponin T 1:500 (Abcam, ab8295). Primary antibodies were diluted in 5% skim milk and incubated 4°C overnight. Western blotting was performed using standard chemiluminescent western blotting procedure.
**Tandem-Mass-Tag proteomics and phosphoproteomics**

A detailed protocol for streamlined Tandem Mass Tag (SL-TMT) quantitative (phospho)proteome profiling using synchronous precursor selection-MS3 is provided in (Navarrete-perea et al., 2018). Briefly, cells were lysed, after which cysteine bonds were reduced with 5mM tris(2-carboxyethyl)phosphine (TCEP) and alkylated with 10mM iodoacetamide that was quenched with 10mM DTT. A total of 100μg of protein per sample methanol-chloroform precipitated to extract proteins, which were subsequently digested using Lys-C (Wako) overnight followed by digestion with trypsin (ThermoFisherScientific) for 6h. The resulting peptides were labeled with tandem-mass-tag (TMT) isobaric tags. To check mixing ratios, 2μg of each sample was pooled, desalted, and analyzed by mass-spectrometry. Using normalization factors calculated from this “label check”, samples were mixed 1:1 across all channels, and a single desalting step was performed using a 100mg Sep-Pak solid-phase extraction column (Waters). The dried, mixed, and desalted sample was subjected to centrifugation-based phosphopeptide enrichment using the Pierce High-Select Fe-NTA phosphopeptide enrichment kit (ThermoFisher Scientific). Enriched phosphopeptides were desalted for SPS-MS3 analysis. The flow-through unbound fraction and washes from this enrichment were combined, desalted, and fractionated by basic pH reverse-phase HPLC. The resulting fractions were desalted by StageTip and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Mass spectrometric data were collected on an Orbitrap Fusion mass spectrometer in line with a Proxeon NanolC-1200 UHPLC. Database searching and reporter-ion quantification was performed using an in-house SEQUEST-based pipeline. Each protein was scaled such that the summed
signal-to-noise for that protein across all channels was 100, thereby generating a relative abundance measurement.

**RNA sequencing**

hiPSC-CMs and mouse left ventricular tissue were lysed in Trizol Reagent (Life Technologies); mouse tissue was subjected to additional two minutes of mechanical bead homogenization using a TissueLyser (Qiagen); RNA was isolated by chloroform:isopropyl alcohol extraction. RNA quality (RIN) and quantity were assessed on the TapeStation 2200 (Agilent). Two rounds of mRNA purification were performed on total RNA (1 ug) using Dynabeads mRNA DIRECT Kit (Invitrogen). The Superscript III First-Strand Synthesis System (Invitrogen) was used to generate double-stranded cDNA. cDNA libraries were constructed using the Nextera XT DNA Library Preparation Kit (Illumina). Libraries were sequenced on the Illumina NextSeq500 platform. 75 bp paired-end reads were aligned to the human reference genome hg38 using Spliced Transcripts Alignment to a Reference (STAR). Raw reads were normalized to the total number of reads per kilobase of transcript per million (RPKM).

**Mouse echocardiography**

Cardiac function was evaluated by echocardiography on mice anesthetized under isoflurane vaporizer (VetEquip) (newborn pups were not anesthetized). Each limb was placed on the ECG leads on a Vevo Mouse Handling Table (FujiFilm VisualSonics) to maintain body temperature at 37°C during the study. Chest hair was removed with depilatory cream to obtain clear images. Anesthesia was terminated after a mouse was properly positioned for imaging, and all measurements were performed with a heart rate between 300 and 550 beats per minute. Two-dimensional and M-mode images of the
left ventricle (parasternal long axis and short axis) were obtained. Measurements were averaged from M-mode tracings of 3 consecutive heart beats including left ventricular end-diastolic dimension (LVEDD), end-systolic dimension (LVESD), left-ventricular posterior wall thickness (LVPW), interventricular septal thickness (IVS). Left ventricular fractional shortening was calculated by this equation: \( FS = 100 \times \frac{(LVEDD-LVESD)}{LVEDD} \). Echocardiography was blinded to mouse genotype.

**Isolation of murine cardiomyocytes and evaluation of single myocyte contractile function**

Cardiomyocytes were isolated from 7- to 16-week old mice by rapid explantation and aortic cannulation on a Langendorff apparatus for perfusion with enzyme buffer using previously described methods (Toepfer, Wakimoto, et al., 2019).

Isolated cardiomyocytes were placed in wells of a glass bottom 12-well imaging place (MatTek) that had been precoated with laminin. Laminin coating was performed for 2 hours before cardiomyocyte introduction into wells using previously described methods (Toepfer, Wakimoto, et al., 2019). Cells were imaged using a Keyence BZ-X710 microscope using a Nikon 40X/0.65 numerical aperture (NA) objective. Cells were kept at 30°C using microscope specific incubation chamber that was also used to deliver 20% O₂ and 5% CO₂ to the experimental chamber. Cells were paced at 1 Hz using custom-built electrodes hooked up to a pacing unit (Pulsar 6i, FHC) delivering 20V. Movies were acquired at 29 frames/second for 5 seconds (five contractile cycles). An ImageJ plugin, SarCoptiM, was used to track SLs during contractile cycles, and calculate cellular shortening (%) (Pasqualin et al., 2016).
**Immunofluorescence**

hiPSC-CMs were replated onto Matrigel-coated glass-bottom imaging-optimized 12-well plates (Mat-Tek). hiPSC-CMs were cultured in RPMI B27 plus insulin containing 5μM ROCK inhibitor and 20% FBS for 2 days post-replating, after which they were cultured in RPMI B27 plus insulin for 1 week prior to imaging analysis. hiPSC-CMs were fixed with 4% PFA, permeabilized with 0.2% Triton-100, and incubated with primary antibody overnight at 4°C in 3% BSA in PBS (mouse anti-FLAG M2, Sigma-Aldrich F1804 1:1000; rabbit MYOM2, abcam ab93915 1:100; mouse mMαC myomesin B4 0.3μg/mL, Developmental Studies Hybridoma Bank). Secondary antibodies with conjugated fluorophore (1:1000) were incubated for 2 hr at room temperature. A spinning disk confocal microscope (Yokogawa CSU-W1 Spinning Disk on Nikon T1) was used to acquire images. Images were analyzed in Fiji, using the same brightness/contrast scaling for wild-type and mutant images.

**Quantification and Statistical Analysis**

All statistical comparisons and sample sizes are included in the Figures and Figure Legends. Unpaired t-tests were used to determine p-values for all proteomics/phosphoproteomics analysis. DESeq2 (Love, Huber and Anders, 2014) was used to analyze RNA sequencing data and determine differentially expressed RNAs; p-values were adjusted for multiple-testing using the Benjamini-Hochberg method. No samples were excluded. Data were visualized and graphics were generated using the RStudio programming environment and Prism 8 software (GraphPad). Non-linear lines-of-best-fit for mouse echocardiography parameters of cardiac function were fitted using
Prism 8. In experiments using cell lines, “n” indicates independent differentiations into hiPSC-CMs. In experiments using mouse tissue, “n” refers to number of animals.
Contributions

Radhika Agarwal, Christine Seidman, and Jonathan Seidman designed the study. R.A. drafted the manuscript and performed all experiments and computational analyses unless otherwise noted. Hiroko Wakimoto, Christopher Toepfer, and Mingyue Lun assisted with mouse genotyping and echocardiography and mouse cardiomyocyte isolation and contractility analyses. Joao Paulo and Steve Gygi provided resources and technical expertise for proteomic and phosphoproteomic studies. Joshua Gorham, Steve DePalma, and Angela Tai assisted with sequencing of MiSeq and RNA libraries and aligning reads. Qi Zhang provided assistance with cell culture. Arun Sharma provided technical expertise in immunostaining protocols.
Chapter 4: Reflections and Future Directions
Regardless of the cause, in the clinic, heart failure is treated as a singular disease and managed with the same, limited arsenal of pharmacologic agents. Perhaps one of the most valuable lessons from the study of inherited cardiomyopathies is that heart failure is not just one disease. Many roads lead to a broken heart. As such, there should be many ways to fix them.

**Filamin C and Z-discopathies as protein aggregation diseases of the heart**

In chapter 2, I discovered that *FLNC* truncating variants modeling those observed in human DCM induced profound protein accumulation and increased lysosome content in hiPSC-CMs. While partially truncated or mutant missense proteins can adopt altered secondary structures that can induce misfolding and aggregation, I was surprised to find that even *FLNC* nonsense variants, which resulted in half-normal FLNC levels, caused protein accumulation. This finding underscores the importance of the many FLNC protein interactions at the Z-disc, and demonstrates that alterations in the stoichiometry of large protein complexes can lead to protein accumulation and cardiomyocyte toxicity. The importance of Z-disc protein interactions is further illustrated by the observation that damaging variants in other Z-disc proteins including desmin (DES), myotilin (MYOT), Z-band alternatively spliced PDZ motif-containing protein (ZASP), and four and a half LIM domain protein (FHL1), have all been associated with either myofibrillar myopathy or DCM, and many of these mutant proteins induce aggregation (Ruparelia et al., 2016). Furthermore, loss-of-function variants in genes encoding critical chaperone proteins - αβ-crystallin (*CRYAB*), a desmin chaperone, and bcl-2-associated athanogene 3 (*BAG3*), a FLNC chaperone – also cause DCM. As the Z-disc is a structure that is under
constant mechanical strain in actively contracting muscle, accumulating proteins at these sites are particularly susceptible to damage, and chaperone-mediated mechanisms are crucial for mediating their refolding or disposal (Arndt et al., 2010). Thus, FLNC is one member of a whole family of muscle diseases whose pathologic mechanisms relate to protein misfolding and aggregation (‘Z-discopathies’) (Knoll and Buyandelger, 2011).

Protein aggregation can lead to disease via a variety of mechanisms, including loss-of-function effects, dominant-negative effects, and via direct toxicity of protein aggregates that can disrupt critical cellular functions (“gain-of-toxicity effects”) (Valastyan and Lindquist, 2014). The observation that the phenotypes and proteomes of FLNC heterozygote and homozygote null hiPSC-CMs were largely distinct, indicated that the protein aggregation that occurred in the context of FLNC heterozygote truncating variants was directly toxic to cardiomyocyte function via gain-of-toxicity mechanisms that could not be entirely explained by a simple loss-of-function model.

A wide range of human diseases are associated with the accumulation of protein aggregates. Precisely why protein aggregates are toxic to human cells is debated. Recent studies demonstrated that the expression of aggregation-prone proteins that were artificially created and had no evolved biological functions, interfered with nucleocytoplasmic protein and RNA transport in HEK293T cells. This gain-of-toxicity was also observed when mutant huntingtin protein (associated with Huntington disease) and mutant TAR DNA binding protein-43 (associated with amyotrophic lateral sclerosis and frontotemporal dementia) were expressed. Further characterization revealed that protein aggregates sequestered additional proteins containing disordered and low-
complexity sequences, including those of the nuclear import and export machinery (Woerner et al., 2016). Importantly, these studies demonstrate that protein aggregates can serve as a nidus for the sequestration of a diverse range of additional proteins, resulting in unanticipated pathologic effects. It is possible that similar sequestration mechanisms underlie the toxicity of protein aggregate formation in Z-discopathies.

Recently, chemically-controllable fluorescent proteins have been developed that allow the rapid production of small aggregates in living cells on the order of seconds, and also enable the monitoring of dynamic trafficking of these aggregates into lysosomes and other structures (Miyazaki et al., 2016; Janssen et al., 2018). Expressing such constructs in cardiomyocytes, and temporally characterizing their immediate and long-term effects on cardiomyocyte gene expression and function, may reveal additional mechanisms as to how and why aggregates may be toxic to muscle cells.

The process of protein aggregation has been extensively studied in the field of neurodegeneration, as several devastating neurodegenerative diseases, including Alzheimer’s, Parkinson’s, and Huntington disease, have all been associated with protein aggregates. The penetrance of these diseases is age-related, which has led to hypotheses that the capacity of pathways that regulate cellular protein homeostasis declines with age (Eisele et al., 2015). Similar to neurodegenerative diseases, Z-discopathies typically do not manifest until adulthood, suggesting that a similar age-related decline in the efficiency of protein homeostasis pathways may contribute to disease pathology.

Mammalian cells have highly-conserved mechanisms for combating protein misfolding and aggregation. First, constitutively-expressed chaperones can be further
induced in response to the accumulation of unfolded proteins. In the endoplasmic reticulum, this is called the unfolded protein response (UPR), while in the nucleus and cytoplasm, it is called the heat shock response (HSR). If a misfolded protein cannot be properly refolded, other systems, such as the proteasome, autophagy, and ER-associated degradation (ERAD) are activated for protein disposal. Dysfunction in any of these pathways, can lead to, or exacerbate, protein-misfolding pathologies (Valastyan and Lindquist, 2014). It may be that in cardiomyocytes, there exists additional specialized protein degradation machinery that supports the rapid turnover of damaged sarcomere proteins subject to constant mechanical strain. The identification of sarcomere-associated chaperones and ubiquitin ligases, including the muscle-specific MuRF family of ubiquitin ligases, supports this hypothesis (Pizon et al., 2002; Willis et al., 2009; Pagan et al., 2013; Ranek et al., 2017).

Further studies are necessary to uncover which protein homeostasis pathways are activated (or not activated) in FLNC-DCM. However, the discovery that proteins involved in autophagy were dysregulated in FLNC-variant hiPSC-CMs in proportion to the severity of protein accumulation, in addition to the observation that lysosome content was increased in these cells, suggests that autophagy may be a critically-involved process, a finding that is consistent with studies of other Z-discopathies (Mclendon and Robbins, 2019). There are three types of autophagy in mammalian cells: macroautophagy, microautophagy, and chaperone-mediated autophagy, all of which result in protein degradation in the lysosome (Pohl and Dikic, 2019). An additional type, chaperone-mediated-selective autophagy (CASA), is active in muscle cells and results in the targeted degradation of sarcomere proteins which include FLNC (Arndt et al.,
Strategies to increase autophagic flux via modulation of one, or several of these pathways, or to induce the expression of specific muscle chaperones, may be broadly therapeutic for Z-discopathies.

There are several such modulators of protein homeostasis in various stages of preclinical and clinical development, that have been applied to other diseases of protein aggregation. Pharmacologic agents demonstrated to induce autophagy include mTORC1 inhibitors, activators of AMPK signaling, and small-molecules that appear to mimic the effects of caloric restriction – another intervention that induces autophagic pathways and may have clinical utility (Galluzzi et al., 2017). Small molecule activators of the UPR and HSR are also in preclinical development for the treatment of various forms of neurodegenerative disease (Eisele et al., 2015). Recently, pharmacological chaperone therapy (PCT) has emerged as a promising therapeutic approach for certain lysosomal storage disorders. These pharmacologic chaperones are small-molecule ligands that can selectively bind to and stabilize mutant proteins, increasing their cellular levels, and improving their proper trafficking and activity (Parenti, Andria and Valenzano, 2015). As chaperone proteins appear to be critically involved in efficient protein turnover at the sarcomere, identifying such pharmacologic chaperones for critical Z-disc proteins such as FLNC may be broadly therapeutic for the treatment of Z-discopathies. Alternatively, identifying compounds that can induce the expression or activity of known cardiac chaperone proteins, such as the FLNC chaperone BAG3, may also be therapeutic. Further characterizing the mechanisms of protein degradation in muscle cells, and how these pathways are perturbed in Z-discopathies, can further inform these efforts.
In addition to testing available therapies, as described above, small-molecule screening platforms can be applied to potentially identify novel compounds that could reduce pathologic Z-disc protein aggregation and restore muscle function. The observed effects of FLNC-variants in increasing lysosome content in our models could enable the following screening strategy. One could create an hiPSC line carrying an endogenous fluorescent protein tag on a lysosomal membrane component, such as LAMP1. FLNC truncating variants that induce protein aggregation could be introduced in this line, and an increase lysosome content should be measurable by tracking fluorescence intensity. These cells could then be used to screen for small molecules that reduce fluorescent signal intensity, indicating a decrease in lysosome content, signaling amelioration of protein aggregation. While this approach is indirect, there may be other ways to directly monitor protein aggregation. For instance, one could create an hiPSC-line containing an endogenous FLNC-GFP reporter. Given that the C-terminus of FLNC is involved in protein dimerization and mediates the majority of FLNC protein interactions, the fluorescent reporter should be introduced at the N-terminus, so as to minimize the probability of FLNC-GFP misfolding. FLNC truncating variants could then be introduced in this cell that induced aggregation of the FLNC-GFP fusion protein, which should alter GFP signal intensity. One could then screen for small molecules that ameliorate protein aggregation, by monitoring FLNC-GFP signal intensity.

The development of additional in vivo models will be important for testing the efficacy and potential toxicities of various therapeutic strategies. The ease of genetic manipulation, low cost, and short generation time of zebrafish make this model system particularly attractive for candidate drug screening. Of course, the utility of any in vivo
genetic model is contingent on being able to observe a clear phenotype. Given that FLNC-myopathies and other Z-discopathies generally do not manifest until adulthood, it may be necessary to create a double mutant of two different Z-disc proteins – for instance, a truncating variant in FLNC, plus a loss of function variant in the FLNC chaperone, BAG3 – in order to create a penetrant phenotype. Mutant fish could then be screened using a variety of small molecules as discussed above.

**α-kinase-3: nucleus-to-M-band signaling and mechanisms of sarcomere protein turnover**

In chapter 3, I demonstrated that ALPK3 is a catalytically-inactive pseudokinase that scaffolds myomesin assembly and affects thick filament protein turnover. In characterizing the function of ALPK3, I identified an entire network of molecules - involving ALPK3, MYOM1, and MYOM2 - that can localize to both the nucleus and the M-band. Common to this class of molecules, is the presence of multiple immunoglobulin (Ig) domains. The Ig domains of MYOM1 have previously been demonstrated to rapidly unfold and refold, and buffer biomechanical forces generated during sarcomere contraction (Tskhovrebova and Trinick, 2012; Xiao and Grater, 2014). Is it possible that this entire class of nucleus-M-band-dual-localizers is involved in sensing and responding to biophysical forces?

This work provokes several additional questions regarding the structural and functional connections between the nucleus and the sarcomere. (1) How are biomechanical forces that are generated by sarcomeres signaled to mediate changes in gene expression in the nucleus? Conversely, how are signals from the nucleus
transduced to mediate changes in sarcomere protein turnover and/or function? (2) How are nuclei positionally anchored and maintained in the complex, dynamic cardiomyocyte cytoskeleton? How is nuclear membrane integrity protected in an environment of constant mechanical insult? Could ‘shock absorbing’ Ig-domains of MYOM2 fulfill this function? (3) Are nuclei coupled to sarcomeres themselves via specialized complexes, similar to sarcomere coupling to the sarcolemma at costameres? Nucleus-M-band-dual-localizers may be ideal candidates to fulfill one or more of the above functions. Several approaches may be applied to further investigate the roles of these proteins in cardiomyocyte health and disease.

First, CRISPR/Cas9-based methods can be employed to systematically knockout nucleus-M-band-localizers in hiPSC-CMs and study the resultant effects on nuclear membrane integrity and sarcomere assembly. Genetic approaches to abrogate the expression of MYOM1 and MYOM2 in hiPSC-CMs will reveal whether these proteins are required for M-band formation and thick filament assembly into sarcomeres. Mutation of specific domains that are involved in the assembly of MYOM1/MYOM2 into the sarcomere may allow analyses of MYOM1/2 nuclear functions independent of sarcomere functions. Past studies have suggested that all of these proteins are developmentally regulated (Grove et al., 1985; Obermann et al., 1996), suggesting that the expression of these proteins, and their turnover, at key stages of cardiomyocyte development may be critical for cardiomyocyte function. As hiPSC-CMs recapitulate features of human cardiomyocyte development during cardiomyocyte differentiation, hiPSC-CMs provide a unique model to study possible perturbations in cardiomyocyte developmental trajectories and monitor sarcomere assembly (Chopra et al., 2018).
Furthermore, the identification of ALPK3 as a new protein that localizes to both the nuclear and M-band compartments provides the opportunity to identify additional key molecules that participate in this network. In particular, investigating the protein interactions of ALPK3, MYOM1, and MYOM2 will likely reveal additional players that participate in nuclear-M-band signaling. Recent studies have suggested that one major function of the M-band may be to scaffold ubiquitin ligases that can dynamically regulate sarcomere protein turnover in response to biomechanical stress (Lange et al., 2019). Interestingly, the TTN pseudokinase domain at the M-band is important for the recruitment of one such ubiquitin ligase, TRIM63 (also called MuRF2) and ALPK3 active-site variants dysregulated TRIM63 expression, suggesting the possibility that pseudokinase domains may be broadly important for the recruitment and activity of ubiquitin ligases in the sarcomere (Pizon et al., 2002; Lange et al., 2005; Bogomolovas et al., 2014). Thus, further understanding the participation of pseudokinase proteins in scaffolding protein interactions may reveal additional critical mechanisms of sarcomere protein turnover.

There are a large number of methods to screen for protein-protein interactions (PPIs). Co-immunoprecipitation experiments typically only identify high-affinity stable PPIs. As sarcomeres are dynamic structures, it is likely that many proteins interact only transiently with its components only transiently. Employing alternative strategies such as proximity-dependent labeling approaches can capture both stable and transient protein interactions over a period of time. These methods rely on fusing proteins to enzymes that generate reactive molecules such as biotin, which covalently label nearby proteins, enabling their isolation and identification by mass spectrometry (Kim and
Roux, 2016; Branon et al., 2017). It may be especially revealing to introduce such a fusion protein on a nuclear membrane and/or M-band component such as ALPK3, MYOM1, or MYOM2 which may identify proteins that are involved in directly linking these two structures.
A final note:

Cardiomyocytes are terminally-differentiated cells that stop dividing shortly after birth. Yet, these cells, which must constantly beat under biomechanical strain, can survive an entire lifetime. This indicates that cardiomyocytes must possess extraordinary systems to maintain and repair their complex architectures through mechanisms that integrate signals from the extracellular environment, the contractile apparatus, and the nucleus. Understanding how these specialized processes are coordinated will likely reveal fundamental biological insights that have not yet been elucidated by the study of other cell types. Already, basic understandings of how myosins function in contraction and relaxation, how intercalated disc components are trafficked, and how sarcomeres are assembled and repaired, are fueling the discovery of entirely new classes of therapeutic agents. Listening to and learning from patients, and questioning why and how they develop disease presents a window of opportunity for understanding these mechanisms and translating discoveries back to the bedside. Understanding the genetic determinants of cardiac disease in populations across the world will be important for understand the diverse causes of cardiomyopathy and ensuring that advances in genetic diagnostics and therapeutics reach affected communities globally. Today, a diagnosis of heart failure carries with it a dismal prognosis. It is my hope that the fundamental mechanistic understandings inspired by the study of patients and their diseases will change this reality. In the words of Albert Einstein, “learn from yesterday, live for today, hope for tomorrow. The important thing is to not stop questioning.”
References


Bogomolovas, J. et al. (2014) ‘Titin kinase is an inactive pseudokinase scaffold that supports MuRF1 recruitment to the sarcomeric M-line’.


Report From the American Heart Association. doi: 10.1161/CIR.0000000000000659.


doi: 10.1371/journal.pgen.1006880.


Stephan, J. S. et al. (2009) 'The Tor and PKA signaling pathways independently target the Atg1 / Atg13 protein kinase complex to control autophagy'.


Towbin, J. A. (2000) 'Molecular Genetics of Hypertrophic Cardiomyopathy'.


Tu, Y. et al. (2003) '45 Migfilin and Mig-2 link focal adhesions to.pdf', 113, pp. 37–47.


Gamez, A et al. (2018) 'Protein misfolding diseases : Prospects of pharmacological


Zhang, M. et al. (2007) ‘Identification of CAP as a Costameric Protein that Interacts with

Appendix 1: Material related to Chapter 2
Figure S2.1. MiSeq sequence validation of subcloned FLNC mutant hiPSC-CM lines after CRISPR/Cas9-mediated genome-editing. Integrated genome viewer (IGV) traces of next-generation sequencing (MiSeq) of WT and FLNC mutant hiPSC-CM lines. All mutant cell lines were created with a CRISPR/Cas9 guideRNA that targeted FLNC exon 30. A homology-directed repair (HDR) template was used to create the G1674S edit (first G>A change); the HDR template also included a second, silent mutation to disrupt the protoscaler adjacent motif (PAM) site (second G>A change), reducing the likelihood of re-cleavage by Cas9.
Figure S2.2. Assessment of differentiation efficiency in WT and FLNC mutant hiPSC-CMs and identification of batch effects.
(A) RNA sequencing was performed on wild-type hiPSCs at select stages of cardiomyocyte differentiation (iPSC, day 4: cardiac mesoderm, day 8: cardiac progenitor, day 17: early cardiomyocyte, day 30: cardiomyocyte). The resultant RNA expression profiles were analyzed by principal component analysis. The majority of variance (71%) is explained by the first principal component, which separates samples by stage of cardiomyocyte maturation. Wild-type hiPSC-CMs at day 30 and all mutant hiPSC-CM cell lines (FLNC−/−, FLNC+/−, FLNC+/Δ7aa, and FLNCG1674S/G1674S) cluster tightly at day 30.
(B) Marker RNAs specific for the cardiomyocyte lineage are comparably expressed between WT and FLNC mutant hiPSC-CMs at day 30 of cardiomyocyte differentiation. **(Figure S2.2 continued)** Values in table represent reads per kilobase of transcript per million mapped reads (RPKM), shown as mean ± SEM. Number of replicates (n) is indicated in table.

(C) Quantitative proteomics was performed on protein extracts harvested from 5 paired differentiation batches of WT and FLNC−/− hiPSC-CMs. Hierarchical clustering via Ward’s method shows that samples cluster by genotype, except for one batch (batch A), which was excluded from subsequent analyses.

(D) Quantitative proteomics was performed on protein extracts harvested from independent differentiation batches of WT, FLNC<sup>G1674S/G1674S</sup>, FLNC<sup>+/−</sup>, and FLNC<sup>+/Δ7aa</sup> hiPSC-CMs. Hierarchical clustering via Ward’s method shows that samples cluster by genotype without evidence of batch effects.
Figure S2.3. Analysis of sarcomere content in FLNC−/− hiPSC-CMs. hiPSC-CMs were replated onto fibronectin-coated patterned surfaces at day 39 of cardiomyocyte differentiation and the fraction of striated sarcomeres within cells was quantified using a FFT-based MATLAB-algorithm. Sarcomeres in hiPSC-CMs are identifiable due to an endogenous eGFP-TTN tag, marking Z-discs. Each point in boxplots (A-B) represents the fraction of striated sarcomeres identified in a single patterned cardiomyocyte. (A) Fraction of hiPSC-CMs with striated sarcomeres, 2 days post-patterning (WT n = 24, FLNC−/− n = 18 cells, unpaired t-test) (B) Fraction of hiPSC-CMs with striated sarcomeres, 8 days post-patterning (WT n = 23 cells, FLNC−/− n = 30, unpaired t-test) (C) Protein expression of TTN in WT and FLNC−/− hiPSC-CMs, assessed by mass spectrometry. (WT n=4, FLNC−/− n=4, ns = not significantly different) (D) Fluorescence intensity of eGFP-TTN signal per cell in WT and FLNC−/− hiPSC-CMs, assessed via flow cytometry (WT n=3, FLNC−/− n=4, ns = not significantly different).
Figure S2.4. Comparison of differentially expressed proteins in FLNC<-/> and FLNC heterozygote mutant hiPSC-CMs and assessment of lysosomal RNA and protein levels.

(A) Heatmap of differentially expressed (DE) proteins in FLNC<-/>, FLNC<+/->, and FLNC<+/Δ7aa> hiPSC-CMs, scaled by row.

(B) Expression of lysosome proteins, assessed via mass spectrometry, in WT and FLNC mutant hiPSC-CMs. Data are shown as mean
(Figure S2.4 continued)
(C) Expression of lysosomal RNAs, assessed via RNA sequencing in WT and FLNC mutant hiPSC-CMs.
(D) Expression of TFEB assessed via RNA sequencing and mass spectrometry. (Boxplots and barplots are depicted as mean ± SEM; For RNA sequencing: WT n = 3, FLNC_{G1764S/G1674S} n = 1, FLNC^{+/-} n = 2, FLNC^{+/\Delta7aa} n = 2; for mass spectrometry: FLNC^{-/-} n=4, FLNC_{G1674S/G1674S} n=3, FLNC^{+/-} n = 3, FLNC^{+/\Delta7aa} n = 2; p-values calculated from unpaired t-tests).
Figure S2.5. Original Western Blots of FLNC protein expression in WT and FLNC mutant hiPSC-CMs.

(A) Original western blot of hiPSC-CMs using an N-terminal FLNC antibody.

(B) Original western blot of hiPSC-CMs using a C-terminal FLNC antibody. Loading controls: β-tubulin (TUBB) and cardiac troponin T2 (TNNT2). Ladder: HiMark pre-stained protein standard.
Appendix 2: Material related to Chapter 3
Figure S3.1. Multiple sequence alignment of the α-kinase domain amongst human α-kinases and their orthologs. Clustal Omega alignment of all human α-kinase domains (ALPK3, ALPK2, ALPK1, TRPM6, TRPM7, eEF2K), their orthologs in (m.m. = mus musculus [mouse], g.g. = gallus gallus [chicken], x.l. = xenopus laevis [frog] d.r. = danio rerio [zebrafish], and the evolutionary-ancient myosin-heavy-chain kinases from dictyostelium discoideum (d.d, slime mold). Uniprot accession sequences and protein names are indicated. Green bars indicate residues that are invariant in all α-kinase domains except for ALPK3 (ALPK3-variant).
Figure S3.2. Validation of ALPK3 mutations in hiPSCs.

(A) Genotype of ALPK3-null hiPSC-CMs validated with next-generation sequencing (MiSeq). Representative IGV trace is shown depicting 31bp deletion (black) and 1bp insertion (purple).

(B) Genotype of ALPK3-active-site hiPSC-CMs validated by Sanger Sequencing of multiple subclones to confirm population clonality. Representative Sanger trace, and corresponding amino acid changes, are shown.

(C) Principal component analysis of RNAseq from WT and ALPK3-null hiPSCs at various stages of cardiomyocyte differentiation (iPSC, day 4 = cardiac mesoderm, day 8 = early cardiac progenitor, day 17 = early cardiomyocyte, day 30 = cardiomyocyte).

(D) RNA expression of cardiomyocyte marker genes in WT and ALPK3+/− hiPSC-CMs. 

(E) RNA expression of Alpk3 in day 30 WT and Alpk3-null hiPSC-CMs, as determined by RNAseq. Data are depicted as mean±SEM. (unpaired t-test, p = 4.4E-5). (WT n=7; Alpk3+/− n=4)
A

B

C

D

<table>
<thead>
<tr>
<th>Gene</th>
<th>WT (RPKM)</th>
<th>ALPK3&lt;sup&gt;−/−&lt;/sup&gt; (RPKM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTN</td>
<td>18±8</td>
<td>22±5</td>
</tr>
<tr>
<td>PLN</td>
<td>360±172</td>
<td>372±116</td>
</tr>
<tr>
<td>ATP2A2 (SERCA2)</td>
<td>128±44</td>
<td>198±47</td>
</tr>
</tbody>
</table>

E

ALPK3 (RPKM)

WT | ALPK3<sup>−/−</sup>
Figure S3.3 Confirmation of full-length ALPK3 expression in HEK293 transient transfections.

(A) Western blot of protein lysates harvested from ALPK3-FLAG transfected (T) and control (C) HEK293 cells. Blot was probed using an anti-FLAG antibody. ALPK3 MW: 201kDa.

(B) Relative ALPK3 peptide expression in ALPK3-transfected versus control HEK293 cells, quantified by mass-spectrometry. A total of 33 peptides mapped uniquely to ALPK3; the start position of each of these peptides in the ALPK3 protein is indicated on the x-axis. Data depicted as mean±SEM.
Figure S3.4. Enrichment analyses of differentially expressed proteins in ALPK3-null hiPSC-CMs. The 347 differentially expressed proteins in ALPK3-null hiPSC-CMs were analyzed using DAVID enrichment analysis. The top enriched terms are shown in table.
Figure S3.5. Enrichment analyses of differentially expressed RNAs and proteins in $\text{Alpk3}$ homozygote mouse left ventricle at postnatal day 8.

(A) ShinyGO enrichment analysis of the 273 differentially expressed RNAs in $\text{Alpk3}^{-/-}$ P8 mouse left ventricular tissue.

(B) The 64 differentially expressed proteins in $\text{Alpk3}^{-/-}$ P8 mouse left ventricular tissue were analyzed using STRING protein-protein interaction networks functional enrichment analysis. Top enriched terms are shown in the table. The protein-protein interactions (PPI) map is shown on the right. There were significantly more interactions than expected by chance (PPI enrichment $p$-value $< 1.0E-16$).
Figure S3.6. Analysis of myomesin expression and splice isoforms.

(A) Western blot of MYOM2 in hiPSC-CMs. Western blot is consistent with MYOM2 quantification by mass spectrometry, validating MYOM2 antibody specificity.

(B) Representative IGV sashimi plot of MYOM1 splice isoforms in hiPSC-CMs. The shorter splice isoform results in the exclusion of MYOM1 exon 18.

(C) Western blot of MYOM1 expression in left ventricular tissue of P8 mice. Replicates represent protein lysates from different mice. Loading control: TNNT2.